Towards fully automated high-dimensional parameterized macromodeling

Towards fully automated high-dimensional parameterized macromodeling

Alessandro Zanco, *Graduate Student Member, IEEE*, Stefano Grivet-Talocia, *Fellow, IEEE*

***Abstract*—This paper presents a fully automated algorithm for the extraction of parameterized macromodels from frequency responses. The reference framework is based on a frequency-domain rational approximation combined with a parameter-space expansion into Gaussian Radial Basis Functions (RBF). An iterative least-squares fitting with positivity constraints is used to optimize model coefficients, with a guarantee of uniform stability over the parameter space. The main novel contribution of this work is a set of algorithms, supported by strong theoretical arguments with associated proofs, for the automated determination of all the hyper-parameters that define model orders, placement and width of RBFs. With respect to standard approaches, which tune these parameters using time-consuming tentative model extractions following a trial-and-error strategy, the presented technique allows much faster tuning of the model structure. The numerical results show that models with up to ten independent parameters are easily extracted in few minutes.**

I. Introduction

Compact macromodeling techniques are quite common in Computer-Aided Design (CAD) flows in the electronic industry. Complex system-level verification via numerical simulation may be drastically simplified by using macromodels, resulting in major speedup in runtime at almost no loss of accuracy [1]. Therefore, reliable and robust algorithms for macromodel extraction are required to enable such optimized workflows. This paper presents a number of new results and related algorithms in this context.

Most recent macromodeling approaches were developed for passive device and interconnects [2]–[5]. Among the data-driven model extraction methods on which this work is focused, we mention the popular approaches based on Vector Fitting [4] and Loewner matrix frameworks [6]–[8], [18], which construct a rational approximation to the system's transfer functions through numerically robust algorithms. The resulting mathematical expressions are easily cast as state-space systems of Ordinary Differential Equations (ODEs) and can be synthesized as SPICE equivalent netlists [9], compatible with any standard circuit solver.

The remarkable robustness of these algorithms led to a number of notable extensions in the recent years. Examples are distributed circuits modeled through delay-rational functions [10]–[13], and multivariate extensions allowing to embed in the model a closed-form dependence on external parameters, possibly related to geometry, materials, ambient, or bias conditions [1], [14]–[16]. This work concentrates on the latter extension, by providing a novel and fully-automated approach for model extraction, which has the potential to scale very favorably with the number of independent parameters.

We assume that the responses of the devices under modeling can be accurately reproduced by the class of Linear Time Invariant (LTI) systems. Therefore, the proposed approach is applicable both to natively linear structures such as all types of interconnects, but also to active nonlinear devices represented by small-signal transfer functions. The proposed approach builds on several known results and assumptions. For instance, as model functional form we exploit the so-called rational barycentric model structure, originally introduced in [17]–[19] and extended to the parameterized framework in [16].

Standard parameterized macromodeling schemes have proven to be extremely effective when dealing with few independent parameters (up to five) but, on the other hand, have shown poor scalability properties in case of larger dimensions. In [20], [21], the Authors proposed a novel parameterized macromodeling framework with excellent properties in terms of model compactness, breaking the dependence of the model complexity on the parameter space dimension. The main enabling factor is the adoption of an unstructured approach based on Radial Basis Functions (RBFs), see e.g. [22], [23], to capture parametric variability in the model structure.

The main drawback is that an RBF-based approach requires the selection of a number of free *hyper-parameters*, whose optimization is not trivial. Indeed, considering the class of Gaussian RBFs, the approximation capabilities strongly depend upon their shape (how "fat" or "thin" are the RBFs), as well as the number and position of their centers. Concerning the optimization of the RBF shape, classical techniques range from simple grid search [20], [21] to a more sophisticated optimization of cross validation indicators and maximum likelihood estimators [24]–[26]. These strategies, however, require the extraction of many intermediate macromodels that, in our setting, would require unacceptable computational efforts. In regard to the selection of the RBF number and centers, a common choice is to pick (or generate) data samples from the input space according to some predefined rule [27]–[29]. Other techniques, instead, rely on more advanced schemes, such as clustering algorithms [30]–[32] and forward (or backward) selection through Recursive Orthogonal Least Squares [33].

In the framework of parameterized macromodeling, we have recently proposed a number of heuristic strategies to optimize both the RBFs shape [21] and centers [34]. In these previous works, the free hyper-parameters were deliberately chosen as common between model numerator and denominator in order to simplify the problem. Even though this choice does not seem to affect the model accuracy, it may drastically increase its complexity, still requiring time-consuming repeated model extractions.

This work tries to overcome the above limitations, by:

1) providing a comprehensive and detailed analysis on how the RBF shape affects both the model accuracy and the

numerical conditioning of the associated fitting problem;
2) exploiting these results to set-up a sub-optimal algorithm (already presented in preliminary form in [48]) to choose an adequate RBF shape;
3) describing a simple yet effective iterative procedure aimed at selecting the least necessary number of RBFs as well as their centers, in order to appropriately capture the parametric variability;
4) finally and most importantly, providing a thorough theoretical background in support of the presented strategies.

The proposed automated approach is validated on various linear passive and nonlinear (linearized) active test-cases, with up to ten independent parameters. The results show that compact, accurate, uniformly stable parameterized macromodels are extracted in few minutes, thus validating the overall framework.

II. General Framework

We start by introducing the general framework on which the proposed automated macromodeling flow builds. This framework is not new [16], [20], [35]. We summarize here the main assumptions, definitions, and consolidated formulations, in order to set notation and make this paper reasonably self-contained.

We consider a device, component, interconnect or subsystem with P interface ports, and we assume that a LTI representation is adequate for capturing the dominant physical processes that have an influence on its behavior. The $P \times P$ transfer function of the system depends on frequency $s = \mathrm{j}\omega$ and on ρ additional parameters, collected in vector $\boldsymbol{\vartheta} = [\vartheta_1, \ldots, \vartheta_\rho] \in \Theta$. These parameters may be related to geometry (e.g., width or separation of traces), materials (e.g., relative permittivity), or ambient (e.g., temperature).

We assume an available characterization of the system response through a set of tabulated samples $\breve{\mathbf{H}}_{k,m} = \breve{\mathbf{H}}(s_k, \boldsymbol{\vartheta}_m)$ known at a discrete set of $\bar{k}$ frequency $s_k = \mathrm{j}\omega_k$ and $\bar{m}$ parameter points $\boldsymbol{\vartheta}_m$. For modeling purposes, the set of all available parametric responses corresponding to parameter values in $\mathcal{R} = \{\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{\bar{m}}\}$ is split in mutually disjoint training $\mathcal{T}$ (with cardinality $\bar{m}_t$) and validation $\mathcal{V}$ (with cardinality $\bar{m}_v$) subsets, i.e. $\mathcal{T} \cup \mathcal{V} = \mathcal{R}$. A common scenario enabling this characterization is a multi-dimensional frequency and parameter sweep calling a frequency-domain field solver. Throughout this work, we assume this initial dataset to be precomputed and available, e.g. as a set of Touchstone files.

The main objective is the extraction of a black-box parameterized model $\mathbf{H}(s, \boldsymbol{\vartheta})$, whose frequency- and parameter-dependent response matches as closely as possible the available data

$$\mathbf{H}(s_k, \boldsymbol{\vartheta}_m) \approx \breve{\mathbf{H}}_{k,m} \quad \forall k,\ \forall \boldsymbol{\vartheta}_m \in \mathcal{T} \tag{1}$$

over the parameter space and frequency band of interest. The model should be not only accurate, but it should have a structure allowing a synthesis into an equivalent parameterized SPICE netlist, to be used as a parameterized library component for simulation-based design optimization, what-if, sensitivity, and Monte Carlo analyses.

A. Model structure

We adopt the well consolidated model structure [16]

$$\mathbf{H}(s; \boldsymbol{\vartheta}) = \frac{\mathbf{N}(s; \boldsymbol{\vartheta})}{D(s; \boldsymbol{\vartheta})} = \frac{\sum_{n=0}^{\bar{n}} \sum_{\ell=1}^{\bar{\ell}_N} \mathbf{R}_{n,\ell}\, \xi_\ell(\boldsymbol{\vartheta})\, \varphi_n(s)}{\sum_{n=0}^{\bar{n}} \sum_{\ell=1}^{\bar{\ell}_D} r_{n,\ell}\, \xi_\ell(\boldsymbol{\vartheta})\, \varphi_n(s)}. \tag{2}$$

Frequency dependence is captured by the basis functions $\varphi_n(s)$, here assumed to be the standard partial fraction basis[1] $\varphi_n(s) = (s - q_n)^{-1}$ for $n = 1, \ldots, \bar{n}$ associated with prescribed Vector Fitting (VF) "basis" poles q_n, with $\varphi_0(s) = 1$. Parameter variability is captured by the basis functions $\xi_\ell(\boldsymbol{\vartheta})$, discussed below. Given this model structure, for any choice of the free coefficients $\mathbf{R}_{n,\ell}$, $r_{n,\ell}$, we have that $\mathbf{H}(s; \boldsymbol{\vartheta})$ is a rational function of frequency, which enables a straightforward translation to an equivalent SPICE netlist [9].

B. Model parameterization

This paper specifically addresses high-dimensional parameterized models, where the number ρ of independent parameters can grow beyond few units. With the exception of [20], [21], previous works on parameterized modeling provided only examples with up to $\rho = 2$ and in some cases $\rho = 3$ parameters. This limitation was due to several factors, including model structure and specific choice of parameter-dependent basis functions $\xi_\ell(\boldsymbol{\vartheta})$. In particular, any choice of multivariate polynomials based on a (structured) Cartesian product of univariate polynomials is likely to be poorly scalable to high parameter dimensions. Here, we want to overcome this limitation and provide a formulation that is limited only by the data variability in the parameter space and not on the parameter space dimension.

It has been already shown in [20], [21] that high-dimensional modeling is only possible through mesh-free and unstructured descriptions, such as provided by Radial Basis Functions (RBF). The results in [20] confirm that up to $\rho = 10$ independent parameters can be efficiently handled. Therefore, we will adopt as parameter basis functions the Gaussian RBFs, defined as

$$\xi_\ell^\varepsilon(\boldsymbol{\vartheta}) = e^{-\varepsilon^2 \left\| \boldsymbol{\vartheta} - \mathring{\boldsymbol{\vartheta}}_\ell \right\|^2} \tag{3}$$

where $\mathring{\boldsymbol{\vartheta}}_\ell$ denotes their centers, and the *shape parameter* ε controls their width hence their smoothness. Both number and location of centers and shape parameters are free variables in the determination of the model structure; this paper proposes a general algorithm for their automated determination.

C. Model identification

Given a set of basis functions, with predefined basis poles q_n, RBF centers $\mathring{\boldsymbol{\vartheta}}_\ell$ and shape parameter ε, the free model coefficients $\mathbf{R}_{n,\ell}$, $r_{n,\ell}$ are determined using the well-established *Parameterized Sanathanan Koerner* (PSK) iteration[2] [16], [37]. Model-data error minimization is achieved

[1] Suitable modifications apply in case of complex-valued basis poles $\alpha_n \pm \mathrm{j}\beta_n$.

[2] The PSK framework, as well as standard Vector Fitting, are generally not guaranteed to converge [36]. However, lack of convergence can be usually associated to the presence of large noise components in the data. Assuming sufficient data quality, no convergence issues are expected.

through an iteration $\mu = 1, 2, \ldots$, that solves at each step a linear relaxation of the data fitting (1)

$$\frac{\mathbf{N}^{\mu}(\mathrm{j}\omega_k;\boldsymbol{\vartheta}_m) - D^{\mu}(\mathrm{j}\omega_k;\boldsymbol{\vartheta}_m)\breve{\mathbf{H}}_{k,m}}{D^{\mu-1}(\mathrm{j}\omega_k;\boldsymbol{\vartheta}_m)} \approx 0. \tag{4}$$

A compact matrix notation for (4) is readily obtained by collecting all elements of $\mathbf{R}_{n,\ell}$, $r_{n,\ell}$ at iteration μ in a vector $\mathbf{x}^{\mu}$, obtaining

$$\boldsymbol{\Psi}^{\mu}(\varepsilon)\mathbf{x}^{\mu} \approx \mathbf{0}, \tag{5}$$

to be suitably constrained to avoid the trivial all-zero solution $\mathbf{x}^{\mu} = \mathbf{0}$, e.g. by setting $\|\mathbf{x}^{\mu}\| = 1$. Note that we highlight the dependence on all matrices on ε in order to enable the developments of Sec. IV.

The matrix $\boldsymbol{\Psi}^{\mu}(\varepsilon) \in \mathbb{C}^{\bar{k}\bar{m}_t P^2 \times (\bar{n}+1)(P^2\bar{\ell}_N + \bar{\ell}_D)}$ is defined as

$$\boldsymbol{\Psi}^{\mu}(\varepsilon) = \begin{pmatrix} \boldsymbol{\Gamma}^{\mu}(\varepsilon) & & \boldsymbol{\Xi}_{(1,1)}(\varepsilon) \\ & \ddots & \vdots \\ & \boldsymbol{\Gamma}^{\mu}(\varepsilon) & \boldsymbol{\Xi}_{(P,P)}(\varepsilon) \end{pmatrix} \tag{6}$$

where individual blocks read

$$\begin{aligned} \boldsymbol{\Gamma}^{\mu}(\varepsilon) &= \mathbf{W}^{\mu-1}\mathbf{X}_N(\varepsilon), \\ \boldsymbol{\Xi}^{\mu}_{(i,j)}(\varepsilon) &= -\mathbf{W}^{\mu-1}\breve{\mathbf{H}}_{(i,j)}\mathbf{X}_D(\varepsilon). \end{aligned} \tag{7}$$

The two matrices $\mathbf{W}^{\mu-1}$ and $\breve{\mathbf{H}}_{(i,j)}$ are diagonal and collect, respectively, the values of (known) denominator available from previous iteration $D^{\mu-1}(\mathrm{j}\omega_k, \boldsymbol{\vartheta}_m)$ for all k, m, and all frequency-parameter samples of the response $H_{i,j}(\mathrm{j}\omega_k, \boldsymbol{\vartheta}_m)$. Matrices

$$\mathbf{X}_N(\varepsilon) = \boldsymbol{\Phi} \otimes \mathbf{K}_N(\varepsilon), \quad \mathbf{X}_D(\varepsilon) = \boldsymbol{\Phi} \otimes \mathbf{K}_D(\varepsilon) \tag{8}$$

where $\otimes$ is the Kronecker product, collect products of frequency and parameter basis functions, individually stored in

$$\boldsymbol{\Phi} = \begin{pmatrix} \varphi_0(s_1) & \cdots & \varphi_{\bar{n}}(s_1) \\ \vdots & & \vdots \\ \varphi_0(s_{\bar{k}}) & \cdots & \varphi_{\bar{n}}(s_{\bar{k}}) \end{pmatrix} \tag{9}$$

$$\mathbf{K}_{N,D}(\varepsilon) = \begin{pmatrix} \xi^{\varepsilon}_1(\boldsymbol{\vartheta}_1) & \cdots & \xi^{\varepsilon}_{\bar{\ell}}(\boldsymbol{\vartheta}_1) \\ \vdots & & \vdots \\ \xi^{\varepsilon}_1(\boldsymbol{\vartheta}_{\bar{m}_t}) & \cdots & \xi^{\varepsilon}_{\bar{\ell}}(\boldsymbol{\vartheta}_{\bar{m}_t}) \end{pmatrix} \tag{10}$$

where $\bar{\ell} = \bar{\ell}_N$ and $\bar{\ell} = \bar{\ell}_D$ for $\mathbf{K}_N$ and $\mathbf{K}_D$, respectively.

D. Fast PSK implementation

The solution of the fully-coupled system (5) is usually achieved through the so-called *Fast PSK iteration* [35]. For each input-output port combination (i, j) we compute independently the following QR factorizations (we omit the dependence on ε for readability)

$$\mathbf{Q}^{\mu}_{(i,j)}\mathbf{R}^{\mu}_{(i,j)} = \begin{pmatrix} \boldsymbol{\Gamma}^{\mu} & \boldsymbol{\Xi}^{\mu}_{(i,j)} \end{pmatrix} \tag{11}$$

where

$$\mathbf{R}^{\mu}_{(i,j)} = \begin{pmatrix} \mathbf{R}^{\mu}_{(i,j;1,1)} & \mathbf{R}^{\mu}_{(i,j;1,2)} \\ \mathbf{0} & \mathbf{R}^{\mu}_{(i,j;2,2)} \end{pmatrix} \tag{12}$$

When the bottom-right blocks $\mathbf{R}^{\mu}_{(1,1;2,2)}, \ldots, \mathbf{R}^{\mu}_{(P,P;2,2)}$ are available, the least squares problem

$$\begin{pmatrix} \mathbf{R}^{\mu}_{(1,1;2,2)} \\ \vdots \\ \mathbf{R}^{\mu}_{(P,P;2,2)} \end{pmatrix} \mathbf{d}^{\mu} = \bar{\mathbf{R}}^{\mu}\mathbf{d}^{\mu} \approx \mathbf{0} \tag{13}$$

is assembled and solved, where $\mathbf{d}^{\mu}$ collects all denominator coefficients $r_{n,\ell}$ at iteration μ. These coefficients allow in turn the setup of the PSK system for the next iteration, by updating the weighting matrix $\mathbf{W}^{\mu}$. When the coefficient estimates $\mathbf{d}^{\mu}$ stabilize at iteration $\bar{\mu}$, the numerator coefficients are found as

$$\boldsymbol{\Gamma}^{\bar{\mu}}\mathbf{C} \approx \mathbf{B} \tag{14}$$

where the matrix unknown is

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}_{(1,1)} & \cdots & \mathbf{c}_{(P,P)} \end{pmatrix} \tag{15}$$

with vectors $\mathbf{c}_{(i,j)}$ collecting all elements (i, j) of numerator matrix coefficients $\mathbf{R}_{n,\ell}$, and where

$$\mathbf{B} = \begin{pmatrix} -\boldsymbol{\Xi}_{(1,1)}\mathbf{d}^{\bar{\mu}} & \cdots & -\boldsymbol{\Xi}_{(P,P)}\mathbf{d}^{\bar{\mu}} \end{pmatrix}. \tag{16}$$

E. Uniform stability

In order to guarantee reliable time-domain simulations, the macromodel must be guaranteed stable for all the possible parameter instances $\boldsymbol{\vartheta}$ in the parameter space (uniform stability)

$$\Re\{p_n(\boldsymbol{\vartheta})\} < 0 \quad \forall \boldsymbol{\vartheta} \in \Theta, \tag{17}$$

where $p_n(\boldsymbol{\vartheta})$ are the parameter-dependent model poles of the model (2) written in pole-residue form

$$\mathbf{H}(s;\boldsymbol{\vartheta}) = \boldsymbol{\Upsilon}_0(\boldsymbol{\vartheta}) + \sum_{n=1}^{\bar{n}} \frac{\boldsymbol{\Upsilon}_n(\boldsymbol{\vartheta})}{s - p_n(\boldsymbol{\vartheta})}. \tag{18}$$

We remark that the adopted model structure (2) parameterizes the poles $p_n(\boldsymbol{\vartheta})$ only implicitly. These poles can only be computed a-posteriori by fixing $\boldsymbol{\vartheta}$ and evaluating the zeros of the denominator in (2). Therefore, direct constraints in the form (17) cannot be exploited as in standard (non-parameterized) VF. However, it has been shown [9] that uniform stability is guaranteed by a strictly positive-real denominator

$$\Re\{D(s,\boldsymbol{\vartheta})\} > \gamma \quad \Re\{s\} > 0, \forall \boldsymbol{\vartheta} \in \Theta \tag{19}$$

for any given positive constant $\gamma > 0$. In turn, this condition is guaranteed [20], [21] by the simpler algebraic conditions

$$\begin{cases} r_{n,\ell} > \gamma \\ -\alpha_n \cdot r'_{n,\ell} \pm \beta_n \cdot r''_{n,\ell} > \gamma \end{cases} \tag{20}$$

on the denominator coefficients associated to real q_n and complex $\alpha_n \pm \mathrm{j}\beta_n$ basis poles, with corresponding real $r_{n,\ell}$ and complex conjugate $r'_{n,\ell} \pm \mathrm{j}r''_{n,\ell}$ residues, respectively, provided that positive definite basis functions $\xi_{\ell}(\boldsymbol{\vartheta}) > 0$, $\forall \boldsymbol{\vartheta}$ (such as the Gaussian RBFs) are used. Uniformly stable models are therefore obtained by solving a convex optimization problem that minimizes the (quadratic) least squares error associated to (13) subject to the linear inequality constraints (20).

III. Problem statement

The general framework discussed in Section II provides a state-of-the-art flow for the generation of parameterized models based on a model structure (2) with Gaussian RBF bases (3), provided that

1) the dynamic order $\bar{n}$ and the basis poles q_n are known;
2) the number $\bar{\ell}_N$ and $\bar{\ell}_D$ of RBF basis functions for numerator and denominator is fixed;
3) the centers $\mathring{\boldsymbol{\vartheta}}_\ell$ where these basis functions are placed are given;
4) the shape parameter ε is fixed.

All these elements are collectively denoted as *hyper-parameters* in the following. Under these assumptions, no user interaction is required for model generation, and all steps discussed in Sec. II are well defined.

Unfortunately, the four above assumptions are not satisfied in practical situations. All hyper-parameters are generally not known to the user and should be initialized before actual model identification can start. In some cases, this selection can be based on some prior knowledge of the structure under modeling. For instance, a fixed and predefined number of poles can be used for well-defined components such as inductors or capacitors. However, the general selection of hyper-parameters is still regarded as an open problem, for which either time-consuming trial and error procedures are routinely applied, or alternatively heuristic selections are performed, which are far from being optimal.

The main contribution of this paper is a full set of algorithms providing an automated suboptimal choice of all these hyper-parameters. With the proposed automated framework, high-dimensional parameterized model generation becomes a "one-click" process, with obvious speedup in model extraction with respect to trial and error attempts. Before proceeding, we emphasize that

- all derivations are based on a fixed initial dataset of samples $\breve{\mathbf{H}}_{k,m}$. No interaction between model generation and initial data calculation (e.g. via an on-demand request of new samples from a field solver) is considered.
- model passivity verification and enforcement is not considered in this work. Several results are available in the context of (low-dimensional) parameterized macromodeling [38], [39]. Extension to a high-dimensional setting is still an open problem under the computational cost perspective. This problem is not discussed here, except for few remarks on the numerical examples in Sec. VI-D.

The proposed contributions are organized as follows. Section IV concentrates on the automated determination of the shape parameter ε once all other hyper-parameters are fixed. Section V-A discusses determination of dynamic model order (number of poles). Section V-B discusses an efficient (greedy) strategy for the selection of appropriate number and location of RBFs for both numerator and denominator.

IV. Automated selection of the Gaussian shape parameter

The approximation capabilities of a finite set of basis functions $\xi_\ell^\varepsilon(\boldsymbol{\vartheta})$ strongly depends on the free parameter ε, which must be carefully selected. In the past years, many efforts have been devoted to establish a direct link between the RBF approximation error and the Gaussian shape parameter ε. For instance, some error bounds have been found as a function of the shape parameter and data point distribution (see [40]). Others [41]–[43] found some peculiar properties of infinitely flat ($\varepsilon \to 0$) RBFs, exploited in [44]–[47] to develop numerically stable algorithms for RBF-based approximation. The main drawback of such techniques is that they do not scale favourably in high-dimensional spaces and, thus, turn out to be of limited use in our scope.

A standard approach in machine learning is to infer a suitable shape parameter value based on statistical considerations (see [24], [25]). In principle, these techniques rely on optimizing some predictor function, such as cross validation or maximum likelihood estimators. All these strategies require therefore the repeated extraction of intermediate macromodels, evaluated for different values of ε. In our framework, this procedure may take unacceptably long runtimes. In the following, we discuss a very efficient strategy, which was presented heuristically in preliminary form in [48]. Here, we provide a solid ground for the basic algorithm, by proving a number of results on the spectral properties of all PSK matrices introduced in Sec. II viewed as a function of ε. The main algorithm is motivated and presented without proofs in the following subsections. All the theoretical results in support of this algorithm are collected in the Appendix, which may be skipped at first reading.

A. Shape parameter, model accuracy, and conditioning

In this section, we investigate the relation between the shape parameter ε, the expected model accuracy, and the condition number of the PSK matrix (6) on which the model identification is based. The various observations that follow are supported by Fig. 1, which refers to an integrated inductor parameterized by its sidelength. The observations are however general and apply to all test cases included in this work.

We analyze the two asymptotic cases.

1) Case $\varepsilon \to \infty$: The shape parameter ε is inversely proportional to the standard deviation of the Gaussian RBF, viewed as a probability density function. Therefore, when $\varepsilon \to \infty$ the width of $\xi_\ell^\varepsilon(\boldsymbol{\vartheta})$ tends to vanish, and the RBF concentrates more and more around its center $\mathring{\boldsymbol{\vartheta}}_\ell$. The asymptotic limit is

$$\xi_\ell^\infty(\boldsymbol{\vartheta}) = \begin{cases} 1 & \text{if } \boldsymbol{\vartheta} = \mathring{\boldsymbol{\vartheta}}_\ell \\ 0 & \text{if } \boldsymbol{\vartheta} \neq \mathring{\boldsymbol{\vartheta}}_\ell \end{cases} \tag{21}$$

A corresponding model provides exact interpolation at the RBF centers, but it is identicaly vanishing elsewhere; it will be unable to provide a continuous parameterization, leading the overall model accuracy to be poor. This behavior is highlighted in the top panel of Figure 1, region C, which reports the model-data error as a function of ε on an illustrative test case.

On the other hand, when approaching this limit the resulting fitting problem is expected to be very well conditioned. Indeed,

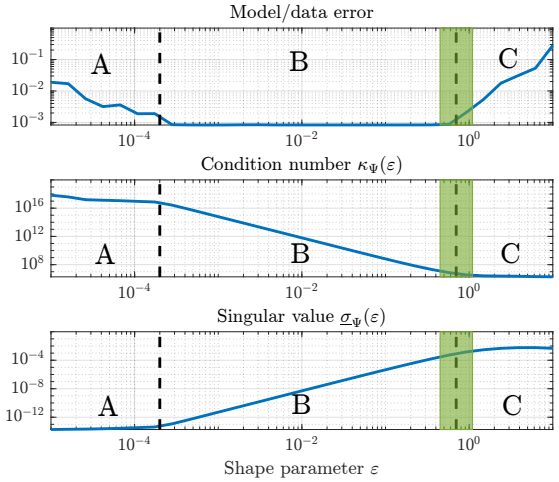

Fig. 1. Dependence on the shape parameter ε of model-data error (top panel), condition number of matrix $\mathbf{\Psi}(\varepsilon)$ (middle panel), and least singular value of matrix $\mathbf{\Psi}(\varepsilon)$ (bottom panel). For all panels three regions are highlighted: A (numerical instabilities associated with small ε), B (acceptable model accuracy), C (loss of approximation capabilities). Green region: candidate sub-optimal shape parameter values.

under suitable rows and columns permutations, the kernel matrix $\mathbf{K}_{N,D}(\varepsilon)$ becomes

$$\mathbf{K}_{N,D}(\infty) = \begin{pmatrix} \mathbf{I}_{\bar{\ell}} \\ \mathbf{0} \end{pmatrix} \tag{22}$$

where $\mathbf{I}_{\bar{\ell}}$ denotes the identity matrix of dimension $\bar{\ell}$. A matrix of this form has optimal condition number $\kappa_{\mathbf{K}}(\infty) = 1$. Thus, with the assumption that matrix $\mathbf{\Phi}$ collecting the basis poles is numerically well-conditioned, such good numerical properties are inherited by $\mathbf{\Psi}^\mu(\varepsilon)$ (or, equivalently by $\bar{\mathbf{R}}^\mu$). The middle panel of Figure 1 confirms this statement (see region C, for large ε).

2) Case $\varepsilon \to 0$: When $\varepsilon \to 0$ the Gaussian RBF becomes increasingly flat around its center. A set of remarkable results [41]–[43] show that under these conditions a Gaussian RBF approximant is equivalent to the De Boor/Ron "least polynomial" interpolant. Therefore, polynomial accuracy is expected for small ε, a condition that is particularly attractive when approximating smooth functions.

However, as Figure 1 suggests, this limit is computationally unattainable. As $\varepsilon \to 0$ the conditioning of the fitting problems degrades (Fig. 1, middle panel, region A), resulting in increasingly inaccurate models (Fig. 1, top panel, region A). This result is very intuitive, since when $\varepsilon \to 0$ all basis functions approach uniformly constant planes and become linearly dependent. The kernel matrix approaches the limit

$$\mathbf{K}_{N,D}(0) = \mathbf{1}_{\bar{m}\times\bar{\ell}} \tag{23}$$

where $\mathbf{1}_{\bar{m}\times\bar{\ell}}$ is a matrix of all ones, which has only rank one.

It seems therefore that a good compromise is to choose ε sufficiently small to ensure accuracy, but not too small in order to prevent numerical conditioning issues. These statements are made more precise in the following.

B. Singular values as a proxy for model accuracy

The behavior of the condition number $\kappa_\Psi(\varepsilon)$ depicted in Fig. 1 is typical in Gaussian RBF approximation. Let us recall that $\kappa_\Psi(\varepsilon) = \bar{\sigma}_\Psi(\varepsilon)/\underline{\sigma}_\Psi(\varepsilon)$, where $\bar{\sigma}_\Psi(\varepsilon)$ and $\underline{\sigma}_\Psi(\varepsilon)$ are, respectively, the leading and the least singular values of matrix $\mathbf{\Psi}(\varepsilon)$. These two singular values are analyzed below.

1) The least singular value: A complete characterization of $\underline{\sigma}_\Psi(\varepsilon)$ requires an in-depth analysis of the singular values spectra of PSK matrices, which is provided in the Appendix. The main result (see Theorem 2 in the Appendix) states that, for $\varepsilon \to 0$, the least singular value $\underline{\sigma}_\Psi(\varepsilon)$ decreases with an integer power τ that is at least 4, provided that the number of RBFs is larger than the dimension of the embedding space plus two, $\bar{\ell} > \rho + 2$. More precisely,

$$\underline{\sigma}_\Psi(\varepsilon) = O(\varepsilon^\tau) \quad \text{as } \varepsilon \to 0, \quad \tau \in \mathbb{N},\ \tau \geq 4. \tag{24}$$

Hence, as confirmed by the bottom panel of Figure 1 (region B), the least singular value $\underline{\sigma}_\Psi(\varepsilon)$ decays linearly when visualized in a log-log scale, with a well-defined slope τ (the flattening in region A is determined solely by machine precision and truncation errors when evaluating the singular values).

2) The leading singular value: The leading singular value $\bar{\sigma}_\Psi(\varepsilon)$ is not expected to vary significantly as ε changes and approaches 0. Considering the two limit cases of the kernel matrix $\mathbf{K}_{N,D}(\varepsilon)$ in (22) and (23), for which the leading singular value is analytically known respectively as 1 and $\sqrt{\bar{m}\bar{\ell}}$, it is expected that $\bar{\sigma}_{\mathbf{K}}(\varepsilon)$ remains bounded in the interval $\left[1,\ \sqrt{\bar{\ell}\bar{m}}\right]$ for $0 \leq \varepsilon \leq \infty$. Based on (8), we conclude that also the variations of $\bar{\sigma}_\Psi(\varepsilon)$ are bounded within an interval that does not depend on ε. Hence, we can conclude that

$$\kappa_\Psi(\varepsilon) = \frac{\bar{\sigma}_\Psi(\varepsilon)}{\underline{\sigma}_\Psi(\varepsilon)} \sim \frac{C}{\underline{\sigma}_\Psi(\varepsilon)} \quad \varepsilon \to 0 \tag{25}$$

where C is a positive constant.

Recalling now (24), we conclude that the condition number $\kappa_\Psi(\varepsilon)$ blows up as $O(1/\varepsilon^\tau)$ for $\varepsilon \to 0$, as depicted in region B of Figure 1. Middle and bottom panels confirm that condition number and least singular value are nearly inversely proportional within region B (results in region A are unreliable due to machine precision). In summary,

1) large values of ε lead to inaccurate models (Fig. 1, region C), since the RBF are too narrow;
2) small values of ε result in accurate models but the associated fitting problem is prone to numerical instabilities when ε is too small (Fig. 1, region A).

We can thus identify the region B of Fig. 1 as a good candidate on which to search for a sub-optimal value of ε, that

1) must be small enough to ensure a proper model parameterization;
2) must be sufficiently large to guarantee a proper conditioning of the fitting problem.

These two requirements are clearly conflicting. In addition, in order to guarantee the best possible numerical conditioning and robustness, we aim at picking ε that minimizes $\kappa_\Psi(\varepsilon)$. In this scope, we narrow our search to the green-shaded area at the interface of regions B and C, where both the requirements on numerical conditioning and model accuracy are fulfilled.

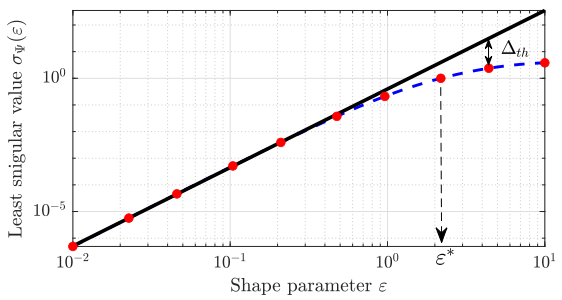

Fig. 2. Algorithm 1: iterative regression (black thick line) on shape parameter-singular value pairs $\{\epsilon_t,\ \varrho_t\}$ (red dots).

C. An algorithm for choosing the shape parameter

We now describe our proposed fast algorithm for the selection of a sub-optimal shape parameter ε^*. With reference to Fig. 1, we aim at identifying the green-shaded area at the intersection between regions B and C. Using the least singular value $\underline{\sigma}_{\mathbf{\Psi}}(\varepsilon)$ as a proxy for the condition number $\kappa_{\mathbf{\Psi}}(\varepsilon)$ of the fitting matrix, we locate the point where its trajectory in a log-log scale begins to flatten. This point is at the corner between the scaling range of $\underline{\sigma}_{\mathbf{\Psi}}(\varepsilon)$ as $O(\varepsilon^\tau)$, and the range where it becomes nearly constant.

As initially proposed in [48], we compute in advance a set of pairs (red dots in Figure 2)

$$\{(\varepsilon_t, \underline{\sigma}_{\mathbf{\Psi}}(\varepsilon_t)),\ t = 1, \dots, \bar{t}\}$$

on a set of increasing logarithmically spaced candidate shape parameter values $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{\bar{t}}\}$ (3–4 points per decade are sufficient). Accordingly, we define $\epsilon_t = \log \varepsilon_t$ and $\varrho_t = \log \underline{\sigma}_{\mathbf{\Psi}}(\varepsilon_t)$. To localize the corner point we exploit the result in Theorem 2, that guarantees a power-law decay of $\underline{\sigma}_{\mathbf{\Psi}}(\varepsilon)$ for small values of ε. Hence, we set up an iterative scheme with iteration index J. At each step, a log-log regression line $\zeta_J(\varepsilon) = \alpha_J \log \varepsilon + \beta_J$ (black thick line in Figure 2) is computed on the pairs $\{(\epsilon_t, \varrho_t),\ t = 1, \dots, J\}$. Then, if the relative deviation $\Delta_{J+1} = |\zeta_J(\varepsilon_{J+1}) - \varrho_{J+1}| / \varrho_{J+1}$ exceeds a predefined threshold Δ_{th} we assume to be in correspondence of the corner point we are looking for. Thus, assuming that at iteration J^* we have $\Delta_{J^*+1} > \Delta_{th}$, we stop the iterations and we set $\varepsilon^* = \varepsilon_{J^*}$. Algorithm 1 reports a high-level pseudo-code for this procedure.

D. From regressor to kernel matrices

As described in Sec. IV-C and Algorithm 1, the identification of the optimal shape parameter is based on the least singular value of the regressor matrix $\mathbf{\Psi}(\varepsilon)$ in (6). This matrix is iteration-dependent and large-sized. Therefore, an expensive and repeated (least) singular value calculation would be required multiple times at each PSK iteration, in order to estimate the appropriate shape parameter for the model. Fortunately, this calculation is not necessary. In fact:

- Although the $\mathbf{\Psi}^\mu(\varepsilon)$ changes at each PSK iteration μ, its spectral properties and in particular the dependence on ε of its least singular value are practically invariant through iterations. Therefore, Algorithm 1 can be run only once at the first iteration $\mu = 1$, and the corresponding optimal ε^* reused at all subsequent iterations.
- The behavior of the least singular value of $\mathbf{\Psi}^\mu(\varepsilon)$ is inherited, by construction, also by matrix $\bar{\mathbf{R}}^\mu(\varepsilon)$ in (13) resulting from the Fast PSK formulation, whose size is significantly smaller. Therefore, evaluating the singular values of $\bar{\mathbf{R}}^\mu(\varepsilon)$ instead of $\mathbf{\Psi}^\mu(\varepsilon)$ enables a significant speed-up of the entire procedure.
- A third significantly faster implementation is based on the kernel matrices $\mathbf{K}_N(\varepsilon)$, $\mathbf{K}_D(\varepsilon)$ associated to numerator and denominator, respectively. These matrices are significantly smaller than $\mathbf{\Psi}^\mu(\varepsilon)$ and $\bar{\mathbf{R}}^\mu(\varepsilon)$, so that their construction and singular values computation requires negligible time. As we will prove in the Appendix, also these kernel matrices share similar spectral properties with $\mathbf{\Psi}^\mu(\varepsilon)$ and $\bar{\mathbf{R}}^\mu(\varepsilon)$. In particular, their least singular values $\underline{\sigma}_{\mathbf{K}_N}(\varepsilon)$ and $\underline{\sigma}_{\mathbf{K}_D}(\varepsilon)$ have the the same dependence with ε as $\underline{\sigma}_{\mathbf{\Psi}}(\varepsilon)$, $\underline{\sigma}_{\bar{\mathbf{R}}}(\varepsilon)$.

Based on the above considerations, our proposed implementation applies Algorithm 1 to matrices $\mathbf{K}_N(\varepsilon)$ and $\mathbf{K}_D(\varepsilon)$ for the identification of the sub-optimal shape parameters ε_N^* and ε_D^* defining the numerator and denominator Gaussian RBFs, respectively.

Algorithm 1 Sub-optimal shape parameter selection [48]

Require: Log-spaced samples $\{\varepsilon_1, \dots, \varepsilon_{\bar{t}}\}$.
Require: Stopping threshold Δ_{th}
Require: Reference matrix $\mathbf{Z}(\varepsilon)$: $\mathbf{\Psi}(\varepsilon)$, $\mathbf{R}(\varepsilon)$, or $\mathbf{K}_{N,D}(\varepsilon)$
Compute $\epsilon_t = \log \varepsilon_t$ and $\varrho_t = \log \underline{\sigma}_{\mathbf{Z}}(\varepsilon_t)$ for $t = 1, \dots, \bar{t}$
for $J = 2, \dots, \bar{t}$ **do**
  Build regression line $\zeta_J(\varepsilon) = \alpha_J \log \varepsilon + \beta_J$ using pairs $\{(\epsilon_t, \varrho_t),\ t = 1, \dots, J\}$
  Evaluate $\Delta_{J+1} = |\zeta_J(\varepsilon_{J+1}) - \varrho_{J+1}| / \varrho_{J+1}$
  if $\Delta_{J+1} > \Delta_{th}$ **then**
    Set $\varepsilon^* = \varepsilon_J$
    break
  end if
end for
return Sub-optimal shape parameter ε^*.

V. Model Order Selection

In order to properly capture the broadband and parametric variations of the system responses, the number and the position of the Radial Basis Functions as well as the number and location of the basis poles should be carefully determined. A direct search among all the possible triplets $\{\bar{n}, \bar{\ell}_N, \bar{\ell}_D\}$ while maximizing the model accuracy is clearly unfeasible.

A. Dynamic order selection

In order to estimate the proper dynamic order (number of model poles), we setup an iterative scheme, that is run in a preprocessing stage. A set of candidate dynamic orders is first collected in ascending order as $\{\bar{n}_1, \dots, \bar{n}_{\bar{q}}\}$. Then, we identify a suitable subset of frequency responses corresponding to some predefined parameter values

$$\mathcal{P} = \{\boldsymbol{\vartheta}_z \in \mathcal{T} :\ z = 1, \dots, \bar{z}\}. \tag{26}$$

In our strategy, $\mathcal{P}$ contains the data points in the training set corresponding to the nearest neighbors to the parameter space vertices as well as its center. This choice is not restrictive.

The well-established VF algorithm [4] is applied independently to all frequency responses corresponding to the $\bar{z}$ parameter values in set $\mathcal{P}$, by iteratively increasing the dynamic order $\bar{n} = \bar{n}_1, \bar{n}_2, \ldots$. As the iterations proceed, since the number of poles is gradually increased, the worst case model-data error $\mathcal{E}_I$ among all the possible $\bar{z}$ models is expected to reduce. The process stops at iteration I^* when the error $\mathcal{E}_{I^*}$ is uniformly below a prescribed threshold $\mathcal{E}_{th}$, and the corresponding dynamic order $\bar{n} = \bar{n}_{I^*}$ is selected. The set of basis poles $\{q_n\}$ is then initialized by picking the poles of the VF model corresponding to the nearest neighbor of the parameter space center.

We remark that the above procedure is conventional, heuristic, and well established in common macromodeling flows and even in commercial tools, at least for non-parameterized macromodeling. It is included here in order to provide the Reader with a complete set of "recipes" for setting up a parameterized macromodeling code.

B. Parameter order and RBF selection

A greedy approach for selecting both number and location of RBFs was proposed by the Authors in [34], leading to quite accurate results. However, this approach proves to be computationally inefficient and, more importantly, it may lead to models that are likely to be more complex than strictly required (see Section VI-B for a numerical example).

Here, we propose a different strategy in the interest of model compactness and required computational runtime. First, we do not seek for an optimal placement of the RBFs in the parameter space. Although several techniques have been proposed in the past years [27]–[32] for this optimal placement, they are inapplicable to the adopted rational model structure, which presents RBF expansions both at numerator and denominator. In fact, many numerical experiments suggest that, in the framework of parameterized macromodeling, the accuracy improvement deriving from optimizing the position of the centers is not justified by the required computational overhead. Instead, we assume an a-priori distribution of the RBF centers, with the only requirement that they cover uniformly the parameter space. With this assumption, the only free hyper-parameters are the number of RBFs. Differently from [34], we allow for different RBF orders $\bar{\ell}_N$, $\bar{\ell}_D$ of numerator and denominator, preventing the model to be unnecessarily complex.

The proposed search for RBF orders is performed on two sets of candidate numerator and denominator orders $\Lambda_N = \{\bar{\ell}_N^1, \ldots, \bar{\ell}_N^{\bar{r}}\}$ and $\Lambda_D = \{\bar{\ell}_D^1, \ldots, \bar{\ell}_D^{\bar{t}}\}$, whose elements are sorted in ascending order. Each element of these sets is associated with a set of predefined RBF centers

$$\begin{aligned} \mathcal{C}_N^r &= \{\boldsymbol{\vartheta}_m \in \mathcal{T} :\ m = 1, \ldots, \bar{\ell}_N^r\} \quad r = 1, \ldots, \bar{r} \\ \mathcal{C}_D^t &= \{\boldsymbol{\vartheta}_m \in \mathcal{T} :\ m = 1, \ldots, \bar{\ell}_D^t\} \quad t = 1, \ldots, \bar{t} \end{aligned} \tag{27}$$

for numerator and denominator, respectively, whose elements are selected as the nearest neighbors in $\mathcal{T}$ of a truncated space-filling Sobol sequence [49]. This choice guarantees uniform coverage of the parameter space, thus improving both model uniform model accuracy and numerical conditioning of the fitting problem. Once numerator and denominator orders are chosen, the associated RBF center location is automatically determined via (27).

All elements in sets $\Lambda_{N,D}$ must fulfill two types of constraints. On one hand, each of the candidate RBF orders must ensure applicability of the proposed algorithm for shape parameter estimation, see Sec. IV. Therefore, we must have $\bar{\ell}_{N,D}^1 > \rho + 2$. On the other hand, the least squares problem (1) must be sufficiently over-determined to avoid overfitting conditions. This is achieved by setting $2(\bar{\ell}_N^{\bar{r}} + \bar{\ell}_D^{\bar{t}}) < \bar{m}_t$. Therefore, the maximum allowed order of numerator and denominator expansions is not predefined, but their cumulative order is. The main objective is to choose appropriate model orders $\bar{\ell}_N$, $\bar{\ell}_D$ from sets Λ_N and Λ_D that minimize the difference between model and data.

The scheme we use here is similar to [50], which was proposed in the context of a multivariate Loewner framework. Tentative models are generated by increasing either numerator or denominator order, until the model-data error is acceptable. In principle, this process would require the repeated extraction of many sub-models, each evaluated for different orders $\bar{\ell}_N$, $\bar{\ell}_D$. In turn, each model extraction may require several PSK iterations. In order to boost the performances, we do not wait for PSK convergence and we perform only one PSK iteration. The resulting model-data error is considered as a proxy of the model accuracy at convergence. This is justified noting that the accuracy at convergence is higher than the one at the first iteration.

The model orders are determined using a sub-optimal greedy approach. We proceed iteratively, by updating a current pair of model orders, denoted as $\bar{\ell}_N^r$, $\bar{\ell}_D^t$ and indexed by pair (r, t). Let us denote with $E(\bar{\ell}_N, \bar{\ell}_D, \varepsilon)$ the model-data error (at the first PSK iteration) evaluated with orders $\bar{\ell}_N$ and $\bar{\ell}_D$. Moreover, we denote with $E^*(\bar{\ell}_N, \bar{\ell}_D) = \min_\varepsilon E(\bar{\ell}_N, \bar{\ell}_D, \varepsilon)$ as obtained by the shape parameter optimization based on Algorithm 1. This algorithm is actually applied to the small-size Gaussian kernel matrices $\mathbf{K}_{N,D}(\varepsilon)$ and requires negligible runtime. At each iteration, we test the errors

$$E^*(\bar{\ell}_N^{r+\delta}, \bar{\ell}_D^t) \quad \text{and} \quad E^*(\bar{\ell}_N^r, \bar{\ell}_D^{t+\delta}), \tag{28}$$

where $\delta = 1, \ldots, \bar{\delta}$ is a step size, and we update indices r and or t depending on which combination of orders results in smaller error. This strategy amounts to increasing either the numerator or the denominator order according to the best local model improvement. Initial orders are initialized as $\bar{\ell}_N^1 = 4\rho$, $\bar{\ell}_D^1 = 2\rho$, and the process is then repeated until a sub-optimal order pair $\{\bar{\ell}_N,\ \bar{\ell}_D\}$ is found.

Figure 3 illustrates the RBF order selection algorithm on the structure presented in Section VI-A. Starting from the initial guess $\bar{\ell}_N^1 = 5$ and $\bar{\ell}_D^1 = 5$, the algorithm follows the steepest descent path, highlighted with a red thick line, until it reaches the minimum among the available choices in Λ_N and Λ_D.

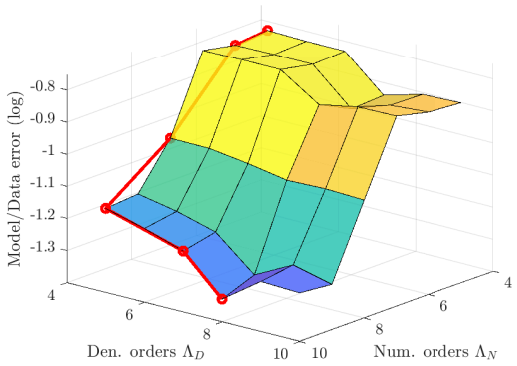

Fig. 3. Steps of the greedy RBF order selection algorithm. The surface represent the error $E^*(\bar{\ell}_N, \bar{\ell}_D)$ after shape parameter optimization; the red thick line shows the steepest descent path followed by the algorithm.

TABLE I
HYPER-PARAMETERS SELECTED BY PROPOSED ALGORITHM FOR THE MULTI-BOARD INTERCONNECT.

Number of poles	$\bar{n}$	24
Numerator parameter order	$\bar{\ell}_N$	10
Denominator parameter order	$\bar{\ell}_D$	8
Numerator shape parameter	ε_N^*	0.4642
Denominator shape parameter	ε_D^*	0.4642

VI. EXAMPLES

A. A multi-board interconnect

We start with a low-complexity two-dimensional example ($\rho = 2$) to illustrate the proposed automated procedure for hyper-parameter selection. The structure under modeling is a multiboard link that interconnects two multilayer PCBs [51] through stripline segments, a connector, and its corresponding via fields. The link is parameterized by the via pad (ϑ^1) and anti-pad (ϑ^2) radii, which vary in the ranges $[100,\ 300]$ μm and $[400,\ 600]$ μm, respectively. A parameterized model was constructed starting from a dataset of $\bar{m} = 81$ parametric frequency responses spanning the band $[0, 5]$ GHz (courtesy of Prof. C. Schuster and Dr. J. Preibisch, Technische Universität Hamburg-Harburg, Hamburg, Germany). Half of these responses was used for training the model, leaving the second half for self-validation.

The proposed automated parameterized macromodeling flow was executed as follows:

1) Dynamic order selection: the optimal number of poles was selected among the candidate orders $\{10, \ldots, 30\}$, using a stopping threshold $\mathcal{E}_{th} = 10^{-3}$.

2) Parameter order selection: numerator and denominator orders were chosen from the sets $\Lambda_N = \Lambda_D = \{5, \ldots, 10\}$ using a step size $\bar{\delta} = 2$ and a stopping threshold 10^{-2}.

3) Shape parameter selection: the RBF width was optimized in terms of the shape parameter ε among a set of 10 logarithmically spaced candidates ranging from 10^{-3} to 10, using a stopping threshold $\Delta_{th} = 4 \times 10^{-2}$.

Table I lists the optimal values returned by proposed scheme. The whole procedure required about 41 seconds to find an appropriate model structure and synthesize the model,

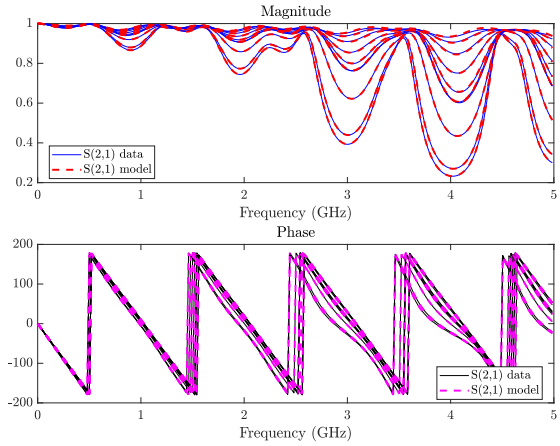

Fig. 4. Model responses compared with validation data for the multi-board interconnect example. Blue lines: randomly-selected validation data. Red dashed lines: corresponding model responses.

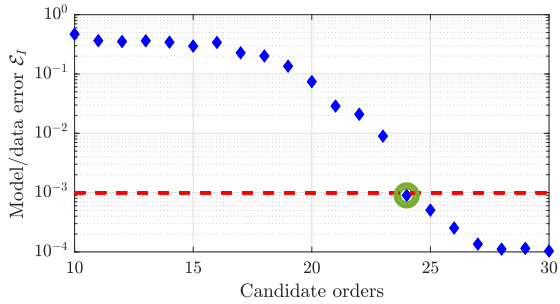

Fig. 5. Multiboard interconnect: iterations of the dynamic order selection algorithm. Blue markers: worst-case model data error (among the selected parameter values) for subsequent candidate number of poles; red-dashed line: stopping threshold; green circle: selected dynamic order.

with a worst-case error of 7×10^{-3} with respect to the validation samples (not used for model generation). Figure 4 confirms the model accuracy by plotting the model responses corresponding to a randomly selected subset of validation data.

We now focus on each step of proposed algorithm, documenting the optimality of its automated choices by comparing to the results from a full grid search.

Dynamic order selection: Fig. 5 depicts the individual steps. Starting from $\bar{n}_1 = 10$, the algorithm keeps increasing the dynamic order until the worst-case model data error $\mathcal{E}_I$ is uniformly below the threshold $\mathcal{E}_{th} = 10^{-3}$. In the presented example, this occurs at iteration $I^* = 15$ (highlighted with a green circle) corresponding to a number of poles $\bar{n} = 24$. Note that the error saturates to 10^{-4} for large candidate orders, denoting the presence of noise and/or non-causal components in the data. In such scenarios this saturation imposes a bound for the attainable accuracy in any rational approximation applied to this dataset, and $\mathcal{E}_{th}$ must be set to a larger value.

Parameter order selection: Figure 3 shows adaptive steps of the greedy algorithm proposed in Sec. V-B to determine numerator and denominator RBF orders. As depicted in Figure 3, the algorithm follows the steepest descent path until, in 5 iterations, it stops at $\bar{\ell}_N^5 = 10$, $\bar{\ell}_D^5 = 8$. The surface (not used for actual computations) represents the results of a standard grid search over all candidate orders. This figure confirms that

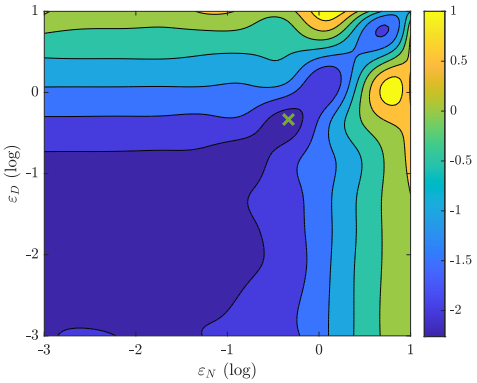

Fig. 6. Shape parameter selection algorithm. The contour plot represents the model-data error as a function of numerator and denominator shape parameter values ε_N, ε_D. The green cross pinpoints the sub-optimal combination ε_N^*, ε_D^* selected by the proposed algorithm

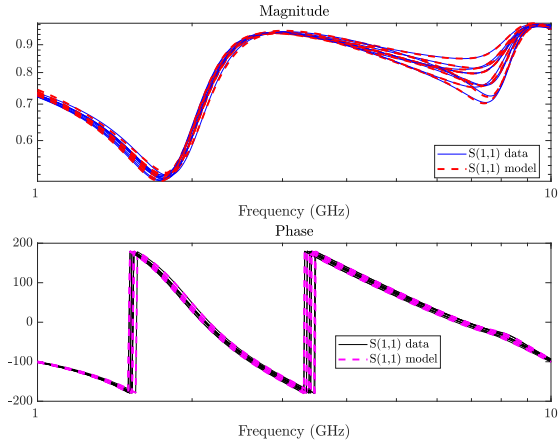

Fig. 7. LNA test case: validation of model responses (res dashed lines) against corresponding reference data (blue solid lines), randomly selected in the 10-dimensional parameter space.

the proposed algorithm was able to find the global minimum in the search set.

Shape parameter selection: Figure 6 depicts the model-data error evaluated on a uniform grid spanning the same shape parameter values ε_N, ε_D used to run the automated selection algorithm. The green cross corresponds to the selected sub-optimal combination $\varepsilon_N^* = \varepsilon_D^* = 0.4642$, which coincides with the lowest error instance among all the available samples in the grid.

In terms of computational performance, the presented strategy provides several advantages. In particular, with respect to a standard non-optimized grid search, the proposed parameter order selection algorithm is twice as fast. The major advantage is however due to the identification of the shape parameters ε_N^* and ε_D^*: the proposed algorithm results 464 times faster with respect to a brute-force search based on repeated model-data error computation over the two-dimensional grid.

B. A Parameterized Low Noise Amplifier

In this second example, we compare the performance of the proposed automated strategy with respect to the preliminary technique presented in [34]. The latter was based on a search for sub-optimal RBF centers without any additional optimization on the associated shape parameters. The results will confirm that proposed algorithm provides accurate macromodels with a reduced model complexity and in a fraction of time.

We consider the same Low Noise Amplifier (LNA) distributed circuit adopted in [34] and originally presented in [52]. The scattering responses of this structure depend on 10 independent parameters, six of which are lumped parasitics of BJTs and the other four refer to the geometry of distributed transmission lines (for additional information on the parameterization see [20]). The training dataset includes $\bar{m} = 2000$ frequency responses, each with 701 linearly spaced frequency samples in the range $1 - 10$ GHz. These $\bar{m}$ responses are distributed in the parameter space according to a Latin Hypercube sampling strategy [53] for best coverage.

The hyper-parameter selection algorithm was run in the same configuration as in Sec. VI-A, but with a more stringent stopping threshold $\mathcal{E}_{th} = 10^{-5}$. The proposed scheme resulted in $\bar{n} = 14$ poles, $\bar{\ell}_N = 72$ and $\bar{\ell}_D = 17$ RBF orders for numerator and denominator, respectively, with associated shape parameter values $\varepsilon_N^* = \varepsilon_D^* = 0.0215$. The final 10-dimensional parameterized macromodel was characterized by a worst-case absolute model-data error below 1%. Figure 7 compares model and reference data responses for a randomly-selected subset among the available $\bar{m} = 2000$ parameter values, confirming uniform model accuracy throughout the parameter space. The total model extraction time was 7.5 minutes: 6.3 minutes were required to optimize the model structure, while the model extraction took only the remaining 75 seconds. As a comparison, using the strategy presented in [34], we were able to extract a parameterized model of comparable accuracy in approximately 45 minutes which is about 6 times slower than than proposed automated algorithm.

As anticipated, the presented strategy is able to optimize model structure and is likely to produce more compact (or, equivalently, less complex) macromodels. Indeed, defining the model complexity as $\eta = \bar{\ell}_N + \bar{\ell}_D$, the approach in [34] provided $\eta = 144$, in contrast with proposed approach for which $\eta = 89$ for the same final model accuracy, resulting in a compression of approximately 40%.

C. A transmission line network

As a final example, we consider a distributed network with four cascaded lossy microstrip segments (with length l_i) and three internal stubs (with length s_i) terminated with purely resistive loads R_i. The copper microstrip conductor has width w and thickness $t = 30$ μm and is placed on a dielectric substrate with height h, relative dielectric constant $\varepsilon_r = 4.1$ and loss tangent $\tan\delta = 0.02$. The input/output responses depend on ten free parameters ϑ_i: Table II reports their nominal values $\bar{\vartheta}_i$ and the associated relative variation ranges $\Delta\vartheta_i$. The parameters l_i are constrained to ensure that the total line length remains constant and equal to 15 mm.

A 10-dimensional parameterized model was constructed starting from a set of $\bar{m} = 300$ scattering frequency responses, each composed of $\bar{k} = 401$ linearly spaced samples in the band $\Omega = [0,\ 20]$ GHz. These responses were pre-computed

TABLE II
TRANSMISSION LINE NETWORK PARAMETERS. l_i, s_i: LENGTH OF TRANSMISSION LINE AND STUB SEGMENTS; w: CONDUCTOR WIDTH; h: SUBSTRATE HEIGHT; R_i: STUB TERMINATION RESISTANCES.

#	Parameter ϑ_i	$\bar{\vartheta}_i$	$\Delta\vartheta_i$
1	l_1 (mm)	10	5 %
2	l_2 (mm)	10	5 %
3	l_3 (mm)	10	5 %
4	l_4 (mm)	10	5 %
5	$s_{1,2,3}$ (mm)	3.75	2 %
6	w (μm)	150	10 %
7	h (μm)	200	10 %
8	R_1 (Ω)	50	10 %
9	R_2 (Ω)	50	10 %
10	R_3 (Ω)	50	10 %

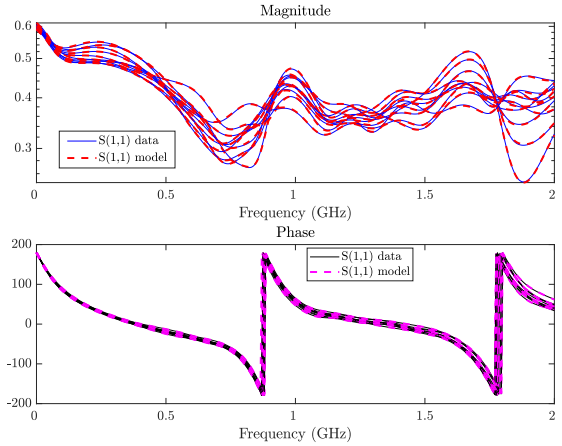

Fig. 8. Transmission line network: validation of the parameterized model responses (red dashed lines) through corresponding reference data (blue solid lines).

using the HSPICE circuit solver [54]. The proposed algorithm automatically determined a model structure with $\bar{n} = 27$ poles, numerator and denominator orders $\bar{\ell}_N = 40$ and $\bar{\ell}_D = 27$, respectively, with associated shape parameter values $\varepsilon_N^* = \varepsilon_D^* = 0.0599$. The modeling procedure required 14 minutes to optimize the model structure and only 30 seconds to synthesize a uniformly stable model with complexity $\eta = 69$ and whose accuracy with respect to validation samples is 11.2×10^{-3}. Figure 8 reports a model-data comparison over a subset of these validation samples. Adopting the approach presented in [34], a model of comparable accuracy was obtained in 1.5 hours, with a complexity $\eta = 112$. Thus, the presented strategy enables for a $6\times$ speed-up in terms of runtime and approximately a 39% compression in terms of model complexity.

D. Passivity

When dealing with passive EM structures (as in Examples A and C), their models should reflect this property. Therefore, we applied an (experimental) modified version of the sampling-based multivariate passivity check of [57]. Both the models A and C presented marginal passivity violations in the fitting band and relevant out-of-band violations, with largest singular values equal to 7 and 6, respectively, located at infinite frequency. An experimental modification of the multivariate passivity enforcement scheme [39] combined with [57] was thus applied to both cases, resulting in perturbed models of comparable accuracy, for which the check [57] did not find residual passivity violations. However, due to the experimental nature of the latter algorithm, we can not exclude the presence of residual violations in these final models.

VII. CONCLUSIONS

This paper presented a complete framework for the generation of high-dimensional parameterized macromodels. With respect to earlier documented work, we provided a number of theoretical results that support a fully automated flow for model extraction from a given set of precomputed frequency responses in the parameter space of interest. In particular, the proposed framework automatically selects number, placement, and width of Gaussian Radial Basis Functions enabling the model to represent variations in a possibly high-dimensional parameter space.

The presented approach is able to guarantee uniform model stability for any parameter combination. Model passivity is instead not guaranteed and, if appropriate, should be enforced as a post-processing step, as common in well-established macromodeling flows. The state of the art algorithms for passivity enforcement are not expected to scale favorably with parameter space dimension. Therefore, our future research efforts will be dedicated to incorporate passivity constraints in model generation, so that a good scalability of the entire passive macromodeling flow can be ensured.

APPENDIX

We provide here the theoretical background in support of the proposed algorithms for shape parameter selection. The results are presented and proved formally, following a bottom-up approach.

A. Euclidean Distance Matrices

We start with two preliminary results on Euclidean Distance Matrices (EDM), which form a basis for the main developments. The following Lemma collects some known results [55] on rank and column space of square EDMs.

Lemma 1. *Consider a set of known vectors* $\{\boldsymbol{\vartheta}_m \in \mathbb{R}^\rho,\ m = 1\ldots,\bar{m}\}$*, and define the square Euclidean Distance Matrix* $\bar{\mathbf{D}} \in \mathbb{R}^{\bar{m}\times\bar{m}}$ *whose* (i,j)*-th element is* $\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_j\|^2$*. Then,*

1) *there exists a unique vector* $\mathbf{p} \in \mathbb{R}^{\bar{m}}$ *such that*

$$\bar{\mathbf{D}}\,\mathbf{p} = \mathbf{1}_{\bar{m}} \tag{29}$$

2) $\operatorname{rank}(\bar{\mathbf{D}}) \leq \rho + 2$.

Proof. See [55]. □

The following result builds on Lemma 1 and extends it to non-square EDMs.

Lemma 2. *Consider a set of known vectors* $\{\boldsymbol{\vartheta}_m \in \mathbb{R}^\rho,\ m = 1,\ldots,\bar{m}\}$ *and assume* $\rho+2 < \bar{\ell} < \bar{m}$*. Define the rectangular Euclidean Distance Matrix* $\mathbf{D}_1 \in \mathbb{R}^{\bar{m}\times\bar{\ell}}$*, whose* (m,ℓ)*-th element is* $\|\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}_\ell\|^2$*. Then,*

1) $\operatorname{rank}(\mathbf{D}_1) \leq \rho + 2$.
2) *denoting as* $\mathcal{V}_1$ *the null space of* $\mathbf{D}_1$*, it holds that* $\mathcal{V}_1 \neq \emptyset$ *and* $\mathcal{V}_1 \perp \mathbf{1}_{\bar{\ell}}$.

Proof. Consider a square Euclidean Distance Matrix $\bar{\mathbf{D}} \in \mathbb{R}^{\bar{m}\times\bar{m}}$ generated by vectors $\boldsymbol{\vartheta}_m$. From Lemma 1 we know that $\operatorname{rank}(\bar{\mathbf{D}}) \leq \rho + 2$. The rectangular matrix $\mathbf{D}_1 \in \mathbb{R}^{\bar{m}\times\bar{\ell}}$ is obtained from $\bar{\mathbf{D}}$ by deleting $\bar{m} - \bar{\ell}$ columns. Therefore, its rank cannot be larger than the rank of $\bar{\mathbf{D}}$. Hence, we conclude that

$$\operatorname{rank}(\mathbf{D}_1) \leq \rho + 2 \tag{30}$$

which proves 1).

Under the assumption that $\bar{\ell} > \rho + 2$, the $\bar{\ell}$ columns of matrix $\mathbf{D}_1$ are linearly dependent. Therefore, its null space $\mathcal{V}_1 \neq \emptyset$ and

$$\dim \mathcal{V}_1 \geq \bar{\ell} - \rho - 2. \tag{31}$$

Applying now Lemma 1 to $\bar{\mathbf{D}}$ and deleting the last $\bar{m} - \bar{\ell}$ rows from (29), proves the existence of a vector $\mathbf{p} \in \mathbb{R}^{\bar{m}}$ such that $\mathbf{D}_1^{\mathsf{T}}\mathbf{p} = \mathbf{1}_{\bar{\ell}}$. Hence, the Fundamental Theorem of Linear Algebra [56] states that

$$\mathcal{V}_1 \perp \mathbf{1}_{\bar{\ell}} \tag{32}$$

which proves 2). □

B. Least Singular Values of Gaussian Kernel Matrices

Using now Lemma 2, we can state and prove one of the main results of this paper, which establishes that the least singular value of a non-square Gaussian kernel matrix $\mathbf{K}(\varepsilon)$ (without loss of generality, from now on we will omit the subscripts (N, D) when referring to the kernel matrix $\mathbf{K}(\varepsilon)$) decays sharply as the shape parameter $\varepsilon \to 0$. This is the main theoretical result that supports the proposed Algorithm 1. We have the following

Theorem 1. *Consider a Gaussian Kernel Matrix* $\mathbf{K}(\varepsilon)$ *as defined in* (10)*, with* $\bar{m} > \bar{\ell} > \rho + 2$*. For* $\varepsilon \to 0$*, the least singular value* $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$ *decays asymptotically as*

$$\underline{\sigma}_{\mathbf{K}}(\varepsilon) = O(\varepsilon^{\nu}) \quad \text{as } \varepsilon \to 0, \;\; \nu \in \mathbb{N}, \;\; \nu \geq 4. \tag{33}$$

Proof. Consider the Euclidean Distance Matrix $\mathbf{D}_1 \in \mathbb{R}^{\bar{m}\times\bar{\ell}}$, whose (m, ℓ)-th element is $\|\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}_\ell\|^2$. The Gaussian kernel matrix $\mathbf{K}(\varepsilon)$ defined in (10) is obtained by $\mathbf{D}_1$ by multiplying each element by $-\varepsilon^2$ and taking the exponential. Since we are interested in $\varepsilon \to 0$, we can form the elementwise MacLaurin expansion, which can be compactly written as

$$\mathbf{K}(\varepsilon) = \sum_{p=0}^{\infty} \frac{(-1)^p}{p!}\varepsilon^{2p}\mathbf{D}_p \tag{34}$$

where $\mathbf{D}_p$ denotes the p-th element-wise (Hadamard) power of matrix $\mathbf{D}_1$, i.e. $(\mathbf{D}_p)_{m,\ell} = \|\boldsymbol{\vartheta}_m - \boldsymbol{\vartheta}_\ell\|^{2p}$.

The least singular value $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$ of $\mathbf{K}(\varepsilon)$ is characterized as

$$\underline{\sigma}_{\mathbf{K}}(\varepsilon) = \min_{\|\mathbf{z}\|=1} \|\mathbf{K}(\varepsilon)\mathbf{z}\|. \tag{35}$$

Using (34) we can write (35) as

$$\underline{\sigma}_{\mathbf{K}}(\varepsilon) = \min_{\|\mathbf{z}\|=1} \left\| \sum_{p=0}^{\infty} \frac{(-1)^p}{p!}\varepsilon^{2p}\mathbf{D}_p\mathbf{z} \right\|, \tag{36}$$

which expresses the least singular value as the minimum length of a vector expressed as an even power series of ε. Therefore, the leading power of $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$ for $\varepsilon \to 0$ corresponds to the first term $\nu = 2p$ in (36) for which $\mathbf{D}_p\mathbf{z}$ that does not vanish identically, for any possible choice of the testing vector $\mathbf{z}$. This information is obtained by characterizing the null spaces $\mathcal{V}_p$ of matrices $\mathbf{D}_p$ for increasing p.

- $p = 0$. The null space $\mathcal{V}_0$ is readily obtained noting that $\mathbf{D}_0 = \mathbf{1}_{\bar{m}\times\bar{\ell}}$ is the matrix of all ones. Thus, the null space $\mathcal{V}_0$ is composed by all the vectors $\mathbf{z} \in \mathbb{R}^{\bar{\ell}}$ that are orthogonal to $\mathbf{1}_{\bar{\ell}}$, i.e.

  $$\mathcal{V}_0 = \{\mathbf{z} \in \mathbb{R}^{\bar{\ell}} \;:\; \mathbf{z}^{\mathsf{T}} \cdot \mathbf{1}_{\bar{\ell}} = 0\}. \tag{37}$$

  Taking any $\mathbf{z} \in \mathcal{V}_0$ with $\|\mathbf{z}\| = 1$ leads to $\mathbf{D}_p\mathbf{z} = \mathbf{0}$. This proves that the least singular value $\underline{\sigma}_{\mathbf{K}}(\varepsilon) \to 0$ for $\varepsilon \to 0$ with a leading power at least $\nu = 2$.
- $p = 1$. Since by assumption $\bar{\ell} > \rho + 2$, Lemma 2 states that the null space $\mathcal{V}_1$ of $\mathbf{D}_1$ is non-empty with dimension larger than $\bar{\ell} - \rho - 2$. Lemma 2 further states that $\mathcal{V}_1$ is orthogonal to the vector $\mathbf{1}_{\bar{\ell}}$. Recalling definition (37), we conclude that $\mathcal{V}_1 \subseteq \mathcal{V}_0$ and that $\mathcal{V}_1 \cap \mathcal{V}_0 = \mathcal{V}_1 \neq \emptyset$. Thus, it is always possible to find a nontrivial vector $\mathbf{z} \in \mathcal{V}_1$ that cancels both the terms $\mathbf{D}_1\mathbf{z}$ and $\mathbf{D}_0\mathbf{z}$ in (36). This proves that the leading power of $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$ for $\varepsilon \to 0$ is at least $\nu = 4$.
- $p = 2$. An explicit characterization of the null space $\mathcal{V}_2$ of matrix $\mathbf{D}_2$ is not available, so we cannot draw any conclusion on whether $\mathbf{D}_2\mathbf{z}$ vanishes for some nontrivial vector $\mathbf{z}$.

We therefore conclude that the scaling exponent for the least singular value $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$ for $\varepsilon \to 0$ is at least $\nu \geq 4$. □

C. From Kernel to Regressor Matrices

Given the asymptotic scaling estimate for the least singular value of $\mathbf{K}(\varepsilon)$ provided in Theorem 1, we now show that the same scaling exponent propagates to the various PSK matrices and applies also to the regressor matrix $\boldsymbol{\Psi}(\varepsilon)$. We proceed in steps, since the latter is constructed by assembling block matrices with a particular structure, see Sec. II.

First, we consider how the least singular value $\underline{\sigma}_{\boldsymbol{\Phi}\otimes\mathbf{K}}(\varepsilon)$ of matrix $\boldsymbol{\Phi} \otimes \mathbf{K}(\varepsilon)$ behaves as $\varepsilon \to 0$. A standard result in linear algebra states that the singular values of a Kronecker product $\boldsymbol{\Phi} \otimes \mathbf{K}(\varepsilon)$ are obtained as all product combinations between the singular values of the two individual matrices $\boldsymbol{\Phi}$ and $\mathbf{K}(\varepsilon)$. Therefore, the least singular value $\underline{\sigma}_{\boldsymbol{\Phi}\otimes\mathbf{K}}(\varepsilon)$ is the product of the least singular values of both matrix terms

$$\underline{\sigma}_{\boldsymbol{\Phi}\otimes\mathbf{K}}(\varepsilon) = \underline{\sigma}_{\mathbf{K}}(\varepsilon) \cdot \underline{\sigma}_{\boldsymbol{\Phi}} \tag{38}$$

where $\underline{\sigma}_{\boldsymbol{\Phi}}$ is constant with respect to ε. This implies that

$$\underline{\sigma}_{\boldsymbol{\Phi}\otimes\mathbf{K}}(\varepsilon) = O(\varepsilon^{\nu}) \quad \text{as } \varepsilon \to 0, \tag{39}$$

with the same power law $\nu \geq 4$ that applies to $\underline{\sigma}_{\mathbf{K}}(\varepsilon)$.

The above consideration enables us to proceed with matrices $\boldsymbol{\Gamma}^{\mu}(\varepsilon)$ and $\boldsymbol{\Xi}^{\mu}_{(i,j)}(\varepsilon)$ forming the building blocks of the regressor matrix in (6). These matrices share the same basic structure, addressed in the following

Lemma 3. *Let $\mathbf{\Phi}$ and $\mathbf{K}(\varepsilon)$ be defined as in* (9) *and* (10)*. Assuming $\mathbf{\Delta}$ to be a generic real-valued nonsingular square matrix of compatible size, for $\varepsilon \to 0$ the least singular value of $\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon))$ decays as ε^ν, $\nu \in \mathbb{N}$, $\nu \geq 4$.*

Proof. For any fixed ε, let us denote the singular value decomposition of $\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon)$ as $\mathbf{U}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{V}^{\mathsf{T}}(\varepsilon)$. The singular values, stacked in descending order along the diagonal of matrix $\mathbf{\Sigma}(\varepsilon)$, are denoted as $\sigma_t(\varepsilon)$, $t = 1, \dots, \bar{t}$, and $\mathbf{U}^{\mathsf{T}}(\varepsilon)\mathbf{U}(\varepsilon) = \mathbf{I}$, $\mathbf{V}^{\mathsf{T}}(\varepsilon)\mathbf{V}(\varepsilon) = \mathbf{I}$. We know from (39) that $\sigma_{\bar{t}}(\varepsilon) = \underline{\sigma}_{\mathbf{\Phi}\otimes\mathbf{K}}(\varepsilon)$ decays as $O(\varepsilon^\nu)$, with $\nu \geq 4$, for $\varepsilon \to 0$.

Consider now

$$\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon)) = \mathbf{\Delta} \cdot \mathbf{U}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{V}^{\mathsf{T}}(\varepsilon) \tag{40}$$

and evaluate the QR decomposition $\mathbf{\Delta} \cdot \mathbf{U}(\varepsilon) = \mathbf{Q}(\varepsilon)\mathbf{R}(\varepsilon)$, where $\mathbf{R}(\varepsilon) \in \mathbb{R}^{\bar{t}\times\bar{t}}$ is upper triangular and $\mathbf{Q}^{\mathsf{T}}(\varepsilon)\mathbf{Q}(\varepsilon) = \mathbf{I}$. We have

$$\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon)) = \mathbf{Q}(\varepsilon)\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{V}(\varepsilon)^{\mathsf{T}}. \tag{41}$$

For later use, we denote the columns of $\mathbf{R}(\varepsilon)$ as $\mathbf{r}_t$, i.e. $\mathbf{R}(\varepsilon) = \begin{pmatrix}\mathbf{r}_1 & \cdots & \mathbf{r}_{\bar{t}}\end{pmatrix}$. Since the columns of $\mathbf{U}(\varepsilon)$ are mutually orthogonal and $\mathbf{\Delta}$ is nonsingular, then $\mathbf{\Delta}\mathbf{U}(\varepsilon)$ has full column rank, which implies that $\mathbf{R}(\varepsilon)$ is square and invertible and its columns $\mathbf{r}_t$ are linearly independent.

Now, we evaluate the singular value decomposition of the inner matrix product $\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon) = \hat{\mathbf{U}}(\varepsilon)\hat{\mathbf{\Sigma}}(\varepsilon)\hat{\mathbf{V}}^{\mathsf{T}}(\varepsilon)$, whose singular values are denoted as $\hat{\sigma}_t(\varepsilon)$, for $t = 1, \dots, \bar{t}$ with $\hat{\mathbf{U}}^{\mathsf{T}}(\varepsilon)\hat{\mathbf{U}}(\varepsilon) = \mathbf{I}$ and $\hat{\mathbf{V}}^{\mathsf{T}}(\varepsilon)\hat{\mathbf{V}}(\varepsilon) = \mathbf{I}$. A direct substitution in (41) leads to

$$\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon)) = \widetilde{\mathbf{U}}(\varepsilon)\hat{\mathbf{\Sigma}}(\varepsilon)\widetilde{\mathbf{V}}^{\mathsf{T}}(\varepsilon) \tag{42}$$

where matrices $\widetilde{\mathbf{U}}(\varepsilon) = \mathbf{Q}(\varepsilon)\hat{\mathbf{U}}(\varepsilon)$ and $\widetilde{\mathbf{V}}(\varepsilon) = \mathbf{V}(\varepsilon)\hat{\mathbf{V}}(\varepsilon)$ inherit the orthogonality conditions on their factors, as $\widetilde{\mathbf{U}}^{\mathsf{T}}(\varepsilon)\widetilde{\mathbf{U}}(\varepsilon) = \mathbf{I}$ and $\widetilde{\mathbf{V}}^{\mathsf{T}}(\varepsilon)\widetilde{\mathbf{V}}(\varepsilon) = \mathbf{I}$. Thus, expression (42) provides a singular value decomposition of $\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon))$, which is unique up to a reordering of the singular values. As usual, we assume the latter to be sorted in descending order. The result in (42) highlights that $\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)$ and $\mathbf{\Delta}\cdot(\mathbf{\Phi}\otimes\mathbf{K}(\varepsilon))$ share the same singular values. Hence, the asymptotic behavior of the least singular value of $\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon))$ can be inferred from the behavior of the least singular value $\hat{\sigma}_{\bar{t}}(\varepsilon)$ of $\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)$, which is characterized as

$$\hat{\sigma}_{\bar{t}}(\varepsilon) = \min_{\|\mathbf{w}\|=1} \|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}\|. \tag{43}$$

Let us now define $\mathbf{w}_{\bar{t}} = (0, \cdots, 0, 1)^{\mathsf{T}}$, so that any unit length vector $\mathbf{w}$ can be written as

$$\mathbf{w} = \alpha\mathbf{w}_{\bar{t}} + \beta\mathbf{v}, \quad \mathbf{v} \perp \mathbf{w}_{\bar{t}}, \quad \|\mathbf{v}\| = 1, \quad \alpha^2 + \beta^2 = 1. \tag{44}$$

Applying matrix product $\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)$ to this vector leads to

$$\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w} = \alpha\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}_{\bar{t}} + \beta\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{v} \tag{45}$$

Noting that $\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}_{\bar{t}} = \sigma_{\bar{t}}(\varepsilon)\mathbf{r}_{\bar{t}}(\varepsilon)$, we can write

$$\|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}\| = \left\|\alpha\sigma_{\bar{t}}(\varepsilon)\mathbf{r}_{\bar{t}} + \beta\sum_{t<\bar{t}}\mathbf{r}_t\sigma_t(\varepsilon)v_t\right\| \tag{46}$$

where v_t is the t-th component of vector $\mathbf{v}$. We know that for $\varepsilon \to 0$ the least (and last) singular value $\sigma_{\bar{t}}(\varepsilon)$ decays as $O(\varepsilon^\nu)$ with $\nu \geq 4$, and that any other singular value $\sigma_t(\varepsilon)$ for $t < \bar{t}$ cannot decay faster than $\sigma_{\bar{t}}(\varepsilon)$. We have two possible cases

1) $\sigma_t(\varepsilon) = O(\varepsilon^\nu)$, $\forall t$
   In this case it holds that $\|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}\| = O(\varepsilon^\nu)$ for any $\mathbf{w}$ written as in (44), since all terms in (46) have the same scaling law for $\varepsilon \to 0$.
2) $\sigma_t(\varepsilon) = O(\varepsilon^\gamma)$, with $\gamma < \nu$ for some t
   In this case there exists a vector $\mathbf{w}_t$ for which $\eta_t(\varepsilon) = \|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}_t\| = O(\varepsilon^\gamma)$ (pick an all-zero vector with a single unit entry at component t). For this choice of vector, the term $\eta_t(\varepsilon) = \|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}_t\|$ decays slower than $\eta_{\bar{t}}(\varepsilon) = \|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}_{\bar{t}}\|$ for $\varepsilon \to 0$. However, even if for some ε it may be the case that $\eta_t(\varepsilon) < \eta_{\bar{t}}(\varepsilon)$, there exists a value ε^* so that $\eta_{\bar{t}}(\varepsilon) < \eta_t(\varepsilon)$ for $\varepsilon < \varepsilon^*$. Therefore, asymptotically the vector $\mathbf{w} = \mathbf{w}_{\bar{t}}$ leads to the smallest value of $\eta(\varepsilon)$ for $\varepsilon \to 0$.

These considerations lead us to conclude that

$$\min_{\|\mathbf{w}\|=1} \|\mathbf{R}(\varepsilon)\mathbf{\Sigma}(\varepsilon)\mathbf{w}\| = \underline{\sigma}\{\mathbf{\Delta} \cdot (\mathbf{\Phi} \otimes \mathbf{K}(\varepsilon))\} = O(\varepsilon^\nu) \tag{47}$$

with a scaling exponent $\nu \geq 4$. □

To keep our analysis as general as possible, we should recall that matrices $\mathbf{\Gamma}^\mu(\varepsilon)$ and $\mathbf{\Xi}^\mu_{(i,j)}(\varepsilon)$ in (7) are built, respectively, from Gaussian kernel matrices $\mathbf{K}_N(\varepsilon)$ and $\mathbf{K}_D(\varepsilon)$ associated to numerator and denominator, which may be different. Therefore, if we assume that for $\varepsilon \to 0$ the associated least singular values $\underline{\sigma}_{\mathbf{K}_N}(\varepsilon)$ and $\underline{\sigma}_{\mathbf{K}_D}(\varepsilon)$ decay as ε^{ν_N} and ε^{ν_D}, respectively, Lemma 3 ensures that

$$\underline{\sigma}_{\mathbf{\Gamma}}(\varepsilon) = O(\varepsilon^{\nu_N}), \quad \underline{\sigma}_{\mathbf{\Xi}_{(i,j)}}(\varepsilon) = O(\varepsilon^{\nu_D}) \tag{48}$$

as $\varepsilon \to 0$, with both $\nu_D \geq 4$ and $\nu_N \geq 4$.

D. Assembling regressor matrices

Now that all the main building blocks have been characterized, we can derive an asymptotic estimate for the singular value $\underline{\sigma}_{\mathbf{\Psi}}(\varepsilon)$ of matrix $\mathbf{\Psi}^\mu(\varepsilon)$ in (6). To this end, we define two auxiliary matrices

$$\mathbf{\Psi}^\mu_L(\varepsilon) = \begin{pmatrix}\mathbf{\Gamma}^\mu(\varepsilon) & & \\ & \ddots & \\ & & \mathbf{\Gamma}^\mu(\varepsilon)\end{pmatrix}, \quad \mathbf{\Psi}^\mu_R(\varepsilon) = \begin{pmatrix}\mathbf{\Xi}^\mu_{(1,1)}(\varepsilon) \\ \vdots \\ \mathbf{\Xi}^\mu_{(P,P)}(\varepsilon)\end{pmatrix} \tag{49}$$

so that

$$\mathbf{\Psi}^\mu(\varepsilon) = \begin{pmatrix}\mathbf{\Psi}^\mu_L(\varepsilon) & \mathbf{\Psi}^\mu_R(\varepsilon)\end{pmatrix} \tag{50}$$

We study $\mathbf{\Psi}^\mu_L(\varepsilon)$ and $\mathbf{\Psi}^\mu_R(\varepsilon)$ separately.

By construction, the singular values of $\mathbf{\Psi}^\mu_L(\varepsilon)$ are the same of matrix $\mathbf{\Gamma}^\mu(\varepsilon)$, replicated P^2 times. Therefore, it is trivial to see that also $\underline{\sigma}_{\mathbf{\Psi}_L}(\varepsilon) = O(\varepsilon^{\nu_N})$ as $\varepsilon \to 0$.

In case of matrix $\mathbf{\Psi}^\mu_R(\varepsilon)$, we have the following result

Lemma 4. *Let $\mathbf{\Psi}^\mu_R(\varepsilon)$ be defined as in* (49)*. Assuming that for all $i,\ j = 1, \dots, P$ the least singular value of each block $\underline{\sigma}_{\mathbf{\Xi}_{(i,j)}(\varepsilon)} = O(\varepsilon^{\nu_D})$ as $\varepsilon \to 0$, then also*

$$\underline{\sigma}_{\mathbf{\Psi}_R}(\varepsilon) = O(\varepsilon^{\nu_D}), \quad \varepsilon \to 0. \tag{51}$$

Proof. From Section II we recall that

$$\boldsymbol{\Xi}_{(i,j)}(\varepsilon) = -\breve{\mathbf{H}}_{(i,j)}\mathbf{W}^{\mu-1}\mathbf{X}_D(\varepsilon) \tag{52}$$

Let us denote the singular value decomposition of $\mathbf{X}_D(\varepsilon)$ as $\mathbf{U}(\varepsilon)\boldsymbol{\Sigma}(\varepsilon)\mathbf{V}^{\mathsf{T}}(\varepsilon)$. The singular values, stacked in descending order along the diagonal of matrix $\boldsymbol{\Sigma}(\varepsilon)$, are denoted as $\sigma_t(\varepsilon)$, $t = 1, \dots, \bar{t}$, with $\mathbf{U}^{\mathsf{T}}(\varepsilon)\mathbf{U}(\varepsilon) = \mathbf{I}$ and $\mathbf{V}^{\mathsf{T}}(\varepsilon)\mathbf{V}(\varepsilon) = \mathbf{I}$. We can thus write

$$\boldsymbol{\Psi}_R^{\mu}(\varepsilon) = \underbrace{\begin{pmatrix} -\breve{\mathbf{H}}_{(1,1)}\mathbf{W}^{\mu-1}\mathbf{U}(\varepsilon) \\ \vdots \\ -\breve{\mathbf{H}}_{(P,P)}\mathbf{W}^{\mu-1}\mathbf{U}(\varepsilon) \end{pmatrix}}_{\mathbf{L}(\varepsilon)} \boldsymbol{\Sigma}(\varepsilon)\mathbf{V}^{\mathsf{T}}(\varepsilon) \tag{53}$$

which has the same structure as (40) with $\mathbf{L}(\varepsilon)$ replacing $\boldsymbol{\Delta} \cdot \mathbf{U}(\varepsilon)$. Therefore, following the same steps in the proof of Lemma 3 leads to the conclusion

$$\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon) = O(\varepsilon^{\nu_D}) \tag{54}$$

as $\varepsilon \to 0$. □

Now that we have a complete characterization of the singular values of $\underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon)$ and $\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon)$, we can state our final result

Theorem 2. *Let $\boldsymbol{\Psi}^{\mu}(\varepsilon)$ be defined as in* (50)*. If for $\varepsilon \to 0$ it holds that $\underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon) = O(\varepsilon^{\nu_N})$ and $\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon) = O(\varepsilon^{\nu_D})$, then*

$$\underline{\sigma}_{\boldsymbol{\Psi}}(\varepsilon) = O(\varepsilon^{\tau}), \quad \tau = \max\{\nu_N,\ \nu_D\}. \tag{55}$$

Proof. The least singular value to be characterized is

$$\underline{\sigma}_{\boldsymbol{\Psi}}(\varepsilon) = \min_{\|\mathbf{w}\|=1} \|\boldsymbol{\Psi}^{\mu}(\varepsilon)\mathbf{w}\|. \tag{56}$$

We assume

$$\mathbf{w} = \begin{pmatrix} \alpha\mathbf{w}_L \\ \beta\mathbf{w}_R \end{pmatrix} \tag{57}$$

with $\|\mathbf{w}_L\| = \|\mathbf{w}_R\| = 1$ and $\alpha^2 + \beta^2 = 1$ to ensure $\|\mathbf{w}\| = 1$. Defining the the following singular value decompositions

$$\begin{aligned} \boldsymbol{\Psi}_L^{\mu}(\varepsilon) &= \mathbf{U}_L(\varepsilon)\boldsymbol{\Sigma}_L(\varepsilon)\mathbf{V}_L^{\mathsf{T}}(\varepsilon) \\ \boldsymbol{\Psi}_R^{\mu}(\varepsilon) &= \mathbf{U}_R(\varepsilon)\boldsymbol{\Sigma}_R(\varepsilon)\mathbf{V}_R^{\mathsf{T}}(\varepsilon) \end{aligned} \tag{58}$$

we can write

$$\begin{aligned} \boldsymbol{\Psi}^{\mu}(\varepsilon)\mathbf{w} = \alpha\mathbf{U}_L(\varepsilon)\boldsymbol{\Sigma}_L(\varepsilon)\mathbf{V}_L^{\mathsf{T}}(\varepsilon)\mathbf{w}_L \\ + \beta\mathbf{U}_R(\varepsilon)\boldsymbol{\Sigma}_R(\varepsilon)\mathbf{V}_R^{\mathsf{T}}(\varepsilon)\mathbf{w}_R \end{aligned} \tag{59}$$

For any fixed ε, let us consider the first term in (59), and pick $\mathbf{w}_L = \underline{\mathbf{v}}_L$, where $\underline{\mathbf{v}}_L$ is the last column of $\mathbf{V}_L(\varepsilon)$. Then, $\mathbf{V}_L^{\mathsf{T}}(\varepsilon)\underline{\mathbf{v}}_L = (0, \dots, 1)^{\mathsf{T}}$. The matrix $\boldsymbol{\Sigma}_L(\varepsilon)$ contains the singular values of $\boldsymbol{\Psi}_L^{\mu}(\varepsilon)$ sorted in descending order, therefore $\boldsymbol{\Sigma}_L(\varepsilon)\mathbf{V}_L^{\mathsf{T}}(\varepsilon)\underline{\mathbf{v}}_L = (0, \dots, \underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon))^{\mathsf{T}}$ and the first term of (59) reduces to

$$\alpha\mathbf{U}_L(\varepsilon)\boldsymbol{\Sigma}_L(\varepsilon)\mathbf{V}_L^{\mathsf{T}}(\varepsilon)\underline{\mathbf{v}}_L = \alpha\underline{\mathbf{u}}_L\underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon) \tag{60}$$

where $\underline{\mathbf{u}}_L$ is the last column of $\mathbf{U}_L(\varepsilon)$. Repeating the same procedure on the second term of (59), we choose vector $\mathbf{w}_R = \underline{\mathbf{v}}_R$ to be the last column of $\mathbf{V}_R(\varepsilon)$, obtaining

$$\beta\mathbf{U}_R(\varepsilon)\boldsymbol{\Sigma}_R(\varepsilon)\mathbf{V}_R^{\mathsf{T}}(\varepsilon)\underline{\mathbf{v}}_R = \beta\underline{\mathbf{u}}_R\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon) \tag{61}$$

where $\underline{\mathbf{u}}_R$ is the last column of $\mathbf{U}_R(\varepsilon)$. For this particular choice of $\mathbf{w}$ we have

$$\mathbf{w} = \begin{pmatrix} \alpha\underline{\mathbf{v}}_L \\ \beta\underline{\mathbf{v}}_R \end{pmatrix} \ \rightarrow\ \boldsymbol{\Psi}^{\mu}(\varepsilon)\mathbf{w} = \alpha\underline{\mathbf{u}}_L\underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon) + \beta\underline{\mathbf{u}}_R\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon) \tag{62}$$

where all vector components $\mathbf{w}$, $\underline{\mathbf{v}}_{L,R}$, $\underline{\mathbf{u}}_{L,R}$ have unit norm and the two least singular value terms $\underline{\sigma}_{\boldsymbol{\Psi}_{L,R}}(\varepsilon)$ scale as $O(\varepsilon^{\nu_{N,D}})$ for $\varepsilon \to 0$, respectively. Therefore, when $\mathbf{w}$ is chosen as in (62), we have

$$\min_{\|\mathbf{w}\|=1} \left\|\alpha\underline{\mathbf{u}}_L\underline{\sigma}_{\boldsymbol{\Psi}_L}(\varepsilon) + \beta\underline{\mathbf{u}}_R\underline{\sigma}_{\boldsymbol{\Psi}_R}(\varepsilon)\right\| = O(\varepsilon^{\tau}) \tag{63}$$

with $\tau = \max\{\nu_N,\ \nu_D\}$, where the maximum is attained by choosing $\alpha = 1, 0$ and $\beta = 0, 1$ respectively.

Using a similar argument as in the proof of Lemma 3, it is easy to show that any different choice of $\mathbf{w}$ leads to a scaling exponent for $\varepsilon \to 0$ that is cannot be larger than τ. Therefore, the minimum is attained by (63) and we conclude that

$$\underline{\sigma}_{\boldsymbol{\Psi}}(\varepsilon) = \min_{\|\mathbf{w}\|=1} \|\boldsymbol{\Psi}^{\mu}(\varepsilon)\mathbf{w}\| = O(\varepsilon^{\tau}) \tag{64}$$

with $\tau = \max\{\nu_N,\ \nu_D\}$. □

Since both $\nu_N \geq 4$ and $\nu_D \geq 4$, we also have $\tau \geq 4$, which provides the sharp least singular value decay estimate that supports and justifies in the procedure proposed in Section IV for shape parameter estimation.

REFERENCES

[1] T. Bradde, P. Toledo, M. De Stefano, A. Zanco, S. Grivet-Talocia, and P. Crovetti, "Enabling fast power integrity transient analysis through parameterized small-signal macromodels," in *EMC Europe 2019*, Barcellona (Spain), 2–6 Sept, 2019, pp. 759–764.

[2] C. P. Coelho, J. R. Phillips, and L. M. Silveira, "Passive constrained rational approximation algorithm using nevanlinna-pick interpolation," in *Proceedings 2002 Design, Automation and Test in Europe Conference and Exhibition*, Paris (France), 4–8 March, 2002, pp. 923–930

[3] M. Elzinga, K. L. Virga, and J. L. Prince, "Improved global rational approximation macromodeling algorithm for networks characterized by frequency-sampled data," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 9, pp. 1461–1468, Jan., 2000.

[4] B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Trans. Power Del*, vol. 14, no. 3, pp. 1052–1061, Jul., 1999.

[5] R. Achar, P. K. Gunupudi, M. Nakhla, and E. Chiprout, "Passive interconnect reduction algorithm for distributed/measured networks," *IEEE Trans. Circuits Syst. II. Analog Digit. Signal Process.*, vol. 47, no. 4, pp. 287–301, Apr., 2000.

[6] B. Anderson and A. Antoulas, "Rational interpolation and state-variable realizations," *Linear Algebra and its Applications*, vol. 137, pp. 479–509, Sept., 1990.

[7] A. Mayo and A. Antoulas, "A framework for the solution of the generalized realization problem," *Linear algebra and its applications*, vol. 425, no. 2-3, pp. 634–662, Sept., 2007.

[8] M. T. Kassis, M. Kabir, Y. Q. Xiao, and R. Khazaka, "Passive reduced order macromodeling based on loewner matrix interpolation," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 8, pp. 2423–2432, Aug., 2016.

[9] S. Grivet-Talocia and R. Trinchero, "Behavioral, parameterized, and broadband modeling of wired interconnects with internal discontinuities," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 1, pp. 77–85, Jul., 2018.

[10] N. M. Nakhla, A. Dounavis, R. Achar, and M. S. Nakhla, "Depact: Delay extraction-based passive compact transmission-line macromodeling algorithm," *IEEE Trans. Adv. Packag*, vol. 28, no. 1, pp. 13–23, Feb., 2005.

[11] A. Charest, D. Saraswat, M. Nakhla, R. Achar, and N. Soveiko, "Compact macromodeling of high-speed circuits via delayed rational functions," *IEEE Microw. Wireless Compon. Lett*, vol. 17, no. 12, pp. 828–830, Dec., 2007.

[12] B. Zhong, T. Hu, D. Fu, S. L. Dvorak, and J. L. Prince, "A study of a hybrid phase-pole macromodel for transient simulation of complex interconnects structures," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 8, pp. 1250–1261, Jul., 2005.

[13] A. Chinea, P. Triverio, and S. Grivet-Talocia, "Delay-based macromodeling of long interconnects from frequency-domain terminal responses," *IEEE Trans. Adv. Packag*, vol. 33, no. 1, pp. 246–256, Apr., 2010.

[14] F. Ferranti, L. Knockaert, and T. Dhaene, "Passivity-preserving parametric macromodeling by means of scaled and shifted state-space systems," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 10, pp. 2394–2403, Sept., 2011.

[15] E. R. Samuel, L. Knockaert, F. Ferranti, and T. Dhaene, "Guaranteed passive parameterized macromodeling by using sylvester state-space realizations," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 4, pp. 1444–1454, Apr., 2013.

[16] P. Triverio, S. Grivet-Talocia, and M. S. Nakhla, "A parameterized macromodeling strategy with uniform stability test," *IEEE Trans. Adv. Packag*, vol. 32, no. 1, pp. 205–215, Feb., 2009.

[17] D. Deschrijver, L. Knockaert, and T. Dhaene, "A barycentric vector fitting algorithm for efficient macromodeling of linear multiport systems," *IEEE Microw. Wireless Compon. Lett*, vol. 23, no. 2, pp. 60–62, Feb., 2013.

[18] A. C. Antoulas and B. D. O. Anderson, "On the scalar rational interpolation problem," *IMA Journal of Mathematical Control and Information*, vol. 3, no. 2-3, pp. 61–88, Sept., 1986.

[19] A. C. Antoulas, *Approximation of large-scale dynamical systems*. Society for Industrial and Applied Mathematics, 2005.

[20] A. Zanco, S. Grivet-Talocia, T. Bradde, and M. De Stefano, "Uniformly stable parameterized macromodeling through positive definite basis functions," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 10, no. 11, pp. 1782–1794, Jul., 2020.

[21] A. Zanco and S. Grivet-Talocia, "High-dimensional parameterized macromodeling with guaranteed stability," in *2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, Montreal (Canada), 6–9 Oct., 2019, pp. 1–3.

[22] R. Schaback, "A practical guide to radial basis functions," University of Göttingen, Tech. Rep., 2007. [Online]. Available: https://num.math.uni-goettingen.de/schaback/teaching/sc.pdf

[23] D. Wirtz, N. Karajan, and B. Haasdonk, "Surrogate modeling of multiscale models using kernel methods," *International Journal for Numerical Methods in Engineering*, vol. 101, no. 1, pp. 1–28, Jan., 2015.

[24] S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation," *Advances in Computational Mathematics*, vol. 11, no. 2-3, pp. 193–210, Nov., 1999.

[25] M. Mongillo, "Choosing basis functions and shape parameters for radial basis function methods," *SIAM undergraduate research online*, vol. 4, pp. 190–209, Oct., 2011.

[26] F. Marchetti, "The extension of Rippa's algorithm beyond LOOCV," *Applied Mathematics Letters*, vol. 120, pp. 1–8, Oct., 2021.

[27] S. Elanayar V.T. and Y. C. Shin, "Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 594–603, Jul., 1994.

[28] W. Kaminski and P. Strumillo, "Kernel orthonormalization in radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1177–1183, Sept., 1997.

[29] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks," Royal Signals and Radar Establishment Malvern (UK), Tech. Rep., Mar., 1988.

[30] S. Chen, "Nonlinear time series modelling and prediction using gaussian RBF networks with enhanced clustering and RLS learning," *Electronics letters*, vol. 31, no. 2, pp. 117–118, Jan., 1995.

[31] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural comput.*, vol. 1, no. 2, pp. 281–294, Jun., 1989.

[32] W. Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 4, pp. 601–612, Jul., 1998.

[33] D. Yu, J. Gomm, and D. Williams, "A recursive orthogonal least squares algorithm for training rbf networks," *Neural Processing Letters*, vol. 5, no. 3, pp. 167–176, Jun., 1997.

[34] A. Zanco and S. Grivet-Talocia, "A mesh-free adaptive parametric macromodeling strategy with guaranteed stability," in *2020 International Symposium on Electromagnetic Compatibility - EMC EUROPE*, Rome (Italy), 23–25 Sept., 2020, pp. 1–6.

[35] T. Bradde, S. Grivet-Talocia, M. De Stefano, and A. Zanco, "A scalable reduced-order modeling algorithm for the construction of parameterized interconnect macromodels from scattering responses," in *2018 IEEE Symposium on EMC, SI & PI)*, Long Beach, (CA, USA), 30 Jul. – 3 Aug., 2018, pp. 650–655.

[36] S. Lefteriu and A. C. Antoulas, "On the convergence of the vector-fitting algorithm," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 4, pp. 1435–1443, Apr., 2013.

[37] C. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. Autom. Control*, vol. 8, no. 1, pp. 56–58, Jan., 1963.

[38] S. Grivet-Talocia, "A perturbation scheme for passivity verification and enforcement of parameterized macromodels," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 7, no. 11, pp. 1869–1881, Nov., 2017.

[39] A. Zanco, S. Grivet-Talocia, T. Bradde, and M. De Stefano, "Enforcing passivity of parameterized lti macromodels via hamiltonian-driven multivariate adaptive sampling," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst*, vol. 39, no. 1, pp. 225–238, Nov., 2018.

[40] A.-D. Cheng, "Multiquadric and its shape parameter—a numerical investigation of error estimate, condition number, and round-off error by arbitrary precision computation," *Engineering analysis with boundary elements*, vol. 36, no. 2, pp. 220–239, Feb., 2012.

[41] R. Schaback, "Multivariate interpolation by polynomials and radial basis functions," *Constructive Approximation*, vol. 21, no. 3, pp. 293–317, Feb., 2005.

[42] B. Fornberg, G. Wright, and E. Larsson, "Some observations regarding interpolants in the limit of flat radial basis functions," *Computers & mathematics with applications*, vol. 47, no. 1, pp. 37–55, Jan., 2004.

[43] E. Larsson and B. Fornberg, "Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions," *Computers & Mathematics with Applications*, vol. 49, no. 1, pp. 103–130, Jan., 2005.

[44] B. Fornberg and G. Wright, "Stable computation of multiquadric interpolants for all values of the shape parameter," *Computers & Mathematics with Applications*, vol. 48, no. 5-6, pp. 853–867, Sept., 2004.

[45] B. Fornberg, E. Larsson, and N. Flyer, "Stable computations with gaussian radial basis functions," *SIAM Journal on Scientific Computing*, vol. 33, no. 2, pp. 869–892, Jan., 2011.

[46] R. Cavoretto, G. E. Fasshauer, and M. McCourt, "An introduction to the hilbert-schmidt svd using iterated brownian bridge kernels," *Numerical Algorithms*, vol. 68, no. 2, pp. 393–422, Feb. 2015.

[47] A. Yurova and K. Kormann, "Stable evaluation of gaussian radial basis functions using hermite polynomials," *arXiv preprint arXiv:1709.02164*, Sept., 2017.

[48] A. Zanco and S. Grivet-Talocia, "Hyperparameter determination in multivariate macromodeling based on radial basis functions," in *2020 IEEE 29th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, San Jose (CA, USA), 5–7 Oct., 2020, pp. 1–3.

[49] P. Bratley and B. L. Fox, "Algorithm 659: Implementing Sobol's quasirandom sequence generator," *ACM Transactions on Mathematical Software (TOMS)*, vol. 14, no. 1, pp. 88–100, Mar., 1988.

[50] Y. Q. Xiao, S. Grivet-Talocia, P. Manfredi, and R. Khazaka, "A novel framework for parametric loewner matrix interpolation," *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 9, no. 12, pp. 2404–2417, Dec., 2019.

[51] J. Preibisch, T. Reuschel, K. Scharff, J. Balachandran, B. Sen, and C. Schuster, "Exploring efficient variability-aware analysis method for high-speed digital link design using pce," *DesignCon*, Santa Clara (CA, USA), 31 Jan. – 2 Feb., 2017.

[52] T. Buss, "2ghz low noise amplifier with the bfg425w," Philips Semiconductors, B.V., Nijmegen, The Netherlands, Tech. Rep., 1996. [Online]. Available: http://application-notes.digchip.com/004/4-7999.pdf

[53] M. D. McKay, R. J. Beckman, and W. J. Conover, "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, May., 1979.

[54] "HSPICE, Synopsis [online] http://www.synopsis.com."

[55] A. Alfakih, "On the nullspace, the rangespace and the characteristic polynomial of euclidean distance matrices," *Linear algebra and its applications*, vol. 416, no. 2-3, pp. 348–354, Jul., 2006.

[56] G. Strang, "The fundamental theorem of linear algebra," *The American Mathematical Monthly*, vol. 100, no. 9, pp. 848–855, Nov., 1993.

[57] M. De Stefano, S. Grivet-Talocia, "A multivariate adaptive sampling scheme for passivity characterization of parameterized macromodels," *SPI 2021 – 25th IEEE workshop on signal and power integrity*, pp. 1–4, 10–12 May, 2021.

Alessandro Zanco received the Bachelor Degree in Electrical Engineering in 2015 and the Master degree in Mechathronic Engineering in 2018 from Politecnico di Torino, Torino, Italy. He is currently pursuing his Ph.D in Electrical, Electronics and Communication Engineering, at Politecnico di Torino. His research interests include high-dimensional parameterized black-box modeling for EMC.

He is co-recipient of the 2018 Best Paper Award of the IEEE INTERATIONAL SYMPOSIUM ON ELECTROMAGNETIC COMPATIBILITY and of the 2019 Best Student Paper of the IEEE WORKSHOP ON SIGNAL AND POWER INTEGRITY

Stefano Grivet-Talocia (M'98–SM'07–F'18) received the Laurea and Ph.D. degrees in electronic engineering from the Politecnico di Torino, Turin, Italy. From 1994 to 1996, he was with the NASA/Goddard Space Flight Center, Greenbelt, MD, USA. He is currently a Full Professor of electrical engineering with the Politecnico di Torino. He co-founded the academic spinoff company IdemWorks in 2007, serving as the President until its acquisition by CST in 2016. He has authored about 200 journal and conference papers. His current research interests include passive macromodeling of lumped and distributed interconnect structures, model-order reduction, modeling and simulation of fields, circuits, and their interaction, wavelets, time-frequency transforms, and their applications. Dr. Grivet-Talocia was a co-recipient of the 2007 Best Paper Award of the IEEE TRANSACTIONS ON ADVANCED PACKAGING. He received the IBM Shared University Research Award in 2007, 2008, and 2009. He was an Associate Editor of the IEEE TRANSACTIONS ON ELECTROMAGNETIC COMPATIBILITY from 1999 to 2001 and He is currently serving as Associate Editor for the IEEE TRANSACTIONS ON COMPONENTS, PACKAGING AND MANUFACTURING TECHNOLOGY. He was the General Chair of the 20th and 21st IEEE Workshops on Signal and Power Integrity (SPI2016 and SPI2017).