Hyperparameter determination in multivariate macromodeling based on radial basis functions

Hyperparameter determination in multivariate macromodeling based on radial basis functions

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

***Abstract*—This paper introduces a simple and effective algorithm for the automated selection of Radial Basis Function hyperparameters in the context of high-dimensional multivariate macromodeling. Numerical results show an average speedup of at least one order of magnitude with respect to direct hyperparameter optimization.**

I. Introduction

Multivariate (parameterized) macromodels aim at reproducing the frequency behavior of complex structures through surrogate compact dynamical systems [1], whose coefficients depend on a number of free variables related to geometry, material, processes, temperature, etc. Such models are usually identified from sampled responses through a multivariate approximation, whose accuracy and compactness depend on the specific choice of some basis function sets embedded in the model equations.

It has been shown that Radial Basis Functions (RBFs) provide an excellent candidate for scalability to high parameter dimension, thanks to their mesh-free structure. RBFs are usually defined in terms of one or more *hyperparameters*, which should be carefully optimized for best performance. Various techniques exist for selecting appropriate values, usually based on optimization of cross validation and maximum likelihood estimators [2], [3] or direct optimization or grid search [4]–[6]. Here, we propose a simple yet very effective algorithm that, with negligible loss of accuracy and no overhead, allows a sub-optimal determination of hyperparameters related to Gaussian RBFs.

II. Background and Formulation

We consider a generic P-port electronic, electrical or electromagnetic structure, whose behavior depends on ρ geometric or physical parameters collected in vector $\boldsymbol{\vartheta} = [\vartheta_1, \ldots, \vartheta_\rho] \in \Theta$. The structure is characterized through its frequency responses $\breve{\mathbf{H}}_{k,m} = \breve{\mathbf{H}}(s_k, \boldsymbol{\vartheta}_m)$ known at a discrete set of $\bar{k}$ frequency $s_k = \mathrm{j}\omega_k$ and $\bar{m}$ parameter samples $\boldsymbol{\vartheta}_m$. We seek for a surrogate parametric model $\mathbf{H}(s, \boldsymbol{\vartheta})$ such that

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

in order to approximate with a controlled error both the parametric and the broadband frequency behavior of the data.

The model structure we use is well established [1]

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

where frequency dependence is captured by the basis functions $\varphi(s)$, which correspond to the standard partial fraction basis associated with Vector Fitting (VF) "basis" poles, and parameter variability is captured by the basis functions $\xi_\ell(\boldsymbol{\vartheta})$. The model coefficients $\mathbf{R}_{n,\ell}$, $r_{n,\ell}$ are computed through the so-called *Parameterized Sanathanan Koerner* (PSK) algorithm [1], [7], which is an iterative linear relaxation which converts (1) into a sequence of linear least squares problems

$$\boldsymbol{\Gamma}^\mu \mathbf{c}^\mu = \mathbf{b}, \quad \mu = 1, 2, \ldots \tag{3}$$

where $\boldsymbol{\Gamma}^\mu \in \mathbb{C}^{\bar{k}\bar{m}, (P+1)(\bar{n}+1)\bar{\ell}}$ contains, suitably ordered, products of basis functions $\varphi_n(s)$ and $\xi_\ell(\boldsymbol{\vartheta})$, see [5]. The iterations stop when the model coefficients $\mathbf{R}_{n,\ell}$, $r_{n,\ell}$, stored in unknown vector $\mathbf{c}$, stabilize.

III. Problem Statement

Several choices have been documented for the selection of the parameter basis functions $\xi_\ell(\boldsymbol{\vartheta})$, including orthogonal polynomials, Bernstein polynomials, spectral (trigonometric) expansions, and Radial Basis Functions (RBF) [4], [5], [8]. Only the latter enable mesh-free (unstructured) expansions providing a good scalability to high dimension [4], [5]. Therefore, this work focuses on Gaussian RBFs, defined as

$$\xi_\ell(\boldsymbol{\vartheta}) = e^{-\varepsilon \|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_\ell\|^2} \tag{4}$$

where $\boldsymbol{\vartheta}_\ell$ denotes the center and $\varepsilon > 0$ is the *shape parameter*. Figure 1 depicts through a projection onto a single dimension the effect of ε, which determines how "fat" (ε small) or "thin" (ε large) is the RBF.

The approximation capability of the macromodel is strongly dependent on ε, as well as the numerical steps needed to construct the model (the regressor matrix $\boldsymbol{\Gamma}^\mu = \boldsymbol{\Gamma}^\mu(\varepsilon)$ has a strong dependence on ε). The main objective of this work is to find a sub-optimal value of ε that provides accurate macromodels, with limited overhead required for its determination. In particular, we would like to avoid a direct search on the shape parameter, which would require repeated model construction for different values of ε within an optimization loop.

IV. Sub-optimal hyper-parameter selection

Let us consider the two asymptotic cases $\varepsilon \to 0$ and $\varepsilon \to \infty$.

- For $\varepsilon \to 0$, Gaussian RBFs become increasingly flat (see Fig. 1). It is well known that, under suitable conditions, RBF interpolation in this limit approaches a polynomial accuracy [9], [10], so that the model is expected to be very accurate. Unfortunately, as ε decreases, the condition number of $\boldsymbol{\Gamma}^\mu(\varepsilon)$ grows exponentially fast, making the solution of (3) numerically unstable. Figure 2 illustrates the dependence of the model-data error on ε. In case of small ε (region A), such numerical problems are evident.

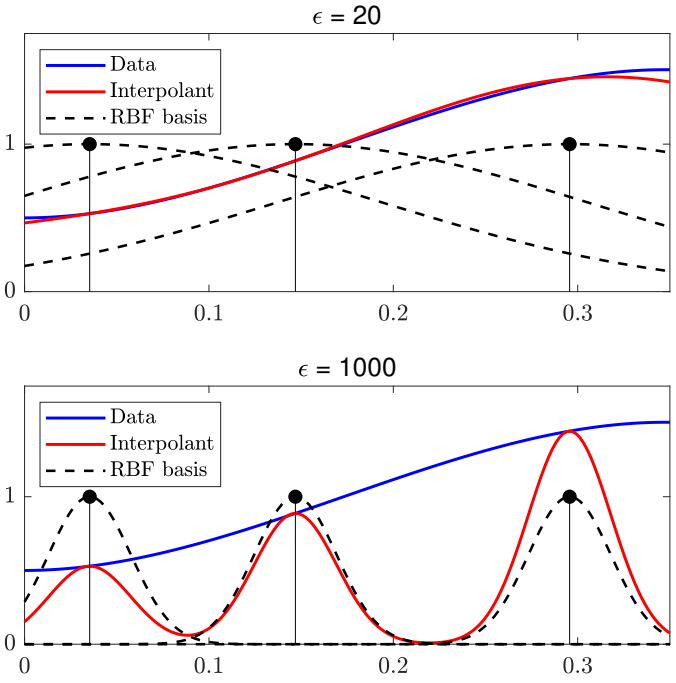

Fig. 1. Comparison of Gaussian RBF interpolants for different shape parameter values. Top panel (ε small): the interpolant accurately approximates the data. Bottom panel (ε large): the basis functions are too narrow to capture the data variability.

- When $\varepsilon \to \infty$, the Gaussian RBF $\xi_\ell(\boldsymbol{\vartheta})$ concentrates more and more around its center $\boldsymbol{\vartheta}_\ell$, until it reaches the asymptotic limit

$$\xi_\ell(\boldsymbol{\vartheta}) = \begin{cases} 1 & \text{if } \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_\ell \\ 0 & \text{if } \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_\ell \end{cases} \tag{5}$$

  A corresponding model is thus expected to have perfect accuracy only at the RBF centers, but it will be unable to provide a continuous parameterization. As expected, the model-data error increases (Fig. 2, region C).

Candidate sub-optimal shape parameters that minimize model-data error are located in the region B of Fig. 2, where they are

1) sufficiently large to avoid numerical instabilities, and
2) sufficiently small to effectively parameterize the model.

It is also advisable to minimize the condition number of $\boldsymbol{\Gamma}^\mu(\varepsilon)$ for numerical robustness, therefore, we aim at selecting the largest admissible ε located at the interface between regions B and C. This selection must be performed without explicitly computing the model-data error, which in turn would require multiple model estimation and long runtime.

As a proxy to such (inverse) condition number, we consider the least singular value $\underline{\sigma}(\varepsilon)$ of $\boldsymbol{\Gamma}^\mu(\varepsilon)$, whose typical behavior is depicted in Fig. 3. It can be proved that $\underline{\sigma}(\varepsilon)$ increases exponentially in regions A and B, until it stabilizes to a finite value for large ε in region C. It is thus sufficient to identify the corner point where the $\underline{\sigma}(\varepsilon)$ trajectory begins to flatten, in a log-log scale. To this end,

1) we precompute a set of $\bar{t}$ least singular values $\{\underline{\sigma}(\varepsilon_1), \ldots, \underline{\sigma}(\varepsilon_{\bar{t}})\}$ of matrix $\boldsymbol{\Gamma}^\mu(\varepsilon)$;

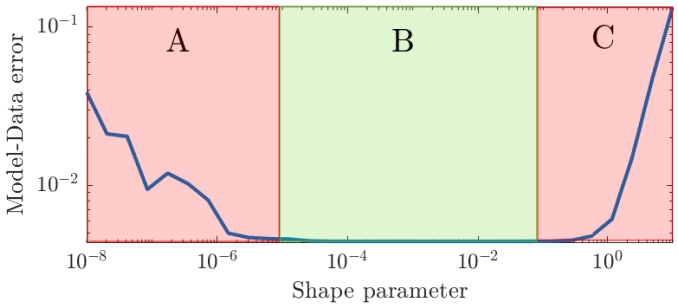

Fig. 2. Model-data error vs shape parameter ε. Region A: numerical instabilities associated with small ε. Region B: candidate sub-optimal shape parameter values. Region C: loss of parameterization capabilities.

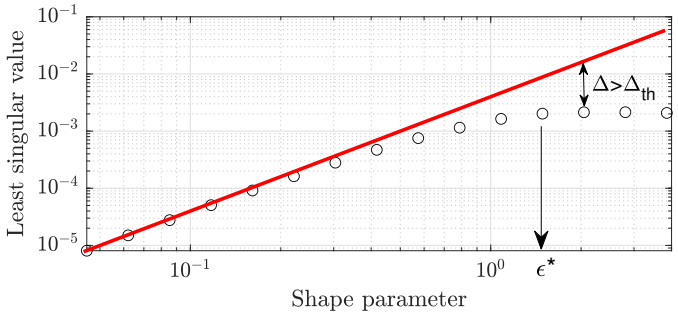

Fig. 3. Black dots: least singular values $\underline{\sigma}(\varepsilon)$ samples. Red solid line: log-log regression line evaluated on the samples $\varepsilon \leq \varepsilon^*$.

2) we iteratively build a log-log regression line on the pairs

$$\{(\varepsilon_t, \log_{10} \underline{\sigma}(\varepsilon_t)),\ t = 1, 2, \ldots, t_i\}, \quad i = 2, 3, \ldots.$$

This process stops at iteration i^* when the relative deviation Δ of the regression line with respect to the singular value sample t_{i^*+1} exceeds a predefined threshold Δ_{th}. See Fig. 3 for a graphical illustration. The sub-optimal shape parameter ε^* is thus selected as

$$\varepsilon^* = \varepsilon_{t_{i^*}} \tag{6}$$

The above procedure still requires the construction of the regression matrix $\boldsymbol{\Gamma}^\mu(\varepsilon)$ at each iteration, leading to a potentially different $\varepsilon^* = \varepsilon^*(\mu)$ and consequently different RBF expansions at each iteration. This is not even necessary, since the spectral properties of matrix $\boldsymbol{\Gamma}^\mu(\varepsilon)$ are not expected to change significantly throughout the PSK iterations. Therefore, our proposed algorithm predetermines the sub-optimal ε^* at the first iteration using $\boldsymbol{\Gamma}^1(\varepsilon)$, thus significantly improving runtime.

A second and even more efficient implementation considers the Kernel matrix

$$\mathbf{K}(\varepsilon) = \begin{pmatrix} \xi_1(\boldsymbol{\vartheta}_1) & \cdots & \xi_{\bar{\ell}}(\boldsymbol{\vartheta}_1) \\ \vdots & & \vdots \\ \xi_1(\boldsymbol{\vartheta}_{\bar{m}}) & \cdots & \xi_{\bar{\ell}}(\boldsymbol{\vartheta}_{\bar{m}}) \end{pmatrix} \in \mathbb{R}^{\bar{m} \times \bar{\ell}} \tag{7}$$

whose size is significantly smaller than $\boldsymbol{\Gamma}^1(\varepsilon)$, and whose construction requires negligible time. It can be shown (details will be presented in a forthcoming report) that the least singular value of $\mathbf{K}(\varepsilon)$ is strongly related to the corresponding

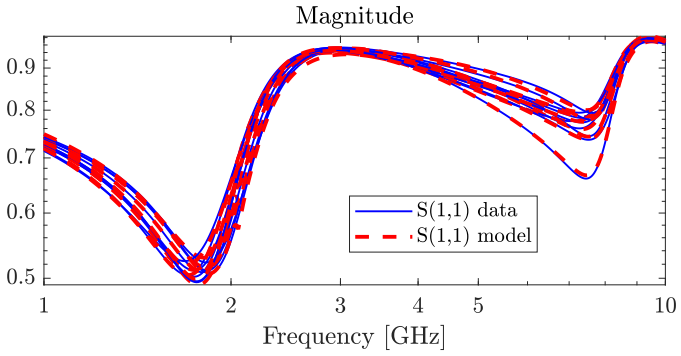

Fig. 4. Model responses (red dashed lines) compared with validation data samples (blue solid lines)

least singular value of $\mathbf{\Gamma}^1(\varepsilon)$ and, in particular, has the same dependence with ε as depicted in Fig. 3. Therefore, our second proposed implementation considers matrix $\mathbf{K}(\varepsilon)$ for the identification of the sub-optimal ε^*.

V. Numerical examples

We illustrate the proposed approach on a 10-parameter benchmark structure, the Low Noise Amplifier (LNA), originally presented in [11]. For details on the parameterization see [5]. We aim at comparing the presented automated approach with a direct search on the shape parameter ε in terms of runtime and model accuracy.

The structure is known through $\bar{m} = 2000$ frequency responses uniformly covering the parameter space, each including $\bar{k} = 701$ frequency samples in the range [1 Hz, 10 GHz]. The model to be constructed has $\bar{n} = 16$ poles and is parameterized by means of $\bar{\ell} = 40$ and $\bar{\ell} = 4$ Gaussian basis functions at numerator and denominator, respectively. The RBFs centers were randomly selected among the available samples, as in [5].

The proposed algorithm was applied to find the optimal shape parameter ε in the range $[0.001, 10]$. A total of $\bar{t} = 20$ log-spaced candidate points were used, with a stopping threshold $\Delta_{th} = 10^{-2}$. A single model extraction required 99 s. The overhead to estimate the sub-optimal shape parameter was 35 s using matrix $\mathbf{\Gamma}^1(\varepsilon)$ and negligible using the kernel matrix $\mathbf{K}(\varepsilon)$. Both approaches led to the same $\varepsilon^* = 2.98 \cdot 10^{-2}$, with a corresponding model-data error $9.2 \cdot 10^{-3}$. Figure 4 compares the model responses (red dashed lines) with data (blue solid lines) for 10 randomly chosen validation samples. As a comparison, a standard shape parameter optimization based on direct search on the same set of candidate shape parameters required 20 minutes, leading to $\varepsilon = 2.60 \cdot 10^{-3}$ and a model-data error $7.9 \cdot 10^{-3}$. Therefore, the proposed approach was able to construct a sub-optimal model of comparable accuracy with a $9.2\times$ and $12.5\times$ speedup, see Table I.

Extensive application to a larger set of benchmark structures (10 different test cases depending on 1 up to 10 parameters) led to an average speed up of about $11\times$ using $\mathbf{\Gamma}^1(\varepsilon)$ and of $21\times$ using $\mathbf{K}(\varepsilon)$. These results will be documented in a forthcoming report.

TABLE I
Performance of proposed shape parameter identification in terms of model-data error and runtime.

Method	ε^*	Error	Runtime(s)	Speed-up
Grid search	$2.60 \cdot 10^{-3}$	$7.9 \cdot 10^{-3}$	1243	-
Proposed with $\mathbf{\Gamma}(\varepsilon)$	$2.98 \cdot 10^{-2}$	$9.2 \cdot 10^{-3}$	134	$9.2\times$
Proposed with $\mathbf{K}(\varepsilon)$	$2.98 \cdot 10^{-2}$	$9.2 \cdot 10^{-3}$	99	$12.5\times$

VI. Conclusions

In the context of high-dimensional parameterized macromodeling, this paper introduced a heuristic yet effective strategy to select a sub-optimal shape parameter ε of the Gaussian Radial Basis Functions forming the model structure. The proposed method exploits the spectral properties of a small-size kernel matrix as a criterion to optimize the shape parameter. This strategy avoids a direct search, which would require repeated model construction within an optimization loop, and leads to accurate models in significantly reduced runtime.

The presented approach effectively solves only one problem in mesh-free parameterized macromodeling based on unstructured RBF expansions, but several open problems remain to be addressed, such as the automated selection of both number and centers of Gaussian RBFs.

References

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

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

[3] 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, 1999.

[4] 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)*, 2019, pp. 1–3.

[5] A. Zanco, S. Grivet-Talocia, T. Bradde, and M. De Stefano, "Uniformly stable parameterized macromodeling through positive definite basis funtions," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2020, submitted.

[6] A. Zanco and S. Grivet-Talocia, "A mesh-free adaptive parametric macromodeling strategy with guaranteed stability," in *2020 International Symposium on Electromagnetic Compatibility, Rome, Italy, 7–11 Sept.* IEEE, 2020, in press.

[7] C. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *Automatic Control, IEEE Transactions on*, vol. 8, no. 1, pp. 56–58, jan 1963.

[8] S. Grivet-Talocia and E. Fevola, "Compact parameterized black-box modeling via fourier-rational approximations," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 4, pp. 1133–1142, 2017.

[9] 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, 2004.

[10] 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, 2011.

[11] T. Buss, "2 GHz 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