EPFL

Thèse n° 9271

Model order reduction based on functional rational approximants for parametric PDEs with meromorphic structure

Présentée le 17 décembre 2021

Faculté des sciences de base
Calcul scientifique et quantification de l'incertitude - Chaire CADMOS
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Davide PRADOVERA

Acceptée sur proposition du jury

Prof. D. Kressner, président du jury
Prof. F. Nobile, directeur de thèse
Prof. G. Rozza, rapporteur
Dr L. Feng, rapporteuse
Prof. S. Deparis, rapporteur

École polytechnique fédérale de Lausanne

2021

"I'd far rather be happy than right any day."
"And are you?"
"No. That's where it all falls down, of course."

— Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

Abstract

Many engineering fields rely on frequency-domain dynamical systems for the mathematical modeling of physical (electrical/mechanical/etc.) structures. With the growing need for more accurate and reliable results, the computational burden incurred by frequency sweeps has increased too: in many practical cases, a direct frequency-response analysis over a wide range of frequencies is prohibitively expensive. In this respect, model order reduction (MOR) methods are very appealing, as they allow to replace the costly solves of the original problem with a cheap-to-evaluate surrogate model.

In this work, we describe a MOR approach, dubbed "minimal rational interpolation" (MRI), that builds a rational interpolant of the frequency response of the dynamical system. In MRI, we build a surrogate model in a data-driven fashion, starting from only few (very expensive) solves of the original problem at well-chosen frequencies. Notably, we do not need any knowledge of (nor access to) the underlying structure of the original problem, so that MRI can be described as a "non-intrusive" method. We perform a theoretical analysis of MRI, showing that it converges to the exact frequency response in a quasi-optimal way, in an "approximation theory" sense. We also describe how this approach can be complemented by adaptive sampling strategies, which, relying on *a posteriori* error estimators, automatically select the "best" sampling frequencies.

Oftentimes, the underlying problem does not depend on frequency alone, but also on additional parameters, which might represent uncertain features of the physical system or design parameters that have to be optimized. This is the so-called "parametric" case, which is much more complex than the non-parametric one, especially if a modest number of parameters is involved. As a way to tackle the parametric setting, we propose a MOR approach based on marginalization: we use MRI to build local frequency surrogates at different parameter configurations, and then we combine these local surrogates to obtain a global reduced model. Several issues arise when carrying out this "combination" step. In this thesis, we propose a practical algorithm for this, relying on matching the partial fraction expansions of the local surrogates term-by-term.

Several numerical experiments are carried out as a way to showcase the effectiveness of our proposed approaches, both in the non-parametric and parametric settings. Our "case studies" are selected as simplified versions of problems of practical interest. Notably, we include examples of resonant behavior of mechanical structures with uncertain material properties, and of impedance modeling of distributed electrical circuits with a modest number of design parameters.

Keywords: model order reduction, rational interpolation, frequency response of parametric dynamical systems, greedy algorithm.

Résumé

De nombreux domaines de l'ingénierie s'appuient sur les systèmes dynamiques dans le domaine des fréquences pour la modélisation mathématique des structures physiques (électriques/mécaniques/etc.). La nécessité croissante d'obtenir des résultats plus précis et plus fiables s'accompagne d'une augmentation de la charge de calcul liée aux balayages des fréquences : dans de nombreux cas pratiques, une analyse directe de la réponse en fréquence sur une large gamme de fréquences est d'un coût prohibitif. À cet égard, les méthodes de "model order reduction" (MOR) sont très intéressantes, car elles permettent de remplacer les solutions coûteuses du problème original par un modèle reduit qui peut être évalué rapidement.
Dans ce travail, nous décrivons une approche MOR, appelée "minimal rational interpolation" (MRI), qui construit une interpolation rationnelle de la réponse en fréquence du système dynamique. En MRI, nous construisons un modèle reduit "data-driven", à partir de seulement quelques solutions (très coûteuses) du problème original à des fréquences bien choisies. Notamment, nous n'avons pas besoin de connaître (ni d'accéder à) la structure sous-jacente du problème original, de sorte que MRI peut être décrite comme une méthode "non-intrusive". Nous effectuons une analyse théorique de la méthode MRI, montrant qu'elle converge vers la réponse en fréquence exacte d'une manière quasi-optimale, dans le sens de la "théorie des approximations". Nous décrivons également comment cette approche peut être complétée par des stratégies d'échantillonnage adaptatives, qui, en s'appuyant sur des estimateurs *a posteriori* d'erreur, permettent de sélectionner automatiquement les "meilleures" fréquences d'échantillonnage.
Souvent, le problème sous-jacent ne dépend pas seulement de la fréquence, mais aussi de paramètres supplémentaires, qui peuvent représenter des caractéristiques incertaines du système physique ou des paramètres de conception qui doivent être optimisés. C'est ce que l'on appelle le cas "paramétrique", qui est beaucoup plus complexe que le cas non-paramétrique, surtout si un nombre modeste de paramètres est impliqué. Pour aborder le cas paramétrique, nous proposons une approche MOR basée sur la marginalisation : nous utilisons MRI pour construire des modèles reduits locaux en fréquence à différentes configurations de paramètres, puis nous combinons ces modèles locaux pour obtenir un modèle réduit global. Plusieurs difficultés apparaissent lors de cette étape de "combinaison". Dans cette thèse, nous proposons un algorithme pratique pour cette étape, qui repose sur la mise en correspondance terme par terme des décompositions en fractions partielles des modèles reduits locaux.
Plusieurs expériences numériques sont réalisées afin de démontrer l'efficacité de nos approches, à la fois dans le cadre non-paramétrique et paramétrique. Nos "études de cas" sont choisies comme des versions simplifiées de problèmes d'intérêt pratique. Notamment, nous incluons des exemples de comportement résonnant de structures mécaniques avec des propriétés matérielles incertaines, et de modélisation de l'impédance de circuits électriques distribués avec un nombre modeste de paramètres de conception.

Mots clés : model order reduction, interpolation rationnelle, réponse en fréquence des systèmes dynamiques paramétriques, algorithme glouton.

Acknowledgements

This work has been supported by the Swiss National Science Foundation (SNF) through project "Model order reduction based on functional rational approximants for parametric PDEs with meromorphic structure" (project number: 182236, url: http://p3.snf.ch/project-182236).

Contents

List of acronyms

AAA	Adaptive Antoulas–Anderson
DAE	Differential Algebraic Equation
FEM	Finite Element Method
FOM	Full Order Model
LS	Least Squares
MC	Monte Carlo
MOR	Model Order Reduction
MRI	Minimal Rational Interpolation
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
PDF	Probability Density Function
pMOR	parametric Model Order Reduction
POD	Proper Orthogonal Decomposition
QoI	Quantity of Interest
RB	Reduced Basis
ROM	Reduced Order Model
SISO	Single-Input Single-Output
UQ	Uncertainty Quantification
VF	Vector Fitting

1 Introduction

The numerical simulation of dynamical systems in frequency domain is of utmost importance in several engineering fields, among which electronic circuit design, acoustics, resonance modeling and control for large structures, and many others. The computational burden of such simulations has kept increasing in the last decades: on one hand, the problem size has been growing because of the need for higher numerical resolution; on the other hand, the necessity to tune design parameters and model uncertain features has lead researchers to tackle parametric models, possibly with a large number of parameters.

The purpose of *model order reduction* (MOR) in general, and of *parametric MOR* (pMOR) in the specific case of dynamical systems in the presence of parameters, is to alleviate this computational load. The main strategy to reach this goal relies on building a surrogate model (*reduced order model*, ROM), which mimics accurately the original problem, but which can be solved at a much reduced cost. In the last two decades, the field of pMOR has thrived, leading to the development, analysis, and application of a wide collection of surrogate modeling strategies. In general, we can assign each of these methods to one of two main categories:

- **Projection-based pMOR.** The surrogate model is built by restricting the original problem onto a suitable subspace, computed from a set of solutions (most often, snapshots of the system state) of the full problem. This requires access to the operators of the full model, which are not necessarily available in applications, for instance in the case of a black-box solver, or if the system operators never get fully assembled in the solution process. Some subcategories of projective pMOR can be identified depending on whether a global basis (such as POD/Reduced Basis [Bau+11] or multi-parameter multi-moment-matching [BF14; Wei+99]) or a collection of local bases (e.g., manifold interpolation of local bases [AF08] or of reduced system matrices [AF11; LEP09; Pan+10]) are employed.

- **Non-intrusive pMOR.** The surrogate model is constructed by interpolation or regression of a set of solutions (usually, output samples) of the full problem. As long as the system state is not necessary for the application at hand, it is common to work directly with the system output. In this case as well, the methods can be further split into two subgroups, although the boundary between the two is more vague: some approaches set up the surrogate by solving a unique global interpolation problem [GTT18; IA14; LAI11], whereas others build it by first constructing several (rational) models in frequency only, and then combining them over parameter space [FKD11; YFB19a; YFB19b].

Due to its weaker assumptions, in this thesis we consider mostly the latter case. Notably, our final goal is to provide a data-driven non-intrusive pMOR framework that relies as little as possible on the specific dependence of the problem on the parameters. Some of the challenges that we face are the following:

- Taking snapshots of the full model incurs in a high computational cost. As such, we set the objective of taking *as few samples as possible*. More specifically, we wish to devise a MOR approach that achieves the highest possible accuracy, given the available snapshots. Our "non-intrusiveness" constraint certainly does not help us in this endeavor. Indeed, in most state-of-the-art non-intrusive MOR approaches, some snapshots must, in effect, be sacrificed in order to keep the method non-intrusive.
- Still related to the subject of "effective sampling", we wish to design a strategy for *adaptive* sampling. In practice, given a (preliminary) surrogate model, we want to be able to tell if it is sufficiently accurate for the specific application at hand. If this is not the case, we wish to identify which not-yet-sampled parameter can provide the "highest amount of information" when used to update the current surrogate. For both these objectives, our target is the design of *a posteriori* indicators of the "goodness of approximation" of the surrogate. Once again, staying non-intrusive will prove to be a hindrance to us in this task.
- The surrogate modeling of parametric frequency-response problems has intrinsic difficulties, related, for the most part, to the curse of dimension, which makes most operations increase considerably in cost when a modest number of parameters is considered. Oftentimes, this issue manifest itself both in the training of the surrogate model and in its evaluation (once available), thus reducing the usefulness of MOR. In this context, our target is to reduce the training and evaluation costs as much as possible. On the training side, this requires the design of effective strategies for high-dimensional sampling and for the assembly of the surrogate. On the evaluation side, we make an effort to reduce the computational cost by keeping the surrogate as "explicit" as possible, thus lowering the number of operations required for its evaluation.

1.1 Example: analysis of resonances under uncertainty in structural mechanics

We proceed by providing a simple prototype of a parametric problem for which we want to build a surrogate model. Since this test will be considered in depth in Section 7.2, here we skip most mathematical details and keep only the most significant aspects. We wish to mention that, for the sake of reproducibility, the code used to obtain our results (here and throughout this thesis) is made publicly available in [Pra21].

Consider the tuning fork depicted in Figure 1.1, subject to a time-harmonic pressure pulse, which is applied near the top of the fork. If the fork is made of a linearly elastic material, and if the applied pressure pulse is weak enough, the fork will undergo a time-harmonic deformation, synchronized with the pulse. As the frequency of the pulse changes, the induced deformation varies too. In particular, we will observe an amplified effect if the frequency is close to a resonating frequency of the tuning fork. Such resonating frequencies depend on the properties of the fork, so that, e.g., changing the density of the material of which the fork is composed will change the resonating frequencies. This can be observed in Figure 1.1, where we show the deformations resulting from 9 different combinations of frequency and material density. In concrete terms, we

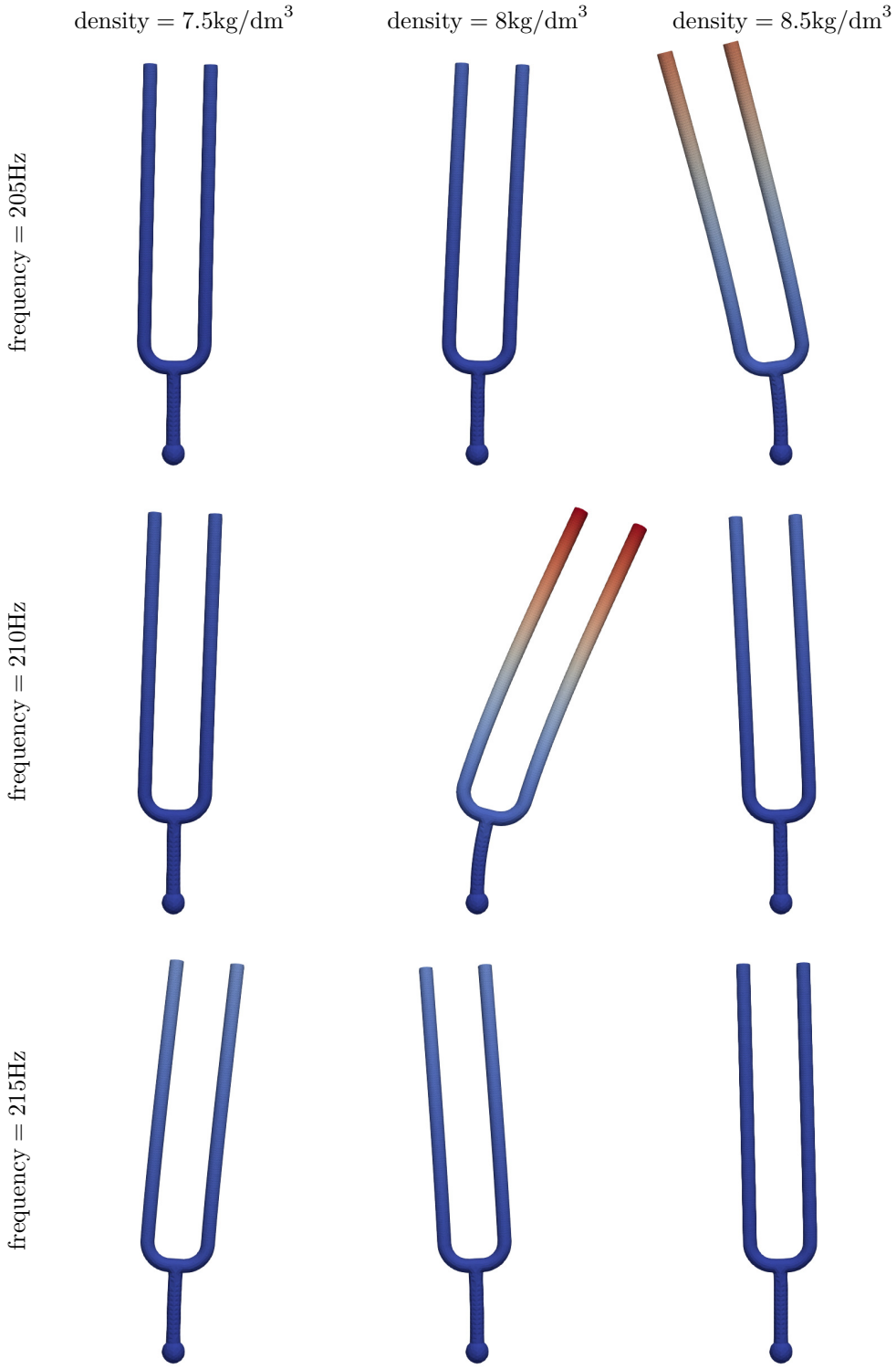


Figure 1.1 – Snapshots (at time $t = 0$) of the deformation induced for different values of the frequency (see on the left) and of the density (see on top).

say that the *response* of the tuning fork depends on both frequency and material parameters.

In practice, one is sometimes interested not in the whole deformation field, but only in a quantity of interest (QoI), which, in effect, is a (scalar) function of frequency and parameters. For instance, as QoI, we could take the maximal deformation that a point of the tuning fork undergoes. Evaluating the QoI requires *first* finding the deformation field and *then* extracting the maximal deformation from it. In MOR, we replace the computation of the "true" deformation field (which requires a considerable computational effort) with the (much cheaper) evaluation of a surrogate deformation field. Provided our surrogate is accurate enough for our purposes, this results in a considerable time save.

In many cases, the evaluation of the (surrogate) QoI is needed in a so-called *outer-loop application*, e.g., in optimal design, where we wish to identify the material properties that yield the best value of the QoI, or in uncertainty quantification (UQ), where the material properties are random variables, and we wish to quantify statistical features of the QoI. In Section 7.2 we will consider a specific instance of this latter case, where we estimate the distribution of QoI based on that of the material properties. Specifically, we use Monte Carlo (MC) sampling for probability density estimation, comparing the results of sampling the "true" QoI and the surrogate one.

1.2 Thesis outline

This thesis is organized as follows.

In Chapter 2 we recall some results from the literature in several areas: polynomial and rational approximation theory for scalar functions of a complex variable, dynamical systems in frequency domain and their infinite-dimensional extensions, and the most popular MOR approaches for their approximation (in absence of parameters).

In Chapter 3 we introduce a novel approach for rational approximation of vector-valued functions of a complex variable, and we develop a convergence theory for it, showing *maximal* convergence of the approximation for a certain class of target functions. This is our proposed approach for non-intrusive approximation of non-parametric dynamical systems. For convenience, we include the proofs of our theoretical results in a separate chapter, namely, Chapter 4.

In Chapter 5 we discuss some of the most relevant issues that arise when applying our proposed technique in practical applications on a computer. We also describe some useful extensions of our method. Among them, we provide a strategy for adaptive frequency sampling that is quite practical in many applications. To conclude this section, we also include some numerical experiments as proof of the good properties of our approach.

In Chapter 6 we move to parametric dynamical systems. First, we outline the main difficulties of generalizing MOR methods from the non-parametric to the parametric setting, and we summarize the state-of-the-art ways in which one usually tries to overcome them. Then, we propose a strategy to extend our non-parametric approach to the parametric framework, relying on a "marginalization" idea and on a specific expansion of the frequency-dependent rational surrogate. Some ideas for adaptive parameter sampling based on locally adaptive sparse grids are also described.

In Chapter 7 we carry out several detailed numerical experiments involving parametric dynamical systems in different settings: vibrations of a membrane, structural mechanics, and circuit modeling.

These simulations allow us to verify the quality of the approximation and the computational efficiency of the surrogate model obtained with our proposed approach.

In Chapter 8 we conclude the thesis. We also include an outline of some possible future research directions in the field.

2 Preliminaries

In this chapter, we introduce some notation and preliminary results that will be useful in the following sections. In order not to clutter the discussion with too many references, we open each section with a list of general sources concerning the covered material.

2.1 Polynomial and rational interpolation (in 1D)

Sources: [AKT14; BGM96; BLM18; Cla76; GPT11; GMSV84; HKD12; Kle12; QSS07; SV13; Saf72; Sta97; Sta98; Tre13; Wal29; Wal60; Wal79]

The topic of approximation of data by polynomial and rational functions is way too vast to be properly reviewed in an introductory section like this one. As such, we limit our discussion to the fundamental building blocks on which we will rely in the next sections. We present our results assuming the data to come from a univariate $\mathcal{V}$-valued function, with $\mathcal{V}$ a general complex-valued Banach space with norm $\|\cdot\|_{\mathcal{V}}$. When reviewing rational approximation, we will only consider the case $\mathcal{V} \subseteq \mathbb{C}$. Indeed, as we will show in the next chapters, one of the main aims of this thesis is to extend rational approximation to high-(or even ∞-)dimensional spaces $\mathcal{V}$.

In our review of the already available results concerning polynomial and rational approximation, we will focus on the cases of holomorphic and meromorphic functions only, respectively. Notably, we mostly ignore the much more complicated case of (global or local) lower regularity. The main motivation supporting this choice is that the functions that we wish to approximate are usually holomorphic or meromorphic (at least when restricted to bounded domains), cf. Sections 2.2 and 2.3.

Moreover, we note that, for the most part, we concentrate on the "interpolation" problem, where values of the target function (and of its derivatives) must be exactly recovered by the approximant. This is a special case of the more general "approximation" problem, where the approximant is not necessarily required to interpolate the data.

2.1.1 Polynomial interpolation (of scalar functions in 1D)

With the term "polynomial interpolation", we mean the task of finding a polynomial of suitable degree that recovers the values of the target function, and, possibly, of its derivatives, at a set of

sample points. Our first result is one of existence and uniqueness, which we state as a definition for convenience.

Definition 2.1 (Lagrange-Hermite interpolation)**.** *Take sample points $Z = \{z_j\}_{j=1}^{S} \subset \mathbb{C}$ (not necessarily distinct). We denote the distinct elements of Z by $\{\mathring{z}_1, \dots, \mathring{z}_{S'}\}$, with $\mathring{z}_j$ appearing $E_j + 1$ times in Z, so that $S = \sum_{j=1}^{S'} E_j + S'$. Let $v : \mathbb{C} \to \mathcal{V}$ be defined at all the sample points, and admit at least E_j (complex) derivatives at each $\mathring{z}_j$. The Lagrange(-Hermite) interpolant of v at Z is* the *polynomial*

$$I^Z(v) \in \mathbb{P}_{S-1}(\mathbb{C}; \mathcal{V}) = \left\{ \sum_{n=0}^{S-1} p_n z^n : \{p_n\}_{n=0}^{S-1} \subset \mathcal{V} \right\}$$

such that

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} I^Z(v)\bigg|_{\mathring{z}_j} \stackrel{!}{=} \frac{\mathrm{d}^n}{\mathrm{d}z^n} v\bigg|_{\mathring{z}_j} \qquad \forall j = 1, \dots, S',\ \forall n = 0, \dots, E_j \tag{2.1}$$

(with the symbol $\stackrel{!}{=}$ we denote a constraint that must be enforced, as opposed to an identity).

Note that $I^{\{z_0,\dots,z_0\}}(v)$, with z_0 repeated $E+1$ times, is the Taylor polynomial of degree E of v at z_0, whereas $I^Z(v)$, with Z having distinct elements, is the Lagrangian interpolant of v at Z, which may be expressed in the extremely useful barycentric form:

$$I^Z(v)(z) = \omega^Z(z) \sum_{j=1}^{S} \frac{v(z_j)}{(z - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \quad \text{with} \quad \omega^Z(z) = \prod_{j=1}^{S} (z - z_j). \tag{2.2}$$

We mention that a generalized barycentric form exists also for repeated sample points, although it is a bit heavy in notation, see (5.4).

In order to understand how well functions can be approximated by polynomials, we must first introduce some tools from complex analysis.

Definition 2.2 (Logarithmic capacity)**.** *Let $A \Subset \mathbb{C}$ be compact, containing infinitely many points. We define its* logarithmic capacity *(or* transfinite diameter*) as*

$$\mathrm{Cap}(A) = \lim_{n\to\infty} \inf_{P \in \mathbb{P}_n^{\mathrm{I}}(\mathbb{C};\mathbb{C})} \sup_{z \in A} |P(z)|^{1/n} \tag{2.3}$$

with $\mathbb{P}_n^{\mathrm{I}}(\mathbb{C};\mathbb{C})$ the set of monic polynomials of exact degree n. If $A \Subset \mathbb{C}$ contains only finitely many points, its logarithmic capacity is 0.

The following bounds hold:

$$\left(\frac{\ell(A)}{\pi}\right)^{1/2} \leq \mathrm{Cap}(A) \leq \frac{1}{2} \max_{z,z' \in A} |z - z'|\,, \tag{2.4}$$

with ℓ the 2D Lebesgue measure. For further use, we mention that the capacity of a disk is its radius, and both bounds in (2.4) are sharp. Note that zero-measure sets may have non-zero capacity: the capacity of a line segment of length L is $L/4$.

Now that we know how to measure sets appropriately, we define a way to "expand" them.

Definition 2.3 (Conformal extension via Green's potential)**.** *Let $A \Subset \mathbb{C}$ be compact, and assume that $A^C = \mathbb{C} \setminus A$ is connected. Identifying $\mathbb{C}$ with $\mathbb{R}^2$ by $x + \mathrm{i}y \leftrightarrow (x, y)$, assume that A^C admits a*

Green's function $G(x, y)$ with pole at ∞, i.e., G is real-valued and harmonic over A^C, continuous over $A^C \cup \partial A$, $G|_{\partial A} = 0$, and $G(x, y) \sim \frac{1}{2}\log(x^2 + y^2)$ as $|x + \mathrm{i}y| \to \infty$. We define the Green's potential *of A, $\Phi_A : \mathbb{C} \to \mathbb{R}$, as*

$$\Phi_A(x + \mathrm{i}y) = \begin{cases} \mathrm{Cap}(A)\exp(G(x, y)) & \text{if } x + \mathrm{i}y \in A^C, \\ \mathrm{Cap}(A) & \text{if } x + \mathrm{i}y \in A. \end{cases} \tag{2.5}$$

Given $\rho \geq \mathrm{Cap}(A)$, we define a conformal extension A_ρ of A as the sub-level set $\{z \in \mathbb{C} : \Phi_A(z) \leq \rho\}$. By construction, it turns out that $\mathrm{Cap}(A_\rho) = \rho$.

The important case of A being a line segment unfortunately is not covered by this definition, since its complement does not admit a proper Green's function. Still, segments *can* be conformally extended, although this requires a slight generalization of Definition 2.2. For simplicity, we provide an *ad hoc* definition just for segments.

Definition 2.4 (Green's potential of line segments)**.** *The Green's potential of a line segment $[a, b]$, with $a, b \in \mathbb{C}$, is defined as*

$$\Phi_{[a,b]}(z) = \frac{|b - a|}{2}\Phi_{[-1,1]}\left(\frac{z - \frac{a+b}{2}}{\frac{b-a}{2}}\right) \quad \text{with } \Phi_{[-1,1]}(z) = \frac{1}{2}\max\left|z + \sqrt{z^2 - 1}\right|, \tag{2.6}$$

where the maximum is taken over the two realizations of the complex square root.

The Green's potential is continuous over $\mathbb{C}$ and increases "as we move away from A". As could be guessed, disks get conformally extended to disks with larger radii, since the potential of a disk centered at z_0 equals $|z - z_0|$ for z outside the disk. Conformal extensions of segments are more interesting: given $A = [-1, 1]$, its conformal extensions are

$$A_\rho = \left\{\frac{z + z^{-1}}{2} : z \in \mathbb{C}, |z| \leq 2\rho\right\} \quad \text{for } \rho > \frac{1}{2}, \tag{2.7}$$

i.e., Bernstein ellipses with horizontal and vertical axes of lengths $\rho \pm \frac{1}{\rho}$, respectively.

The main point of these definitions is that, when interpolating samples of v taken over A, the interpolants can be shown to converge to v over conformal extensions of A. We make this statement rigorous in the following theorem.

Theorem 2.1 (Maximal convergence [Wal60, Section 4.7])**.**
Let $A \Subset \mathbb{C}$ admit a Green's potential, and assume v to be analytic over A. Let $\bar{\rho} > \mathrm{Cap}(A)$ (possibly infinite) be the largest number such that v is analytic over the interior of $A_{\bar{\rho}}$. For all $\mathrm{Cap}(A) < \rho < \bar{\rho}$, there exist a constant C_ρ and a sequence of polynomials $\{P_n\}_{n=1}^\infty$ of increasing degree (with $\deg(P_n) \leq n$) such that

$$\|v(z) - P_n(z)\|_{\mathcal{V}} \leq C_\rho\left(\frac{\mathrm{Cap}(A)}{\rho}\right)^n \qquad \forall z \in A. \tag{2.8}$$

These polynomial approximations also satisfy

$$\|v(z) - P_n(z)\|_{\mathcal{V}} \leq C_\rho\left(\frac{r}{\rho}\right)^n \qquad \forall z \in A_r \quad \forall\, \mathrm{Cap}(A) < r < \rho. \tag{2.9}$$

On the other hand, for all $\rho > \bar{\rho}$, there do not exist a constant C_ρ and a sequence of polynomials $\{P_n\}_{n=1}^\infty$ of increasing degree (with $\deg(P_n) \leq n$) for which (2.8) holds.

The theorem above provides a "barrier" not only for the region of good approximation by polynomials, but also for the convergence rate that can be achieved. Moreover, (2.8) suggests that we should expect the approximation error to behave somewhat uniformly over A, whereas, in the annulus between ∂A and $\partial A_{\bar{\rho}}$, the error progressively increases according to the level curves of the Green's potential, see (2.9).

Note that Theorem 2.1 does not hold for all sets A, since the concept of maximal convergence cannot be extended to sets that are "too rough". However, generalizations to some cases of practical interest, e.g., a single point $A = \{z_0\}$ or finite unions of points $A = \{z_j\}_{j=1}^{S}$, can be obtained by replacing Green's potential with lemniscates, cf. [Wal60, Sections 3.3–3.5]. We skip the details here.

A final question that remains unanswered is how to find a sequence of maximally converging polynomials. The following theorem provides a sufficient condition for this, showing that *interpolatory polynomials can converge maximally.*

Theorem 2.2 (Maximal convergence by interpolants [Wal60, Section 7.2])**.**
Let $A \Subset \mathbb{C}$ admit a Green's potential Φ_A. Define a sequence of sampling sets $\{Z_S\}_{S=1}^{\infty}$, with $Z_S = \{z_j^{(S)}\}_{j=1}^{S} \subset \mathbb{C}$. Assume that $\{Z_S\}_{S=1}^{\infty}$ has no limit points outside A, and that

$$\lim_{S\to\infty} \prod_{j=1}^{S} \left| z - z_j^{(S)} \right|^{1/S} = \Phi_A(z) \quad \textit{uniformly in } z \textit{ over compact subsets of } \mathbb{C} \setminus A. \tag{2.10}$$

Then, the sequence of maximally converging polynomials in Theorem 2.1 may be defined as $P_{S-1} = I^{Z_S}(v)$ for all S.

Sampling schemes satisfying (2.10) are commonly referred to as *Fekete* (or *Fejér*) points, which are quite hard to find for general sets A. Luckily, they are available in closed form for some cases of practical importance:

- for the unit disk, they are the roots of unity: $Z_S = \left\{\exp\left(2\pi \frac{j}{S}\mathrm{i}\right)\right\}_{j=1}^{S}$;
- for the segment $[-1, 1]$, they are the Chebyshev nodes $Z_S = \left\{\cos\left(\frac{2j-1}{2S}\pi\right)\right\}_{j=1}^{S}$.

By rotations, dilations, and translations, the Fekete points for general disks and segments are readily found. Before proceeding, we note that (2.10) is an asymptotic definition, so that the Fekete points are not at all unique. For instance, the Clenshaw-Curtis nodes $Z_S = \left\{\cos\left(\frac{j-1}{S-1}\pi\right)\right\}_{j=1}^{S}$ are also Fekete points for the interval $[-1, 1]$.

Remark 2.1. *The Fekete points of $A \Subset \mathbb{C}$, if they exist, can be chosen so that they lie only on ∂A. More generally, Fekete points accumulate at all points of ∂A, while they* cannot *have any limit point in $\mathbb{C} \setminus \partial A$. This is consistent with the "electrostatic potential" interpretation of the sampling points, where each sampling point is an electron, free to move within the perfect conductor A. If the electrons are allowed to move due to electric repulsion forces until the total energy of the system is minimized (i.e., they are at rest in the "electrostatic ground state"), then the points corresponding to their rest positions are Fekete.*

2.1.2 Rational interpolation (of scalar functions in 1D)

With the term "rational interpolation", we mean the task of finding a rational function (i.e., the ratio of two polynomials) of suitable type (see definition below) that recovers the values of the target function, and, possibly, of its derivatives, at a set of sample points. Like for polynomial approximation, our first result is one of existence and uniqueness.

Definition 2.5 (Lagrange-Hermite rational interpolation)**.** *Take sample points $Z = \{z_j\}_{j=1}^{S} \subset \mathbb{C}$ (not necessarily distinct). We denote the distinct elements of Z by $\{\mathring{z}_1, \dots, \mathring{z}_{S'}\}$, with $\mathring{z}_j$ appearing $E_j + 1$ times in Z. Let $v : \mathbb{C} \to \mathbb{C}$ be defined at all the sample points, and admit at least E_j (complex) derivatives at each $\mathring{z}_j$. Also, let $N \in \{0, \dots, S-1\}$, and define $M = S - 1 - N$. The type $[M/N]$ Lagrange(-Hermite) rational interpolant of v at Z is* the *rational function*

$$R^Z_{[M/N]} = \frac{P^Z_{[M/N]}}{Q^Z_{[M/N]}}, \quad \text{with } P^Z_{[M/N]} \in \mathbb{P}_M(\mathbb{C};\mathbb{C}),\ Q^Z_{[M/N]} \in \mathbb{P}_N(\mathbb{C};\mathbb{C}),$$

such that $Q^Z_{[M/N]} \not\equiv 0$ and

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} P^Z_{[M/N]}\bigg|_{\mathring{z}_j} \overset{!}{=} \frac{\mathrm{d}^n}{\mathrm{d}z^n}\left(Q^Z_{[M/N]} v\right)\bigg|_{\mathring{z}_j} \qquad \forall j = 1, \dots, S',\ \forall n = 0, \dots, E_j. \tag{2.11}$$

(To avoid an overly heavy notation for $R^Z_{[M/N]}$, we do not explicitly indicate the dependence on the target function v.)

In this thesis, we call *type* of a rational function P/Q any 2-tuple $[M/N]$ such that $\deg(P) \le M$ and $\deg(Q) \le N$. We will never require the concept of *exact type* $[M'/N']$ such that $\deg(P) = M'$ and $\deg(Q) = N'$.

The S *linearized order conditions* (2.11) aim at interpolation of v and its derivatives. We note that the extreme case of all sample points coalescing in one, i.e., $Z = \{z_0, \dots, z_0\}$, is commonly known under the name of *Padé approximation*, which generalizes the Taylor polynomial.

We remark that, while $R^Z_{[M/N]}$ is unique (when simplified to lowest terms), $P^Z_{[M/N]}$ and $Q^Z_{[M/N]}$ might not be. For instance, let $v(z) = 1 + z^2$, $Z = \{0, 0, 0\}$, and $[M/N] = [1/1]$. Then, a simple calculation shows that

$$R^Z_{[M/N]}(z) = \frac{\alpha z}{\alpha z} = 1 \qquad \forall \alpha \in \mathbb{C}.$$

The same example also shows that the linearized order conditions do not necessarily imply the (non-linearized) order conditions

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} R^Z_{[M/N]}\bigg|_{\mathring{z}_j} \overset{!}{=} \frac{\mathrm{d}^n}{\mathrm{d}z^n} v\bigg|_{\mathring{z}_j} \qquad \forall j = 1, \dots, S', \forall n = 0, \dots, E_j, \tag{2.12}$$

since $0 = \frac{\mathrm{d}^2}{\mathrm{d}z^2} R^Z_{[M/N]} \neq \frac{\mathrm{d}^2}{\mathrm{d}z^2} v = 2$ in our example.

In general, one might wonder if a rational approximant of type $[S - 1 - N/N]$ could be designed (using a different definition) so that (2.12) is satisfied, for all $N \in \{1, \dots, S-2\}$ (the cases $N = 0$ and $N = S - 1$, by reciprocity, are covered by Definition 2.1). The general answer is no, as our simple example above shows. However, if the S sample points are distinct, then (2.11) and (2.12) are equivalent: a *Lagrange* (i.e., without derivatives) rational interpolant of given type $[S - 1 - N/N]$ always exists.

Due to non-uniqueness and non-existence (when considering the non-linearized problem) issues, a convergence analysis of rational approximants is trickier than that of polynomial approximants. However, it turns out that, as long as the denominator degree stays fixed, the convergence analysis proceeds without snags. First, we mention a generalization of Theorem 2.1.

Theorem 2.3 (Maximal convergence [Wal60, Appendix 4])**.**
Let $A \Subset \mathbb{C}$ admit a Green's potential, and assume $v : \mathbb{C} \to \mathcal{V}$ to be analytic over A. Fix $N \geq 0$ and define $\bar{\rho} > \mathrm{Cap}(A)$ (possibly infinite) as the largest number such that v is meromorphic over the interior of $A_{\bar{\rho}}$, with exactly N poles counting multiplicity, i.e., there exist poles $\Lambda = \{\lambda_j\}_{j=1}^N \subset A_{\bar{\rho}} \setminus A$ and v_H holomorphic over $A_{\bar{\rho}}$, such that

$$v(z) = \frac{v_H(z)}{\prod_{j=1}^N (z - \lambda_j)} \qquad \forall z \in A_{\bar{\rho}} \setminus \{\lambda_j\}_{j=1}^N.$$

For all $\mathrm{Cap}(A) < \rho < \bar{\rho}$, there exist a constant C_ρ and a sequence of rational functions $\{R_M\}_{M=1}^\infty$ of increasing numerator degree ($\mathrm{type}(R_M) = [M/N]$) such that

$$\|v(z) - R_M(z)\|_{\mathcal{V}} \leq \frac{C_\rho}{\prod_{j=1}^N |z - \lambda_j|} \left(\frac{\mathrm{Cap}(A)}{\rho}\right)^M \qquad \forall z \in A. \tag{2.13}$$

These rational approximations also satisfy

$$\|v(z) - R_M(z)\|_{\mathcal{V}} \leq \frac{C_\rho}{\prod_{j=1}^N |z - \lambda_j|} \left(\frac{r}{\rho}\right)^M \qquad \forall z \in A_r \setminus \Lambda \quad \forall\, \mathrm{Cap}(A) < r < \rho. \tag{2.14}$$

On the other hand, for all $\rho > \bar{\rho}$, there do not exist a constant C_ρ and a sequence of rational functions $\{R_M\}_{M=1}^\infty$ of increasing numerator degree ($\mathrm{type}(R_M) = [M/N]$) for which (2.13) holds.

Requiring v to be holomorphic over the sampling region A is quite restrictive. However, since its main purpose is avoiding sampling at the poles of v, such condition can be weakened to: v is meromorphic over A and its poles lie at strictly positive distance from ∂A, cf. [Wal79, Theorem 1]. Note that, under this relaxed assumption, A must be replaced by $A \setminus \Lambda$ in (2.13).

Moreover, we note that (2.14) concerns the convergence of the approximation error over a region where poles are present. This can be done because the poles of the maximally converging rational functions are assigned so as to cancel the N poles of v. This choice is *ad hoc*, and cannot be applied in practice without having *a priori* knowledge of the locations of the poles of v. Assuming that the poles of v are available in advance defeats the purpose of rational approximation, since the problem effectively simplifies to a polynomial approximation one. As such, we must weaken the convergence bound (2.14) to allow for inexact pole approximation, as stated by the following result.

Theorem 2.4 (Quasi-maximal convergence by interpolants [Saf72, Theorem 2])**.**
Let $A \Subset \mathbb{C}$ admit a Green's potential Φ_A and take its Fekete points $\{Z_S\}_{S=1}^\infty$, cf. Theorem 2.2. Let the scalar-valued function $v : \mathbb{C} \to \mathbb{C}$ satisfy the hypotheses of Theorem 2.3. The sequence of maximally converging rational functions in Theorem 2.3 may be defined as $R_M = R_{[M/N]}^{Z_{M+N+1}}$ (the Lagrange rational interpolant of v, see Definition 2.5) for all M, provided bound (2.14) is replaced by the weaker condition

$$\|v(z) - R_M(z)\|_{\mathcal{V}} \leq C_\rho(\mathring{A}_r) \left(\frac{r}{\rho}\right)^M \qquad \forall z \in \mathring{A}_r \quad \forall \mathring{A}_r \Subset A_r \setminus \Lambda \quad \forall\, \mathrm{Cap}(A) < r < \rho. \tag{2.15}$$

Note that C_ρ depends on $\mathring{A}_r$, and implicitly contains information on the exact poles Λ. The poles of R_M converge to Λ.

As a final step, we consider the rather complicated case of N not being fixed. In most applications, the number of poles of the target function v is unknown, and it makes practical sense to increase N together with S. This gives rise to diagonal approximants $[N/N]$, with $S = 2N + 1$, and sub-/super-diagonal approximants $[N+\delta/N]$, with $S = 2N+\delta+1$. These cases are, unfortunately, affected by the appearance of so-called *spurious poles*, namely, numerical roots of $Q^Z_{[M/N]}$ that do not approximate any pole of v. In some situations, the pole is accompanied by a zero (i.e., a root of $P^Z_{[M/N]}$), so that zero and pole cancel each other (or, in finite precision, they almost cancel each other). This zero-pole pair is commonly called *Froissart doublet*.

These spurious effects prevent uniform convergence of the approximant. However, a weaker notion of convergence can be proven instead.

Theorem 2.5 (Convergence in capacity [Wal79, Theorem 2])**.**
Let $A \Subset \mathbb{C}$ admit a Green's potential and take its Fekete points $\{Z_S\}_{S=1}^{\infty}$, cf. Theorem 2.2. Let the scalar-valued function $v : \mathbb{C} \to \mathbb{C}$ satisfy the hypotheses of Theorem 2.3. Note, in particular, that we are fixing a priori *(but arbitrarily large) N and $\bar{\rho}$. Let $A' \Subset A_{\bar{\rho}}$ be compact and $\varepsilon > 0$. Then*

$$\lim_{N'\to\infty} \mathrm{Cap}\left\{ z \in A' : \left\| v(z) - R^{Z_{2N'+1}}_{[N'/N']}(z) \right\|_{\mathcal{V}} > \varepsilon^{N'} \right\} = 0. \tag{2.16}$$

with Cap *the logarithmic capacity, see Definition 2.2, and $R^{Z_{2N'+1}}_{[N'/N']}$ the Lagrange rational interpolant of v, see Definition 2.5.*

This result in capacity allows spurious poles to appear, but constrains their locations to a set of asymptotically zero capacity. Also, we note that Theorem 2.5 is stated for diagonal approximants. Generalizations to other types $[M_i/N_i]$ are possible, as long as M_i and N_i diverge when $i \to \infty$.

As already mentioned in passing, we have restricted our presentation on rational approximation to the scalar-valued case. Generalizations of rational approximants to $\mathcal{V}$-valued target functions (with $\mathcal{V}$ a Banach space of dimension > 1) are possible, but not at all straightforward. To start with, the numerator $P^Z_{[M/N]}$ must become a $\mathcal{V}$-valued polynomial, but the denominator $Q^Z_{[M/N]}$ must, obviously, remain $\mathbb{C}$-valued. This causes an imbalance in the amount of degrees of freedom that pertains to numerator and denominator. As a consequence, the linearized order conditions (2.11) are necessarily overdetermined. As a way to solve this issue, one customarily casts the linearized order conditions in least squares (LS) form, so that interpolation of values and derivatives is no longer guaranteed. As a notable consequence of this, neither the quasi-maximal convergence nor the convergence in capacity of rational interpolants can be directly extended to the higher-dimensional setting, since the involved "interpolants" do not exist. In Chapter 3, we will address this problem, by proposing a technique for maximally converging rational interpolation in the non-scalar-valued case, under some restrictions on the approximated quantity.

2.2 Dynamical systems in frequency domain

Sources: [Ant05; GTG15; Rug96; Ske88; Son98]

Linear time invariant (LTI) systems are ubiquitous in engineering applications, where they are used to model the evolution of an (electrical, mechanical, chemical, etc.) system, usually as a

result of external inputs, e.g., forces/displacements or voltages/currents. Such systems are usually expressed in the form of a so-called *descriptor system*

$$\begin{cases} E\frac{\mathrm{d}\widehat{v}}{\mathrm{d}t}(t) = A\widehat{v}(t) + B\widehat{u}(t) & \text{for } t > 0, \\ \widehat{y}(t) = C\widehat{v}(t) & \text{for } t \geq 0, \end{cases} \tag{2.17}$$

with $E, A \in \mathbb{C}^{n_v \times n_v}$, $B \in \mathbb{C}^{n_v \times n_u}$, $C \in \mathbb{C}^{n_y \times n_v}$, $\widehat{u} \in \mathbb{C}^{n_u}$, $\widehat{v} \in \mathbb{C}^{n_v}$, and $\widehat{y} \in \mathbb{C}^{n_y}$. The vectors $\widehat{u}$, $\widehat{v}$, and $\widehat{y}$ are actually "signals", i.e., functions from the *time domain* $[0, \infty)$ to a vector space of suitable dimension. We will refer to them as "input", "state", and "output" of the system, respectively. Note that we are employing non-standard notation for the state (which is usually denoted by $\hat{x}$) to avoid clashing with the PDE setting that will be introduced in the next section. Moreover, we indicate time-domain quantities with a hat accent since they are not the main focus of our work, and will appear much less often than frequency-domain objects. Conversely, we will represent quantities in frequency domain (for which the hat accent is customarily reserved) as plain cursive letters.

System (2.17) is of first order, since only the first time derivative appears. However, this form is general enough, since a descriptor system of any (finite) order can be brought to first order by so-called *augmentation*: for instance, consider

$$\begin{cases} \sum_{n=1}^{S} E_n \frac{\mathrm{d}^n\widehat{v}}{\mathrm{d}t^n}(t) = A\widehat{v}(t) + B\widehat{u}(t) & \text{for } t > 0, \\ \widehat{y}(t) = \sum_{n=0}^{S-1} C_n \frac{\mathrm{d}^n\widehat{v}}{\mathrm{d}t^n}(t) & \text{for } t \geq 0, \end{cases}$$

with, possibly, some of the E_n's and C_n's being zero. We define the augmented state $\widehat{v}' = \left(\widehat{v}_0^\top, \widehat{v}_1^\top, \dots, \widehat{v}_{S-1}^\top\right)^\top \in \mathbb{C}^{Sn_v}$, with $\widehat{v}_n = \frac{\mathrm{d}^n\widehat{v}}{\mathrm{d}t^n}$. This yields the following first order system:

$$\begin{cases} \sum_{n=1}^{S} E_n \frac{\mathrm{d}\widehat{v}_{n-1}}{\mathrm{d}t}(t) = A\widehat{v}_0(t) + B\widehat{u}(t) & \text{for } t > 0, \\ \frac{\mathrm{d}\widehat{v}_{n-1}}{\mathrm{d}t}(t) = \widehat{v}_n(t) & \text{for } n = 1, \dots, S-1, \text{ for } t > 0, \\ \widehat{y}(t) = \sum_{n=0}^{S-1} C_n \widehat{v}_n(t) & \text{for } t \geq 0, \end{cases}$$

whose matrix representation (2.17) reads, in block form,

$$E' = \begin{bmatrix} E_1 & \cdots & E_{S-1} & E_S \\ I & & & \\ & \ddots & & \\ & & I & \end{bmatrix}, \; A' = \begin{bmatrix} A & & & \\ & I & & \\ & & \ddots & \\ & & & I \end{bmatrix}, \; B' = \begin{bmatrix} B \\ \\ \\ \\ \end{bmatrix}, \text{ and } C' = \begin{bmatrix} C_0 & \cdots & C_{S-1} \end{bmatrix}.$$

Other augmentation options are possible as well [Gui99; HMT09].

To have a well-posed problem, an initial condition for the state must be prescribed, thus obtaining a Cauchy problem, e.g.,

$$\begin{cases} E\frac{\mathrm{d}\widehat{v}}{\mathrm{d}t}(t) = A\widehat{v}(t) + B\widehat{u}(t) & \text{for } t > 0, \\ \widehat{v}(0) = 0, \end{cases} \quad \text{and} \quad \widehat{y}(t) = C\widehat{v}(t) \text{ for } t \geq 0.$$

For simplicity, we are setting to zero the initial condition of the Cauchy problem, but, of course, this should not be seen as a limitation. Note that, if the system was obtained by augmentation of an order-S system, then the initial condition involves also the time derivatives of $\widehat{v}$, up to order $S-1$. Some cases deserving a special mention are: $n_y = n_u = 1$, i.e., single-input single-output

(SISO) systems, and $C = I$ (the identity matrix of size $n_y = n_v$), so that the output coincides with the system state.

We note that, if E is full rank, we can invert it and obtain a (vectorial) ordinary differential equation (ODE): $\frac{\mathrm{d}\widehat{v}}{\mathrm{d}t} = E^{-1}A\widehat{v} + E^{-1}B\widehat{u} = A'\widehat{v} + B'\widehat{u}$. Instead, if E is not invertible, then we obtain a differential algebraic equation[1] (DAE). Let $E = U\Sigma V^H$ be the singular value decomposition (SVD) of E, so that

$$U = \left[U_r \middle| U_r'\right], \quad V = \left[V_r \middle| V_r'\right], \quad \text{and} \quad \Sigma = \begin{bmatrix}\Sigma_r & \\ & \end{bmatrix},$$

with $U_r, V_r \in \mathbb{C}^{n_v \times r}$ and $\Sigma_r \in \mathbb{C}^{r\times r}$, r being the rank of E. Given

$$\widehat{v}_r(t) = V_r^H \widehat{v}(t) \in \mathbb{C}^r \quad \text{and} \quad \widehat{v}_r'(t) = {V_r'}^H \widehat{v}(t) \in \mathbb{C}^{n_v - r},$$

we can separate the differential and the algebraic parts:

$$\begin{cases} \frac{\mathrm{d}\widehat{v_r}}{\mathrm{d}t}(t) = \Sigma_r^{-1}U_r^H A V_r \widehat{v}_r(t) + \Sigma_r^{-1}U_r^H A V_r' \widehat{v}_r'(t) + \Sigma_r^{-1}U_r^H B\widehat{u}(t) \\ 0 = {U_r'}^H A V_r \widehat{v}_r(t) + {U_r'}^H A V_r' \widehat{v}_r'(t) + {U_r'}^H B\widehat{u}(t) \end{cases} \qquad \text{for } t > 0.$$

Assuming invertibility of ${U_r'}^H A V_r'$, we can solve for $\widehat{v}_r'$ in the algebraic equation, and plug its expression in the differential equation, thus obtaining a non-singular system of size r.

Under minor regularity assumptions on the input, we can obtain a complementary viewpoint of the descriptor system by taking the Laplace transform: the Laplace transform of a signal $\widehat{w}$ ($\widehat{w}$ being either $\widehat{u}$, $\widehat{v}$, or $\widehat{y}$) is defined as

$$w(z) = \int_0^\infty \widehat{w}(t) e^{-zt} \mathrm{d}t, \tag{2.18}$$

for all *(complex) frequencies* $z \in \mathbb{C}$ for which the integral is defined. Note that we use the non-standard symbol z to denote the frequency, instead of the more common notations s or $\omega\mathrm{i}$, because, throughout this thesis, we will take z as the independent variable in an approximation problem like those presented in the previous section. Since the Laplace transform of $\frac{\mathrm{d}\widehat{v}}{\mathrm{d}t}(t)$ is $zv(z)$, the descriptor system (2.17) in *frequency domain* reads

$$\begin{cases} zEv(z) = Av(z) + Bu(z), \\ y(z) = Cv(z). \end{cases} \tag{2.19}$$

Note that, depending on the form of the input and on the spectral properties of the *matrix pencil* (A, E), see below, (2.19) might be valid only for frequencies in a subset of $\mathbb{C}$.

In its basic form (2.19), the frequency-domain formulation of a system only focuses on the "long-term" behavior of the system. In this regard, a critical quantity to study the system response to a given input is the *transfer function*, i.e., the $n_y \times n_u$ matrix

$$H(z) = C(zE - A)^{-1}B = \frac{C\,\mathrm{adj}(zE - A)B}{\det(zE - A)}, \quad \text{so that } y(z) = H(z)u(z), \tag{2.20}$$

where we denote by $\mathrm{adj}(\cdot)$ the adjugate matrix (not to be confused with the adjoint matrix),

[1] We are excluding the trivial case of $E = 0$, the zero matrix, which yields a linear system that can be solved (independently) at each time.

i.e., $\operatorname{adj}(A) = \det(A)A^{-1}$ for an invertible matrix A. By definition, $z \mapsto \det(zE - A)$ belongs to $\mathbb{P}_{n_v}(\mathbb{C};\mathbb{C})$. Similarly, by the Cayley-Hamilton theorem,

$$z \mapsto \operatorname{adj}(zE - A) \in \mathbb{P}_{n_v-1}(\mathbb{C};\mathbb{C}^{n_v\times n_v}) = \left\{ z \mapsto \sum_{n=0}^{n_v-1} C_n z^n : \{C_n\}_{n=0}^{n_v-1} \subset \mathbb{C}^{n_v\times n_v} \right\},$$

i.e., the space of matrix-valued polynomials of degree $n_v - 1$. Accordingly, the transfer function is a rational matrix-valued function, with poles at the eigenvalues of the pencil (A, E). By extension, we will call the poles of H the *poles* or *resonating frequencies of the system*. Now, we have two cases depending on the invertibility of E.

- If E is invertible, we may write $H(z) = C(zI - A')^{-1}B'$, with $A' = E^{-1}A$ and $B' = E^{-1}B$. As such, the poles of the system are the eigenvalues of A', and, if A' is diagonalizable, we have the *partial fraction form*

$$H(z) = \sum_{j=1}^{n_v} \frac{CP_jB'}{z - \lambda_j} = \sum_{j=1}^{n_v} \frac{r_j}{z - \lambda_j}, \tag{2.21}$$

  with $\{\lambda_j\}_{j=1}^{n_v}$ being the eigenvalues of A', with $\{P_j\}_{j=1}^{n_v}$ the corresponding family of spectral projectors ($A'P_j = \lambda_j P_j$). Note that, if λ_j is semisimple, e.g., $\lambda_j = \lambda_{j+1}$, multiple terms with the same denominator, but with projectors onto different spaces, will appear in (2.21). As such, if A' is diagonalizable, *all the poles of H are simple*. On the other hand, multiple poles may appear whenever A' is not diagonalizable, leading to the more cumbersome expression

$$H(z) = \sum_{j=1}^{n_v'} \sum_{k=1}^{d_j+1} \frac{r_{j,k}}{(z - \lambda_j)^k}, \tag{2.22}$$

  with $n_v' < n_v$ the number of distinct eigenvalues $\{\lambda_j\}_{j=1}^{n_v'}$, each associated to a spectral defect $d_j \geq 0$, i.e., the difference between its algebraic and geometric multiplicity.

- If E is not invertible, we (somewhat formally) say that the pencil (A, E) has ∞ as eigenvalue, with multiplicity at least one. To see what this means, we can process E as we did in the DAE case above. Splitting the contributions from non-singular and singular parts, we obtain

$$H(z) = CV \begin{bmatrix} z\Sigma_r - U_r^H AV_r & -U_r^H AV_r' \\ -{U_r'}^H AV_r & -{U_r'}^H AV_r' \end{bmatrix}^{-1} U^H B.$$

  By Schur complement, assuming ${U_r'}^H AV_r'$ to be invertible, we can obtain an expression of the form

$$H(z) = C'(zI - A')^{-1}B' + H_\infty, \tag{2.23}$$

  with $C' \in \mathbb{C}^{n_y\times r}$, $A' \in \mathbb{C}^{r\times r}$, $B' \in \mathbb{C}^{r\times n_u}$, and $H_\infty \in \mathbb{C}^{n_y\times n_u}$. The constant term H_∞ is the limit of H as $|z| \to \infty$, and, in a formal way, is the residue corresponding to the $n_v - r$ "missing" ∞ eigenvalues. Depending on the diagonalizability of A', the same observations on simple poles as in the previous case apply.

2.3 PDEs in frequency domain

Sources: [Eva10; GT01; Mat08; Qua09; Ram86; Sal+13]

Several applications of frequency-domain dynamical systems in engineering are discrete in nature, e.g. in lumped circuit modeling, so that *finite-dimensional* dynamical systems (2.19) suffice to study them. However, particularly in the fields of wave propagation and structural analysis, the interest often lies in understanding continuous *infinite-dimensional* systems. In this section, we introduce some archetypal examples of such models and we describe how some of the most important properties of frequency-domain dynamical systems can be extended from the finite-dimensional framework to the infinite-dimensional one.

To begin with, we define a spatial domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, i.e., an open bounded set, with sufficiently smooth boundary. As in the previous section, we denote by $\widehat{v}$ the time-dependent state of the system of interest. Only, this time, $\widehat{v}$ does not take values in the finite-dimensional vector space $\mathbb{C}^{n_v}$, but in a Hilbert function space defined over Ω, for instance the space of complex-valued square-(Lebesgue-)integrable functions

$$L^2(\Omega) = \left\{ \phi : \Omega \to \mathbb{C} : \phi \text{ measurable}, \int_\Omega |\phi(x)|^2 \, \mathrm{d}x < \infty \right\}, \tag{2.24}$$

or a more general Sobolev space

$$H^k(\Omega) = \left\{ \phi \in L^2(\Omega) : \sum_{\substack{\alpha_1,\dots,\alpha_d \geq 0 \\ \alpha_1+\dots+\alpha_d \leq k}} \int_\Omega \left| \frac{\partial^{\alpha_1+\dots+\alpha_d} \phi}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(x) \right|^2 \mathrm{d}x < \infty \right\}, \tag{2.25}$$

for $k \geq 0$, with derivatives that should be interpreted in the distributional sense. Negative and fractional Sobolev indices are also possible, according to the usual definitions. From here onward, we will denote by $\mathcal{V}$ the function space where $\widehat{v}$ takes values, i.e., $\widehat{v} : [0, \infty) \to \mathcal{V}$.

As in the previous section, we assume that $\widehat{v}$ is not available in closed form, being given only implicitly, as the (or a) solution of the partial differential equation (PDE)

$$\begin{cases} \frac{\partial \widehat{v}}{\partial t}(x,t) = \mathcal{L}(\widehat{v})(x,t) + \mathcal{B}(\widehat{u})(x,t) & \text{for } (x,t) \in \Omega \times (0,\infty), \\ \mathcal{F}(\widehat{v})(x,t) = 0 & \text{for } (x,t) \in \partial\Omega \times (0,\infty), \\ \widehat{v}(x,0) = 0 & \text{for } x \in \Omega, \end{cases} \tag{2.26}$$

where the calligraphic capital letters denote the following time-independent operators:

- $\mathcal{L}$ a differential operator from $\mathcal{V}$ to the dual space $\mathcal{V}^\star$;
- $\mathcal{B}$ a "forcing term" operator from some Hilbert space $\mathcal{W}_u$ to $\mathcal{V}^\star$; we might have a scalar/vector input ($\mathcal{W}_u = \mathbb{C}^{n_u}$) like in the finite-dimensional case;
- $\mathcal{F}$ an operator from $\mathcal{V}$ to some function space $\mathcal{V}_\partial$ over $\partial\Omega$, e.g., $H^{1/2}(\partial\Omega)$; for instance, $\mathcal{F}(\phi)$ may contain the trace of ϕ, or of its normal/tangential derivative, or a combination of them, on $\partial\Omega$; note that we may use $\mathcal{F}$ to denote non-homogeneous boundary conditions, e.g., by making $\mathcal{F}$ affine rather than linear.

As in the previous section, for simplicity, we have set the initial condition of the Cauchy problem to zero. In many applications, we also have a "measurement operator" $\mathcal{C} : \mathcal{V} \to \mathcal{W}_y$ (with $\mathcal{W}_y$ some Hilbert space) that defines the system output $\widehat{y} = \mathcal{C}(\widehat{v}) : (0, \infty) \to \mathcal{W}_y$. As in the finite-dimensional case, we might have a scalar/vector output ($\mathcal{W}_y = \mathbb{C}^{n_y}$ and $\mathcal{C} \in [\mathcal{V}^\star]^{n_y}$) or the state itself as the output ($\mathcal{W}_y = \mathcal{V}$ and $\mathcal{C}(\phi) = \phi$).

We list here some illustrative specific instances of (2.26).

- **Heat equation.** In the heat equation, $\widehat{v}(t) \in H^1(\Omega)$ denotes the temperature at a point of Ω. The operator $\mathcal{L}$ encodes diffusion by Fick's law: $\mathcal{L}(\phi) = \operatorname{div}(K \operatorname{grad} \phi)$, with K a (potentially, x-dependent) $d \times d$ positive-definite matrix, representing thermal diffusivity. The forcing term could, for instance, be a superposition of localized Gaussian heat sources with variable intensity: given source locations $\{\bar{x}_j\}_{j=1}^{n_u}$ and $\delta > 0$,

$$\mathcal{B}(\widehat{u})(x,t) = \sum_{j=1}^{n_u} \frac{\widehat{u}_j(t)}{(2\pi\delta^2)^{d/2}} \exp\left(-\frac{\|x - \bar{x}_j\|^2}{2\delta^2}\right).$$

  The boundary conditions might be of mixed Dirichlet-Neumann-Robin type, used to model a fixed temperature T_D, a fixed thermal flux g_N, and a fixed radiation coefficient k_R with respect to a reference temperature T_R, respectively: given a partition $\partial\Omega = \Gamma_D \sqcup \Gamma_N \sqcup \Gamma_R$ (with $\sqcup$ denoting disjoint union),

$$\mathcal{F}(\phi) = \begin{bmatrix} (\phi - T_D)|_{\Gamma_D} \\ (-(K \operatorname{grad} \phi) \cdot \nu - g_N)|_{\Gamma_N} \\ (k_R(\phi - T_R) - (K \operatorname{grad} \phi) \cdot \nu)|_{\Gamma_R} \end{bmatrix} \in H^{1/2}(\Gamma_D) \times H^{-1/2}(\Gamma_N) \times H^{-1/2}(\Gamma_R).$$

  with ν the outer normal to $\partial\Omega$.

- **Scalar wave equation.** The scalar wave equation is a second-order PDE, which, however, can be cast in first-order form by augmentation, see Section 2.2. The state $\widehat{v}(t) \in H^1(\Omega)$ represents excitations in the field through which the wave propagates, e.g., the air pressure for acoustic waves, or the normal plate displacement for (linearly) elastic (small) vibrations of a thin plate. The differential equation reads

$$\frac{\partial^2 \widehat{v}}{\partial t^2} = c^2 \Delta \widehat{v} + \widehat{f},$$

  with $\Delta = \operatorname{div} \operatorname{grad}$ the Laplace operator, c the (potentially, x-dependent) wavespeed, and $\widehat{f}$ a forcing term. As boundary conditions, we may take mixed Dirichlet-Neumann-Robin ones, with various physical interpretations (pressure, momentum, stress, etc.) depending on the wave medium. The corresponding $\mathcal{F}$ usually has a form similar to that for the heat equation. One important exception to this arises when modeling wave propagation on unbounded domains, e.g., through the *exterior scattering problem*:

$$\begin{cases} \frac{\partial^2 \widehat{v}}{\partial t^2} = c^2 \Delta \widehat{v} + f & \text{in } \mathbb{R}^d \times (0, \infty), \\ \lim_{\|x\| \to \infty} \|x\|^{(d-1)/2} \left(c(x) \frac{\partial}{\partial \|x\|} + \frac{\partial}{\partial t} \right) \widehat{v}(x, \cdot) = 0 & \text{in } (0, \infty), \\ \widehat{v} = \frac{\partial \widehat{v}}{\partial t} = 0 & \text{in } \mathbb{R}^d \times \{0\}. \end{cases} \tag{2.27}$$

  The "boundary condition" above is the so-called *Sommerfeld radiation condition*, which ensures that no "wave sources at ∞" exist. The characteristic that sets it apart from the other boundary conditions introduced above is the appearance of the time derivative in it. Before proceeding, we note that the exterior scattering problem above does not fall in our PDE framework, since it is formulated on an unbounded domain. However, it is standard, for computational reasons, to solve this class of problems by truncating the domain, thus enforcing the Sommerfeld condition on a finite boundary and recovering our desired PDE formulation on a bounded domain.

- **Elastic body vibrations.** The vibration of a 3D elastic body is a second-order PDE obtained by combining the conservation of momentum principle and an elastic constitutive law. The system state $\widehat{v}(t) \in [H^1(\Omega)]^d$, $d = 3$, denotes the displacement vector with respect to the undeformed configuration Ω. Restricting our focus to the small-deformation regime, the differential equation reads

$$\frac{\partial^2 \widehat{v}}{\partial t^2} + \eta \frac{\partial \widehat{v}}{\partial t} = \frac{1}{\rho} \operatorname{div} \sigma(\widehat{v}) + \widehat{f},$$

  with η a damping coefficient (measured in Hz), ρ the material density, $\widehat{f}(t)$ a force per unit mass, and σ the $d \times d$ stress tensor, which, in linear elasticity, can be expressed as

$$\sigma(\phi) = \mu \left(\operatorname{grad} \phi + (\operatorname{grad} \phi)^\top \right) + \lambda (\operatorname{div} \phi) I,$$

  μ and λ being the (potentially, x-dependent) Lamé constants of the material. For simplicity, we only consider Rayleigh mass damping, so that η is a constant. Note that more complicated damping models might not fall in the theoretical framework of Section 2.3.2 below.

  The Dirichlet-Neumann-Robin boundary conditions, encoded by the operator

$$\mathcal{F}(\phi) = \begin{bmatrix} (\phi - \delta_D)|_{\Gamma_D} \\ (\sigma(\phi)\nu - g_N)|_{\Gamma_N} \\ (k_R(\phi - X_R) + \sigma(\phi)\nu)|_{\Gamma_R} \end{bmatrix} \in [H^{1/2}(\Gamma_D)]^3 \times [H^{-1/2}(\Gamma_N)]^3 \times [H^{-1/2}(\Gamma_R)]^3,$$

  here assume the connotations of fixed displacement δ_D, external surface forces per unit area g_N, and external spring-like force with spring stiffness k_R and rest position at X_R.

We note that, with the exception of the heat equation, which is parabolic, the other problems presented above are hyperbolic. In particular, the operator $\mathcal{L}$ in first-order form (2.26) is symmetric and semi-positive definite for the heat equation, whereas it has a symplectic (Hamiltonian) structure in the other cases, e.g., for the scalar wave equation,

$$\frac{\partial}{\partial t} \begin{bmatrix} \phi \\ \phi' \end{bmatrix} = \begin{bmatrix} 0 & c \\ -(-c\Delta) & 0 \end{bmatrix} \begin{bmatrix} \phi \\ \phi' \end{bmatrix} + \begin{bmatrix} 0 \\ \widehat{f} \end{bmatrix}.$$

This has relevant consequences on the spectral properties of the system, as we will see in the upcoming sections.

The frequency-domain formulation of the time-domain PDE (2.26) is obtained with the same tools and aims as in the finite-dimensional case:

$$\begin{cases} zv(x,z) = \mathcal{L}(v)(x,z) + \mathcal{B}(u)(x,z) & \text{for } x \in \Omega, \\ \mathcal{F}(v)(x,z) = 0 & \text{for } x \in \partial\Omega. \end{cases} \tag{2.28}$$

Note that, to obtain the expression above, we have relied on the assumption that time appears in the PDE only in the left-hand-side of the differential equation. This is not the case if, e.g., the PDE operators are time-dependent or if time derivatives of $\widehat{v}$ appear in the forcing term or boundary conditions. Both cases can be handled by a careful calculation of the proper Laplace transforms. We provide in Section 2.3.3 some details on the latter case, in the specific instance of the Sommerfeld condition.

We proceed by deriving some useful properties for the frequency-domain versions of the PDEs

introduced above.

2.3.1 Simultaneously diagonalizable first-order systems

The heat equation can be cast in the form (2.28) by setting $\mathcal{L}(\phi) = \operatorname{div}(K \operatorname{grad} \phi)$. If K is symmetric and positive definite (which is the usual case in practical applications), then $\mathcal{L}$ (endowed with the boundary conditions) is an elliptic semi-negative definite operator, with compact inverse. As such, we can employ the spectral theorem to identify a basis of $H^1(\Omega)$, composed of eigenfunctions of $\mathcal{L}$. More specifically, there exist a discrete spectrum $\Lambda = \{\lambda_j\}_j \subset \mathbb{R}_{\leq 0}$ and a corresponding $H^1(\Omega)$-orthonormal basis $\Phi = \{\phi_\lambda\}_{\lambda\in\Lambda} \subset H^1(\Omega)$, such that $\lim_{j\to\infty} \lambda_j = -\infty$ and

$$\mathcal{L}\left(\sum_{\lambda\in\Lambda} \alpha_\lambda \phi_\lambda\right) = \sum_{\lambda\in\Lambda} \alpha_\lambda \lambda \phi_\lambda, \tag{2.29}$$

for "nice enough" coefficient sequences $\{\alpha_\lambda\}_{\lambda\in\Lambda}$. Notably, the identity above should be interpreted as: if the left series converges in the $H^1(\Omega)$ topology, then the right series converges in the $H^{-1}(\Omega)$ topology, and equality holds.

Consequently, the frequency-domain PDE can be cast over the basis Φ as

$$z \sum_{\lambda\in\Lambda} v_\lambda(z)\phi_\lambda = \sum_{\lambda\in\Lambda} v_\lambda(z)\lambda\phi_\lambda + \sum_{\lambda\in\Lambda} b_\lambda(u,z)\phi_\lambda, \tag{2.30}$$

with all series converging in the $H^{-1}(\Omega)$ topology, and

$$v_\lambda(z) = \langle v(\cdot,z), \phi_\lambda\rangle_{\mathcal{V}} \quad \text{and} \quad b_\lambda(u,z) = {}_{\mathcal{V}^\star}\langle \mathcal{B}(u)(\cdot,z), \phi_\lambda\rangle_{\mathcal{V}}.$$

Since Φ is a basis, we can solve (2.30) component by component, leading to a *spectral expansion*

$$v(\cdot,z) = \sum_{\lambda\in\Lambda} \frac{b_\lambda(u,z)}{z-\lambda}\phi_\lambda, \tag{2.31}$$

which generalizes (2.21). Accordingly, if $\mathcal{B}(u)(\cdot,z)$ is holomorphic or meromorphic with respect to z over some compact $A \Subset \mathbb{C}$, then $v(\cdot,z)$ and $y(z) = \mathcal{C}v(\cdot,z)$ are also meromorphic there. This justifies a rational approximation approach. We remark that, by symmetry of $\mathcal{L}$, (2.31) is an $H^1(\Omega)$-orthogonal expansion. This property will be crucial in the next sections.

To conclude, we note that, if $0 \notin \Lambda$ (e.g., if the Dirichlet boundary Γ_D is non-empty), all the poles of v are located in the left half $\mathbb{C}$-plane, so that the system state is bounded at all frequencies with non-negative real parts.

2.3.2 Simultaneously diagonalizable second-order systems

Analyzing the second-order PDEs introduced in Section 2.3 is a bit trickier, since the operator $\mathcal{L}$ in their first-order augmented formulation is not symmetric. However, it turns out that, in several cases, we can recycle the results for first-order problems. To this aim, assume that the boundary conditions of the frequency-domain PDE are independent of z, or, equivalently, that the boundary conditions of the time-domain PDE are time-independent and do not contain time derivatives of $\widehat{v}$. Without loss of generality, we develop the theory in the linear elasticity case. Similar conclusions can be obtained for the frequency-domain scalar wave equation, i.e., the

so-called *Helmholtz equation*: $z^2 v = c^2 \Delta v + f$.

Under standard assumptions on the PDE data (ρ, μ, and λ), we can apply the spectral theorem to the operator $\mathcal{L}$ appearing in the *second-order* formulation. This leads to an expression of the form

$$(z^2 + \eta z) \sum_{\lambda \in \Lambda} v_\lambda(z) \phi_\lambda = \sum_{\lambda \in \Lambda} v_\lambda(z) \lambda \phi_\lambda + \sum_{\lambda \in \Lambda} \mathcal{V}^\star \langle f(\cdot, z), \phi_\lambda \rangle_{\mathcal{V}} \phi_\lambda, \tag{2.32}$$

and a second-order $H^1(\Omega)$-orthogonal spectral expansion holds:

$$v(\cdot, z) = \sum_{\lambda \in \Lambda} \frac{\mathcal{V}^\star \langle f(\cdot, z), \phi_\lambda \rangle_{\mathcal{V}}}{z^2 + \eta z - \lambda} \phi_\lambda. \tag{2.33}$$

As in the previous case, the rational approximation endeavor is justified if $\mathcal{B}(u)(\cdot, z)$ is holomorphic or meromorphic with respect to z.

Note that, in the undamped case ($\eta = 0$), the poles of v are located on the imaginary axis. On the other hand, if damping is applied ($\eta > 0$), all the resonating frequencies (with the possible exception of $z = 0$) are located in the left half $\mathbb{C}$-plane, and the system response is bounded at frequencies with positive real parts.

As in the previous case, since the operator $\mathcal{L}$ from the second-order formulation is normal, (2.33) is an $H^1(\Omega)$-orthogonal expansion.

2.3.3 Non-simultaneously diagonalizable systems

Both cases described above rely on the spectral theorem to expand the PDE onto an eigenbasis of $\mathcal{L}^{-1}$. Notably, the eigendecomposition requires the compactness and normality of $\mathcal{L}^{-1}$ to be applied. While compactness can be guaranteed in most PDE applications by Sobolev embedding, normality is not always present. In the class of non-normal problems, the ones that concern us the most are exterior scattering problems, which, in frequency domain, can be cast in the Helmholtz form

$$\begin{cases} \left(-\Delta + \frac{z^2}{c^2}\right) v(\cdot, z) = f(\cdot, z) & \text{in } \Omega, \\ \left(\frac{\partial}{\partial \nu} + \frac{z}{c}\right) v(\cdot, z) = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.34}$$

with ν the outer normal to $\partial\Omega$. We note that it is customary to choose Ω as a ball in $\mathbb{R}^d$ centered at 0, so that $\frac{\partial}{\partial \nu} = \frac{\partial}{\partial \|x\|}$.

Since the frequency appears in the boundary conditions, we cannot repeat the same derivation as before, since applying the spectral theorem to the Laplacian operator endowed with z-dependent boundary conditions will necessarily yield a z-dependent spectrum and eigenbasis. Thus, alternative approaches become necessary to obtain a rational-like expansion of v. In [Bon+20b], we show the following result.

Theorem 2.6 (Meromorphicity of scattering frequency response [Bon+20b, Proposition 5.3])**.** *Fix an arbitrary bounded open set $A \subset \mathbb{C}$, and let $v(\cdot, z) \in \mathcal{V} = H^1(\Omega)$ be the solution of (2.34), with $f(\cdot, z)$ holomorphic with respect to z over A. Then, there exist two functions P and Q, with $P : A \to \mathcal{V}$ holomorphic over A, and $Q : A \to \mathbb{C}$ a polynomial (of finite degree) such that $v(\cdot, z) = P(z)/Q(z)$ for all $z \in A$. The denominator Q can be chosen so that all of its roots lie in the left half $\mathbb{C}$-plane, with the possible exception of $z = 0$.*

Sketch of proof. First, one shows that $v(z)$ is defined and bounded for all z in the right half $\mathbb{C}$-plane and on the imaginary axis, except at $z = 0$. This is a classical result in PDE theory, and can be done using standard Hilbert-space tools.

Then, one uses Riesz representation theory to cast (2.34) in $\mathcal{V}$, rather than in $\mathcal{V}^\star$, as

$$(\mathcal{I} + \mathcal{T}(z))v(\cdot, z) = F(z).$$

Above, $\mathcal{I} : \mathcal{V} \to \mathcal{V}$ is the identity operator in $\mathcal{V}$, so that $\mathcal{I}\phi$ is the Riesz representer of $-\Delta\phi + \phi \in \mathcal{V}^\star = H^{-1}(\Omega)$, whereas $\mathcal{T}(z) : \mathcal{V} \to \mathcal{V}$ contains the Riesz representer of the remaining terms of (2.34), i.e.,

$$\langle \mathcal{T}(z)\phi, \psi \rangle_{\mathcal{V}} = \langle \left(\frac{z^2}{c^2} - 1 \right) \phi, \psi \rangle_{L^2(\Omega)} + \langle \frac{z}{c}\phi, \psi \rangle_{L^2(\partial\Omega)} \quad \forall \phi, \psi \in \mathcal{V}.$$

Note that the boundary conditions are included in $\mathcal{T}(z)$. Moreover, $F(z) \in \mathcal{V}$ is the Riesz representer of $f(\cdot, z) \in \mathcal{V}^\star$, which is also holomorphic with respect to z. The perturbation $\mathcal{T}(z)$ is analytic with respect to z (it is a quadratic polynomial) and is compact (intuitively, by Sobolev embedding, because $\mathcal{T}$ only contains up to first derivatives). To obtain the claim, it suffices to apply [Ste68, Theorem 1]. □

We observe that the theorem does not provide any information on the pole orders, i.e., the exponents in the denominators of a partial fraction expansion like (2.22). Indeed, it only guarantees that the "total pole order" of A, i.e., the sum of all the orders of the poles in A is finite, for all bounded A. This implies that, if the poles $\Lambda = \{\lambda_j\}_j$ are countably infinite[2], then $\lim_{j\to\infty} |\lambda_j| = \infty$.

As a final note, we remark that, by the Fredholm alternative, if z is a resonating frequency, then there exists a nontrivial solution to the homogeneous problem obtained by setting $f = 0$ in (2.34), and vice-versa. In particular, if $\lambda \in \Lambda$, let $\phi_\lambda \in \mathcal{V} \setminus \{0\}$ be a solution of the corresponding homogeneous problem. Then, using $\bar{\cdot}$ to denote complex conjugation,

$$\begin{cases} \left(-\Delta + \frac{\overline{\lambda}^2}{c^2}\right) \overline{\phi_\lambda} = \overline{\left(-\Delta + \frac{\lambda^2}{c^2}\right) \phi_\lambda} = 0 & \text{in } \Omega, \\ \left(\frac{\partial}{\partial\nu} + \frac{\overline{\lambda}}{c}\right) \overline{\phi_\lambda} = \overline{\left(\frac{\partial}{\partial\nu} + \frac{\lambda}{c}\right) \phi_\lambda} = 0 & \text{on } \partial\Omega, \end{cases}$$

so that $\overline{\lambda}$ is a resonating frequency too. This allows us to conclude that resonating frequencies are either real (≤ 0) or come in complex conjugate pairs (with negative real parts).

2.4 MOR approaches for frequency-response problems

In applications, it is often required to evaluate the transfer function (2.20) of a system at many frequencies, often spanning several orders of magnitude. Due to the necessity to invert the frequency-dependent operator $zE - A$ to obtain the resolvent, this can incur in a substantial computational cost. Such concerns are especially relevant if the state dimension n_v is large, which is usually the case in practice, e.g., when the descriptor system (2.19) stems from a discretization of a PDE over a fine grid, cf. Section 2.3. In particular, we put ourselves in the framework where performing a generalized eigendecomposition of the pencil (A, E) (thus identifying explicitly poles and residues of H) is unfeasible, e.g., because of numerical instabilities, or just because of the sheer size of the matrices involved.

[2]Since Theorem 2.6 holds uniformly over compact sets in $\mathbb{C}$, the poles of v can be at most countably infinite.

In order to alleviate the computational burden, it might make sense to replace the original transfer function H (the *full order model*, FOM) with a surrogate version $\widetilde{H}$ (the *reduced order model*, ROM), which should satisfy the following properties:

- Evaluating the ROM at all the desired frequencies saves times over simply evaluating the FOM at all the desired frequencies. It is customary to split MOR approaches (at least, for frequency-domain applications) into two parts: first, the ROM $\widetilde{H}$ is built starting from samples of H at few representative frequencies (the so-called *snapshots* of the FOM); then, once $\widetilde{H}$ has been constructed, it can be evaluated at any (not yet sampled) frequency of interest. The first phase, dubbed *offline phase*, is expensive due to the necessity to sample the FOM, and also due to the operations needed to build $\widetilde{H}$ itself, but it only needs to be carried out once. The second phase, dubbed *online phase*, is cheaper, since it is independent of the FOM. If the complexity of the online phase is independent of n_v, the ROM is said to be *online-efficient*.
- The ROM approximates the FOM with acceptable accuracy. Since we know that H is a rational function, cf. (2.21) and (2.22), it is customary to confer a rational structure to $\widetilde{H}$ too. The task of "achieving acceptable accuracy" can then be translated as approximating well the poles (in $\mathbb{C}$) and the residues (in $\mathbb{C}^{n_y \times n_u}$) of H. It should be noted that approximating well *all* pole/residue pairs prevents online efficiency. However, fortunately, practical applications require evaluating H (thus, by proxy, $\widetilde{H}$) only at frequencies in a certain region of interest, what we will refer to as "frequency range". For instance, in order to analyze long-term resonant behavior in hyperbolic PDEs, it is common to study purely imaginary frequencies (corresponding to harmonic modes) in a certain range $z \in [z_{\min}\mathrm{i}, z_{\max}\mathrm{i}]$. As such, the pole/residue pairs can be sorted according to their "relevance" over the frequency range (as we will see in the next chapters, relevance could be measured, e.g., using the Green's potential of the frequency range). Then, we can formalize our accuracy requirement by asking that the ROM approximates well the *most relevant* pole/residue pairs.

Before proceeding to detail the state-of-the-art of MOR approaches for frequency-domain problems, we wish to comment on the fact that, in the previous paragraphs, we have implied that the snapshots of the FOM are samples of the transfer function H. This is not at all a necessity in a MOR approach. In fact, only some of the methods that we present in the upcoming sections build a ROM from samples of H. The other methods use samples of the system state $v(z)$ or of its derivatives, or even trajectories $\widehat{v}(t)$ of the state in the time-domain formulation of the system.

2.4.1 Data-driven rational approximation

Sources: [ABG20; Ant05; BG17; GS99; GTG15; IA14; NST18; Xia+19]

One of the most natural (and also most popular) ways of building the ROM $\widetilde{H}$ is to follow an "approximation theory" approach: since we know from system theory that the transfer function is rational, we are behooved to build the ROM $\widetilde{H}$ by rational approximation, relying on snapshots of H at few sampling frequencies $\{z_j\}_{j=1}^{S}$. This family of approaches is usually described as "data-driven" and "non-intrusive", because neither knowledge nor access to the matrices defining the system (E, A, B, and C) is required. As such, this class of methods is especially fitting (pun intended) for applications where data is experimental, coming from simulations in a lab rather than on a computer. More generally, the methods that we present here can, in principle, be applied to any kind of data, not necessarily snapshots of a first-order frequency-domain LTI system.

However, in such cases, one should worry about whether the data at hand is "well-approximable" by rational functions.

Let us first consider the case of SISO systems, whose transfer function is scalar-valued ($n_y = n_u = 1$), and assume that S snapshots at distinct frequencies $Z = \{z_j\}_{j=1}^S \subset \mathbb{C}$ have been computed. The simplest option is to define the ROM using Lagrange rational interpolation of H, as $\widetilde{H} = R^Z_{[S-1-N/N]}$, for some N. Considering the partial fraction decomposition (2.21), it is customary to choose $N \approx S/2$, so that $M \approx N$. While this approach sounds good in theory, it is often disastrous in practice (in its naive formulation for general sampling schemes), due to the (usually) terrible conditioning of the Vandermonde matrices that are necessary to express the linearized order conditions (2.11). Several countermeasures to improve the numerical stability of the problem have been explored:

- Select good sampling frequencies and a good polynomial basis. In few cases, the Vandermonde matrices of the rational interpolation problem are not too badly conditioned, usually thanks to discrete orthogonality conditions of the polynomial basis of choice. For instance, monomials are perfectly conditioned (unit condition number) over the roots of unity, as are the Chebyshev polynomials over the Chebyshev nodes. Unfortunately, it is not always possible to select "nice" sampling points, e.g., due to the necessity to sample across different orders of magnitude.
- Use the *barycentric* (sometimes called *partial fraction*) basis $\phi_{z_k^\star}(z) = (z - z_k^\star)^{-1}$, with $\{z_k^\star\}_k$ a collection of support points. This is not a polynomial basis but, by simple algebraic manipulations, it can be shown to be equivalent to one, at least in the specific case of diagonal $[N/N]$ rational interpolation. This idea defines the so-called *Loewner framework* [Ant05; ABG20; BG17; IA14], which works as follows:
  (i) Given $S = 2N + 1$, set aside $N + 1$ sample points $\{z_j\}_{j=1}^{N+1}$ to be support points.
  (ii) Define $\widetilde{H}(z) = \left(\sum_{j=1}^{N+1} \frac{c_j H(z_j)}{z - z_j}\right) \Big/ \left(\sum_{j=1}^{N+1} \frac{c_j}{z - z_j}\right)$, with $\{c_j\}_{j=1}^{N+1} \subset \mathbb{C}$ to be found. (Note that, with this choice, the values of H at the support points are interpolated automatically, as long as the coefficients c_j are non-zero.)
  (iii) Find $\{c_j\}_{j=1}^{N+1}$ by imposing linearized interpolation conditions

$$\widetilde{H}(z_k) \sum_{j=1}^{N+1} \frac{c_j}{z_k - z_j} = \sum_{j=1}^{N+1} \frac{c_j H(z_j)}{z_k - z_j} \overset{!}{=} H(z_k) \sum_{j=1}^{N+1} \frac{c_j}{z_k - z_j} \qquad \forall k = N+2, \dots, S \tag{2.35}$$

  (under a normalization condition to avoid $c_1 = \ldots = c_{N+1} = 0$).
- Oversampling. This means casting the Lagrange rational interpolation problem in LS form, so that, using S samples, we compute a rational approximant of type $[M/N]$, with $M + N + 1 < S$. To this aim, it suffices to replace the $M + N + 1$ linearized order conditions with the minimization of the weighted ℓ^2 linearized interpolation error:

$$\min_{\substack{P \in \mathbb{P}_M(\mathbb{C};\mathbb{C}) \\ Q \in \mathbb{P}_N(\mathbb{C};\mathbb{C})}} \sum_{k=1}^{S} w_k^2 \left| P(z_k) - Q(z_k) H(z_k) \right|^2 . \tag{2.36}$$

  Note that we have added real weights $\{w_k\}_{k=1}^S$, denoting, in some sense, the relative importance of the sample points. Moreover, we remark that a normalization condition (usually set on Q) is still necessary to avoid the trivial solution. Having increased the number

of rows of the Vandermonde matrices involved, the MOR procedure is more numerically stable. The price to pay is that of the additional snapshots. As a notable element of the LS school, the vector fitting (VF) [GTG15; GS99] approach sets the weights as $w_k = |Q(z_k)|^{-1}$, with the aim of recovering the "true" (non-linearized) interpolation error

$$\min_{\substack{P\in\mathbb{P}_M(\mathbb{C};\mathbb{C})\\ Q\in\mathbb{P}_N(\mathbb{C};\mathbb{C})}} \sum_{j=1}^{S} \left| \frac{P(z_j)}{Q(z_j)} - H(z_j) \right|^2 . \tag{2.37}$$

Due to the non-linearity (with respect to the coefficients of Q) introduced in problem (2.37), iterative methods (usually based on Picard iteration) are necessary to solve it approximately. As a final note, we mention that there exists also a LS version of the Loewner framework [NST18], where the interpolation conditions (2.35) are replaced by

$$\min_{(c_j)_{j=1}^{N+1}\in\mathbb{C}^{N+1}} \sum_{k=N+2}^{S} w_k^2 \left| \sum_{j=1}^{N+1} \frac{c_j(H(z_j) - H(z_k))}{z_k - z_j} \right|^2 , \tag{2.38}$$

under some normalization constraint to avoid the trivial solution $c_1 = \ldots = c_{N+1} = 0$.

Extensions to the non-scalar setting, when at least one of n_y and n_u is larger than 1, are almost immediate, but come at a small price. Indeed, as already mentioned in passing in Section 2.1.2, the linearized order conditions (2.11) of Lagrange rational interpolation lead to an overdetermined system when applied to vector-/matrix-valued functions, so that it becomes necessary to replace them by their LS version (2.36), with the absolute value being replaced by the $(n_y \times n_u)$-Frobenius norm. This means that, in general, in the vector/matrix setting, exact interpolation at all sample points is impossible using standard methods. Still, we may have interpolation at a subset of them, e.g., at the support points if the Loewner framework is employed.

Before proceeding, we note that, as an alternative way of solving the overdetermination issue, a matrix version of the Loewner framework can be defined, where interpolation of the full transfer function H is replaced by "tangential interpolation" of $e_j^L H(z)$ and $H(z)e_j^R$, with $\{e_j^L\}_j \subset \mathbb{C}^{1\times n_y}$ and $\{e_j^R\}_j \subset \mathbb{C}^{n_u\times 1}$ being user-defined tangential interpolation directions.

2.4.1.1 Model selection: adaptive VF and AAA

As mentioned in the previous section, an online-efficient rational surrogate has the purpose of approximating the most relevant poles/residues of v. Using a rational surrogate of type $[M/N]$ (for simplicity, we assume $M = N$ for the rest of this section), only up to N poles/residues of v can be estimated. However, in most applications, the user is unaware *a priori* of how many poles are relevant, and it is not at all obvious how a correct choice of the degree N could be performed. One might argue that choosing N as large as possible is the best option, since it maximizes the chances of identifying *at least* the relevant poles and residues. However, this leads to two issues, especially when the number of samples S is very large:

- Overfitting. Due to noise in the snapshots, resulting, e.g., from measurement noise or even from simple round-off errors, making the ROM too rich could lead to unstable and/or unreliable results due to overfitting.
- Lack of efficiency. A large surrogate is more expensive to evaluate than a smaller one.

This, sometimes, can hinder online efficiency. Moreover, storing larger surrogates might be inconvenient (this matters mostly for cases with large n_y or n_u).

For these reasons, one should try to find a "sweet spot" in the accuracy versus surrogate size trade-off. To this aim, both the Loewner framework and VF can be equipped with a "model selection" routine, whose objective is to explore different rational types $[N/N]$ and find the "best" one. All of this usually happens over a fixed set of sample points S, whose snapshots are pre-computed. Note that, here, we are vitally assuming to be in a data-rich setting, i.e., that S is large enough to identify the most relevant features of v. For this reason, we use the terms "model selection" rather than "adaptivity", which, in this thesis, we reserve for algorithms that add new sample points until an acceptable accuracy of the ROM is reached.

The model selection for the VF method, see, e.g., [GTG15; Xia+19], essentially works by trial-and-error, by testing different values of N, building a different VF surrogate for each of them, and then picking the one yielding the smallest LS misfit, defined as the minimal value of (2.37). More refined techniques have also been proposed, that progressively increase N, stopping as soon as the relative fitting error is below a prescribed tolerance (a typical value is 10^{-3}).

Model selection in the Loewner framework works quite similarly. However, one should note that increasing N in the Loewner framework requires turning sample points into support points, so that the number of addends in the outer sum in (2.38) decreases. A quite popular flavor of model selection for the Loewner framework is the AAA algorithm [NST18], where the selection of which sample points should be turned into support points is carried out in a greedy way, based on the pointwise misfit between FOM and ROM at the sample points.

2.4.2 State-based intrusive methods: (Petrov-)Galerkin projection

Sources: [Ant05; BF14; BHM18; Fre03; Gri97; GTG15; GW08; QMN15; RHP08]

A fairly complementary view of the MOR endeavor is taken in projective approaches. Here, the core objective is building a surrogate for the system state v. The surrogate for the transfer function is obtained only afterwards, as a post-processing step, essentially by multiplying the ROM of v by the matrix C from the left. In this context, values of the transfer function H are not enough to build the ROM, and snapshots of the state v are necessary.

The main idea to obtain the surrogate state $\widetilde{v}$ is to perform a (Petrov-)Galerkin projection of the state equation $(zE - A)v(z) = Bu$ onto two linear subspaces $\widetilde{\mathcal{V}}, \widetilde{\mathcal{W}} \subset \mathcal{V}$ of size $R < n_v$, from the right and left, respectively. Note that $\widetilde{\mathcal{V}} = \widetilde{\mathcal{W}}$ in the Galerkin case, and that the two subspaces might have different dimensions in an LS Petrov-Galerkin setting. For simplicity, we ignore the latter case, since extending our discussion to it is mostly trivial.

In summary, the projection of the FOM can be carried out by following these steps:

- Find bases of $\widetilde{\mathcal{V}}$ and $\widetilde{\mathcal{W}}$, and use them as columns of the rectangular projection matrices $\widetilde{V}, \widetilde{W} \in \mathbb{C}^{n_v \times R}$, respectively.

- Define the surrogate as a (z-dependent) element $\widetilde{v}(z)$ of $\widetilde{\mathcal{V}}$, whose expansion in the chosen basis of $\widetilde{\mathcal{V}}$ reads $\widetilde{v}(z) = \widetilde{V}\widetilde{\alpha}(z)$, with $\widetilde{\alpha}(z) \in \mathbb{C}^R$.

- Find the surrogate coefficients as the solution of the reduced system

$$\widetilde{W}^H(zE - A)\widetilde{v}(z) = \widetilde{W}^H(zE - A)\widetilde{V}\widetilde{\alpha}(z) \overset{!}{=} \widetilde{W}^H Bu. \tag{2.39}$$

The reduced problem (2.39) has dimension R, but it requires matrix multiplications of size n_v, thus preventing an (online-)efficient evaluation of the ROM. To overcome this issue, it suffices to apply the distributive property of matrix multiplication, transforming (2.39) into

$$(z\widetilde{E} - \widetilde{A})\widetilde{\alpha}(z) = (z\widetilde{W}^H E\widetilde{V} - \widetilde{W}^H A\widetilde{V})\widetilde{\alpha}(z) \overset{!}{=} \widetilde{W}^H Bu = \widetilde{B}u. \tag{2.40}$$

Note that all terms denoted by a tilde in the outer left- and right-hand-sides have dimensions independent of n_v. Moreover, they can be precomputed once and for all in the offline phase. Consequently, in order to evaluate the ROM at some new frequency z in the online phase, it suffices to build the $R \times R$ matrix $z\widetilde{E} - \widetilde{A}$, and then solve (2.40). In this way, we can achieve online efficiency.

Note that a closed-form expression for the surrogate state is available, as

$$\widetilde{v}(z) = \widetilde{V}\widetilde{\alpha}(z) = \widetilde{V}(z\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}u.$$

In particular, by projecting also the matrix C, as $\widetilde{C} = C\widetilde{V} \in \mathbb{C}^{n_y \times R}$, we can derive a reduced expression for the system output and transfer function:

$$\widetilde{y}(z) = \widetilde{C}(z\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}u \qquad \text{and} \qquad \widetilde{H}(z) = \widetilde{C}(z\widetilde{E} - \widetilde{A})^{-1}\widetilde{B}.$$

This shows that a projective MOR approach yields a rational surrogate, successfully mimicking the structure of the FOM. In particular, the surrogate model is of sub-diagonal type $[R - 1/R]$, cf. (2.20).

Now that we have outlined the skeleton of projective MOR approaches, the only piece that is still missing is: how should the subspaces $\widetilde{\mathcal{V}}$ and $\widetilde{\mathcal{W}}$ be chosen to obtain a "good" ROM? We will discuss a few specific state-of-the-art strategies next. For simplicity of presentation, we mostly consider just the Galerkin case.

However, before proceeding, we wish to note that we have only given a cursory overview of projective methods for first-order LTI dynamical systems. Projective methods can be employed over a much larger range of problems of the general form $A(z)v(z) = B(z)u$. In particular, we remark that, in order to guarantee online efficiency, the FOM should depend on the parameter(s) z in a *separable* way, e.g.,

$$\left(\sum_{i=1}^{n_A} \theta_i(z) A_i\right) v(z) = \sum_{i=1}^{n_B} \vartheta_i(z) B_i u, \tag{2.41}$$

with $\{\theta_i\}_{i=1}^{n_A} \cup \{\vartheta_i\}_{i=1}^{n_b}$ *scalar* functions, and $\{A_i\}_{i=1}^{n_A}$ and $\{B_i\}_{i=1}^{n_b}$ parameter-*independent* matrices of suitable sizes. The corresponding reduced model is obtained by projecting each term in the sums onto the chosen subspace(s):

$$\left(\sum_{i=1}^{n_A} \theta_i(z) \underbrace{\widetilde{W}^H A_i \widetilde{V}}_{\widetilde{A}_i}\right) v(z) = \sum_{i=1}^{n_B} \vartheta_i(z) \underbrace{\widetilde{W}^H B_i}_{\widetilde{B}_i} u. \tag{2.42}$$

The FOM decomposition (2.41) is commonly referred to as *affine* in z by MOR practitioners

[QMN15; RHP08]. Due to the risk of confusing such term with its more standard mathematical meaning (of a constant plus a linear term), we will denote the "MOR" meaning as "affine$^{\text{MOR}}$" and the "standard" one as just "affine". Note that first-order frequency-domain LTI systems are affine in both senses. If (2.41) does not hold, there exist approaches to find an affine$^{\text{MOR}}$ approximation of the FOM, e.g., the empirical interpolation method [Bar+04].

In the next sections, we describe some popular and practical ways to choose the reduced spaces $\widetilde{\mathcal{V}}$ and $\widetilde{\mathcal{W}}$.

2.4.2.1 Krylov methods and implicit moment matching

By looking at the reduced system (2.40) from an eigenproblem viewpoint, we may say that we are approximating the eigenvalues of the pencil (A, E) by its so-called *Ritz eigenvalues*. Standard methods for computing Ritz eigenvalues are available in the linear algebra literature, with one of the most notable ones being the Krylov method. In this approach, given $z_0 \in \mathbb{C}$ and a vector $b \in \mathbb{C}^{n_v}$, one defines the subspace $\widetilde{\mathcal{V}}$ of dimension $R = S$ as

$$\mathcal{K}((z_0E - A)^{-1}, b, S) = \text{Span}\left\{(z_0E - A)^{-j}b\right\}_{j=1}^{S}. \tag{2.43}$$

Sometimes, the vector b is also appended to the family of generators of $\mathcal{K}((z_0E-A)^{-1}, b, S)$, since, after all, its "computation" requires no effort. A useful observation for practical computations involving Krylov subspaces is that an orthonormal basis for $\mathcal{K}((zE - A)^{-1}, b, S)$ can be built in a numerically stable fashion by the Arnoldi algorithm. Such basis can then be used to define in a numerically robust way the projection matrix $\widetilde{V}$, see, e.g., [Ant05; Fre03; Gri97; GW08].

This idea can be applied for our MOR purposes. For single-input systems, it is customary to choose $b = B$, whereas for multi-input systems one may, e.g., consider the union of the Krylov spaces obtained using the different columns of B as vectors b. With these choices, it turns out that the vectors spanning $\widetilde{\mathcal{V}}$ are actually the (columns of the) state $v(z_0)$ and its derivatives $\frac{\mathrm{d}^j v}{\mathrm{d}z^j}(z_0)$, for $j = 1, \ldots, S-1$. For this reason, one can actually show that the ROM is guaranteed to interpolate v and its derivatives at z_0, up to order $S - 1$, leading to the name *implicit moment matching*, see, e.g., [BF14], for the overall MOR approach.

In the implicit moment matching approach, the snapshots are the state v and its derivatives at the single point z_0 (note that this is not exactly the case if the Arnoldi procedure is applied, but the same conclusions apply). An extension to a "multi-point" method is possible, where, instead of using all the computational budget on a "deep" Krylov space $\mathcal{K}((z_0E - A)^{-1}, b, S)$, we define $\widetilde{\mathcal{V}}$ as the sum of Krylov spaces with different centers $\{z_i\}_{i=1}^{T} \subset \mathbb{C}$:

$$\widetilde{\mathcal{V}} = \bigoplus_{i=1}^{T} \mathcal{K}\left((z_iE - A)^{-1}, b, \frac{S}{T}\right).$$

This allows for "distributed" information over the frequency range, and results in Hermite-Lagrange interpolation of the system state v.

In the limit case $T = S$, no derivatives of v are considered. In such situation, the ROM approach changes name to *reduced basis* (RB) method [QMN15; RHP08]. Interestingly, for SISO systems, a parallelism could be drawn between the RB method and the (interpolatory) Loewner framework, see Section 2.4.1, since both methods interpolate H at all the sample points. However, RB is intrusive while the Loewner framework is not. The main price that one pays for the difference

in flexibility of the methods is that, by using the Loewner framework, only up to $\sim S/2$ poles can be approximated from S samples of v, whereas RB, with the same number of samples, can approximate twice as many poles.

As a final remark, we note that, in the Petrov-Galerkin case, the space $\mathcal{W}$ is defined by applying Krylov to the dual problem: $\mathcal{W} = \mathcal{K}((z_0E^H - A^H)^{-1}, c, S)$, with c, usually, set equal to C^H, or to the columns of C^H for multi-output systems. In the Petrov-Galerkin setting, RB and the Loewner framework can, in principle, achieve the same peak "snapshot usage efficiency", since $2R$ snapshots are necessary to build an approximation of type $[R-1/R]$ in both approaches (in the projective case, we are assuming that half of the samples are used to build each of $\widetilde{\mathcal{V}}$ and $\widetilde{\mathcal{W}}$).

2.4.2.2 Proper orthogonal decomposition

In the previous section, we have introduced methods that build a reduced space $\widetilde{\mathcal{V}}$ of dimension R starting from $S = R$ snapshots (for simplicity, we are only considering a Galerkin setting). However, the RB method (and, up to a point, the implicit moment matching method too) can be extended to an LS-like MOR approach, where $S > R$ snapshots are used to build the reduced model. The motivations for doing this are mainly robustness and online efficiency, as already discussed more thoroughly in Section 2.4.1.

The most widespread way of extracting the "best" dimension-R subspace from the S snapshots is a principal component analysis (PCA) of the snapshot matrix, which contains the snapshots as columns. Such PCA is usually computed via SVD, leading to the so-called *SVD-proper orthogonal decomposition* (SVD-POD) method [Ant05; QMN15], which works as follows:

- Given the snapshots $\{v(z_j)\}_{j=1}^S \subset \mathbb{C}^{n_v}$, assemble the snapshot matrix $X = [v(z_1)|\dots|v(z_S)] \in \mathbb{C}^{n_v \times S}$ and the *snapshot Gramian*

$$G \in \mathbb{C}^{S\times S}, \qquad (G)_{j'j} = \langle v(z_j), v(z_{j'})\rangle_{\mathcal{V}}, \qquad j, j' = 1, \dots, S. \tag{2.44}$$

- Compute the SVD of $G = U\Sigma U^H$, which is just an eigendecomposition (with $U \in \mathbb{C}^{S\times S}$) where the S diagonal elements of $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_S)$ are in non-increasing order.
- Define $\widetilde{U} = XU\Sigma^{-1/2} \in \mathbb{C}^{n_v \times S}$.
- Define the projection matrix $\widetilde{V}$ by extracting the first R columns of $\widetilde{U}$. The reduced space $\widetilde{\mathcal{V}}$ is the span of the columns of $\widetilde{V}$.

Note that, if $M \in \mathbb{C}^{n_v \times n_v}$ is the positive-definite matrix representing the $\mathcal{V}$-inner product, we can equivalently find $\widetilde{U}$ as matrix containing the left singular vectors appearing in the SVD of $M^{1/2}X$.

In our description above, we have implicitly assumed that R is fixed in advance, prescribing the dimension of the reduced system. On the other hand, in practice, it usually makes more sense to fix a tolerance $\epsilon > 0$, and choose R as a function of ϵ, as the smallest integer such that

$$\sum_{i=R+1}^{S} \sigma_i \leq \epsilon \sum_{i=1}^{S} \sigma_i.$$

By the Eckart-Young theorem and some algebraic manipulations, this is equivalent to: the R

columns of $\widetilde{V}$ must be able to capture well the range of X, up to a tolerance ϵ in the column-wise $\mathcal{V}$-norm, i.e.,

$$\sum_{j=1}^{S} \left\| v(z_j) - \sum_{k=1}^{R} (\widetilde{V})_{:k} \langle v(z_j), (\widetilde{V})_{:k} \rangle_{\mathcal{V}} \right\|_{\mathcal{V}}^2 \leq \epsilon^2 \sum_{j=1}^{S} \|v(z_j)\|_{\mathcal{V}}^2 ,$$

where we use $(\widetilde{V})_{:k}$ to denote the k-th column of $\widetilde{V}$.

We observe that, in this form, POD displays some similarities to the "model selection" approaches presented in Section 2.4.1.1, since it is able to find the "sweet spot" between the dimension and the accuracy of the ROM, with the snapshots being fixed. As a consequence, for POD to work well, we must assume to be in a data-rich environment, i.e., that the S snapshots are sufficient to identify the relevant features (poles and residues) of v. We will make an effort to remove this assumption in the next section.

As a final remark, it is rather interesting to note that a different version of SVD-POD is also available for frequency-domain problems: assume that snapshots from a *time-domain simulation* of $\widehat{v}(t)$, cf. Section 2.2, are available, forming the snapshot matrix $X = [\widehat{v}(t_1)| \dots |\widehat{v}(t_S)]$. The SVD-based construction described above can be carried out using such X, thus building a reduced frequency-domain model from time-domain data. This class of methods is sometimes applied under the name "dynamic mode decomposition" [BHM18].

2.4.2.3 Adaptive frequency sampling: the weak-greedy reduced basis method

In many practical cases, taking snapshots is expensive (hence the need for MOR) and one does not know *a priori* how many snapshots are "enough" to guarantee a good approximation accuracy. For this reason, MOR approaches that allow an adaptive selection of the number and locations of the snapshots are quite useful, particularly because they limit the risk of oversampling. We remark that we are not just interested in "model selection" here, since we do not assume to be in a data-rich framework. Instead, we wish to determine whether a ROM has good approximation properties *using only the (few) snapshots that have already been computed*. For frequency-response problems, this is rather tricky, due to the meromorphicity of the response, which, notably, prevents a uniform convergence of $\widetilde{v}$ to v on neighborhoods of the poles, cf. Section 2.1.2.

This problem is quite complicated, since we are asking for a *certification* of the ROM, i.e., a guarantee of the goodness of approximation. The MOR strategy that is most commonly used to this aim is the *weak-greedy* RB method [QMN15; RHP08]. We proceed by explaining the reason behind the two terms in its name:

- "Weak" refers to the fact that the FOM *residual*, which will be defined shortly, is employed to drive the adaptivity rather than the FOM *error* $\widetilde{v}(z) - v(z)$, used in the *strong-greedy* RB method. This is done with the objective of (offline) efficiency, since it turns out that evaluating the error norm is usually prohibitively expensive, whereas the residual (dual) norm can be obtained cheaply, as we now show. Given a generic parametric problem $A(z)v(z) = B(z)u$ whose solution is approximated by the surrogate $\widetilde{v}(z)$, the corresponding "residual" is defined as $A(z)\widetilde{v}(z) - B(z)u \in \mathcal{V}^\star$ or, equivalently (assuming that $v(z)$ exists), $A(z)(\widetilde{v}(z) - v(z))$. Note that, if $A(z)$ is almost singular, as it is for dynamical systems when z is near a pole, then the residue can be small while the error is not. More specifically, if A and B are locally bounded, the residue is bounded whenever $\widetilde{v}$ is, whereas the error is bounded whenever $\widetilde{v}$ *and* v are.

For an affine$^{\mathrm{MOR}}$ problem (2.41), we can express the residual norm squared as

$$\left\| \left(\sum_{i=1}^{n_A} \theta_i(z) A_i \right) \widetilde{v}(z) - \sum_{i=1}^{n_B} \vartheta_i(z) B_i u \right\|_{\mathcal{V}^\star}^2 = \sum_{i,i'=1}^{n_A} \theta_i(z)\overline{\theta_{i'}(z)} \langle A_i \widetilde{V}\widetilde{\alpha}(z), A_{i'}\widetilde{V}\widetilde{\alpha}(z)\rangle_{\mathcal{V}^\star} - 2\mathrm{Re}\left(\sum_{i,i'=1}^{n_A,n_b} \theta_i(z)\overline{\vartheta_{i'}(z)} \langle A_i \widetilde{V}\widetilde{\alpha}(z), B_{i'}u\rangle_{\mathcal{V}^\star} \right) + \sum_{i,i'=1}^{n_B} \vartheta_i(z)\overline{\vartheta_{i'}(z)} \langle B_i u, B_{i'}u\rangle_{\mathcal{V}^\star}, \tag{2.45}$$

with $\langle v, w\rangle_{\mathcal{V}^\star}$ being the inner product over $\mathcal{V}^\star$ (which exists by Riesz representation theory). By precomputing the terms $\langle A_i v(z_j), A_{i'} v(z_{j'})\rangle_{\mathcal{V}^\star}$, $\langle A_i v(z_j), B_{i'}u\rangle_{\mathcal{V}^\star}$, and $\langle B_i u, B_{i'}u\rangle_{\mathcal{V}^\star}$ for all i, i', j, and j' (with $\{v(z_j)\}_j$ being the snapshots), one can make the cost of evaluating (2.45) independent of n_v.

- "Greedy" refers to the fact that only one new sample point is added at a time, at the location that maximizes a "greedy indicator", e.g., the dual norm of the residual. We summarize the resulting procedure in Algorithm 1. Since the new snapshot does not affect the previous ones, we do not need to rebuild the ROM from scratch at every iteration. Indeed, looking at the reduced system (2.42), each term of the right-hand-side only gains a row, while each term of the left-hand-side gains a thin "border" of new entries: if $\widetilde{V}$ and $\widetilde{W}$ are the $n_v \times (S-1)$ projection matrices at step $S-1$, and ϕ and ψ are the columns that are about to be appended to $\widetilde{V}$ and $\widetilde{W}$, respectively, then

$$\widetilde{B}_i^{\mathrm{next}} = \left[\begin{array}{c} \widetilde{B}_i^{\mathrm{previous}} \\ \hline \psi^H B_i \end{array} \right] \qquad \text{and} \qquad \widetilde{A}_i^{\mathrm{next}} = \left[\begin{array}{c|c} \widetilde{A}_i^{\mathrm{previous}} & \widetilde{W}^H A_i \phi \\ \hline \psi^H A_i \widetilde{V} & \psi^H A_i \phi \end{array} \right]. \tag{2.46}$$

Algorithm 1 Weak-greedy RB

Require: distinct initial sample points $Z = \{z_1, \ldots, z_{S_0}\} \subset \mathbb{C}$, tolerance ϵ
Require: affine$^{\mathrm{MOR}}$ problem left-hand-side $A = A(z)$ and right-hand-side $b = b(z)$
Require: distinct test points $Z_{\mathrm{test}} = \{z_1, \ldots, z_T\} \subset \mathbb{C} \setminus Z$
  for $j = 1, \ldots, S_0 - 1$ **do**
    compute snapshot $v(z_j) = A(z_j)^{-1} b(z_j)$
  end for
  for $S = S_0, S_0 + 1, \ldots$ **do**
    compute snapshot $v(z_S) = A(z_S)^{-1} b(z_S)$
    build the projection matrix $\widetilde{V} = [v(z_1)| \ldots |v(z_S)]$
    Optional: orthonormalize the projection matrix
    build the ROM (2.42), using (2.46) if $S > S_0$
    evaluate the greedy indicator $\eta(z)$ at all $z \in Z_{\mathrm{test}}$
    find the point of worst approximation $z^\star = \arg\max_{z \in Z_{\mathrm{test}}} \eta(z)$
    if $\eta(z^\star) < \epsilon$ **then**
      return the current ROM
    end if
    move $z^\star$ from Z_{test} to Z, where it acquires the name z_{S+1}
  end for

We note that, in Algorithm 1, we have included an optional orthonormalization step for the projection matrix. This is to improve the stability of the projection. Normalizing the snapshots

has relevance *per se*, since the system state is meromorphic, so that samples near and far from the poles might have different orders of magnitude. Notably, we remark that $\widetilde{V}$ can be kept orthonormal at all iterations without impacting the incremental construction of the ROM (2.46), e.g., by the Gram-Schmidt procedure.

Now, as even a cursory glance at the MOR literature (for "standard", e.g., elliptic and parabolic, problems) could reveal, it is not customary to use the plain (dual) norm of the residual $\eta(z) = \|A(z)\widetilde{v}(z) - b(z)\|_{\mathcal{V}^\star}$ as greedy indicator, as we set out to do. Indeed, for coercive and inf-sup stable problems, one can theoretically prove the existence of a (z-dependent) constant $c(z)$ such that

$$\|\widetilde{v}(z) - v(z)\|_{\mathcal{V}} \leq c(z) \, \|A(z)\widetilde{v}(z) - b(z)\|_{\mathcal{V}^\star} \, . \tag{2.47}$$

For instance, in the context of elliptic PDEs, $c(z)$ might be the ratio of the continuity and coercivity constants involved in the Lax-Milgram theorem. When (2.47) holds, one usually employs as greedy indicator the quantity $\eta(z) = c(z) \, \|A(z)\widetilde{v}(z) - b(z)\|_{\mathcal{V}^\star}$, which provides a guaranteed bound for the ROM error. In practice, $c(z)$ is not explicitly computable without a significant (often, unfeasible) computational effort. So, it is customary to settle for an upper bound for c, say $\widetilde{c}$, that is cheaply computable *a posteriori*, and then set $\eta(z) = \widetilde{c}(z) \, \|A(z)\widetilde{v}(z) - b(z)\|_{\mathcal{V}^\star}$. Several techniques to this aim can be found in the literature, e.g., the successive constraint method [Huy+07; Huy+10].

That being said, frequency-domain applications are not "standard" MOR problems, since the factor c is often not very well-behaved: as we mentioned in passing above, $c(\lambda) = \infty$ at all poles λ. In particular, if v has poles in the target frequency range, or close to it, then most MOR approaches for finding the upper bound $\widetilde{c}$ fail. In fact, one might even wonder if a uniform control over the *error* is the correct measure of goodness of approximation, considering that v is unbounded, cf. Theorem 2.4. For these reasons, one usually falls back onto using the "bare" residual dual norm, since it is well-behaved at the poles of the system and solves the issue of the unboundedness of v. Accordingly, this is the *de facto* standard for weak-greedy MOR of frequency-domain applications in acoustics, electronics, optics, etc., see, e.g., [ABG20; Bay+20; RM18; RRM09; Rub14], whenever one anticipates poles in, or close to, the frequency range. One might (and, if possible, should) apply the standard indicator, containing the factor $\widetilde{c}$, whenever such term is bounded and can be computed reliably, e.g., for parabolic problems (2.26) and scattering problems (2.27), provided they are uniformly stable over the frequency range. We refer to Section 5.5.4 for a numerical example.

3 The minimal rational interpolation method

The *minimal rational interpolation* (MRI) method was proposed by the thesis author in [Pra20], combining some of the features of projection methods and of rational approximation approaches. In brief, MRI builds a surrogate model by constructing a rational approximation of the state v, starting from snapshots of v at some sample points. The key property of MRI, which distinguishes it from the rational approximation methods described in Section 2.4.1, is encoded by the "minimal" in the name of the method: the information provided by the snapshots is exploited as much as possible. For instance, this means that, for a fixed number of snapshots, MRI can build an approximant with double the rational type than, e.g., VF can. This "optimal snapshot usage" is, in some cases, a feature shared by MRI and projective approaches, see Section 2.4.2. However, as opposed to those methods, MRI is non-intrusive. Particularly, it does not require access to an affine$^{\text{MOR}}$ decomposition/approximation of the FOM, so that even problems with non-affine$^{\text{MOR}}$ frequency dependencies or problems with black-box solvers are amenable to MRI.

In this chapter, we provide a rigorous definition of MRI, and describe *a priori* convergence results, which show that MRI has good approximation properties. (To simplify our presentation, we postpone the proofs of such results until Chapter 4.) In particular, while MRI can be applied, in principle, for the approximation of any univariate function (even ∞-dimensional-valued ones), our convergence theory, at the moment, relies on some assumptions on the state v: more explicitly, we will require v to admit a (finite or infinite) partial fraction expansion with simple poles

$$v(z) = \sum_{\lambda \in \Lambda} \frac{r_\lambda}{\lambda - z}, \tag{3.1}$$

where $\Lambda \subset \mathbb{C}$ denotes the set of poles of v, each with a corresponding residue

$$r_\lambda = \lim_{z \to \lambda} (\lambda - z) v(z), \tag{3.2}$$

see also Section 2.2. The theory below will be shown to hold only when the "most relevant" residues (see below for a rigorous definition) form a linearly independent set. Crucially, this property cannot be satisfied if MRI is applied to scalar quantities of interest (QoIs), or to vector QoIs with too few components (when compared to the number of relevant residues). Consequently, while MRI can be applied even in such situations, the approximation quality is usually poor.

For simplicity of exposition, we start our discussion from a special case of MRI, the *fast LS Padé approximation* method, which was originally introduced in [Bon+20a]. This entails major simplifications in the definitions, statements, and derivations. An extension to the general MRI

approach will follow in Section 3.2.

3.1 The single-point case: fast LS Padé approximation

Fast LS Padé approximation generalizes Padé approximation to the vector setting. For simplicity, we assume the target function to take values in a Hilbert space, rather than in a Banach one, since having an inner product simplifies matters significantly. Note that even an "infinite-dimensional" functional setting is allowed. We proposed this technique in [Bon+20a], using the following definition.

Definition 3.1 (Fast LS Padé approximation [Bon+20a, Definition 4.1])**.** *Let $\mathcal{V}$ be a complex-valued Hilbert space with inner product $\langle v, w\rangle_{\mathcal{V}}$ and induced norm $\|v\|_{\mathcal{V}} = \langle v, v\rangle_{\mathcal{V}}^{1/2}$, and fix $z_0 \in \mathbb{C}$ and $M, N, E \geq 0$, with $E \geq \max\{M, N\}$. Also, let $v : \mathbb{C} \to \mathcal{V}$ be a $\mathcal{V}$-valued function of a complex variable, continuous and differentiable (in the complex sense) at least E times at z_0. An $[M/N]$ fast LS Padé approximant of v centered at z_0 (dependent on E) is a rational function*

$$v_{[M/N]}^{z_0} = \frac{P_{[M/N]}^{z_0}}{Q_{[M/N]}^{z_0}}, \tag{3.3}$$

such that

$$P_{[M/N]}^{z_0} \in \mathbb{P}_M(\mathbb{C}; \mathcal{V}) = \left\{ \sum_{n=0}^{M} p_n (z - z_0)^n : \{p_n\}_{n=0}^{M} \subset \mathcal{V} \right\}, \tag{3.4a}$$

$$Q_{[M/N]}^{z_0} \in \mathbb{P}_N^{z_0}(\mathbb{C}; \mathbb{C}) = \left\{ z \mapsto \sum_{n=0}^{N} q_n (z - z_0)^n : \{q_n\}_{n=0}^{N} \subset \mathbb{C}, \sum_{n=0}^{N} |q_n|^2 = 1 \right\}, \tag{3.4b}$$

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} P_{[M/N]}^{z_0} \bigg|_{z_0} \overset{!}{=} \frac{\mathrm{d}^n}{\mathrm{d}z^n} (Q_{[M/N]}^{z_0} v) \bigg|_{z_0} \quad \forall n = 0, \dots, M, \tag{3.4c}$$

$$J_E(Q_{[M/N]}^{z_0}) \leq J_E(Q) := \left\| \frac{\mathrm{d}^E}{\mathrm{d}z^E} (Qv) \bigg|_{z_0} \right\|_{\mathcal{V}} \quad \forall Q \in \mathbb{P}_N^{z_0}(\mathbb{C}; \mathbb{C}). \tag{3.4d}$$

An $[M/N]$ fast LS Padé approximant is actually computable (in an efficient way) given only the Taylor coefficients $v(z_0), \frac{\mathrm{d}v}{\mathrm{d}z}(z_0), \dots, \frac{\mathrm{d}^E v}{\mathrm{d}z^E}(z_0)$. Indeed, (3.4c) and (3.4d) can be expressed in terms of such coefficients by the Leibniz rule: for $n \in \mathbb{N}$,

$$\frac{\mathrm{d}^n}{\mathrm{d}z^n} (Qv) \bigg|_{z_0} = \sum_{m=0}^{\min\{N,n\}} \binom{n}{m} \frac{\mathrm{d}^m Q}{\mathrm{d}z^m}(z_0) \frac{\mathrm{d}^{n-m} v}{\mathrm{d}z^{n-m}}(z_0). \tag{3.5}$$

Notably, we will see that the identification of $Q_{[M/N]}^{z_0}$ through (3.4d) can be carried out by solving an eigenproblem. More details on this will be given in Section 5.1.

We also note that fast LS Padé approximants were introduced as an improvement over a prior similar version of the method, dubbed LS Padé approximation in [BNP18, Definition 4.2]. The main improvements over this latter technique are discussed in [Bon+20a].

In Definition 3.1, (3.4b) corresponds to a normalization condition on $Q_{[M/N]}^{z_0}$, which prevents the trivial solution $Q_{[M/N]}^{z_0} \equiv 0$. In particular, we remark that the normalized polynomials $\mathbb{P}_N^{z_0}(\mathbb{C}; \mathbb{C})$ are not invariant under centered dilations $z \mapsto z_0 + \alpha(z - z_0)$. For instance, this

means that rescaling the frequency, in general, changes the surrogate: if $\widetilde{v}(z) = v(\alpha z)$, then $\widetilde{v}^0_{[M/N]}(z) \neq v^0_{[M/N]}(\alpha z)$ in general. Moreover, let

$$\langle Q, Q' \rangle_{z_0} = \sum_{n=0}^{\infty} q_n \overline{q'}_n, \qquad \text{where } Q(z) = \sum_{n=0}^{\infty} q_n (z - z_0)^n \text{ and } Q'(z) = \sum_{n=0}^{\infty} q'_n (z - z_0)^n, \tag{3.6}$$

with induced norm $\|Q\|_{z_0} = \langle Q, Q \rangle_{z_0}^{1/2}$. In this metric, $\mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$ is the unit sphere in $\mathbb{P}_N(\mathbb{C};\mathbb{C})$.

The minimality condition in (3.4d) is effectively responsible for the identification of the surrogate denominator $Q^{z_0}_{[M/N]}$, and is pivotal for showing that fast LS Padé approximants have good properties. In this regard, we note that we say "a", rather than "the", fast LS Padé approximant, since, in general, neither the denominator $Q^{z_0}_{[M/N]}$ nor the rational function $P^{z_0}_{[M/N]}/Q^{z_0}_{[M/N]}$ is uniquely determined. Indeed, on one hand, $Q^{z_0}_{[M/N]}$ is never unique because it can be multiplied by an arbitrary unit complex number (in such case the surrogate stays unchanged, since $P^{z_0}_{[M/N]}$ is also multiplied by the same factor, see (3.4c)). On the other hand, J_E might have two (or more) linearly independent minimizers in $\mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$, leading to different fast LS Padé approximants. This is especially the case in a numerical framework, if round-off becomes relevant, see Section 5.2.1.

As already mentioned, the "FOM information" needed to compute a fast LS Padé approximant is the collection of Taylor coefficients of v at z_0, of order up to E. This allows to discuss (qualitatively) the topic of "optimal snapshot usage".

Remark 3.1. *Assume that we are interested in constructing an $[M/N]$ fast LS Padé approximant, so that, according to Definition 3.1, at least* $\max\{M,N\}+1$ *Taylor coefficients of v are necessary. (Note that this is already an improvement on standard rational approximation, where $M+N+1$ Taylor coefficients are necessary to achieve the same rational type, cf. Definition 2.5.) In order to exploit the information on v (i.e., the snapshots) "optimally", one should only take as many samples as necessary, i.e., choose $E = \max\{M,N\}$. Setting E any larger leads to an LS-like Taylor method and "wastes" snapshots. This is not really advisable, particularly because it does not alleviate numerical instabilities (see, e.g., the numerical tests in Section 5.5), differently from the usual stabilizing effect of adding snapshots in classical methods, e.g., VF. Accordingly, the choice $E = \max\{M,N\}$ is the* de facto *standard in our applications, thus justifying the absence of the parameter E in our notation $v^{z_0}_{[M/N]}$ for fast LS Padé approximants.*

Akin to standard Padé approximants (for scalar functions), fast LS Padé approximants rely on information on the target function at a single point. Notably, the value of v and of its derivatives up to order M at z_0 is recovered exactly by the surrogate, as encoded (in a linearized way) by (3.4c). Accordingly, we may expect the approximation quality to degrade if we move farther away from z_0. Indeed, this behavior can be rigorously proven, even though the theoretical framework requires some assumptions.

Assumption 3.1 (Local simple partial fraction expansion)**.** *Let $B^{z_0}(R) = \{z \in \mathbb{C} : |z - z_0| < R\}$. We assume that v admits the simple-pole partial fraction expansion* (3.1) *over $B^{z_0}(R)$, with distinct poles and $z_0 \notin \Lambda$. If the poles of v are (countably) infinite, we require the partial fraction expansion to be an absolutely convergent series in the $\mathcal{V}$-sense, so that, in particular,*

$$\|v(z)\|_{\mathcal{V}} \leq \sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z|} < \infty \qquad \forall z \in B^{z_0}(R) \setminus \Lambda.$$

Moreover, let $\Lambda^{z_0}(R) = \Lambda \cap B^{z_0}(R)$ be the poles of v inside $B^{z_0}(R)$, those that we dubbed "relevant"

above. We assume that $\Lambda^{z_0}(R)$ has a finite number of elements $N^{z_0}(R)$. In particular, we employ the pole ordering

$$v(z) = \sum_{j=1}^{N^{z_0}(R)} \frac{r_{\lambda_j}}{\lambda_j - z} + \sum_{\lambda \in \Lambda \setminus \Lambda^{z_0}(R)} \frac{r_\lambda}{\lambda - z}, \tag{3.7}$$

with $|\lambda_1 - z_0| \leq |\lambda_2 - z_0| \leq \ldots \leq |\lambda_{N^{z_0}(R)} - z_0| < R$.

Note that a function v satisfying Assumption 3.1 over $B^0(R)$ is meromorphic there, but not necessarily over any $B^{z_0}(R+\varepsilon)$, for $\varepsilon > 0$: for instance

$$v(z) = \sum_{n=0}^{\infty} \frac{1/n!}{1 + e^{-n} - z} \tag{3.8}$$

has a branch point at $z = 1$ (encoded by a sequence of poles with 1 as limit point), and satisfies Assumption 3.1 over $B^0(1)$ but not over any $B^0(1+\varepsilon)$. Incidentally, this example shows why we have not introduced an ordering in the poles outside $B^{z_0}(R)$, since it might not exist.

Additionally, to develop our theory, we will require the residues to be orthogonal, in the following sense.

Assumption 3.2 (Orthogonal partial fraction residues)**.** *Assumption 3.1 holds. The residues $\{r_\lambda\}_{\lambda \in \Lambda}$ form a $\mathcal{V}$-orthogonal family, i.e., $\langle r_\lambda, r_{\lambda'} \rangle_{\mathcal{V}} = 0$ for all $\lambda \neq \lambda'$.*

A discussion on the possibility of extending our results to the weaker assumption of linearly independent, rather than orthogonal, residues will follow in Section 3.1.3.

In the upcoming sections, we show *a priori* convergence results for fast LS Padé approximants, first in the approximation of poles of v, and then in the approximation of v itself.

3.1.1 Pole convergence

Throughout the present section, we assume that Assumption 3.1 holds. Our main aim here is to show that the fast LS Padé denominator $Q^{z_0}_{[M/N]}$ is close to the "ideal" denominator $g^{z_0}_N$, which, for $N \leq N^{z_0}(R)$, is defined as an element of $\mathbb{P}^{z_0}_N(\mathbb{C};\mathbb{C})$ such that $g^{z_0}_N(\lambda_j) = 0$ for all $j = 1, \ldots, N$. The polynomial $g^{z_0}_N$ identifies exactly the N most relevant (with respect to z_0) poles of v.

Our first result states a bound for the values $|Q^{z_0}_{[M/N]}(\lambda_j)|$, for $j = 1, \ldots, N$, at least under Assumption 3.2. Since $g^{z_0}_N(\lambda_j) = 0$ for those values of j, this corresponds to bounding the absolute error in the denominator $|Q^{z_0}_{[M/N]} - g^{z_0}_N|$ at the relevant poles of v.

Lemma 3.1 (Denominator value at poles [Bon+20a, Lemma 5.4] (extended))**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq N^{z_0}(R)$. Assume that $Q^{z_0}_{[M,N]}$ is computed using Definition 3.1, relying on E Taylor coefficients of v. Also, let $\bar{R}_N = R$ if $N \geq N^{z_0}(R)$ and $\bar{R}_N = |\lambda_{N+1} - z_0|$ otherwise. Then, for all $j = 1, \ldots, N$, we have the bound

$$\left| Q^{z_0}_{[M/N]}(\lambda_j) \right| \leq C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N} \right)^{2E}, \tag{3.9}$$

with C_j independent of M (in fact, the value of M is irrelevant here) and E, but dependent on N.

Proof. See Section 4.1.4. □

The result above shows that, if N is fixed, the value of the fast LS Padé denominator at a (relevant) exact pole converges to 0, exponentially in the number of snapshots E. Thanks to the normalization of $Q^{z_0}_{[M/N]}$ and to the fact that the poles of v are simple, we can convert this into convergence (at the same rate) of the surrogate poles to the relevant exact ones.

Theorem 3.1 (Pole convergence [Bon+20a, Theorem 5.5])**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq N^{z_0}(R)$. Assume that $Q^{z_0}_{[M,N]}$ is computed using Definition 3.1, relying on E Taylor coefficients of v. Also, let $\bar{R}_N$ be as in Lemma 3.1. If $|\lambda_N - z_0| < \bar{R}_N$, then, for all $j = 1, \ldots, N$ and E large enough,

$$\min_{\lambda' : Q^{z_0}_{[M/N]}(\lambda') = 0} |\lambda' - \lambda_j| \leq C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N} \right)^{2E}, \tag{3.10}$$

with C_j independent of M (in fact, the value of M is irrelevant here) and E, but dependent on N.

Proof. See Section 4.1.5. □

This result shows that fast LS Padé approximation can be employed as a fairly good eigensolver, as far as the approximation of eigenvalues close to z_0 is concerned.

Given the (exponential) convergence result for fixed N, we can ask whether convergence is guaranteed also when N varies. In particular, the case of practical interest is that of N diverging to ∞, which happens, for instance, when using diagonal fast LS Padé approximants of type $[E/E]$, with $E \to \infty$. This case is much trickier, and it turns out that the rate of convergence depends in a fairly complicated way on the pattern of the poles of v: notably, if poles are sufficiently sparse, we can even expect super-exponential convergence, as shown in our numerical results in Section 5.5. Since, usually, *a priori* information on the poles of v is not available, we are content with a "plain" convergence result in the general case. However, we require the following global assumption.

Assumption 3.3 (Global simple orthogonal partial fraction expansion)**.** *We assume that v admits the simple-pole partial fraction expansion* (3.1) *over all compact subsets of $\mathbb{C}$, with distinct poles and $z_0 \notin \Lambda$. Moreover, we employ the pole ordering*

$$v(z) = \sum_{j=1}^{\infty} \frac{r_{\lambda_j}}{\lambda_j - z}, \tag{3.11}$$

with $|\lambda_1 - z_0| \leq |\lambda_2 - z_0| \leq \ldots$. If Λ is finite with m elements, then we set formally $\lambda_{m+1} = \lambda_{m+2} = \ldots = \infty$. Otherwise, if the poles are (countably) infinite, we interpret the series in the $\mathcal{V}$-convergence sense as in Assumption 3.1, and we also require the poles Λ to not have any finite limit point, i.e., we assume that $\lim_{j\to\infty} |\lambda_j - z_0| = \infty$. In addition, we ask that the residues $\{r_\lambda\}_{\lambda\in\Lambda}$ form a $\mathcal{V}$-orthogonal family in the sense of Assumption 3.2.

Note that Assumption 3.3 implies Assumption 3.2 over $B^{z_0}(R)$, for all $R > 0$.

Theorem 3.2 (Global pole convergence [Bon+20a, Theorem 5.7])**.**
Let Assumption 3.3 be valid, and consider a sequence $\{(N_i, E_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i \leq$

$E_i \leq E_{i+1}$ for all $i = 2, 3, \ldots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $j = 1, 2, \ldots$, we have

$$\lim_{i\to\infty} \min_{\lambda' : Q^{z_0}_{[M/N_i]}(\lambda')=0} |\lambda' - \lambda_j| = 0, \tag{3.12}$$

where $Q^{z_0}_{[M/N_i]}$ denotes the fast LS Padé denominator of degree N_i, computed from E_i Taylor coefficients of v. (Note that the value of M is irrelevant.)

Proof. See Section 4.1.6. □

Remark 3.2. *In the previous theorem, if the denominator degrees in $\{(N_i, E_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$ are bounded, i.e., $\lim_{i\to\infty} N_i = \bar{N} < \infty$, then the claim still holds for all $j = 1, \ldots, \bar{N}$, provided $|\lambda_{\bar{N}} - z_0| < \left|\lambda_{\bar{N}+1} - z_0\right|$. To show this, it suffices to apply the "fixed N" result, Theorem 3.1.*

Note that, if N_i diverges, we have shown that all exact poles are approximated by surrogate ones, as long as we allow the denominator to have sufficiently large degree. On the other hand, we have no guarantee of the converse, i.e., that all surrogate poles are converging to exact ones. Indeed, in general, spurious poles may appear at arbitrary locations, even in areas where v behaves smoothly. This is a common issue in (diagonal) rational approximation. In the next section, we show that the (possible) appearance of such spurious effects does not disrupt the good approximation properties of the rational approximant.

3.1.2 Error convergence

Also throughout the present section, we assume that Assumption 3.1 holds. We shift our focus from the fast LS Padé denominator $Q^{z_0}_{[M/N]}$ to the fast LS Padé approximant itself, with the objective of showing convergence of $v^{z_0}_{[M/N]}$ to v. Due to the (assumed) meromorphic structure of v, this, in general, will not be possible uniformly with respect to $z \in B^{z_0}(R)$. Instead, it proves necessary to show convergence in a weaker sense, to avoid the singular points of v. See Theorem 2.4 for a classical result with the same limitation.

Theorem 3.3 (Error convergence [Bon+20a, Lemma 6.1])**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq \min\{N^{z_0}(R), M+1\}$. Also, let $\bar{R}_N$ be as in Lemma 3.1, and consider the punctured domain $B_N = B^{z_0}(\bar{R}_N) \setminus \{\lambda_j\}_{j=1}^{N}$. Then, if $v^{z_0}_{[M/N]}$ is the $[M/N]$ fast LS Padé approximant computed with $E = \max\{M, N\}$,

$$\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq \frac{C}{d(z)\left|Q^{z_0}_{[M/N]}(z)\right|} \left(\frac{|z - z_0|}{\bar{R}_N}\right)^E \qquad \forall z \in B_N, \tag{3.13}$$

with C independent of M, E and z, and

$$d(z) = \min_{z' \in \mathbb{C}\setminus B_N} |z - z'| = \min\left\{\bar{R}_N - |z - z_0|, |z - \lambda_1|, \ldots, \left|z - \lambda_{N^{z_0}(R)}\right|\right\}.$$

Additionally, let B' be an arbitrary compact subset of B_N. We have uniform exponential convergence over B' for fixed *N:*

$$\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq C_{B'} \left(\frac{\max_{z\in B'} |z - z_0|}{\bar{R}_N}\right)^E \qquad \forall z \in B', \text{ for large } E, \tag{3.14}$$

with $C_{B'}$ independent of E but, notably, dependent on N and B'.

Proof. See Section 4.1.7. □

Note that, as in Theorem 2.4, we are excluding the exact poles from the "convergence set" B_N, since it is not useful to talk of convergence there. On the other hand, the term $d(z)$ encodes the distance from z to the exterior of B_N, with $d(z) = |z - \lambda|$ in a neighborhood of each of the N most relevant poles, denoted by λ. This shows that the error bound (3.13) diverges with order 1 as z approaches an exact (relevant) pole. A similar diverging behavior can be observed when z approaches the surrogate poles, i.e., the roots of $Q^{z_0}_{[M/N]}$. This is to be expected, since the surrogate itself is unbounded there.

Before proceeding further, we remark that the uniform bound (3.14) is the natural extension of the maximal convergence result from Theorem 2.4, showing that fast LS Padé approximants converge maximally over disks around the center of approximation. Notably, fast LS Padé approximants achieve this result relying on N less Taylor coefficients of v than standard rational (Padé) approximants.

Like for poles, the (exponential) convergence result for fixed N can be extended to variable N. As in the previous section, we can obtain a general convergence result, under global assumptions on v. Unfortunately, if N diverges, it is impossible to prove uniform convergence over compact subsets of the punctured domain B_N, mainly due to the possible appearance of spurious poles even in regions where v is regular. Instead, the convergence must be stated in a weaker sense.

Theorem 3.4 (Global error convergence [Bon+20a, Theorem 6.3] (extended))**.**
Let Assumption 3.3 be valid, and consider a sequence $\{(M_i, N_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i \leq M_i + 1 \leq M_{i+1} + 1$ for all $i = 2, 3, \ldots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $R > 0$ and $\varepsilon > 0$, we have

$$\lim_{i\to\infty} \mathrm{Cap}\left(\left\{z \in B^{z_0}(R) : \left\|v^{z_0}_{[M_i/N_i]}(z) - v(z)\right\|_{\mathcal{V}} > \varepsilon^{N_i}\right\}\right) = 0, \tag{3.15}$$

where $v^{z_0}_{[M_i/N_i]}$ is the $[M_i/N_i]$ fast LS Padé approximant computed from $E_i = \max\{M_i, N_i\}$ Taylor coefficients of v and Cap *is the logarithmic capacity, see Definition 2.2.*

Proof. See Section 4.1.8. □

Remark 3.3. *In the previous theorem, if the denominator degrees in $\{(M_i, N_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$ are bounded, i.e., $\lim_{i\to\infty} N_i = \bar{N} < \infty$, then the claim still holds. To see this, it suffices to use the "fixed N" result, Theorem 3.3, together with the lemniscate argument in the proof of Theorem 3.4.*

This, being a sort of generalization of Theorem 2.5, is a sharp result, in the sense that convergence in a stronger metric cannot be guaranteed. Some counterexamples to convergence in a stronger metric (in the classical setting) can be found in [BGM96; Lub03].

3.1.3 Extension to the non-orthogonal case

Most, if not all, of the convergence results for fast LS Padé approximation described until now rely quite heavily on the assumption of orthogonal residues, encoded by Assumption 3.2, mainly through the extremely powerful expansion (4.6) given in Section 4.1.2. In a qualitative way, poles

N	λ'_1 ($\|r'_1\|_2$)	λ'_2 ($\|r'_2\|_2$)	λ'_3 ($\|r'_3\|_2$)	λ'_4 ($\|r'_4\|_2$)	λ'_5 ($\|r'_5\|_2$)
1	+1.4939-0.2027i (+2.3333e+00)				
2	+1.0148-0.0127i (+1.1203e+00)	+0.0825-2.3462i (+2.3205e+00)			
3	+1.0001-0.0000i (+1.1422e+00)	-0.1070-1.3828i (+7.2808e-01)	+3.3077-0.0055i (+1.2333e+01)		
4	+1.0000-0.0000i (+1.1333e+00)	+0.0413-1.4819i (+1.0042e+00)	+3.0000-0.0000i (+1.1339e+02)	+4.0000-0.0000i (+1.9485e+02)	
5	+1.0000+0.0000i (+1.0197e+00)	+0.0391-1.5167i (+5.9199e-01)	-1.5355+1.0307i (+1.2842e+00)	+3.0000+0.0000i (+8.3797e+01)	+4.0000-0.0000i (+1.5516e+02)

Table 3.1 – Roots of $Q^0_{[N/N]}$ for an artificial 4D v with 2 interfering residues. The norms of the corresponding surrogate residues are also included.

corresponding to orthogonal residues are easier to identify through J_E, since the different residues do not interfere with each other: in some sense, the only source of noise when building fast LS Padé approximants comes from the "irrelevant" poles, i.e., those that, being located outside the convergence region, are not targeted by the approximation effort. And, since this noise decreases exponentially in E, cf. the proof of Lemma 3.1, good approximation properties can be deduced.

However, residual orthogonality is a quite restrictive assumption, essentially requiring the symmetry/self-adjointness of the underlying problem, whose state is v. As such, it makes sense to ask if our convergence results can be extended to the non-orthogonal case. In trying to answer this question, one important limitation should be kept in mind: fast LS Padé approximants are incapable of distinguishing collinear residues, due to the definition of optimization problem (3.4d). We illustrate this with three simple examples. (Note that the code used to obtain the results below is publicly available as part of [Pra21].)

Let $\mathcal{V} = \mathbb{C}^4$ with the usual Euclidean inner product $\langle\cdot,\cdot\rangle_2$ and norm $\|\cdot\|_2$. We define

$$v(z) = \sum_{j=1}^{4} \frac{\mathbf{e}_j}{j - z} + \frac{\mathbf{e}_2}{3\mathrm{i}/2 - z},$$

where $\mathbf{e}_j \in \mathcal{V}$ is the j-th element of the canonical basis, i.e., $(\mathbf{e}_j)_i = \delta_{ij}$. We compute $[N/N]$ fast LS Padé approximants of v centered at 0, using N Taylor coefficients of v, for $N = 1, \ldots, 5$. We show in Table 3.1 the surrogate poles. We can observe that the poles $\lambda \in \{1, 3, 4\}$ are well approximated for N large enough, since, after all, they are separated from the rest of the spectrum. On the other hand, the remaining poles at $\lambda \in \{2, \frac{3}{2}\mathrm{i}\}$ are not properly identified by the fast LS Padé approximants. This will be the case, no matter how large we make N (or E). As such, the region of convergence, cf. Theorem 3.3, will necessarily have to stop at a distance of (at most) 1.5 from the center $z_0 = 0$, even though we have convergence of some poles outside it.

On a side note, we remark that the lack of linear independence also impacts the approximation quality at the poles that are properly identified. Indeed, in Table 3.1 we can also see the norms of the residues corresponding to each surrogate pole, computed as

$$r'_j = \lim_{z \to \lambda'_j} (\lambda'_j - z) v^{z_0}_{[M/N]}(z) = -P^{z_0}_{[M/N]}(\lambda'_j) \Bigg/ \left(\frac{\mathrm{d}}{\mathrm{d}z} Q^{z_0}_{[M/N]} \Big|_{\lambda'_j} \right).$$

By definition of v, all exact residues equal 1, but the numerical ones appear incorrect, particularly for poles 3 and 4, which lie outside the effective convergence region. As N (or E) increase, we can expect only the residue corresponding to $\lambda = 1$ to converge to the exact value.

N	λ'_1 ($\|r'_1\|_2$)	λ'_2 ($\|r'_2\|_2$)	λ'_3 ($\|r'_3\|_2$)	λ'_4 ($\|r'_4\|_2$)	λ'_5 ($\|r'_5\|_2$)
1	+5.5855-5.2388i (+4.5555e+01)				
2	+0.8916+0.0541i (+7.9284e-01)	-0.2640-2.0161i (+3.1099e+00)			
3	+0.9748-0.1424i (+1.8885e+00)	+1.2706+0.4407i (+4.6792e+00)	-0.0122-1.6920i (+2.2926e+00)		
4	+0.9554+0.0466i (+1.4828e+00)	+1.2458-0.6219i (+6.2230e+00)	+0.0518-1.5713i (+2.0335e+00)	+1.5141+1.1144i (+1.2201e+01)	
5	+0.9770+0.0209i (+8.9124e-01)	-1.3339+0.3563i (+6.0324e-01)	+1.4303-0.5989i (+6.8311e+00)	+0.0077-1.5626i (+1.5906e+00)	+1.6326+0.9452i (+1.1857e+01)

Table 3.2 – Roots of $Q^0_{[N/N]}$ for an artificial 4D v with interference between all residues. The norms of the corresponding surrogate residues are also included.

N	λ'_1 ($\|r'_1\|_2$)	λ'_2 ($\|r'_2\|_2$)	λ'_3 ($\|r'_3\|_2$)	λ'_4 ($\|r'_4\|_2$)	λ'_5 ($\|r'_5\|_2$)
1	+1.6457+0.0000i (+4.0537e+00)				
2	+1.0307+0.0000i (+1.1686e+00)	+2.9094+0.0000i (+4.5077e+00)			
3	+1.0005-0.0000i (+1.0035e+00)	+2.0703+0.0000i (+1.7533e+00)	+3.6249-0.0000i (+3.2830e+00)		
4	+1.0000-0.0000i (+1.0000e+00)	+2.0000+0.0000i (+1.4142e+00)	+3.0000-0.0000i (+1.7321e+00)	+4.0000+0.0000i (+2.0000e+00)	
5	+1.0000+0.0000i (+1.0000e+00)	-1.1270+0.0000i (+2.2012e-15)	+2.0000+0.0000i (+1.4142e+00)	+3.0000+0.0000i (+1.7321e+00)	+4.0000+0.0000i (+2.0000e+00)

Table 3.3 – Roots of $Q^0_{[N/N]}$ for an artificial 4D v with linearly independent residues. The norms of the corresponding surrogate residues are also included.

Now we increase the amount of "interference" between residues by setting

$$v(z) = \sum_{j=1}^{4} \frac{\mathbf{e}_j}{j - z} + \frac{\sum_{j=1}^{4} \mathbf{e}_j}{3\mathrm{i}/2 - z},$$

so that the residue at $3\mathrm{i}/2$ is not orthogonal to any of the others. In this case, the results are even worse, as can be seen in Table 3.2. The "collinearity noise" affects all poles, and the approximation quality degrades as a consequence: only $\lambda = 1$ seems to be identified, and not so well at that.

As a final note, we observe that, in contrast, had the residues been orthogonal (or even just linearly independent), fast LS Padé approximants would have recovered the *exact* poles and residues of v starting from $N = \dim(\mathcal{V})$. We show this by building fast LS Padé approximants for

$$v(z) = \sum_{j=1}^{4} \frac{\sum_{k=1}^{j} \mathbf{e}_k}{j - z},$$

whose 4 residues are linearly independent, albeit not orthogonal. The surrogate poles and residues are shown in Table 3.3. Notably, we see the appearance of a Froissart doublet, i.e., a pole with (numerically) vanishing residue, for $N = 5$. Other than that, the poles and residues are identified exactly ($\|r_j\|_2 = \sqrt{j}$), as predicted.

These simple examples show that the lack of linear independence in the residues can have ruinous effects on fast LS Padé approximants, making a weakened version of Assumption 3.2 necessary to hope for any convergence result similar to those of the previous section. Such extensions are presented (without proof) in [Pra20] for the "fixed N" results from the previous sections, namely, Theorems 3.1 and 3.3. We summarize here the main claims in the form of a conjecture.

Conjecture 1 ([Pra20])**.** *Let Assumption 3.1 hold for all R, with the spectrum Λ having no finite*

limit points, cf. Assumption 3.3. Moreover, assume that the residues $\{r_\lambda\}_{\lambda\in\Lambda}$ *form a linearly independent family, i.e., that, if* $\left\|\sum_{\lambda\in\Lambda}\alpha_\lambda r_\lambda\right\|_{\mathcal{V}} = 0$ *for some sequence* $\{\alpha_\lambda\}_{\lambda\in\Lambda}\subset\mathbb{C}$*, then* $\alpha_\lambda = 0$ *for all* λ*. Then, we expect Theorem 3.1 to hold, with a larger constant* C_j *and with halved exponent (this is in accordance to the common decrease in convergence rate for non-symmetric and non-Hermitian eigensolvers). Also, we expect Theorem 3.3 to hold, with larger constants* C *and* C_A*.*

We perform some numerical verifications of these claims in Section 5.5.

3.2 The general MRI algorithm

The minimal rational interpolation (MRI) algorithm is born as an extension of fast LS Padé approximation, by allowing the sample points to be distinct, distributed over a region of $\mathbb{C}$, rather than concentrated at a single point z_0. In fact, MRI allows coalescence of the sample points, so that fast LS Padé approximation (with $E = M$) is a special case of MRI.

From a different viewpoint, MRI is also a generalization of scalar Lagrange(-Hermite) rational interpolation, see Definition 2.5, to the high-dimensional setting, providing an alternative to the LS approaches described in Section 2.4.1. In particular, the term "minimal" in the name of MRI reveals that, in this technique, we strive to exploit the samples of v as much as possible. More specifically, we will see that a rational approximation of type $[S-1/S-1]$ can be built starting from S snapshots, whereas, with the classical interpolation approaches described in Section 2.4.1, building a rational surrogate of the same type requires at least $2S-1$ snapshots.

The definition of MRI, as given by the thesis author in [Pra20], follows.

Definition 3.2 (Minimal rational interpolation [Pra20, Definition 2.1])**.** *Let* $\mathcal{V}$ *be as in Definition 3.1. We consider sample points* $Z = \{z_j\}_{j=1}^S \in \mathbb{C}$ *(not necessarily distinct) and an integer* $N \geq 0$*, with* $N+1 \leq S$*. We require a basis* $\Psi_N = \{\psi_i\}_{i=0}^N$ *of the polynomial space* $\mathbb{P}_N(\mathbb{C};\mathbb{C})$*. Also, let* $v : \mathbb{C} \to \mathcal{V}$ *be a* $\mathcal{V}$*-valued function such that* $I^Z(v)$ *exists, according to Definition 2.1. An* $[S-1/N]$ *MRI of* v *based on samples at* Z *(dependent on* Ψ_N*) is a rational function*

$$v^Z_{[S-1/N]} = \frac{I^Z(Q^Z_{[S-1/N]}v)}{Q^Z_{[S-1/N]}}, \tag{3.16}$$

such that

$$Q^Z_{[S-1/N]} \in \mathbb{P}_N^{\Psi_N}(\mathbb{C};\mathbb{C}) = \left\{\sum_{n=0}^N q_n\psi_n : \{q_n\}_{n=0}^N \subset \mathbb{C}, \sum_{n=0}^N |q_n|^2 = 1\right\}, \tag{3.17a}$$

$$J^Z(Q^Z_{[S-1/N]}) \leq J^Z(Q) := \left\|\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv)\right\|_{\mathcal{V}} \quad \forall Q \in \mathbb{P}_N^{\Psi_N}(\mathbb{C};\mathbb{C}). \tag{3.17b}$$

As already mentioned, MRI includes fast LS Padé approximation as a special case, since I^Z coincides with a truncated Taylor series when $Z = \{z_0,\dots,z_0\}$, and the basis Ψ_N can be chosen as the shifted monomials $\{(\cdot - z_0)^i\}_{i=0}^N$, cf. Definition 3.1. Note that, as for fast LS Padé approximants, uniqueness cannot in general be guaranteed. Moreover, it is interesting to observe that the term $\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(\phi)$ in (3.17b) corresponds to (a multiple of) the leading coefficient of the interpolant of ϕ, extending the equivalent term $\frac{\mathrm{d}^E\phi}{\mathrm{d}z^E}(z_0)$ in fast LS Padé approximation, see

(3.4d).

The normalization condition (3.17a) accounts for the MRI denominator $Q^Z_{[S-1/N]}$ being expressed in the general polynomial basis Ψ_N, whose choice will, in practice, depend on how the sample points are distributed. The role that Ψ_N plays in the numerical properties of MRI is discussed in Section 5.2.3.

All the theory for fast LS Padé approximation can be generalized to MRI, under similar assumptions on the target function v. However, before presenting the main results, we note that, with respect to fast LS Padé approximation, in MRI we have the added difficulty due to the Ψ_N-based normalization. For most of our results, namely, those for fixed N, no issues due to this normalization arise. However, when investigating the convergence of MRI as $N \to \infty$, we will need to postulate that Ψ_N behaves nicely *uniformly in* N. More specifically, we formalize this by requiring the existence of uniform (in N) upper and lower bounds for the value of normalized polynomials $Q \in \mathbb{P}^{\Psi_N}_N(\mathbb{C};\mathbb{C})$.

Assumption 3.4. *Let $z_0 \in \mathbb{C}$ be fixed. For all $N \in \{1,2,\dots\}$, let $\mathbb{P}^{\Psi_N,z_0}_N(\mathbb{C};\mathbb{C})$ be the set of polynomials $Q \in \mathbb{P}^{\Psi_N}_N(\mathbb{C};\mathbb{C})$, such that $Q(z_0) \neq 0$. There exist positive constants ρ^{z_0}, c^{z_0}, and C^{z_0} (independent of z and N) such that*

$$(c^{z_0})^N \prod_{j=1}^{N'} \frac{|z-z_j|}{\rho^{z_0}+|z_0-z_j|} \leq |Q(z)| \leq (C^{z_0})^N \prod_{j=1}^{N'} \left|\frac{z-z_j}{z_0-z_j}\right| \qquad \forall Q \in \mathbb{P}^{\Psi_N,z_0}_N(\mathbb{C};\mathbb{C}) \quad \forall z \in \mathbb{C} \quad \forall N \in \mathbb{N}, \quad (3.18)$$

where $\{z_j\}_{j=1}^{N'}$ are the roots of Q, repeated according to multiplicity.

We refer to the results proven in Sections 4.1.1 and 4.2.1 for a motivation behind the specific structure of the bounds in the assumption above. Note that, in [Pra20], a stronger assumption is considered: namely, both c^{z_0} and C^{z_0} must appear *without the exponent* N in (3.18). However, it turns out that the weaker Assumption 3.4 is sufficent to develop our theory, cf. Sections 4.2.6 and 4.2.8.

One can prove that some widely used polynomial bases satisfy Assumption 3.4. Among those, we can find the (scaled and shifted) monomial basis $\Psi_N = \{\left(\frac{\cdot-z_0}{R}\right)^i\}_{i=0}^N$ (see Lemma 4.1 for a proof). Moreover, the (scaled and shifted) Chebyshev and Legendre polynomials ($\Psi_N = \{T_i\left(\frac{\cdot-z_0}{R}\right)\}_{i=0}^N$ and $\Psi_N = \{L_i\left(\frac{\cdot-z_0}{R}\right)\}_{i=0}^N$, respectively) satisfy Assumption 3.4 for some choices of z_0. We refer to Appendix A for a proof of this fact.

3.2.1 Pole convergence

With the objective of extending the results on pole convergence from Section 3.1.1, we redefine "pole relevance" to account for the different sampling scheme. To this aim, we must first fix a "sampling set" $A \Subset \mathbb{C}$, with the convention that we are allowed to take samples of v only at points of ∂A, namely, at Fekete points of A. Note that this is required only in our theoretical analysis, and should not be seen as a constraint on the applicability of the MRI method.

In fast LS Padé approximation, pole relevance was set based on how distant each pole of v was from $A = \{z_0\}$. When A is not a single point, it turns out that the best way to extend this concept is through the Green's potential Φ_A, see Definition 2.3, whenever Φ_A is well-defined.

Accordingly, we generalize Assumption 3.1 as follows.

Assumption 3.5 (Local simple partial fraction expansion)**.** *Consider a sampling set $A \Subset \mathbb{C}$ with Green's potential Φ_A. Given a fixed $R > \mathrm{Cap}(A)$, let A_R be the conformal extension of A, according to Definition 2.3. We assume that v admits the simple-pole partial fraction expansion* (3.1) *over A_R, with distinct poles and $\partial A \cap \Lambda = \emptyset$ (so as to avoid sampling at a pole). If the poles of v are (countably) infinite, we require the partial fraction expansion to be an absolutely convergent series in the $\mathcal{V}$-sense, so that, in particular,*

$$\|v(z)\|_{\mathcal{V}} \leq \sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z|} < \infty \qquad \forall z \in A_R \setminus \Lambda.$$

Moreover, let $\Lambda^A(R) = \Lambda \cap A_R$ be the poles of v inside A_R. We assume that $\Lambda^A(R)$ has a finite number of elements $N^A(R)$. In particular, we employ the pole ordering

$$v(z) = \sum_{j=1}^{N^A(R)} \frac{r_{\lambda_j}}{\lambda_j - z} + \sum_{\lambda \in \Lambda \setminus \Lambda^A(R)} \frac{r_\lambda}{\lambda - z}, \tag{3.19}$$

with $\Phi_A(\lambda_1) \leq \Phi_A(\lambda_2) \leq \ldots \leq \Phi_A(\lambda_{N^A(R)}) < R$.

We note that poles inside A can be ordered in an arbitrary way, since $\Phi_A(\lambda) = \mathrm{Cap}(A)$ for all $\lambda \in A$. An obvious generalization of Assumption 3.2 is also available.

Assumption 3.6 (Orthogonal partial fraction residues)**.** *Assumption 3.5 holds. The residues $\{r_\lambda\}_{\lambda \in \Lambda}$ form a $\mathcal{V}$-orthogonal family, i.e., $\langle r_\lambda, r_{\lambda'} \rangle_{\mathcal{V}}$ for all $\lambda \neq \lambda'$.*

Now we are ready to state our results, which we present in the same order as for fast LS Padé approximation. We start with the convergence of the MRI denominator values for fixed N.

Lemma 3.2 (Denominator value at poles [Pra20, Lemma 3.6])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A, and let Assumption 3.6 be valid over A_R. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S (we remind the reader that Z_S has S elements, cf. Theorem 2.2). Take a fixed $N^A(\mathrm{Cap}(A)) \leq N \leq N^A(R)$. Also, let $\bar{R}_N = R$ if $N \geq N^A(R)$ and $\bar{R}_N = \Phi_A(\lambda_{N+1})$ otherwise. For all $j = 1, \ldots, N$, $\rho < \bar{R}_N$, and S large enough (depending on ρ and N), we have the bound

$$\left| Q^{Z_S}_{[S-1/N]}(\lambda_j) \right| \leq C_j \left(\frac{\Phi_A(\lambda_j)}{\rho} \right)^{2S}, \tag{3.20}$$

with C_j independent of M and S.

Proof. See Section 4.2.4. □

From here, we can get to more "practical" results, starting from pole convergence for fixed N.

Theorem 3.5 (Pole convergence [Pra20, Theorem 3.7])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A, and let Assumption 3.6 be valid over A_R. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S. Take a fixed *$N \leq N^A(R)$. Also, set $\bar{R}_N = R$ if $N = N^A(R)$ and $\bar{R}_N = \Phi_A(\lambda_{N+1})$ otherwise. Assume that*

$\Phi_A(\lambda_N) < \rho < \bar{R}_N$, *with* ρ *arbitrary. Then, for all* $j = 1, \dots, N$ *and* S *large enough (depending on* ρ *and* N*),*

$$\min_{\lambda' : Q^{Z_S}_{[S-1/N]}(\lambda')=0} |\lambda' - \lambda_j| \leq C_j \left(\frac{\Phi_A(\lambda_j)}{\rho} \right)^{2S}, \tag{3.21}$$

with C_j *independent of* S*.*

Proof. See Section 4.2.5. □

As could have been expected, we see that the Green's potential of A determines the convergence rate. Note, in particular, that all poles within A can be expected to converge at the same rate, since $\Phi_A(\lambda) = \mathrm{Cap}(A)$ for all $\lambda \in A$.

In the following result, we generalize the global pole convergence of Theorem 3.2. We note that, since we are showing convergence of all poles, ordering the poles using the Green's potential is unnecessary.

Theorem 3.6 (Global pole convergence [Pra20, Theorem 3.8])**.**
Let $A \Subset \mathbb{C}$ *be the sampling set, with Fekete points* $\{Z_S\}_{S=1}^{\infty}$*, with* $Z_S \subset \partial A$ *for all* S*. Let Assumptions 3.3 and 3.4 be valid for some* $z_0 \in \mathbb{C}$*, and consider a sequence* $\{(N_i, S_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$*, such that* $N_{i-1} \leq N_i < S_i \leq S_{i+1}$ *for all* $i = 2, 3, \dots$*. Further, assume that* $\lim_{i\to\infty} N_i = \infty$*, i.e., both the number of snapshots and the denominator degree diverge. For all* $j = 1, 2 \dots$*, we have*

$$\lim_{i\to\infty} \min_{\lambda' : Q^{Z_{S_i}}_{[S_i - 1/N_i]}(\lambda')=0} |\lambda' - \lambda_j| = 0. \tag{3.22}$$

Proof. See Section 4.2.6. □

3.2.2 Error convergence

Next comes the error convergence, whose rate is, again, determined by the Green's potential of A. Note that, once more, we are relying heavily on Assumption 3.5.

Theorem 3.7 (Error convergence [Pra20, Theorem 3.9])**.**
Let $A \Subset \mathbb{C}$ *be the sampling set, with Green's potential* Φ_A*, and let Assumption 3.6 be valid over* A_R*. Take a sequence of Fekete points* $\{Z_S\}_{S=1}^{\infty}$*, with* $Z_S \subset \partial A$ *for all* S *(we remind the reader that* Z_S *has* S *elements, cf. Theorem 2.2). Take a* fixed $N \leq N^A(R)$*. Also, let* $\bar{R}_N$ *be as in Lemma 3.2, and consider an arbitrary* $\mathrm{Cap}(A) < \rho < \bar{R}_N$*. Define the punctured domain* B_N *as the interior of* $A_\rho \setminus \{\lambda_j\}_{j=1}^N$*. Then, for all* $\varepsilon > 0$ *and* S *large enough (depending on* N *and* ε*),*

$$\left\| v^{Z_S}_{[S-1/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq \frac{C}{d(z) \left| Q^{Z_S}_{[S-1/N]}(z) \right|} \left(\frac{\Phi_A(z)}{\rho} \right)^S \qquad \forall z \in B_N, \tag{3.23}$$

with C *independent of* S*, and* $d(z) = \min_{z' \in \mathbb{C} \setminus B_N} |z - z'|$*.*

Additionally, let B' *be an arbitrary compact subset of* B_N*. We have uniform exponential convergence over* B' for fixed N*:*

$$\left\| v^{Z_S}_{[S-1/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq C_{B'} \left(\frac{\max_{z \in B'} \Phi_A(z)}{\rho} \right)^S \qquad \forall z \in B', \tag{3.24}$$

with $C_{B'}$ independent of S but, notably, dependent on B'.

Proof. See Section 4.2.7. □

Note that the approximation error at all points of $A \setminus \Lambda$ can be expected to converge at the same rate, since $\Phi_A(z) = \mathrm{Cap}(A)$ for all $z \in A$. Moreover, comparing (3.23) with the "maximal convergence" in Theorem 2.4, we conclude that MRI converges maximally under Assumption 3.6.

Finally, we have global convergence in capacity.

Theorem 3.8 (Global error convergence [Pra20, Theorem 3.10] (extended))**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S. Let Assumptions 3.3 and 3.4 be valid for some $z_0 \in \mathbb{C}$, and consider a sequence $\{(N_i, S_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i < S_i \leq S_{i+1}$ for all $i = 2, 3, \ldots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $R > \mathrm{Cap}(A)$ and $\varepsilon > 0$, we have

$$\lim_{i\to\infty} \mathrm{Cap}\left(\left\{z \in A_R : \left\| v^{Z_{S_i}}_{[S_i - 1/N_i]}(z) - v(z) \right\|_{\mathcal{V}} > \varepsilon^{N_i}\right\}\right) = 0, \tag{3.25}$$

where Cap *is the logarithmic capacity, see Definition 2.2.*

Proof. See Section 4.2.8. □

As a final note, we remark that the considerations on residue collinearity presented in Section 3.1.3 for fast LS Padé approximation apply also to MRI.

4 Proofs of convergence results for MRI

In this chapter we report all the proofs of the results from Sections 3.1 and 3.2. For convenience, we split them depending on whether they pertain to fast LS Padé approximation or MRI.

4.1 Proofs of results for the single-point case

We start from the results related to fast LS Padé approximation. The main reference throughout this chapter is [Bon+20a].

4.1.1 Auxiliary result: bounds for normalized polynomials

First, it is useful to explore some properties of the set $\mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$, where $Q_{[M/N]}^{z_0}$ is sought. Notably, we can show some bounds on values of normalized polynomials.

Lemma 4.1 (Normalization of nodal polynomials [Bon+20a, Lemma 5.1])**.**
Let $Q \in \mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$. Then

$$|Q(z)| \leq \left(\sum_{n=0}^{N} |z-z_0|^{2n}\right)^{1/2} = \begin{cases} \sqrt{N+1} & \text{if } |z-z_0| = 1, \\ \left(\frac{|z-z_0|^{2N+2}-1}{|z-z_0|^2-1}\right)^{1/2} & \text{if } |z-z_0| \neq 1. \end{cases} \tag{4.1}$$

Moreover, assume that Q has exact degree $N' \leq N$, with roots $\{z_j\}_{j=1}^{N'}$ (repeated according to multiplicity), all different from z_0. Then,

$$\prod_{j=1}^{N'} \frac{|z-z_j|}{1+|z_0-z_j|} \leq |Q(z)| \leq \prod_{j=1}^{N'} \left|\frac{z-z_j}{z_0-z_j}\right| \leq \prod_{j=1}^{N'} \left(1+\left|\frac{z-z_0}{z_0-z_j}\right|\right). \tag{4.2}$$

Remark 4.1. *In the second part of the lemma, one could equivalently (formally) set the missing $N-N'$ roots to ∞.*

Proof. The first claim follows easily by the Cauchy-Schwarz inequality:

$$|Q(z)| = \left|\sum_{n=0}^{N} \frac{1}{n!} \frac{\mathrm{d}^n Q}{\mathrm{d}z^n}(z_0)(z-z_0)^n\right| \leq$$

$$\leq \left(\sum_{n=0}^{N} \left| \frac{1}{n!} \frac{\mathrm{d}^n Q}{\mathrm{d}z^n}(z_0) \right|^2 \right)^{1/2} \left(\sum_{n=0}^{N} |z - z_0|^{2n}\right)^{1/2} = \left(\sum_{n=0}^{N} |z - z_0|^{2n}\right)^{1/2}.$$

Given $Q \in \mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$ with roots $\{z_j\}_{j=1}^{N'}$, let $\omega(z) = \prod_{j=1}^{N'}(z - z_j)$, so that $Q = \tau\omega$ for some $\tau \in \mathbb{C} \setminus \{0\}$. By the Hadamard multiplication theorem [Tit78, Section 4.6], we have the identity

$$\begin{aligned} |\tau|^{-2} = |\tau|^{-2} \sum_{i=0}^{N} |q_i|^2 = |\tau|^{-2} \sum_{i=0}^{N} \left| \frac{1}{i!} \frac{\mathrm{d}^i Q}{\mathrm{d}z^i}(z_0) \right|^2 &= |\tau|^{-2} \int_0^1 \left| Q(z_0 + e^{2\pi \mathrm{i}\theta}) \right|^2 \mathrm{d}\theta \\ &= \int_0^1 \prod_{j=1}^{N'} \left| z_0 + e^{2\pi\mathrm{i}\theta} - z_j \right|^2 \mathrm{d}\theta. \end{aligned}$$

On one hand, by the triangular inequality, we have

$$|\tau|^{-2} \leq \int_0^1 \prod_{j=1}^{N'} \left(|z_0 - z_j| + \left|e^{2\pi\mathrm{i}\theta}\right| \right)^2 \mathrm{d}\theta = \prod_{j=1}^{N'} \left(|z_0 - z_j| + 1 \right)^2.$$

The lower bound in (4.2) follows.

On the other hand, by the Cauchy-Schwarz inequality in $L^2((0,1))$, we have

$$\begin{aligned} |\tau|^{-2} &\geq \left| \int_0^1 \prod_{j=1}^{N'} \left(z_0 + e^{2\pi\mathrm{i}\theta} - z_j \right) \mathrm{d}\theta \right|^2 \\ &= \left| \int_0^1 \left(\prod_{j=1}^{N'} (z_0 - z_j) + \sum_{j=1}^{N'} c_j e^{2\pi \mathrm{i} j \theta} \right) \mathrm{d}\theta \right|^2 \\ &= \left| \prod_{j=1}^{N'} (z_0 - z_j) + \sum_{j=1}^{N'} c_j \int_0^1 e^{2\pi \mathrm{i} j \theta} \mathrm{d}\theta \right|^2, \end{aligned}$$

for some θ-independent coefficients $\{c_j\}_{j=1}^{N'}$. Since $\int_0^1 e^{2\pi \mathrm{i} j\theta} \mathrm{d}\theta = 0$ for all integers $j \geq 1$, we have $|\tau|^{-2} \geq \prod_{j=1}^{N'} |z_0 - z_j|$, which implies the first upper bound in (4.2). The second upper bound follows by applying the triangular inequality within each term of the product: $|z - z_j| \leq |z_0 - z_j| + |z - z_0|$.

We note that, alternatively, the second half of the claim could have been proven via Lemmas A.1 and A.2, since monomials are orthogonal over the unit circle (with unit weight). □

Now, let $g_N^{z_0}$ be defined as in Section 3.1.1, i.e., as an element of $\mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C})$ such that $g_N^{z_0}(\lambda_j) = 0$ for all $j = 1, \ldots, N$. By applying (4.2) to $g_N^{z_0}$, we obtain (note that z_0 is not a root of $g_N^{z_0}$ by Assumption 3.1)

$$\prod_{j=1}^{N} \frac{|z - \lambda_j|}{1 + |z_0 - \lambda_j|} \leq |g_N^{z_0}(z)| \leq \prod_{j=1}^{N} \left| \frac{z - \lambda_j}{z_0 - \lambda_j} \right| \leq \prod_{j=1}^{N} \left(1 + \left| \frac{z - z_0}{z_0 - \lambda_j} \right| \right). \tag{4.3}$$

4.1.2 Auxiliary result: alternative expressions of target functional

We can obtain alternative expressions for the quantity J_E in (3.4d).

Lemma 4.2 (Alternative expressions of J_E [Bon+20a, Section 5] and [Pra20, Section 3.4])**.** *Let Assumption 3.1 be valid over $B^{z_0}(R)$ for small enough R, and take an arbitrary $Q \in \mathbb{P}_N(\mathbb{C};\mathbb{C})$. Then,*

$$\frac{1}{E!}J_E(Q) = \left(\sum_{\lambda,\lambda' \in \Lambda} \frac{\langle r_\lambda, r_{\lambda'} \rangle_{\mathcal{V}}}{(\lambda - z_0)^{E+1}\overline{(\lambda' - z_0)}^{E+1}} Q(\lambda)\overline{Q(\lambda')} \right)^{1/2}, \tag{4.4}$$

and

$$\frac{1}{E!}J_E(Q) \leq \sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|^{E+1}} |Q(\lambda)|. \tag{4.5}$$

If, in addition, Assumption 3.2 holds, then

$$\frac{1}{E!}J_E(Q) = \left(\sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} |Q(\lambda)|^2 \right)^{1/2}. \tag{4.6}$$

Proof. Since the derivatives of v are

$$\frac{\mathrm{d}^n v}{\mathrm{d}z^n}(z_0) = n! \sum_{\lambda \in \Lambda} \frac{r_\lambda}{(\lambda - z_0)^{n+1}}, \tag{4.7}$$

we can apply the Leibniz rule to J_E to obtain

$$\begin{aligned} J_E(Q) = \left\| \sum_{n=0}^{N} \binom{E}{n} \frac{\mathrm{d}^n Q}{\mathrm{d}z^n}(z_0) \frac{\mathrm{d}^{E-n} v}{\mathrm{d}z^{E-n}}(z_0) \right\|_{\mathcal{V}} &= E! \left\| \sum_{\lambda \in \Lambda} \frac{r_\lambda}{(\lambda - z_0)^{E+1}} \sum_{n=0}^{N} \frac{1}{n!} \frac{\mathrm{d}^n Q}{\mathrm{d}z^n}(z_0)(\lambda - z_0)^n \right\|_{\mathcal{V}} \\ &= E! \left\| \sum_{\lambda \in \Lambda} \frac{r_\lambda}{(\lambda - z_0)^{E+1}} Q(\lambda) \right\|_{\mathcal{V}} \end{aligned}$$

(we have implicitly relied on $E \geq N$ in applying the Leibniz rule). The upper bound follows by the triangular inequality, whereas the other claims follow by expanding the $\mathcal{V}$-norm in terms of the $\mathcal{V}$-inner product. □

4.1.3 Auxiliary result: optimal value of target functional

We can build a bound for the minimal value of J_E in Definition 3.1, namely, $J_E(Q^{z_0}_{[M/N]})$.

Lemma 4.3 (Minimal value of J_E [Bon+20a, Lemma 5.3])**.** *Let Assumption 3.1 be valid over $B^{z_0}(R)$, and assume that $Q^{z_0}_{[M,N]}$ is computed using Definition 3.1, relying on E Taylor coefficients of v. Then,*

$$\frac{1}{E!}J_E(Q^{z_0}_{[M,N]}) \leq C \left(\bar{R}_N\right)^{-E}, \tag{4.8}$$

where $\bar{R}_N = R$ if $N \geq N^{z_0}(R)$ and $\bar{R}_N = |\lambda_{N+1} - z_0|$ otherwise. The constant C is independent of M and E.

Proof. Let $N_{\text{eff}} = \min\{N, N^{z_0}(R)\}$. We set $\Lambda_{N_{\text{eff}}} = \{\lambda_j\}_{j=1}^{N_{\text{eff}}}$ as the "relevant" poles, which are

the roots of the (effective) target polynomial $g^{z_0}_{N_{\text{eff}}} \in \mathbb{P}^{z_0}_{N_{\text{eff}}}(\mathbb{C};\mathbb{C}) \subset \mathbb{P}^{z_0}_{N}(\mathbb{C};\mathbb{C})$. Moreover, we define the "irrelevant" poles as $\bar{\Lambda} = \Lambda \setminus \Lambda_{N_{\text{eff}}}$. Note that $\min_{\lambda\in\bar{\Lambda}} |\lambda - z_0| \geq \bar{R}_N$.

By (3.4d), $Q^{z_0}_{[M/N]}$ yields a lower value of J_E than $g^{z_0}_{N_{\text{eff}}}$. This, together with (4.3) and (4.5), gives

$$\begin{aligned}\frac{1}{E!}J_E(Q^{z_0}_{[M/N]}) \leq& \frac{1}{E!}J_E(g^{z_0}_{N_{\text{eff}}}) \leq \sum_{\lambda\in\Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|^{E+1}} \left|g^{z_0}_{N_{\text{eff}}}(\lambda)\right| \\ &= \sum_{\lambda\in\bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|^{E+1}} \left|g^{z_0}_{N_{\text{eff}}}(\lambda)\right| \leq \sum_{\lambda\in\bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|^{E+1}} \prod_{j=1}^{N_{\text{eff}}} \left(1 + \left|\frac{\lambda - z_0}{\lambda_j - z_0}\right|\right) \\ &\leq \sup_{\lambda\in\bar{\Lambda}} \left(|\lambda - z_0|^{N_{\text{eff}}-E} \prod_{j=1}^{N_{\text{eff}}} \left(\frac{1}{|\lambda - z_0|} + \frac{1}{|\lambda_j - z_0|}\right)\right) \sum_{\lambda\in\bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|}.\end{aligned}$$

Since $|\lambda_j - z_0| \leq |\lambda - z_0|$ for all $j = 1, \ldots, N_{\text{eff}}$ and $\lambda \in \bar{\Lambda}$, we have

$$\prod_{j=1}^{N_{\text{eff}}} \left(\frac{1}{|\lambda - z_0|} + \frac{1}{|\lambda_j - z_0|}\right) \leq \frac{2^{N_{\text{eff}}}}{\prod_{j=1}^{N_{\text{eff}}} |\lambda_j - z_0|}.$$

Moreover, since $E \geq N \geq N_{\text{eff}}$,

$$\sup_{\lambda\in\bar{\Lambda}} |\lambda - z_0|^{N_{\text{eff}}-E} = \left(\inf_{\lambda\in\bar{\Lambda}} |\lambda - z_0|\right)^{N_{\text{eff}}-E} \leq \left(\bar{R}_N\right)^{N_{\text{eff}}-E}.$$

The claim follows. □

4.1.4 Proof of Lemma 3.1

For convenience, we report a copy of the statement in question, copied from Section 3.1.1.

Lemma 3.1 (Denominator value at poles [Bon+20a, Lemma 5.4] (extended))**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq N^{z_0}(R)$. Assume that $Q^{z_0}_{[M,N]}$ is computed using Definition 3.1, relying on E Taylor coefficients of v. Also, let $\bar{R}_N = R$ if $N \geq N^{z_0}(R)$ and $\bar{R}_N = |\lambda_{N+1} - z_0|$ otherwise. Then, for all $j = 1, \ldots, N$, we have the bound

$$\left|Q^{z_0}_{[M/N]}(\lambda_j)\right| \leq C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N}\right)^{2E}, \tag{3.9}$$

with C_j independent of M (in fact, the value of M is irrelevant here) and E, but dependent on N.

Proof. Based on (4.6), we define the following Hermitian quadratic form over $\mathbb{P}_N(\mathbb{C};\mathbb{C}) \times \mathbb{P}_N(\mathbb{C};\mathbb{C})$:

$$b_E(Q, Q') = \sum_{\lambda\in\Lambda} \frac{\|r_\lambda\|^2_{\mathcal{V}}}{|\lambda - z_0|^{2E+2}} Q(\lambda)\overline{Q'(\lambda)},$$

which is obviously semi-positive definite. By Lemma 4.2, $J_E(Q) = E! b_E(Q, Q)^{1/2}$. Accordingly,

we define the Hermitian eigenproblem

$$\text{find } \sigma \geq 0 \text{ and } Q_\sigma \in \mathbb{P}_N^{z_0}(\mathbb{C};\mathbb{C}) \ : \ b_E(Q_\sigma, Q) = \sigma \langle Q_\sigma, Q\rangle_{z_0} \qquad \forall Q \in \mathbb{P}_N(\mathbb{C};\mathbb{C}). \tag{4.9}$$

Note that we are using the inner product $\langle \cdot, \cdot \rangle_{z_0}$ defined in (3.6).

Thanks to the standard properties of quadratic forms over finite-dimensional spaces, (3.4d) can be reinterpreted as: $Q^{z_0}_{[M/N]}$ is a solution of (4.9) with minimal eigenvalue σ. Moreover, we note that Lemma 4.3 provides an upper bound for (the square root of) such minimal eigenvalue, since

$$b_E(Q^{z_0}_{[M/N]}, Q) = \frac{1}{E!^2} J_E(Q^{z_0}_{[M/N]})^2 \langle Q^{z_0}_{[M/N]}, Q\rangle_{z_0} \qquad \forall Q \in \mathbb{P}_N(\mathbb{C};\mathbb{C}). \tag{4.10}$$

Now, let $\Lambda_N = \{\lambda_j\}_{j=1}^N$ denote the relevant poles. A truncated version of the quadratic form can be similarly defined as

$$b_E^{(N)}(Q, Q') = \sum_{\lambda \in \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} Q(\lambda)\overline{Q'(\lambda)}.$$

Semi-positive definiteness is retained, since

$$b_E^{(N)}(Q, Q) = \left\| \sum_{\lambda \in \Lambda_N} \frac{r_\lambda}{(\lambda - z_0)^{E+1}} Q(\lambda) \right\|_{\mathcal{V}}^2.$$

However, the minimal eigenvalue of $b_E^{(N)}$ is certainly 0, for, since λ_i only ranges over the relevant poles, $b_E^{(N)}(g_N^{z_0}, g_N^{z_0}) = 0$. On the other hand, since, by Assumption 3.2, the relevant residues are linearly independent, there cannot exist a polynomial Q of degree $\leq N$, linearly independent from $g_N^{z_0}$, such that $b_E^{(N)}(Q, Q) = 0$, i.e., $b_E^{(N)}$ has rank exactly N.

For each $j \in \{1, \dots, N\}$, define the monic polynomial $\gamma_j \in \mathbb{P}_{N-1}(\mathbb{C};\mathbb{C})$ as $\gamma_j(z) = \prod_{i=1, i\neq j}^N (z - \lambda_i)$. The set $\{g_N^{z_0}\} \cup \{\gamma_j\}_{j=1}^N$ is a basis of $\mathbb{P}_N(\mathbb{C};\mathbb{C})$, as can be deduced by direct inspection of the roots of its elements. Accordingly, we may write $Q^{z_0}_{[M/N]} = \alpha_0 g_N^{z_0} + \sum_{j=1}^N \alpha_j \gamma_j$, and, for $j = 1, \dots, N$,

$$\left| Q^{z_0}_{[M/N]}(\lambda_j) \right| = |\alpha_j \gamma_j(\lambda_j)| \leq |\alpha_j| \prod_{\substack{i=1 \\ i\neq j}}^N |\lambda_j - \lambda_i| \,.$$

Thus, it only remains to bound $|\alpha_j|$ from above, for each j. By construction and by semilinearity, for all $j = 1, \dots, N$,

$$\begin{aligned} b_E^{(N)}(Q^{z_0}_{[M/N]}, \gamma_j) =& b_E^{(N)}\left(\alpha_0 g_N^{z_0} + \sum_{j'=1}^N \alpha_{j'} \gamma_{j'}, \gamma_j \right) = \sum_{j'=1}^N \alpha_{j'} b_E^{(N)}\left(\gamma_{j'}, \gamma_j\right) \\ =& \sum_{j'=1}^N \alpha_{j'} \sum_{\lambda \in \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} \gamma_{j'}(\lambda)\overline{\gamma_j(\lambda)} \\ =& \alpha_j \frac{\|r_{\lambda_j}\|_{\mathcal{V}}^2}{|\lambda_j - z_0|^{2E+2}} \left|\gamma_j(\lambda_j)\right|^2 \end{aligned}$$

by discrete orthogonality of γ_j and $\gamma_{j'}$ over Λ_N. It follows that

$$|\alpha_j| = \frac{|\lambda_j - z_0|^{2E+2}}{\|r_{\lambda_j}\|_{\mathcal{V}}^2 |\gamma_j(\lambda_j)|^2} \left| b_E^{(N)}(Q^{z_0}_{[M/N]}, \gamma_j) \right| \leq C'_j |\lambda_j - z_0|^{2E} \left| b_E^{(N)}(Q^{z_0}_{[M/N]}, \gamma_j) \right|,$$

with C'_j independent of E.

We need to find a bound for $b_E^{(N)}(Q^{z_0}_{[M/N]}, \gamma_j)$: to this aim, we observe that

$$\left| b_E^{(N)}(Q^{z_0}_{[M/N]}, \gamma_j) \right| \leq \left| b_E(Q^{z_0}_{[M/N]}, \gamma_j) \right| + \sup_{\|Q\|_{z_0}=1} \left| b_E^{(N)}(Q, \gamma_j) - b_E(Q, \gamma_j) \right|.$$

In order to bound the first term, we apply (4.10), the Cauchy-Schwarz inequality, and (4.8):

$$\left| b_E(Q^{z_0}_{[M/N]}, \gamma_j) \right| \leq C^2 \left(\bar{R}_N\right)^{-2E} \|\gamma_j\|_{z_0} = C' \left(\bar{R}_N\right)^{-2E}.$$

Regarding the second term, by (4.2) and the triangular inequality, we have

$$\begin{aligned}
\left| b_E(Q, \gamma_j) - b_E^{(N)}(Q, \gamma_j) \right| &\leq \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} \left| Q(\lambda) \overline{\gamma_j(\lambda)} \right| \\
&\leq \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} \left(\sum_{n=0}^{N} |\lambda - z_0|^{2n} \right)^{1/2} \prod_{\substack{i=1 \\ i \neq j}}^{N} |\lambda - \lambda_i| \\
&\leq \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} \left(\sum_{n=0}^{N} |\lambda - z_0|^{2n} \right)^{1/2} \prod_{\substack{i=1 \\ i \neq j}}^{N} (|\lambda - z_0| + |\lambda_i - z_0|) \\
&\leq \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2E+2}} \left(\sum_{n=0}^{N} |\lambda - z_0|^{2n} \right)^{1/2} 2^{N-1} |\lambda - z_0|^{N-1}.
\end{aligned}$$

As in the proof of Lemma 4.3, we have

$$\left| b_E(Q, \gamma_j) - b_E^{(N)}(Q, \gamma_j) \right| \leq 2^{N-1} \sup_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\left(\sum_{n=0}^{N} |\lambda - z_0|^{2n} \right)^{1/2} |\lambda - z_0|^{N-1}}{|\lambda - z_0|^{2E}} \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^2}.$$

Since $E \geq N$, the term inside the supremum, i.e.,

$$\frac{1}{|\lambda - z_0|^{2E-N+1}} \left(\sum_{n=0}^{N} |\lambda - z_0|^{2n} \right)^{1/2} = \frac{1}{|\lambda - z_0|^{2E-2N+1}} \left(\sum_{n=-N}^{0} |\lambda - z_0|^{2n} \right)^{1/2},$$

is decreasing with respect to $|\lambda - z_0|$, so that

$$\begin{aligned}
\left| b_E(Q, \gamma_j) - b_E^{(N)}(Q, \gamma_j) \right| &\leq 2^{N-1} \frac{\left(\sum_{n=0}^{N} \left(\bar{R}_N\right)^{2n} \right)^{1/2} \left(\bar{R}_N\right)^{N-1}}{\left(\bar{R}_N\right)^{2E}} \sum_{\lambda \in \Lambda \setminus \Lambda_N} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^2} \\
&= C'' \left(\bar{R}_N\right)^{-2E},
\end{aligned}$$

and the claim follows. □

4.1.5 Proof of Theorem 3.1

For convenience, we report a copy of the statement in question, copied from Section 3.1.1.

Theorem 3.1 (Pole convergence [Bon+20a, Theorem 5.5])**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq N^{z_0}(R)$. Assume that $Q^{z_0}_{[M,N]}$ is computed using Definition 3.1, relying on E Taylor coefficients of v. Also, let $\bar{R}_N$ be as in Lemma 3.1. If $|\lambda_N - z_0| < \bar{R}_N$, then, for all $j = 1, \dots, N$ and E large enough,

$$\min_{\lambda' : Q^{z_0}_{[M/N]}(\lambda') = 0} |\lambda' - \lambda_j| \leq C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N} \right)^{2E}, \tag{3.10}$$

with C_j independent of M (in fact, the value of M is irrelevant here) and E, but dependent on N.

Proof. Fix $j \in \{1, \dots, N\}$, and let $\{\lambda'_j\}_{j=1}^{N'}$ be the roots of $Q^{z_0}_{[M/N]}$, where $N' \leq N$ is the exact degree of $Q^{z_0}_{[M/N]}$. By Lemmas 3.1 and 4.1 and the triangular inequality, we have the bound

$$\begin{aligned} C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N} \right)^{2E} &\geq \left| Q^{z_0}_{[M/N]}(\lambda_j) \right| \geq \prod_{j'=1}^{N'} \frac{\left|\lambda'_{j'} - \lambda_j\right|}{1 + \left|\lambda'_{j'} - z_0\right|} \\ &\geq \prod_{j'=1}^{N'} \phi_j\left(\left|\lambda'_{j'} - \lambda_j\right|\right) \geq \left(\min_{j'=1,\dots,N'} \phi_j\left(\left|\lambda'_{j'} - \lambda_j\right|\right) \right)^{N'}, \end{aligned} \tag{4.11}$$

with $\phi_j(x) = x/(1 + |\lambda_j - z_0| + x)$ a strictly increasing bounded continuous function over $x \geq 0$, such that $\phi_j(0) = 0$ and $\phi'(0) = 1/(1 + |\lambda_j - z_0|)$. In particular, by monotonicity,

$$\min_{j'=1,\dots,N'} \phi_j\left(\left|\lambda'_{j'} - \lambda_j\right|\right) = \phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right| \right).$$

Since $|\lambda_j - z_0| \leq |\lambda_N - z_0| < \bar{R}_N$, we conclude that

$$\lim_{E\to\infty} \left(\phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right| \right) \right)^{N'} \leq \lim_{E\to\infty} C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N} \right)^{2E} = 0.$$

In particular, it follows that $N' > 0$ for large E, and $\left|\lambda'_{j'} - \lambda_j\right|$ converges to 0 as $E \to \infty$ by continuity of ϕ_j. Moreover, since this holds for all j, we have $N' = N$ for large E.

Now it only remains to prove the convergence rate, which we do for the exact pole with index $j \in \{1, \dots, N\}$. This is trivial if $N = 1$ (it suffices to apply ϕ_j^{-1} to (4.11)), so we restrict our attention to $N > 1$. Let r be half the minimal distance between a couple of relevant poles, i.e.,

$$2r = \min_{1 \leq j < j' \leq N} |\lambda_j - \lambda_{j'}|.$$

Since each of the surrogate poles is converging to a *different* exact pole, for large E, the surrogate

poles can be partitioned as $\{\lambda'_{j'}\}_{j'=1}^{N} = \{\lambda^\star\} \sqcup \{\lambda^\star_{j'}\}_{j'=1}^{N-1}$, in such a way that

$$\min_{j'=1,\dots,N} \left|\lambda'_{j'} - \lambda_j\right| = |\lambda^\star - \lambda_j| < r < \left|\lambda^\star_{j'} - \lambda_j\right| \quad \forall j' = 1, \dots, N-1.$$

This allows to refine (4.11):

$$C_j \left(\frac{|\lambda_j - z_0|}{\bar{R}_N}\right)^{2E} \geq \phi_j\left(|\lambda^\star - \lambda_j|\right) \prod_{j'=1}^{N-1} \phi_j\left(\left|\lambda^\star_{j'} - \lambda_j\right|\right) \geq \phi_j\left(|\lambda^\star - \lambda_j|\right) \phi_j\left(r\right)^{N-1}.$$

Dividing by the E-independent term $\phi_j\left(r\right)^{N-1}$ and applying ϕ_j^{-1}, we obtain

$$|\lambda^\star - \lambda_j| \leq \phi_j^{-1}\left(\frac{C_j}{\phi_j\left(r\right)^{N-1}} \left(\frac{|\lambda_j - z_0|}{\bar{R}_N}\right)^{2E}\right).$$

Note that, when applying ϕ_j^{-1}, we are assuming that the argument above is smaller than 1, which is the case for large E.

To obtain the desired rate, it suffices to note that

$$\phi_j^{-1}(x) = (1 + |\lambda_j - z_0|)\frac{x}{1-x} \leq 2(1 + |\lambda_j - z_0|)x \quad \forall 0 \leq x \leq \frac{1}{2},$$

where, without loss of generality, we can assume that $x \leq \frac{1}{2}$ for E large enough, since our x converges to 0 as $E \to \infty$. □

4.1.6 Proof of Theorem 3.2

For convenience, we report a copy of the statement in question, copied from Section 3.1.1.

Theorem 3.2 (Global pole convergence [Bon+20a, Theorem 5.7])**.**
Let Assumption 3.3 be valid, and consider a sequence $\{(N_i, E_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i \leq E_i \leq E_{i+1}$ for all $i = 2, 3, \dots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $j = 1, 2, \dots$, we have

$$\lim_{i\to\infty} \min_{\lambda' : Q^{z_0}_{[M/N_i]}(\lambda') = 0} |\lambda' - \lambda_j| = 0, \tag{3.12}$$

where $Q^{z_0}_{[M/N_i]}$ denotes the fast LS Padé denominator of degree N_i, computed from E_i Taylor coefficients of v. (Note that the value of M is irrelevant.)

Proof. For simplicity, we only show the result for $E_i = N_i$, and we drop the index i, so that the claim is equivalent to convergence as $N \to \infty$. The general case is covered in full in [Bon+20a], and only requires some minor changes to the proof.

Let $j \in \{1, 2, \dots\}$ be fixed once and for all. Moreover, let $\{\lambda'_{j'}\}_{j'=1}^{N'}$ be the roots of $Q^{z_0}_{[M/N]}$, with $N' \leq N$. From Lemmas 4.2 and 4.3, we know that

$$\frac{\|r_{\lambda_j}\|_{\mathcal{V}}^2}{|\lambda_j - z_0|^{2N+2}} \left|Q^{z_0}_{[M/N]}(\lambda_j)\right|^2 \leq \sum_{\lambda\in\Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\lambda - z_0|^{2N+2}} \left|Q^{z_0}_{[M/N]}(\lambda)\right|^2$$

$$= \frac{1}{E!^2} J_E (Q^{z_0}_{[M/N]})^2 \leq C^2 \left(\bar{R}_N\right)^{-2N},$$

so that

$$\left|Q^{z_0}_{[M/N]}(\lambda_j)\right| \leq C \frac{|\lambda_j - z_0|^{N+1}}{\|r_{\lambda_j}\|_{\mathcal{V}} \left(\bar{R}_N\right)^N}.$$

By inspection of the proof of Lemma 4.3, we note that C can be chosen as

$$C = \frac{(2\bar{R}_N)^N}{\prod_{j'=1}^N |\lambda_{j'} - z_0|} \sum_{\lambda \in \bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|}, \tag{4.12}$$

with $\bar{R}_N = |\lambda_{N+1} - z_0|$ and $\bar{\Lambda} = \{\lambda_{j'}\}_{j'=N+1}^{\infty}$.

On the other hand, from the proof of Theorem 3.1, we have

$$\left|Q^{z_0}_{[M/N]}(\lambda_j)\right| \geq \prod_{j'=1}^{N'} \phi_j\left(\left|\lambda'_{j'} - \lambda_j\right|\right) \geq \phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right|\right)^{N'},$$

with $\phi_j(x) = x/(1 + |\lambda_j - z_0| + x)$. In particular, since $\phi_j(x) < 1$ for all $x > 0$, we have

$$\left|Q^{z_0}_{[M/N]}(\lambda_j)\right| \geq \phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right|\right)^{N}.$$

In summary, we see that

$$\begin{aligned}\phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right|\right) &\leq \left(\frac{(2\bar{R}_N)^N}{\prod_{j'=1}^N |\lambda_{j'} - z_0|} \sum_{\lambda \in \bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|} \frac{|\lambda_j - z_0|^{N+1}}{\|r_{\lambda_j}\|_{\mathcal{V}} \left(\bar{R}_N\right)^N}\right)^{1/N} \\ &= 2\left(\frac{|\lambda_j - z_0|}{\|r_{\lambda_j}\|_{\mathcal{V}}} \sum_{\lambda \in \bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|}\right)^{1/N} \prod_{j'=1}^{N} \left|\frac{\lambda_j - z_0}{\lambda_{j'} - z_0}\right|^{1/N}, \end{aligned} \tag{4.13}$$

and, by applying ϕ_j^{-1}, the claim follows if we can show that the right-hand-side above converges to 0 as $N \to \infty$.

If Λ is finite, i.e., $\lambda_m = \infty$ for m large enough, then the claim holds trivially, e.g., because $\sum_{\lambda \in \bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|} = 0$ for N large enough. Otherwise, it is easy to see that many of the terms above are bounded uniformly with respect to N, so that

$$\phi_j\left(\min_{j'=1,\dots,N'} \left|\lambda'_{j'} - \lambda_j\right|\right) \leq (2 + \varepsilon) \prod_{j'=1}^{N} \left|\frac{\lambda_j - z_0}{\lambda_{j'} - z_0}\right|^{1/N}$$

for an arbitrary $\varepsilon > 0$, for N large enough (depending on ε). But, by the Stolz-Cesàro theorem [ABC12],

$$\lim_{N \to \infty} \prod_{j'=1}^{N} \left|\frac{\lambda_j - z_0}{\lambda_{j'} - z_0}\right|^{1/N} = \lim_{N \to \infty} \left|\frac{\lambda_j - z_0}{\lambda_N - z_0}\right| = 0,$$

and the claim follows. □

4.1.7 Proof of Theorem 3.3

For convenience, we report a copy of the statement in question, copied from Section 3.1.2.

Theorem 3.3 (Error convergence [Bon+20a, Lemma 6.1])**.**
Let Assumption 3.2 be valid over $B^{z_0}(R)$, and take $N \leq \min\{N^{z_0}(R), M+1\}$. Also, let $\bar{R}_N$ be as in Lemma 3.1, and consider the punctured domain $B_N = B^{z_0}(\bar{R}_N) \setminus \{\lambda_j\}_{j=1}^N$. Then, if $v^{z_0}_{[M/N]}$ is the $[M/N]$ fast LS Padé approximant computed with $E = \max\{M, N\}$,

$$\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq \frac{C}{d(z) \left| Q^{z_0}_{[M/N]}(z) \right|} \left(\frac{|z - z_0|}{\bar{R}_N} \right)^E \qquad \forall z \in B_N, \tag{3.13}$$

with C independent of M, E and z, and

$$d(z) = \min_{z' \in \mathbb{C} \setminus B_N} |z - z'| = \min \left\{ \bar{R}_N - |z - z_0|, |z - \lambda_1|, \dots, \left| z - \lambda_{N^{z_0}(R)} \right| \right\}.$$

Additionally, let B' be an arbitrary compact subset of B_N. We have uniform exponential convergence over B' for fixed *N:*

$$\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq C_{B'} \left(\frac{\max_{z \in B'} |z - z_0|}{\bar{R}_N} \right)^E \qquad \forall z \in B', \text{ for large } E, \tag{3.14}$$

with $C_{B'}$ independent of E but, notably, dependent on N and B'.

Proof. We start from the case $N \leq M = E$. For convenience, we set

$$P^{z_0}_{[M/N]}(z) = \sum_{n=0}^{M} p_n (z - z_0)^n \quad \text{and} \quad Q^{z_0}_{[M/N]}(z) = \sum_{n=0}^{M} q_n (z - z_0)^n,$$

with $q_{N+1} = \ldots = q_M = 0$, i.e., we add $M - N$ zero terms to the expression of $Q^{z_0}_{[M/N]}$. First, we note that, by (3.4c) and (4.7) we have

$$p_n = \sum_{m=0}^{n} \frac{1}{(n-m)!} q_m \frac{\mathrm{d}^{n-m} v}{\mathrm{d} z^{n-m}}(z_0) = \sum_{\lambda \in \Lambda} \sum_{m=0}^{n} \frac{r_\lambda}{(\lambda - z_0)^{n+1}} q_m (\lambda - z_0)^m,$$

so that

$$\begin{aligned}
P^{z_0}_{[M/N]}(z) &= \sum_{\lambda \in \Lambda} \sum_{n=0}^{M} \sum_{m=0}^{n} \frac{r_\lambda}{(\lambda - z_0)^{n+1}} q_m (\lambda - z_0)^m (z - z_0)^n \\
&= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{\lambda - z_0} \sum_{m=0}^{M} q_m (\lambda - z_0)^m \sum_{n=m}^{M} \left(\frac{z - z_0}{\lambda - z_0} \right)^n \\
&= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{\lambda - z_0} \sum_{m=0}^{M} q_m (\lambda - z_0)^m \frac{\left(\frac{z - z_0}{\lambda - z_0} \right)^{M+1} - \left(\frac{z - z_0}{\lambda - z_0} \right)^m}{\frac{z - z_0}{\lambda - z_0} - 1} \\
&= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z - \lambda} \left(\left(\frac{z - z_0}{\lambda - z_0} \right)^{M+1} \sum_{m=0}^{M} q_m (\lambda - z_0)^m - \sum_{m=0}^{M} q_m (z - z_0)^m \right)
\end{aligned}$$

$$= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} Q^{z_0}_{[M/N]}(\lambda) + Q^{z_0}_{[M/N]}(z) v(z).$$

This allows us to obtain an alternative representation for the approximation error:

$$v^{z_0}_{[M/N]}(z) - v(z) = \frac{P^{z_0}_{[M/N]}(z) - Q^{z_0}_{[M/N]}(z) v(z)}{Q^{z_0}_{[M/N]}(z)} = \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} \frac{Q^{z_0}_{[M/N]}(\lambda)}{Q^{z_0}_{[M/N]}(z)}.$$

By taking the $\mathcal{V}$-norm and exploiting the Pythagorean theorem, we obtain

$$\begin{aligned}
\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} &= \left(\sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|^2_{\mathcal{V}}}{|z-\lambda|^2} \left| \frac{z-z_0}{\lambda - z_0} \right|^{2M+2} \left| \frac{Q^{z_0}_{[M/N]}(\lambda)}{Q^{z_0}_{[M/N]}(z)} \right|^2 \right)^{1/2} \\
&\leq \frac{|z-z_0|^{M+1}}{d(z) \left| Q^{z_0}_{[M/N]}(z) \right|} \left(\sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|^2_{\mathcal{V}}}{|\lambda - z_0|^{2M+2}} \left| Q^{z_0}_{[M/N]}(\lambda) \right|^2 \right)^{1/2} \\
&= \frac{|z-z_0|^{M+1}}{M!\, d(z) \left| Q^{z_0}_{[M/N]}(z) \right|} J_M(Q^{z_0}_{[M/N]})
\end{aligned} \tag{4.14}$$

by Lemma 4.2. Since $E = M$, the first claim follows by Lemma 4.3.

In the alternative case $N = M + 1 = E + 1$, only two minor tweaks to the proof are necessary: firstly, the expansion of $Q^{z_0}_{[M/N]}$ and $P^{z_0}_{[M/N]}$ become

$$Q^{z_0}_{[M/N]}(z) = \sum_{n=0}^{M} q_n (z-z_0)^n + q_{M+1}(z-z_0)^{M+1},$$

and, following the same steps as above,

$$\begin{aligned}
P^{z_0}_{[M/N]}(z) &= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} \sum_{m=0}^{M} q_m (\lambda - z_0)^m - \sum_{m=0}^{M} q_m (z-z_0)^m \right) \\
&= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} Q^{z_0}_{[M/N]}(\lambda) + v(z) Q^{z_0}_{[M/N]}(z) \\
&\quad - \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} q_{M+1}(\lambda - z_0)^{M+1} + q_{M+1}(z-z_0)^{M+1} \right) \\
&= \sum_{\lambda \in \Lambda} \frac{r_\lambda}{z-\lambda} \left(\frac{z-z_0}{\lambda - z_0} \right)^{M+1} Q^{z_0}_{[M/N]}(\lambda) + Q^{z_0}_{[M/N]}(z) v(z),
\end{aligned}$$

respectively. This is the same expansion as in the case $N \leq M$, so that (4.14) holds. However, since $M < E$, Lemma 4.3 cannot be applied. Instead, we need the alternative bound

$$\begin{aligned}
\left\| v^{z_0}_{[M/N]}(z) - v(z) \right\|_{\mathcal{V}} &= \left(\sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|^2_{\mathcal{V}}}{|\lambda - z_0|^{2M+4}} \frac{|z-z_0|^{2M+2} |\lambda - z_0|^2}{|z-\lambda|^2} \left| \frac{Q^{z_0}_{[M/N]}(\lambda)}{Q^{z_0}_{[M/N]}(z)} \right|^2 \right)^{1/2} \\
&\leq \frac{|z-z_0|^{M+1}}{(M+1)! \left| Q^{z_0}_{[M/N]}(z) \right|} \sup_{\lambda \in \Lambda} \left| \frac{\lambda - z_0}{z-\lambda} \right| J_{M+1}(Q^{z_0}_{[M/N]}).
\end{aligned}$$

The triangular inequality and the definition of $d(z)$ lead to

$$\left|\frac{\lambda - z_0}{z - \lambda}\right| \leq 1 + \left|\frac{z - z_0}{z - \lambda}\right| \leq 1 + \frac{\bar{R}_N}{d(z)} \leq 2\frac{\bar{R}_N}{d(z)},$$

which then yields the claim by Lemma 4.3.

Now, let N be fixed. To obtain the uniform convergence result, it suffices to bound from below d and $\left|Q^{z_0}_{[M/N]}\right|$, uniformly over B'. The former is trivially bounded away from 0 since B' is a *compact* subset of B_N. Concerning the latter, we know from Theorem 3.1 that the roots of $Q^{z_0}_{[M/N]}$ converge to exact poles of v, which lie outside of B'. Hence, for E large enough, all roots of $Q^{z_0}_{[M/N]}$ lie within $B^{z_0}(\bar{R}_N)$, at distance at least ε from B', for some small enough ε. Then, Lemma 4.1 gives the desired result: for all $z \in B'$,

$$\left|Q^{z_0}_{[M/N]}(z)\right| \geq \prod_{j=1}^{N'} \frac{\left|z - \lambda'_j\right|}{1 + \left|z_0 - \lambda'_j\right|} \geq \left(\frac{\varepsilon}{1 + \bar{R}_N}\right)^{N'} \geq \min\left\{1, \frac{\varepsilon}{1 + \bar{R}_N}\right\}^N > 0,$$

where $\{\lambda'_j\}_{j=1}^{N'}$, $N' \leq N$, are the roots of $Q^{z_0}_{[M/N]}$. □

4.1.8 Proof of Theorem 3.4

For convenience, we report a copy of the statement in question, copied from Section 3.1.2.

Theorem 3.4 (Global error convergence [Bon+20a, Theorem 6.3] (extended))**.**
Let Assumption 3.3 be valid, and consider a sequence $\{(M_i, N_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i \leq M_i + 1 \leq M_{i+1} + 1$ for all $i = 2, 3, \ldots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $R > 0$ and $\varepsilon > 0$, we have

$$\lim_{i\to\infty} \mathrm{Cap}\left(\left\{z \in B^{z_0}(R) : \left\|v^{z_0}_{[M_i/N_i]}(z) - v(z)\right\|_{\mathcal{V}} > \varepsilon^{N_i}\right\}\right) = 0, \tag{3.15}$$

where $v^{z_0}_{[M_i/N_i]}$ is the $[M_i/N_i]$ fast LS Padé approximant computed from $E_i = \max\{M_i, N_i\}$ Taylor coefficients of v and Cap *is the logarithmic capacity, see Definition 2.2.*

Proof. For simplicity, we assume that $N_i \leq M_i = E_i$ for all (large enough) i. The general case is analogous, and is covered in [Bon+20a]. The main idea of the proof is to define a sequence of sets $\{A_i\}_{i=1}^{\infty}$, with $A_i \subset B^{z_0}(R)$ for all i, such that

(a) $\left\{z \in B^{z_0}(R) : \left\|v^{z_0}_{[M_i/N_i]}(z) - v(z)\right\|_{\mathcal{V}} > \varepsilon^{N_i}\right\} \subset A_i$ for all i,

(b) $\lim_{i\to\infty} \mathrm{Cap}(A_i) = 0$.

Since $0 \leq \mathrm{Cap}(A) \leq \mathrm{Cap}(B)$ if $A \subset B$, the claim follows.

Define $\eta > 0$ small enough and $R' > 0$ such that $R + \eta \leq R'$. Moreover, let $\bar{R}_{N_i} = |\lambda_{N_i+1} - z_0|$, with $R' < \bar{R}_{N_i}$ for large enough i. Let $N^{z_0}(R')$ be as in Assumption 3.1, i.e., $N^{z_0}(R')$ is the number of elements of $\Lambda \cap B^{z_0}(R')$. Also, let $\{\lambda'_{j,i}\}_{j=1}^{N'_i}$ be the poles of $Q^{z_0}_{[M_i/N_i]}$, with $N'_i \leq N_i$, which we sort in such a way that

$$\left|\lambda'_{1,i} - z_0\right| \leq \cdots \leq \left|\lambda'_{N'_i,i} - z_0\right|.$$

Also, we set $N_i'' \in \{0, \dots, N_i'\}$ as the largest index j such that $\left|\lambda_{j,i}' - z_0\right| \leq 2R$. We define *ad hoc* A_i as

$$A_i = \left\{ z \in B^{z_0}(R) : \left(\prod_{j=1}^{N^{z_0}(R')} |z - \lambda_j| \right) \left(\prod_{j=1}^{N_i''} \left|z - \lambda_{j,i}'\right| \right) \leq \delta_i^{N^{z_0}(R') + N_i''} \right\}, \tag{4.15}$$

with

$$\delta_i = \left(\frac{R^{M_i+1}(1+2R)^{N_i}(2R')^{N^{z_0}(R')}}{\varepsilon^{N_i} \min\{1, R^{N_i}\} \left(\bar{R}_{N_i}\right)^{M_i - N_i} \min\{2R', \eta\}} \sum_{\lambda \in \bar{\Lambda}_i} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|} \prod_{j=1}^{N_i} \frac{2}{|\lambda_j - z_0|} \right)^{1/(N^{z_0}(R') + N_i'')}.$$

Now it just remains to show the two claims above.

(a) We assume from here onward that $z \in B^{z_0}(R)$. By Lemma 4.1, we have

$$\left|Q^{z_0}_{[M_i/N_i]}(z)\right| \geq \prod_{j=1}^{N_i'} \frac{\left|\lambda_{j,i}' - z\right|}{1 + \left|\lambda_{j,i}' - z_0\right|}.$$

For $j = 1, \dots, N_i''$, we have the bound

$$\frac{\left|\lambda_{j,i}' - z\right|}{1 + \left|\lambda_{j,i}' - z_0\right|} \geq \frac{\left|\lambda_{j,i}' - z\right|}{1 + 2R},$$

whereas, for $j = N_i'' + 1, \dots, N_i'$, we have, by the triangular inequality,

$$\frac{\left|\lambda_{j,i}' - z\right|}{1 + \left|\lambda_{j,i}' - z_0\right|} \geq \frac{\left|\lambda_{j,i}' - z_0\right|}{1 + \left|\lambda_{j,i}' - z_0\right|} - \frac{|z - z_0|}{1 + \left|\lambda_{j,i}' - z_0\right|} \geq \frac{2R}{1+2R} - \frac{R}{1+2R} = \frac{R}{1+2R},$$

since $x \mapsto x/(1+x)$ is an increasing function. In summary,

$$\left|Q^{z_0}_{[M_i/N_i]}(z)\right| \geq \frac{R^{N_i' - N_i''} \prod_{j=1}^{N_i''} \left|\lambda_{j,i}' - z\right|}{(1+2R)^{N_i'}} \geq \frac{\min\{1, R^{N_i}\}}{(1+2R)^{N_i}} \prod_{j=1}^{N_i''} \left|\lambda_{j,i}' - z\right|. \tag{4.16}$$

Moreover let $d(z)$ be as in Theorem 3.3, but with R' replacing R:

$$d(z) = \min\left\{R' - |z - z_0|, |z - \lambda_1|, \dots, \left|z - \lambda_{N^{z_0}(R')}\right|\right\}.$$

Also, define $j^\star(z) \in \{1, \dots, N^{z_0}(R')\}$ as the index of the closest pole to z, i.e., $\left|\lambda_{j^\star(z)} - z\right| \leq |\lambda_j - z|$ for $j = 1, \dots, N^{z_0}(R')$. Then

$$\frac{\prod_{j=1}^{N^{z_0}(R')} |z - \lambda_j|}{d(z)} = \begin{cases} \prod_{j=1, j \neq j^\star(z)}^{N^{z_0}(R')} |z - \lambda_j| & \text{if } \left|z - \lambda_{j^\star(z)}\right| \leq R' - |z - z_0|, \\ \frac{\prod_{j=1}^{N^{z_0}(R')} |z - \lambda_j|}{R' - |z - z_0|} & \text{if } \left|z - \lambda_{j^\star(z)}\right| > R' - |z - z_0|. \end{cases}$$

Both cases can be easily bounded from above by the triangular inequality:

$$\prod_{\substack{j=1\\ j\neq j^\star(z)}}^{N^{z_0}(R')} |z-\lambda_j| \leq \prod_{\substack{j=1\\ j\neq j^\star(z)}}^{N^{z_0}(R')} (|z-z_0|+|\lambda_j-z_0|) \leq (2R')^{N^{z_0}(R')-1}$$

and

$$\frac{\prod_{j=1}^{N^{z_0}(R')}|z-\lambda_j|}{R'-|z-z_0|} \leq \frac{\prod_{j=1}^{N^{z_0}(R')}(|z-z_0|+|\lambda_j-z_0|)}{R'-R} \leq \frac{(2R')^{N^{z_0}(R')}}{\eta},$$

so that

$$\frac{\prod_{j=1}^{N^{z_0}(R')}|z-\lambda_j|}{d(z)} \leq \frac{(2R')^{N^{z_0}(R')}}{\min\{2R',\eta\}}. \tag{4.17}$$

Now we take (4.14) and then use (4.12) as an upper bound for the constant in (4.8). In this fashion, we deduce that, for large i,

$$\begin{aligned}\left\|v^{z_0}_{[M_i/N_i]}(z)-v(z)\right\|_{\mathcal{V}} \leq & \frac{(2\bar{R}_{N_i})^{N_i}|z-z_0|^{M_i+1}}{d(z)\left|Q^{z_0}_{[M/N]}(z)\right|\left(\bar{R}_{N_i}\right)^{M_i}\prod_{j=1}^{N_i}|\lambda_j-z_0|}\sum_{\lambda\in\bar{\Lambda}_i}\frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}\\ \leq & \frac{R^{M_i+1}}{d(z)\left|Q^{z_0}_{[M/N]}(z)\right|\left(\bar{R}_{N_i}\right)^{M_i-N_i}}\sum_{\lambda\in\bar{\Lambda}_i}\frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}\prod_{j=1}^{N_i}\frac{2}{|\lambda_j-z_0|},\end{aligned} \tag{4.18}$$

with $\bar{\Lambda}_i=\{\lambda_j\}_{j=N_i+1}^{\infty}$. Putting (4.16) to (4.18) together, we obtain

$$\begin{aligned}&\left(\prod_{j=1}^{N^{z_0}(R')}|z-\lambda_j|\right)\left(\prod_{j=1}^{N_i''}|\lambda'_{j,i}-z|\right)\leq\\ &\leq\frac{R^{M_i+1}(1+2R)^{N_i}(2R')^{N^{z_0}(R')}}{\left\|v^{z_0}_{[M_i/N_i]}(z)-v(z)\right\|_{\mathcal{V}}\min\{1,R^{N_i}\}\left(\bar{R}_{N_i}\right)^{M_i-N_i}\min\{2R',\eta\}}\sum_{\lambda\in\bar{\Lambda}_i}\frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}\prod_{j=1}^{N_i}\frac{2}{|\lambda_j-z_0|}.\end{aligned}$$

Now, let $z\in B^{z_0}(R)$ be such that $\left\|v^{z_0}_{[M_i/N_i]}(z)-v(z)\right\|_{\mathcal{V}}>\varepsilon^{N_i}$. Then the right-hand-side above is $\leq\delta_i^{N^{z_0}(R')+N_i''}$, and (a) follows, since $z\in A_i$ by construction.

(b) A_i is a subset of the interior of a lemniscate:

$$A_i'=\left\{z\in\mathbb{C}:\left(\prod_{j=1}^{N^{z_0}(R')}|z-\lambda_j|\right)\left(\prod_{j=1}^{N_i''}|z-\lambda'_{j,i}|\right)\leq\delta_i^{N^{z_0}(R')+N_i''}\right\}.$$

The logarithmic capacity of A_i' equals δ_i by [BGM96, Theorem 6.6.3], so that we just have to show that $\lim_{i\to\infty}\delta_i=0$. To this aim, we first consider

$$\Delta_i=\delta_i^{N^{z_0}(R')+N_i''}=\frac{R(2R')^{N^{z_0}(R')}}{\min\{2R',\eta\}}\sum_{\lambda\in\bar{\Lambda}_i}\frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}\left(\frac{R}{\bar{R}_{N_i}}\right)^{M_i-N_i}\prod_{j=1}^{N_i}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}|\lambda_j-z_0|}.$$

We see that $\left(R/\bar{R}_{N_i}\right)^{M_i-N_i}$ converges to 0 because its base does, while

$$\lim_{i\to\infty}\prod_{j=1}^{N_i}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}\,|\lambda_j-z_0|}=0$$

since $\lim_{j\to\infty}|\lambda_j-z_0|=\infty$, so that

$$\begin{aligned}
0&\le\lim_{i\to\infty}\left(\prod_{j=1}^{N_i}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}\,|\lambda_j-z_0|}\right)^{1/(N^{z_0}(R')+N_i'')}\\
&\le\lim_{i\to\infty}\left(\prod_{j=1}^{N_i}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}\,|\lambda_j-z_0|}\right)^{1/(N^{z_0}(R')+N_i)}\\
&=\lim_{i\to\infty}\left(\left(\prod_{j=1}^{N_i}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}\,|\lambda_j-z_0|}\right)^{1/N_i}\right)^{N_i/(N^{z_0}(R')+N_i)}\\
&=\lim_{j\to\infty}\frac{2(1+2R)R}{\varepsilon\min\{1,R\}\,|\lambda_j-z_0|}=0
\end{aligned}$$

by the Stolz-Cesàro theorem [ABC12].

The claim follows. □

4.2 Proofs of results for the general case

Now we move to the results pertaining to MRI. The main reference throughout this chapter is [Pra20]. For some of the latter results, we only report a sketch of the proof, due to the similarities with the proofs of the equivalent results for fast LS Padé approximation.

4.2.1 Auxiliary result: bounds for normalized polynomials

In order to better understand how the Ψ_N-based normalization (3.17a) works, we generalize Lemma 4.1.

Lemma 4.4 (Normalization of nodal polynomials [Pra20, Lemma 3.4])**.**
Let $Q\in\mathbb{P}_N^{\Psi_N}(\mathbb{C};\mathbb{C})$ and $z_0\in\mathbb{C}$. Then, there exists C^{z_0,Ψ_N} (independent of z) such that

$$|Q(z)|\le C^{z_0,\Psi_N}\left(\sum_{n=0}^{N}|z-z_0|^{2n}\right)^{1/2}=\begin{cases}C^{z_0,\Psi_N}\sqrt{N+1} & \text{if } |z-z_0|=1,\\ C^{z_0,\Psi_N}\left(\frac{|z-z_0|^{2N+2}-1}{|z-z_0|^2-1}\right)^{1/2} & \text{if } |z-z_0|\neq 1.\end{cases}\tag{4.19}$$

Moreover, assume that Q has exact degree $N'\le N$, with roots $\{z_j\}_{j=1}^{N'}$ (repeated according to multiplicity), all different from z_0. Then, there exists c^{z_0,Ψ_N} (independent of z) such that

$$c^{z_0,\Psi_N}\prod_{j=1}^{N'}\frac{|z-z_j|}{1+|z_0-z_j|}\le|Q(z)|\le C^{z_0,\Psi_N}\prod_{j=1}^{N'}\left|\frac{z-z_j}{z_0-z_j}\right|\le C^{z_0,\Psi_N}\prod_{j=1}^{N'}\left(1+\left|\frac{z-z_0}{z_0-z_j}\right|\right)\ \forall z\in\mathbb{C}.\tag{4.20}$$

Note that c^{z_0,Ψ_N} and C^{z_0,Ψ_N} depend on N.

Proof. Lemma 4.1 shows that the claim holds true for the shifted monomial basis $\{(\cdot - z_0)^i\}_{i=0}^N$. The result for general bases follows, since, for a given N, all norms over the finite-dimensional space $\mathbb{P}_N(\mathbb{C};\mathbb{C})$ are equivalent. □

4.2.2 Auxiliary result: alternative expressions of target functional

We can obtain alternative expressions for the quantity J^Z in (3.17b).

Lemma 4.5 (Alternative expressions of J^Z [Pra20, Lemma 3.5 (first claim)])**.**
Let Assumption 3.5 be valid over A_R for small enough R, and take an arbitrary $Q \in \mathbb{P}_N(\mathbb{C};\mathbb{C})$. Also, define the nodal polynomial $\omega^Z(z) = \prod_{z' \in Z}(z - z')$. Then,

$$\frac{1}{(S-1)!} J^Z(Q) = \left(\sum_{\lambda,\lambda' \in \Lambda} \frac{\langle r_\lambda, r_\lambda \rangle_{\mathcal{V}}}{\omega^Z(\lambda)\overline{\omega^Z(\lambda')}} Q(\lambda)\overline{Q(\lambda')} \right)^{1/2}, \tag{4.21}$$

and

$$\frac{1}{(S-1)!} J^Z(Q) \leq \sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\omega^Z(\lambda)|} |Q(\lambda)| . \tag{4.22}$$

If, in addition, Assumption 3.6 holds, then

$$\frac{1}{(S-1)!} J^Z(Q) = \left(\sum_{\lambda \in \Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}^2}{|\omega^Z(\lambda)|^2} |Q(\lambda)|^2 \right)^{1/2}. \tag{4.23}$$

Proof. The proof is similar to that of Lemma 4.2, which can be found in Section 4.1.2.

Assume that the sample points are distinct. Then, by the barycentric formula (2.2), we have

$$\begin{aligned} J^Z(Q) =& \left\| \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv) \right\|_{\mathcal{V}} = \left\| \sum_{\lambda \in \Lambda} \sum_{j=1}^{S} \frac{Q(z_j) r_\lambda}{(\lambda - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} \left(\frac{\omega^Z}{\cdot - z_j} \right) \right\|_{\mathcal{V}} \\ =& (S-1)! \left\| \sum_{\lambda \in \Lambda} \frac{r_\lambda}{\omega^Z(\lambda)} \sum_{j=1}^{S} \frac{\omega^Z(\lambda) Q(z_j)}{(\lambda - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \right\|_{\mathcal{V}} = (S-1)! \left\| \sum_{\lambda \in \Lambda} \frac{r_\lambda}{\omega^Z(\lambda)} Q(\lambda) \right\|_{\mathcal{V}} . \end{aligned}$$

Note that the last step follows by the exactness of the interpolation operator on degree-$(S-1)$ polynomials. Also, we can divide by $\omega^Z(\lambda)$ since, by assumption, no pole of v is on ∂A. The first claim follows. The two bounds can be obtained with the same steps as in Section 4.1.2.

If the sample points are not distinct, it becomes necessary to use the generalized barycentric form (5.4) instead. The rest of the proof is similar, albeit much more notation-heavy. Alternatively, the proof of the general case can be carried out by exploiting the exactness of the interpolation operator I^Z as in the proof of Lemma 5.1. We skip the details here. □

4.2.3 Auxiliary result: optimal value of target functional

We can bound the optimal value of J^Z.

Lemma 4.6 (Minimal value of J^Z [Pra20, Lemma 3.5 (last claim)])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S (we remind the reader that Z_S has S elements, cf. Theorem 2.2). Let Assumption 3.5 be valid over A_R for some $R > \mathrm{Cap}(A)$. Fix $N \geq N^A(\mathrm{Cap}(A))$ (with $N^A(\mathrm{Cap}(A))$ the number of poles of v in A). Also, let $\bar{R}_N = R$ if $N \geq N^A(R)$ and $\bar{R}_N = \Phi_A(\lambda_{N+1})$ otherwise. Then, for all $\rho < \bar{R}_N$ and S large enough (depending on ρ and N),

$$\frac{1}{(S-1)!} J^{Z_S}(Q^{Z_S}_{[S-1/N]}) \leq C\rho^{-S}, \tag{4.24}$$

The constant C is independent of M and S.

Proof. The proof is fairly similar to that of Lemma 4.3, which can be found in Section 4.1.3.

Let $z_0 \in A \setminus \Lambda$, $N_{\mathrm{eff}} = \min\{N, N^A(R)\}$, and set $\Lambda_{N_{\mathrm{eff}}} = \{\lambda_j\}_{j=1}^{N_{\mathrm{eff}}}$ and $\bar{\Lambda} = \Lambda \setminus \Lambda_{N_{\mathrm{eff}}}$. In particular, note that $\min_{\lambda\in\bar{\Lambda}} \Phi_A(\lambda) \geq \bar{R}_N$. Also, we define the (effective) target polynomial $g^A_{N_{\mathrm{eff}}}$ as an element of $\mathbb{P}^{\Psi_N}_N(\mathbb{C};\mathbb{C})$ with exact degree N_{eff} and roots $\Lambda_{N_{\mathrm{eff}}}$. First, we observe that we can apply (4.20) to $g^A_{N_{\mathrm{eff}}}$ to obtain

$$\left|g^A_{N_{\mathrm{eff}}}(z)\right| \leq C^{z_0,\Psi_N} \prod_{j'=1}^{N_{\mathrm{eff}}} \left|\frac{z-\lambda_{j'}}{\lambda_{j'}-z_0}\right|.$$

By (3.17b), $Q^{Z_S}_{[S-1/N]}$ yields a lower value of J^{Z_S} than $g^A_{N_{\mathrm{eff}}}$. This, together with (4.22), gives

$$\begin{aligned}
\frac{1}{(S-1)!} J^Z(Q^{Z_S}_{[S-1/N]}) \leq& \sum_{\lambda\in\Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\omega^{Z_S}(\lambda)|} \left|g_{N_{\mathrm{eff}}} A(\lambda)\right| = \sum_{\lambda\in\bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\omega^{Z_S}(\lambda)|} \left|g_{N_{\mathrm{eff}}} A(\lambda)\right| \\
\leq& \frac{C^{z_0,\Psi_N}}{\prod_{j'=1}^{N_{\mathrm{eff}}} |\lambda_{j'}-z_0|} \sum_{\lambda\in\bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|} \frac{|\lambda-z_0| \prod_{j'=1}^{N_{\mathrm{eff}}} |\lambda-\lambda_{j'}|}{|\omega^{Z_S}(\lambda)|} \\
\leq& \frac{C^{z_0,\Psi_N}}{\prod_{j'=1}^{N_{\mathrm{eff}}} |\lambda_{j'}-z_0|} \sup_{\lambda\in\bar{\Lambda}} \Bigg(\underbrace{\frac{|\lambda-z_0| \prod_{j'=1}^{N_{\mathrm{eff}}} |\lambda-\lambda_{j'}|}{|\omega^{Z_S}(\lambda)|}}_{r^S(\lambda)} \Bigg) \sum_{\lambda\in\Lambda} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}.
\end{aligned} \tag{4.25}$$

It remains to capture the behavior of the supremum. For S large enough, $S > N_{\mathrm{eff}} + 1$ and, due to the degree difference between numerator and denominator in r^S, $\lim_{|\lambda|\to\infty} r^S(\lambda) = 0$. Thus, for large enough S, we have two cases:

- the supremum is attained at some $\lambda \in \bar{\Lambda}$, which is bounded uniformly in S;
- the supremum is the limit of a sequence $\{r^S(\bar{\lambda}_i)\}_{i=1}^{\infty}$, with $\bar{\Lambda} \supset \{\bar{\lambda}_i\}_{i=1}^{\infty} \to \bar{\lambda}$, where $\bar{\lambda}$ is bounded uniformly in S.

In both cases, we conclude that there exists a finite $\delta > 0$ such that $B = (A_{\bar{R}_N+\delta} \setminus A_{\bar{R}_N}) \cup \partial A_{\bar{R}_N}$ contains all supremizers (or sequences of supremizer convergents) for large S. Thus, we can apply (2.10) over B: given $0 < \varepsilon = 1 - \rho/\bar{R}_N$, for S large enough, we have

$$(1-\varepsilon)^S \Phi_A(z)^S \leq \left|\omega^{Z_S}(z)\right| \leq (1+\varepsilon)^S \Phi_A(z)^S \qquad \forall z \in B,$$

so that

$$\sup_{\lambda\in\bar{\Lambda}} r^S(\lambda) \leq \sup_{z\in B}\left(\frac{|z-z_0|\prod_{j'=1}^{N_{\text{eff}}}|z-\lambda_{j'}|}{|\omega^{Z_S}(z)|}\right) \leq \frac{\sup_{z\in B}\left(|z-z_0|\prod_{j'=1}^{N_{\text{eff}}}|z-\lambda_{j'}|\right)}{(1-\varepsilon)^S\left(\inf_{z\in B}\Phi_A(z)\right)^S}$$
$$= \frac{\sup_{z\in B}\left(|z-z_0|\prod_{j'=1}^{N_{\text{eff}}}|z-\lambda_{j'}|\right)}{(1-\varepsilon)^S\left(\bar{R}_N\right)^S} = \frac{\sup_{z\in B}\left(|z-z_0|\prod_{j'=1}^{N_{\text{eff}}}|z-\lambda_{j'}|\right)}{\rho^S}.$$

The claim follows. □

By inspection of the proof, we can see that, in general, (4.24) might not hold if ρ is replaced by its upper bound $\bar{R}_N$. This is due to the necessity to replace the nodal polynomial ω^{Z_S} in bound (4.22) with its limit $\Phi_A(\cdot)^S$, see Theorem 2.2. Note that any extension of Lemma 4.3 must be weakened in this way, due to the added complexity of "distributed" sampling schemes.

4.2.4 Proof of Lemma 3.2

For convenience, we report a copy of the statement in question, copied from Section 3.2.1.

Lemma 3.2 (Denominator value at poles [Pra20, Lemma 3.6])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A, and let Assumption 3.6 be valid over A_R. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S (we remind the reader that Z_S has S elements, cf. Theorem 2.2). Take a fixed $N^A(\mathrm{Cap}(A)) \leq N \leq N^A(R)$. Also, let $\bar{R}_N = R$ if $N \geq N^A(R)$ and $\bar{R}_N = \Phi_A(\lambda_{N+1})$ otherwise. For all $j = 1,\dots,N$, $\rho < \bar{R}_N$, and S large enough (depending on ρ and N), we have the bound

$$\left|Q^{Z_S}_{[S-1/N]}(\lambda_j)\right| \leq C_j\left(\frac{\Phi_A(\lambda_j)}{\rho}\right)^{2S}, \tag{3.20}$$

with C_j independent of M and S.

Sketch of proof. The proof is fairly similar to that of Lemma 3.1, which can be found in Section 4.1.4. As such, here we only summarize the main differences, referring to [Pra20, Theorem 3.6] for the full proof.

Let j and ρ be fixed once and for all. The quadratic forms are slightly different, since (4.21) must be used instead of (4.4). Moreover, Lemma 4.6 and (4.20) must replace Lemma 4.3 and (4.2), respectively. That being said, we obtain a bound of the form: for all $\bar{\rho} < \bar{R}_N$ and S large enough (depending on $\bar{\rho}$ and N)

$$\left|Q^{Z_S}_{[S-1/N]}(\lambda_j)\right| \leq C_j\frac{\left|\omega^{Z_S}(\lambda_j)\right|^2}{\bar{\rho}^{2S}}.$$

In particular, we choose $\bar{\rho} = \rho(1+\varepsilon) > \rho$, with $\varepsilon > 0$ small enough (e.g., $\varepsilon = \frac{\bar{R}_N-\rho}{2\rho}$). In order to obtain the claim, we must use (2.10) on the numerator: for the chosen ε, we have

$$\frac{\left|\omega^{Z_S}(\lambda_j)\right|^2}{\bar{\rho}^{2S}} \leq (1+\varepsilon)^{2S}\frac{\Phi_A(\lambda_j)^{2S}}{\bar{\rho}^{2S}} = \frac{\Phi_A(\lambda_j)^{2S}}{\rho^{2S}},$$

for S large enough (depending on ε, i.e., on N and ρ). □

4.2.5 Proof of Theorem 3.5

For convenience, we report a copy of the statement in question, copied from Section 3.2.1.

Theorem 3.5 (Pole convergence [Pra20, Theorem 3.7])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A, and let Assumption 3.6 be valid over A_R. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S. Take a fixed *$N \leq N^A(R)$. Also, set $\bar{R}_N = R$ if $N = N^A(R)$ and $\bar{R}_N = \Phi_A(\lambda_{N+1})$ otherwise. Assume that $\Phi_A(\lambda_N) < \rho < \bar{R}_N$, with ρ arbitrary. Then, for all $j = 1, \dots, N$ and S large enough (depending on ρ and N),*

$$\min_{\lambda' : Q^{Z_S}_{[S-1/N]}(\lambda') = 0} |\lambda' - \lambda_j| \leq C_j \left(\frac{\Phi_A(\lambda_j)}{\rho} \right)^{2S}, \tag{3.21}$$

with C_j independent of S.

Sketch of proof. The proof is very similar to that of Theorem 3.1, which can be found in Section 4.1.5. As such, we only summarize it here, referring to [Pra20, Theorem 3.7] for all the details.

First, we show that the N poles of the MRI surrogate converge to the exact ones, and then we obtain the convergence rate by exploiting the fact that the poles are simple, so that they are separated by a (uniformly) lower-bounded distance. Note that Lemmas 3.1 and 4.1 need to be replaced by Lemmas 3.2 and 4.4, respectively. □

4.2.6 Proof of Theorem 3.6

For convenience, we report a copy of the statement in question, copied from Section 3.2.1.

Theorem 3.6 (Global pole convergence [Pra20, Theorem 3.8])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S. Let Assumptions 3.3 and 3.4 be valid for some $z_0 \in \mathbb{C}$, and consider a sequence $\{(N_i, S_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i < S_i \leq S_{i+1}$ for all $i = 2, 3, \dots$. Further, assume that $\lim_{i\to\infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $j = 1, 2 \dots$, we have

$$\lim_{i\to\infty} \min_{\lambda' : Q^{Z_{S_i}}_{[S_i - 1/N_i]}(\lambda') = 0} |\lambda' - \lambda_j| = 0. \tag{3.22}$$

Sketch of proof. The proof is fairly similar to that of Theorem 3.2, see Section 4.1.6. As such, we only summarize the main steps here, referring to [Pra20, Theorem 3.8] for the full proof. Note, however, that [Pra20, Theorem 3.8] uses a stronger version of Assumption 3.4 than ours (with c^{z_0} and C^{z_0} appearing without exponents in (3.18)). For this reason, we will briefly outline the main differences between the two cases at the end of the proof.

Let $j \in \{1, 2, \dots\}$ be fixed, and assume that the poles $\{\lambda_{j'}\}_{j'=1}^{\infty}$ are sorted according to their distance from z_0. First, using the upper bound in (3.18) and Lemmas 3.2, 4.5, and 4.6, we obtain a bound on $Q^{Z_{S_i}}_{[S_i - 1/N_i]}(\lambda_j)$:

$$\left| Q^{Z_{S_i}}_{[S_i - 1/N_i]}(\lambda_j) \right| \leq \frac{(C^{z_0})^{N_i}}{\|r_{\lambda_j}\|_{\mathcal{V}}} \sum_{\lambda \in \bar{\Lambda}} \frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda - z_0|} \frac{\left|\omega^{Z_S}(\lambda_j)\right|}{\prod_{j'=1}^{N_i} |\lambda_{j'} - z_0|} \sup_{\lambda \in \bar{\Lambda}} \frac{|\lambda - z_0| \prod_{j'=1}^{N_i} |\lambda - \lambda_{j'}|}{|\omega^{Z_S}(\lambda)|},$$

with $\bar{\Lambda} = \{\lambda_{j'}\}_{j'=N+1}^{\infty}$. Then, we use the lower bound in (3.18) to show that

$$\left|Q^{Z_{S_i}}_{[S_i-1/N_i]}(\lambda_j)\right| \geq (c^{z_0})^{N_i}\phi_j\left(\min_{j'=1,\dots,N'}\left|\lambda'_{j'}-\lambda_j\right|\right)^{N_i},$$

with $\{\lambda'_{j'}\}_{j'=1}^{N'_i}$ the roots of $Q^{Z_{S_i}}_{[S_i-1/N_i]}$ and $\phi_j(x) = x/(\rho^{z_0}+|\lambda_j - z_0|+x)$ an increasing function taking values in $[0,1[$.

Due to the properties of ϕ_j, the claim follows if we can show that

$$\lim_{i\to\infty}\frac{C^{z_0}}{c^{z_0}}\left(\frac{1}{\left\|r_{\lambda_j}\right\|_{\mathcal{V}}}\sum_{\lambda\in\bar{\Lambda}}\frac{\|r_\lambda\|_{\mathcal{V}}}{|\lambda-z_0|}\frac{\left|\omega^{Z_S}(\lambda_j)\right|}{\prod_{j'=1}^{N_i}|\lambda_{j'}-z_0|}\sup_{\lambda\in\bar{\Lambda}}\frac{|\lambda-z_0|\prod_{j'=1}^{N_i}|\lambda-\lambda_{j'}|}{|\omega^{Z_S}(\lambda)|}\right)^{1/N_i} = 0.$$

To this aim, we first bound the argument of the supremum via the triangular inequality, obtaining a simpler expression, whose supremizer is λ_{N_i+1}. Then, we note that most of the terms appearing in the resulting bound are bounded uniformly in N, so that we may write

$$\phi_j\left(\min_{j'=1,\dots,N'_i}\left|\lambda'_{j'}-\lambda_j\right|\right) \leq C_j\frac{C^{z_0}}{c^{z_0}}\prod_{j'=1}^{N_i}\left(\frac{|\lambda_j-z_0|+\max_{z\in A}|z-z_0|}{|\lambda_{j'}-z_0|}\right)^{1/N_i},$$

with C_j bounded and independent from i. The claim then follows, since, by the Stolz-Cesàro theorem [ABC12], the product above converges to 0 as $i\to\infty$.

These same steps are followed in the proof of [Pra20, Theorem 3.8]. There, however, a condition stronger than Assumption 3.4 was used, so that the final bound is of the form

$$\phi_j\left(\min_{j'=1,\dots,N'_i}\left|\lambda'_{j'}-\lambda_j\right|\right) \leq C\left(\frac{C^{z_0}}{c^{z_0}}\right)^{1/N_i}\prod_{j'=1}^{N_i}\left(\frac{|\lambda_j-z_0|+\max_{z\in A}|z-z_0|}{|\lambda_{j'}-z_0|}\right)^{1/N_i}.$$

Since C^{z_0}/c^{z_0} is bounded (by the weaker version of Assumption 3.4), the absence of the exponent $1/N_i$ does not prevent us from showing the desired claim. □

4.2.7 Proof of Theorem 3.7

For convenience, we report a copy of the statement in question, copied from Section 3.2.2.

Theorem 3.7 (Error convergence [Pra20, Theorem 3.9])**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Green's potential Φ_A, and let Assumption 3.6 be valid over A_R. Take a sequence of Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S (we remind the reader that Z_S has S elements, cf. Theorem 2.2). Take a fixed *$N \leq N^A(R)$. Also, let $\bar{R}_N$ be as in Lemma 3.2, and consider an arbitrary $\mathrm{Cap}(A) < \rho < \bar{R}_N$. Define the punctured domain B_N as the interior of $A_\rho \setminus \{\lambda_j\}_{j=1}^{N}$. Then, for all $\varepsilon > 0$ and S large enough (depending on N and ε),*

$$\left\|v^{Z_S}_{[S-1/N]}(z) - v(z)\right\|_{\mathcal{V}} \leq \frac{C}{d(z)\left|Q^{Z_S}_{[S-1/N]}(z)\right|}\left(\frac{\Phi_A(z)}{\rho}\right)^S \qquad \forall z\in B_N, \tag{3.23}$$

with C independent of S, and $d(z) = \min_{z'\in\mathbb{C}\setminus B_N}|z-z'|$.

Additionally, let B' be an arbitrary compact subset of B_N. We have uniform exponential conver-

gence over B' for fixed N*:*

$$\left\| v^{Z_S}_{[S-1/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq C_{B'} \left(\frac{\max_{z \in B'} \Phi_A(z)}{\rho} \right)^S \qquad \forall z \in B', \tag{3.24}$$

with $C_{B'}$ independent of S but, notably, dependent on B'.

Sketch of proof. The proof is fairly similar to that of Theorem 3.3, see Section 4.1.7. As such, we only summarize it here, referring to [Pra20, Theorem 3.9] for all the details.

First, we rewrite the error norm by employing the barycentric formula (2.2), making the quantity $J^{Z_S}(Q^{Z_S}_{[S-1/N]})$ appear: more specifically, the bound (4.14) obtains a generalization in

$$\left\| v^{Z_S}_{[S-1/N]}(z) - v(z) \right\|_{\mathcal{V}} \leq \frac{\left| \omega^{Z_S}(z) \right|}{d(z) \left| Q^{Z_S}_{[S-1/N]}(z) \right|} J^{Z_S}(Q^{Z_S}_{[S-1/N]}). \tag{4.26}$$

Then, it suffices to bound this quantity by employing Lemma 4.6 (whereas, in the equivalent theorem for fast LS Padé approximation, Lemma 4.3 was used). □

4.2.8 Proof of Theorem 3.8

For convenience, we report a copy of the statement in question, copied from Section 3.2.2.

Theorem 3.8 (Global error convergence [Pra20, Theorem 3.10] (extended))**.**
Let $A \Subset \mathbb{C}$ be the sampling set, with Fekete points $\{Z_S\}_{S=1}^{\infty}$, with $Z_S \subset \partial A$ for all S. Let Assumptions 3.3 and 3.4 be valid for some $z_0 \in \mathbb{C}$, and consider a sequence $\{(N_i, S_i)\}_{i=1}^{\infty} \subset \mathbb{N}^2$, such that $N_{i-1} \leq N_i < S_i \leq S_{i+1}$ for all $i = 2, 3, \ldots$. Further, assume that $\lim_{i \to \infty} N_i = \infty$, i.e., both the number of snapshots and the denominator degree diverge. For all $R > \mathrm{Cap}(A)$ and $\varepsilon > 0$, we have

$$\lim_{i \to \infty} \mathrm{Cap}\left(\left\{ z \in A_R : \left\| v^{Z_{S_i}}_{[S_i - 1/N_i]}(z) - v(z) \right\|_{\mathcal{V}} > \varepsilon^{N_i} \right\} \right) = 0, \tag{3.25}$$

where Cap *is the logarithmic capacity, see Definition 2.2.*

Sketch of proof. The proof is fairly similar to that of Theorem 3.4, see Section 4.1.8. As such, we only summarize it here, referring to [Pra20, Theorem 3.10] for all the details.

We define a sequence of sets $\{A_i\}_{i=1}^{\infty}$ where the approximation error is large (more properly, we should say that we define them as the complements of sets where the error is small) and then we prove that their capacities converge to 0. In doing this, we rely on Assumption 3.4, (4.26), and Lemma 3.2.

To conclude this (sketch of) proof, we note that, as in the proof of Theorem 3.6, in [Pra20], we use a stronger version of Assumption 3.4 than ours (with c^{z_0} and C^{z_0} appearing without exponents in (3.18)) to prove an equivalent convergence result. In practical terms, our "weakened" version of Assumption 3.4 makes the sets A_i larger than the corresponding ones in [Pra20, Theorem 3.10]. Other than this, the proof remains unchanged. □

5 Additional aspects of MRI

In Chapter 3, we have defined and analyzed the MRI approach for rational function approximation. In the upcoming sections, we discuss some more practical aspects, useful when applying MRI in a MOR framework. An implementation of MRI is publicly available as part of the open-source Python package `RROMPy`, developed by the thesis author. The source code can be found at c4science.ch/source/RROMPy.

Note that, since fast LS Padé approximation is a special case of MRI, we do not lose any generality by considering only MRI in our discussion.

5.1 Implementation

The first issue that we face is: given Definition 3.2, how can an MRI be built in a practical application on a computer? Let us assume that the sample points $Z = \{z_j\}_{j=1}^S$ are fixed in advance (see Section 5.3 for more details on sample point selection), and that an external (PDE/linear system) solver has computed the corresponding snapshots $\{v_j\}_{j=1}^S$, with $v_j = v(z_j) \in \mathcal{V}$. Note that, by setting $v_j = v(z_j)$ for all j, we are implicitly assuming the sample points to be distinct. The general case is discussed in Section 5.1.1.

Now that the snapshots are available, we can set up the minimization problem that will allow us to find the MRI denominator $Q^Z_{[S-1/N]}$. For simplicity, we assume that the degree $N < S$ has been fixed in advance, and we refer to Section 5.2.1 for a discussion on on-the-fly selection of N. To construct the target functional J^Z, see (3.17b), we first look at the quantity $\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv)$, with Q a generic polynomial of degree $\leq N$. Thanks to the barycentric expansion (2.2), we have

$$\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv) = \sum_{j=1}^{S} \frac{Q(z_j)v(z_j)}{\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} \left(\frac{\omega^Z}{(\cdot - z_j)} \right) = \sum_{j=1}^{S} c_j Q(z_j) v(z_j), \tag{5.1}$$

with $\omega^Z(z) = \prod_{j=1}^S (z - z_j)$ the nodal polynomial and $c_j = (S-1)!/\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)$.

By expanding Q as $Q = \sum_{i=0}^N q_i \psi_i$, we can see that $J^Z(\cdot)^2$ is a quadratic form with respect to

the expansion coefficients $\{q_i\}_{i=0}^N$:

$$J^Z(Q)^2 = \left\| \sum_{j=1}^{S} c_j Q(z_j) v(z_j) \right\|_{\mathcal{V}}^2 = \left\| \sum_{i=0}^{N} \sum_{j=1}^{S} c_j \psi_i(z_j) v(z_j) q_i \right\|_{\mathcal{V}}^2$$
$$= \sum_{i,i'=0}^{N} \sum_{j,j'=1}^{S} c_j \overline{c_{j'}} \psi_i(z_j) \overline{\psi_{i'}(z_{j'})} \langle v(z_j), v(z_{j'}) \rangle_{\mathcal{V}} q_i \overline{q_{i'}}.$$

Accordingly, the representative matrix of $J^Z(\cdot)^2$ can be expressed as $\Upsilon_N^H C^H G C \Upsilon_N$, with

$$\Upsilon_N \in \mathbb{C}^{S \times (N+1)}, \qquad (\Upsilon_N)_{ji} = \psi_i(z_j), \qquad j = 1, \dots, S,\ i = 0, \dots, N, \tag{5.2}$$

the (generalized) Vandermonde matrix associated to Ψ_N at Z,

$$C \in \mathbb{C}^{S \times S}, \qquad (C)_{jj'} = c_j \delta_{jj'}, \qquad j, j' = 1, \dots, S, \tag{5.3}$$

a scaling diagonal matrix, and G the snapshot Gramian (2.44).

This expansion was judiciously chosen so that the normalization of Q, see (3.17a), turns (3.17b) into a minimal eigenvalue problem for $\Upsilon_N^H C^H G C \Upsilon_N$, with the expansion coefficients of MRI denominator $Q^Z_{[S-1/N]}$ forming the corresponding eigenvector. This allows for an efficient numerical solution via an "off-the-shelf" eigensolver.

Once $Q^Z_{[S-1/N]}$ has been found, the corresponding numerator can be built by solving a polynomial interpolation problem for $Q^Z_{[S-1/N]} v$. This can be done, e.g., by inverting the generalized square Vandermonde matrix Υ_{S-1}, cf. (5.2), with $\{\psi_i\}_{i=0}^{S-1}$ being either the same basis as for $Q^Z_{[S-1/N]}$ (with the addition of some elements if $N < S-1$) or a different basis entirely. Note that the interpolation problem involves linear combinations of the snapshots, so that the coefficients of $I^Z(Q^Z_{[S-1/N]} v)$ have the same size as v.

Overall, excluding the computation of the snapshots, the whole MRI-building procedure has complexity $\mathcal{O}\left(S^2(S + \dim(\mathcal{V}))\right)$. We summarize it in Algorithm 2, where, for notational simplicity, we use the same basis $\Psi_N \subset \Psi_{S-1}$ (with $\deg(\psi_i) \leq N$ for $i = 0, \dots, N$) for numerator and denominator.

Algorithm 2 MRI for distinct points

Require: distinct sample points $Z = \{z_1, \dots, z_S\} \subset \mathbb{C}$, sampler v
Require: basis $\Psi_{S-1} = \{\psi_0, \dots, \psi_{S-1}\}$, denominator degree $N < S$
  assemble Υ_N and C as in (5.2) and (5.3)
  for $j = 1, \dots, S$ **do**
    compute snapshot $v_j = v(z_j)$
  end for
  assemble the $S \times S$ snapshot Gramian G as in (2.44)
  find the minimal eigenvector $(q_0, \dots, q_N)^\top \in \mathbb{C}^{N+1}$ of $\Upsilon_N^H C^H G C \Upsilon_N$, e.g., via `eig`
  set $Q^Z_{[S-1/N]} = \sum_{i=0}^N q_i \psi_i$
  assemble Υ_{S-1} as in (5.2) and compute $\Upsilon_{S-1}^{-1} =: (\gamma_{ij})_{i=0,j=1}^{S-1,S}$
  set $P^Z_{[S-1/N]} = \sum_{i=0}^{S-1} \left(\sum_{j=1}^S \gamma_{ij} Q^Z_{[S-1/N]}(z_j) v_j \right) \psi_i$
  return $P^Z_{[S-1/N]} / Q^Z_{[S-1/N]}$

5.1.1 Implementation for coalesced points

When sample points appear multiple times in Z, e.g., when using fast LS Padé approximation, the general outline of the algorithm stays the same, but the matrices involved change slightly. Let us assume that the distinct elements of Z are $\{z_1, \dots, z_{S'}\}$, with z_j appearing $E_j + 1$ times in Z. To begin with, we need to extend the barycentric formula to this case: given the nodal polynomial $\omega^Z(z) = \prod_{j=1}^{S'}(z - z_j)^{E_j+1}$, we have

$$I^Z(\psi)(z) = \omega^Z(z) \sum_{j=1}^{S'} \sum_{l=0}^{E_j} \sum_{k=0}^{E_j-l} \frac{w_{jk}}{(z - z_j)^{E_j-l-k+1}} \frac{1}{l!} \frac{\mathrm{d}^l \psi}{\mathrm{d}z^l}(z_j), \tag{5.4}$$

with $\{w_{jk}\}_{j=1,k=0}^{S',E_j}$ *unique* weights that are available only "implicitly" as the solution of suitable Hermite interpolation conditions [SV13], which we do not report here for brevity.

Now, let the snapshots be indexed as $\{v_{jm}\}_{j=1,m=0}^{S',E_j}$, with $v_{jm} = \frac{\mathrm{d}^m v}{\mathrm{d}z^m}(z_j)$. By the Leibniz rule, we have

$$I^Z(Qv) = \omega^Z(z) \sum_{j=1}^{S'} \sum_{l=0}^{E_j} \sum_{k=0}^{E_j-l} \sum_{m=0}^{l} \frac{w_{jk} v_{jm}}{(z - z_j)^{E_j-l-k+1}} \frac{1}{m!(l-m)!} \frac{\mathrm{d}^{l-m} Q}{\mathrm{d}z^{l-m}}(z_j), \tag{5.5}$$

so that

$$\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv) = \sum_{j=1}^{S'} \sum_{l=0}^{E_j} \sum_{k=0}^{E_j-l} \sum_{m=0}^{l} \frac{w_{jk} v_{jm}}{m!(l-m)!} \frac{\mathrm{d}^{l-m} Q}{\mathrm{d}z^{l-m}}(z_j) \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} \left(\frac{\omega^Z}{(\cdot - z_j)^{E_j-l-k+1}} \right). \tag{5.6}$$

Whenever $E_j - l - k + 1 > 1$ in (5.6), $(\cdot - z_j)^{-E_j+l+k-1}\omega^Z$ has degree smaller than $S-1$, and the $(S-1)$-th derivative vanishes. Hence, we are only left with the terms corresponding to $l + k = E_j$, and

$$\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv) = \sum_{j=1}^{S'} \sum_{l=0}^{E_j} \sum_{m=0}^{l} c_{jlm} v_{jm} \frac{\mathrm{d}^{l-m} Q}{\mathrm{d}z^{l-m}}(z_j), \tag{5.7}$$

with $c_{jlm} = \frac{(S-1)!}{m!(l-m)!} w_{j(E_j-l)}$.

After setting $Q = \sum_{i=0}^{N} q_i \psi_i$ as usual, we obtain

$$J^Z(Q)^2 = \left\| \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} I^Z(Qv) \right\|_{\mathcal{V}}^2 = \left\| \sum_{i=0}^{N} \sum_{j=1}^{S'} \sum_{l=0}^{E_j} \sum_{m=0}^{l} c_{jlm} \frac{\mathrm{d}^{l-m} \psi_i}{\mathrm{d}z^{l-m}}(z_j) v_{jm} q_i \right\|_{\mathcal{V}}^2,$$

so that we may repeat the construction carried out in the Lagrangian case. We have two notable differences: the scaling matrix C, see (5.3), is no longer diagonal, and the generalized Vandermonde matrix Υ_N, see (5.2), contains also derivatives of the basis polynomials.

The case $S' = 1$, i.e., of fully coalesced points $Z = \{z_1, \dots, z_1\}$, admits a simplified formulation, and is a rather special case, corresponding to fast LS Padé approximation. For both these reasons, we report here the expressions of the weights $w_{1k} = \delta_{k0}$ for $k = 0, \dots, S-1$, of the generalized Vandermonde matrix

$$\Upsilon_N \in \mathbb{C}^{S \times (N+1)}, \qquad (\Upsilon_N)_{ji} = \frac{\mathrm{d}^{S-j} \psi_i}{\mathrm{d}z^{S-j}}(z_1), \qquad j = 1, \dots, S,\ i = 0, \dots, N, \tag{5.8}$$

of the lower-triangular scaling matrix

$$C \in \mathbb{C}^{S\times S}, \qquad (C)_{mj} = \begin{cases} c_{1(S-j+m)m} & \text{if } j > m, \\ 0 & \text{otherwise,} \end{cases} \qquad m = 0,\dots,S-1,\ j = 1,\dots,S, \quad (5.9)$$

and of the snapshot Gramian

$$G \in \mathbb{C}^{S\times S}, \qquad (G)_{m'm} = \langle v_{1m}\,, v_{1m'}\rangle_{\mathcal{V}}, \qquad m, m' = 0,\dots,S-1. \quad (5.10)$$

To conclude this section, we remark that repeated sample points are allowed only if we can query the solver for derivatives of v too. This is not always possible, especially if the dependence of the FOM on z is complicated, or if the solver is truly "black-box". In such cases, approximations of the derivatives of v (obtained, e.g., by finite differences) might be used instead. On top of this, we note that taking (too many) derivatives of meromorphic functions is not the most stable idea to begin with, since the effect of the most relevant poles is usually amplified, see, e.g., (4.7).

5.2 Numerical matters

The approach presented above, as summarized in Algorithm 2, represents the simplest possible algorithm for MRI. Still, before we can happily employ it in an industrial application, we should answer several questions. Among those, we consider the most important ones in the following sections: is the method numerically robust? What metric over $\mathcal{V}$ should we choose? What basis Ψ_N should we choose?

The even more crucial (in some sense) matter of how to choose Z follows in Section 5.3.

5.2.1 Conditioning and instabilities

We first explore the issue of numerical stability, assuming that all parameters of MRI (v, $\|\cdot\|_{\mathcal{V}}$, Z, N, Ψ_N) are fixed. For simplicity, we ignore stability issues in the sampling phase, since those should be dealt with by the sampler, which is responsible for providing us with the snapshots $\{v_j\}_{j=1}^S$.

We can identify three main steps that could be numerically problematic. We deal with them one by one.

5.2.1.1 Assembly of the scaling matrix C, see (5.3)

This requires computing the coefficients c_j. It turns out that building C is no more unstable that inverting the generalized square Vandermonde matrix Υ_{S-1}, an item that we will discuss in the next point. Indeed, for instance, when the sample points are distinct, the weights $\{c_j\}_{j=1}^S$ are actually the entries of the vector

$$(c_j)_{j=1}^S = \Upsilon_{S-1}^{-\top}\Upsilon_{S-1}^{(\text{lead})}, \qquad \text{with } \Upsilon_{S-1}^{(\text{lead})} = \left[\frac{\mathrm{d}^{S-1}\psi_0}{\mathrm{d}z^{S-1}},\dots,\frac{\mathrm{d}^{S-1}\psi_{S-1}}{\mathrm{d}z^{S-1}}\right]^\top \in \mathbb{C}^S. \quad (5.11)$$

To see this, it suffices to note that

$$c_j = \frac{(S-1)!}{\prod_{j'=1,j'\neq j}^{S}(z_j - z_{j'})} = \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}}\left(\frac{\omega^Z}{(\cdot - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)}\right) = \frac{\mathrm{d}^{S-1}\ell_j}{\mathrm{d}z^{S-1}},$$

with $\ell_j \in \mathbb{P}_{S-1}(\mathbb{C};\mathbb{C})$ the j-th Lagrange polynomial. Now, ℓ_j can be found by inversion of Υ_{S-1}, as the unique polynomial of degree smaller than S that satisfies $\ell_j(z_k) = \delta_{jk}$, the Kronecker delta:

$$\ell_j(z) = \sum_{i=0}^{S-1} \left(\Upsilon_{S-1}^{-1}\mathbf{e}_j\right)_i \psi_i(z), \quad \text{with } (\mathbf{e}_j)_i = \delta_{ij}.$$

(5.11) follows by taking $S-1$ derivatives in the identity above.

Note that, if sample points are repeated, more coefficients might need to be computed, and the weights w_{jk} must also be found, see Section 5.1.1. Regardless, similar arguments apply in that case too.

5.2.1.2 Inversion of the generalized square Vandermonde matrix Υ_{S-1}, see (5.2)

The robustness of this operation rests mostly on the choice of Z and Ψ_{S-1}, which will be discussed in Section 5.2.3. Still, without changing Z or Ψ_{S-1}, there is a (quite obvious) alternative for improving numerical stability: moving from interpolation to LS approximation. Indeed, in Definition 3.2, we define the MRI numerator as the degree-$(S-1)$ interpolant $I^Z(Q^Z_{[S-1/N]}v)$ in order to exploit the available information as much as possible. However, in practice, it might make sense to sacrifice interpolation in the interest of stability. This can be achieved by tweaking Definition 3.2, replacing the interpolation operator I^S with an LS polynomial approximation operator: e.g., given distinct sample points Z, $M < S-1$, and $\psi : \mathbb{C} \to \mathcal{V}$ such that $I^Z(\psi)$ is defined, we consider

$$I^Z_M(\psi) := \operatorname*{arg\,min}_{P\in\mathbb{P}_M(\mathbb{C};\mathcal{V})} \sum_{j=1}^{S} w_j \left\|P(z_j) - \psi(z_j)\right\|^2_{\mathcal{V}}, \tag{5.12}$$

with $\{w_j\}_{j=1}^S \subset \mathbb{R}_{>0}$ suitable weights. The derivative order $S-1$ in (3.17b) should then be set to M in order to capture correctly the leading coefficient of I^Z_M.

Note that, when using an "LS version of MRI", the optimal snapshot usage is lost, thus making the resulting approach more similar to standard rational approximation MOR techniques, see Section 2.4.1.

5.2.1.3 Solution of minimal eigenproblem for $\Upsilon_N^H C^H G C \Upsilon_N$

The numerical stability of eigenpair computations is usually analyzed using Weyl's inequality or similar tools, which allow to bound the spectrum of a (Hermitian) matrix affected by numerical noise by using the spectrum of the unperturbed matrix and the spectrum of the noise. However, this has limitations for minimal eigenvalues, as we proceed to showcase with a fairly general example.

Let the Hermitian matrix A have eigenvalues $0 \leq \sigma_n(A) \leq \ldots \leq \sigma_1(A)$, and consider the Hermitian perturbation Δ, with eigenvalues $\sigma_n(\Delta) \leq \ldots \leq \sigma_1(\Delta)$, which we assume small (in

magnitude). By Weyl's law, the two smallest eigenvalues of $A + \Delta$ admit the bounds

$$\sigma_n(A) + \sigma_n(\Delta) \leq \sigma_n(A + \Delta) \leq \sigma_n(A) + \sigma_1(\Delta)$$

and

$$\sigma_{n-1}(A) + \sigma_n(\Delta) \leq \sigma_{n-1}(A + \Delta) \leq \sigma_{n-1}(A) + \sigma_1(\Delta).$$

In particular, if $\sigma_1(\Delta) - \sigma_n(\Delta)$ is larger than the *spectral gap* for $\sigma_n(A)$, namely, $\sigma_{n-1}(A) - \sigma_n(A)$, then $\sigma_{n-1}(A + \Delta)$ and $\sigma_n(A + \Delta)$ might be arbitrarily close, or even coincide. This, generally, is not an unsurmountable issue for minimal eigen*value* approximation, since the dependence of the minimal eigen*value* on the perturbation is continuous. However, approximating the corresponding minimal eigen*vector* is troublesome, as it does not depend continuously on Δ (in fact, it is not even uniquely defined when the minimal eigenvalue is semisimple, i.e., $\sigma_{n-1}(A+\Delta) = \sigma_n(A+\Delta)$).

One fairly straightforward check that can be introduced to detect such situations is: given $A := \Upsilon_N^H C^H G C \Upsilon_N$, compute $\sigma_n(A)$ and the corresponding eigenvector (which gives the coefficients of $Q^Z_{[S-1/N]}$), but also $\sigma_{n-1}(A)$ and $\sigma_1(A)$. If the relative spectral gap $\frac{\sigma_{n-1}(A)-\sigma_n(A)}{\sigma_1(A)-\sigma_n(A)}$ is too small, e.g., below a user-defined tolerance (a typical value is 10^{-14}), mark $Q^Z_{[S-1/N]}$ as unreliable, decrease N, and repeat the calculation. This, effectively, allows an on-the-fly selection of N, based on the conditioning of the problem. As a side note, we observe that one may also decrease M together with N, thus making the approach non-interpolatory, cf. Section 5.2.1.2.

Before proceeding to other issues, we wish to describe a strategy that can considerably help reduce the ill-conditioning of the eigenproblem. The key idea is: since G is a Gramian matrix, $A = \Upsilon_N^H C^H G C \Upsilon_N$ has itself a Gramian structure. More precisely, let

$$U = \left[v_1 \middle| \cdots \middle| v_S \right] \tag{5.13}$$

be the *snapshot quasi-matrix*, whose columns are elements of $\mathcal{V}$, and let

$$U = \left[w_1 \middle| \cdots \middle| w_S \right] R, \qquad \text{with } \langle w_j \, , w_{j'} \rangle_{\mathcal{V}} = \delta_{jj'} \text{ and } R \in \mathbb{C}^{S \times S}, \tag{5.14}$$

be its QR decomposition, obtained, e.g., by Householder triangularization [Tre09]. Then $G = R^H R$, and

$$A = \left(RC\Upsilon_N \right)^H \left(RC\Upsilon_N \right),$$

so that the eigenvalues and eigenvectors of A are squared singular values and right singular vector pairs for $RC\Upsilon_N$, respectively. Crucially, this means that the conditioning of the problem, e.g., as measured by the relative spectral gap, necessarily improves, since

$$\frac{\sigma_{n-1}(A) - \sigma_n(A)}{\sigma_1(A) - \sigma_n(A)} < \frac{\sqrt{\sigma_{n-1}(A)} - \sqrt{\sigma_n(A)}}{\sqrt{\sigma_1(A)} - \sqrt{\sigma_n(A)}} = \frac{\sigma_{n-1}(RC\Upsilon_N) - \sigma_n(RC\Upsilon_N)}{\sigma_1(RC\Upsilon_N) - \sigma_n(RC\Upsilon_N)}$$

whenever $\sigma_1(A) > \sigma_{n-1}(A)$.

5.2.2 Choice of metric

In the definition of MRI, namely Definition 3.2, the "most well-hidden" parameter on which MRI depends is the norm $\|\cdot\|_{\mathcal{V}}$, which affects the rational interpolant in a rather convoluted way, by defining the metric in which the optimization target J^Z is measured, see (3.17b).

In most applications, one can identify the most natural choice for this metric. For instance, in finite-dimensional non-homogeneous eigenproblem-like settings, e.g.,

$$(zE - A)v(z) = b \qquad \text{with } A, E \in \mathbb{C}^{n\times n} \text{ and } v(z), b \in \mathbb{C}^n = \mathcal{V},$$

with E positive definite, it is reasonable to take the inner product induced by E, namely, $\langle v, w\rangle_{\mathcal{V}} = w^H E v$ for $v, w \in \mathcal{V}$. Note that frequency-domain LTI dynamical systems usually fall in this category.

On the other hand, in some cases, a functional viewpoint should be taken, notably when $v(z)$ is the solution field of a PDE, with $\mathcal{V}$ an infinite-dimensional Hilbert space, e.g., $H^1(\Omega)$ in some of the examples from Section 2.3. In such settings, one should select $\|\cdot\|_{\mathcal{V}}$ as a functional norm with respect to which $\mathcal{V}$ is Banach, e.g., $\|v\|_{H^1(\Omega)} = \left(\||\mathrm{grad}\, v|\|^2_{L^2(\Omega)} + \|v\|^2_{L^2(\Omega)}\right)^{1/2}$ for $\mathcal{V} = H^1(\Omega)$.

All this being said, nothing prevents choosing a different norm, or, in fact, even a seminorm, in the definition of MRI. Of course, this is accompanied by the observation that, if the chosen (semi)norm is not Hilbertian, i.e., if there is no associated inner product, then finding the minimum of J^Z might be trickier. In particular, expressing the minimization problem as a minimal eigenproblem, cf. Section 5.1, may not be possible.

Until here, we have only discussed what norms *can* be used, but now we turn to the issue of what norms *should* be chosen, so as to improve the approximation quality. In general, the choice of $\|\cdot\|_{\mathcal{V}}$ enters MRI in an extremely complicated way, so that a precise answer is impossible. However, we can obtain a partial indication from the theoretical results from Sections 3.1.1, 3.1.2, 3.2.1, and 3.2.2: if at all possible, one should choose an inner product that makes the residues of v orthogonal, since this improves the convergence rate of the poles, as well as the constant in the error convergence (at least in theory, cf. Conjecture 1).

Unfortunately, this is usually impossible, since, in most cases, either (i) one cannot tell *a priori* if such an inner product exists, or (ii) one can show *a priori* that such an inner product does not exists. However, there are some families of frequency-domain problems where this is a valid option: for instance, Hermitian dynamical systems with invertible E, cf. Section 2.2, or self-adjoint (or normal) PDEs, cf. Sections 2.3.1 and 2.3.2. In such cases, one should strive to select an inner product that makes the problem Hermitian/self-adjoint, or, more generally, normal.

As a concluding remark, in the author's experience, it seems that the actual impact of the choice of metric on MRI is less than could be expected. Indeed, we have observed that target functions v that are "nice", e.g., that have orthogonal residues in a certain metric, are approximated well even if MRI is applied with a different metric, with respect to which the residues are not orthogonal. Some numerical experiments in this direction are carried out in Sections 5.5.2 and 7.1.

5.2.3 Choice of polynomial basis and barycentric extension

Here we explore the question of what metric should be used on the polynomial space $\mathbb{P}_N(\mathbb{C};\mathbb{C})$ to define the unit ball $\mathbb{P}_N^{\Psi_N}(\mathbb{C};\mathbb{C})$, the search space for the optimal denominator $Q^Z_{[S-1/N]}$. We note that this question is complementary to that considered in the previous section, where the metric on the target space was discussed. This time, the theory cannot really give us any meaningful insights (except that we might want to try to satisfy Assumption 3.4 if we plan to have large values of N). Instead, in trying to identify the "best" polynomial basis Ψ_N, which induces the sought-after metric through (3.17a), we aim mostly at good numerical properties.

In this context, the main objective is ensuring that the generalized Vandermonde matrices Υ_N and Υ_{S-1} are reasonably well-conditioned, since, after all, an unstable construction of the numerator $I^Z(Q^Z_{[S-1/N]}v)$ could jeopardize the whole rational approximant. To this aim, one should follow the usual practices for numerical polynomial approximation and interpolation. For instance, if Z are located at the S-th roots of unity, then monomials can be recommended. Similarly, if Z are the Chebyshev nodes, then Chebyshev polynomials are the natural choice.

In both above-mentioned cases, a "good" choice of Ψ_N is possible as a direct consequence of the sample points being "well-chosen" themselves. However, this is not the case in many practical situations. For instance, sample points may have unfavorable positions when the data comes from experimental measurements, or when sample points are added adaptively, see Section 5.3. Luckily, a universal answer exists, guaranteeing well-posedness of the interpolation endeavor: the (sample points-dependent) Lagrangian basis, defined as

$$\Psi_{S-1} = \Psi_{S-1}(Z) = \left\{ \frac{\omega^Z}{(\cdot - z_i)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_i)} \right\}_{i=1}^{S} = \left\{ \prod_{\substack{j=1 \\ j\neq i}}^{S} \frac{\cdot - z_j}{z_i - z_j} \right\}_{i=1}^{S}. \tag{5.15}$$

Note that this expression only holds for distinct sample points. A Hermite-Lagrangian basis is available in the general case, cf. the generalized barycentric form (5.4). By definition, the generalized square Vandermonde matrix for this basis is $\Upsilon_{S-1} = I$, the identity matrix.

Unfortunately, this incurs in a few issues. On one hand, the basis is not hierarchical, so that, if $N < S-1$, a basis for $\mathbb{P}_N(\mathbb{C};\mathbb{C})$ cannot be built from elements of Ψ_{S-1}, and it becomes necessary to select a different basis for $Q^Z_{[S-1/N]}$. Note that, in this regard, one can choose a different basis entirely, or again a Lagrangian one, by picking $N+1$ support points out of the S sample points. On the other hand, and more importantly, even if the construction of the MRI surrogate proceeds smoothly thanks the Lagrangian basis, the evaluation of polynomials in the Lagrangian basis is usually highly unstable, thus making the evaluation of the ROM in the online phase unreliable.

In an almost miraculous way, there exists a way to take advantage of the Lagrangian property without sacrificing online stability: a barycentric formulation, taking inspiration from the Loewner framework and AAA, see Section 2.4.1. We provide here a definition of the barycentric formulation of MRI. Note that, for conciseness, we restrict our focus to distinct sample points, but generalizations to coalesced sample points are possible, although quite heavy in notation.

Definition 5.1 (Barycentric MRI)**.** *Let $\mathcal{V}$ be as in Definition 3.2. We consider distinct sample points $Z = \{z_j\}_{j=1}^S \in \mathbb{C}$. Let $v : \mathbb{C} \to \mathcal{V}$ be a $\mathcal{V}$-valued function such that $I^Z(v)$ exists, according to Definition 2.1. An $[S-1/S-1]$ barycentric MRI of v based on samples at Z is a rational function*

$$v^Z_{[S-1]}(z) = \left(\sum_{j=1}^{S} \frac{(Q^Z_{[S-1]})_j v(z_j)}{z - z_j} \right) \Bigg/ \left(\sum_{j=1}^{S} \frac{(Q^Z_{[S-1]})_j}{z - z_j} \right), \tag{5.16}$$

such that

$$Q^Z_{[S-1]} \in \partial B^{\mathbf{0}}(1) = \left\{ [q_1, \dots, q_S]^\top \in \mathbb{C}^S : \sum_{j=1}^{S} |q_j|^2 = 1 \right\}, \tag{5.17a}$$

$$\bar{J}^Z(Q^Z_{[S-1]}) \leq \bar{J}^Z(Q) := \left\| \sum_{j=1}^{S} (Q)_j v(z_j) \right\|_{\mathcal{V}} \quad \forall Q \in \partial B^{\mathbf{0}}(1). \tag{5.17b}$$

At first blush, it might not be obvious that $v^Z_{[S-1]}$ is a rational function. Still, it suffices to multiply both numerator and denominator of (5.16) by the nodal polynomial $\omega^Z(z) = \prod_{j=1}^S (z - z_j)$ to make the "true" polynomial numerator and denominator emerge:

$$v^Z_{[S-1]}(z) = \frac{\omega^Z(z) \sum_{j=1}^S \frac{(Q^Z_{[S-1]})_j v(z_j)}{z-z_j}}{\omega^Z(z) \sum_{j=1}^S \frac{(Q^Z_{[S-1]})_j}{z-z_j}} = \frac{\sum_{j=1}^S (Q^Z_{[S-1]})_j v(z_j) \prod_{j'=1,j'\neq j}^S (z - z_{j'})}{\sum_{j=1}^S (Q^Z_{[S-1]})_j \prod_{j'=1,j'\neq j}^S (z - z_{j'})}. \tag{5.18}$$

We can see that the rational form employs a scaled Lagrangian basis

$$\Psi_{S-1} = \Psi_{S-1}(Z) = \left\{ \frac{\omega^Z}{(\cdot - z_i)} \right\}_{i=1}^S = \left\{ \prod_{\substack{j=1 \\ j\neq i}}^S (\cdot - z_j) \right\}_{i=1}^S \tag{5.19}$$

that satisfies $\psi_i(z_{i'}) = \delta_{ii'} \prod_{j=1,j\neq i}^S (z_i - z_j)$. It is customary to use the barycentric form (5.16) rather than the rational one (5.18), since only the former can be reliably evaluated online in a robust manner, see, e.g., [BT04; NST18].

We observe that Definition 5.1 is, indeed, an extension of MRI, since (i) the denominator is found by minimizing the leading coefficient of the (polynomial) numerator

$$\frac{1}{(S-1)!} \frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}} \left(\sum_{j=1}^S (Q^Z_{[S-1]})_j v(z_j) \prod_{\substack{j'=1 \\ j'\neq j}}^S (\cdot - z_{j'}) \right) = \sum_{j=1}^S (Q^Z_{[S-1]})_j v(z_j),$$

cf. (5.18), and (ii) the approximation is interpolatory:

$$\lim_{z\to z_j} v^Z_{[S-1]}(z) = \lim_{z\to z_j} \frac{\frac{(Q^Z_{[S-1]})_j v(z_j)}{z-z_j}}{\frac{(Q^Z_{[S-1]})_j}{z-z_j}} = v(z_j).$$

In particular, we note that barycentric MRI is, by definition, a diagonal rational approximation, since it forces $N = S - 1$. Still, technically, one could reduce the degree of the denominator to $N < S$ by adding $S - 1 - N$ linear constraints to the space $B^{\mathbf{0}}(1)$ where $Q^Z_{[S-1]}$ is sought, cf. (5.17a), as is sometimes done in scalar barycentric rational interpolation, see, e.g., [Kle12]. For simplicity, we ignore this possibility here.

As anticipated, the construction of a barycentric MRI requires no (inversions of) Vandermonde matrices, since interpolation is guaranteed by definition, and the functional $\bar{J}^Z(\cdot)^2$ is a quadratic form, with the snapshot Gramian G, see (2.44), as representative matrix. This, notably, leads to the quite remarkable conclusion that the denominator coefficients $Q^Z_{[S-1]}$ can be found in a very straightforward way as a minimal eigenvector of the snapshot Gramian.

Concerning the theoretical properties of barycentric MRI, we first note that, since we are forcing $N = S - 1$, the "fixed N" convergence theory from Section 3.2 is irrelevant here. On the other hand, it is sensible to ask whether the "$N \to \infty$" results, namely, Theorems 3.6 and 3.8, can be extended to barycentric MRI. The answer is not trivial, since the scaled Lagrangian basis (5.19) does not necessarily satisfy Assumption 3.4. Notably, showing whether the lower bound in Assumption 3.4 holds or not is a rather tricky issue. For this reason, we believe that it would be

more appropriate to develop a theory for barycentric MRI "from scratch", rather than trying to make the method fit in the framework from Section 3.2. To this aim, in the author's opinion, it should be possible to extend the proofs of Theorems 3.6 and 3.8 by relying on the specific nodal form of the barycentric basis (5.19), rather than on Assumption 3.4. For the moment, we leave this as a conjecture.

5.3 Adaptive frequency sampling

Until now, we have assumed that the sample points Z are fixed in advance by the user. This is a common framework when the user wishes to limit the number of times the FOM is solved. However, in practice, the number of samples might not be high enough to provide a good approximation of the target function over the region of interest. For instance, if the sampling set A contains N' poles of v (counting multiplicity), then at least $S = N' + 1$ snapshots are necessary to build a surrogate of the appropriate type, and a (slightly) higher number of them might be advisable to improve the approximation quality. However, in practice, since N' is not known in advance, this does not help us in choosing S.

In this context, an adaptive addition of sample points, in the same spirit as the weak-greedy RB method, see Section 2.4.2.3, seems appropriate. We summarize the skeleton of the z-adaptive scheme in Algorithm 3. Note that we need a polynomial basis *generator*, rather than just a basis, since the size of the basis will necessarily change at each iteration. To this aim, a natural choice consists in taking a hierarchical basis (e.g., monomials or Chebyshev polynomials). Also, we remark that, for simplicity, we always set $N = S - 1$, i.e., we force a diagonal approximation. In general, a different choice of N (which should increase with S) could be prescribed by the user. Still, if the automatic N-reduction from Section 5.2.1 is applied, the algorithm should be able to adapt N on-the-fly, thus preserving numerical stability. When looking at offline efficiency, one should note that, as the iterations proceed, the MRI surrogate must be rebuilt from scratch each time. However, starting from the second iteration, most of the snapshots have already been computed at a previous iteration, so that, effectively, each call to Algorithm 2 "costs" only one snapshot and the construction of the surrogate.

Algorithm 3 z-adaptive MRI (for distinct points)

Require: distinct initial sample points $Z = \{z_1, \dots, z_{S_0}\} \subset \mathbb{C}$, sampler v, tolerance ϵ
Require: distinct test points $Z_{\text{test}} = \{z_1, \dots, z_T\} \subset \mathbb{C} \setminus Z$
Require: basis generator $\Psi(S) = \{\psi_0, \dots, \psi_{S-1}\}$ for all $S \geq S_0$
  for $S = S_0, S_0 + 1, \dots$ **do**
    build MRI $v^Z_{[S-1/S-1]}$ via Algorithm 2 using the basis $\Psi(S)$
    evaluate the greedy indicator $\eta(z)$ at all $z \in Z_{\text{test}}$
    find the point of worst approximation $z^\star = \arg\max_{z \in Z_{\text{test}}} \eta(z)$
    if $\eta(z^\star) < \epsilon$ **then**
      return $v^Z_{[S-1/S-1]}$
    end if
    move $z^\star$ from Z_{test} to Z
  end for

The key step of Algorithm 3 is the evaluation of the error indicator η at all points of the test Z_{test}, which should be a fine grid of points of the sampling set A. Obviously, for the sake of offline efficiency, the computation of $\eta(z)$ must not require the snapshot $v(z)$. For this reason, we use the term "error indicator" rather than just "error", since the latter cannot be evaluated

efficiently, whereas the former might. Several possible definitions of η have been proposed by us in [PN20]. We proceed by summarizing them here.

5.3.1 Intrusive exact affine$^{\mathrm{MOR}}$ residual

Assume that the problem depends on frequency in an affine$^{\mathrm{MOR}}$ way. We recall the residual expansion (2.45), which, in the scope of the RB method, allows to evaluate the residual of the surrogate model at arbitrary (test) frequencies, with a computational cost that is independent of the size of the original problem. This comes at the price of pre-computing (at modest computational cost) the terms appearing in the affine$^{\mathrm{MOR}}$ expansion (2.45) of the residual.

Such expansion can be recycled for MRI. To this aim, it suffices to observe that, by definition of the MRI numerator, we can always expand $v^Z_{[S-1/N]}$ on the snapshot basis: e.g., if the sample points are distinct, the barycentric formula (2.2) yields

$$v^Z_{[S-1/N]}(z) = \frac{1}{Q^Z_{[S-1/N]}(z)} \sum_{j=1}^{S} \frac{\omega^Z(z) Q^Z_{[S-1/N]}(z_j) v_j}{(z - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} = \sum_{j=1}^{S} \alpha_j(z) v_j. \tag{5.20}$$

Notably, for MRI, the coefficients $\{\alpha_j\}_{j=1}^S$ are rational functions, available in closed form, whereas, for projective methods, they are given implicitly as the solution of the reduced system. As such, in general, their evaluation is cheaper for MRI. Note that using the affine$^{\mathrm{MOR}}$ residual expansion (2.45) necessarily makes the overall approach intrusive.

5.3.2 Partially intrusive affine residual

Consider a problem of the form $(zE - A)v(z) = b + zb'$, i.e., a problem with affine dependence on z in both left- and right-hand-sides. Note that this family includes frequency-domain LTI dynamical systems (provided the forcing term is affine). We have the following result.

Lemma 5.1. *Let v be the solution of a linear problem with affine dependence on z of both left- and right-hand-side: $(zE - A)v(z) = b + zb'$. Assume that $v^Z_{[S-1/N]}$ is the MRI obtained from samples at the distinct points Z, according to Definition 3.2. Then, we can express the residual as*

$$(zE - A)v^Z_{[S-1/N]}(z) - b - zb' = c\frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)}, \tag{5.21}$$

with $c \in \mathcal{V}^\star$ independent of z. An explicit expression for c is given in the proof.

Proof. Let $\omega^Z(z) = \prod_{z' \in Z}(z - z')$. We apply the barycentric formulas (2.2) and (5.20) several times to obtain

$$\begin{aligned}
(zE - A)v^Z_{[S-1/N]}(z) =& \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)} \sum_{j=1}^{S} \frac{Q^Z_{[S-1/N]}(z_j)(zE - A)v_j}{(z - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \\
=& \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)} \sum_{j=1}^{S} \frac{Q^Z_{[S-1/N]}(z_j)\left((z - z_j)Ev_j + (z_jE - A)v_j\right)}{(z - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)} \\
=& \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)} \underbrace{\sum_{j=1}^{S} \frac{Q^Z_{[S-1/N]}(z_j)Ev_j}{\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)}}_{c'} + \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)} \sum_{j=1}^{S} \frac{bQ^Z_{[S-1/N]}(z_j)}{(z - z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)}
\end{aligned}$$

$$+\frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)}\sum_{j=1}^{S}\frac{z_j b' Q^Z_{[S-1/N]}(z_j)}{(z-z_j)\frac{\mathrm{d}\omega^Z}{\mathrm{d}z}(z_j)}$$
$$=\frac{\omega^Z(z)c' + I^Z\left(bQ^Z_{[S-1/N]}\right)(z) + I^Z\left(b'\widetilde{Q}^Z_{[S-1/N]}\right)(z)}{Q^Z_{[S-1/N]}(z)},$$

with $\widetilde{Q}^Z_{[S-1/N]}$ denoting the polynomial $z \mapsto zQ^Z_{[S-1/N]}(z)$. Note that we have applied the definition of the snapshots $v_j = v(z_j)$ when using the identity $(z_jE - A)v_j = b + z_jb'$.

By exactness of polynomial interpolation, $I^Z(bQ^Z_{[S-1/N]}) = bQ^Z_{[S-1/N]}$. In general, the same cannot be said about $I^Z(b'\widetilde{Q}^Z_{[S-1/N]})$, since $\widetilde{Q}^Z_{[S-1/N]}$ is a polynomial whose degree might be larger than $S-1$. Still, given $\gamma = \frac{1}{(S-1)!}\frac{\mathrm{d}^{S-1}}{\mathrm{d}z^{S-1}}Q^Z_{[S-1/N]}$ (which is zero if $\deg(Q^Z_{[S-1/N]}) < S-1$), we have

$$zb'Q^Z_{[S-1/N]}(z) = \gamma b'\omega^Z(z) + b'\left(zQ^Z_{[S-1/N]}(z) - \gamma\omega^Z(z)\right) = \gamma b'\omega^Z(z) + b'\widehat{Q}^Z_{[S-1/N]}(z),$$

with $\deg(\widehat{Q}^Z_{[S-1/N]}) \leq S-1$ since ω^Z is monic. Since ω^Z vanishes over Z, it follows that $I^Z\left(\gamma b'\omega^Z\right) = 0$, and we obtain

$$I^Z\left(b'\widetilde{Q}^Z_{[S-1/N]}\right)(z) = b'\widehat{Q}^Z_{[S-1/N]}(z) = zb'Q^Z_{[S-1/N]}(z) - \gamma b'\omega^Z(z).$$

Putting everything together, we obtain the desired identity

$$(zE-A)v^Z_{[S-1/N]}(z) - b - zb' = \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)}\left(c' - \gamma b'\right).$$

□

Remark 5.1. *The proof above can be adjusted to show some potentially useful additional results. Namely, let v be the solution of $(zE-A)v(z) = b + zb'$, and consider the interpolation operator I^Z from Definition 2.1, with Z being distinct points. If $\widetilde{v} = I^Z(vQ)/Q$ for some polynomial $Q \in \mathbb{P}_{S-1}(\mathbb{C};\mathbb{C}) \setminus \{0\}$, then we have*

$$(zE-A)\widetilde{v}(z) - b - zb' = c\frac{\omega^Z(z)}{Q(z)} \quad \text{with } c \in \mathcal{V}^\star \text{ independent of } z. \tag{5.22}$$

Equivalently stated, the choice of the denominator is irrelevant for Lemma 5.1.

Moreover, if v is the solution of $(zE-A)v(z) = b$ and $\widetilde{v} = I^Z(vQ)/Q$ with $Q \in \mathbb{P}_S(\mathbb{C};\mathbb{C}) \setminus \{0\}$, then (5.22) holds. Quite remarkably, this implies that, when $b' = 0$, (5.22) holds for the RB method as well, since the RB surrogate relying on S snapshots is an interpolatory rational function of type $[S-1/S]$, see Section 2.4.2.

Lemma 5.1 makes the z-dependence of the residual norm explicit:

$$\eta(z) = \left\|(zE-A)v^Z_{[S-1/N]}(z) - b - zb'\right\|_{\mathcal{V}^\star} = C\left|\frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)}\right|, \tag{5.23}$$

with $C = \|c\|_{\mathcal{V}^\star}$ independent of z. (Note that, for barycentric MRI, this simplifies even further:

$\eta(z) = C \left| \sum_{j=1}^{S} (Q^Z_{[S-1]})_j / (z - z_j) \right|^{-1}$.) As such, the point $z^\star$ that maximizes η over Z_{test} can be found only using scalar operations, as

$$z^\star = \arg\max_{z \in Z_{\text{test}}} \left| \frac{\omega^Z(z)}{Q^Z_{[S-1/N]}(z)} \right|. \tag{5.24}$$

This being said, we still need to compute C in order to tell whether the tolerance ϵ is attained. To this aim, it suffices to perform a single residual evaluation, e.g., at $z^\star$, and then set

$$C := \left\| (z^\star E - A) v^Z_{[S-1/N]}(z^\star) - b - z^\star b' \right\|_{\mathcal{V}^\star} \left| \frac{Q^Z_{[S-1/N]}(z^\star)}{\omega^Z(z^\star)} \right|.$$

This, in principle, does not require an intrusive access to the system matrices, but assumes that we can query the solver for (the dual norm of) a residual, something that might not be possible in a truly non-intrusive setting.

Before continuing, we note that this strategy (and the upcoming ones too) could be applied in a heuristic way also to more general problems, whose z-dependence is more complicated than just affine. If the "second variations" of left- and right-hand-sides of the problem are small, we can expect the point $z^\star$ in (5.24) to be close to the maximizer of the actual residual norm.

5.3.3 Non-intrusive affine error

Let us consider the framework of the "partially intrusive affine residual" estimator introduced above, but assume that, in the scope of identifying the scaling factor C in (5.23), it is not possible to evaluate the residual at $z^\star$. The point $z^\star$ can still be computed in a non-intrusive fashion, only relying on Z and $Q^Z_{[S-1/N]}$, since it does not depend on C. So, the only issue is determining whether to terminate the greedy iterations or to add $z^\star$ to Z and continue with the next iteration. To this aim, we can employ, somewhat heuristically, the error norm $e(z^\star) = \left\| v^Z_{[S-1/N]}(z^\star) - v(z^\star) \right\|_{\mathcal{V}}$, rather than the residual norm, in the termination criterion. More precisely, we follow these steps:

- Find $z^\star$ as in (5.24).
- Compute the error $e(z^\star)$.
- If $e(z^\star)$ is smaller than the tolerance, terminate. Otherwise, continue with the next iteration.

What we are doing is, in effect, "borrowing" the z-dependence from the residual and the scaling factor from the error.

We remark that error and residual depend differently on z, since the error depends on several additional quantities that are not available *a priori*, notably, on the exact poles of the system. In particular, the residual maximizer $z^\star$, see (5.24), might not coincide with the error maximizer. This makes our approach heuristic, since, even if $e(z^\star) < \epsilon$, the tolerance might not be attained at all other points of Z_{test}.

Note that, in order to compute $e(z^\star)$, we are forced to take a snapshot at $z^\star$. Still, this should not be seen as a waste of resources, for a snapshot at $z^\star$ is precisely the one needed at the next

greedy iteration. Only the snapshot at the last iteration is effectively wasted. At the same time, we observe that one could actually build a new MRI surrogate in post-processing, including also the extra sample point. In the `RROMPy` package, we carry out this extra step by default. However, it is quite interesting to note that the "updated" MRI surrogate is, in general, not guaranteed to perform better than the "non-updated" one. This is due to the non-monotonicity of e and η with respect to S, which, in turn, is caused by the meromorphicity of v and $v^Z_{[S-1/N]}$.

5.3.4 Non-intrusive affine collinearity

In the previous paragraphs, we have described how an exact (for affine problems) expression of the residual can be employed to find $z^\star$, while a different indicator is used to determine whether convergence has been reached. Here, we use the same idea, with a non-standard convergence indicator, based on snapshot collinearity. More precisely, given the span of the snapshots $\widetilde{\mathcal{V}} = \mathrm{Span}\{v_j\}_{j=1}^S$ and the new point $z^\star$, we compute the new snapshot $v(z^\star)$, which will potentially become v_{S+1}. We determine that the greedy algorithm has converged if

$$\left\| v(z^\star) - P_{\widetilde{\mathcal{V}}} v(z^\star) \right\|_{\mathcal{V}} = \min_{w \in \widetilde{\mathcal{V}}} \|v(z^\star) - w\|_{\mathcal{V}} < \epsilon \|v(z^\star)\|_{\mathcal{V}}, \tag{5.25}$$

with $P_{\widetilde{\mathcal{V}}}$ the $\mathcal{V}$-orthogonal projection onto $\widetilde{\mathcal{V}}$. This kind of condition was first considered in [RM18], and corresponds to the idea that, the "farther" the new snapshot is from the span of the previous ones, the more information it provides to the surrogate. A rigorous motivation based on ideas from POD, see Section 2.4.2.2, and on the linear independence of the most relevant residues of v, see Assumption 3.2, is possible, but we skip it here.

Note that, as in the previous case, we can recycle the final snapshot (taken for testing purposes), by building a new MRI surrogate in a post-processing step, including also the last sample point.

5.4 Adaptive frequency range partitioning

The adaptive sampling strategy presented in the previous section has quite useful features, which make it applicable in many practical settings. However, it also has some limitations. The main one is that, although the sample points are chosen optimally for approximation purposes, they might not have good numerical stability properties. In particular, the related Vandermonde matrix might be rather ill-conditioned, cf. Section 5.2.1.2. Moreover (and even if the barycentric basis is used to ensure good Vandermonde-related conditioning) instabilities might appear when considering the eigenproblem whose solution determines $Q^Z_{[S-1/N]}$, cf. Section 5.2.1.3. In the latter case, we have proposed in Section 5.2.1.3 to reduce the number N of approximated poles as a way to improve the conditioning. Still, the number of exact poles might be larger than the largest N that the eigensolver can handle, resulting necessarily in exact poles that cannot be approximated.

In case of numerical instability, we propose in [PN21] a simple, yet effective, solution, based on an adaptive partitioning of the frequency range. Our suggested method can handle rather large frequency ranges, over which a single MRI could not be built in a numerically stable fashion. The main price to pay is the increased complexity of the overall surrogate, as well as its reduced global regularity (it is only piecewise-rational rather than rational). Before proceeding, we note that similar ideas can be found in [HDO11] for weak-greedy RB.

For simplicity, assume that the frequency range is a line segment[1] $A = [z_\mathrm{L}, z_\mathrm{R}]$. If any instability arises when computing the MRI surrogate on A, then we split A at some point $z_\mathrm{C} \in A$, into two sub-intervals $A_1 = [z_\mathrm{L}, z_\mathrm{C}]$ and $A_2 = [z_\mathrm{C}, z_\mathrm{R}]$. Then, we try to build a new MRI surrogate on each sub-interval. Note that, for the sake of (offline) efficiency, we should try to reuse the already-computed snapshots in the sub-interval to which they belong. If we encounter any additional instability, we partition further, and so on. At the end of this algorithm (assuming that an end comes), we have a partition $A = \bigsqcup_{j=1}^{T} A_j$, such that each sub-interval is associated to the MRI surrogate H_j. In the online phase, if we wish to evaluate the overall surrogate at $z \in \mathbb{C}$, we just find the sub-interval $A_{j'(z)}$ that is closest to z, i.e.,

$$j'(z) = \operatorname*{arg\,min}_{j=1,\dots,T} \min_{z' \in A_j} |z - z'| \,,$$

and use $H_{j'(z)}(z)$ as surrogate value.

The only missing ingredient that we must specify is how to choose the splitting point z_C. Several options are possible: for instance, we can pick the midpoint $\frac{z_\mathrm{L}+z_\mathrm{R}}{2}$, or the sample point closest to the midpoint. Alternatively, if A spans several orders of magnitude (as it may happen, e.g., when making Bode diagrams in frequency-response analysis), then one might want to replace the arithmetic midpoint $\frac{z_\mathrm{L}+z_\mathrm{R}}{2}$ with the geometric one $\sqrt{z_\mathrm{L} z_\mathrm{R}}$. Moreover, note that, if the frequency range is split at a sample point, then the corresponding sample is shared between the two sub-surrogates, so that the continuity of the overall surrogate at the shared sample is ensured by the interpolation property.

Algorithm 4 z-adaptive MRI with automatic range partitioning

Require: frequency range $A = [z_\mathrm{L}, z_\mathrm{R}] \subset \mathbb{C}$, distinct initial sample points $Z = \{z_1, \dots, z_{S_0}\} \subset A$
Require: sampler v, tolerance ϵ, distinct test points $Z_\mathrm{test} = \{z_1, \dots, z_T\} \subset A \setminus Z$
Require: basis generator $\Psi(S) = \{\psi_0, \dots, \psi_{S-1}\}$ for all $S \geq S_0$
  take initial snapshots $v(z_1), \dots, v(z_{S_0})$
  initialize set of unexplored sub-intervals as $P = \{A\}$
  while $P \neq \emptyset$ **do**
    select an arbitrary element $A' = [z'_\mathrm{L}, z'_\mathrm{R}]$ of P and remove it from P
    try to build a z-adaptive MRI via Algorithm 3 over A'
      (using ϵ, Ψ, initial sample points $Z \cap A'$, and test points $Z_\mathrm{test} \cap A'$)
    add the newly explored sample points to Z
    if a numerical instability is detected **then**
      find $z'_\mathrm{C} \in A'$ by using one of the strategies detailed above, e.g., $z'_\mathrm{C} = \sqrt{z'_\mathrm{L} z'_\mathrm{R}}$
      add $[z'_\mathrm{L}, z'_\mathrm{C}]$ and $[z'_\mathrm{C}, z'_\mathrm{R}]$ to P
      discard the local MRI surrogate over A'
    else
      store the "good" local MRI surrogate and the interval A' for further use
    end if
  end while
  return overall surrogate obtained by patching together all the "good" local surrogates

We summarize the procedure in Algorithm 4. Note that, when building the local surrogates by z-adaptive MRI, see Section 5.3, we use only the portion of sample and test points that belong to the current sub-interval. If the initial test set Z_test has few elements, or if the sub-interval Z' is particularly small, then the local test set might be rather small too. As such, in order to prevent

[1] This is actually the most common case in practical applications. Generalizations to 2-dimensional frequency ranges are possible, e.g., by using sparse grids over a bounding box of A, cf. [HDO11].

a preemptive termination of the local z-adaptive algorithm, it might make sense to enrich the local test set with some arbitrary extra points. Moreover, we note that, when choosing a basis generator for a sub-interval, Ψ might not always be the best choice, e.g., in terms of numerical stability. To this aim, one may want to shift and rescale ("normalize") the polynomial basis to conform to the local sub-interval.

The piecewise-rational surrogate resulting from Algorithm 4 attains the prescribed tolerance over the whole A and, on each sub-interval, the usual MRI properties hold. However, the overall surrogate is not necessarily continuous across sub-intervals, except when a shared sample point is present. Moreover, the "poles" of a piecewise-rational surrogate are not necessarily well-defined. For instance, taking the union of all the surrogate poles might result in multiple distinct approximations of the same exact pole. On the other hand, one might try taking

$$\widetilde{\Lambda} = \bigcup_{j=1}^{T} \left\{ \lambda_i^{(j)} : \lambda_i^{(j)} \text{ is a pole of } H_j \text{ and } j'(\lambda_i^{(j)}) = j \right\}$$

as surrogate poles (i.e., we accept a pole iff its closest sub-interval is the one of the surrogate to which it belongs), but this could result in missing some not-so-well-approximated poles near the intersections between sub-intervals.

To conclude, we note that the above-mentioned instabilities can appear, more generally, even without adaptive frequency sampling, i.e., in the case of sample points fixed *a priori*. Our proposed partitioning approach may be successfully employed to obtain a piecewise-rational approximation even in such situations. For simplicity, we skip the details here.

5.5 Numerical tests

In this section, we carry out some numerical experiments to validate our theoretical results for fast LS Padé approximation and MRI. To this aim, we consider a synthetic problem in $\mathbb{C}^n$, $n = 100$, which we endow with the standard Euclidean inner product:

$$\begin{cases} zv(z) = Av(z) + b, \\ y(z) = c^H v(z), \end{cases} \qquad \text{with } A = U \operatorname{diag}(\lambda_1, \dots, \lambda_n) U^{-1}. \tag{5.26}$$

In (5.26), $\Lambda = \{\lambda_j\}_{j=1}^n$ are randomly generated from a uniform distribution over the square $[-5, 5] \oplus [-5, 5]\mathrm{i} \subset \mathbb{C}$, see Figure 5.1 ($\star$), and $b, c \in \mathbb{C}^n$ are two random (complex) Gaussian vectors. By similarity, the columns of the matrix $U \in \mathbb{C}^{n \times n}$ are (not necessarily normalized) eigenvectors of A, with eigenvalues Λ. Depending on how we select U, v might or might not satisfy some of the assumptions from Chapter 3. As long as U is invertible, Assumptions 3.1 and 3.5 are satisfied with $R = \infty$ (due to the finite-dimensional nature of the problem), and the corresponding residues form a linearly independent family. However, Assumptions 3.2, 3.3, and 3.6 are satisfied iff the columns of U are orthogonal, that is, iff A is normal: $A^H A = A A^H$.

Our main objective is showcasing the effectiveness of MRI. For this purpose, we study the pole approximation error

$$\min_{\widetilde{\lambda} \in \widetilde{\Lambda}} \left| \widetilde{\lambda} - \lambda_j \right| \tag{5.27}$$

($\widetilde{\Lambda}$ being the surrogate poles) and the relative output approximation error

$$\frac{|\widetilde{y}(z) - y(z)|}{|y(z)|} \tag{5.28}$$

($\widetilde{y}$ being the ROM), both for fixed and variable denominator degree. Moreover, we compare MRI with some state-of-the-art MOR competitors, see Section 2.4. More precisely, we apply the following approaches:

(a) Fast LS Padé approximation centered at $z_0 = 0$, with $N = 10$ (fixed) and $M = E \geq 10$ (the number of snapshots is $S = E + 1$).

(b) Diagonal fast LS Padé approximation centered at $z_0 = 0$, with $M = N = E \geq 10$.

(c) MRI with samples at the roots of unity (which are Fekete for the unit disk), with $N = 10$ (fixed) and $M + 1 = S > 10$.

(d) Diagonal MRI with samples at the roots of unity, with $M + 1 = N + 1 = S > 10$.

(e) Implicit moment matching centered at $z_0 = 0$, with $S = R \geq 10$.

(f) RB with samples at the roots of unity, with $R = S \geq 10$.

(g) The Loewner framework with samples (of y) at the roots of unity, with $2N + 1 = S > 20$.

Note that, except for the Loewner framework, all methods build first a ROM $\widetilde{v}$ for v using snapshots of v, and only afterwards obtain the ROM for y as $\widetilde{y} = c^H \widetilde{v}$.

For the sake of reproducibility, the code used to obtain all the results below is made publicly available as part of [Pra21].

5.5.1 MRI for a normal problem

We start from the best case (at least for MRI): we set U as a randomly generated matrix with orthonormal columns, obtained by QR factorization of a random Gaussian matrix. In Figure 5.1, we show the relative approximation error over the square $[-3, 3] \oplus [-3, 3]\mathrm{i}$, in a logarithmic color scale, for all methods. Note that we compare the methods for a fixed number of snapshots $S = 21$.

The "single-point" methods (a), (b) and (e) are quite recognizable since their error reaches zero at $z = 0$. The approaches based on samples at the roots of unity show an error that is almost uniform on the unit disk, in accordance with approximation theory, see, e.g., Theorem 2.3. A notable exception to this is the Loewner framework surrogate, which coincides with the Lagrange rational interpolant of type [10/10], and whose results appear a bit worse than the rest. This is due to the rational type (specifically, the numerator degree) being too small to identify well the features of y. Indeed, the main issue with standard non-intrusive methods is their non-optimal use of the snapshots, see Section 2.4.1. At the same time, the non-intrusive MRI approach, by relying on samples of v rather than of y, seems to be very comparable to intrusive projective methods, at least qualitatively. Note that methods (a) to (f) perform similarly outside the unit disk. Our theory from Chapter 3 provides a justification (at least for the MRI flavors with fixed N): since the Green's potential for the unit disk is the complex magnitude (outside the disk),

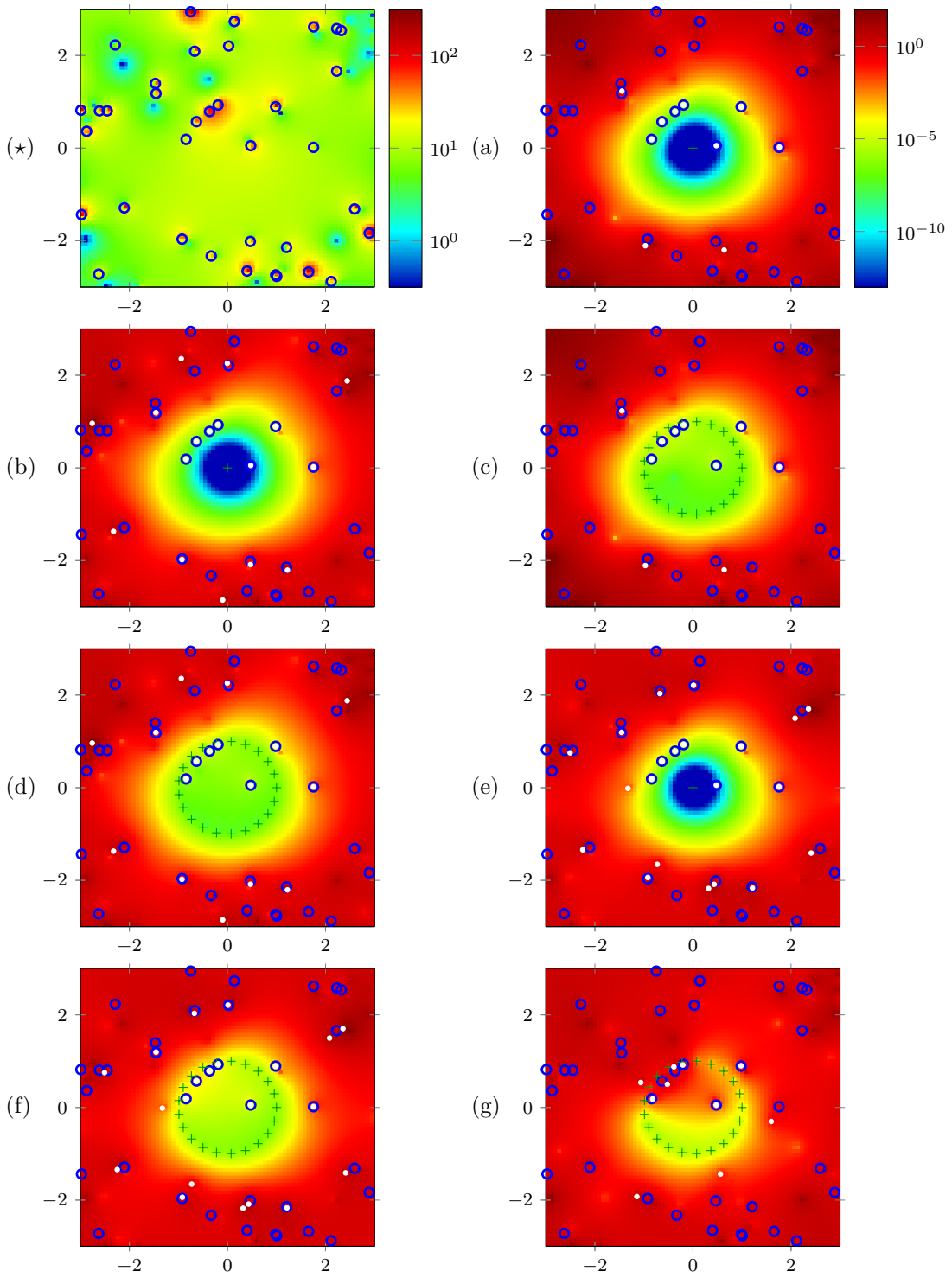


Figure 5.1 – Qualitative results in the normal setting. Figure (⋆): plot of exact output magnitude $|y(z)|$. Figures (a)–(g): plot of the relative error (5.28). Sample points, exact poles, and surrogate poles are denoted by green pluses, blue circles, and white disks, respectively. All the errors have the same color scale, reported next to Figure (a). All plots have $\mathrm{Re}(z)$ on the x-axis and $\mathrm{Im}(z)$ on the y-axis.

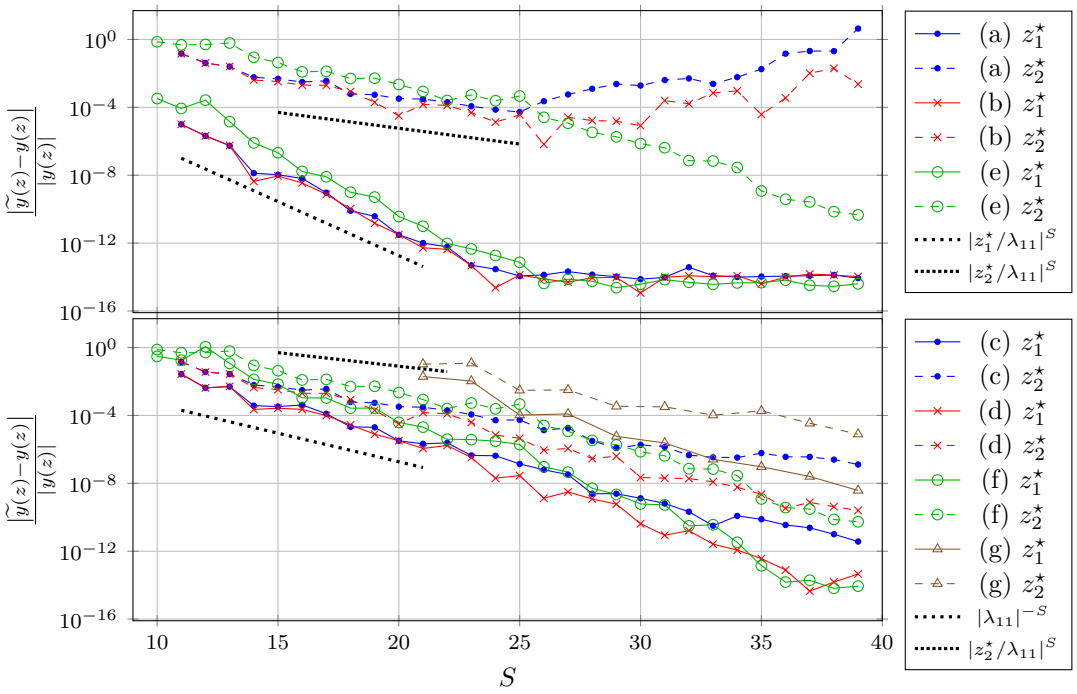


Figure 5.2 – Error convergence at $z_1^\star = \frac{\mathrm{i}}{2}$ and $z_2^\star = 1+\mathrm{i}$ in the normal setting. Theoretical convergence rates (for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(z_1^\star) = 1$).

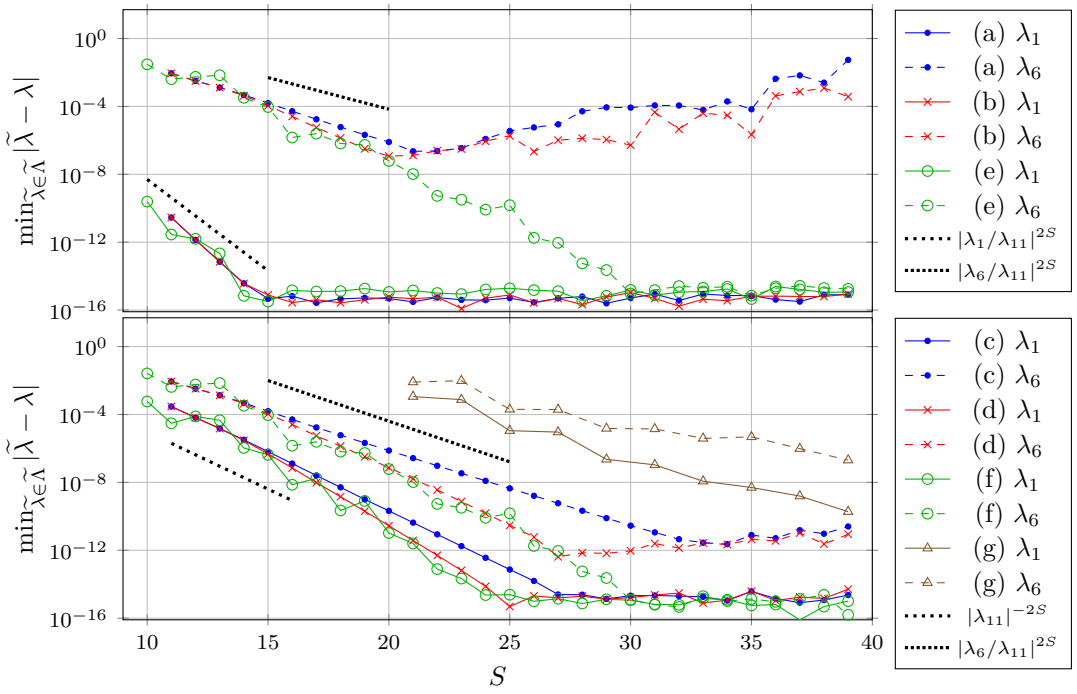


Figure 5.3 – Convergence of pole approximation error at λ_1 and λ_6 in the normal setting. Theoretical convergence rates (for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(\lambda_1) = 1$).

the convergence rate for the MRI error, see Theorem 3.7, has the same form as the convergence rate for fast LS Padé approximation, see Theorem 3.3.

We make this analysis quantitative by studying the convergence of the approximation error at the (arbitrarily selected) points $z_1^\star = \frac{\mathrm{i}}{2}$ and $z_2^\star = 1 + \mathrm{i}$, as the number of snapshots increases. We note that $z_1^\star$ is in the unit disk while $z_2^\star$ is not. We show the results in Figure 5.2, where, to avoid cluttering, we separate the single-point methods from the interpolatory ones. We can observe that the asymptotic convergence rate predicted by Theorem 3.7 is achieved by all methods. Notably, the methods with variable denominator degree N seem to converge as fast as those with fixed N, if not a bit faster, even though their theory does not allow any foretelling of their convergence rate, see Theorems 3.4 and 3.8.

Concerning single-point approximations, we note that the convergence appears to stagnate after $S = 25$. In fact, the fast LS Padé approximation error at $z_2^\star$ even increases for larger values of S. This is due to numerical issues in the computation of the derivatives of v, which become more and more collinear as the order increases, cf. (4.7). Conversely, the implicit moment matching method does not seem to suffer from the same issues. This is due to the Arnoldi algorithm, which allows to perform projections in a (more) stable fashion even when the Krylov subspace order becomes large.

On the interpolatory side, we confirm that the Loewner framework performs significantly worse than the rest. This is due to its rational type being smaller than that achievable by the other methods, for fixed S. A convergence analysis with respect to the numerator degree, which, for (g), equals $(S-1)/2$, would show that all methods perform similarly.

In Figure 5.3 we study the convergence of the pole approximation error, for the (arbitrarily chosen) poles $\lambda_1 \approx 0.475 + 0.053\mathrm{i}$ and $\lambda_6 \approx 0.983 + 0.893\mathrm{i}$ (we order the poles according to their distance from $z = 0$). The observed behavior is similar to that of the approximation error. Notably, the observed convergence rates are in agreement with Theorems 3.1 and 3.5.

In conclusion, implicit moment matching is the best-performing single-point method, whereas, among distributed approaches, diagonal MRI and RB perform similarly well. Still, as opposed to the other two methods, diagonal MRI manages to achieve positive results without being intrusive.

5.5.2 MRI for a non-normal problem

Now we move to a more difficult case: we set U as

$$U = \begin{bmatrix} 1 & \delta & & & \\ \delta & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \in \mathbb{C}^{n \times n}, \tag{5.29}$$

so that the first two eigenvectors are non-orthogonal, and thus A is non-normal, whenever $\delta \neq 0$. Notably, the closer δ is to ± 1, the more difficult it will be for MRI to distinguish the first two poles and residues. We repeat the same numerical experiment as before with this U, for $\delta = 0.9$, and we report the results in Figures 5.4 and 5.5. As expected, cf. Conjecture 1, the error behaves similarly to the normal case, for all methods. The pole convergence displays the same behavior as before too. This is quite remarkable, since Conjecture 1 predicts a decrease of the convergence

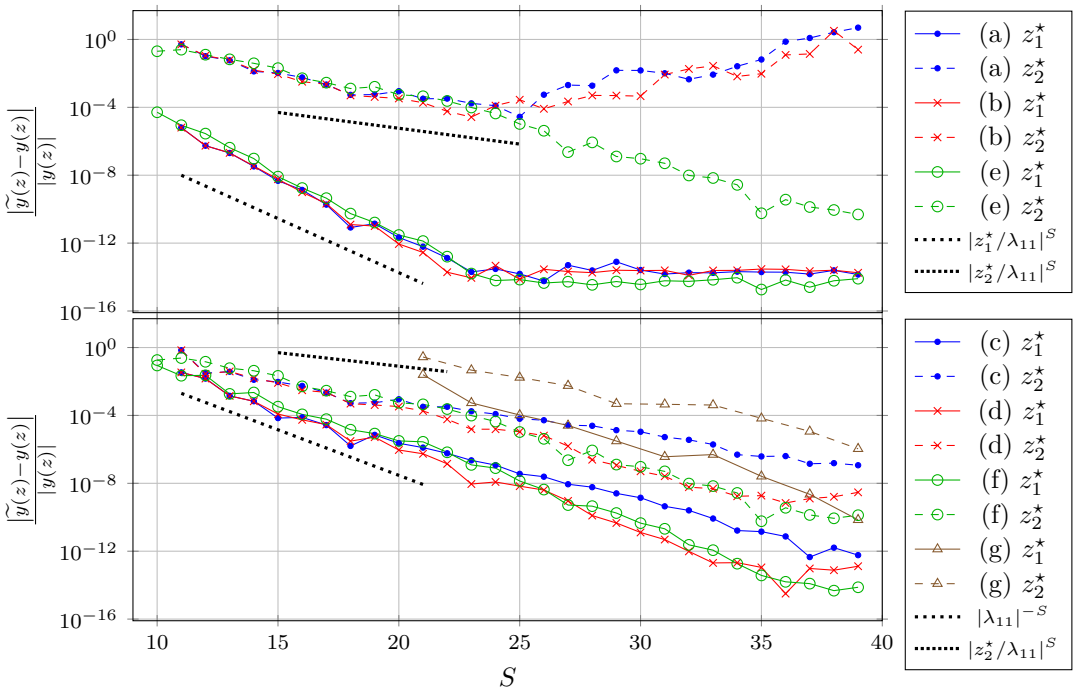


Figure 5.4 – Error convergence at $z_1^\star = \frac{\mathrm{i}}{2}$ and $z_2^\star = 1 + \mathrm{i}$ for $\delta = 0.9$. Theoretical convergence rates (for $\delta = 0$, for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(z_1^\star) = 1$).

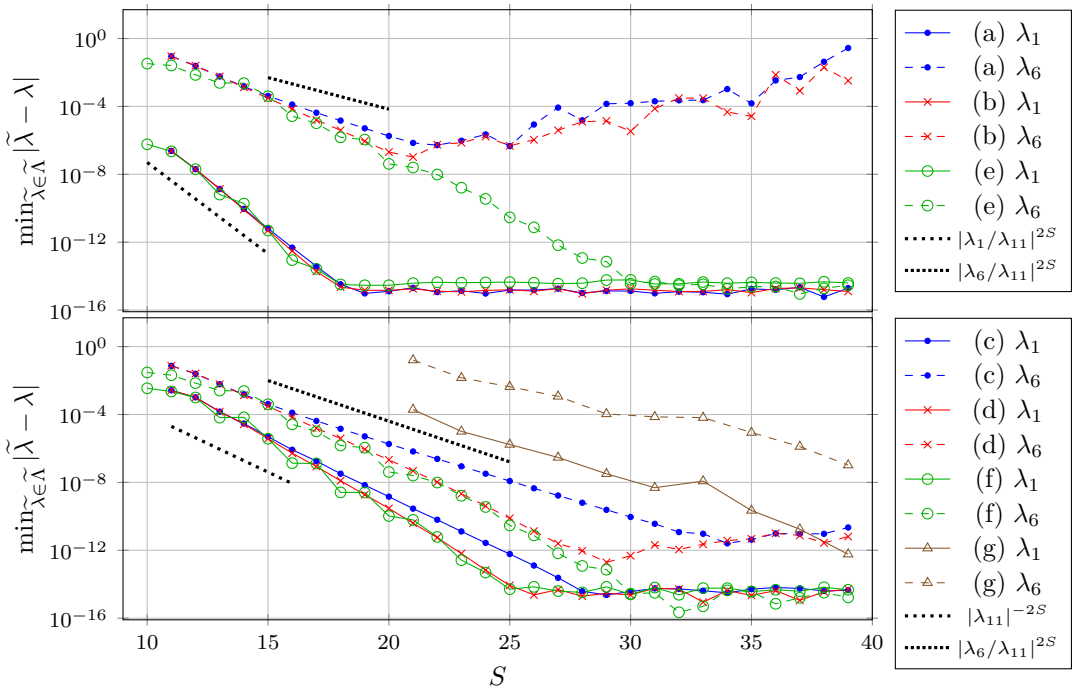


Figure 5.5 – Convergence of pole approximation error at λ_1 and λ_6 for $\delta = 0.9$. Theoretical convergence rates (for $\delta = 0$, for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(\lambda_1) = 1$).

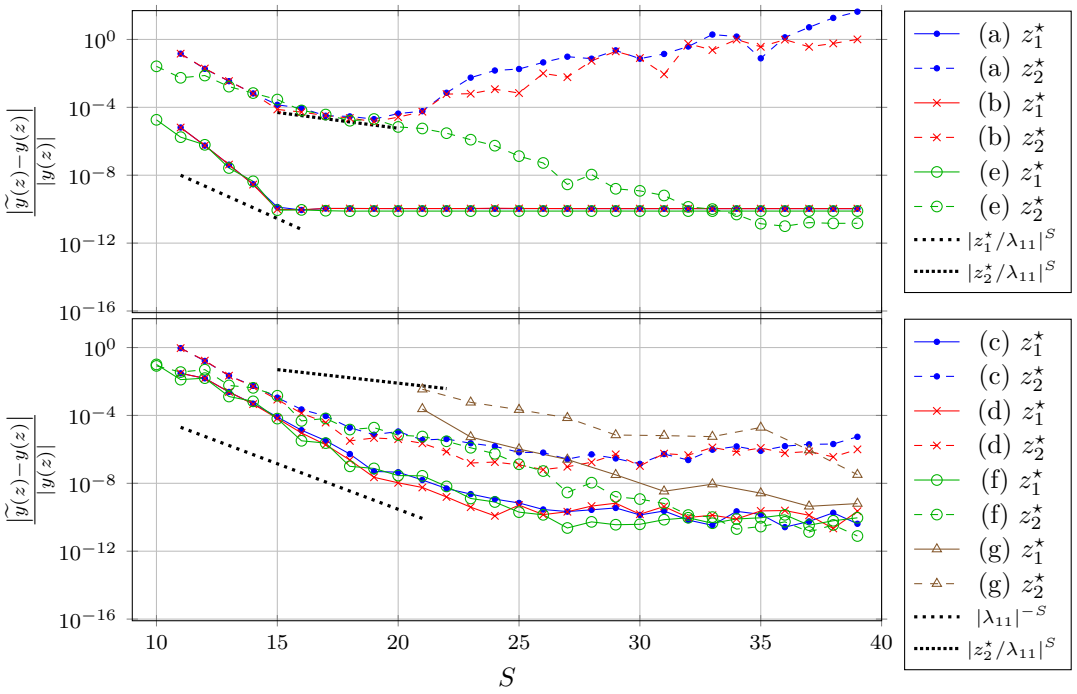


Figure 5.6 – Error convergence at $z_1^\star = \frac{\mathrm{i}}{2}$ and $z_2^\star = 1 + \mathrm{i}$ for $\delta = 0.999$. Theoretical convergence rates (for $\delta = 0$, for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(z_1^\star) = 1$).

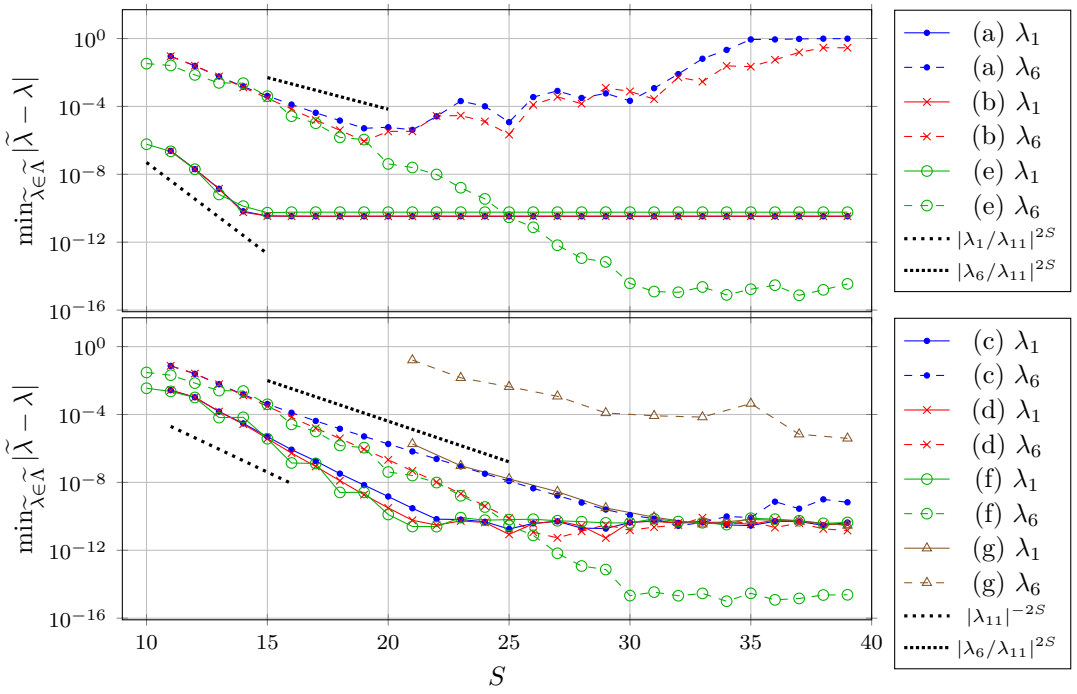


Figure 5.7 – Convergence of pole approximation error at λ_1 and λ_6 for $\delta = 0.999$. Theoretical convergence rates (for $\delta = 0$, for MRI with fixed denominator degree) are also included in black (note that $\Phi_{B^0(1)}(\lambda_1) = 1$).

rate (at least for λ_1), which we do not observe here.

We further increase the difficulty of the problem by setting $\delta = 0.999$. In this case, one should note that the conditioning of the matrix $zI - A$ might start to have a relevant effect, by introducing numerical noise in the snapshots. We show the convergence results in Figures 5.6 and 5.7.

In single-point approximation, the error seems to be greatly affected by the almost-collinear eigenvectors. Particularly, the error saturates at a smaller S, at a much larger value. This is probably due to a combination of noise in the snapshots due to ill-conditioning of the problem and "collinearity noise", cf. Section 3.1.3. It is rather interesting to note that implicit moment matching seems to overcome this "noise" when approximating λ_6, whose eigenspace is orthogonal to the others. More generally, implicit moment matching seems able to approximate well poles whose residues behave well (i.e., they are well-separated from the rest), whereas fast LS Padé approximation does not.

Interpolatory approaches yield fairly similar results. Overall, we note that the lack of normality seems to be an issue also for projective approaches, slowing down their convergence. Still, implicit moment matching and RB appear to perform robustly enough. Moreover, we can observe that, at least in the regions of exponential convergence, the convergence rate (notably, for poles) does not seem affected by the lack of normality, contrary to what Conjecture 1 could have lead to believe. Indeed, the bounds in Theorems 3.1 and 3.7 seem to be always valid.

5.5.3 MRI with greedy sampling

As a last test involving (5.26), we apply MRI with adaptive frequency sampling as described in Section 5.3. We consider both the normal case and the non-normal cases $\delta \in \{0.9, 0.999\}$ introduced in the previous sections. We set the algorithm parameters as follows:

- The initial sample points are $Z = \{-1, 1\}$.
- The tolerance ϵ is 10^{-3}.
- The test points are approximately 10^4 points on a Cartesian grid within the unit disk $B^0(1)$, whose spacing is approximately 0.018.
- We use the barycentric basis, see Section 5.2.3.
- We use the hybrid error estimator based on the evaluation of the *relative* approximation error (on the state v) in a "look-ahead" fashion. More precisely, we use the approach presented in Section 5.3.3, but we replace the convergence check $e(z^\star) < \epsilon$ with its "relative" version $e(z^\star) < \epsilon \left\| v(z^\star) \right\|_{\mathcal{V}}$.

In Figure 5.8, we show some snapshots of the y-error field (with respect to z) as the greedy iterations proceed, in the normal case. We can observe that the approximation error is quite large at the beginning, but it decreases as more and more samples are added. In particular, the aim of the greedy algorithm is to achieve a (relative) error that is uniformly small over the sampling set $B^0(1)$ (technically, only over the discrete test set). We see that this is achieved at $S = 17$, with the error tolerance being uniformly attained over $B^0(1)$, at least in the "eyeball-norm".

It is rather interesting to observe that some of the sample points seem to coincide with the poles of v. Indeed, from Lemma 5.1, we know that additional samples should be added at the

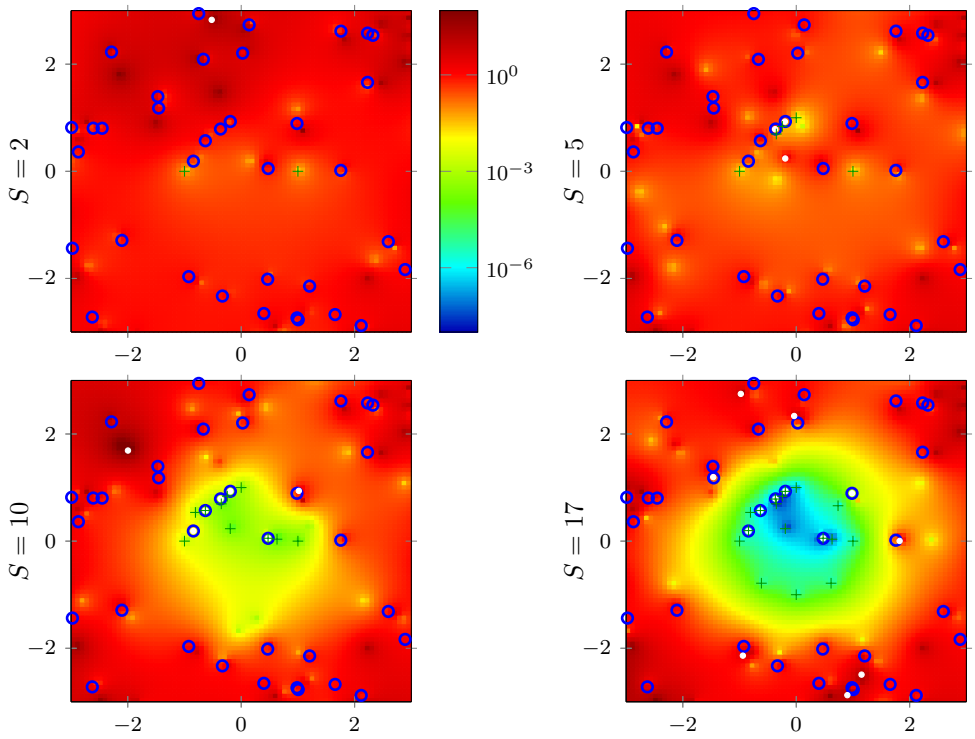


Figure 5.8 – Qualitative results of the greedy algorithm in the normal setting. Plot of the relative error on y (5.28) at different iterations. The sample points are denoted by green pluses. All the errors have the same color scale. All plots have $\mathrm{Re}(z)$ on the x-axis and $\mathrm{Im}(z)$ on the y-axis.

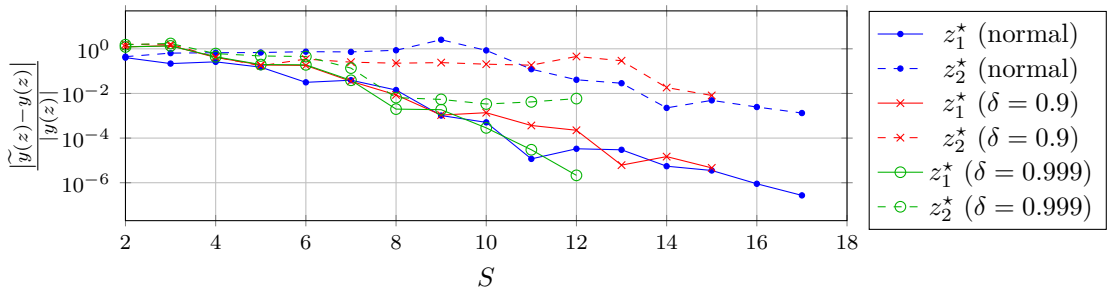


Figure 5.9 – Error convergence at $z_1^\star = \frac{\mathrm{i}}{2}$ and $z_2^\star = 1 + \mathrm{i}$.

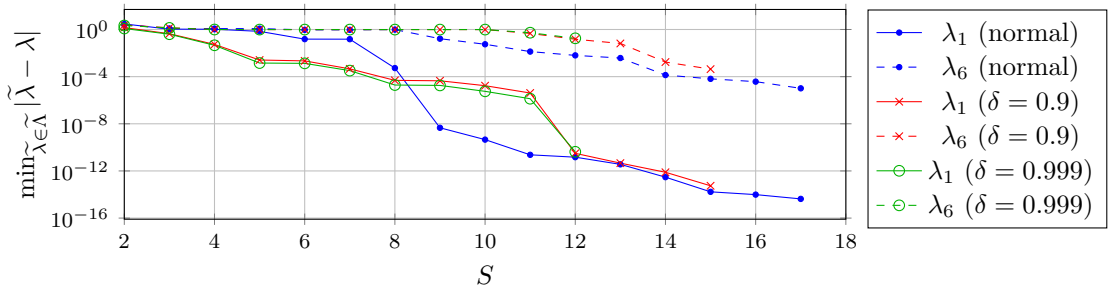


Figure 5.10 – Convergence of pole approximation error at λ_1 and λ_6.

approximate poles, while, from Theorem 3.6, we know that the approximate poles converge to the exact ones. On one hand, this makes sense: sampling very close to an exact pole gives extremely accurate information on such pole, and on the corresponding residue as well. On the other hand, this can easily lead to issues in computing the snapshot, since the FOM will generally be rather ill-conditioned near the resonating frequencies. If an iterative linear system solver is applied to compute the snapshot, a large number of iterations might become necessary. More generally, the snapshots might be affected by significant numerical noise, regardless of the solver.

Given our non-intrusive paradigm, the simple solution is to ignore these issues by blaming the FOM solver. However, this is not really satisfactory, as it puts rather important limitations to the applicability and robustness of the z-adaptive MRI approach. In our numerical experiments, we do not observe any numerical issues related to noisy samples. A more theoretically motivated discussion on the effect of noisy samples on MRI can be found in Section 8.1.

Other than near the poles of v, the adaptive sampling seems to take samples mostly near the boundary of the sampling region, namely, close to some of the roots of unity, which are Fekete points for the unit disk, cf. Theorem 2.2. This behavior (of sampling first around the poles and then at "approximately Fekete" points) is somehow expected when using the error indicator from Lemma 5.1. Indeed, the greedy algorithm first tries to "cancel out" each surrogate pole with a root of the nodal polynomial and then takes on the task of minimizing (uniformly) the nodal polynomial itself, which can be done exactly by putting samples at the Fekete points. We note that all these observations are in agreement with the usual numerical considerations on the "magic points" of the empirical interpolation method [Bar+04], which are chosen in a similar greedy fashion and are also observed to form an "approximately Fekete" sequence.

We display in Figures 5.9 and 5.10 the relative approximation error and the pole approximation error as the greedy iterations proceed, respectively. We can see that the normal and non-normal cases are quite similar. In fact, somewhat surprisingly, the greedy algorithm terminates in less iterations in the non-normal case. In all settings, we can observe that the errors form a plateau for small values of S and then start to decrease in a way that is not as neat as in the previous sections. This is to be expected, since sample points may be added both close and far from the location where the error is evaluated, thus preventing any hope for a monotonic error decay. A similar argument can also be used to explain, e.g., the sharp drop in the pole error at $S = 12$ in the non-normal cases: quite simply, the twelfth sample point was added very close to λ_1.

To conclude, we note that, by comparing Figures 5.9 and 5.10 and Figures 5.2 to 5.7, we see that the convergence of the relative y-approximation error (at least at $z_1^\star$ and $z_2^\star$) and of the pole approximation error (at least for λ_1 and λ_6) is faster with greedy sampling than with Fekete sampling. This provides empirical evidence of the good approximation properties of the greedily chosen sample points.

5.5.4 MRI for a scattering problem

As a further (more complicated and slightly less academic) non-normal example, we move to an example from the field of PDEs. More precisely, we consider a frequency-domain scattering problem of the form (2.34), which we recall here for convenience:

$$\begin{cases} \left(-\Delta + \frac{z^2}{c^2}\right) v(z) = f(\cdot, z) & \text{in } \Omega, \\ \left(\frac{\partial}{\partial \nu} + \frac{z}{c}\right) v(z) = 0 & \text{on } \partial\Omega. \end{cases}$$

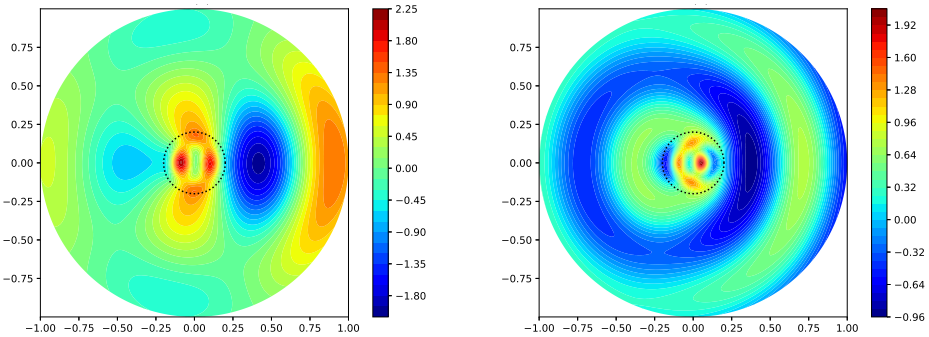


Figure 5.11 – Solution (real part) of the scattering problem for $z = 7\mathrm{i}$ (left) and $z = 3\pi\mathrm{i}$ (right). We also superimpose the boundary of the scatterer in black.

We fix the 2D unit disk $\Omega = B^{\mathbf{0}}(1)$ as spatial domain, with normal $\nu|_x = x$, and we set the wave speed and the forcing term as

$$c(x) = \begin{cases} 5 & \text{if } |x| < \frac{1}{5} \\ 1 & \text{otherwise} \end{cases} \quad \text{and} \quad f(x,z) = z^2 \left(\frac{1}{c(x)^2} - 1 \right) \exp(z x^{(1)}),$$

respectively. This corresponds to the field scattered by a sound-permeable circular scatterer (with radius $\frac{1}{5}$ and contrast 5) when the horizontal plane wave $u_{\mathrm{inc}}(x,z) = \exp(z x^{(1)})$ impinges on it from the right. We discretize the PDE using piecewise-linear finite elements, over a fine uniform triangulation of Ω. We note that, for this step, we employ the FEniCS library [Aln+15], which is fully compatible with the `RROMPy` package. The finite element space has approximately $2 \cdot 10^4$ degrees of freedom. Two sample solutions at $z = 7\mathrm{i}$ and $z = 3\pi\mathrm{i}$ are shown in Figure 5.11.

Our approximation target is the state $v(z) \in H^1(\Omega)$, for $z \in A = [5\mathrm{i}, 10\mathrm{i}]$. We employ MRI with S samples at the Chebyshev points of A, i.e.,

$$Z = \left\{ \frac{7}{2}\mathrm{i} + \frac{5}{2}\mathrm{i} \cos\left(\frac{2j-1}{2S}\pi \right) \right\}_{j=1}^{S}.$$

The (optimally conditioned) Chebyshev polynomial basis is used to expand numerator and denominator. From the well-posedness of the problem for imaginary frequencies, cf. Section 2.3.3, we expect the system to have no poles within A. Also, we note that the forcing term f depends on z. On one hand, this introduces an additional complexity in the state v, so that, with respect to a simpler (e.g., constant) forcing term, we are likely to need more snapshots to achieve a given accuracy. On the other hand, we see that the dependence of f on z is holomorphic, and looking for a rational approximation is still justified.

In Figure 5.12, we show the results obtained for $S = 21$ and $S = 31$. In particular, we use the energy seminorm $\|v\|_{H^1_0(\Omega)} = \|\|\mathrm{grad}\, v\|\|_{L^2(\Omega)}$ to measure the magnitude of v. From the plot, we see that some poles are extremely close to A. This is due to internal reflections inside the scatterer, cf. Figure 5.11 (left). Overall, we see that the approximation quality is fairly good, but the relative error is not that small locally around some of the poles.

We compare these results with those obtained with a z-adaptive MRI approach. The greedy parameters are the same as in Section 5.5.3, except for:

- The initial sample points are $Z = \{5\mathrm{i}, 10\mathrm{i}\}$.
- The test points are 10^4 uniformly spaced points over A.
- We use the Legendre polynomial basis to expand numerator and denominator.

Concerning the error estimator, we employ once more the relative "look-ahead" one from Section 5.3.3. Note that, here, its theoretical foundations do not hold because neither the problem nor the forcing term are affine in z. Still, we choose to employ it in a heuristic way.

The algorithm terminates at the 29-th snapshot, and yields the surrogate shown in Figure 5.13. We can observe that the relative error behaves much more uniformly than with Chebyshev sampling. This is reasonable, since, after all, the z-greedy algorithm tries to enforce the tolerance constraint uniformly over the test set. In particular, this has the consequence of keeping the error under control even near the poles. This being said, we see that the tolerance is not attained over the whole domain, since there are some regions where the relative error is between 10^{-3} and 10^{-2}. This is due to the above-mentioned heuristic nature of the error estimator in the greedy algorithm. Still, we can be quite satisfied with the results, considering that the tolerance is exceeded only for few frequencies, and only by a small margin.

As a final measure of the quality of the surrogate, we evaluate the relative approximation error on a set of 25 quasi-randomly generated points in $A \setminus Z_{\text{test}}$ (we use the Halton scheme, see [Hal64]). In Figure 5.14, we plot the maximum of these 25 evaluations. In particular, we show how such quantity evolves as the z-greedy iterations proceed. For small values of S, we see that the error stagnates around 1, showing that the amount of samples is not yet sufficient for a good approximation of all the "relevant" poles. Then, the error starts to decrease until the prescribed tolerance is attained. Note that the error becomes smaller than the threshold ϵ exactly at the last iteration of the adaptive algorithm. This is an empirical verification that our non-intrusive greedy estimator, albeit heuristic, works well in this not-so-straightforward example.

In the same plot, we also show the same error quantity for the case of Chebyshev samples. Note, in particular, that the greedy samples form a nested sequence as S increases, but the Chebyshev ones do not. We can observe that the error measure stagnates (or maybe converges extremely slowly) over the whole range of values of S. This provides a numerical proof of the fact that greedily selected sample points perform better than ones fixed *a priori*.

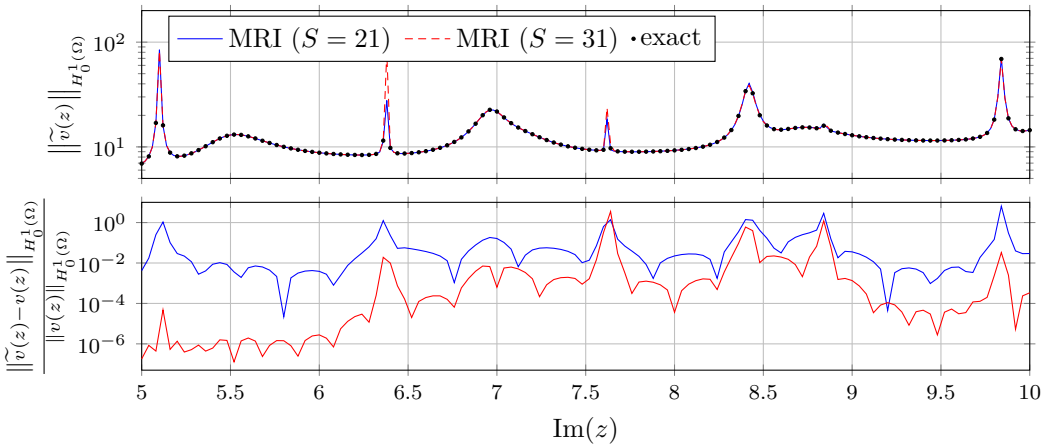


Figure 5.12 – Results for the scattering problem. Surrogate (using Chebyshev sample points) and exact norm of v in the top plot. In the bottom plot, the corresponding approximation errors.

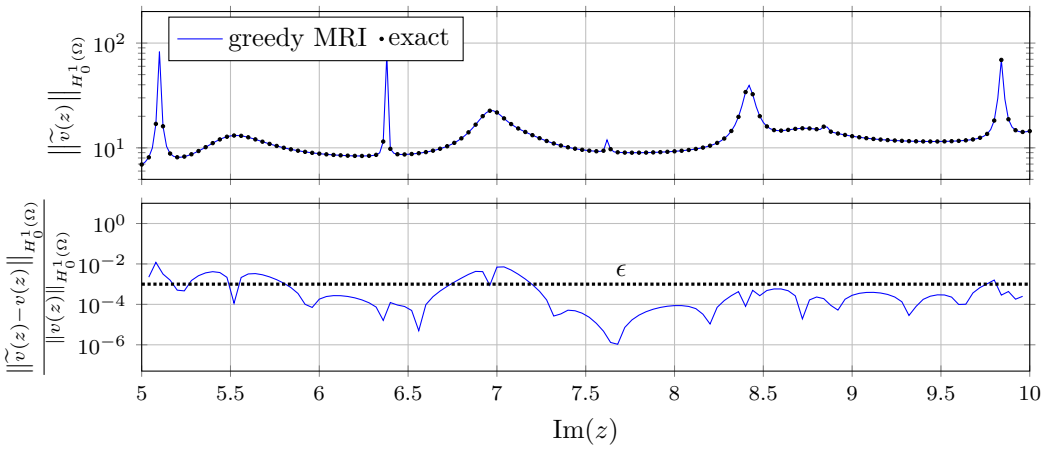


Figure 5.13 – Results for the scattering problem. Surrogate (via greedy MRI) and exact norm of v in the top plot. In the bottom plot, the corresponding approximation error.

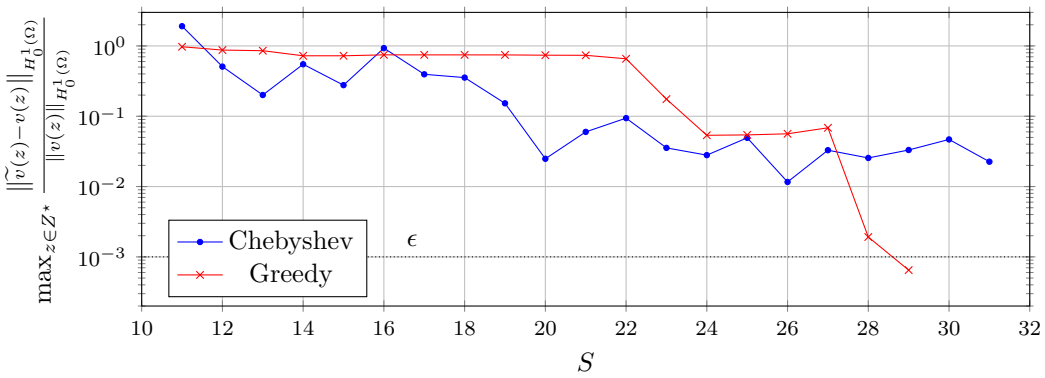


Figure 5.14 – Relative error convergence for the scattering problem. We plot the maximum error over a set of 25 random frequencies in A.

6 MOR approaches for parametric frequency-response problems

In the previous chapters, we have focused on building a surrogate for a problem depending on a single parameter, namely, the frequency. More generally, the MRI method introduced in Chapter 3 can be applied to produce a reduced model for any univariate function, albeit we have shown that the quality of the approximation depends on some characteristics of the target function, i.e., its approximability by a rational function and the non-collinearity of information coming from different poles. A naturally arising question is, then: can this concept be extended to the multivariate case? After all, quoting [ABG20], "anything worth simulating will have [additional] parameters", which might represent, e.g., control variables, design parameters, or uncertainties, depending on the application.

In this chapter, we try to provide a partial answer to this question, which, as it turns out, is rather complex. We start from an overview of the state-of-the-art MOR methods for parametric frequency-response problems. Then we focus on a specific family of methods, namely, marginalized pMOR approaches based on interpolation of surrogate poles and residues, and describe our contributions to the field.

6.1 Parametric dynamical systems in frequency domain

Let there be n_θ parameters, which we denote by $\theta = \left(\theta^{(i)}\right)_{i=1}^{n_\theta}$. We will assume that θ belongs to some compact parameter range $\Theta \subset \mathbb{C}^{n_\theta}$. In dynamical systems, it is quite common for parameter dependence to appear in the matrices defining the system. As such, we will take the following frequency-domain LTI system as model problem:

$$\begin{cases} zE(\theta)v(z,\theta) = A(\theta)v(z,\theta) + B(\theta)u(z,\theta), \\ y(z,\theta) = C(\theta)v(z,\theta). \end{cases} \tag{6.1}$$

Infinite-dimensional extensions of parametric problems in the field of PDEs are, of course, also possible, see, e.g., our numerical examples in Sections 7.1 and 7.2. Throughout this section, we will assume that all θ-dependent quantities are continuous with respect to θ. In trying to understand the properties of (6.1), we wish to recycle all our considerations for non-parametric dynamical systems. To this aim, it suffices to fix θ to obtain a non-parametric system, whose

transfer function H, notably, depends on z in a meromorphic way:

$$H(z,\theta) = \frac{C(\theta)\operatorname{adj}(zE(\theta) - A(\theta))B(\theta)}{\det(zE(\theta) - A(\theta))} = \sum_{j=1}^{n_v} \frac{r_j(\theta)}{z - \lambda_j(\theta)}. \tag{6.2}$$

Note that, above, the first identity, giving the rational form of H, is always valid, whereas the second one requires $E(\theta)$ to be invertible and $E(\theta)^{-1}A(\theta)$ to be diagonalizable, cf. Section 2.2.

Since, by the Caley-Hamilton theorem, both adjugation and determinant are smooth functions (in fact, polynomials) of the matrix entries, we can deduce that regularity with respect to θ is passed on from system matrices to the numerator and denominator of the rational form. However, in this context, we note that any θ-dependence might be raised to the n_v-th power, e.g., $\operatorname{adj}(z\theta I) = (z\theta)^{n_v-1}I$ and $\det(z\theta I) = (z\theta)^{n_v}$.

From (6.2) we can also draw conclusions on poles and residues of the system. Indeed, using standard results from perturbation theory, see, e.g., [Kat95], one can conclude that poles and residues are continuous (but possibly multivalued) functions of θ over all compact sets where the partial fraction expansion (6.2) is valid. In fact, the poles of the system always depend continuously on θ. Moreover, generally speaking, smoothness is inherited by poles and residues, as long as the poles stay well-separated. Unfortunately, smoothness (except the continuity of the poles) might be completely lost when poles cross. We proceed to show this through a simple example. Let $n_\theta = 1$,

$$A(\theta) = \begin{bmatrix} 1 & \theta \\ \theta & -1 \end{bmatrix}, \; E(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; B(\theta) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } C(\theta) = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Simple calculations show that the rational form of H is smooth: notably, H is a multivariate rational function

$$H(z,\theta) = C(zE - A(\theta))^{-1}B = \frac{z+1}{z^2 - 1 - \theta^2}.$$

(More generally, H is a multivariate rational function whenever E, A, B, and C are polynomial or rational functions of θ.)

On the other hand, we can verify that the partial fraction form is not as smooth

$$H(z,\theta) = \frac{\frac{1}{2}\left(1 + \frac{1}{\sqrt{1+\theta^2}}\right)}{z - \sqrt{1+\theta^2}} + \frac{\frac{1}{2}\left(1 - \frac{1}{\sqrt{1+\theta^2}}\right)}{z + \sqrt{1+\theta^2}},$$

with $\sqrt{\cdot}$ being some continuous instance of the complex square root (notably, the same instance throughout the expression above). Since the complex square root induces branch points at $\theta = \pm\mathrm{i}$ in poles and residues, they are intrinsically multi-valued. Moreover, we see that A fails to be diagonalizable at the branch points, and the simple-pole partial fraction expansion does not exist there, so that we must replace it by

$$H(z;\pm\mathrm{i}) = \frac{1}{z} + \frac{1}{z^2}.$$

Accordingly, the residues in the simple-pole expansion are unbounded at $\theta = \pm\mathrm{i}$.

Due to the (potential) lack of regularity of poles and residues, many pMOR approaches choose to avoid a partial fraction expansion of H, working in the rational form, or even the system one (6.1), instead. However, this may lead to some limitations, as we proceed to explain.

6.2 Overview of MOR for parametric frequency-domain problems

In the literature, one can find several approaches for building a surrogate for parametric problems like (6.1). Broadly speaking, we can split these methods into "non-intrusive" and "intrusive" just like in the non-parametric case. However, for parametric systems, an additional distinction must be made, depending on whether the frequency z is treated together with the other parameters θ or not. Note, however, that the gap between these two families is not as clear-cut as that between non-intrusive and intrusive methods. We outline the two categories, as well as their main pro and cons, separately.

6.2.1 State-of-the-art global pMOR approaches

In global pMOR, one builds a single ROM that provides a map from (z, θ) to $H(z, \theta)$, thus treating frequency and parameters "jointly". The two main instances of global approaches are, on the data-driven side, multivariate rational approximation [GTT18; IA14; LAI11; MGT20; Xia+19] and, on the projective side, RB/POD on a global basis [Bau+11; BF14; Dan+04; DEF09; Wei+99]. Both of them are rather straightforward generalizations of their non-parametric counterparts. However, some technical observations are in order:

- In multivariate rational approximation, one must choose how to balance the degrees of numerator and denominator with respect to z and (each component of) θ. In this context, it is customary to use the barycentric basis for z, in Loewner-style, cf. Section 2.4.1, whereas standard polynomial bases (monomials, Chebyshev, etc.) are employed to model the θ-dependence, usually with a constraint on the total degree with respect to θ. Overall, the resulting surrogate is of the form

$$\widetilde{H}(z,\theta) = \frac{\sum_{j=1}^{N_z+1} \sum_{\delta\in\Delta} p_{j,\delta} \left(\theta^{(1)}\right)^{\delta_1} \cdots \left(\theta^{(n_\theta)}\right)^{\delta_{n_\theta}} /(z - z_j)}{\sum_{j=1}^{N_z+1} \sum_{\delta\in\Delta} q_{j,\delta} \left(\theta^{(1)}\right)^{\delta_1} \cdots \left(\theta^{(n_\theta)}\right)^{\delta_{n_\theta}} /(z - z_j)},$$

  with the index set Δ being, e.g.,

$$\Delta = \left\{ (\delta_i)_{i=1}^{n_\theta} \in \{0, 1, \ldots\}^{n_\theta} : \sum_{i=1}^{n_\theta} \delta_i \leq N_\theta \right\}.$$

  This corresponds to an approximation of the form P/Q, where P and Q belong to some tensor space $\mathbb{P}_{N_z}(\mathbb{C};\mathbb{C}) \otimes \mathbb{P}_\Delta(\mathbb{C}^{n_\theta};\mathbb{C})$.

- In the projective case, several options are possible for determining the reduced basis that is used to project the system. A fairly popular choice is the concatenation of reduced bases obtained by non-parametric MOR (e.g., by implicit moment matching) at few parameter values $\{\theta_j\}_{j=1}^T$. This is usually followed by POD, since many of the basis vectors might be almost collinear. An alternative is to carry out implicit moment matching jointly in (z, θ), so that the basis is composed of partial derivatives of v with respect to z *and* θ at one or more (z, θ)-points.

- Both strategies allow for *a posteriori* model selection, using the ideas already discussed in Sections 2.4.1.1 and 2.4.2.2 for the non-parametric case.

The greatest advantage of global approaches is their simplicity, both of implementation (especially in the projective case) and of use (since they are almost as straightforward as their non-parametric analogues). However, they have some relevant drawbacks. The biggest one is the lack of efficiency when dealing with even a modest (~ 5) number of parameters. This issue manifests itself in two different ways:

- For non-intrusive methods, the number of coefficients in numerator and denominator suffers from the curse of dimension, increasing at an exponential rate in n_θ. By enforcing sparsity of the coefficients, see, e.g., [CL11], it might be possible to improve the situation in some cases. However, in general applications, one should not expect the denominator coefficients to be sparse. For instance, consider the simple case of affinely drifting poles:

$$v(z,\theta) = \left(zI - \begin{bmatrix} \alpha_0^{(1)} & & \\ & \ddots & \\ & & \alpha_0^{(n_v)} \end{bmatrix} + \sum_{i=1}^{n_\theta} \alpha_i \theta^{(i)} I \right)^{-1} b = \sum_{j=1}^{n_v} \frac{b_j}{z - \alpha_0^{(j)} - \sum_{i=1}^{n_\theta} \alpha_i \theta^{(i)}}. \tag{6.3}$$

  Then, approximating well N poles of the system requires as denominator a full polynomial in z *and* θ of total degree N, which has $\binom{N+n_\theta+1}{N} \gtrsim n_\theta^N/N!$ coefficients.

- For intrusive methods, a ROM with good approximation properties usually comes at the price of a (very) large reduced size R, thus hindering online efficiency. This is due to the fact that, in general, the "best" basis for projection changes with θ, cf. the partial fraction expansion (6.2). Accordingly, it becomes necessary to add many global (θ-independent) basis elements just to follow the evolution of a *single* θ-dependent basis element. We illustrate this with a small, but rather perverse, example: consider the matrix $A(\theta) = a(\theta)a(\theta)^H$, with

$$a(\theta) = \begin{bmatrix} 1 & \theta & \cdots & \theta^N & 0 & \cdots & 0 \end{bmatrix}^\top \in \mathbb{C}^{n_v}.$$

  $A(\theta)$ has rank 1 for all θ, but a global basis must necessarily be of size $N+1$ in order to recover (by projection) the range of A for all $\theta \in [0,1]$.

6.2.2 State-of-the-art marginalized pMOR approaches

In marginalized pMOR, one builds a ROM in two steps: first, surrogates *in frequency only* are built at (few) representative parameter values $\widetilde{\Theta} = \{\theta_j\}_{j=1}^T$, and *then* the different z-ROMs are combined, usually by some kind of interpolation, over Θ to obtain the overall surrogate. The difference between global and marginalized pMOR might not emerge clearly from this description. To make the matter a bit clearer, we consider the four main instances of marginalized pMOR, which differ in terms of the quantity that gets interpolated over Θ.

6.2.2.1 Interpolation of local reduced bases

Using a projective approach at each $\theta_j \in \widetilde{\Theta}$, we identify (e.g., by implicit moment matching) a collection of local reduced spaces $\{\widetilde{\mathcal{V}}_j\}_{j=1}^T$ and corresponding projection matrices $\{\widetilde{V}_j\}_{j=1}^T \subset \mathbb{C}^{n_v \times R}$. Note that, for simplicity, we are assuming that all reduced bases have the same size. Rather than concatenating the projection matrices and applying POD, cf. Section 6.2.1, we build a θ-dependent reduced space of size R, which interpolates the local spaces $\widetilde{\mathcal{V}}_j$. Then, we project the FOM onto this θ-dependent reduced space to obtain the final surrogate.

The most critical step to ensure good approximation (other than choosing the local bases in a judicious way) is the interpolation of the local spaces. In order to avoid rank-deficiencies and instabilities, this operation is usually done on a manifold of fixed-rank subspaces. In practical terms, one usually employs the Grassmann manifold, with each reduced space $\widetilde{\mathcal{V}}_j$ being represented by its projection matrix $\widetilde{V}_j$. We refer to [AF08] for more details on the manifold interpolation step. Also, we note that there are some similarities between this approach and the dynamical low rank (DLR) method [KL07]. However, in DLR the interpolation is carried out over a single parameter (time) while all the other parameters are already considered in the local surrogates.

The main advantage of this approach is that it solves the issue of an overly large global basis, since the basis size is kept fixed to R. The biggest disadvantage is that the manifold interpolation of the reduced basis over Θ is not at all straightforward, and usually prevents online efficiency, since identifying the interpolated basis at a new parameter $\theta^\star$ and projecting the FOM are intrinsically high-dimensional operations. We note that online efficiency might be recovered in some cases by introducing (in an offline pre-processing step) a low-dimensional approximation of the manifold interpolation operator [Son13]. Still, the online efficiency comes at the price of an inexact interpolation, which might have disastrous effects on the quality of the surrogate, especially if the number of parameters is large.

6.2.2.2 Interpolation of local transfer functions

Assume that, at each $\theta_j \in \widetilde{\Theta}$, a surrogate $\widetilde{H}_j$ for the transfer function $H(\cdot;\theta_j)$ has been built (note that it does not matter whether this was done via a non-intrusive or an intrusive approach). Then, the overall surrogate is obtained by polynomial (or rational) interpolation of $\{\widetilde{H}_j\}_{j=1}^T$, e.g., when $n_\theta = 1$, by setting

$$\widetilde{H}(z,\theta) = I^{\widetilde{\Theta}}(\widetilde{H}_\bullet(z))(\theta) = \sum_{j=1}^{T} \frac{\omega^{\widetilde{\Theta}}(\theta)}{(\theta-\theta_j)\frac{\mathrm{d}\omega^{\widetilde{\Theta}}}{\mathrm{d}z}(\theta_j)} \widetilde{H}_j(z),$$

or, more generally, using a separable ("polynomial chaos"-like) expansion

$$\widetilde{H}(z,\theta) = \sum_{j=1}^{T} \alpha_j(\theta)\widetilde{H}_j(z), \qquad \text{with } \alpha_j : \Theta \to \mathbb{C}\ \forall j. \tag{6.4}$$

The weights α_j may be found, e.g., by using "hat functions" on a triangulation of Θ, or radial basis functions. See Section 6.3.1.3 below for more details.

This idea is extremely simple, but has some quite serious issues. The most notable one is that the poles of the global ROM are *static*, i.e., they do not move as θ varies. In particular, the poles of $\widetilde{H}(\cdot,\theta)$ are the union of *all* the poles of *all* the local surrogates $\widetilde{H}_j$ whose weight $\alpha_j(\theta)$ is non-zero. This leads to the proliferation of poles, providing a rather compelling motivation for using locally supported basis functions (e.g., hat functions) as weights in the expansion (6.4).

Some instances of this class of methods can be found in [BB09; FKD11; Spi+15].

6.2.2.3 Interpolation of local reduced system matrices

Let a reduced system

$$\begin{cases} z\widetilde{E}_j\widetilde{v}_j(z) = \widetilde{A}_j\widetilde{v}_j(z) + \widetilde{B}_j u(z,\theta_j), \\ \widetilde{y}_j(z) = \widetilde{C}_j\widetilde{v}_j(z), \end{cases} \tag{6.5}$$

be available at each $\theta_j \in \widetilde{\Theta}$, where $\widetilde{y}_j(z) \approx y(z,\theta_j)$. Again, for simplicity we are assuming that all reduced systems have the same size R. Note that (6.5) might be obtained quite naturally by projecting the system (6.1) at $\theta = \theta_j$ onto some subspace $\widetilde{\mathcal{V}}_j$. However, one might also build (6.5) in a non-intrusive fashion, by finding first a surrogate $\widetilde{H}_j$ for the transfer function, and then constructing some system that has $\widetilde{H}_j$ as transfer function. For instance, the simple scalar transfer function

$$\widetilde{H}_j(z) = \sum_{i=1}^{R} \frac{\widetilde{r}_i^{(j)}}{z - \widetilde{\lambda}_i^{(j)}} \tag{6.6}$$

might be represented in the form (6.5) by setting, e.g.,

$$\widetilde{E}_j = I, \quad \widetilde{A}_j = \operatorname{diag}(\widetilde{\lambda}_1^{(j)}, \dots, \widetilde{\lambda}_R^{(j)}), \quad \widetilde{B}_j = \begin{bmatrix} \widetilde{r}_1^{(j)} \\ \vdots \\ \widetilde{r}_R^{(j)} \end{bmatrix}, \quad \text{and} \quad \widetilde{C}_j = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}. \tag{6.7}$$

Note that a similar construction can be carried out also for systems with more than one input/output, where, however, $\widetilde{C}_j$ might have a more complicated form and the size of the system matrices (6.7) is not necessarily R, but $\sum_{i=1}^{R} \operatorname{rank}(\widetilde{r}_i^{(j)})$, with "rank" the matrix rank. In particular, each pole $\widetilde{\lambda}_i^{(j)}$ appears $\operatorname{rank}(\widetilde{r}_i^{(j)})$ times in the diagonal of A.

Once (6.5) is available for all j, a global surrogate can be built by interpolating the reduced matrices, e.g., as

$$\begin{cases} z\left(\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{E}_j\right)\widetilde{v}(z,\theta) = \left(\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{A}_j\right)\widetilde{v}(z,\theta) + \left(\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{B}_j\right)u(z,\theta), \\ \widetilde{y}(z,\theta) = \left(\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{C}_j\right)\widetilde{v}(z,\theta). \end{cases} \tag{6.8}$$

with $\{\alpha_j\}_{j=1}^{T}$ weight functions, with $\alpha_j : \Theta \to \mathbb{C}$ for all j, as in Section 6.2.2.2.

Considering that the interpolated objects are reduced quantities of size R, online efficiency is guaranteed. However, a new issue emerges, due to the so-called *freedom introduced by realization* [YFB19b]: the expression of the reduced system (6.5) is not unique. Indeed, general changes of basis can be applied to the state $\widetilde{v}_j$ and also to the state equation itself: e.g., the system

$$\begin{cases} z(P_j\widetilde{E}_jQ_j^{-1})\widetilde{v}'_j(z) = (P_j\widetilde{A}_jQ_j^{-1})\widetilde{v}'_j(z) + (P_j\widetilde{B}_j)u(z,\theta_j), \\ \widetilde{y}_j(z) = (\widetilde{C}_jQ_j^{-1})\widetilde{v}'_j(z), \end{cases} \qquad \text{with } \widetilde{v}'_j(z) = Q_j\widetilde{v}_j(z),$$

has the same exact transfer function as (6.5), for all invertible P_j and Q_j. Still, the global surrogate (6.8) depends on the specific realization of the surrogates, so that it becomes necessary to pre-process the systems (6.5), making all their realizations compatible, before they can be combined.

Several approaches have been proposed for this purpose: for instance, one may set for simplicity

$P_j = Q_j$, and find Q_j by solving a generalized (non-orthogonal) Procrustes problem[1]

$$Q_j = \underset{Q:|\det(Q)|=1}{\arg\min} \left\| Q\widetilde{E}_j Q^{-1} - \widetilde{E}_{j'} \right\|_F^2 + \left\| Q\widetilde{A}_j Q^{-1} - \widetilde{A}_{j'} \right\|_F^2 + \\ + \left\| Q\widetilde{B}_j - \widetilde{B}_{j'} \right\|_F^2 + \left\| \widetilde{C}_j Q^{-1} - \widetilde{C}_{j'} \right\|_F^2, \quad (6.9)$$

with the aim of "matching" surrogates j and j' by a transformation of model j only. Note that weights might be introduced to balance the relative importance of the four terms in (6.9). Alternatively, in a projective setting, one may solve an (orthogonal) Procrustes problem involving the projection matrices $\widetilde{V}_j$ instead.

This step is then repeated with a sufficient number of surrogate pairs, until all models are matched. For more details on this intermediate step, we refer to [AF11; DVW10; LEP09; Pan+10].

6.2.2.4 Interpolation of local poles and residues

Assume that, at each $\theta_j \in \widetilde{\Theta}$, a surrogate $\widetilde{H}_j$ for the transfer function $H(\cdot;\theta_j)$ has been built, with simple partial fraction decomposition (6.6). Once more, we assume that the reduced size R is the same for all j. To define the overall ROM, we interpolate over Θ the poles and residues of the local surrogates:

$$\widetilde{H}(z,\theta) = \sum_{i=1}^{R} \frac{\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{r}_i^{(j)}}{z - \sum_{j=1}^{T} \alpha_j(\theta)\widetilde{\lambda}_i^{(j)}} = \sum_{i=1}^{R} \frac{\widetilde{r}_i(\theta)}{z - \widetilde{\lambda}_i(\theta)}, \quad (6.10)$$

with $\{\alpha_j\}_{j=1}^T$ scalar-valued weight functions, as in Section 6.2.2.2.

As in Section 6.2.2.3, here too we must worry about the freedom introduced by realization before interpolating. However, here this issue manifests itself in a milder way, only as the possibility of permuting the terms of the partial fraction decomposition. In order to find an "optimal" ordering of the addends, it is customary, see, e.g., [YFB19b], to solve the minimization problem

$$\min_{\sigma\in(1:R)!} \sum_{i=1}^{R} \left(\left| \widetilde{\lambda}_{\sigma_i}^{(j)} - \widetilde{\lambda}_i^{(j')} \right| + \left\| \widetilde{r}_{\sigma_i}^{(j)} - \widetilde{r}_i^{(j')} \right\| \right), \quad (6.11)$$

where $(1:R)!$ denotes the set of permutations of the tuple $(1,\dots,R)$. Note that weights might be introduced to balance the relative importance of poles and residues in (6.11). Then,

$$\widetilde{H}_j(z) = \sum_{i=1}^{R} \frac{\widetilde{r}_{\sigma_i}^{(j)}}{z - \widetilde{\lambda}_{\sigma_i}^{(j)}}$$

is the "best" reordering of the j-th surrogate with respect to the j'-th one. Note that, although (6.11) is a combinatorial problem (the discrete search space for σ has size $R!$), there are some ways to reduce the cost of its solution, see Section 6.3.1.2. A few alternative matching strategies

[1] Note that a closed-form solution of the nonlinear problem (6.9) is not available. A common way to simplify matters, see, e.g., [Pan+10], is to consider the *linearized unconstrained* problem

$$Q_j = \underset{Q}{\arg\min} \left\| Q\widetilde{E}_j - \widetilde{E}_{j'}Q \right\|_F^2 + \left\| Q\widetilde{A}_j - \widetilde{A}_{j'}Q \right\|_F^2 + \left\| Q\widetilde{B}_j - \widetilde{B}_{j'} \right\|_F^2 + \left\| \widetilde{C}_j - \widetilde{C}_{j'}Q \right\|_F^2$$

instead, which can be solved, e.g., by vectorization of Q.

are discussed in [YFB18]. This step is repeated with a sufficient number of surrogate pairs, until all models are matched.

We remark that, interestingly, one could have obtained (almost) the same method by considering the system representation of $\widetilde{H}_j$, see (6.7), and solving (6.9) under the constraint of Q being a permutation matrix. Moreover, note that, for simplicity, we have ignored additional constant or polynomial terms in the partial fraction decomposition (6.6), which might appear in applications whenever a rational approximation $\widetilde{H}_j$ is built with type $[M/N]$, $M \geq N$. Such terms do not need to be matched, since they can be interpolated over Θ as they are.

As opposed to the superposition of the transfer functions, see Section 6.2.2.2, here the number of poles stays fixed as θ varies. Indeed, continuous or, more generally, smooth functions (depending on the weight functions $\{\alpha_j\}_{j=1}^T$) are employed to model the behavior of poles and residues with respect to θ. As such, this class of surrogates might encounter some issues in approximating irregularities in poles and residues due to pole intersections, as we will discuss in the next section.

6.3 Additional aspects and improvements to pole/residue matching

Due to their favorable properties, in this section we restrict our attention to marginalized pMOR approaches based on interpolation of surrogate poles and residues, and we investigate some of their features and limitations. We note that, from this section onward, the content is, for the most part, original, and takes [NP21] as main reference.

6.3.1 Implementation aspects

From a practical viewpoint, the construction of the global surrogate $\widetilde{H}$ from the local ones $\{\widetilde{H}_j\}_{j=1}^T$ can be split in three steps.

6.3.1.1 Conversion to partial fraction form

In order to match and interpolate poles and residues of the surrogates, one has to find them first.

Assume that $\widetilde{H}_j$ is available in rational form $\widetilde{H}_j = \widetilde{P}_j/\widetilde{Q}_j$, with $\widetilde{P}_j$ and $\widetilde{Q}_j$ polynomials of respective degrees M and N, with $M \geq N-1$. This is the case, e.g., when using MRI. We wish to convert $\widetilde{H}_j$ to the partial fraction form

$$\widetilde{H}_j(z) = \sum_{i=1}^{N} \frac{\widetilde{r}_i^{(j)}}{z - \widetilde{\lambda}_i^{(j)}} + \sum_{i=0}^{M-N} \widetilde{r}_{-i}^{(j)} z^i \tag{6.12}$$

(note that the monomial basis in the second sum may be replaced by a general polynomial basis). The poles $\{\widetilde{\lambda}_i^{(j)}\}_{i=1}^N$ can be found as the roots of $\widetilde{Q}_j$, using an off-the-shelf root-finding algorithm. Note that, if $\widetilde{H}_j$ is given in barycentric coordinates (5.16), specialized (stable) root-finding algorithms are available [Kle12; NST18].

Once the poles are available, the residues can then be found in one of two ways:

- First, the proper residues $\{\widetilde{r}_i^{(j)}\}_{i=1}^N$ are found one by one, by using the formula

$$\widetilde{r}_i^{(j)} = \lim_{z \to \widetilde{\lambda}_i^{(j)}} (z - \widetilde{\lambda}_i^{(j)}) \widetilde{H}_j(z) = \frac{\widetilde{P}_j(\widetilde{\lambda}_i^{(j)})}{\frac{\mathrm{d}\widetilde{Q}_j}{\mathrm{d}z}(\widetilde{\lambda}_i^{(j)})}. \tag{6.13}$$

  Then the remaining improper residues $\{\widetilde{r}_{-i}^{(j)}\}_{i=0}^{M-N}$, if any, are found by imposing the interpolation conditions

$$\widetilde{H}_j(\widetilde{z}) = \sum_{i=1}^{N} \frac{\widetilde{r}_i^{(j)}}{\widetilde{z} - \widetilde{\lambda}_i^{(j)}} + \sum_{i=0}^{M-N} \widetilde{r}_{-i}^{(j)} \widetilde{z}^i \qquad \forall \widetilde{z} \in Z^\star, \tag{6.14}$$

  with $Z^\star \subset \mathbb{C} \setminus \{\widetilde{\lambda}_i^{(j)}\}_{i=1}^N$ a set of $M-N+1$ arbitrary distinct points. This requires evaluating the surrogate at $M - N + 1$ points and solving a linear system of the same size.

- All the residues can be found at the same time by solving the interpolation problem (6.14) with $Z^\star \subset \mathbb{C} \setminus \{\widetilde{\lambda}_i^{(j)}\}_{i=1}^N$ a set of $M + 1$ arbitrary distinct points. This requires evaluating the surrogate at $M + 1$ points and solving a linear system of the same size.

When projection-based methods are applied to LTI systems, the poles can be found by identifying the spectrum of the reduced matrix pencil $(\widetilde{A}_j, \widetilde{E}_j)$, cf. (6.5). Similarly, the residues can then be computed directly from the eigenvectors of the reduced matrix pencil, see Section 2.2, or by either of the two approaches described above.

Note that, in our presentation, we have ignored the issue of higher-order poles. In some circumstances, cf. the diagonalizable case in Section 2.2, multiple poles do not prevent the existence of the simple expansion (6.12). Indeed, they simply cause the first sum to have less than N terms, since some of the denominators are the same and can be grouped together. Note that this might cause an imbalance in the number of poles and residues. We discuss this at length in Section 6.3.3.

On the other hand, in general, higher-order roots might cause (6.12) to not exist, since higher-order denominators might become necessary. In this regard, we wish to remark that, numerically, due to round-off error in the root-finding algorithm, the poles will be always distinct. For instance, the conversion of z^{-2} to partial fraction form will likely yield two poles at $\pm\varepsilon$, $|\varepsilon| \ll 1$, with large (in magnitude) corresponding residues. Obviously, one might employ countermeasures to identify the pair of nearly identical roots and cluster them together. However, we will see in Section 6.3.2 that, by design, methods based on pole/residue-matching struggle with higher-order poles. Thus, it does not really matter whether higher-order poles are not properly identified, since the approximation quality will be rather poor anyway.

6.3.1.2 Matching the surrogates

Given the surrogates in partial fraction form $\{\widetilde{H}_j\}_{j=1}^T$, we wish to match them pairwise via (6.11) until they are all compatible. Note that a global approach, performing a "global" all-to-all matching of the models is an NP-hard combinatorial problem [Kar72], hence computationally unfeasible.

To this aim, we propose in [NP21] the following alternative approach. As a preliminary step, we interpret the sampled parameters $\{\theta_j\}_{j=1}^T$ as vertices of a complete graph, see Figure 6.1, and we

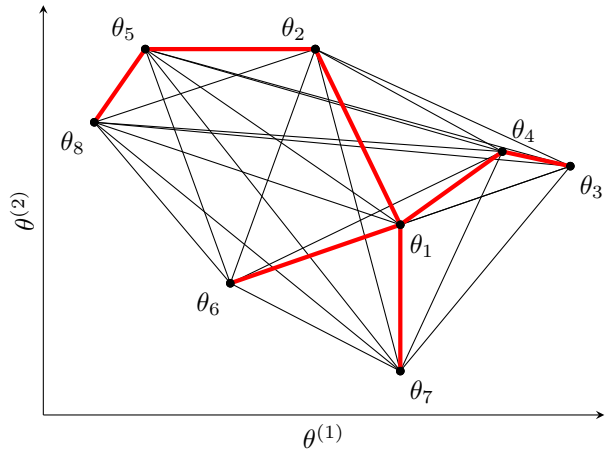


Figure 6.1 – Graph view of the sample points. The red spanning tree has minimal total length.

find a spanning tree of such graph. The $T-1$ indices corresponding to the edges of such graph identify a set of pairs $\{(j_l, j'_l)\}_{l=1}^{T-1} \subset \{1,\dots,T\}^2$. Then, we match in sequence the surrogates corresponding to each pair, by solving $T-1$ different problems of the form (6.11). As a way to help the matching procedure, it makes sense to select the spanning tree by minimizing its total length (in the $\mathbb{C}^{n_\theta}$-Euclidean norm), so that each matching happens between surrogates that are "as close as possible".

Now we turn to the issue of solving each matching problem. An exhaustive search over the possible values of σ is unfeasible due to the overly high number of options, namely, $R!$. In [YFB19a], a branch-and-bound algorithm is proposed to this aim, which terminates in $\mathcal{O}(R^2)$ if the two models to be matched are already "almost matched", i.e., if the optimal σ is close to the identical permutation $(1,\dots,R)$, but has worst-case complexity $\Omega(R!)$. We improved on this in [NP21], describing an algorithm for solving (6.11) in worst-case polynomial time. More specifically, define the $R \times R$ matrix D, as

$$(D)_{ii'} = \left|\widetilde{\lambda}_{\sigma_i}^{(j)} - \widetilde{\lambda}_i^{(j')}\right| + \left\|\widetilde{r}_{\sigma_i}^{(j)} - \widetilde{r}_i^{(j')}\right\|.$$

Each entry of D represents the "cost" of matching pole/residue i in model j to pole/residue i' in model j', cf. (6.11). The matching step can be equivalently characterized as: extract exactly one entry of D per row and per column so as to minimize the sum of the selected entries. This can be achieved by using tools from network flow optimization, as described in [Cro16]. Such algorithm is implemented in the `linear_sum_assignment` function in the `scipy.optimize` module [Vir+20], which takes the cost matrix D as only input. As shown in [Cro16, Section II.C], this method has $\mathcal{O}(R^3)$ worst-case complexity.

6.3.1.3 Interpolating poles and residues

After the models have been properly matched, it remains to actually build the global surrogate (6.10). This can be done either by constructing the T weights $\{\alpha_j\}_{j=1}^T$, with $\alpha_j : \Theta \to \mathbb{C}$, or by directly computing the interpolated poles and residues, namely, $\{\widetilde{\lambda}^{(i)}\}_{j=1}^R$ and $\{\widetilde{r}^{(i)}\}_{j=1}^R$, with $\widetilde{\lambda}^{(i)} : \Theta \to \mathbb{C}$ and $\widetilde{r}^{(i)} : \Theta \to \mathbb{C}^{n_y \times n_u}$, directly. In the following, we will discuss the former case,

since extensions to the latter are usually quite trivial.

Here, the main objective is enforcing, for all $j = 1, \dots, T$, the (Lagrange) interpolation condition

$$\alpha_j(\theta_{j'}) = \delta_{jj'} \qquad \forall j' = 1, \dots, T, \tag{6.15}$$

with δ the Kronecker delta. To this aim, one may, e.g., choose $\{\alpha_j\}_{j=1}^T$ as the family of Lagrange polynomials associated to $\{\widetilde{\theta}_j\}_{j=1}^T$, even though this is not straightforward in high(er) dimensions. Of course, this should be done with care, in order to avoid numerical instabilities, as well as the standard issues that arise in multivariate polynomial interpolation if $n_\theta > 1$. Alternatively, we note that radial basis functions and Gaussian processes might be particularly interesting choices.

Another important approximation class can be used when the sample points are in a regular arrangement, e.g., if they form the vertices of a triangulation of Θ or a sparse grid, see Section 6.4.1. In such situations, one may use locally supported piecewise-linear "hat functions", in "finite element"-style. The resulting interpolation strategy is quite flexible, suitable even for the approximation of non-smooth (but continuous) quantities. Obviously, there is another side to the coin: due to the lack of global smoothness, the piecewise-linear approximation of uniformly (in θ) smooth functions will necessarily yield worse results than, e.g., global polynomials.

As a final note, we mention that, in some cases, one may want to weaken the interpolation constraint (6.15) to an LS version

$$\widetilde{\lambda}^{(i)} = \underset{\widetilde{\lambda} \in \star}{\arg\min} \sum_{j=1}^{T} w_j \left| \widetilde{\lambda} - \widetilde{\lambda}_j^{(i)} \right|^2$$

and

$$\widetilde{r}^{(i)} = \underset{\widetilde{r} \in *}{\arg\min} \sum_{j=1}^{T} w_j \left| \widetilde{r} - \widetilde{r}_j^{(i)} \right|^2,$$

with $\star$ and $*$ suitable interpolation classes, e.g., polynomials of sufficiently low degree. This may be done in the interest of numerical stability, but also as a way to "compress" the resulting ROM, thus improving online efficiency.

6.3.2 Approximation power of parametric partial fraction form

We can expect a pMOR strategy based on interpolation of surrogate poles and residues to yield good results as long as the simple partial fraction expansion (6.2) holds for all $\theta \in \Theta$, with the most relevant poles and residues depending smoothly enough on θ. As usual, pole relevance is determined based on the target frequency range. Conversely, the smoothness of poles and residues determines the approximation class where the interpolation weights $\{\alpha_j\}_{j=1}^T$ should be sought.

That being said, all the interpolation methods that we presented require at least continuity and boundedness of poles and residues. This assumption can be guaranteed if the FOM matrix pencil $(A(\theta), E(\theta))$, cf. (6.2), satisfies some hypotheses, namely:

1) Invertibility of $E(\theta)$ for all $\theta \in \Theta$. This guarantees the boundedness of all poles.

2) Simultaneous diagonalizability of $E(\theta)$ and $A(\theta)$ for all $\theta \in \Theta$ (if E is invertible, this is equivalent to the diagonalizability of $E^{-1}A$). This guarantees the continuity and boundedness of

all residues, and the existence of a simple partial fraction expansion (6.2).

Under these assumptions, even if two (or more) poles intersect at some θ, they remain semisimple throughout Θ, and their order in the partial fraction decomposition remains equal to 1.

Note that 1) and 2) are sufficient conditions for a good behavior of the *relevant* poles and residues, but they are not necessary. Indeed, we can expect our pMOR approach to work well even if some faraway pole/residue pairs are not smooth, as long as we do not ask the ROM to approximate them. Accordingly, a somewhat rigorous statement of our minimal assumptions could read: if we wish to identify R pole/residue pairs, we require the R most important poles (e.g., using the Green's potential of the frequency range to define relative importance), as well as the corresponding residues, to be smooth enough.

In some practical cases, it is actually possible to show that 1) and 2) (or, at least, their weakened "local" versions) hold, by employing *a priori* information on the spectral properties of system. However, in the vast majority of cases, this is not possible. In fact, some situations of interest are known to have polynomial bifurcations in the poles, with couples of real poles becoming complex-conjugate. We showcase this with a simple example.

Consider the parametric system (6.1), with $n_\theta = 1$ and

$$A(\theta) = \begin{bmatrix} 0 & \theta - 0.1 \\ 1 & 0 \end{bmatrix}, \; E(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; B(\theta) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } C(\theta) = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

The corresponding transfer function is readily obtained:

$$H(z,\theta) = \frac{1}{z^2 - (\theta - 0.1)} = \frac{\frac{1}{2\sqrt{\theta-0.1}}}{z - \sqrt{\theta - 0.1}} + \frac{-\frac{1}{2\sqrt{\theta-0.1}}}{z + \sqrt{\theta - 0.1}}. \tag{6.16}$$

We see that $\theta = \theta^\star = 0.1$ is a bifurcation parameter, where A is not diagonalizable, so that the residues become unbounded there. Now, we apply the pole/residue-matching approach using T uniformly spaced sample points $0 = \theta_1 < \ldots < \theta_T = 1$ over $\Theta = [0, 1]$, employing piecewise-linear hat functions to interpolate poles and residues. In what follows, we choose $T \in \{5, 9, 17\}$, and we focus on the approximation of the transfer function over the frequency range $A = [-1, 1]$. Note that, for simplicity, we interpolate the exact poles and residues of the system, rather than surrogate ones, so that we can skip the not-so-interesting (in the present section) construction of the local frequency ROMs. The code used in our experiments below is publicly available as part of [Pra21].

In Figure 6.2, we show the exact response y and the relative approximation error. We notice vertical lines with vanishing error at the θ-sample points, since the local frequency responses are interpolated exactly. If frequency surrogates had been used instead of the exact transfer function, we would see the z-approximation error at the θ-sample points instead. We can see that the error is large near the system poles due to the piecewise-linear approximation of the quadratic pole curve $\lambda(\theta)$. Moreover, the error is also large *for all* z when θ is located in the interval between sample points that contains $\theta^\star$, i.e., $]\theta_1, \theta_2[$ for $T \in \{5, 9\}$ and $]\theta_2, \theta_3[$ for $T = 17$. This is due to the fact that, locally, we are approximating the quadratic bifurcation with a pair of straight (with respect to θ) surrogate poles that "twist" around the branch point in the complex plane. Such poles are not able to provide a faithful approximation of the transfer function (notably, they break z-symmetry), so that the difficulty in approximating the bifurcation can also affect frequencies far from the bifurcation frequency $z = 0$.

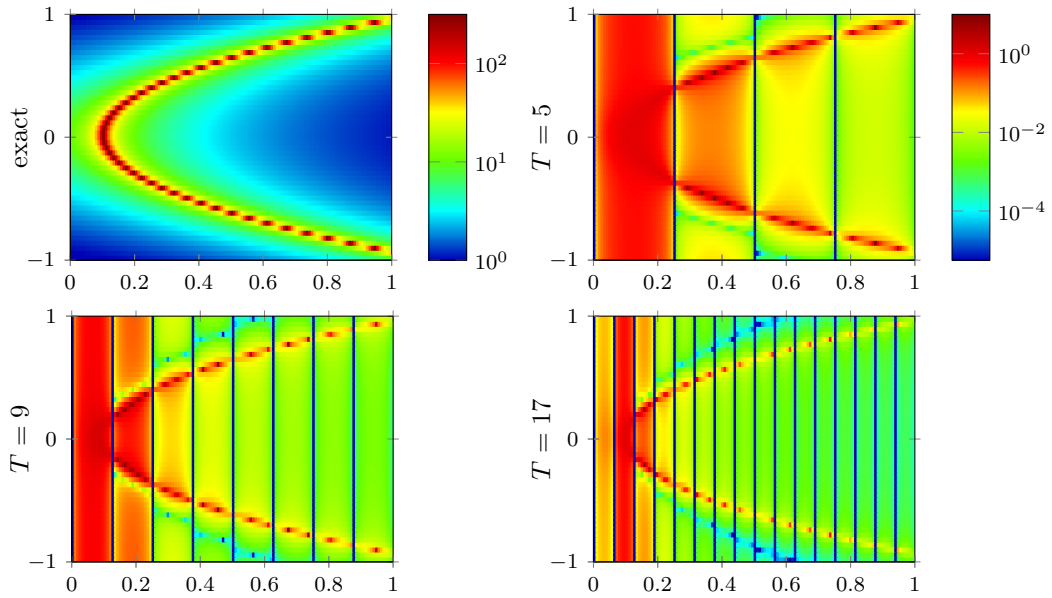


Figure 6.2 – Plot of exact $|y(z, \theta)|$ in the top left plot. In the other plots, we show the relative error for $T \in \{5, 9, 17\}$ samples of θ. Piecewise linear hat functions are used for θ-interpolation. All errors have the same color scale, reported next to the top right plot. All plots have θ on the x-axis and z on the y-axis.

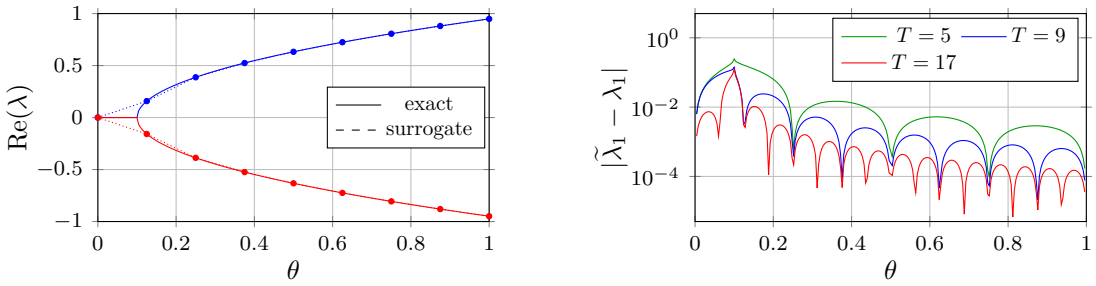


Figure 6.3 – Pole approximations for $T = 9$ (left), where different colors are used for the two poles. Pole approximation error for λ_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for λ_2 is identical. Piecewise linear hat functions are used for θ-interpolation.

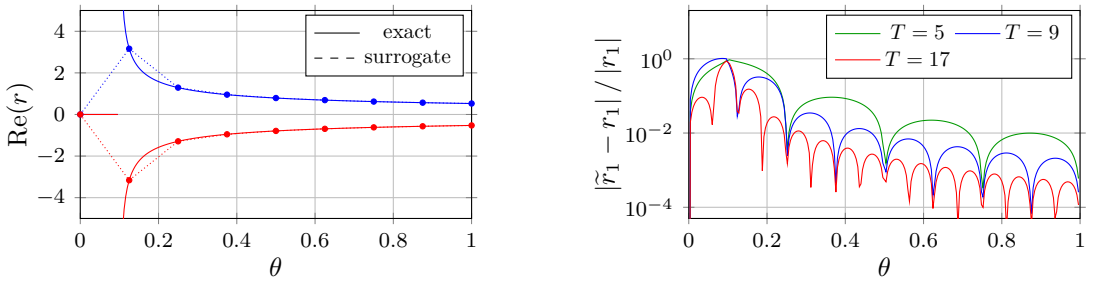


Figure 6.4 – Residue approximations for $T = 9$ (left), where different colors are used for the two residues. Residue approximation error for r_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for r_2 is identical. Piecewise linear hat functions are used for θ-interpolation.

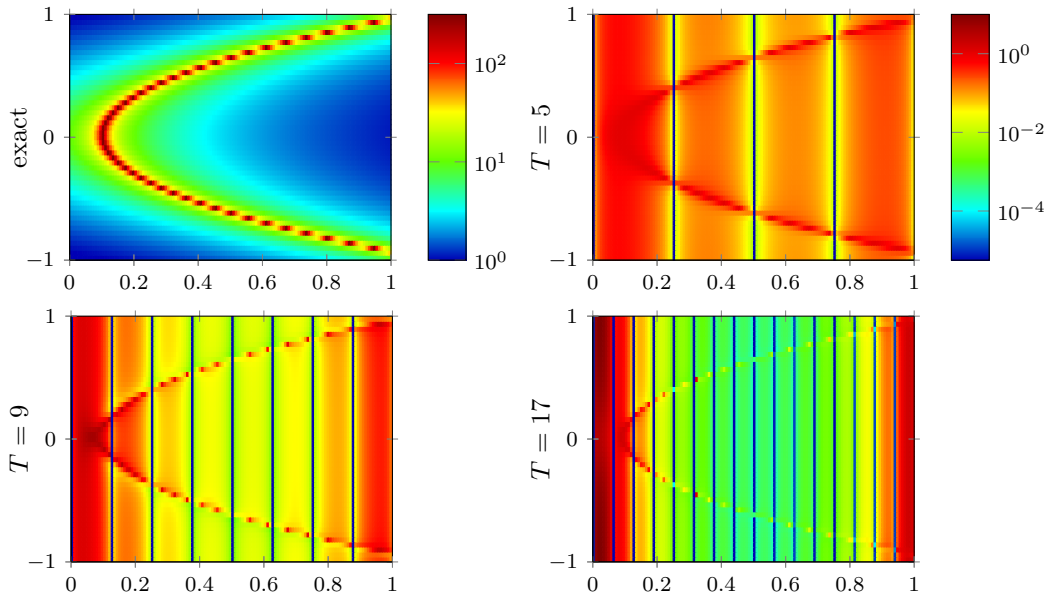


Figure 6.5 – Plot of exact $|y(z, \theta)|$ in the top left plot. In the other plots, we show the relative error for $T \in \{5, 9, 17\}$ samples of θ. Global polynomials are used for θ-interpolation. All errors have the same color scale, reported next to the top right plot. All plots have θ on the x-axis and z on the y-axis.

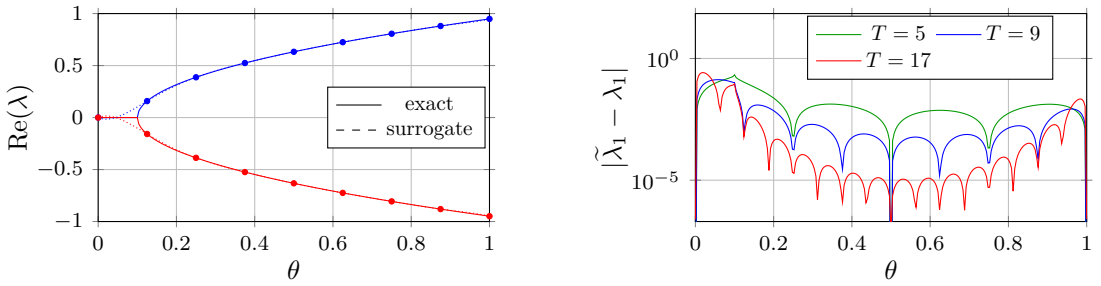


Figure 6.6 – Pole approximations for $T = 9$ (left), where different colors are used for the two poles. Pole approximation error for λ_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for λ_2 is identical. Global polynomials are used for θ-interpolation.

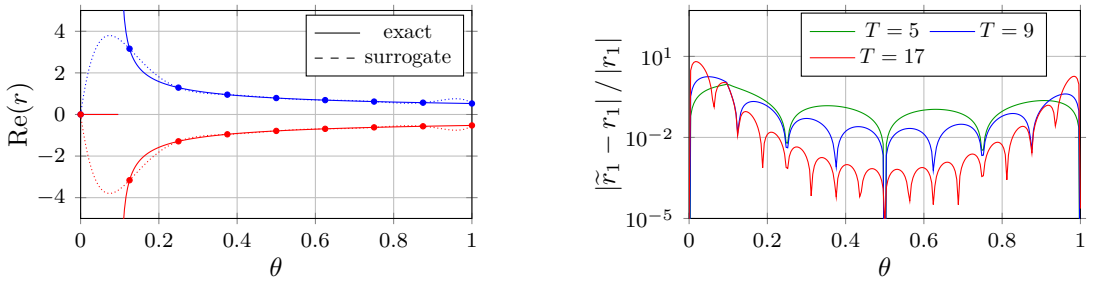


Figure 6.7 – Residue approximations for $T = 9$ (left), where different colors are used for the two residues. Residue approximation error for r_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for r_2 is identical. Global polynomials are used for θ-interpolation.

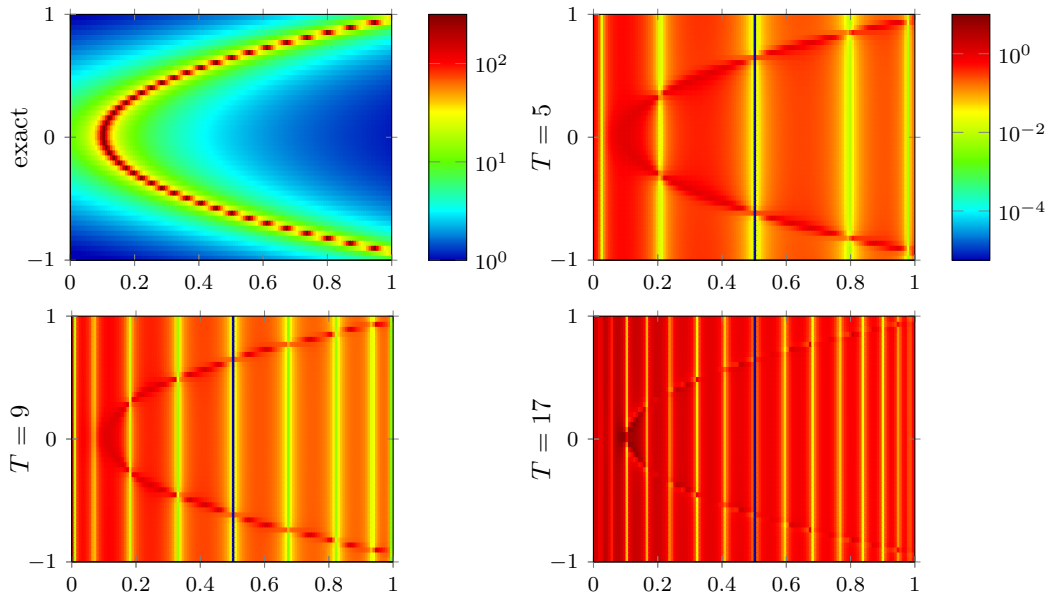


Figure 6.8 – Plot of exact $|y(z, \theta)|$ in the top left plot. In the other plots, we show the relative error for $T \in \{5, 9, 17\}$ samples of θ. Chebyshev polynomials are used for θ-interpolation. All errors have the same color scale, reported next to the top right plot. All plots have θ on the x-axis and z on the y-axis.

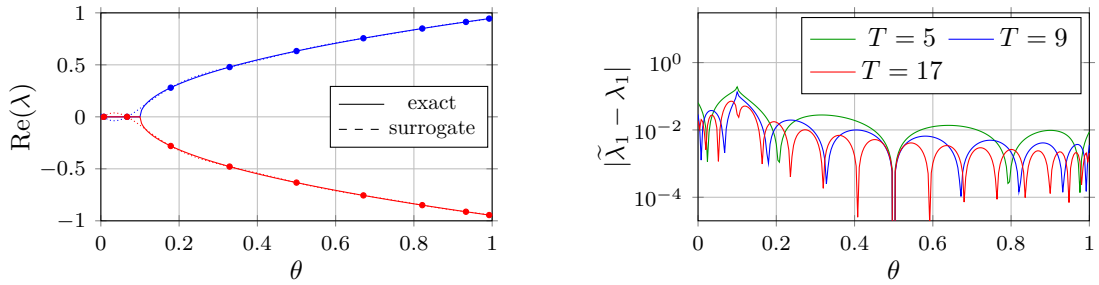


Figure 6.9 – Pole approximations for $T = 9$ (left), where different colors are used for the two poles. Pole approximation error for λ_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for λ_2 is identical. Chebyshev polynomials are used for θ-interpolation.

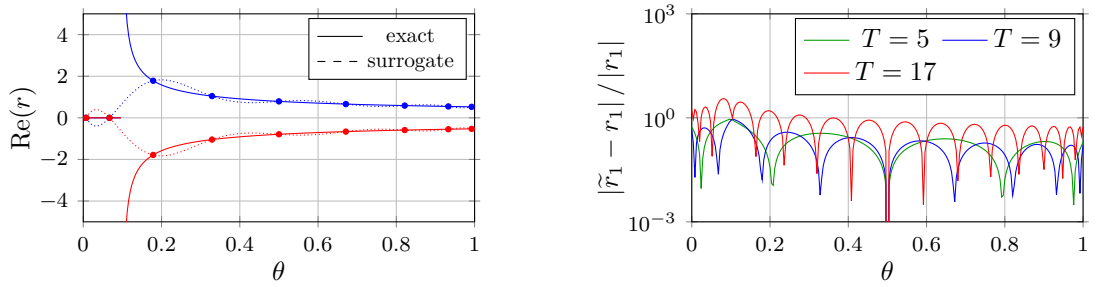


Figure 6.10 – Residue approximations for $T = 9$ (left), where different colors are used for the two residues. Residue approximation error for r_1 (right) for $T \in \{5, 9, 17\}$. By symmetry, the error for r_2 is identical. Global polynomials are used for θ-interpolation.

However, thanks to the small support of the employed hat functions, the instability is only local in θ. We can confirm this by looking at the error in the approximation of poles and residues in Figures 6.3 and 6.4. It is crucial to observe that, if globally supported functions had been used for the θ-interpolation, the approximation could, in general, have been globally inaccurate due to the adverse effect of the bifurcation. We proceed to show this numerically. We build a surrogate using exactly the same information as before. However, this time, the interpolation over Θ is carried out via global polynomials. More specifically, given T uniformly spaced θ-samples, we interpolate poles and residues using polynomials of degree $T-1$. We show the corresponding results in Figures 6.5 to 6.7, where we can observe that the above-mentioned global effect of the bifurcation manifests itself as Gibbs oscillations near the edges of Θ, in accordance to standard polynomial approximation theory. Note that we pick T small enough that the uniform spacing of the sample points is *not* the cause of the instabilities.

We further verify this by repeating the experiment using Chebyshev polynomials over Chebyshev points of Θ, i.e., $\theta_j = \frac{1}{2} - \frac{1}{2}\cos\left(\frac{2j-1}{2T}\pi\right)$ for $j = 1, \dots, T$, so that the interpolation problem is optimally conditioned. The corresponding results, in Figures 6.8 to 6.10, display a uniformly poor approximation quality. Particularly, looking at the residue error plot in Figure 6.10, we see that all the local maxima of the relative error have approximately the same value, in accordance to the optimality property of Chebyshev polynomials. Unfortunately a bad approximation of the bifurcation makes such uniform maximal error quite large. As we increase T, we should *not* expect such value to decrease (by much).

These not-so-satisfactory results are intrinsic to the pole/residue-matching approach with continuous approximation in θ, since we are trying to approximate an unbounded quantity (the residues) with a bounded one (polynomials). In order to alleviate this issue, one could allow rational approximation with respect to θ. However, we note that bifurcations correspond to *essential singularities* in the residues. Accordingly, a rational approximation should not be expected to converge rapidly, if at all.

In our view, the only way to tackle the approximation of singularities in a satisfactory way is by avoiding the partial fraction decomposition (at least, locally around the singularity). Indeed, from (6.16), we see that we can recover the exact transfer function by joint (z, θ)-rational approximation using polynomials of degrees $(2, 1)$. Unfortunately, in general, a joint rational approximation suffers from the issues described in Section 6.2.1. A marginalized rational approximation might yield better results, without incurring in the heavy limitations induced by the curse of dimension. Equivalently, one could try to identify the "bad" bifurcating poles and, locally, group them together. This yields a hybrid joint/marginalized approach. We discuss this further in Section 8.1.

6.3.3 Unbalanced surrogate matching

The matching approach described in Section 6.2.2.4, solvable with the algorithm from Section 6.3.1.2, was presented under the assumption that the two surrogates to be matched have the same size (more precisely, the same number of poles). As we describe in [NP21], this assumption can be weakened to allow the matching of unbalanced surrogates. To this aim, let the two surrogates $\widetilde{H}_j$ and $\widetilde{H}_{j'}$ have R_j and $R_{j'}$ poles, respectively, with $R_j > R_{j'}$. We consider the rectangular matching problem

$$\min_{\sigma \in (1:R_j)!} \sum_{i=1}^{R_{j'}} \left(\left| \widetilde{\lambda}_{\sigma_i}^{(j)} - \widetilde{\lambda}_i^{(j')} \right| + \left\| \widetilde{r}_{\sigma_i}^{(j)} - \widetilde{r}_i^{(j')} \right\| \right), \tag{6.17}$$

which can be solved by applying a slight generalization of the algorithm from Section 6.3.1.2, see [NP21]. Then $(\sigma_1, \dots, \sigma_{R_{j'}})$ gives the desired permutation of the $R_{j'}$ "matched" poles and residues, while the remaining indices $(\sigma_{R_{j'}+1}, \dots, \sigma_{R_j})$ remain free and can be sorted arbitrarily. Note that, by invertibility of the optimal permutation, the case $R_j < R_{j'}$ can be approached by switching the roles of the two surrogates.

Now, it remains to decide how to deal with the $R_j - R_{j'}$ unassigned poles and residues. We propose three different options:

a) We might believe such poles to be spurious in the richer j-th surrogate due to some inaccuracy in building it. Then, we can throw away the corresponding terms from the partial fraction decomposition of $\widetilde{H}_j$. Note that, in order to preserve the good approximation properties at θ_j, one should compensate for the removed terms. For instance, assume that we decide to remove the R_j-th term from (6.12) (with $R_j = N$). Then, we replace $\widetilde{H}_j$ with the adjusted surrogate

$$\widetilde{H}_j(z) = \sum_{i=1}^{R_j-1} \frac{\widetilde{r}_i^{(j)}}{z - \widetilde{\lambda}_i^{(j)}} + \sum_{i=0}^{M-R_j+1} \widetilde{r}_{-i}^{(j)} z^i, \tag{6.18}$$

where the number of smooth terms has been increased by 1. The modified coefficients $\{\widetilde{r}_{-i}^{(j)}\}_{i=0}^{M-R_j+1}$ are chosen so that

$$\frac{\widetilde{r}_{R_j}^{(j)}}{z - \widetilde{\lambda}_{R_j}^{(j)}} + \sum_{i=0}^{M-R_j} \widetilde{r}_{-i}^{(j)} z^i \approx \sum_{i=0}^{M-R_j+1} \widetilde{r}_{-i}^{(j)} z^i$$

in some sense. For instance, if $\widetilde{H}_j$ was built by MRI from frequency samples at Z_j, we might find the adjusted coefficients by enforcing interpolation over Z_j.

b) We might believe such poles to be missing in the poorer j'-th surrogate, due to some inaccuracy in building it. Then, we can copy them from $\widetilde{H}_j$ to $\widetilde{H}_{j'}$. As in the previous case, in order to preserve the good approximation properties at $\theta_{j'}$, this should be accompanied by a modification of the original terms of $\widetilde{H}_{j'}$. However, we note that, here, it is not necessary to increase the number of smooth terms $M - R_j$. See below for a practical instance of this step.

c) We might believe such poles to be missing in $\widetilde{H}_{j'}$ because they are too far away from the frequency range of interest and, as such, they could not be properly identified by the surrogate at $\theta_{j'}$, cf. Theorem 3.5. Then, we can add $R_j - R_{j'}$ poles at ∞ to the j'-th surrogate:

$$\widetilde{H}_{j'}(z) = \sum_{i=1}^{R_{j'}} \frac{\widetilde{r}_i^{(j')}}{z - \widetilde{\lambda}_i^{(j')}} + \sum_{i=R_{j'}+1}^{R_j} \frac{\widetilde{r}_{\sigma_i}^{(j)}}{z - \infty} + \sum_{i=0}^{M-R_{j'}} \widetilde{r}_{-i}^{(j')} z^i, \tag{6.19}$$

where we note that the corresponding residues were copied from the j-th surrogate.

Since the correction is vanishing at $\theta = \theta_{j'}$, this does not require any modification to the original coefficients of $\widetilde{H}_{j'}$. However, it introduces the additional difficulty of having to interpolate unbounded poles over θ. To this aim, we propose to apply a piecewise-polynomial/rational interpolation of the poles, extending the approach based on hat functions described in Section 6.3.1.3. For this, we introduce locally singular (rational) basis functions, e.g., $\alpha_{\theta_j}(\theta) = 1/|\theta - \theta_j|$, to replace the usual polynomial hat functions at locations with unbounded poles.

We compare the three approaches in a simple synthetic example, whose code is available as part of [Pra21]. Our target is the approximation of the transfer function

$$H(z,\theta) = \frac{1}{z - \left(\frac{1}{2} - \frac{1}{4}\theta\right)} + \frac{2}{z - \left(\frac{1}{5}\theta - \frac{1}{3}\right)},$$

with $(z,\theta) \in A \times \Theta = [-1,1] \times [-1,1]$. We note that H has two real non-intersecting linearly drifting poles and constant residues. We take parameter sample points at 5 uniformly spaced points over Θ. We assume that the local frequency surrogate is exact at the 4 points $\theta_1 = -1$, $\theta_3 = 0$, $\theta_4 = \frac{1}{2}$, and $\theta_5 = 1$, i.e.,

$$\widetilde{H}_j(z) = H(z,\theta_j) = \frac{1}{z - \left(\frac{1}{2} - \frac{1}{4}\theta_j\right)} + \frac{2}{z - \left(\frac{1}{5}\theta_j - \frac{1}{3}\right)} \qquad \text{for } j \in \{1,3,4,5\}. \tag{6.20}$$

However, we assume that the local frequency surrogate at $\theta_2 = -\frac{1}{2}$ fails to identify the first pole:

$$\widetilde{H}_2(z) = -\frac{8}{5} + \frac{2}{z + \frac{13}{30}}. \tag{6.21}$$

Note that a constant term has been added so that the bad surrogate, albeit missing a pole, achieves interpolation of H at $(z,\theta) = (0, -\frac{1}{2})$.

In approach a), we remove the first pole (i.e., the positive one) from all surrogates $\widetilde{H}_j$ for $j \in \{1,3,4,5\}$. Following approach b) (resp. c)), we adjust the surrogate $\widetilde{H}_2$ by adding an extra pole/residue term, with pole $\widetilde{\lambda} = \widetilde{\lambda}_1^{(1)} = \frac{3}{4}$ (resp. $\widetilde{\lambda} = \infty$) and residue $\widetilde{r} = \widetilde{r}_1^{(1)} = 1$. The additions/removals of pole/residue terms in approaches a) and b) are followed by adjustments of the local smooth terms so as to guarantee interpolation of H at $z = 0$. More specifically, in approach a), $\widetilde{H}_j$, for $j \in \{1,3,4,5\}$, changes from (6.20) to

$$\widetilde{H}_j(z) = \frac{1}{\frac{1}{4}\theta_j - \frac{1}{2}} + \frac{2}{z - \left(\frac{1}{5}\theta_j - \frac{1}{3}\right)},$$

whereas, in approach b), $\widetilde{H}_2$ changes from (6.21) to

$$\widetilde{H}_2(z) = \frac{1}{z - \frac{3}{4}} + \frac{2}{z + \frac{13}{30}} - \frac{4}{15}.$$

No correction is necessary in approach c). We employ piecewise-linear hat functions for the θ-interpolation in all cases, except in approach c) for the unbounded pole near θ_2, where the singular basis function $\alpha_{\theta_2}(\theta) = 1/|\theta - \theta_2|$ is used instead.

The results are shown in Figure 6.11. We can observe that a) has a single surrogate pole at each θ, whereas the other approaches have two. Since our setup falls exactly in the framework of case b) (i.e., the pole/residues unbalance is caused by a local pole being missed by a local surrogate), it is this method that provides the best result. Still, we can see that, overall, the reconstructed approximation is quite poor. Indeed, we cannot expect to perform better if crucial information is missing from the local surrogates.

Oftentimes, in more realistic applications, imbalances in the number of poles arise *outside* the frequency range, so that the effects of the model correction are less visible. This is particularly the case in two situations of practical interest: when poles of H enter and exit the parameter range as θ varies and when the local surrogates are built by an adaptive method, cf. Sections

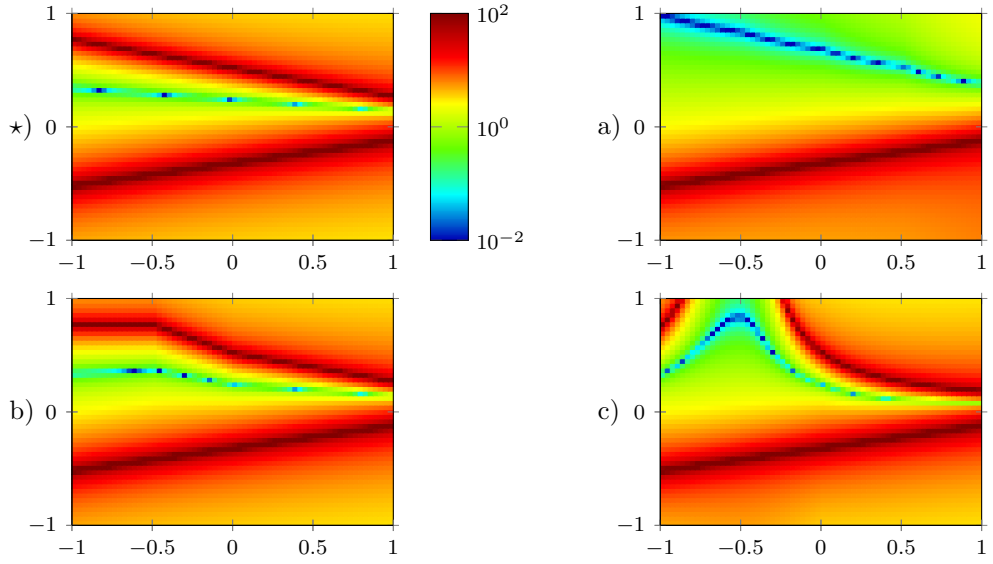


Figure 6.11 – Figure ⋆) (top left): plot of exact $|y(z,\theta)|$. In the other plots, we show the surrogate $|\tilde{y}(z,\theta)|$ for $T = 5$ samples of θ. Different methods are used to deal with the missing top pole at $\theta = -\frac{1}{2}$. All plots have the same color scale. All plots have θ on the x-axis and z on the y-axis.

2.4.1.1, 2.4.2.3, and 5.3. Notably, in the latter case, we should never find ourselves in case b), assuming that a good criterion for model selection/greedy sampling has been used.

In [NP21] we propose a heuristic way to blend the two approaches a) and b), allowing to remove some extra poles *and* reconstruct some missing ones. The main idea is quite simple:

- First, we apply b) to obtain a balanced global surrogate

$$\widetilde{H}(z,\theta) = \sum_{i=1}^{R} \frac{\sum_{j=1}^{T} \alpha_j(\theta)\widetilde{r}_i^{(j)}}{z - \sum_{j=1}^{T} \alpha_j(\theta)\widetilde{\lambda}_i^{(j)}} + \widetilde{r}_0(z,\theta), \tag{6.22}$$

  where $R = \max_j R_j$ and $\widetilde{r}_0$ is smooth, without throwing away any pole.

- Then, for each $i \in \{1,\dots,R\}$, we count how many times we had to copy the i-th pole/residue over from a surrogate to another, i.e., how many elements of $\{\widetilde{\lambda}_i^{(j)}\}_{j=1}^{T}$ were artificially added. If the above count is larger than T times some user-prescribed tolerance (between 0 and 1), then the i-th term is removed from (6.22).

- If any partial fraction term was removed at the above step, apply the necessary adjustments (to the smooth terms) so that the approximation quality is preserved, cf. the discussion above.

We provide a pseudo-code for this procedure in Algorithm 5.

It is interesting to note that the reconstruction of missing poles introduces an asymmetry in the matching procedure, so that the order in which the models are matched matters. For a visual example, see the situation depicted in Figure 6.12. Note that, depending on whether $\delta \gtrless \frac{2}{3}$, one or

Algorithm 5 Pole/residue matching for generic (unbalanced) local frequency surrogates

Require: distinct sample points $\widetilde{\Theta} = \{\theta_1, \dots, \theta_T\} \subset \mathbb{C}^{n_\theta}$
Require: local surrogates $\{\widetilde{H}_1, \dots, \widetilde{H}_T\}$
Require: matching root index $j^\star \in \{1, \dots, T\}$, tolerance for synthetic poles $\delta \in [0, 1]$
  initialize the explored sets as $J = \{j^\star\}$
  while $\#J < T$ **do**
    breadth-first search: $(j, j') \leftarrow \arg\min_{j \in J,\, j' \in \{1,\dots,T\} \setminus J} |\theta_j - \theta_{j'}|$
    if $R_j \geq R_{j'}$ **then**
      find optimal permutation σ by solving (6.17)
      append the synthetic terms $\sum_{i=R_{j'}+1}^{R_j} \widetilde{r}_{\sigma_i}^{(j)} / (z - \widetilde{\lambda}_{\sigma_i}^{(j)})$ to $\widetilde{H}_{j'}$
      correct the smooth terms of $\widetilde{H}_{j'}$ to account for the added terms and update $R_{j'} := R_j$
      apply the inverse permutation of σ to $R_{j'}$
    else
      find optimal permutation σ by solving (6.17) with j and j' switched
      apply the permutation σ to $R_{j'}$
      for $j \in J$ **do**
        append the synthetic terms $\sum_{i=R_j+1}^{R_{j'}} \widetilde{r}_i^{(j')} / (z - \widetilde{\lambda}_i^{(j')})$ to $\widetilde{H}_j$
        correct the smooth terms of $\widetilde{H}_j$ to account for the added terms and update $R_j := R_{j'}$
      end for
    end if
    add j' to J
  end while
  set $I = \emptyset$
  for $i = 1, \dots, R_1$ **do**
    count how many elements of $\{\widetilde{\lambda}_i^{(1)}, \dots, \widetilde{\lambda}_i^{(T)}\}$ are synthetic
    if count$> \delta\, T$ **then**
      add i to I
    end if
  end for
  for $j = 1, \dots, T$ **do**
    remove the terms $\sum_{i \in I} \widetilde{r}_i^{(j)} / (z - \widetilde{\lambda}_i^{(j)})$ from $\widetilde{H}_j$
    correct the smooth terms of $\widetilde{H}_j$ to account for the removed terms and update $R_j := R_j - \#I$
  end for
  return $\{\widetilde{H}_1, \dots, \widetilde{H}_T\}$

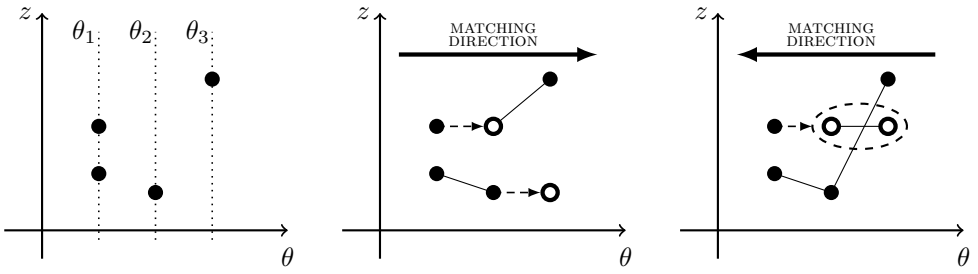


Figure 6.12 – Example of history-dependent matching. Poles of the surrogates are denoted by full dots (left plot). Results by matching left-to-right (middle plot) and right-to-left (right plot). Dashed arrows and empty circles are used to denote pole duplication steps and synthetic poles, respectively.

two poles will withstand the "synthetic tolerance" check if the matching is carried out right-to-left. On the other hand, two or no poles will remain after left-to-right matching, depending on $\delta \gtrless \frac{1}{3}$.

For this reason, it is important to choose well the root in the breadth-first exploration of $\widetilde{\Theta}$, namely, $j^\star$ in Algorithm 5. Unfortunately, there is no practical way to tell what index is optimal, if any. From a computational point of view, it makes sense to choose as root the surrogate with the largest number of poles, so that we never have to retrace our steps to add synthetic poles, i.e., we are always in the first case of the first "if" statement in Algorithm 5.

Moreover, we remark that copying poles over from one surrogate to the next is quite blunt, especially when the parameter resolution is low. If the poles depend smoothly on θ, it is preferable to employ a reconstruction with a larger stencil, for instance global polynomial extrapolation, using information from all the surrogates that contain the missing pole, see Section 6.3.1.3. However, this is not always viable during the matching loop, since we explore $\widetilde{\Theta}$ breadth-first: for instance, if $R_j < R_{j'}$, we are forced to reconstruct poles from the single new model $\widetilde{H}_{j'}$. A similar problem may arise in the case $R_j > R_{j'}$ if the already-explored index set J, cf. Algorithm 5, is too small. Still, it remains feasible to extrapolate all the synthetic poles with higher order after the matching loop is complete.

6.4 Adaptive parameter sampling

Until now, we have assumed the parameter sample points $\widetilde{\Theta}$ to have been fixed in advance. However, in many situations, it proves extremely useful to have some kind of adaptivity included in the sampling of Θ, so that samples may be added only where the surrogate model is particularly inaccurate, e.g., in our case, near pole mismatches or where large θ-interpolation errors occur. Still, it is quite difficult to devise adaptive strategies in non-intrusive pMOR, especially if the number of parameters n_θ is large, since not much is known about the parametric dependence of the problem.

We propose here a technique for adaptive θ-sampling based on *locally refined sparse grids*, closely related to that considered in [Als+19], which, in turn, relies on some ideas from [MZ09; PPB10]. In the next sections, we first introduce locally refined sparse grids and how to interpolate over them. Then, the description of our θ-adaptive pMOR approach follows. Our main reference in this section is [NP21].

6.4.1 Locally refined sparse grids

For simplicity, we carry out our construction in the case $\Theta = [-1, 1]^{n_\theta}$. Generalizations to more complicated parameter domains may be obtained by isomorphism. First, consider the nested ($\Gamma(n) \subseteq \Gamma(n+1)$ for all n) one-dimensional point sets

$$\Gamma(d) = \begin{cases} \emptyset & \text{if } d < 0, \\ \{0\} & \text{if } d = 0, \\ \{2^{1-d} j\}_{j=-2^{d-1}}^{2^{d-1}} & \text{if } d > 0. \end{cases} \tag{6.23}$$

We extend this definition to multiple dimensions by tensorization: for any *level index* $\boldsymbol{d} = (d^{(1)}, \ldots, d^{(n_\theta)}) \in \mathbb{Z}^{n_\theta}$, we define the corresponding *tensor grid* $\Gamma(\boldsymbol{d}) = \Gamma(d^{(1)}) \times \Gamma(d^{(2)}) \times \ldots \times$

$\Gamma(d^{(n_\theta)})$. It is useful to define the infinite discrete point set

$$\Xi = \bigcup_{\boldsymbol{d}\in\mathbb{Z}^{n_\theta}} \Gamma(\boldsymbol{d}) = \bigcup_{d=0}^{\infty} \Gamma(d)^{n_\theta},$$

which is dense in Θ (it coincides with the dyadic rationals in Θ) and also, by construction, a superset of any tensor grid. For our purposes, it suffices to (improperly) define a *sparse grid* as a (finite) subset[2] of Ξ. We will define the set of adaptive sample points using sparse grids.

Now, given any point $\theta \in \Xi$, we can find a unique *depth* $\boldsymbol{d} = \boldsymbol{d}(\theta)$ such that

$$\theta \in \Gamma(\boldsymbol{d}) \setminus \left(\bigcup_{i=1}^{n_\theta} \Gamma(\boldsymbol{d} - \boldsymbol{e}_i) \right), \tag{6.24}$$

with $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{Z}^{n_\theta}$ a zero vector with a single 1 at the i-th component. Equivalently stated, the coordinates of $\theta = \left(j^{(1)}/2^{d^{(1)}-1}, \ldots, j^{(n_\theta)}/2^{d^{(n_\theta)}-1} \right) \in \Xi$ are fractions in lowest terms (with $d^{(i)} = 0$ if $j^{(i)} = 0$).

We define the *forward points* of θ as the ($\leq 2n_\theta$) elements of the discrete neighborhood

$$U(\theta) = \bigcup_{i=1}^{n_\theta} \left\{ \widetilde{\theta} \in \Gamma(\boldsymbol{d} + \boldsymbol{e}_i) : \left| \widetilde{\theta} - \theta \right| = 2^{-d^{(i)}} \right\} = \Theta \cap \bigcup_{i=1}^{n_\theta} \left\{ \theta \pm 2^{-d^{(i)}} \boldsymbol{e}_i \right\},$$

Moreover, to each $\theta \in \Xi$, with (6.24), we associate a *hierarchical hat function* $\varphi_\theta : \Theta \to [0, 1]$ according to the definition

$$\varphi_\theta(\widetilde{\theta}) = \prod_{i=1}^{n_\theta} \widetilde{\varphi}_{\theta^{(i)}, d^{(i)}}(\widetilde{\theta}^{(i)}), \tag{6.25}$$

with $\widetilde{\varphi}_{0,0}(\cdot) = 1$ and, for $d = 1, 2, \ldots$,

$$\widetilde{\varphi}_{\theta,d}(\widetilde{\theta}) = \begin{cases} 1 - 2^{d-1} \left| \widetilde{\theta} - \theta \right| & \text{if } \left| \widetilde{\theta} - \theta \right| < 2^{1-d}, \\ 0 & \text{if } \left| \widetilde{\theta} - \theta \right| \geq 2^{1-d}. \end{cases}$$

Note that hierarchical hat functions might be continuously extended to the whole $\mathbb{C}^{n_\theta}$ through the definition

$$\varphi_\theta(\theta') = \varphi_\theta \left(\arg\min_{\theta'' \in \Theta} |\theta'' - \theta'| \right) \qquad \forall \theta \in \mathbb{C}^{n_\theta}.$$

By construction, φ_θ is hierarchical in the following sense: $\varphi_\theta(\theta') = 0$ at all $\theta' \in \Xi$ of which θ is a forward point (the *backward points* of θ). Moreover, $\varphi_\theta(\theta'') = 0$ also at all backward points $\theta'' \in \Xi$ of such θ', etc., all the way back to $\theta = \mathbf{0}$. We show some two-dimensional examples of forward points and of hierarchical hat functions in Figure 6.13.

We rely on hierarchical hat functions to cast piecewise-linear interpolation problems over sparse grids. More precisely, given sample points $\widetilde{\Theta} = \{\theta_j\}_{j=1}^T \subset \Xi$, and data $\{f(\theta_j)\}_{j=1}^T$, the *piecewise-linear interpolant of f based on samples at $\widetilde{\Theta}$* is the unique element $\widetilde{I^{\widetilde{\Theta}}}(f)$ of $\operatorname{span}\{\varphi_{\theta_j}\}_{j=1}^T$ which interpolates exactly the data: this means that there exist unique coefficients $\{c_j\}_{j=1}^T$, depending

[2]The *classical* definition of sparse grids, see, e.g., [BNR00; MZ09; NTW08], assumes the subset to have a specific structure. Namely, a sparse grid has the form $\bigcup_{\boldsymbol{d}\in\Delta} \Gamma(\boldsymbol{d})$, with $\Delta \subset \mathbb{Z}^{n_\theta}$ an index set.

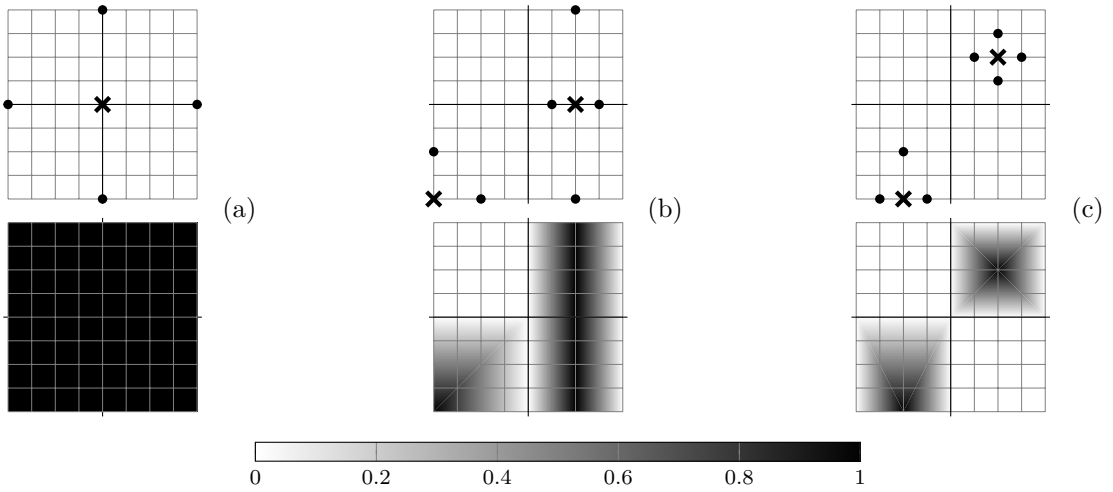


Figure 6.13 – In the top row, the forward points of $(0,0) \in \Gamma(0,0)$ in (a), $(-1,-1) \in \Gamma(1,1)$ and $(1/2,0) \in \Gamma(2,0)$ in (b), and $(-1/2,-1) \in \Gamma(2,1)$ and $(1/2,1/2) \in \Gamma(2,2)$ in (c). In the bottom row, the corresponding hierarchical hat functions.

only on $\widetilde{\Theta}$ and $\{f(\theta_j)\}_{j=1}^T$, such that

$$f(\theta_j) = \widetilde{I}^{\widetilde{\Theta}}(f)(\theta_j) = \sum_{j'=1}^{T} c_{j'} \varphi_{\theta_{j'}}(\theta_j) \qquad \forall j = 1, \dots, T. \tag{6.26}$$

The (Lagrangian) interpolation weight functions (α_j in (6.15)) can then be found as piecewise-linear interpolants of the data $\{f(\theta_{j'}) = \delta_{jj'}\}_{j'=1}^T$. We note that the expression of each hierarchical basis function (6.26) depends only on its support point θ_j, whereas the expression of each Lagrangian weight function depends on the whole set of sample points $\widetilde{\Theta}$.

Let $\widetilde{\Theta} \subset \Xi$ be a *downward-closed set*, i.e., for all $\theta \in \widetilde{\Theta}$, all the backward points of θ are also in $\widetilde{\Theta}$. Then, our definition of interpolation coincides with the standard one on sparse grids [BNR00]. This allows a rather nice interpretation of the expansion coefficients $\{c_j\}_{j=1}^T$ as *hierarchical surpluses*, which provide pointwise information on the approximation error between different sparse grid levels. On the other hand, when $\widetilde{\Theta}$ is not downward-closed, the corresponding interpolation might be quite misbehaved: for instance, if $\mathbf{0} \notin \widetilde{\Theta}$, then $\widetilde{I}^{\widetilde{\Theta}}(f)(\mathbf{0}) = 0$, regardless of f. In the following, we will assume that $\widetilde{\Theta}$ is at least *sequentially hierarchical*, i.e., that, for all $\theta \in \widetilde{\Theta}$, there exists a sequence $\{\widetilde{\theta}_0, \dots, \widetilde{\theta}_L\} \subset \widetilde{\Theta}$, such that $\widetilde{\theta}_0 = \mathbf{0}$, $\widetilde{\theta}_L = \theta$, and $\widetilde{\theta}_{j+1}$ is a forward point of $\widetilde{\theta}_j$ for all j.

To conclude, we wish to mention the existence of an alternative family of sparse grids, those of so-called *Haar*-type [CCS14; MZ09], which can be obtained as above by replacing (6.23) with

$$\Gamma(d) = \begin{cases} \emptyset & \text{if } d \leq 0, \\ \{2^{1-d} j\}_{j=-2^{d-1}+1}^{2^{d-1}-1} & \text{if } d > 0. \end{cases}$$

This prevents any sampling on the boundary of Θ, restricting the sparse grid points to its interior. This class of sparse grids is particularly natural when continuous piecewise-linear interpolation is replaced by discontinuous piecewise-constant interpolation, e.g., through the hierarchical

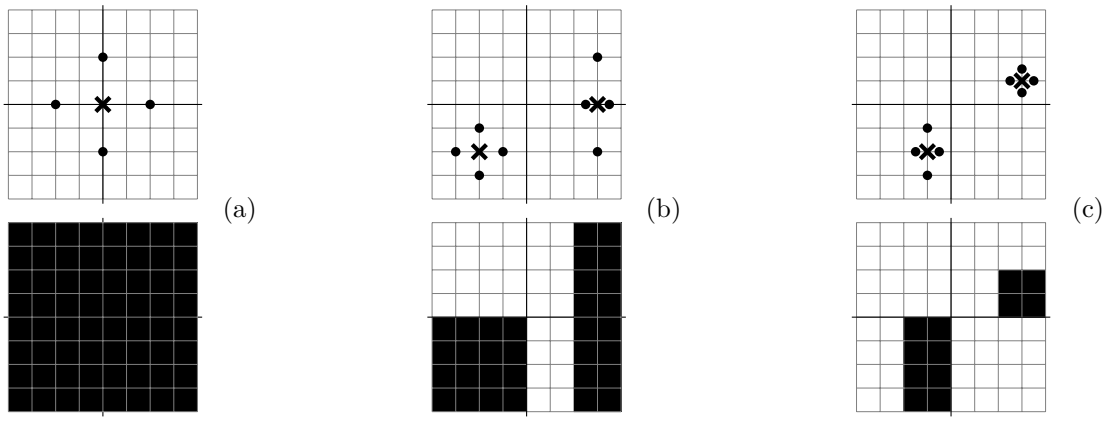


Figure 6.14 – In the top row, the Haar-forward points of $(0,0) \in \Gamma(1,1)$ in (a), $(-1/2,-1/2) \in \Gamma(2,2)$ and $(3/4,0) \in \Gamma(3,1)$ in (b), and $(-1/4,-1/2) \in \Gamma(3,2)$ and $(3/4,1/4) \in \Gamma(3,3)$ in (c). In the bottom row, the corresponding hierarchical piecewise-constant functions (white is 0 and black is 1).

piecewise-constant basis functions

$$\phi_\theta(\widetilde{\theta}) = \prod_{i=1}^{n_\theta} \chi\left(\widetilde{\theta}^{(i)} \in [\theta^{(i)} - 2^{1-d^{(i)}}, \theta^{(i)} + 2^{1-d^{(i)}}]\right)$$

with $(d^{(1)}, \ldots, d^{(n_\theta)})$ the level index associated to θ and χ the indicator function, equal to 1 if its argument is true, and 0 otherwise. See Figure 6.14 for some examples of Haar points and hierarchical basis functions. We remark that employing such bases for a discontinuous interpolation of poles and residues is equivalent to a θ-"nearest neighbor" approach, effectively allowing to skip the pole/residue-matching step completely.

6.4.2 The look-ahead strategy for greedy parameter sampling

Now we are ready to describe our adaptive technique, which is summarized in Algorithm 6. The main idea is to build a (z,θ)-surrogate based on parameter samples $\widetilde{\Theta} = \{\theta_j\}_{j=1}^T$ that are points of Ξ, and then use the forward points of $\widetilde{\Theta}$ as a test set, where the accuracy of the surrogate is verified. If the surrogate model is too inaccurate at some of the test points, they are added to $\widetilde{\Theta}$, a new surrogate is computed, and the test set is enlarged. This loop is repeated until a specified tolerance is achieved at all current test points.

Within each iteration, in order to quantify the accuracy of the current ROM at a test parameter value θ_{test}, we use the following look-ahead strategy:

- We use the current ROM $\widetilde{H}$ (whose training set does not include θ_{test}, nor any of the other test points) to predict the frequency response at θ_{test}, by $\widetilde{H}(\cdot, \theta_{\text{test}})$.

- Via MRI (or any other frequency-domain MOR approach), we build a frequency surrogate $\widetilde{H}_{\text{test}}$ at θ_{test}, which we take as "truth frequency response" at θ_{test}. This requires taking new snapshots at θ_{test}, at as many frequency points as required (note that adaptive frequency sampling may be employed).

Algorithm 6 Adaptive parameter sampling

Require: initial parameter sample points $\widetilde{\Theta} = \{\theta_1, \dots, \theta_{T_0}\} \subset \mathbb{C}^{n_\theta}$ (sequentially hierarchical)
Require: algorithm for building local ROMs, algorithm for combining local ROMs
Require: tolerance $\epsilon > 0$
loop
  for $\theta_j \in \widetilde{\Theta}$ **do**
    build frequency ROM $\widetilde{H}_j \approx H(\cdot, \theta_j)$ via the prescribed algorithm
      (if a frequency ROM has already been built at θ_j, just load it from memory)
  end for
  build a global ROM $\widetilde{H}$ by combining the local ones via the prescribed algorithm
  initialize $\Theta_{\text{next}} = \emptyset$
  define the test set Θ_{test} as $\{\text{forward points of } \theta\}_{\theta \in \widetilde{\Theta}} \setminus \widetilde{\Theta}$
  for $\theta_{\text{test}} \in \Theta_{\text{test}}$ **do**
    evaluate $\widetilde{H}(\cdot, \theta_{\text{test}})$
    build frequency ROM $\widetilde{H}_{\text{test}} \approx H(\cdot, \theta_{\text{test}})$ via the prescribed algorithm
      (if a frequency ROM has already been built at θ_{test}, just load it from memory)
    convert $\widetilde{H}(z, \theta_{\text{test}})$ and $\widetilde{H}_{\text{test}}$ to partial fraction form
    evaluate $\text{dist}(\widetilde{H}(\cdot, \theta_{\text{test}}), \widetilde{H}_{\text{test}})$ as in (6.27)
    if dist $> \epsilon$ **then**
      add θ_{test} to Θ_{next}
    end if
  end for
  if $\Theta_{\text{next}} = \emptyset$ **then**
    return $\widetilde{H}$
  end if
  append Θ_{next} to $\widetilde{\Theta}$
end loop

- We convert the two models to partial fraction form:

$$\widetilde{H}(z, \theta_{\text{test}}) = \sum_{i=1}^{\widetilde{R}} \frac{\widetilde{r}_i}{z - \widetilde{\lambda}_i} + \widetilde{r}_0(z) \quad \text{and} \quad \widetilde{H}_{\text{test}}(z) = \sum_{i=1}^{\widetilde{R}_{\text{test}}} \frac{\widetilde{r}_i^{(\text{test})}}{z - \widetilde{\lambda}_i^{(\text{test})}} + \widetilde{r}_0^{(\text{test})}(z),$$

  with $\widetilde{r}_0$ and $\widetilde{r}_0^{(\text{test})}$ smooth terms, e.g., polynomials. Note that $\widetilde{H}(\cdot, \theta_{\text{test}})$ is already in partial fraction form if the pole/residue-matching strategy is employed.

- We judge the accuracy of $\widetilde{H}$ at θ_{test} by evaluating its "partial fraction distance"[3] from $\widetilde{H}_{\text{test}}$, i.e.,

$$\text{dist}(\widetilde{H}(\cdot, \theta_{\text{test}}), \widetilde{H}_{\text{test}}) = \min_{\sigma \in (1:\widetilde{R})!} \sum_{i=1}^{\widetilde{R}_{\text{test}}} \left(\left| \widetilde{\lambda}_{\sigma_i} - \widetilde{\lambda}_i^{(\text{test})} \right| + \left\| \widetilde{r}_{\sigma_i} - \widetilde{r}_i^{(\text{test})} \right\| \right). \tag{6.27}$$

  This requires the solution of a matching problem, see Section 6.3.1.2. Note that we have assumed, without loss of generality, that $\widetilde{R} \geq \widetilde{R}_{\text{test}}$. If this is not the case, it suffices to switch the roles of the two ROMs.

For the sake of efficiency, it is crucial to observe that, over the different θ-greedy iterations, the

[3] Actually, (6.27) does not define a distance, e.g., because it is only semi-positive definite when $\widetilde{R} \neq \widetilde{R}_{\text{test}}$.

frequency ROM at some $\theta^\star$ might be required several times, not only when evaluating the quality of the current surrogate (if $\theta^\star$ is in the test set), but also when building the global surrogate (if $\theta^\star$ is in the training set). As long as memory is not an issue, one should store frequency surrogates built at previous steps, so that no expensive snapshot is wasted.

An additional observation can be made, again in the context of not wasting snapshots: in Algorithm 6, all the local surrogates for $\theta_{\text{test}} \in \Theta_{\text{test}}$ remain unused when the method terminates. This occurrence is very similar to the single snapshot that gets wasted in the look-ahead z-adaptive sampling proposed in Sections 5.3.3 and 5.3.4. Still, here, the scale is much larger, since the test set can (and usually does) contain quite a large number of parameter values, at each of which several snapshots were taken. For this reason, it makes a lot of sense to carry out a post-processing step, where Θ_{test} is appended to $\widetilde{\Theta}$, and a new global surrogate is built using *all* the computed local surrogates. This is the default behavior in the `RROMPy` package.

6.4.2.1 Features and limitations of look-ahead approach

Our proposed approach, differently from the usual *isotropic* adaptive sparse grid sampling [BNR00; NTW08], in general does not add whole levels $\Gamma(\boldsymbol{d})$, but only subsets of them. In fact, the training set is not even guaranteed to be downward-closed, but only sequentially hierarchical. The matter of missing backward points is discussed to some detail in [Als+19, Section 3.2]. Here, we do *not* require missing backward points to be added to the training set, both for simplicity of exposition and (mainly) to reduce the cost of the offline phase. At the same time, if one can afford a higher offline time, including backward points is advisable, even though the increase in training cost could be significant, since each sparse grid point has up to $2n_\theta$ backward points, and a (costly) frequency model must be built at each of them for error estimation.

The main advantage of locally refined sparse grids is, rather obviously, the possibility of performing local refinements near the parameter values where the current surrogate is worse. Notably, we note that such local refinements are carried out when generating the test set via forward points. Still, our proposed look-ahead adaptive sampling can be generalized to non-local refinement schemes. We proceed by giving more details on one such generalization:

- Let $\{\theta_1, \theta_2, \ldots\} \subset \Theta$ be a sequence of *procedurally generated* sample points. For instance, we may draw each θ_j independently from a certain random distribution over Θ or, more interestingly, we might generate the sample points via a quasi-random (e.g., Sobol or Halton) sequence generator.
- Initialize the training set as $\widetilde{\Theta} = \{\theta\}_{j=1}^{T_0}$, and the test set as $\Theta_{\text{test}} = \{\theta_j\}_{T_0+1}^{T_0+N_{\text{test}}}$.
- Build the surrogate based on local frequency ROMs at $\widetilde{\Theta}$ and find all elements of Θ_{test} that do not satisfy the tolerance.
- Move such test points from Θ_{test} to $\widetilde{\Theta}$, and add new test points $\{\theta_{T_0+N_{\text{test}}+1}, \ldots\}$ to Θ_{test}. Note that we may keep the size N_{test} of the test set fixed or we might increase it as the training set gets larger and larger.
- If no new test points were added, terminate. Otherwise, repeat the loop.

The biggest drawback of this approach is that the new test points are *not* selected adaptively, since they are chosen incrementally from the sequence $\{\theta_1, \theta_2, \ldots\} \subset \Theta$, which was fixed *a priori*.

This motivates the idea of increasing the size of the test set as the algorithm proceeds, as a way to better explore the parameter domain.

On the other hand, this sampling strategy has the favorable property of suffering from the curse of dimension even less than sparse grids. Indeed, in the approximation theory literature, we can find many examples of (quasi-)random sequences of points being successfully used to approximate functions over extremely high-dimensional spaces [Coo+20; Kuo+21; MN15]. Unfortunately, the lack of geometric structure in the sample points makes it tricky to use locally supported (hat-like) basis functions, cf. Section 6.3.1.3, at least in a natural way. Indeed, one could "force" a hyper-triangulation based on the training points, but this can be computationally expensive for modest or large numbers of parameters due, again, to the curse of dimension. As a more favorable alternative, it is customary to employ radial basis functions or, more generally, kernel approximation for the interpolation of the target quantity (here, poles and residues).

As a final note, we wish to stress that look-ahead sampling strategies are, by their very nature, heuristic. In particular, we cannot guarantee that, at the end of the greedy loop, the tolerance will be attained over the whole parameter domain, since we are using a relatively small (and sparse) test set to quantify the approximation error. Representing ("sketching") the parameter domain via the test set can be justified only by assuming the resolution of the test set to be sufficiently fine. However, in practice, this is usually computationally unfeasible (especially if the number of parameters is large, due to the curse of dimension).

6.4.2.2 Dörfler-based adaptivity for parameter sampling

The adaptive strategy described in the previous section employs the partial fraction distance as a measure of "closeness" between truth and surrogate rational models. In particular, the sampling algorithm terminates only when this distance is uniformly small over the test set. However, in some cases, see Section 7.1.2, one can numerically observe that such distance never actually gets below the prescribed tolerance. As we will see, this is usually due to poorly approximated poles *outside* the frequency range, which are positioned far apart in the truth and surrogate models.

One naive way of counteracting this issue is to "cut off" the less relevant poles of the local models before computing their partial fraction distance. For instance, one could remove from the local surrogates all the poles that are too far from the frequency range A, employing the Green's potential of A to define the distance of each pole from A, cf. Assumption 3.5. Note that, in order to guarantee good approximation, see Section 6.3.3, pole removal should be accompanied by an adjustment of the local surrogate. Still, this introduces an additional parameter that needs to be chosen well, namely, the cut-off level. In particular, if the level is too large, the problem is not solved at all. On the other hand, if the level is too small, some relevant poles might be removed.

A different (but not necessarily disjoint) way of solving the problem is to use the partial fraction distance only to determine which test points are badly approximated by the current surrogate, but employ a different strategy to decide when to stop the sampling. To this aim, one could simply fix in advance a "computational budget", which corresponds to how many snapshots we are willing to take. Then, we proceed with the θ-adaptive sampling as described above. Only, this time, we terminate the algorithm if the tolerance is attained *or* if the computational budget is exhausted.

This "computational budget" idea guarantees that the algorithm will terminate. Still, due to the above-mentioned poorly approximated poles (if they exist), the algorithm may perform

refinements only around the parameters to which such "bad" poles belong, leading to rather coarse sampling over the rest of the parameter domain. This behavior is not exactly desirable, since a large portion of the computational budget is (almost) wasted by taking snapshots around the "bad" surrogate(s). Instead, it seems more reasonable to use the extra samples to explore the parameter domain slightly more "uniformly". For this purpose, we describe here a Dörfler-like approach [Dör96]:

- Fix in advance the computational budget and an adaptivity parameter $\eta \in]0, 1[$ (usually denoted by θ in the literature).
- At each iteration of Algorithm 6, compute the partial fraction distance at all test points $\{\mathrm{dist}_{\mathrm{test}}^{(j)}\}_{\theta_{\mathrm{test}}^{(j)} \in \Theta_{\mathrm{test}}}$, with $\mathrm{dist}_{\mathrm{test}}^{(1)} \geq \ldots \geq \mathrm{dist}_{\mathrm{test}}^{(T_{\mathrm{test}})}$.
- Add to the training set the T test points with the largest distances, with T being the smallest integer such that
$$\sum_{j=1}^{T+1} \mathrm{dist}_{\mathrm{test}}^{(j)} > \eta \sum_{j=1}^{T_{\mathrm{test}}} \mathrm{dist}_{\mathrm{test}}^{(j)}.$$
- If the computational budget has been exceeded, terminate. Otherwise, proceed with the next iteration of the sampling loop.

In this approach, we no longer need to fix a tolerance on the partial fraction distance, since test points are added not in function of their *absolute* local error, but depending on their *relative* contribution to the total testing error. Note that the value of the parameter η determines whether we want to carry out more localized (for small η) or more uniform (for large η) θ-refinements, cf. our numerical experiments in Chapter 7.

7 Numerical tests

In this chapter, we apply the pMOR approach described in the previous sections to some applications involving parametric frequency-response problems. For the sake of reproducibility, the code used to obtain all our results is made publicly available as part of [Pra21]. We note that the PDE examples presented in Sections 7.2 and 7.3 are discretized with the finite element (FE) method, using the FEniCS library [Aln+15].

Beside the tests performed below, we remark that other examples of (successful) application of MRI and of our proposed marginalized pMOR method can be found in [Bon+20a; Bon+20b; BP19; Pra20] and [BP21; NP21], respectively.

7.1 Vibrations of a PAC-MAN-like drum

Our first application is rather academic, but we include it nonetheless because it showcases fairly well a few interesting properties of MRI, but also some of the intrinsic difficulties that arise in the parametric case.

Let $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)}) \in \mathbb{R}^3$ be a vector of real parameters. We consider a parametric 2D PAC-MAN-like spatial domain

$$\Omega = \Omega(\theta^{(1)}) = B^{\mathbf{0}}(1) \setminus \left\{ (x^{(1)}, x^{(2)}) : -\tan(\theta^{(1)}) x^{(1)} \leq x^{(2)} \leq \tan(\theta^{(1)}) x^{(1)} \right\}, \tag{7.1}$$

for $0 < \theta^{(1)} < \frac{\pi}{2}$, see Figure 7.1. The associated $\theta^{(1)}$-dependent Hilbert space is

$$\mathcal{V} = \mathcal{V}(\theta^{(1)}) = H_0^1(\Omega(\theta^{(1)})) = \left\{ v \in H^1(\Omega(\theta^{(1)})) : v|_{\partial\Omega(\theta^{(1)})} = 0 \right\}, \tag{7.2}$$

which we endow with the usual inner product $\langle v, w \rangle_{\mathcal{V}} = \langle \operatorname{grad} v, \operatorname{grad} w \rangle_{[L^2(\Omega)]^2}$. Moreover, we define a piecewise-constant parametric forcing term

$$u(\theta^{(2)}) \in L^2(B^{\mathbf{0}}(1)), \quad \text{with} \quad u(\theta^{(2)})\Big|_x = \begin{cases} 1/(\pi(\theta^{(2)})^2) & \text{if } |x - (0, 0.6)| < \theta^{(2)}, \\ 0 & \text{otherwise,} \end{cases} \tag{7.3}$$

for $0 < \theta^{(2)} < 0.4$, and a piecewise-constant parametric sensor (corresponding to a local average)

$$g(\theta^{(3)}) \in L^2(B^{\mathbf{0}}(1)), \quad \text{with} \quad g(\theta^{(3)})\Big|_x = \begin{cases} 1/(\pi(\theta^{(3)})^2) & \text{if } |x-(0,-0.6)| < \theta^{(3)}, \\ 0 & \text{otherwise,} \end{cases} \tag{7.4}$$

for $0 < \theta^{(3)} < 0.4$. We consider the parametric Helmholtz equation with homogeneous Dirichlet boundary conditions

$$\begin{cases} -(\Delta + z)v(z,\theta) = u(\theta^{(2)}) & \text{in } \Omega(\theta^{(1)}), \\ v(z,\theta) = 0 & \text{on } \partial\Omega(\theta^{(1)}), \\ y(z,\theta) = \langle v(z,\theta), g(\theta^{(3)})\rangle_{L^2(\Omega(\theta^{(1)}))}. \end{cases} \tag{7.5}$$

In the following, in order to lighten the notation, we will often omit parametric dependence.

Using standard tools from PDE analysis, namely, Sobolev embeddings and the Fredholm alternative, we can show that, for all $\theta \in]0, \frac{\pi}{2}[\times]0, 0.4[^2$, (7.5) admits a unique solution in $\mathcal{V}$, for all $z \in \mathbb{C} \setminus \Lambda$, with $\Lambda = \Lambda(\theta^{(1)})$ being the (positive and discrete) spectrum of the minus Laplacian operator $-\Delta : \mathcal{V}(\theta^{(1)}) \to \mathcal{V}(\theta^{(1)})^\star = H^{-1}(\Omega(\theta^{(1)}))$. In particular, note that Λ is independent of the input and output parameters $\theta^{(2)}$ and $\theta^{(3)}$, while y obviously is not.

Due to the simplicity of the problem, we can actually pinpoint the spectrum Λ exactly. To this aim, consider the homogeneous version of (7.5), obtained by setting $u = 0$. A complex number z belongs to Λ iff there exists a non-trivial solution to such problem. We move then to polar coordinates, so that Ω maps to the rectangle $(\rho, \phi) \in]0, 1[\times]\theta^{(1)}, 2\pi - \theta^{(1)}[= \Omega'(\theta^{(1)})$. We make a separable ansatz $v(x, y) = w_\rho(\rho)w_\phi(\phi)$, so that the Helmholtz equation can be cast as

$$\begin{cases} -\frac{\mathrm{d}^2 w_\rho}{\mathrm{d}\rho^2}(\rho)w_\phi(\phi) - \frac{1}{\rho}\frac{\mathrm{d}w_\rho}{\mathrm{d}\rho}(\rho)w_\phi(\phi) - \frac{1}{\rho^2}w_\rho(\rho)\frac{\mathrm{d}^2 w_\phi}{\mathrm{d}\phi^2}(\phi) = z w_\rho(\rho)w_\phi(\phi) \quad \text{for } (\rho,\phi) \in \Omega'(\theta^{(1)}), \\ |w_\rho(0)| < \infty, \ w_\rho(1) = 0, \\ w_\phi(\theta^{(1)}) = w_\phi(2\pi - \theta^{(1)}) = 0. \end{cases}$$

Employing the usual separability arguments, see, e.g., [Sal+13, Section 8.5.1], we conclude that, for a non-trivial solution to exist, there must exist $\alpha^2 \in \mathbb{C}$ such that

$$\begin{cases} -\frac{1}{\rho^2}\frac{\mathrm{d}^2 w_\phi}{\mathrm{d}\phi^2}(\phi) = \alpha^2 w_\phi(\phi) & \text{for } \phi \in]\theta^{(1)}, 2\pi - \theta^{(1)}[, \\ w_\phi(\theta^{(1)}) = w_\phi(2\pi - \theta^{(1)}) = 0, \end{cases}$$

and

$$\begin{cases} -\frac{\mathrm{d}^2 w_\rho}{\mathrm{d}\rho^2}(\rho) - \frac{1}{\rho}\frac{\mathrm{d}w_\rho}{\mathrm{d}\rho}(\rho) - \left(\frac{1}{\rho^2} - z\alpha^2\right) w_\rho(\rho) = 0 \quad \text{for } \rho \in]0, 1[, \\ |w_\rho(0)| < \infty, \ w_\rho(1) = 0. \end{cases}$$

The former problem admits a non-trivial solution iff $\frac{\alpha}{2\pi - 2\theta^{(1)}}$ is a non-zero integer, in which case the solution is

$$w_\phi(\phi) = A \sin\left(\pi\alpha\left(\phi - \theta^{(1)}\right)\right) \qquad \forall A \in \mathbb{C}.$$

On the other hand, the latter problem is solved (non-trivially) by multiples of the Bessel function of the first kind $\rho \mapsto J_\alpha(\sqrt{z}\rho)$, under the (boundary) condition that $J_\alpha(\sqrt{z}) = 0$.

Putting everything together, a non-trivial solution to the homogeneous problem is allowed iff

$$\sqrt{z} \text{ is a root of } J_\alpha \text{ for some } \alpha \in \mathbb{R} \text{ such that } \frac{\alpha}{2\pi - 2\theta^{(1)}} \in \mathbb{Z} \setminus \{0\}. \tag{7.6}$$

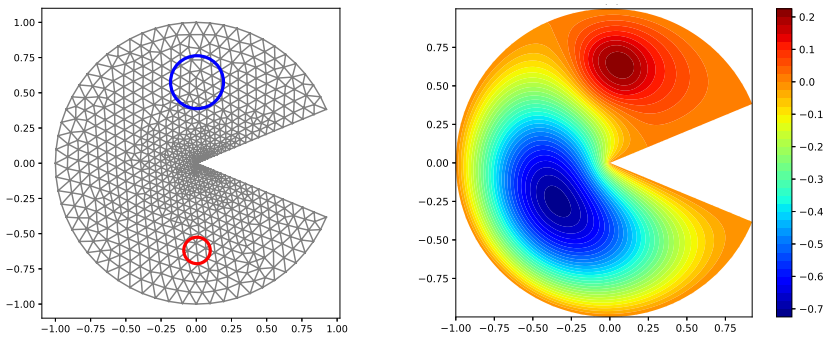


Figure 7.1 – On the left, mesh used to solve the PAC-MAN-like problem for $\theta^{(1)} = \frac{3\pi}{24}$, coarsened by a factor 4 to allow to distinguish the elements by eye. We also superimpose the boundaries of the supports of $u(0.2)$ and $g(0.1)$, in blue and red, respectively. On the right, the solution corresponding to $(z, \theta) = (12, \frac{3\pi}{24}, 0.2, 0.1)$. The value of y is approximately -0.466.

Since the roots of Bessel functions can be approximated to arbitrary precision, this provides an (almost) explicit definition of Λ. Note that, by self-adjointness of the Laplacian operator, the corresponding eigenspaces are orthogonal with respect to both the $L^2(\Omega)$ and $H^1(\Omega)$ inner products. Hence, we can automatically deduce $\mathcal{V}$-orthogonality.

In order to approximate numerically the Helmholtz problem (7.5), we discretize Ω by introducing a mesh $\mathcal{T}$, i.e., a collection of (closed) triangles mutually overlapping at most over edges. Then, we perform a Galerkin projection of the problem onto the finite element (FE) space

$$\mathcal{V}_\mathcal{T} = \left\{ v \in H^1_0(\Omega_\mathcal{T}) : v|_T \text{ is an affine function for all triangles } T \in \mathcal{T} \right\},$$

where the computational domain $\Omega_\mathcal{T}$ is the interior of the union of all triangles in $\mathcal{T}$. Note that:

- due to the curved boundary, $\Omega_\mathcal{T} \neq \Omega$;
- since Ω has a concave angle at $(0, 0)$ (for $0 < \theta^{(1)} < \frac{\pi}{2}$), the triangulation should be properly refined around 0 to account for potential local irregularities in the solution, e.g., unboundedness of its gradient;
- for a fine enough mesh, the spectrum of the discretized problem provides a faithful approximation of the smaller elements of Λ;
- $\mathcal{T}$ and $\mathcal{V}_\mathcal{T}$ depend on $\theta^{(1)}$.

In our simulations, we set the mesh size (namely, the largest side length of a triangle) to be approximately 0.02 near the boundary of $\partial B^{\mathbf{0}}(1)$ and approximately 0.003 near the center $(0, 0)$.

We remark that the triangulation $\mathcal{T}$ is not conforming to the support of the forcing term and of the sensor. This leads to additional (but minor) errors due to the non-resolved sub-mesh scale at the boundary of such supports. It is interesting to note that, from this point of view, we could treat the FE snapshots as affected by a (z, θ)-dependent noise.

From a MOR point of view, we observe that problem (7.5) does not depend on θ in an affine$^{\text{MOR}}$ way. In particular, the parametric spatial domain could be mapped onto a reference (parameter-independent) domain, e.g., moving to scaled polar coordinates. However, this, on one hand, forces

to use a parameter-independent triangulation of the domain, whereas, in practice, it might make sense to refine the mesh more around $(0,0)$ for smaller angles $\theta^{(1)}$. On the other hand, this does not solve the issue of non-affine$^{\text{MOR}}$ dependence on the other parameters $\theta^{(2)}$ and $\theta^{(3)}$. In fact, it makes the matter worse, since the expressions of u and g on the reference spatial domain become $\theta^{(1)}$-dependent.

Note that, had the spatial domain been more complicated, identifying the mapping to the reference domain would have been rather difficult. To this aim, methods as those presented in [JIR14; LR10] could have been employed, at the cost of higher cost, both offline and online. Notably, the increase in online time is due to the higher number of terms that become necessary to obtain an affine$^{\text{MOR}}$ approximation of the mapping.

7.1.1 Non-parametric setting: changing the metric of the Hilbert space

We start by fixing $\theta = \bar{\theta} = (\frac{3\pi}{24}, 0.2, 0.1)$, and by building a surrogate of the state v with respect to z over the frequency range $A = [5, 75]$. Note that the frequency range contains 11 elements of $\Lambda(\bar{\theta}^{(1)})$. To build the surrogate, we apply barycentric MRI with adaptive sampling, using the relative look-ahead estimator from Section 5.3.3 (cf. also Section 5.5.3) with a tolerance $\epsilon = 10^{-2}$, and 10^3 uniformly spaced test frequencies. Note that the problem is affine in z, so that the residual behavior is captured exactly by Lemma 5.1.

Once the surrogate $\widetilde{v}$ has been built, we can derive from it a surrogate $\widetilde{y}$ for the output through

$$\widetilde{y}(z, \bar{\theta}) = \langle \widetilde{v}(z, \bar{\theta}), g(\bar{\theta}^{(3)}) \rangle_{L^2(\Omega(\bar{\theta}^{(1)}))}. \tag{7.7}$$

Since $\widetilde{v}$ is a linear combination of the snapshots, see Definition 3.2, and $\langle \cdot, g(\bar{\theta}^{(3)}) \rangle_{L^2(\Omega(\bar{\theta}^{(1)}))}$ is a linear operation, $\widetilde{y}$ can be made online-efficient:

$$\widetilde{y}(z, \bar{\theta}) = \langle \sum_{j=1}^{S} \widetilde{\alpha}_j(z) v(z_j, \bar{\theta}), g(\bar{\theta}^{(3)}) \rangle_{L^2(\Omega(\bar{\theta}^{(1)}))} = \sum_{j=1}^{S} \widetilde{\alpha}_j(z) \underbrace{\langle v(z_j, \bar{\theta}), g(\bar{\theta}^{(3)}) \rangle_{L^2(\Omega(\bar{\theta}^{(1)}))}}_{y(z_j, \bar{\theta})},$$

with $\{\widetilde{\alpha}_j(z)\}_{j=1}^{S}$ being the coefficients of the expansion of $\widetilde{v}$ onto the snapshot basis. Note that such coefficients are explicitly available scalar rational functions of z.

The initial training set contains just the two extreme frequencies $\{5, 75\}$. The algorithm terminates at the 16-th iteration, yielding (after the post-processing described in Section 5.3.3) a rational surrogate of type [16/16]. We show the resulting surrogate state v and output y in Figure 7.2. Some validation points are also included, showing a good approximation quality.

We compare this approach with two alternatives:

- z-adaptive MRI with the same parameters, but using the polynomial version with the Legendre basis rather than the barycentric one;
- z-weak-greedy RB based on the relative residual estimator

$$\eta(z) = \frac{\left\| (\Delta + z)\widetilde{v}(z, \bar{\theta}) + u(\bar{\theta}^{(2)}) \right\|_{\mathcal{V}'}}{\left\| u(\bar{\theta}^{(2)}) \right\|_{\mathcal{V}'}}$$

  (where we compute the numerator using (2.45)), with the same parameters as before.

Due to the different estimator, RB requires one extra iteration to reach convergence.

We note that the sample points are located differently in the three cases, see Figure 7.3. This is rather interesting if we consider that all three approaches select the next sample point using the same strategy, i.e., by choosing the test point with the maximal residual, cf. Lemma 5.1. This means that the only reasons for the different "sampling histories" of the methods are the (minor) differences in the respective surrogates at each iteration, which get amplified as the algorithm progresses.

The surrogate states and outputs are visually indistinguishable, so we compare the different surrogates by their relative approximation error instead. In Figure 7.4, we show the errors in both state and output. We can see that the results are quite similar, and all the methods achieve the required accuracy in the state relative error. RB seems to have a slight edge on MRI, especially for larger frequencies. However, this is just due to the extra snapshot taken. If the RB method had been stopped one iteration in advance, we would recover an error similar to the two MRI approaches.

Concerning the error, note that, *a priori*, MRI is not guaranteed to attain the prescribed tolerance, due to the partly heuristic nature of the estimator, see Section 5.3.3. On the other hand, RB is not guaranteed to achieve an error below the tolerance either, since it employs the residual as estimator, and it is not straightforward to relate error and residual. Indeed, due to the presence of the resonances, the problem is not uniformly inf-sup stable, so that error and residual are not equivalent near the poles of v.

We plot in Figure 7.5 the singular value decay of the "normalized snapshot Gramian", i.e., the Gramian matrix containing the pairwise inner products between normalized snapshots, which can be obtained from the standard snapshot Gramian (2.44) as

$$G_0 = d_G^{-1/2} G d_G^{-1/2}, \quad \text{with } d_G = \operatorname{diag}((G)_{11}, (G)_{22}, \dots, (G)_{SS}).$$

We can observe a plateau that lasts for approximately 11 singular values, exactly as many as the number of poles of v in A. This is in agreement with the observations in [RM18], since all the relevant residues are, in some sense, equally important for a uniformly good approximation over the frequency range. We can actually try to make this observation quantitative, by introducing

$$\zeta_j = \left(\frac{\operatorname{Cap}(A)}{\Phi_A(\lambda_j)} \right)^S, \qquad j = 1, 2, \dots, \tag{7.8}$$

with Φ_A the Green's potential of the frequency range and $\Lambda = \{\lambda_j\}_{j=1}^{\infty}$ being the pole ordering induced by Φ_A, cf. Assumption 3.5. The intuitive motivation behind the definition of ζ_j is the error bound appearing in Theorem 3.7, where ζ_j encodes the dependence on S of the bound over the frequency range. (In theory, we should use Theorem 3.8 rather than Theorem 3.7 since we are increasing S and N together, but, unfortunately, the former theorem does not give any upper bound on the error.) We can see in Figure 7.5 that (7.8) seems to behave quite similarly to the singular values. Notably, $\zeta_1 = \ldots = \zeta_{11} = 1$ by definition, in agreement with the already-observed plateau.

It is quite interesting to note that these observations justify a (heuristic but not-so-intrusive) way of estimating the number of poles in A by looking at the "width" of the plateau in the singular values of G_0. Before proceeding, we wish to remark that, if we had used the singular values of the non-normalized snapshot Gramian, we would not have observed a behavior as "clean", see

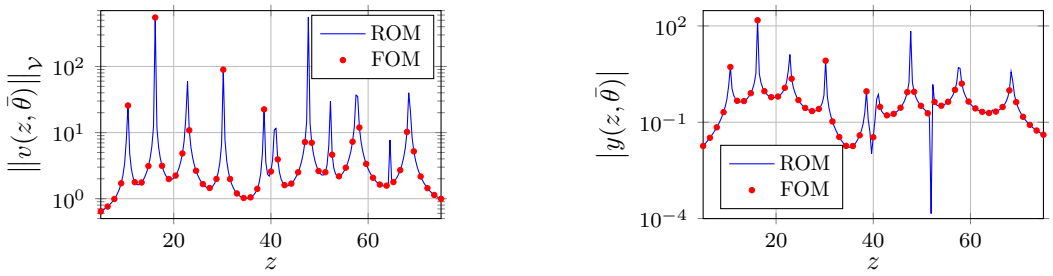


Figure 7.2 – Surrogates for norm of the state (left) and magnitude of the output (right) for the non-parametric PAC-MAN-like problem. Red dots are validation points, obtained by evaluation of the FOM.

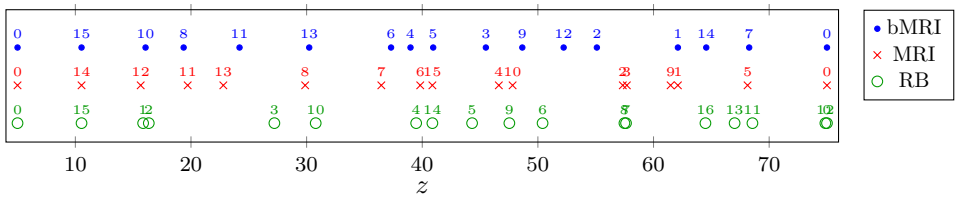


Figure 7.3 – Locations of sample points for barycentric MRI (top), MRI with Legendre basis (middle), and weak-greedy RB (bottom), for the non-parametric PAC-MAN-like problem. The labels indicate on which iteration each point was added, starting from 0.

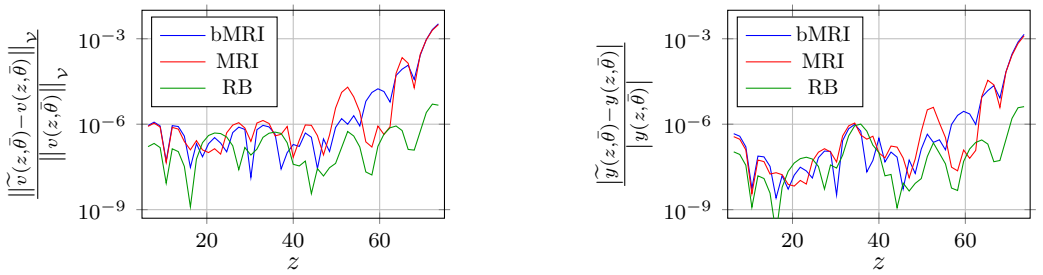


Figure 7.4 – Norm of the relative error in the approximation of the state (left) and output (right) for the non-parametric PAC-MAN-like problem.

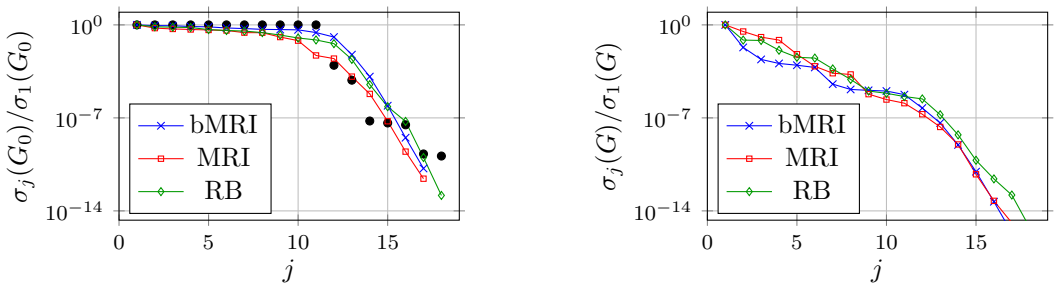


Figure 7.5 – Singular values of the normalized (left) and non-normalized (right) snapshot Gramian for the non-parametric PAC-MAN-like problem in the $H_0^1(\Omega)$ metric. The black dots in the left plot are values of (7.8).

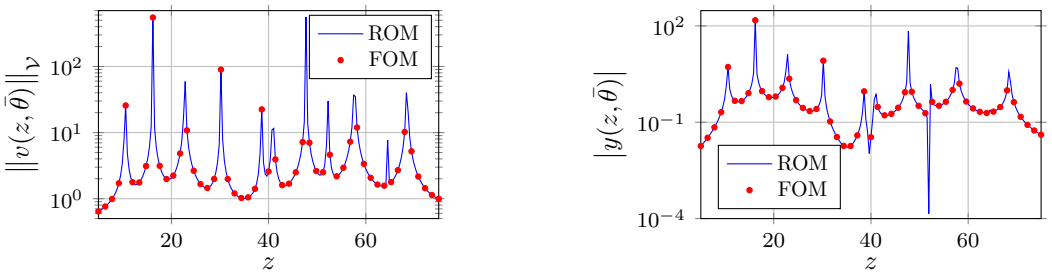


Figure 7.6 – Surrogates for norm of the state (left) and magnitude of the output (right) for the non-parametric PAC-MAN-like problem in the modified metric. Red dots are validation points, obtained by evaluation of the FOM.

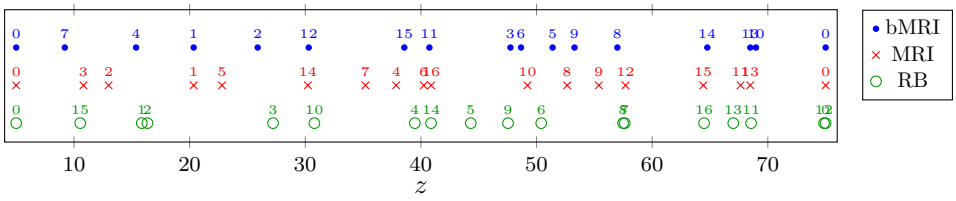


Figure 7.7 – Locations of sample points for barycentric MRI (top), MRI (middle), and weak-greedy RB (bottom), for the non-parametric PAC-MAN-like problem in the modified metric. The labels indicate on which iteration each point was added, starting from 0.

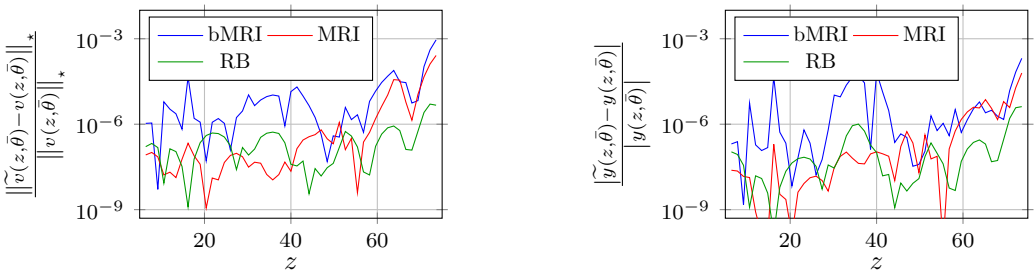


Figure 7.8 – Norm of the relative error in the approximation of the state (left) and output (right) for the non-parametric PAC-MAN-like problem in the modified metric.

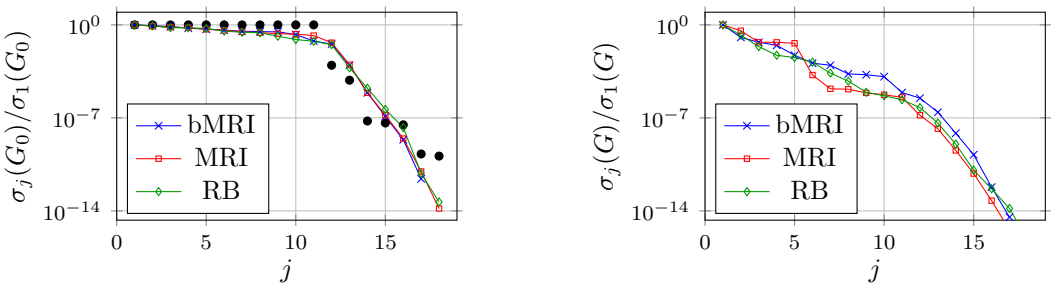


Figure 7.9 – Singular values of the normalized (left) and non-normalized (right) snapshot Gramian for the non-parametric PAC-MAN-like problem in the modified metric. The black dots in the left plot are values of (7.8).

the right plot in Figure 7.5. Indeed, in the snapshot Gramian, each pole λ_j inside the frequency range is given a "relevance" that is inversely proportional to the distance (squared) between λ_j and the closest sample point. Since such "relevance" is j-dependent, the "plateau" turns out not to be very flat.

Now, we wish to investigate whether such satisfactory results rely heavily on the orthogonality of the residues of v. To this aim, we replace the $H_0^1(\Omega_\mathcal{T})$ inner product with a different one on $\mathcal{V}_\mathcal{T}$, namely a weighted ℓ^2 one:

$$\langle v, w \rangle_\star = \sum_{j=1}^{\dim(\mathcal{V}_\mathcal{T})} \gamma_j v(x_j)\overline{w(x_j)},$$

where $\{\gamma_j\}_j$ are independent samples drawn from a uniform distribution on $[0, 1]$ and $\{x_j\}_j \subset \overline{\Omega_\mathcal{T}}$ are the vertices of $\mathcal{T}$. The inner product on the dual space $\mathcal{V}^\star$ is simply obtained by taking the reciprocals of the weights.

We repeat our experiment by employing this metric. This, among other effects, will induce changes in the snapshot Gramians used in the methods. Notably, RB is not too affected by the change of metric. Indeed, thanks to Lemma 5.1, the residual estimator at each iteration remains unchanged, except for the possible multiplicative scaling by a constant. The only change is in the snapshot orthonormalization step, which does not have any effect on the surrogate, apart from a minor one in terms of numerical stability. We confirm this empirically, by noting that the results by RB are the same as with the original metric.

As before, barycentric MRI terminates at the 16-th snapshot. However, this time, MRI with Legendre basis requires as many snapshots as RB (thus yielding a slightly lower approximation error than barycentric MRI). This being said, by looking at Figures 7.6 to 7.9, no significant change can be observed with respect to the orthogonal case, except for some slight increase in the surrogate error. Still, the error stays uniformly below the prescribed tolerance also here. In particular, note that the plateau in the singular value decay of the normalized snapshot Gramian can be observed even in this modified norm.

7.1.2 Parametric setting: adventures in adaptive sampling

Next, we allow the two non-geometric parameters $\theta^{(2)}$ and $\theta^{(3)}$ to vary in the range $[0.1, 0.3]^2$. Note that, since Λ does not depend on those parameters, the poles of v and y are constant. As such, the only variations of v and y over $(\theta^{(2)}, \theta^{(3)})$ are due to changes in the residues. Note, however, that our pMOR approach, in true non-intrusive fashion, will not be aware of this property.

In order to approximate this problem, we apply the pole/residue-matching approach with adaptive parameter sampling, where each local surrogate is built as above via greedy barycentric MRI. In particular, we apply MRI to the system state v, but then we match and interpolate poles and residues of the surrogate system output (7.7). We interpolate over $(\theta^{(2)}, \theta^{(3)})$ via piecewise-linear hat functions. As initial parameter samples, we set $\widetilde{\Theta} = \{(0.2, 0.2)\}$, and we fix the adaptive parameter sampling tolerance $\epsilon = 35 \cdot 10^{-2}$, see Algorithm 6. Note that 35 is a scaling factor that captures the (half-)length of the frequency range, allowing us to "normalize" the partial fraction distance (6.27) in some sense. Due to the adaptivity in frequency, the surrogates might (and do) have different sizes. We choose to deal with this by throwing away any pole that remains unmatched, which, using the notation of Algorithm 6, corresponds to setting $\delta = 0$. Note that, due to the poles being constant, this parameter does not affect much the results.

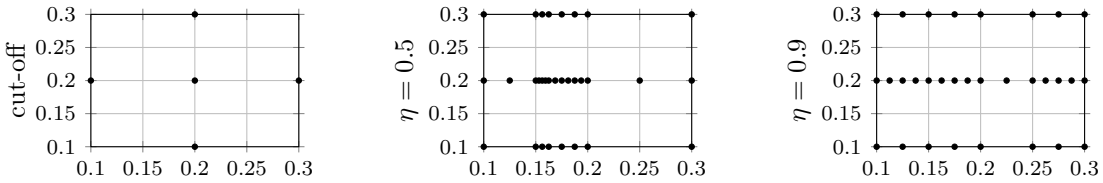


Figure 7.10 – Parameter sample points for the three adaptive methods, applied to the PAC-MAN-like problem with two parameters. All plots have $\theta^{(2)}$ on the x-axis and $\theta^{(3)}$ on the y-axis.

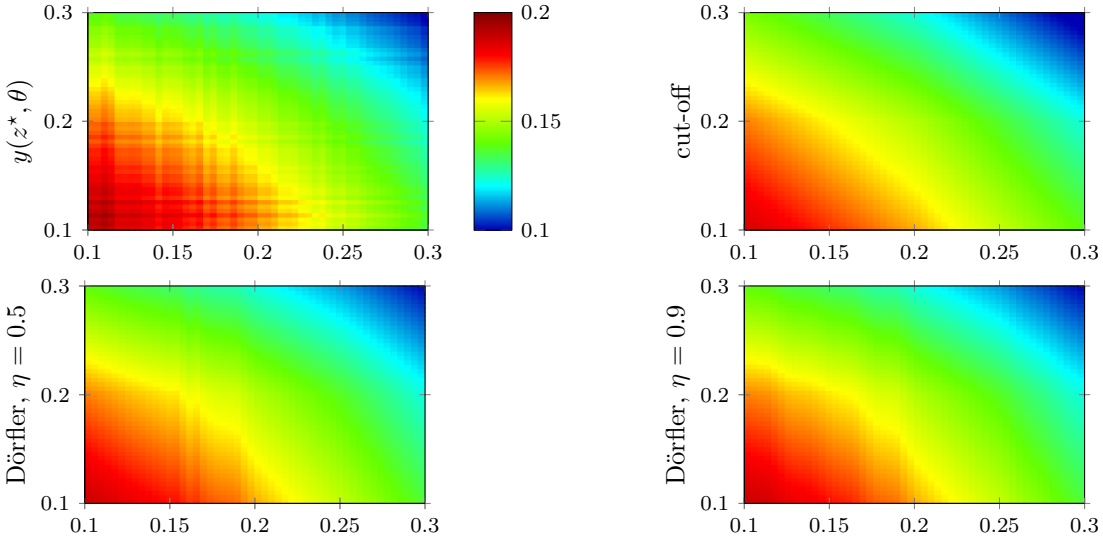


Figure 7.11 – Qualitative results for the PAC-MAN-like problem with two parameters. Plot of exact $y(z^\star, \theta)$ in the top left plot. In the other plots, we show the surrogates $\widetilde{y}(z^\star, \theta)$. All plots have the same color scale, reported next to the top left plot. All plots have $\theta^{(2)}$ on the x-axis and $\theta^{(3)}$ on the y-axis.

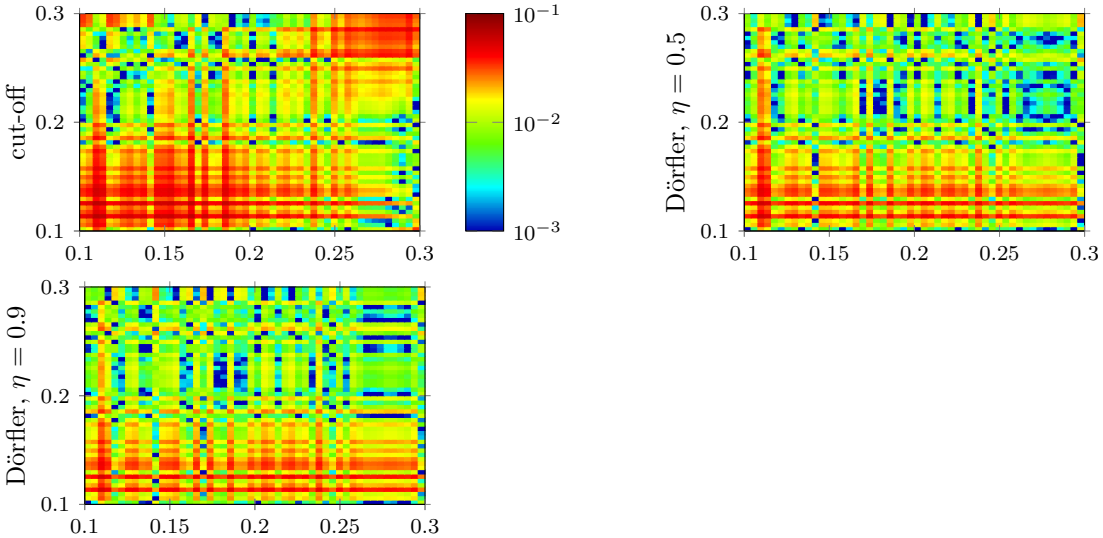


Figure 7.12 – Relative error in the approximation of y at $z = z^\star$ for the PAC-MAN-like problem with two parameters. All errors have the same color scale, reported next to the top left plot. All plots have $\theta^{(2)}$ on the x-axis and $\theta^{(3)}$ on the y-axis.

Unfortunately, the algorithm does not converge, due to the error at all test points with $\theta^{(3)} = 0.2$ never going below the tolerance. In particular, by direct inspection of the local surrogates, we see that this is due to very large distances between poles outside the frequency range. For instance, both surrogates at $(\theta^{(2)}, \theta^{(3)}) \in \{(0.2, 0.2), (0.1, 0.2)\}$ identify very well the 11 poles of v inside the frequency range, but they also have some extra ones: the former has a pole at $z \approx 233$, while the latter at $z \approx 331$. At the first iteration, these faraway poles are matched one to the other, yielding a large partial fraction distance between surrogate and truth models.

To solve this issue, we try to apply some of the ideas from Section 6.4.2.2, namely, introducing a cut-off in the poles and using a Dörfler-inspired approach. We start from the cut-off method. Removing all poles outside the range $[-3.75, 83.75]$ (corresponding to a Green's potential of $2\,\mathrm{Cap}(A) = 35$) still results in a non-converging algorithm, whereas removing the poles outside $[2.1, 77.9]$ (corresponding to a Green's potential of $1.5\,\mathrm{Cap}(A) = 26.25$) converges in a single iteration with only 5 total snapshots, possibly an indication of a too early termination. Indeed, in our experience, we find it useful to start from a slightly larger initial sample set $\widetilde{\Theta}$ whenever choosing an aggressively low cut-off level.

Concerning the Dörfler approach, we set $\eta \in \{0.5, 0.9\}$, while limiting the maximum number of θ-samples to 30, which, considering the results of the previous section, correspond to approximately 500 total snapshots. Since we do not enforce this computational budget too strictly (even if we go over the budget, we still allow the current iteration to continue), the two approaches effectively employ slightly more than 30 parameter samples. More specifically, 31 parameter samples are taken with both $\eta = 0.5$ and $\eta = 0.9$.

We show in Figure 7.10 the locations of the sample points for the three successful algorithms (i.e., cut-off with lower tolerance and the two Dörfler ones). In the Dörfler cases, we can observe the horizontal local refinements at $\theta^{(3)} = 0.2$ due to the above-mentioned issue of faraway poles. Moreover, we verify that larger values of η correspond to more uniform refinements.

In Figure 7.11, we compare the methods based on their approximation of y at the randomly selected frequency $z^\star = 31.2178$. We use a linear color scale since we do not expect any resonant behavior with respect to $(\theta^{(2)}, \theta^{(3)})$. We note that, thanks to the online efficiency of the surrogates, obtaining the values for the top left plot took approximately 500 times longer than obtaining those for the other three plots *combined*. Moreover, we remark that the "jagged" look of the exact output is due to the fact that the triangulation is not conforming to the support of the forcing term and of the sensor, as mentioned above. As a consequence, y is not necessarily smooth with respect to $\theta^{(2)}$ and $\theta^{(3)}$. In contrast, the surrogates, being based on samples at just few values of θ, are smoother.

In Figure 7.12, we provide quantitative information on the approximation error. Out of all methods, the Dörfler approaches seems to perform best. This is reasonable, since, after all, they rely on more snapshots than the other ROM. Note that the error cannot be expected to be too small due to (i) the numerical noise due to the above-mentioned non-conformity of the mesh, (ii) the error in the local surrogates, and (iii) the piecewise-linear interpolation of smooth poles and residues.

7.1.3 Parametric setting: non-affine$^{\mathbf{MOR}}$ parametrization of geometry

Finally, we add also the remaining parameter $\theta^{(1)}$, which we allow to vary within the range $[\frac{\pi}{12}, \frac{\pi}{6}]$. Now the spectrum changes as well, but it turns out that the relevant poles do not intersect over

the chosen parameter range, cf. the analytic formula in Section 7.1 and Figure 7.14. Still, the poles might leave the frequency range as $\theta^{(1)}$ increases.

We apply the same three approaches as before. However, for the Dörfler cases, we increase the computational budget to 50 parameter samples, i.e., approximately 10^3 snapshots overall. Note that, since the triangulation $\mathcal{T}$ is $\theta^{(1)}$-dependent, the system state v lives in different spaces for different values of $\theta^{(1)}$. This has two important consequences:

- Every time we take a snapshot at a new value of $\theta^{(1)}$, it is necessary to create a new mesh and project the PDE onto the discrete space from scratch.
- Snapshots of the state v at different values of $\theta^{(1)}$ are incompatible. Thus, matching and interpolating the residues of (the surrogate of) y, instead of those of v, as we did in the previous section, becomes a necessity rather than a choice.

We display in Figure 7.13 the sample points that are employed in each approach. In both the cut-off case and the Dörfler one with $\eta = 0.5$, the algorithm terminates at 53 snapshots, whereas 65 snapshots are necessary for the Dörfler case with $\eta = 0.9$.

Now we move to the quality of the surrogate. Since it is impossible to make plots in > 3D, we restrict θ to the main diagonal of Θ, i.e.,

$$\theta = \Big(\frac{\pi}{12} + \frac{\pi}{12}\alpha, 0.1 + 0.2\alpha, 0.1 + 0.2\alpha\Big) \quad \text{with } \alpha \in [0, 1].$$

We show in Figure 7.14 the surrogate poles obtained with the three methods. We can observe that most of the exact poles are identified well. Notably, some mismatches happen for the two largest relevant poles in the Dörfler cases. This is due to a poor choice of the relative weights of pole distance and residue distance in the matching optimization problem (6.11): had less importance been given to the residues, the matching would have been correct. Similarly, another way of fixing the incorrect matches is increasing the number of samples of θ, since, in the limit of many samples, we expect a correct matching of poles and residues whenever the poles do not cross.

In Figure 7.15, we show the output $y(z, \theta)$ and we compare it with its surrogates. The corresponding approximation error is plotted in Figure 7.16. We observe that the approximation quality is rather good throughout most of the frequency and parameter ranges. However, the error increases for larger frequencies due to poles slightly larger than 75 that are not approximated so well by the ROM. Moreover, in the Dörfler cases, we note a larger error near the pole mismatches. That being said, thanks to the local support of the θ-interpolation basis, the inaccuracy is localized both in z and θ.

In all cases, the online speedup is approximately 600. Note that we are including the creation of the triangulation and the assembly of the FE system in measuring the FOM solution time.

7.2 Harmonic-elastic deformation of a tuning fork

We move to a 3D problem in linear elasticity, which could be considered as a simple example of frequency response of a mechanical structure under material uncertainties.

Let $\Omega \subset \mathbb{R}^3$ be a region of space occupied by a tuning fork, see Figure 7.17. For reference, the

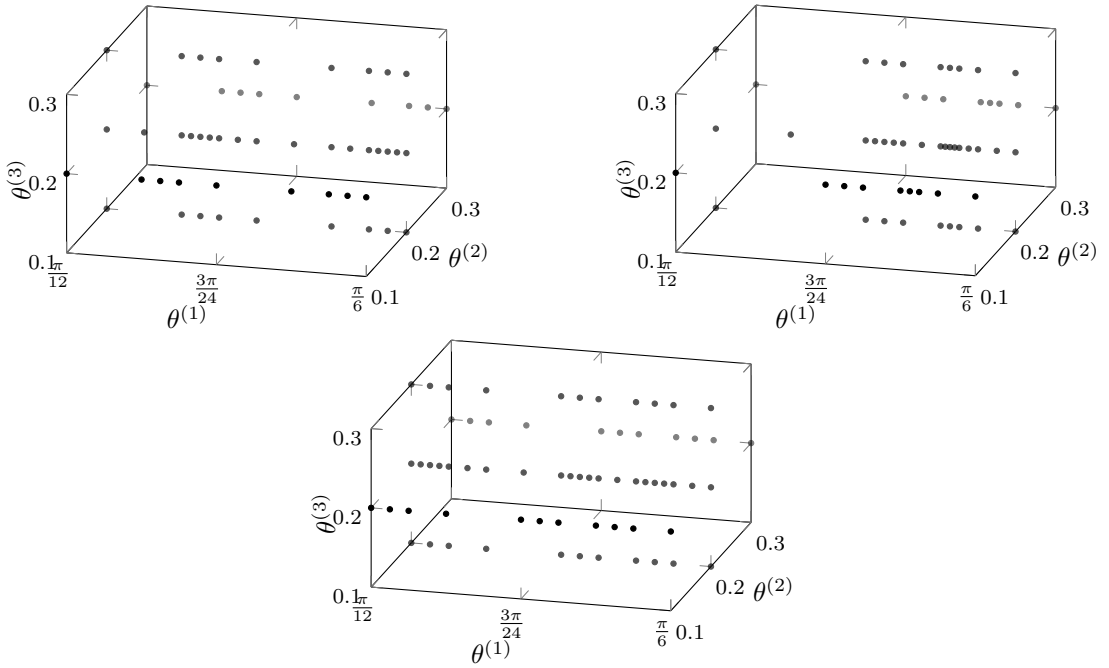


Figure 7.13 – Parameter sample points obtained with the cut-off approach (top left), and with the Dörfler approach with $\eta = 0.5$ (top right) and $\eta = 0.9$ (bottom), for the PAC-MAN-like problem with three parameters.

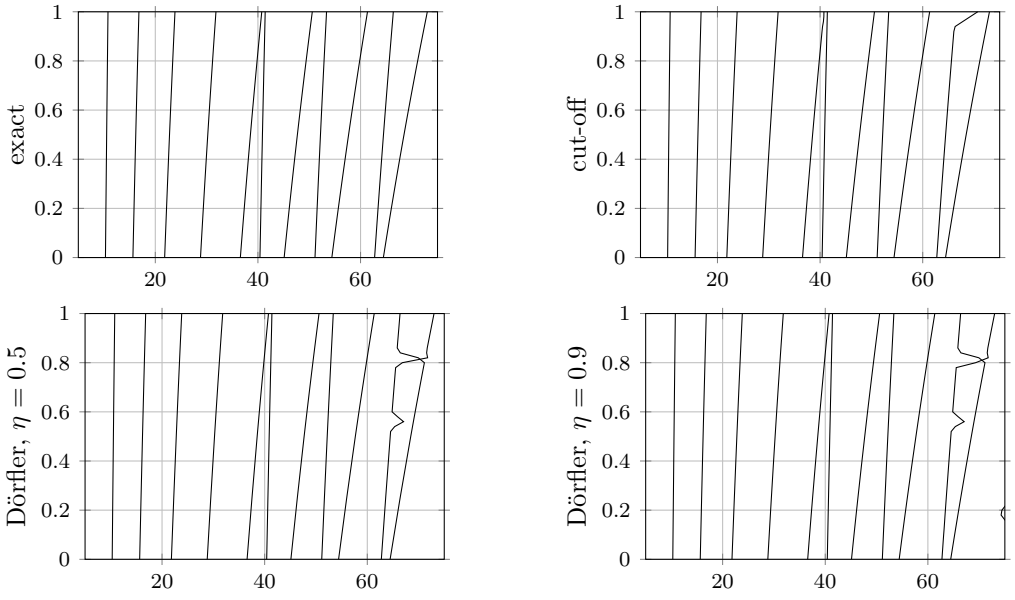


Figure 7.14 – Exact poles of y over the main diagonal of Θ (top left) for the PAC-MAN-like problem with three parameters. The other plots contain the surrogate poles for the same values of θ. All plots have z on the x-axis and α on the y-axis, with $\alpha = 0$ and $\alpha = 1$ corresponding to two vertices of Θ.

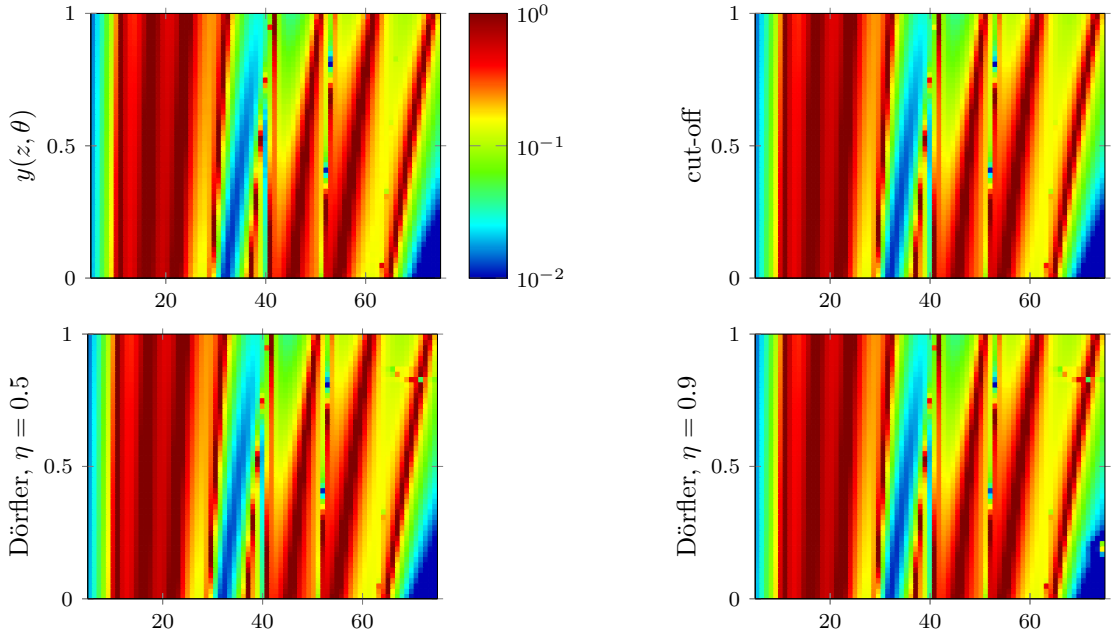


Figure 7.15 – Qualitative results for the PAC-MAN-like problem with three parameters. Plot of exact $|y(z,\theta)|$ in the top left plot. In the other plots, we show the surrogates $|\widetilde{y}(z,\theta)|$. We vary θ over the main diagonal of Θ. All plots have the same color scale, reported next to the top left plot. All plots have z on the x-axis and α on the y-axis, with $\alpha = 0$ and $\alpha = 1$ corresponding to two vertices of Θ.

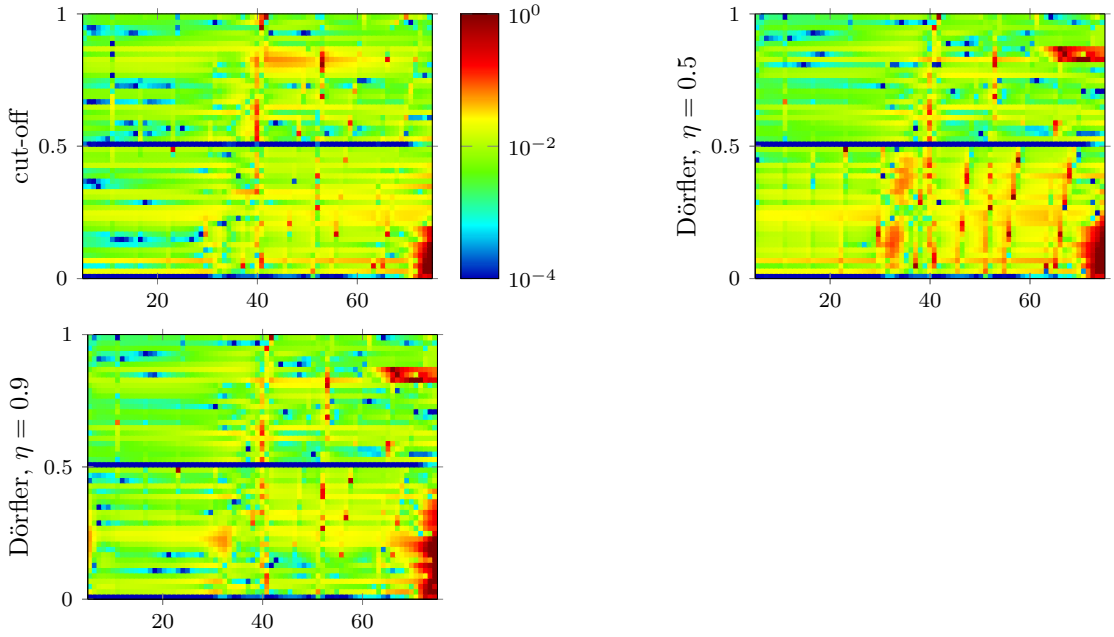


Figure 7.16 – Relative error in the approximation of y over the main diagonal of Θ for the PAC-MAN-like problem with three parameters. All errors have the same color scale, reported next to the top left plot. All plots have z on the x-axis and α on the y-axis, with $\alpha = 0$ and $\alpha = 1$ corresponding to two vertices of Θ.

total length of the tuning fork is approximately 11cm (see [Pra21] for more details on the spatial domain). We partition Ω into three parts $\Omega = \Omega_{\text{pommel}} \sqcup \Omega_{\text{handle}} \sqcup \Omega_{\text{fork}}$, with Ω_{pommel} being the spherical pommel and Ω_{handle} being the remainder of the handle, up to where the fork splits. We partition the boundary into $\partial\Omega = \Gamma_D \sqcup \Gamma_N$, with $\Gamma_D = \partial\Omega_{\text{pommel}} \cap \partial\Omega$.

We assume the tuning fork to be made of two linearly elastic materials (one for pommel and handle, another for the fork) with uniform properties. More precisely, we define piecewise-constant fields for the Young's modulus, the Poisson's ratio, and the material density:

$$E(\theta^{(1)};x) = \begin{cases} E_0 & \text{if } x \in \Omega_{\text{pommel}} \cup \Omega_{\text{handle}}, \\ \theta^{(1)} & \text{if } x \in \Omega_{\text{fork}}, \end{cases} \quad \nu(\theta^{(2)};x) = \begin{cases} \nu_0 & \text{if } x \in \Omega_{\text{pommel}} \cup \Omega_{\text{handle}}, \\ \theta^{(2)} & \text{if } x \in \Omega_{\text{fork}}, \end{cases}$$

$$\text{and } \rho(\theta^{(3)};x) = \begin{cases} \rho_0 & \text{if } x \in \Omega_{\text{pommel}} \cup \Omega_{\text{handle}}, \\ \theta^{(3)} & \text{if } x \in \Omega_{\text{fork}}. \end{cases}$$

We fix $E_0 = 200\text{GPa}$, $\nu_0 = 0.25$, $\rho_0 = 8\text{kg/dm}^3$, corresponding to steel, whereas we model $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)})$ as uniform random variables, taking values in $\Theta = [180, 220]\text{GPa} \times [0.24, 0.26] \times [7.8, 8.2]\text{kg/dm}^3$. We denote their mean value by $\theta_0 = (E_0, \nu_0, \rho_0)$.

We apply a time-harmonic pressure pulse to the top portion of the boundary of one of the two teeth of the tuning fork. The specific expression at the complex (angular) frequency $\mathrm{i}\omega$ is $\widehat{f}(t;x) = f(\mathrm{i}\omega;x)e^{-\mathrm{i}\omega t} \in \mathbb{R}^3$, with

$$f(\mathrm{i}\omega;x) = \frac{T}{2\pi L^2} \exp\left(-\frac{|(x-x_0) - ((x-x_0)\cdot\widetilde{x})\widetilde{x}|^2}{2L^2} + \mathrm{i}\frac{\omega}{c}((x-x_0)\cdot\widetilde{x}) \right) \widetilde{x}.$$

In the expression above, $T = 1\text{kN}$ is the total force applied, $L = 2\text{cm}$ is the pulse width, $x_0 = (-3.1, 10.6, 0)\text{mm}$ is the origin of the pulse, $\widetilde{x} = (\cos(20°)\cos(10°), \sin(10°), \sin(20°)\cos(10°))$ is the direction of the pulse, and $c = 300\text{m/s}$ is the speed of the pulse in the air. Note that f is space-harmonic in the direction $\widetilde{x}$ and Gaussian in any plane orthogonal to $\widetilde{x}$.

By linearity, the long-term behavior of the tuning fork can be analyzed by solving the frequency-domain problem

$$\begin{cases} \frac{1}{\rho(\theta^{(3)})} \operatorname{div} \sigma\left(v(z,\theta),\theta\right) + 2\pi\mathrm{i}\widetilde{\eta} z v(z,\theta) + 4\pi^2 z^2 v(z,\theta) = \mathbf{0} & \text{in } \mathring{\Omega}_{\text{pommel}} \cup \mathring{\Omega}_{\text{handle}} \cup \mathring{\Omega}_{\text{fork}}, \\ v(z,\theta) = \mathbf{0} & \text{on } \Gamma_D, \\ \sigma(v(z,\theta),\theta)\widetilde{\nu} = f(2\pi\mathrm{i}z) & \text{on } \Gamma_N, \end{cases} \tag{7.9}$$

with the additional constraint that $v(z,\theta)$ and the normal component of $\sigma(v(z,\theta),\theta)$ must be continuous across the interfaces between subdomains, i.e., $\overline{\Omega}_{\text{pommel}} \cap \overline{\Omega}_{\text{handle}}$ and $\overline{\Omega}_{\text{handle}} \cap \overline{\Omega}_{\text{fork}}$. In (7.9), we have defined the following quantities:

- The displacement field $v(z,\theta) \in \mathcal{V} = [H^1_{\Gamma_D}(\Omega)]^3$, where $H^1_{\Gamma_D}(\Omega) = \left\{ v \in H^1(\Omega) : v|_{\Gamma_D} = 0 \right\}$. We endow $\mathcal{V}$ with the elastic energy metric, whose inner product reads

$$\langle v, w \rangle_{\mathcal{V}} = \sum_{i,i'=1}^{3} \langle \left(\sigma\left(v,\theta_0\right)\right)_{ii'}, \left(\operatorname{grad} w\right)_{ii'} \rangle_{L^2(\Omega)} = \langle \sigma\left(v,\theta_0\right), \operatorname{grad} w \rangle_{[L^2(\Omega)]^{3\times 3}}.$$

  Note that we are using the mean value of θ to make the metric parameter-independent.

- The stress matrix $\sigma \in \mathbb{C}^{3\times 3}$, whose (linear) constitutive relation is

$$\sigma(v,\theta) = \frac{E(\theta^{(1)})}{1+\nu(\theta^{(2)})}\left(\frac{\operatorname{grad} v + (\operatorname{grad} v)^\top}{2} + \frac{\nu(\theta^{(2)})}{(1-2\nu(\theta^{(2)}))}(\operatorname{div} v)I\right).$$

- The linear frequency $z = \omega/(2\pi) \in A = [50, 1000]\text{Hz}$.

- The Rayleigh (mass) damping coefficient $\widetilde{\eta} = 100\text{Hz}$.

- The outer unit normal vector $\widetilde{\nu}$ to $\partial\Omega$.

As quantity of interest (QoI), we take either of the following two items:

- The maximum displacement magnitude $V_{\max} \in \mathbb{R}_{\geq 0}$, defined as

$$V_{\max} = V_{\max}(z,\theta) = \max_{x\in\Omega} |\, v(z,\theta)|_x |\,, \tag{7.10}$$

  with $|\cdot|$ denoting the usual Euclidean norm in $\mathbb{C}^3$. Note that, by construction, $V_{\max}$ is usually attained near the end of the teeth of the fork, since it is there that the forcing term has its support.

- The location (most importantly, the real part) of the natural resonating frequency of the tuning fork, i.e., in this context, the pole $\lambda_{\text{nat}} = \lambda_{\text{nat}}(\theta) \in \mathbb{C}$ of v with the smallest (positive) real part. Note that the frequency and parameter ranges have been chosen so that $\operatorname{Re}(\lambda_{\text{nat}}(\theta)) \in A$ for all $\theta \in \Theta$.

We remark that the former QoI is a non-linear function of the PDE state $v(z,\theta)$, depending on both z and θ. On the other hand, the latter QoI is a non-linear function of θ only, and for any θ, its value depends on the whole frequency response of the system.

To make the problem treatable on a computer, we introduce a FE discretization $\mathcal{V}_\mathcal{T}$ of $\mathcal{V}$ and then perform a Galerkin projection of (7.9) onto $\mathcal{V}_\mathcal{T}$. The finite-dimensional space $\mathcal{V}_\mathcal{T}$ is composed of piecewise-linear vector-valued functions defined over a tetrahedral discretization of Ω, which has approximately $1.3 \cdot 10^4$ vertices, resulting in $\dim(\mathcal{V}_\mathcal{T}) \approx 3.9 \cdot 10^4$. We note that the mesh size has been chosen according to the target frequency range A, to ensure a good resolution of the relevant frequencies.

Given the solution of the discrete problem $v(z,\theta) \in \mathcal{V}_\mathcal{T}$, the corresponding $V_{\max}$ can be easily found by computing the maximum of $|\, v(z,\theta)|_x |$ over all vertices x of $\mathcal{T}$, due to the discrete solution being piecewise-linear. On the other hand, computing λ_{nat} requires solving a non-Hermitian quadratic eigenproblem involving the system matrices. Here, we find an approximation of λ_{nat} by augmenting the quadratic eigenproblem, cf. Section 2.2, and then applying the (sparse) `eigs` method available in the `scipy.sparse.linalg` library [Vir+20], which implements an implicitly restarted Arnoldi method.

Before proceeding, we wish to mention that approximating this problem via projective MOR is rather complicated, since (i) the problem depends in a non-affine$^{\text{MOR}}$ way on z through the forcing term and (ii) the QoIs are non-linear, hindering online efficiency. In fact, as we will see below, the non-linearity of the first QoI is an issue even for our approach.

7.2.1 UQ of non-linear QoIs via locally adaptive sparse grids

We build a surrogate $\widetilde{v}$ for the state v by the pole/residue-matching approach from Chapter 6. To obtain the local surrogates, we apply adaptive MRI with Legendre polynomials, with starting frequency samples $\{50, 1000\}$Hz and tolerance 10^{-2}, using the relative look-ahead estimator from Section 5.3.3 (cf. also Section 5.5.3). The parameter sample points are selected adaptively starting from the initial set $\widetilde{\Theta} = \{\theta_0\}$, by using the Dörfler idea from Section 6.4.2.2, with adaptivity parameter $\eta = 0.5$. We set the computational budget by forcing the number of θ-samples to be at most 60. Interpolation of poles and residues over θ is carried out by piecewise-linear hat functions.

We deem important to note that, in the pole/residue matching step, we use a weighted version of the partial fraction distance (6.27):

$$\operatorname{dist}(\widetilde{H}(\cdot,\theta_{\text{test}}), \widetilde{H}_{\text{test}}) = \min_{\sigma\in(1:\widetilde{R})!} \sum_{i=1}^{\widetilde{R}_{\text{test}}} \left(\frac{1}{1\text{Hz}} \left| \widetilde{\lambda}_{\sigma_i} - \widetilde{\lambda}_i^{(\text{test})} \right| + \frac{1}{1\text{J}} \left\| \widetilde{r}_{\sigma_i} - \widetilde{r}_i^{(\text{test})} \right\|_{\mathcal{V}} \right).$$

This is necessary to make the units of poles and residues (Hz and J) compatible.

We remove faraway surrogate poles according to the Green's potential of A, keeping only those that have a potential no larger than $2\operatorname{Cap}(A) = 475$Hz. This corresponds to removing all poles outside an ellipse (in $\mathbb{C}$) centered at $z = 525$Hz, with horizontal and vertical semi-axes of lengths 593.75Hz and 356.25Hz, respectively. Only 4 exact poles are present in this area, as shown in Figure 7.19 (left). Note that, for all θ, the relevant poles form two pairs of almost coinciding complex numbers. In particular, all poles must have strictly negative imaginary part due to the Rayleigh damping.

Surrogates of the QoIs can then be found from $\widetilde{v}$, as follows:

- $\widetilde{V}_{\max}(z,\theta)$ can be computed from $\widetilde{v}(z,\theta)$ just like $V_{\max}(z,\theta)$ from $v(z,\theta)$, cf. the end of the previous section. Note that, due to its non-linear nature, $\widetilde{V}_{\max}$ is not online-efficient.
- $\widetilde{\lambda}^\star(\theta)$ is the pole of $\widetilde{v}(\cdot,\theta)$ with smallest positive real part. Note that $\widetilde{\lambda}^\star(\theta)$ is, by construction, online-efficient.

The algorithm terminates after 68 samples of θ (see Section 7.1.2 on why more than 60 samples of θ are taken), whose locations are displayed in Figure 7.18. We can observe that most of the refinement is performed along $\theta^{(1)}$. From this, we can qualitatively conclude that, according to the pMOR method, the Young's modulus is responsible for most of the variability. Overall, 886 snapshots of v are taken.

As in the previous example, the dimension of the parameter space prevents visualization of surrogate and corresponding error over the whole frequency and parameter ranges. In Figures 7.20 and 7.21 we restrict θ to a secondary diagonal of Θ, namely

$$\theta = \left((180 + 40\alpha)\text{GPa}, 0.24 + 0.2\alpha, (8.2 - 0.4\alpha)\text{kg/dm}^3\right) \quad \text{with } \alpha \in [0,1].$$

The first plot compares qualitatively the norm of the state v with its surrogate, whereas the second shows the relative approximation error in the state. Both results are fairly good, showing a good accuracy of the surrogate. We mention that the online speedup is approximately 60, i.e., in the time required to solve the FOM once, the surrogate can be evaluated 60 times. Note that

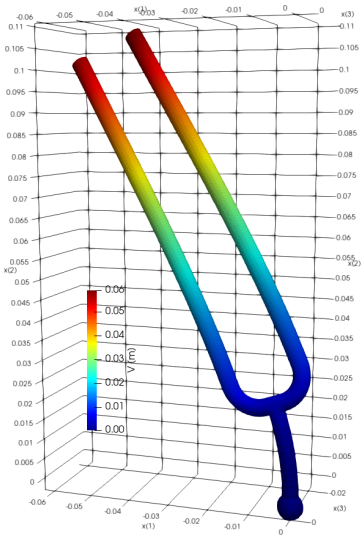


Figure 7.17 – Deformed tuning fork (by the real part of the displacement) at $(z, \theta) = (250\text{Hz}, \theta_0)$.

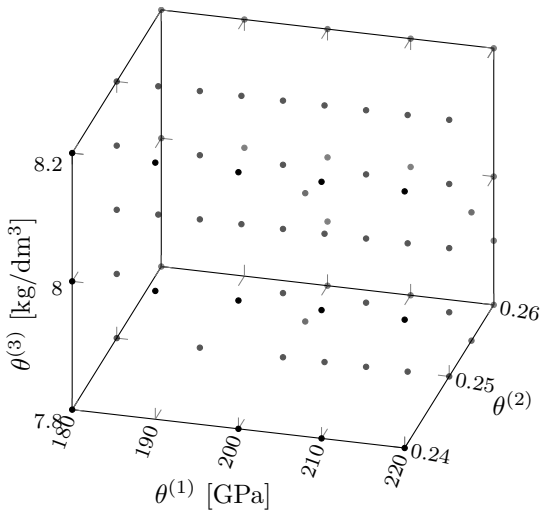


Figure 7.18 – Parameter sample points selected on a sparse grid by the Dörfler approach with $\eta = 0.5$.

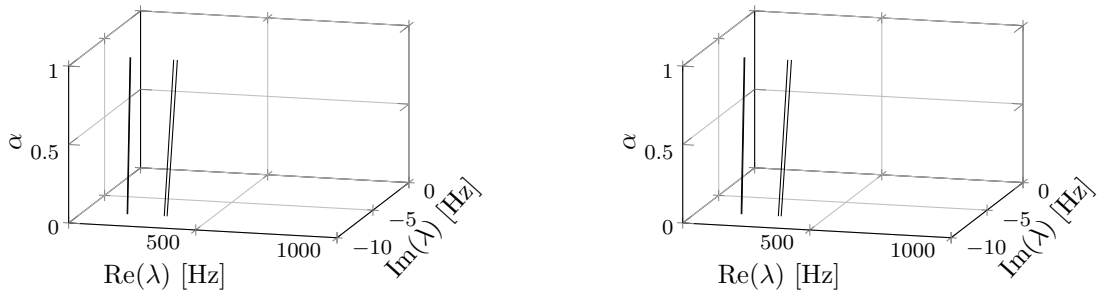


Figure 7.19 – Exact (left) and surrogate (right) poles of y over the main diagonal of Θ. The surrogate is built from parametric samples on a sparse grid. The coordinates $\alpha = 0$ and $\alpha = 1$ correspond to two vertices of Θ.

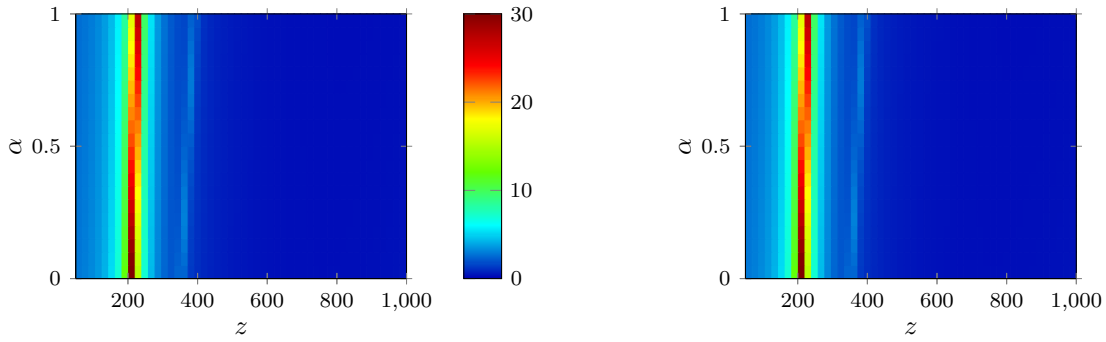


Figure 7.20 – Qualitative results for the ROM built from parametric samples on a sparse grid. Plot of exact elastic energy $\|v(z,\theta)\|_{\mathcal{V}}$ (measured in J) in the left plot. In the right plot, we show the surrogate $\|\widetilde{v}(z,\theta)\|_{\mathcal{V}}$. We vary θ over a secondary diagonal of Θ. All plots have the same color scale, reported next to the left plot. The coordinates $\alpha = 0$ and $\alpha = 1$ correspond to two vertices of Θ.

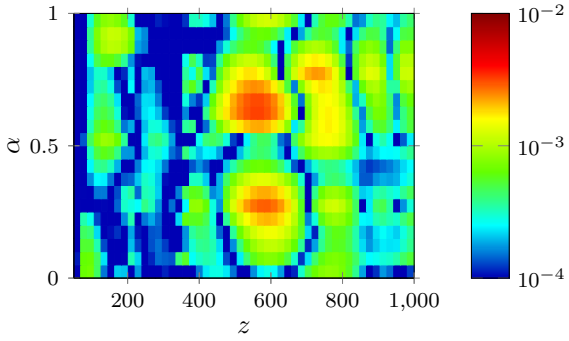
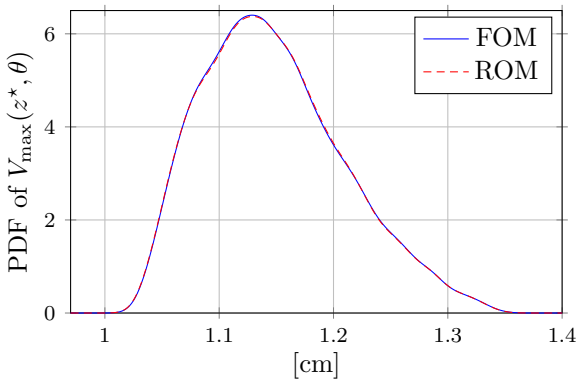
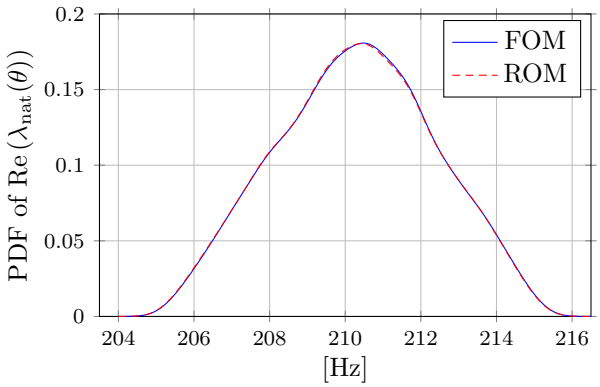


Figure 7.21 – Relative error in the approximation of v over a secondary diagonal of Θ, from parametric samples on a sparse grid. The coordinates $\alpha = 0$ and $\alpha = 1$ correspond to two vertices of Θ.



moment	FOM	ROM	rel. err.
mean [m]	1.15E-2	1.15E-2	8.86E-6
variance [m^2]	3.80E-7	3.80E-7	2.61E-4
skewness	5.16E-1	5.08E-1	1.46E-2
kurtosis	2.79E+0	2.78E+0	1.66E-3

Figure 7.22 – On the left, PDF of maximal displacement $V_{\max}$ at $z = z^\star = 405.8$Hz, approximated via 10000 MC samples. In blue and red, we show the results by using the FOM and the ROM (from parametric samples on a sparse grid), respectively. On the right, the MC approximations of the first moments of $V_{\max}$.



moment	FOM	ROM	rel. err.
mean [Hz]	2.10E+2	2.10E+2	5.44E-5
variance [Hz^2]	4.34E+0	4.35E+0	6.87E-4
skewness	-3.54E-2	-3.52E-2	4.39E-3
kurtosis	2.40E+0	2.40E+0	7.45E-4

Figure 7.23 – On the left, PDF of natural frequency $\mathrm{Re}(\lambda_{\mathrm{nat}})$, approximated via 10000 MC samples. In blue and red, we show the results by using the FOM and the ROM (from parametric samples on a sparse grid), respectively. On the right, the MC approximations of the first moments of $\mathrm{Re}(\lambda_{\mathrm{nat}})$.

this speedup is fairly good, but not as large as in the previous example, because, in the online phase, we are approximating the high-dimensional system state, so that true online efficiency is impossible to achieve.

Next, we move to the first QoI, namely, $V_{\max}$. For simplicity, we restrict our focus to a single frequency $z^\star = 405.8$Hz (selected at random). We use a Monte Carlo (MC) approach to approximate the probability density function (PDF) of $V_{\max}(z^\star, \cdot)$, based on $N_{\mathrm{MC}} = 10^4$ samples $\{V_{\max}(z^\star, \theta_i)\}_{i=1}^{N_{\mathrm{MC}}}$, with $\{\theta_i\}_{i=1}^{N_{\mathrm{MC}}}$ drawn from the distribution of θ. As an approximation, we repeat the experiment by replacing the FOM $V_{\max}$ with its surrogate $\widetilde{V}_{\max}$. In particular, we employ the exact same N_{MC} samples of θ. The PDFs obtained with the two approaches are shown in Figure 7.22, where we can observe good agreement. Note that, in our plot, rather than showing a histogram of the MC samples, we show its smoothed version via Gaussian kernel density estimation (KDE), using the `scipy.stats.kde` function [Vir+20].

Concerning the computational cost of the MC estimation, let us set as reference T_{FOM}, i.e., the time needed for a single FOM solve, which, on a desktop machine with an 8-core 3.60GHz Intel® processor, is approximately 2 seconds. Running MC with the FOM requires $N_{\mathrm{MC}}T_{\mathrm{FOM}}$ (the cost of extracting $V_{\max}$ from v is negligible). On the other hand, using the ROM requires approximately $886T_{\mathrm{FOM}}$ (the overhead for training the ROM) plus $N_{\mathrm{MC}}T_{\mathrm{FOM}}/60$, with 60 being the FOM–ROM speedup. Hence, as long as $N_{\mathrm{MC}} > 886 \Big/ \left(1 - \frac{1}{60}\right) \approx 900$, we can expect to save time.

Note that there is a (somewhat heuristic) way to further improve the online cost of the surrogate $\widetilde{V}_{\max}$ based on a localized approach. Indeed, as already observed, we can expect the maximum displacement to be attained near the top of the tuning fork. Let Γ_{top} be the union of the two flat disks that form the end caps of the tuning fork teeth, i.e.,

$$\Gamma_{\mathrm{top}} = \left\{ x \in \partial\Omega : x^{(2)} = \max_{X \in \partial\Omega} X^{(2)} \right\}.$$

(Technically, for our purposes, it would suffice to take the boundary (in the 2D sense) of Γ_{top}, which is composed of just two circles.) Rather than computing $V_{\max}$ directly from v using (7.10), we can introduce an "intermediate system output" $y(z,\theta) = v(z,\theta)|_{\Gamma_{\mathrm{top}}} \in [H^{1/2}(\Gamma_{\mathrm{top}})]^3$. Note that such y is a *linear* functional of v, whose size is much smaller than v's: specifically, in our simulation, $\dim(\mathcal{V}_\mathcal{T}) \approx 4 \times 10^4$, whereas $\dim(\mathcal{V}_\mathcal{T}|_{\Gamma_{\mathrm{top}}}) \approx 3 \cdot 10^2$. This means that a (mostly) online-efficient surrogate for y can be easily built from $\widetilde{v}$. Then, we can define the alternative surrogate

$$\widetilde{V}'_{\max} = \widetilde{V}'_{\max}(z,\theta) = \max_{x \in \Gamma_{\mathrm{top}}} |\widetilde{y}(z,\theta;x)| \approx V_{\max}(z,\theta).$$

Under the assumption that $\widetilde{V}_{\max}$ is attained over Γ_{top}, we have $\widetilde{V}'_{\max} = \widetilde{V}_{\max}$. Still, while the computation of $\widetilde{V}_{\max}$ requires finding the maximum over a vector of FE nodal values of size $\dim(\mathcal{V}_\mathcal{T})$, computing $\widetilde{V}'_{\max}$ is much faster, since it only involves a vector of few degrees of freedom.

Computing the second QoI λ_{nat}, although very complicated for the FOM, is extremely simple for the ROM. Indeed, it suffices to evaluate the interpolated surrogate pole with the smallest positive real part, cf. the surrogate expression (6.10). Hence, the solution of an augmented non-Hermitian eigenproblem can be replaced by the evaluation of a simple piecewise-linear scalar function. This allows for a significant online speedup. In our simulation, we measured the speedup factor to be approximately 3000, since we move from the approximately $15T_{\mathrm{FOM}}$ cost of solving the eigenproblem to the approximately $T_{\mathrm{FOM}}/200$ cost of evaluating the surrogate $\widetilde{\lambda}_{\mathrm{nat}}$. This

means that, overall, we save time as long as $N_{\text{MC}} > 886 \Big/ \left(15 - \frac{1}{200}\right) \approx 60$. In our experiments, this speedup comes without any significant drawbacks in accuracy, as we proceed to show. In Figure 7.23, we compare the MC estimation (smoothed by KDE) of the PDF of $\text{Re}(\lambda_{\text{nat}})$ using $N_{\text{MC}} = 10^4$ samples of θ. As in the previous case, we observe a good agreement of the two approximations.

7.2.2 UQ of non-linear QoIs via quasi-random samples

We repeat the experiment from the previous section, changing the strategy for the selection of the θ-samples used to build the surrogate via pMOR: we use the same number of samples of θ, but, this time, we identify them via a quasi-random low-discrepancy sequence generator. More specifically, we use the Halton sequence, see [Hal64]. Such points are much less structured and have the property of being quasi-uniformly spaced in Θ. Both features can be observed in Figure 7.24, where we plot the 68 sample points.

Due to the lack of geometric structure, we interpolate poles and residues via radial basis functions with linear bias, namely Wendland C^2 functions [Wen04], whose expression reads

$$\psi_{\theta_j,r}(\theta) = \rho_r \left(\sqrt{\left(\frac{\theta^{(1)} - \theta_j^{(1)}}{20\text{GPa}} \right)^2 + \left(\frac{\theta^{(2)} - \theta_j^{(2)}}{0.1} \right)^2 + \left(\frac{\theta^{(3)} - \theta_j^{(3)}}{0.2\text{kg/dm}^3} \right)^2} \right), \tag{7.11}$$

with

$$\rho_r(x) = \begin{cases} \left(1 - \frac{x}{r}\right)^4 \left(1 + 4\frac{x}{r}\right) & \text{if } x < r, \\ 0 & \text{otherwise.} \end{cases}$$

Above, $\theta_j \in \Theta$ is the center of the function, such that $\psi_{\theta_j,r}(\theta) < \psi_{\theta_j,r}(\theta_j) = 1$ for all $\theta \neq \theta_j$, whereas r is a scaling factor, whose choice will be detailed shortly. Note that, in (7.11), we have already applied a non-isotropic scaling by normalizing each parameter according to its range. The interpolant of the generic θ-dependent quantity ϕ (with $\theta \in \mathbb{C}^3$ as in our case) is

$$\widetilde{I}_r^{\{\theta_1,\dots,\theta_T\}}(\phi)(\theta) = \sum_{j=1}^{T} c_{j,r} \psi_{\theta_j,r}(\theta) + b_{0,r} + \sum_{i=1}^{3} b_{i,r} \theta^{(i)},$$

where the coefficients $\{c_{j,r}\}_{j=1}^T$ and $\{b_{i,r}\}_{i=0}^3$ are found by enforcing interpolation and orthogonality conditions:

$$\begin{cases} \widetilde{I}_r^{\{\theta_1,\dots,\theta_T\}}(\phi)(\theta_j) = \phi(\theta_j) & \text{for } j = 1,\dots,T, \\ \sum_{j=1}^{T} c_{0,r} = 0, & \\ \sum_{j=1}^{T} c_{i,r} \theta_j^{(i)} = 0 & \text{for } i = 1,2,3. \end{cases} \tag{7.12}$$

The above system has size $T + 4$ and is linear in the coefficients, so that it can be solved rather easily. Notably (7.12) has a symmetric saddle-point structure, and its conditioning depends on the choice of r: smaller values of r correspond to very concentrated basis functions, and make the interpolation conditions more "diagonal", leading to a better-conditioned system. Conversely, larger values of r increase the spread of the basis functions, making the problem worse-conditioned. As such, we employ the following "linear search"-like conditioning-motivated idea, fairly common among radial basis practitioners, for choosing the value of r:

- Start from an initial value, e.g., $r = 1$ or r equal to the smallest distance between sample

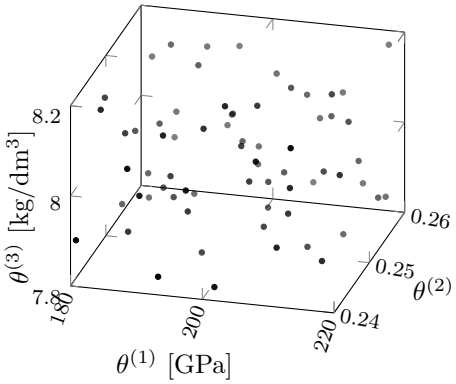


Figure 7.24 – Parameter sample points generated with the Halton scheme.

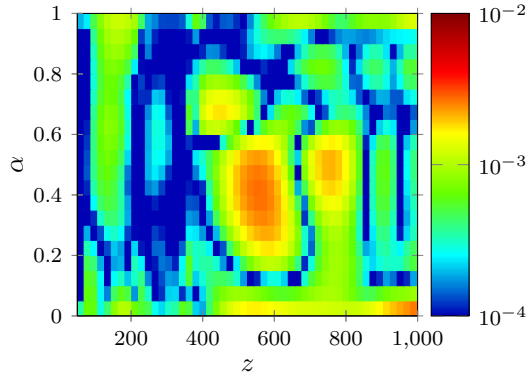
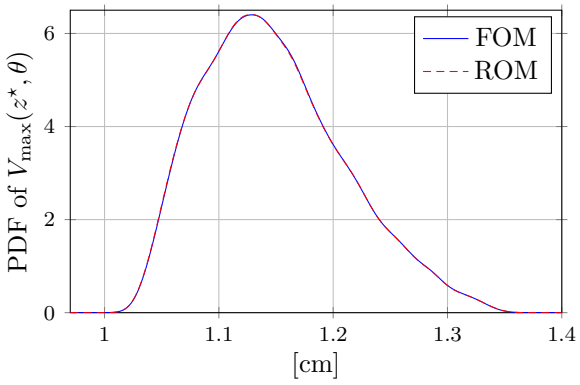
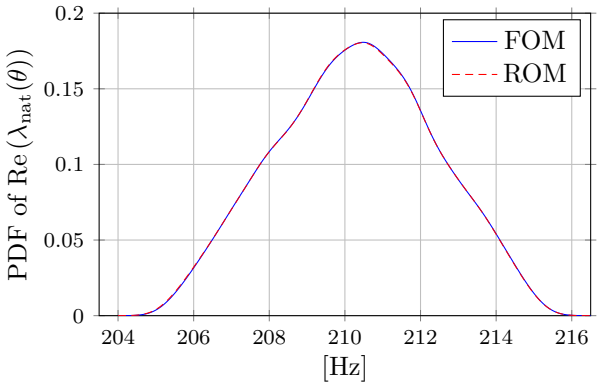


Figure 7.25 – Relative error in the approximation of v over a secondary diagonal of Θ, from quasi-random parametric samples. The coordinates $\alpha = 0$ and $\alpha = 1$ correspond to two vertices of Θ.



moment	FOM	ROM	rel. err.
mean [m]	1.15E-2	1.15E-2	1.30E-4
variance [m^2]	3.80E-7	3.80E-7	1.67E-3
skewness	5.16E-1	5.17E-1	1.18E-3
kurtosis	2.79E+0	2.79E+0	3.34E-4

Figure 7.26 – On the left, PDF of maximal displacement $V_{\max}$ at $z = z^\star = 405.8$Hz, approximated via 10000 MC samples. In blue and red, we show the results by using the FOM and the ROM (from quasi-random parametric samples), respectively. On the right, the MC approximations of the first moments of $V_{\max}$.



moment	FOM	ROM	rel. err.
mean [Hz]	2.10E+2	2.10E+2	3.16E-5
variance [Hz^2]	4.34E+0	4.35E+0	2.58E-3
skewness	-3.54E-2	-3.91E-2	1.06E-1
kurtosis	2.40E+0	2.40E+0	1.34E-3

Figure 7.27 – On the left, PDF of natural frequency $\mathrm{Re}(\lambda_{\mathrm{nat}})$, approximated via 10000 MC samples. In blue and red, we show the results by using the FOM and the ROM (from quasi-random parametric samples), respectively. On the right, the MC approximations of the first moments of $\mathrm{Re}(\lambda_{\mathrm{nat}})$.

points.

- Assemble (7.12) and compute its spectral condition number κ_r, i.e., the ratio of the largest and smallest singular value of the matrix appearing in the left-hand-side of (7.12).
- If κ_r is too small, say $\kappa_r < 10^6$, increase r and retry.
- If κ_r is too large, say $\kappa_r > 10^{12}$, decrease r and retry.

(Note that this strategy is independent of the interpolated quantity ϕ.) This approach (with the above-mentioned bounds on the condition number) yields $r = 16$ for our set of samples.

In Figure 7.25, we show the approximation error over a secondary diagonal of Θ (the same as in the previous example, cf. Figure 7.21). We see that the approximation error is fairly similar to the sparse grid case, if not slightly better, except near the extreme values $\alpha = 0$ and $\alpha = 1$, since, after all, the corresponding vertices of Θ do *not* belong to the training set $\widetilde{\Theta}$ (whereas they always belong to the sparse grid in the previous section). More generally, we may expect the quality of the surrogate to degrade slightly near the boundary of Θ. The results of the MC simulations using the surrogate are also fairly similar to the sparse grid case, as can be seen in Figures 7.26 and 7.27.

Overall, this approach has some advantages:

- Interpolating poles and residues with radial basis functions can be expected to yield better results than with piecewise-linear hat functions whenever the interpolated quantities are smooth functions of θ. This seems to be the case in the present example.
- Due to the low-discrepancy property of the sequence of sample points, approaches based on quasi-random points should be expected to place a higher "density" of sample points *inside* Θ rather than near its boundary. This can be beneficial in the context of UQ, since extreme values of θ (located near the boundary of Θ) are often less likely, particularly if the underlying distribution of θ is not uniform but has a higher concentration near θ_0.

However, there are also some drawbacks:

- Radial basis functions are a bit more costly to evaluate than piecewise-linear hat functions (at least, this is the case in our implementation). Accordingly, the online efficiency might suffer slightly. In this specific example, the speedup factor is approximately 15, as opposed to 60 for sparse grids.
- As already mentioned in Section 6.4.2.1, quasi-random sequences do not allow for local refinements of the test set in an adaptive sampling framework. In particular, a Dörfler approach with quasi-random sequences is essentially the same as just taking the first T elements of the quasi-random sequence, with T being the computational burden.

7.3 Admittance of a transmission line

Our next example comes from circuit modeling and analysis. With it, we wish to show that even a (slightly) larger number of parameters can be handled by our proposed technique. Before proceeding, we wish to mention that a similar example has been presented in [NP21].

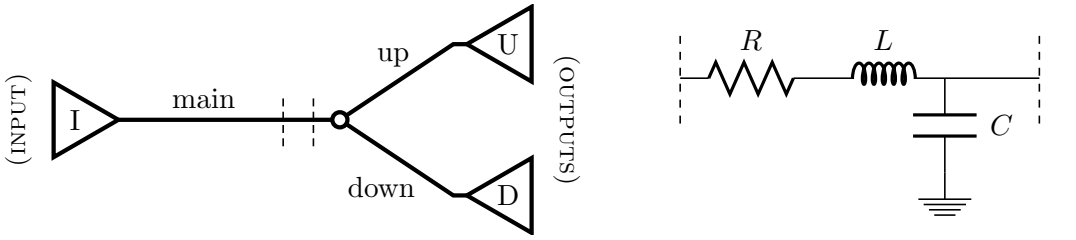


Figure 7.28 – Schematic of the transmission line (left) and circuit representation of a unit RLC cell (right). The degrees of freedom in the modified nodal analysis formulation are the voltage across the capacitor and the current through the resistor (inductor and resistor are lumped together using their total equivalent impedance $R + 2\pi\mathrm{i}L$).

We consider the 3-port Y-shaped transmission line depicted in Figure 7.28. Each branch of the circuit is composed of a series of unit RLC cells: the "main" branch contains 200 cells, whereas the "up" and "down" branches contain 100 cells each. The point of contact between the three branches is a simple Y-junction. The values of resistance, inductance, and capacitance vary both within a branch and from one branch to the others. More specifically, let the cells be indexed from left to right by j, with $j \in \{1, \dots, 200\}$ on the main branch and $j \in \{1, \dots, 100\}$ on the others. Then, we have

$$\begin{aligned} R_j^{\text{main}} &= \xi_j^{(1)}\theta^{(1)}, \quad R_j^{\text{up}} = \xi_j^{(2)}\theta^{(2)}, \quad R_j^{\text{down}} = \xi_j^{(3)}\theta^{(3)}, \\ L_j^{\text{main}} &= \xi_j^{(4)}\theta^{(4)}, \quad L_j^{\text{up}} = \xi_j^{(5)}\theta^{(5)}, \quad L_j^{\text{down}} = \xi_j^{(6)}\theta^{(6)}, \\ C_j^{\text{main}} &= \xi_j^{(7)}\theta^{(7)}, \quad C_j^{\text{up}} = \xi_j^{(8)}\theta^{(8)}, \quad C_j^{\text{down}} = \xi_j^{(9)}\theta^{(9)}, \end{aligned}$$

with $\{\xi_j^{(i)}\}_{j,i}$ being dimensionless scaling factors that we use to model random fluctuations of the nominal values in each cell. In particular, we draw each of them independently from a uniform distribution over $[0.75, 1.25]$. Such random values are drawn once and for all at the beginning of the simulation, and we consider them fixed in the scope of our analysis.

As parameters, we take the mean values of R, L, and C over each branch, namely, $(\theta^{(1)}, \dots, \theta^{(9)})$, which we constrain to the hyper-rectangular parameter domain

$$\theta \in \Theta = ([9, 11]\mathrm{m}\Omega)^3 \times ([450, 550]\mathrm{pH})^3 \times ([450, 550]\mathrm{fF})^3 .$$

We denote by θ_0 the centroid of Θ and we set the frequency range as $z \in A = [0, 8]\mathrm{GHz}$. In A and immediately around it, we can find 12 resonances of the system. Their real parts can be seen in Figure 7.30 (note that there is a double pole with $\mathrm{Re}(\lambda) = 0\mathrm{Hz}$ for all θ) and their imaginary parts are approximately equal to 15MHz.

Our target is the analysis of the admittance parameters $y = y(z, \theta)$ of the system, which form a 3×3 complex matrix (measured in Ω^{-1}). Each entry of the matrix provides information on the current between a port of the circuit and ground, when a unit AC voltage is applied at a single port, while the other ports are shorted to ground. To compute the admittance parameters, we apply modified nodal analysis to the circuit, obtaining a θ-dependent system representation of the form (6.1). In particular, we have $n_u = n_y = 3$, with one column of B and one row of C per port of the circuit, whereas $n_v = 803$, with v containing a collection of current and voltage values within the circuit. Note that the system is, by linearity of the Kirchhoff laws and of the components, affine in θ.

7.3.1 High-dimensional adaptive sampling with modest tolerance

We build a surrogate $\widetilde{y}$ for the admittance parameters y by the pole/residue-matching approach. To obtain the local surrogates, we apply adaptive MRI with Legendre polynomials, with starting frequency samples $\{0, 8\}$GHz and (relative) tolerance 10^{-3}, using the look-ahead estimator from Section 5.3.3. The parameter sample points are selected adaptively starting from the initial set $\widetilde{\Theta} = \{\theta_0\}$. As stopping criterion, we fix the adaptive parameter sampling tolerance $\epsilon = 5 \cdot 10^{-2}$, see Algorithm 6, on a weighted version of the partial fraction distance (6.27):

$$\operatorname{dist}(\widetilde{H}(\cdot, \theta_{\text{test}}), \widetilde{H}_{\text{test}}) = \min_{\sigma \in (1:\widetilde{R})!} \sum_{i=1}^{\widetilde{R}_{\text{test}}} \frac{1}{4} \left((10^{-9}\text{s}) \left| \widetilde{\lambda}_{\sigma_i} - \widetilde{\lambda}_i^{(\text{test})} \right| + (1\Omega) \left\| \widetilde{r}_{\sigma_i} - \widetilde{r}_i^{(\text{test})} \right\|_F \right),$$

with $\|\cdot\|_F$ the Frobenius norm. We do this to normalize the poles, but also to make the units (Hz and Ω^{-1}) compatible by nondimensionalization. Note that the same quantity is used also to match poles and residues. Then, we use piecewise-linear hat functions to interpolate poles and residues.

We remove faraway surrogate poles according to the Green's potential of A, keeping only those that have a potential no larger than $1.5\,\text{Cap}(A) = 3$GHz. This corresponds to removing all poles outside an ellipse (in $\mathbb{C}$) centered at $z = 4$GHz, with horizontal and vertical semi-axes of lengths 5GHz and 1.67GHz, respectively. Note that, due to this fairly strict cut-off, poles that leave the frequency range by more than 1GHz are necessarily removed from the local surrogates. This makes it complicated for the global surrogate to approximate poles that are in A (or close to it) for some parameter samples but then move (slightly) farther away for other values of θ. As a way to counteract this problem, we apply the "unbalanced matching" approach from Algorithm 5, with $\delta = 0.25$. This means that surrogate poles might be synthetic at up to 25% of the sample points.

The parameter sampling loop ends after 3 iterations of Algorithm 6, with a total number of parameter samples equal to 163, corresponding to 3625 total snapshots. In Figure 7.29, we show some components of the sample points. We can observe that refinement are applied mostly across different electric components (e.g., resistors and capacitors) and not between the same components (e.g., resistors and resistors), regardless of branches.

Due to the modest dimension of the parameter space, we restrict θ to 1-dimensional manifolds to make our next visual comparisons. In particular, we consider the principal and a secondary diagonal of Θ:

$$\begin{aligned}\theta_P = \Big(&(9 + 2\alpha_P)\text{m}\Omega, (9 + 2\alpha_P)\text{m}\Omega, (9 + 2\alpha_P)\text{m}\Omega, (450 + 100\alpha_P)\text{pH}, (450 + 100\alpha_P)\text{pH}, \\ &(450 + 100\alpha_P)\text{pH}, (450 + 100\alpha_P)\text{fF}, (450 + 100\alpha_P)\text{fF}, (450 + 100\alpha_P)\text{fF} \Big) \quad \text{for } \alpha_P \in [0, 1]\end{aligned}$$

and

$$\begin{aligned}\theta_S = \Big(&(11 - 2\alpha_S)\text{m}\Omega, (9 + 2\alpha_S)\text{m}\Omega, (9 + 2\alpha_S)\text{m}\Omega, (550 - 100\alpha_S)\text{pH}, (550 + 100\alpha_S)\text{pH}, \\ &(450 + 100\alpha_S)\text{pH}, (550 - 100\alpha_S)\text{fF}, (450 + 100\alpha_S)\text{fF}, (450 + 100\alpha_S)\text{fF} \Big) \quad \text{for } \alpha_S \in [0, 1].\end{aligned}$$

In Figures 7.30 and 7.31, we show the results for principal and secondary diagonals, respectively. For simplicity, we restrict our focus to a single off-diagonal entry of y: the admittance between the

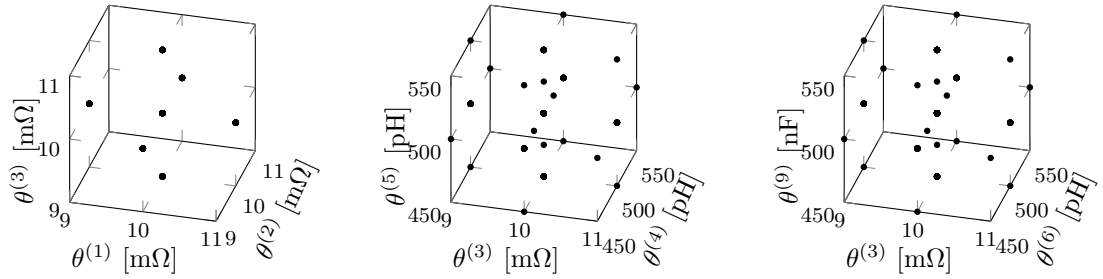


Figure 7.29 – Adaptively selected parameter sample points (using a modest tolerance). Different triples of components are shown in different plots.

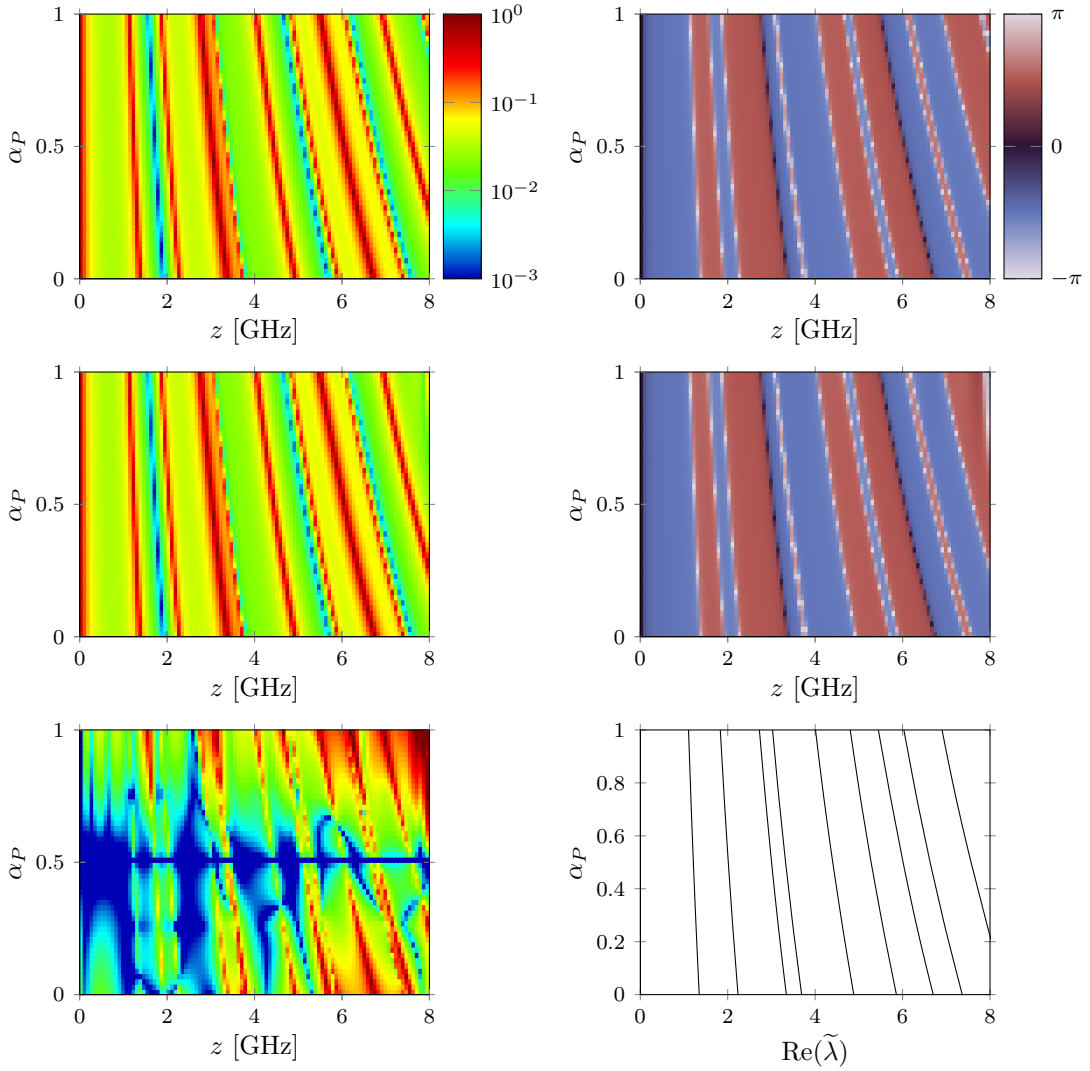


Figure 7.30 – Qualitative results using a modest tolerance along the principal diagonal of Θ. Top left: exact $|y_{\mathrm{UD}}(z, \theta_P)|$ (measured in Ω^{-1}). Top right: exact $\angle y_{\mathrm{UD}}(z, \theta_P)$. Middle left: surrogate $|\widetilde{y}_{\mathrm{UD}}(z, \theta_P)|$. Middle right: surrogate $\angle \widetilde{y}_{\mathrm{UD}}(z, \theta_P)$. Bottom left: relative error $|\widetilde{y}_{\mathrm{UD}}(z, \theta_P) - y_{\mathrm{UD}}(z, \theta_P)| \,/\, |y_{\mathrm{UD}}(z, \theta_P)|$. Bottom right: surrogate poles $\widetilde{\lambda}(\theta_P)$. In each column, all plots have the same color scale, reported next to the top plot.

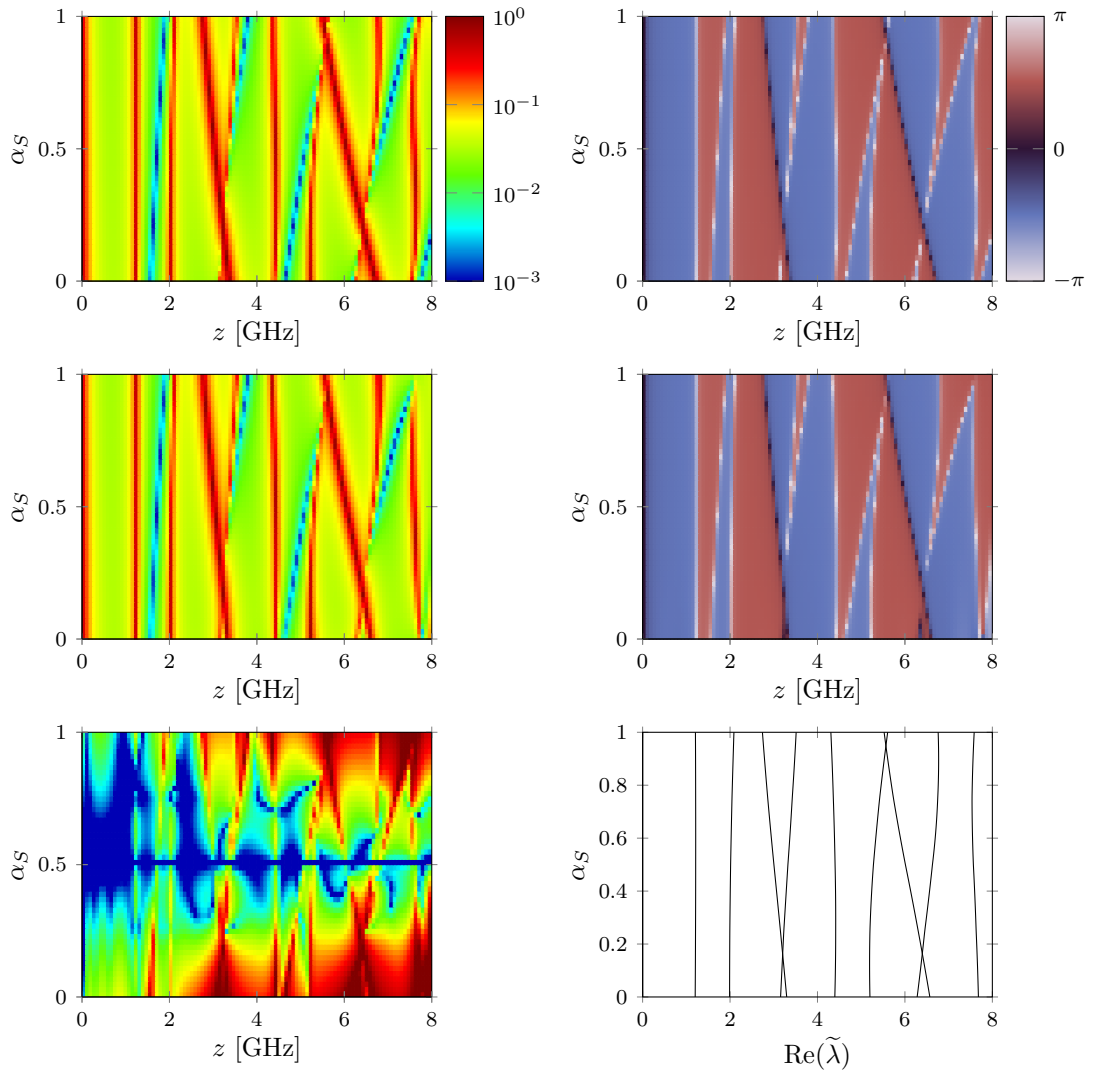


Figure 7.31 – Qualitative results using a modest tolerance along a secondary diagonal of Θ. Top left: exact $|y_{\mathrm{UD}}(z,\theta_S)|$ (measured in Ω^{-1}). Top right: exact $\angle y_{\mathrm{UD}}(z,\theta_S)$. Middle left: surrogate $|\widetilde{y}_{\mathrm{UD}}(z,\theta_S)|$. Middle right: surrogate $\angle\widetilde{y}_{\mathrm{UD}}(z,\theta_S)$. Bottom left: relative error $|\widetilde{y}_{\mathrm{UD}}(z,\theta_S) - y_{\mathrm{UD}}(z,\theta_S)| \,/\, |y_{\mathrm{UD}}(z,\theta_S)|$. Bottom right: surrogate poles $\widetilde{\lambda}(\theta_S)$. In each column, all plots have the same color scale, reported next to the top plot.

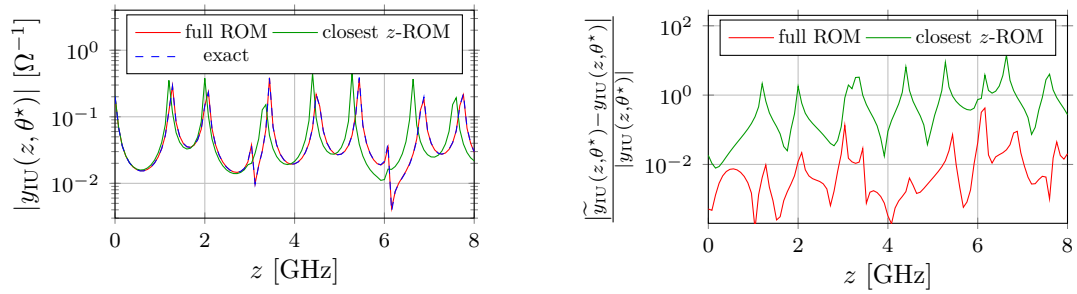


Figure 7.32 – On the left, plot of the magnitude of exact $y_{\mathrm{IU}}(z,\theta^\star)$ and of its surrogates (using a modest tolerance). On the right, the corresponding relative error.

two output ports. In the eyeball-norm, the surrogate seems to provide a fairly good approximation of both magnitude and phase of y_{UD}. We note that it looks like the zeros of y (the blue curves in the magnitude plots) are not approximated too well. After all, MRI focuses mostly on the approximation of the *poles* of the target quantity, rather than of its *zeros*: a good approximation of the zeros derives, somewhat indirectly, from the global convergence results (in capacity), cf. Theorems 3.7 and 3.8.

The relative error plots provide more quantitative results. We can see that the error is rather large (between 0.1 and 1) over a modest portion of the frequency and parameter range (especially in Figure 7.31). This was to be expected due to the intrinsic difficulties in approximating multivariate rational functions in modest and high dimension. In our specific case, the not-so-small error is a direct consequence of the rather large tolerance that was employed. Setting a lower tolerance would decrease the error, at the cost of a (much) larger number of samples. Overall, the observed behavior is consistent with the usual results when approximating high-dimensional functions, where the approximation quality is great in the eyeball-norm, but not that good in the "actual" error norm.

The presence of areas of Θ where the relative error is larger than the prescribed tolerances for z- and θ-adaptivity might seem an indication that the adaptive sampling algorithms, i.e., Algorithms 3 and 6, are not working as they should. However, this is not the case for the following reasons:

- The tolerance on z is guaranteed to be attained only at the training points $\widetilde{\Theta}$ and not on the whole Θ. In fact, even on $\widetilde{\Theta}$, the relative error is necessarily below the tolerance only over the (admittedly, very fine) z-test set.
- The tolerance on θ is guaranteed to be attained only at the training and test points, i.e., in particular, at the forward points of the parameter sample set $\widetilde{\Theta}$. Due to the modest dimension of the parameter space, 3 iterations are not enough to reach the vertices of Θ via forward points. This allows for rather large errors at $\alpha_P, \alpha_S \in \{0, 1\}$, which are actually quite far (at least, in the "forward points" sense) from the training set. One could weaken this issue by starting from a larger initial set $\widetilde{\Theta}$ of training points.
- The tolerance on θ is enforced on the partial fraction distance between surrogate and truth models at the test parameters. While a small partial fraction distance is necessary to have a good approximation (notably, a small relative error), it is not sufficient for it. In particular, this can be observed when some of the exact poles are missing from the surrogate, as it happens near $z = 8\mathrm{GHz}$ and $\alpha_P = 1$.

Unfortunately, these issues are not easily solvable, especially when the number of parameters is large, with the main culprit being the curse of dimension. Indeed, a faithful approximation of a $(1 + n_\theta)$-varied function (like our QoI) will, in general, require a number of samples that increases exponentially with n_θ, even if the target quantity is sufficiently smooth. In devising an adaptive sampling strategy, one should be aware of the fact that a "sufficiently in-depth" exploration of the parameter space is quite often computationally unfeasible. In our experiments, we choose to be conservative on the total number of snapshots taken, accepting the inevitable ensuing decrease in accuracy.

Looking at Figure 7.30, we can see that the surrogate misses a single pole, close to 8GHz for $\alpha_P = 1$. This was to be expected, since the pole is rather far from A (notably, it is outside the cut-off ellipse) for most values of θ, so that it gets eliminated by Algorithm 5. We note that this

causes local inaccuracies, visible mostly in the phase and error plots. Also concerning the poles, we can see in Figure 7.31 that the poles cross. The surrogate is able to recognize this, cf. the crossing surrogate poles in the bottom-right figure, thanks to the presence of the residual distance in the partial fraction distance (6.11) that gets minimized in the matching step.

Since it cannot be seen in the pole plots, we also underline that the double pole at $z = 0\text{Hz}$ is properly identified by the surrogate, which has an (almost) θ-independent pair of poles at $z \approx \pm 1\text{kHz}$.

As a final experiment, we consider the randomly selected parameter point

$$\theta^\star = (9.75\text{m}\Omega, 10.9\text{m}\Omega, 10.5\text{m}\Omega, 510\text{pH}, 466\text{pH}, 466\text{pH}, 456\text{fF}, 537\text{fF}, 510\text{fF}),$$

which does not belong to $\widetilde{\Theta}$. We look at the admittance y_{IU} between input and up ports, comparing the exact and surrogate values. Moreover, we also consider the value of y_{IU} yielded by the local frequency surrogate built at the point

$$\widetilde{\theta}^\star = \operatorname*{arg\,min}_{\widetilde{\theta} \in \widetilde{\Theta}} \left| \widetilde{\theta} - \theta^\star \right|.$$

Such local surrogate is part of the surrogates whose partial fraction expansions are interpolated to give the overall (z, θ)-surrogate. This comparison is a numerical verification of the pole/residue interpolation step: more specifically, we wish to check whether the interpolated poles and residues work better than just using their respective values at the closest parameter sample point, cf. the discussion on Haar grids in Section 6.4.1.

The results are shown in Figure 7.32. A comparison of the surrogate magnitude $|y_{\text{IU}}(\cdot, \theta^\star)|$ shows that the piecewise-constant (in θ) frequency surrogate at $\widetilde{\theta}^\star$ performs rather poorly, with the larger poles being grossly misplaced, whereas the overall piecewise-linear (in θ) surrogate yields a much better approximation. The relative error provides a quantitative verification of this fact. We can see that the global surrogate error is uniformly smaller.

7.3.2 High-dimensional adaptive sampling with low(er) tolerance

We repeat the experiment from the previous section with a reduced θ-sampling tolerance $\epsilon = 10^{-2}$. This time, the algorithm terminates in 8 iterations, after 1945 samples of θ, corresponding to 41428 total snapshots. We can observe in Figure 7.33 some components of the sample points. Again, we can observe that refinements are carried out mostly across electrical components of different type.

In Figures 7.34 and 7.35, we show the results for the principal and secondary diagonal defined in the previous section, respectively. Again, we only look at the admittance between the two output ports. We notice that the approximation error near $z = 8\text{GHz}$ and $\alpha_P = 1$ is still rather large, due to the local pole that is missed by the overall surrogate. Similarly, the approximation error near $z = 8\text{GHz}$ and $\alpha_S = 0$ has not improved much, apparently due to a lack of local refinements. Except for that, the approximation quality seems to be at least slightly better.

We show in Figure 7.36 the surrogate at the randomly selected point $\theta^\star$, cf. the previous section. Again, we compare the overall surrogate with the local frequency surrogate built at the sample point closest to $\theta^\star$. In particular, we note that such closest sample point $\widetilde{\theta}^\star$ is, by chance, the same as in the previous section. The conclusions that we can draw from Figure 7.36 are similar to

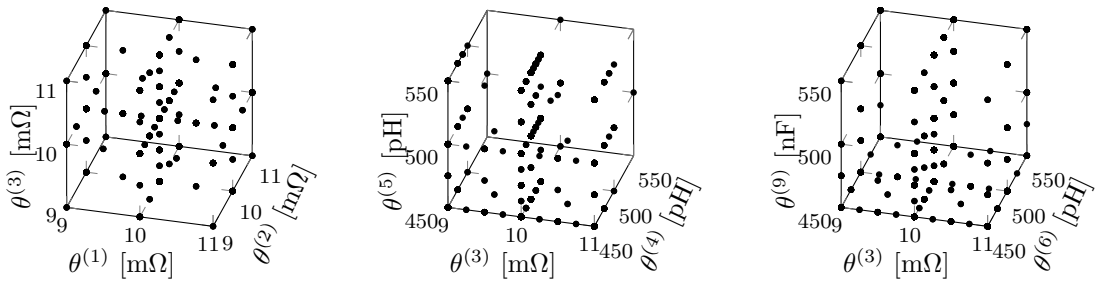


Figure 7.33 – Adaptively selected parameter sample points (using a small tolerance). Different triples of components are shown in different plots.

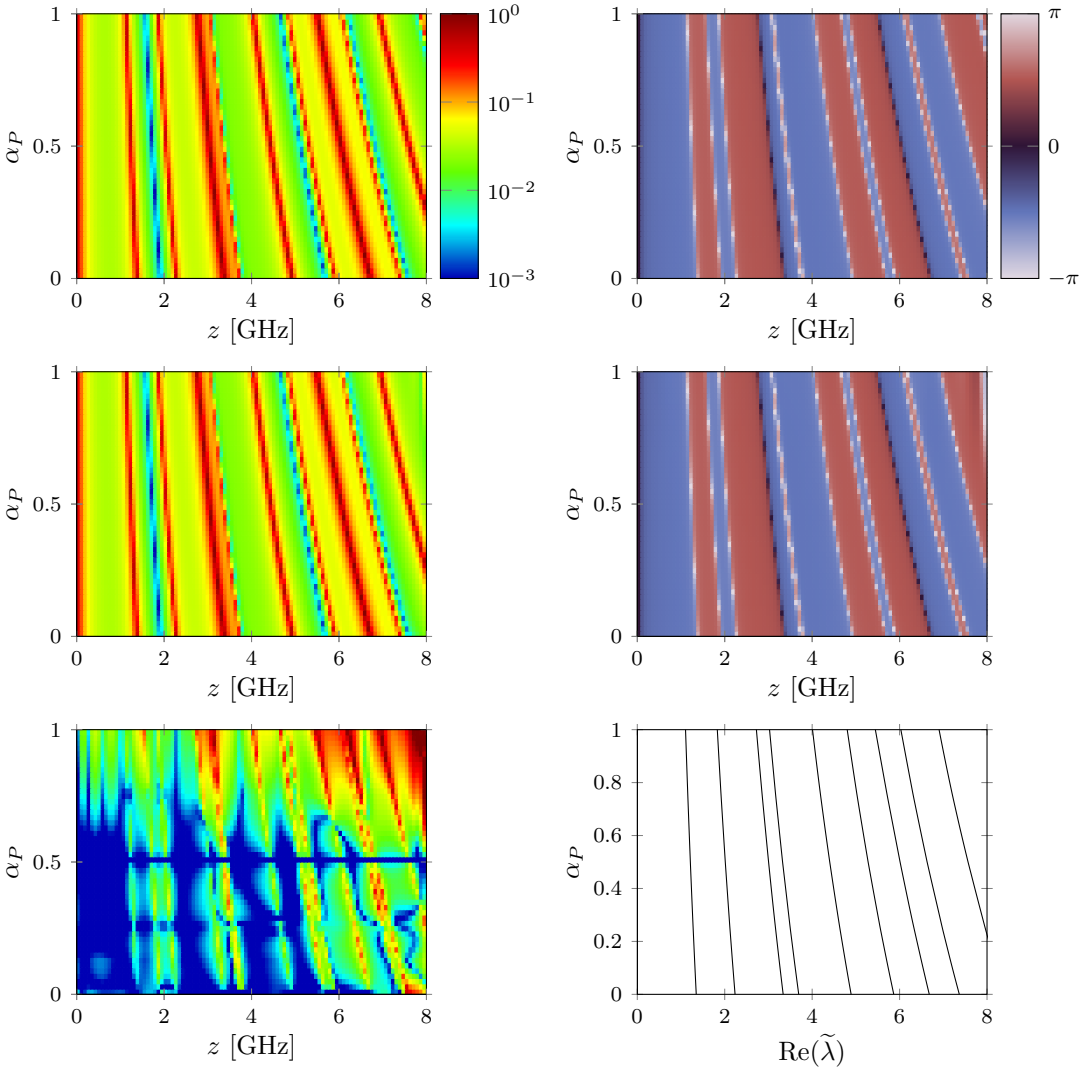


Figure 7.34 – Qualitative results using a small tolerance along the principal diagonal of Θ. Top left: exact $|y_{\mathrm{UD}}(z, \theta_P)|$ (measured in Ω^{-1}). Top right: exact $\angle y_{\mathrm{UD}}(z, \theta_P)$. Middle left: surrogate $|\widetilde{y}_{\mathrm{UD}}(z, \theta_P)|$. Middle right: surrogate $\angle \widetilde{y}_{\mathrm{UD}}(z, \theta_P)$. Bottom left: relative error $|\widetilde{y}_{\mathrm{UD}}(z, \theta_P) - y_{\mathrm{UD}}(z, \theta_P)| \,/\, |y_{\mathrm{UD}}(z, \theta_P)|$. Bottom right: surrogate poles $\widetilde{\lambda}(\theta_P)$. In each column, all plots have the same color scale, reported next to the top plot.

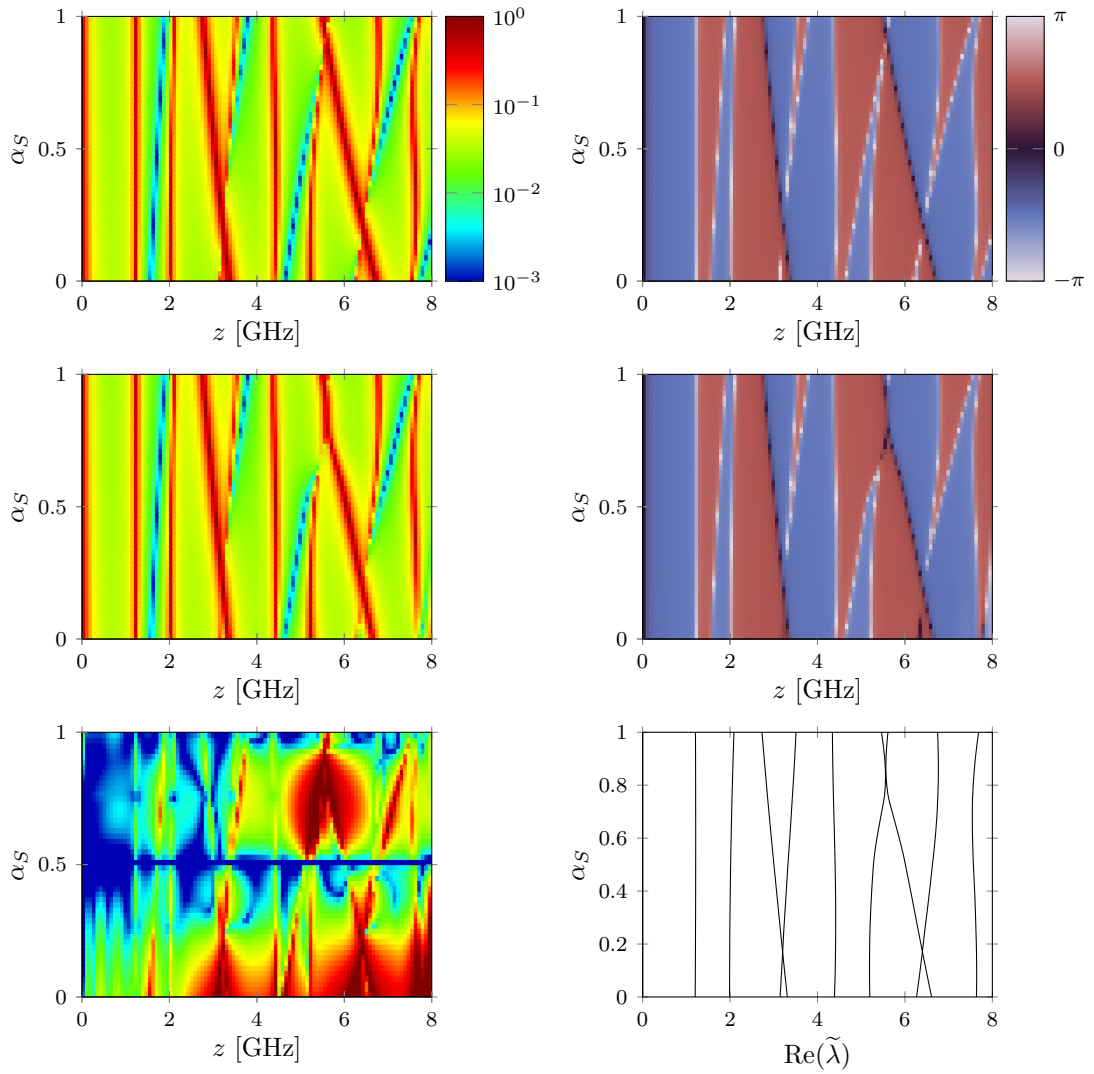


Figure 7.35 – Qualitative results using a small tolerance along a secondary diagonal of Θ. Top left: exact $|y_{\mathrm{UD}}(z,\theta_S)|$ (measured in Ω^{-1}). Top right: exact $\angle y_{\mathrm{UD}}(z,\theta_S)$. Middle left: surrogate $|\widetilde{y}_{\mathrm{UD}}(z,\theta_S)|$. Middle right: surrogate $\angle\widetilde{y}_{\mathrm{UD}}(z,\theta_S)$. Bottom left: relative error $|\widetilde{y}_{\mathrm{UD}}(z,\theta_S) - y_{\mathrm{UD}}(z,\theta_S)| \,/\, |y_{\mathrm{UD}}(z,\theta_S)|$. Bottom right: surrogate poles $\widetilde{\lambda}(\theta_S)$. In each column, all plots have the same color scale, reported next to the top plot.

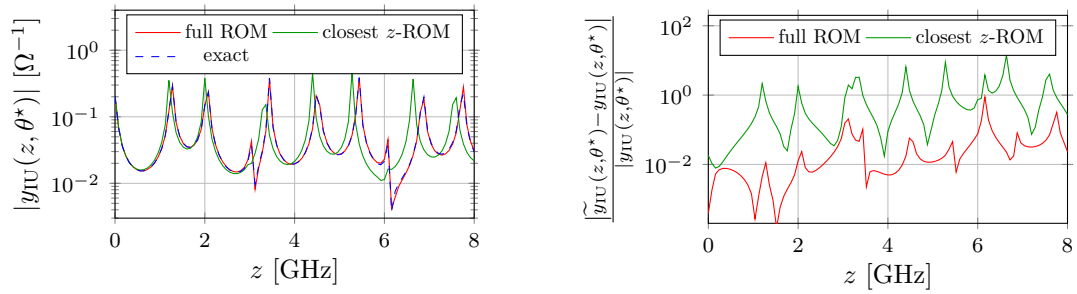


Figure 7.36 – On the left, plot of the magnitude of exact $y_{\mathrm{IU}}(z,\theta^\star)$ and of its surrogates (using a small tolerance). On the right, the corresponding relative error.

those that we already derived from Figure 7.32. Notably, the global surrogate error is comparable, if not slightly worse, than the one from the last section. In our opinion, the reason behind this is related to the pole around $z = 8$GHz that is "missed" near the vertex of Θ corresponding to $\alpha_P = 1$. Indeed, even though, with respect to the results from the previous section, more parameter samples are added near such vertex, the pole is still rejected by Algorithm 6, since it is still missing in more than 25% of the local frequency surrogates. Accordingly, several of the frequency surrogates have to be adjusted to account for the removed pole, as described in Section 6.3.3, introducing additional minor errors for z near 8GHz. Since the number of adjusted surrogates is larger (in proportion) than before, we can expect the effect of this correction to be slightly more noticeable.

Overall, in our view, the "increase" in accuracy was not worth the extra offline cost, especially considering that the accuracy is not guaranteed to increase, cf. Figures 7.32 and 7.36. Indeed, our numerical results can serve as evidence of the fact that, in the parametric setting, the approximation error does not necessarily decrease monotonically as new snapshots are added.

8 Conclusions and outlook

This thesis has covered most of the research work that the author has carried out in his 4 years as a PhD candidate at EPFL.

The first novel contribution is in the form of an original MOR technique, dubbed MRI, aimed at the non-intrusive surrogate modeling of non-parametric frequency-domain applications. We have developed and analyzed MRI in the framework of rational approximation. On the theoretical side, we have taken as reference some classical results in scalar approximation theory [Wal60], and we have shown that the approximation error converges maximally for a suitable class of target functions. Also, we have shown that the approximation of the resonating frequencies of the problem converges at exponential rate. On the applied side, we have provided a practical MRI algorithm and we have investigated the most relevant issues that intrinsically arise when implementing non-intrusive rational MOR techniques like MRI, especially considering finite-precision arithmetic. In addition, we have introduced a strategy for combining MRI with adaptive sampling, which allows to identify the "best" sample point locations in an automated fashion.

Then, we have outlined the main problems that one encounters in moving from the non-parametric setting to the parametric one, and we have described a pMOR approach as a way to solve some of them, relying on a marginalization strategy: MRI is used to build frequency surrogates at different parameter values, which are then combined to obtain a global reduced model. This method is still non-intrusive and, if coupled with a suitable parameter sampling strategy, can be applied even if the number of parameters is modest. The combination of the frequency surrogates is a critical step for a good approximation quality. As such, we have discussed the topic to some length, presenting the biggest limitations of our approach, which are mostly related to the approximability of the problem and to the (possible) incompatibility of the frequency surrogates.

We have complemented our presentation with several numerical experiments to showcase our proposed approaches, both in the non-parametric and parametric settings. In the non-parametric case, we have compared MRI with other MOR methods, showing that the obtained results are mostly similar to (and sometimes better than) the state-of-the-art approaches for the considered problems. Moreover, we have also successfully verified our theoretical claims in an empirical way. In the parametric case, we have provided numerical evidence to support the effectiveness of our technique in few "case studies" of practical interest, whose parametric frequency responses were approximated fairly well by our method. Although the sizes of our examples are rather small, when measured in terms of the cost of solving the FOM, we have still been able to observe a significant ROM-FOM speedup. This allows us to hope for an even greater online efficiency in

more realistic applications.

8.1 Perspectives

We believe that this thesis provides a thorough presentation of our proposed MOR technique, with an incremental discussion going from the single-point non-parametric case to the rather complex parametric setting with fully adaptive sampling. This being said, our investigation has not, by far, reached its conclusion, and several questions remain outstanding.

A first matter that, in our opinion, deserves further study, is that of well-posedness for MRI, in the following sense. Consider the non-parametric setting, and assume that the FOM snapshots are affected by unpredictable noise. If the noise is "small", should we expect the MRI to be close to that obtained in the noise-free case? Many results can be found in the literature for the well-posedness of *polynomial* interpolation, based, e.g., on the concept of Lebesgue constant [Smi06]. As such, we should be able to deal with the stability related to the computation of the MRI numerator. On the other hand, the MRI denominator depends on the snapshots in a rather complicated way, and it is not clear if we should expect well-posedness. Some preliminary investigations on this are actually under way, in the scope of applying MRI to snapshots obtained with h-adaptive FEM, so that the above-mentioned noise corresponds to the numerical discrepancies due to the different h-adapted meshes used to compute the snapshots. Some results in this context are already available for intrusive MOR, see, e.g., [ASU17; URL16], even though the theory only applies to elliptic and parabolic PDEs, and not to problems with resonant behavior like ours.

A second open question concerns the pole/residue matching step described in Section 6.3. Our decision of matching poles and residues of the frequency surrogates one by one, cf. Section 6.3.1.2, was motivated by complexity concerns: the bipartite matching problem can be solved in polynomial time, while the T-partite matching one is NP-hard for $T \geq 3$. However, we should expect the results obtained with a "global" T-partite matching to be more accurate, since this latter approach can (potentially) identify and exploit global features of poles and residues, as opposed to just local ones. Notably, one could, in theory, combine the matching and interpolation steps into one, trying to fit poles and residues with surrogates defined globally over the parameter domain without having to match them first. The main obstacle to this approach is, as already mentioned, the computational complexity. Still, it might be feasible to tackle this matching problem using some ideas from machine learning, whose abilities in pattern recognition are often praised. In particular, we believe that useful tools could be found in the "clustering" and "mixture models" literature, see, e.g., [Bis06].

Related to this, we find our third envisioned line of further inquiry: is the matching and interpolation of poles and residues the best we can do in marginalized pMOR? We have presented several alternatives in Section 6.2.2, each with its own pro and cons. Notably, our selected method based on the simple partial fraction decomposition (6.12) is a special case of the one involving local reduced system matrices, described in Section 6.2.2.3. This latter approach has the advantage of not needing simple poles, but requires the identification of "change of basis" matrices to remove the "freedom introduced by realization", and may struggle in dealing with frequency surrogates of different sizes. Still, due to its generality, we believe it to have great potential. In particular, we note that interpolating the local reduced system matrices is equivalent (under some conditions) to interpolating the coefficients of the numerator and denominator appearing in the rational form of each local surrogate, i.e., the polynomials P_j and Q_j such that $H_j = P_j/Q_j$. Accordingly,

instead of interpolating poles and residues, it might make sense to interpolate numerators and denominators. However, among other issues, we note that numerator and denominator can be both multiplied by an arbitrary constant without changing the rational surrogate, so that such interpolation must necessarily be carried out over a suitable manifold, taking this extra degree of freedom into account.

A Polynomial bases satisfying Assumption 3.4

We start from two auxiliary results, which provide sufficient conditions for the lower and upper bounds in Assumption 3.4 to hold, respectively.

Lemma A.1. *Let $\Gamma \subset \mathbb{C}$ be a piecewise-smooth bounded curve, and take $w : \Gamma \to \mathbb{R}_{\geq 0}$ a weight function, with $\|w\|_{L^1(\Gamma)} = \int_\Gamma w(z)\,|\mathrm{d}l(z)| < \infty$. Consider a family of complex-valued polynomials $\Psi_\infty = \{\psi_i\}_{i=0}^\infty$, hierarchical in the sense that $\deg(\psi_i) = i$ for all i. Moreover, assume that Ψ_∞ is orthogonal over Γ with respect to the w-weighted inner product, i.e.,*

$$\int_\Gamma w(z)\psi_i(z)\overline{\psi_{i'}(z)}\,|\mathrm{d}l(z)| = \gamma_i \delta_{ii'} \qquad \forall i, i' = 0, 1, \ldots.$$

If there exist two positive constants a and b, such that $\gamma_i \geq ab^i$ for all $i = 0, 1, \ldots$, then $\Psi_N = \{\psi_i\}_{i=0}^N$ satisfies the lower bound in Assumption 3.4 for all $z_0 \in \mathbb{C}$, for some (z_0-dependent) c^{z_0} and ρ^{z_0}, whose values are given explicitly in the proof. Namely,

$$|Q(z)| \geq (c^{z_0})^N \prod_{j=1}^{N'} \frac{|z - z_j|}{\rho^{z_0} + |z_0 - z_j|} \quad \forall Q \in \mathbb{P}_N^{\Psi_N, z_0}(\mathbb{C}; \mathbb{C})\ \forall z \in \mathbb{C}\ \forall N \in \mathbb{N},$$

where $\{z_j\}_{j=1}^{N'} \subset \mathbb{C} \setminus \{z_0\}$ are the roots of Q (repeated according to multiplicity).

Proof. Let $N \geq 1$ and z_0 be arbitrary, but fixed. Take $Q \in \mathbb{P}_N^{\Psi_N, z_0}$ a polynomial with roots $\{z_j\}_{j=1}^{N'} \subset \mathbb{C} \setminus \{z_0\}$ (repeated according to multiplicity). Let $Q = \sum_{i=0}^N q_i \psi_i$ and $\omega(z) = \prod_{j=1}^{N'}(z - z_j)$, so that $Q = \tau\omega$ for some $\tau \in \mathbb{C} \setminus \{0\}$. Then, by orthogonality of Ψ_N,

$$\begin{aligned}\int_\Gamma w(z)\,|Q(z)|^2\,|\mathrm{d}l(z)| = \sum_{i,i'=0}^N q_i \overline{q}_{i'} \int_\Gamma w(z)\psi_i(x)\overline{\psi_{i'}(x)}\,|\mathrm{d}l(z)| = \sum_{i=0}^N \gamma_i\,|q_i|^2 \\ \geq a\min\{1, b\}^N \sum_{i=0}^N |q_i|^2 = a\min\{1, b\}^N.\end{aligned}$$

On the other hand, given $\rho^{z_0} = \max_{z\in\Gamma} |z - z_0|$, the triangular inequality gives

$$\int_\Gamma w(z)\,|Q(z)|^2\,|\mathrm{d}l(z)| = |\tau|^2 \int_\Gamma w(z) \prod_{j=1}^{N'} |z - z_j|^2\,|\mathrm{d}l(z)|$$

$$\begin{aligned}&\leq |\tau|^2 \int_\Gamma w(z) \prod_{j=1}^{N'} \left(|z_j - z_0| + \rho^{z_0}\right)^2 |\mathrm{d}l(z)| \\ &= |\tau|^2 \prod_{j=1}^{N'} \left(|z_j - z_0| + \rho^{z_0}\right)^2 \|w\|_{L^1(\Gamma)},\end{aligned}$$

so that

$$|\tau| \geq \left(\frac{a \min\{1,b\}^N}{\|w\|_{L^1(\Gamma)}}\right)^{1/2} \prod_{j=1}^{N'} \frac{1}{\rho^{z_0} + |z_0 - z_j|}.$$

By $|Q| = |\tau|\,|\omega|$, the lower bound in (3.18) follows, with

$$c^{z_0} = \left(\min\{1,b\} \min\left\{1, \frac{a}{\|w\|_{L^1(\Gamma)}}\right\}\right)^{1/2}$$

(this value is chosen so that $(a \min\{1,b\}^N / \|w\|_{L^1(\Gamma)})^{1/2} \geq (c^{z_0})^N$ for all $N \geq 1$). □

Lemma A.2. *Consider a family of complex-valued polynomials $\{\psi_i\}_{i=0}^\infty$, hierarchical in the sense that $\deg(\psi_i) = i$ for all i. Given a fixed $z_0 \in \mathbb{C}$, assume that $\sum_{i=0}^N |\psi_i(z_0)|^2 \leq (C^{z_0})^{2N}$ for some C^{z_0}, for all $N = 0, 1, \ldots$. Then $\Psi_N = \{\psi_i\}_{i=0}^N$ satisfies the upper bound in Assumption 3.4 for the given z_0 and C^{z_0}. Namely,*

$$|Q(z)| \leq (C^{z_0})^N \prod_{j=1}^{N'} \left|\frac{z - z_j}{z_0 - z_j}\right| \quad \forall Q \in \mathbb{P}_N^{\Psi_N, z_0}(\mathbb{C};\mathbb{C})\ \forall z \in \mathbb{C}\ \forall N \in \mathbb{N},$$

where $\{z_j\}_{j=1}^{N'} \subset \mathbb{C} \setminus \{z_0\}$ are the roots of Q (repeated according to multiplicity).

Proof. Let $N \geq 1$ be arbitrary, but fixed. Take $Q \in \mathbb{P}_N^{\Psi_N, z_0}$ a polynomial with roots $\{z_j\}_{j=1}^{N'} \subset \mathbb{C} \setminus \{z_0\}$ (repeated according to multiplicity). Let $Q = \sum_{i=0}^N q_i \psi_i$ and $\omega(z) = \prod_{j=1}^{N'}(z - z_j)$, so that $Q = \tau\omega$ for some $\tau \in \mathbb{C} \setminus \{0\}$. Then,

$$|Q(z)| = |Q(z_0)| \left|\frac{Q(z)}{Q(z_0)}\right| = |Q(z_0)| \prod_{j=1}^{N'} \left|\frac{z - z_j}{z_0 - z_j}\right|$$

(note that $|Q(z_0)| \neq 0$ since z_0 is not a root of Q). Now, by the Cauchy-Schwarz inequality, we have

$$|Q(z_0)| = \left|\sum_{i=0}^N q_i \psi_i(z_0)\right| \leq \left(\sum_{i=0}^N |q_i|^2\right)^{1/2} \left(\sum_{i=0}^N |\psi_i(z_0)|^2\right)^{1/2} \leq \left(\sum_{i=0}^N |\psi_i(z_0)|^2\right)^{1/2} \leq (C^{z_0})^N.$$

The claim follows. □

Now we present some specific instances of bases satisfying Lemmas A.1 and A.2. Consider the Chebyshev polynomials over $A = [-1, 1]$:

$$\begin{cases} \psi_0(z) = 1, \quad \psi_1(z) = z, \\ \psi_{i+2}(z) = 2z\psi_{i+1}(z) - \psi_i(z) \quad \text{for } i \geq 0. \end{cases}$$

It is well known that they satisfy the orthogonality relation

$$\int_{-1}^{1} \frac{\psi_i(x)\overline{\psi_{i'}(x)}}{\sqrt{1-x^2}}\mathrm{d}x = \begin{cases} \pi & \text{if } i = i' = 0, \\ \frac{\pi}{2} & \text{if } i = i' > 0, \\ 0 & \text{if } i \neq i', \end{cases}$$

so that $\gamma_i \geq \pi \cdot 2^{-i}$ for all $i \geq 0$. Moreover, their values at 0 are

$$\psi_i(0) = \begin{cases} (-1)^{i/2} & \text{if } i \text{ is even}, \\ 0 & \text{if } i \text{ is odd}, \end{cases}$$

so that $\sum_{i=0}^{N} |\psi_i(z_0)|^2 = \lfloor \frac{N}{2} \rfloor + 1 \leq 2^{N/2}$ for all $N \geq 0$. Hence, they satisfy Assumption 3.4 with $z_0 = 0$, $\rho_0 = 1$, $c^{z_0} = \pi^{-1/2}$, and $C^{z_0} = 2^{1/4}$. More generally, due to the boundedness of the Chebyshev basis, the same can be said for any $z_0 \in [-1, 1]$, possibly with a larger value of C^{z_0}.

Similar considerations hold for the Legendre polynomials over $A = [-1, 1]$:

$$\begin{cases} \psi_0(z) = 1, \quad \psi_1(z) = z, \\ \psi_{i+2}(z) = \frac{2i+3}{i+2} z\psi_{i+1}(z) - \frac{i+1}{i+2}\psi_i(z) \quad \text{for } i \geq 0. \end{cases}$$

Indeed, they also satisfy an orthogonality relation:

$$\int_{-1}^{1} \psi_i(x)\overline{\psi_{i'}(x)}\mathrm{d}x = \begin{cases} \frac{2}{2i+1} & \text{if } i = i', \\ 0 & \text{if } i \neq i', \end{cases}$$

so that $\gamma_i \geq 2 \cdot 3^{-i}$ for all $i \geq 0$. Moreover, their values at 0 are

$$\psi_i(0) = \begin{cases} \frac{(-1)^{i/2}}{2^i}\binom{i}{i/2} & \text{if } i \text{ is even}, \\ 0 & \text{if } i \text{ is odd}, \end{cases}$$

so that[1] $\sum_{i=0}^{N} |\psi_i(z_0)|^2 \leq (5/4)^{N/2}$ for all $N \geq 0$. Hence, they satisfy Assumption 3.4 with $z_0 = 0$, $\rho_0 = 1$, $c^{z_0} = 3^{-1/2}$, and $C^{z_0} = (5/4)^{1/4}$. More generally, due to the boundedness of the Legendre basis, the same can be said for any $z_0 \in [-1, 1]$, possibly with a larger value of C^{z_0}.

Shifts and dilations of (the independent variable of) the polynomial basis can be handled by employing the following result.

Lemma A.3. *Assume that $\Psi_N = \{\psi_i\}_{i=0}^N$ satisfies Assumption 3.4 for some z_0, ρ^{z_0}, c^{z_0}, and C^{z_0}. Then, given $a, b \in \mathbb{C}$, with $a \neq 0$, $\bar{\Psi}_N = \{\bar{\psi}_i : z \mapsto \psi_i(az + b)\}_{i=0}^N$ also satisfies Assumption 3.4 for $\bar{z}_0 = (z_0 - b)/a$, $\bar{\rho}^{z_0} = \rho^{z_0}/|a|$, $\bar{c}^{z_0} = c^{z_0}$, and $\bar{C}^{z_0} = C^{z_0}$.*

Proof. It suffices to apply a change of variable in (3.18). □

[1] For conciseness, we skip a proof of this fact, which can be easily obtained by induction.

Bibliography

[ABC12] J. M. Ash, A. Berele, and S. Catoiu. “Plausible and Genuine Extensions of L’Hospital’s Rule”. In: *Mathematics Magazine* 85.1 (2012), pp. 52–60.

[ABG20] A. C. Antoulas, C. A. Beattie, and S. Gugercin. *Interpolatory Methods for Model Reduction*. SIAM, 2020. DOI: 10.1137/1.9781611976083.

[AF08] D. Amsallem and C. Farhat. “Interpolation method for adapting reduced-order models and application to aeroelasticity”. In: *AIAA Journal* 46.7 (2008), pp. 1803–1813. DOI: 10.2514/1.35374.

[AF11] D. Amsallem and C. Farhat. “An online method for interpolating linear parametric reduced-order models”. In: *SIAM Journal on Scientific Computing* 33.5 (2011), pp. 2169–2198. DOI: 10.1137/100813051.

[AKT14] A. P. Austin, P. Kravanja, and L. N. Trefethen. “Numerical Algorithms Based on Analytic Function Values at Roots of Unity”. In: *SIAM Journal on Numerical Analysis* 52.4 (2014), pp. 1795–1821. DOI: 10.1137/130931035.

[Aln+15] M. Alnæs et al. “The FEniCS Project version 1.5”. In: *Archive of Numerical Software* 3.100 (2015).

[Als+19] F. Alsayyari et al. “A nonintrusive reduced order modelling approach using Proper Orthogonal Decomposition and locally adaptive sparse grids”. In: *Journal of Computational Physics* 399 (2019). DOI: 10.1016/j.jcp.2019.108912.

[Ant05] A. C. Antoulas. *Approximation of large-scale dynamical systems*. Advances in design and control. Philadelphia, Pa: SIAM, 2005.

[ASU17] M. Ali, K. Steih, and K. Urban. “Reduced basis methods with adaptive snapshot computations”. In: *Advances in Computational Mathematics* 43.2 (2017), pp. 257–294. DOI: 10.1007/s10444-016-9485-9.

[Bar+04] M. Barrault et al. “An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations”. In: *Comptes Rendus Mathematique* 339.9 (2004), pp. 667–672. DOI: 10.1016/J.CRMA.2004.08.006.

[Bau+11] U. Baur et al. “Interpolatory Projection Methods for Parameterized Model Reduction”. In: *SIAM Journal on Scientific Computing* 33.5 (2011), pp. 2489–2518. DOI: 10.1137/090776925.

[Bay+20] S. K. Baydoun et al. “A greedy reduced basis scheme for multifrequency solution of structural acoustic systems”. In: *International Journal for Numerical Methods in Engineering* 121.2 (2020), pp. 187–200. DOI: 10.1002/nme.6205.

[BB09] U. Baur and P. Benner. “Modellreduktion für parametrisierte Systeme durch balanciertes Abschneiden und Interpolation”. In: *At-Automatisierungstechnik* 57.8 (2009), pp. 411–419. DOI: 10.1524/auto.2009.0787.

[BF14] P. Benner and L. Feng. "A Robust Algorithm for Parametric Model Order Reduction Based on Implicit Moment Matching". In: *Reduced Order Methods for Modeling and Computational Reduction*. Springer International Publishing, 2014, pp. 159–185. DOI: 10.1007/978-3-319-02090-7_6.

[BG17] C. A. Beattie and S. Gugercin. "Model Reduction by Rational Interpolation". In: *Model Reduction and Approximation*. Ed. by P. Benner et al. SIAM, 2017. DOI: 10.1137/1.9781611974829.ch7.

[BGM96] G. A. Baker and P. R. Graves-Morris. *Padé approximants*. 2nd. Cambridge University Press, 1996.

[BHM18] P. Benner, C. Himpe, and T. Mitchell. "On reduced input-output dynamic mode decomposition". In: *Advances in Computational Mathematics* 44.6 (2018), pp. 1751–1768. DOI: 10.1007/s10444-018-9592-x.

[Bis06] C. M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. New York: Springer, 2006.

[BLM18] B. Beckermann, G. Labahn, and A. C. Matos. "On rational functions without Froissart doublets". In: *Numerische Mathematik* 138.3 (2018), pp. 615–633. DOI: 10.1007/s00211-017-0917-3.

[BNP18] F. Bonizzoni, F. Nobile, and I. Perugia. "Convergence analysis of Padé approximations for Helmholtz frequency response problems". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 52.4 (2018), pp. 1261–1284. DOI: 10.1051/m2an/2017050.

[BNR00] V. Barthelmann, E. Novak, and K. Ritter. "High dimensional polynomial interpolation on sparse grids". In: *Advances in Computational Mathematics* 12.4 (2000), pp. 273–288. DOI: 10.1023/A:1018977404843.

[Bon+20a] F. Bonizzoni et al. "Fast Least-Squares Padé approximation of problems with normal operators and meromorphic structure". In: *Mathematics of Computation* 89.323 (2020), pp. 1229–1257. DOI: 10.1090/mcom/3511.

[Bon+20b] F. Bonizzoni et al. "Least-Squares Padé approximation of parametric and stochastic Helmholtz maps". In: *Advances in Computational Mathematics* 46.3 (2020), p. 46. DOI: 10.1007/s10444-020-09749-3.

[BP19] F. Bonizzoni and D. Pradovera. "Distributed sampling for rational approximation of the acoustic scattering of an airfoil". In: *PAMM* 19 (2019). DOI: 10.1002/pamm.201900422.

[BP21] F. Bonizzoni and D. Pradovera. "Shape optimization for a noise reduction problem by non-intrusive parametric reduced modeling". In: *WCCM-ECCOMAS2020 Proceedings* (2021). DOI: 10.23967/wccm-eccomas.2020.300.

[BT04] J.-P. Berrut and L. N. Trefethen. "Barycentric Lagrange Interpolation". In: *SIAM Review* 46.3 (2004), pp. 501–517. DOI: 10.1137/s0036144502417715.

[CCS14] A. Chkifa, A. Cohen, and C. Schwab. "High-Dimensional Adaptive Sparse Polynomial Interpolation and Applications to Parametric PDEs". In: *Foundations of Computational Mathematics* 14.4 (2014), pp. 601–633. DOI: 10.1007/s10208-013-9154-z.

[CL11] A. Cuyt and W. S. Lee. "Sparse interpolation of multivariate rational functions". In: *Theoretical Computer Science* 412.16 (2011), pp. 1445–1456. DOI: 10.1016/j.tcs.2010.11.050.

[Cla76] G. Claessens. "The rational Hermite interpolation problem and some related recurrence formulas". In: *Computers & Mathematics with Applications* 2.2 (1976), pp. 117–123. DOI: 10.1016/0898-1221(76)90023-7.

[Coo+20] R. Cools et al. "Fast component-by-component construction of lattice algorithms for multivariate approximation with POD and SPOD weights". In: *Mathematics of Computation* 90.328 (2020), pp. 787–812. DOI: 10.1090/mcom/3586.

[Cro16] D. F. Crouse. "On implementing 2D rectangular assignment algorithms". In: *IEEE Transactions on Aerospace and Electronic Systems* 52.4 (2016), pp. 1679–1696. DOI: 10.1109/TAES.2016.140952.

[Dan+04] L. Daniel et al. "A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 23.5 (2004), pp. 678–693. DOI: 10.1109/TCAD.2004.826583.

[DEF09] R. Dyczij-Edlinger and O. Farle. "Finite element analysis of linear boundary value problems with geometrical parameters". In: *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* 28.4 (2009), pp. 779–794. DOI: 10.1108/03321640910958919.

[Dör96] W. Dörfler. "A convergent adaptive algorithm for Poisson's equation". In: *SIAM Journal on Numerical Analysis* 33.3 (1996), pp. 1106–1124. DOI: 10.1137/0733054.

[DVW10] J. Degroote, J. Vierendeels, and K. Willcox. "Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis". In: *International Journal for Numerical Methods in Fluids* 63.2 (2010), pp. 207–230. DOI: 10.1002/fld.2089.

[Eva10] L. C. Evans. *Partial differential equations*. 2nd ed. Vol. volume 19, Graduate studies in mathematics. Providence Rhode Island: American Mathematical Society, 2010.

[FKD11] F. Ferranti, L. Knockaert, and T. Dhaene. "Passivity-preserving parametric macromodeling by means of scaled and shifted state-space systems". In: *IEEE Transactions on Microwave Theory and Techniques* 59.10 PART 1 (2011), pp. 2394–2403. DOI: 10.1109/TMTT.2011.2164551.

[Fre03] R. W. Freund. "Model reduction methods based on Krylov subspaces". In: *Acta Numerica* 12.2003 (2003), pp. 267–319. DOI: 10.1017/S0962492902000120.

[GMSV84] P. R. Graves-Morris, E. B. Saff, and R. S. Varga. *Rational Approximation and Interpolation*. Lecture Notes in Mathematics. 1984.

[GPT11] P. Gonnet, R. Pachón, and L. Trefethen. "Robust rational interpolation and least-squares". In: *Electronic Transactions on Numerical Analysis* 38 (2011), pp. 146–167.

[Gri97] E. Grimme. "Krylov projection methods for model reduction". PhD thesis. University of Illinois at Urbana-Champaign, 1997.

[GS99] B. Gustavsen and A. Semlyen. "Rational approximation of frequency domain responses by vector fitting". In: *IEEE Transactions on Power Delivery* 14.3 (1999), pp. 1052–1061. DOI: 10.1109/61.772353.

[GT01] D. Gilbarg and N. S. Trudiger. *Elliptic Partial Differential Equations of Second Order*. Springer Berlin Heidelberg, 2001.

[GTG15] S. Grivet-Talocia and B. Gustavsen. *Passive Macromodeling: Theory and Applications*. Wiley series in microwave and optical engineering. Hoboken New Jersey: John Wiley & Sons, Inc, 2015, p. 904. DOI: 10.1002/9781119140931.

[GTT18] S. Grivet-Talocia and R. Trinchero. "Behavioral, Parameterized, and Broadband Modeling of Wired Interconnects with Internal Discontinuities". In: *IEEE Transactions on Electromagnetic Compatibility* 60.1 (2018), pp. 77–85. DOI: 10.1109/TEMC.2017.2723629.

[Gui99] P. Guillaume. "Nonlinear eigenproblems". In: *SIAM Journal on Matrix Analysis and Applications* 20.3 (1999), pp. 575–595. DOI: 10.1137/S0895479897324172.

[GW08] S. Gugercin and K. Willcox. "Krylov projection framework for Fourier model reduction". In: *Automatica* 44.1 (2008), pp. 209–215. DOI: 10.1016/j.automatica.2007.05.007.

[Hal64] J. H. Halton. "Algorithm 247: Radical-inverse quasi-random point sequence". In: *Communications of the ACM* 7.12 (1964), pp. 701–702. DOI: 10.1145/355588.365104.

[HDO11] B. Haasdonk, M. Dihlmann, and M. Ohlberger. "A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space". In: *Mathematical and Computer Modelling of Dynamical Systems* 17.4 (2011), pp. 423–442. DOI: 10.1080/13873954.2011.547674.

[HKD12] K. Hormann, G. Klein, and S. De Marchi. "Barycentric rational interpolation at quasi-equidistant nodes". In: *Dolomites Research Notes on Approximation* 5 (2012), pp. 1–6.

[HMT09] N. J. Higham, D. S. Mackey, and F. Tisseur. "Definite Matrix Polynomials and their Linearization by Definite Pencils". In: *SIAM Journal on Matrix Analysis and Applications* 31.2 (2009), pp. 478–502.

[Huy+07] D. B. P. Huynh et al. "A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants". In: *Comptes Rendus Mathematique* 345.8 (2007), pp. 473–478. DOI: 10.1016/j.crma.2007.09.019.

[Huy+10] D. B. P. Huynh et al. "A natural-norm Successive Constraint Method for inf-sup lower bounds". In: *Computer Methods in Applied Mechanics and Engineering* 199.29-32 (2010), pp. 1963–1975. DOI: 10.1016/j.cma.2010.02.011.

[IA14] A. C. Ionita and A. C. Antoulas. "Data-driven parametrized model reduction in the Loewner framework". In: *SIAM Journal on Scientific Computing* 36.3 (2014), A984–A1007. DOI: 10.1137/130914619.

[JIR14] C. Jäggli, L. Iapichino, and G. Rozza. "An improvement on geometrical parameterizations by transfinite maps". In: *Comptes Rendus Mathematique* 352.3 (2014), pp. 263–268. DOI: 10.1016/j.crma.2013.12.017.

[Kar72] R. M. Karp. "Reducibility Among Combinatorial Problems". In: *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA*. Ed. by R. E. Miller and J. W. Thatcher. The IBM Research Symposia Series. Plenum Press, New York, 1972, pp. 85–103. DOI: 10.1007/978-1-4684-2001-2_9.

[Kat95] T. Kato. *Perturbation Theory for Linear Operators*. Vol. 132. Classics in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995. DOI: 10.1007/978-3-642-66282-9.

[KL07] O. Koch and C. Lubich. "Dynamical low-rank approximation". In: *SIAM Journal on Matrix Analysis and Applications* 29.2 (2007), pp. 434–454. DOI: 10.1137/050639703.

[Kle12] G. Klein. "Applications of Linear Barycentric Rational Interpolation". PhD thesis. University of Fribourg, 2012.

[Kuo+21] F. Y. Kuo et al. “Function integration, reconstruction and approximation using rank-1 lattices”. In: *Mathematics of Computation* 90.330 (2021), pp. 1861–1897. DOI: 10.1090/mcom/3595.

[LAI11] S. Lefteriu, A. C. Antoulas, and A. C. Ionita. “Parametric model reduction in the Loewner framework”. In: *IFAC Proceedings Volumes (IFAC-PapersOnline)* 44.1 PART 1 (2011), pp. 12751–12756. DOI: 10.3182/20110828-6-IT-1002.02651.

[LEP09] B. Lohmann, R. Eid, and H. Panzer. *Efficient Order Reduction of Parametric and Nonlinear Models by Superposition of Locally Reduced Models*. 2009.

[LR10] T. Lassila and G. Rozza. “Parametric free-form shape design with PDE models and reduced basis method”. In: *Computer Methods in Applied Mechanics and Engineering* 199.23-24 (2010), pp. 1583–1592. DOI: 10.1016/j.cma.2010.01.007.

[Lub03] D. S. Lubinsky. “Rogers-Ramanujan and the Baker-Gammel-Wills (Padé) conjecture”. In: *Annals of Mathematics* 157.3 (2003), pp. 847–889. DOI: 10.4007/annals.2003.157.847.

[Mat08] T. P. A. Mathew. *Domain decomposition methods for the numerical solution of partial differential equations*. Vol. 61. Lecture notes in computational science and engineering. Berlin: Springer, 2008.

[MGT20] P. Manfredi and S. Grivet-Talocia. “Rational Polynomial Chaos Expansions for the Stochastic Macromodeling of Network Responses”. In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 67.1 (2020), pp. 225–234. DOI: 10.1109/TCSI.2019.2942109.

[MN15] G. Migliorati and F. Nobile. “Analysis of discrete least squares on multivariate polynomial spaces with evaluations at low-discrepancy point sets”. In: *Journal of Complexity* 31.4 (2015), pp. 517–542. DOI: 10.1016/j.jco.2015.02.001.

[MZ09] X. Ma and N. Zabaras. “An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations”. In: *Journal of Computational Physics* 228.8 (2009), pp. 3084–3113. DOI: 10.1016/j.jcp.2009.01.006.

[NP21] F. Nobile and D. Pradovera. “Non-intrusive double-greedy parametric model reduction by interpolation of frequency-domain rational surrogates”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 55.5 (2021), pp. 1895–1920. DOI: 10.1051/m2an/2021040.

[NST18] Y. Nakatsukasa, O. Sète, and L. N. Trefethen. “The AAA algorithm for rational approximation”. In: *SIAM Journal on Scientific Computing* 40.3 (2018), A1494–A1522. DOI: 10.1137/16M1106122.

[NTW08] F. Nobile, R. Tempone, and C. G. Webster. “A sparse grid stochastic collocation method for partial differential equations with random input data”. In: *SIAM Journal on Numerical Analysis* 46.5 (2008), pp. 2309–2345. DOI: 10.1137/060663660.

[Pan+10] H. Panzer et al. “Parametric model order reduction by matrix interpolation”. In: *At-Automatisierungstechnik* 58.8 (2010), pp. 475–484. DOI: 10.1524/auto.2010.0863.

[PN20] D. Pradovera and F. Nobile. “Frequency-domain non-intrusive greedy Model Order Reduction based on minimal rational approximation”. To appear in: *SCEE 2020 Proceedings*. DOI: 10.5075/epfl-MATHICSE-275533. 2020.

[PN21] D. Pradovera and F. Nobile. “A technique for non-intrusive greedy piecewise-rational model reduction of frequency-domain problems over wide frequency bands”. In preparation. 2021.

[PPB10] D. Pflüger, B. Peherstorfer, and H. J. Bungartz. "Spatially adaptive sparse grids for high-dimensional data-driven problems". In: *Journal of Complexity* 26.5 (2010), pp. 508–522. DOI: 10.1016/j.jco.2010.04.001.

[Pra20] D. Pradovera. "Interpolatory rational model order reduction of parametric problems lacking uniform inf-sup stability". In: *SIAM Journal on Numerical Analysis* 58.4 (2020), pp. 2265–2293. DOI: 10.1137/19M1269695.

[Pra21] D. Pradovera. *Model order reduction based on functional rational approximants for parametric PDEs with meromorphic structure – Numerical tests*. 2021. DOI: 10.5281/zenodo.5358444.

[QMN15] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced basis methods for partial differential equations: An introduction*. UNITEXT. Springer International Publishing, 2015, pp. 1–263. DOI: 10.1007/978-3-319-15431-2.

[QSS07] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Vol. 37. Texts in Applied Mathematics. New York, NY: Springer New York, 2007. DOI: 10.1007/B98885.

[Qua09] A. Quarteroni. *Numerical Models for Differential Problems*. Springer, 2009. DOI: 10.1007/978-88-470-1071-0.

[Ram86] A. G. Ramm. *Scattering by obstacles*. Mathematics and its applications. Reidel, 1986.

[RHP08] G. Rozza, D. B. P. Huynh, and A. T. Patera. "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: Application to transport and continuum mechanics". In: *Archives of Computational Methods in Engineering* 15.3 (2008), pp. 229–275. DOI: 10.1007/s11831-008-9019-9.

[RM18] V. de la Rubia and M. Mrozowski. "A compact basis for reliable fast frequency sweep via the reduced-basis method". In: *IEEE Transactions on Microwave Theory and Techniques* 66.10 (2018), pp. 4367–4382. DOI: 10.1109/TMTT.2018.2865957.

[RRM09] V. de la Rubia, U. Razafison, and Y. Maday. "Reliable fast frequency sweep for microwave devices via the reduced-basis method". In: *IEEE Transactions on Microwave Theory and Techniques* 57.12 (2009), pp. 2923–2937. DOI: 10.1109/TMTT.2009.2034208.

[Rub14] V. de la Rubia. "Reliable reduced-order model for fast frequency sweep in microwave circuits". In: *Electromagnetics* 34.3-4 (2014), pp. 161–170. DOI: 10.1080/02726343.2014.877735.

[Rug96] W. J. Rugh. *Linear System Theory*. 2nd ed. USA: Prentice-Hall, Inc., 1996.

[Saf72] E. B. Saff. "An extension of Montessus de Ballore's theorem on the convergence of interpolating rational functions". In: *Journal of Approximation Theory* 6.1 (1972), pp. 63–67. DOI: 10.1016/0021-9045(72)90081-0.

[Sal+13] S. Salsa et al. *A primer on PDEs : models, methods, simulations*. UNITEXT. Springer Milan, 2013. DOI: 10.1007/978-88-470-2862-3.

[Ske88] R. E. Skelton. *Dynamic Systems Control: Linear Systems Analysis and Synthesis*. USA: John Wiley & Sons, Inc., 1988.

[Smi06] S. J. Smith. "Lebesgue constants in polynomial interpolation". In: *Annales Mathematicae et Informaticae* 33 (2006), pp. 109–123.

[Son13] N. T. Son. "A real time procedure for affinely dependent parametric model order reduction using interpolation on Grassmann manifolds". In: *International Journal for Numerical Methods in Engineering* 93.8 (2013), pp. 818–833. DOI: 10.1002/nme.4408.

[Son98] E. D. Sontag. *Mathematical Control Theory*. 2nd ed. Vol. 6. Texts in Applied Mathematics. New York, NY: Springer New York, 1998. DOI: 10.1007/978-1-4612-0577-7.

[Spi+15] D. Spina et al. "Polynomial chaos-based macromodeling of multiport systems using an input-output approach". In: *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* 28.5 (2015), pp. 562–581. DOI: 10.1002/jnm.2037.

[Sta97] H. Stahl. "The Convergence of Padé Approximants to Functions with Branch Points". In: *Journal of Approximation Theory* 91.2 (1997), pp. 139–204. DOI: 10.1006/jath.1997.3141.

[Sta98] H. Stahl. "Spurious poles in Padé approximation". In: *Journal of Computational and Applied Mathematics* 99 (1998), pp. 511–527. DOI: 10.1016/S0377-0427(98)00180-0.

[Ste68] S. Steinberg. "Meromorphic families of compact operators". In: *Archive for Rational Mechanics and Analysis* 31.5 (1968), pp. 372–379. DOI: 10.1007/BF00251419.

[SV13] B. Sadiq and D. Viswanath. "Barycentric Hermite interpolation". In: *SIAM Journal on Scientific Computing* 35.3 (2013). DOI: 10.1137/110833221.

[Tit78] E. C. Titchmarsh. *The theory of functions*. Second ed. University Press, 1978.

[Tre09] L. N. Trefethen. "Householder triangularization of a quasimatrix". In: *IMA Journal of Numerical Analysis* 29 (2009).

[Tre13] L. N. Trefethen. *Approximation theory and approximation practice*. Applied mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 2013.

[URL16] S. Ullmann, M. Rotkvic, and J. Lang. "POD-Galerkin reduced-order modeling with adaptive finite element snapshots". In: *Journal of Computational Physics* 325 (2016), pp. 244–258. DOI: 10.1016/j.jcp.2016.08.018.

[Vir+20] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: https://doi.org/10.1038/s41592-019-0686-2.

[Wal29] J. L. Walsh. "The approximation of harmonic functions by harmonic polynomials and by harmonic rational functions". In: *Bulletin of the American Mathematical Society* 35.4 (1929), pp. 499–544. DOI: 10.1090/S0002-9904-1929-04753-0.

[Wal60] J. L. Walsh. *Interpolation and Approximation by Rational Functions in the Complex Domain*. 3rd ed. American Mathematical Society, 1960.

[Wal79] H. Wallin. "Potential theory and approximation of analytic functions by rational interpolation". In: *Complex Analysis Joensuu 1978*. Ed. by I. Laine, O. Lehto, and T. Sorvali. Berlin, Heidelberg: Springer Berlin Heidelberg, 1979, pp. 434–450. DOI: 10.1007/bfb0064002.

[Wei+99] D. S. Weile et al. "A method for generating rational interpolant reduced order models of two-parameter linear systems". In: *Applied Mathematics Letters* 12.5 (1999), pp. 93–102. DOI: 10.1016/S0893-9659(99)00063-4.

[Wen04] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004. DOI: 10.1017/cbo9780511617539.

[Xia+19] Y. Q. Xiao et al. “A Novel Framework for Parametric Loewner Matrix Interpolation”. In: *IEEE Transactions on Components, Packaging and Manufacturing Technology* 9.12 (2019). DOI: 10.1109/tcpmt.2019.2948802.

[YFB18] Y. Yue, L. Fengy, and P. Benner. “Interpolation of Reduced-Order Models Based on Modal Analysis”. In: *2018 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization, NEMO 2018* 5 (2018), pp. 6–9. DOI: 10.1109/NEMO.2018.8503114.

[YFB19a] Y. Yue, L. Feng, and P. Benner. *An Adaptive Pole-Matching Method for Interpolating Reduced-Order Models*. ArXiv preprint: 1908.00820. 2019.

[YFB19b] Y. Yue, L. Feng, and P. Benner. “Reduced-order modelling of parametric systems via interpolation of heterogeneous surrogates”. In: *Advanced Modeling and Simulation in Engineering Sciences* 6.1 (2019). DOI: 10.1186/s40323-019-0134-y.

Davide Pradovera

MA B2 435 (Bâtiment MA), Station 8
CH-1015 Lausanne

Mobile: +41 077 95 88 993

Email: davide.pradovera@epfl.ch
URL: http://people.epfl.ch/davide.pradovera

Born: October 9, 1993—Piacenza, Italy
Nationality: Italian

Current position

Doctoral researcher, CSQI, EPFL, Lausanne

Areas of specialisation

Numerical mathematics for partial differential equations, approximation theory, model order reduction, scattering problems

Appointments held

2014–2017 Special courses teacher, Piacenza (I)
2016 Developer intern, Iren S.p.A., Piacenza (I)
2017–present Doctoral assistant, EPFL, Lausanne (CH)

Education

2013–2015 B.Sc. in Applied Mathematics (*cum laude*), Politecnico di Milano, Milan (I)
Thesis: "A mathematical justification of the momentum operator in quantum mechanics", advisor: Prof. M. Verri

2015–2017 M.Sc. in Computational Science and Engineering, EPFL, Lausanne (CH)
Project: "Implementation of smooth contact mechanics with the mortar method", advisor: Prof. G. Anciaux
Project: "Finite elements-based Padé approximants for Helmholtz frequency response problems", advisor: Prof. F. Nobile
Thesis: "Randomized low-rank approximation of matrices and tensors", advisor: Prof. D. Kressner

2017–present Ph.D. in Mathematics, EPFL, Lausanne (CH)
Thesis: "Model order reduction based on functional rational approximants for parametric PDEs with meromorphic structure", advisor: Prof. F. Nobile

Grants, honours, & awards

2011 3rd place at the "Hong Kong International Science Fair"
2013&2014 4th & 5th places in the "Championnat International des Jeux Mathématiques et Logiques"
2017 Douchet prize for best GPA, MATH-EPFL

Publications and talks

Journal articles

2019 F. Bonizzoni and D. Pradovera, "Distributed sampling for rational approximation of the acoustic scattering of an airfoil", PAMM 19.
2020 F. Bonizzoni, F. Nobile, I. Perugia, and D. Pradovera, "Fast Least-Squares Padé approximation of problems with normal operators and meromorphic structure", Math. Comput. 89.
F. Bonizzoni, F. Nobile, I. Perugia, and D. Pradovera, "Least-Squares Padé approximation of parametric and stochastic Helmholtz maps", Adv. Comput. Math. 46.
D. Pradovera, "Minimal rational model order reduction of parametric problems lacking uniform inf-sup stability", SIAM J. Numer. Anal. 58.
2021 F. Bonizzoni and D. Pradovera, "Shape optimization for a noise reduction problem by non-intrusive parametric reduced modeling", Proc. WCCM-ECCOMAS2020.
F. Nobile and D. Pradovera, "Non-intrusive double-greedy parametric model reduction by interpolation of frequency-domain rational surrogates", ESAIM:M2AN 55.

Pending articles

2020 D. Pradovera and F. Nobile, "Frequency-domain non-intrusive greedy Model Order Reduction based on minimal rational approximation", to appear in SCEE 2020 Proc..
2021 D. Pradovera and F. Nobile, "A technique for non-intrusive greedy piecewise-rational model reduction of frequency response problems over wide frequency bands", under review.

Presentations at conferences

2019 D. Pradovera, F. Nobile, F. Bonizzoni, and I. Perugia, "A technique for rational model order reduction of parametric problems lacking uniform inf-sup stability", GAMM 2019, Vienna (A).
D. Pradovera, F. Nobile, F. Bonizzoni, and I. Perugia, "A technique for rational model order reduction of parametric problems lacking uniform inf-sup stability", ICIAM 2019, Valencia (E).
D. Pradovera and F. Nobile, "Interpolatory rational model order reduction of parametric problems lacking uniform inf-sup stability", ENUMATH 2019, Egmond aan Zee (NL).
2021 D. Pradovera, F. Nobile, and F. Bonizzoni, "Non-intrusive model reduction of parametric frequency response problems via minimal rational interpolation", ICOSAHOM 2020/2021 (virtual), Vienna (A).
D. Pradovera and F. Nobile, "Non-intrusive model reduction of parametric frequency-response problems – with applications to UQ", SIMAI 2020+2021, Parma (I).

Others

2018 F. Bonizzoni, I. Perugia, F. Nobile, and D. Pradovera, "An efficient algorithm for Padé-type approximation of the frequency response for the Helmholtz problem", poster, MoRePaS IV, Nantes (F).
F. Bonizzoni, I. Perugia, F. Nobile, and D. Pradovera, "An efficient algorithm for Padé-type approximation of the frequency response for the Helmholtz problem", poster, Swiss Numerics Day 2018, Zurich (CH).

2020 D. Pradovera and F. Nobile, "Frequency-domain non-intrusive greedy Model Order Reduction based on minimal rational approximation", poster, SCEE 2020, Eindhoven (NL).
D. Pradovera, "Padé approximation: a quick overview", seminar (virtual), CSQI talks, Lausanne (CH).
D. Pradovera, "From Padé approximation to rational interpolation", seminar (virtual), CSQI talks, Lausanne (CH).
D. Pradovera and F. Nobile, "Frequency-domain non-intrusive greedy Model Order Reduction based on minimal rational approximation", poster (virtual), MORSS 2020, Lausanne (CH).
D. Pradovera, "Minimal rational approximation", seminar (virtual), CSQI talks, Lausanne (CH).
D. Pradovera, "Minimal rational approximation: a model reduction tool for parametrized PDEs with resonances", seminar (virtual), PDE Afternoons, Vienna (A).

2021 D. Pradovera, "Matching-based pMOR for dynamical systems", seminar (virtual), CSQI talks, Lausanne (CH).

Teaching experience

2017 Analyse avancée I, Mathematics, EPFL
2018 Analyse numerique, Mechanical Engineering, EPFL
Analyse fonctionnelle, Mathematics, EPFL
2019 Introduction to partial differential equations, Mathematics, EPFL
2019–2021 Parallel and high-performance computing, Computational Sciences, EPFL

Other experiences

2020 Conference organizer, Model Order Reduction Summer School 2020 (virtual event).

Computer skills

Advanced Matlab, C/C++, OpenMP, MPI, Python, FreeFem++, LaTeX
Intermediate CUDA, C#, HTML
Basic R, OpenFOAM, Fluent, Fortran, Java

Languages

Italian:	Mothertongue	English:	Fluent
French:	Intermediate	Japanese:	Basic

Ce document a été imprimé au Centre d’impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.