Rational kernel-based interpolation for complex-valued frequency response functions

Julien Bect, Niklas Georg, Ulrich Römer, Sebastian Schöps

RATIONAL KERNEL-BASED INTERPOLATION FOR COMPLEX-VALUED FREQUENCY RESPONSE FUNCTIONS*

JULIEN BECT[‡], NIKLAS GEORG[†§], ULRICH RÖMER[†], AND SEBASTIAN SCHÖPS[§]

Abstract. This work is concerned with the kernel-based approximation of a complex-valued function from data, where the frequency response function of a partial differential equation in the frequency domain is of particular interest. In this setting, kernel methods are employed more and more frequently, however, standard kernels do not perform well. Moreover, the role and mathematical implications of the underlying pair of kernels, which arises naturally in the complex-valued case, remain to be addressed. We introduce new reproducing kernel Hilbert spaces of complex-valued functions, and formulate the problem of complex-valued interpolation with a kernel pair as minimum norm interpolation in these spaces. Moreover, we combine the interpolant with a low-order rational function, where the order is adaptively selected based on a new model selection criterion. Numerical results on examples from different fields, including electromagnetics and acoustic examples, illustrate the performance of the method, also in comparison to available rational approximation methods.

Key words. Complex-valued kernel methods, dynamical systems, frequency response function, model selection, rational approximation

AMS subject classifications. 65N99, 60G15, 46E22

1. Introduction.

We consider dynamical systems of the form

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{D}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{g}(t), \tag{1.1}$$

to be endowed with initial conditions and $\mathbf{K}, \mathbf{D}, \mathbf{M} \in \mathbb{R}^{n_h \times n_h}$, $\mathbf{u}(t), \mathbf{g}(t) \in \mathbb{R}^{n_h}$. We are in particular interested in approximating scalar time-dependent quantities derived from the solution, of the form

$$f(t) = \mathbf{j}^{\mathrm{T}}\mathbf{u}(t), \qquad \mathbf{j} \in \mathbb{R}^{n_h}, \tag{1.2}$$

which are commonly used to assess engineering designs. System (1.1) may stem from a partial differential equation after spatial discretization with n_h degrees of freedom. In a mechanics context, $\mathbf{K}, \mathbf{D}, \mathbf{M}$ are referred to as stiffness, damping and mass matrix, but problems arising in many areas of science and engineering can be brought into this form. Our numerical results will cover electromagnetic and acoustic field problems in particular. In view of the linearity of the equation, a frequency domain analysis is often adopted. Assuming for simplicity that $\mathbf{u}$ and $\dot{\mathbf{u}}$ vanish at $t = 0$, the (one-sided) Laplace transform of (1.1)–(1.2) with respect to the time variable t is

$$\begin{aligned} \left(s^2\mathbf{M} + s\mathbf{D} + \mathbf{K}\right)\, \hat{\mathbf{u}}(s) &= \hat{\mathbf{g}}(s), \\ \hat{f}(s) &= \mathbf{j}^{\mathrm{T}}\hat{\mathbf{u}}(s), \end{aligned} \tag{1.3}$$

*Submitted to the editors DATE.

Funding: The work of N. Georg and U. Römer was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – RO4937/1-1. The work of N. Georg is also supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt.

[†]Institut für Dynamik und Schwingungen, Technische Universität Braunschweig, Braunschweig, Germany (n.georg@tu-braunschweig.de, u.roemer@tu-braunschweig.de).

[‡]Laboratoire des signaux et systèmes, CentraleSupélec, Univ. Paris-Sud, CNRS, Université Paris-Saclay, Paris, France (julien.bect@centralesupelec.fr).

[§]Computational Electromagnetics Group, Technische Universität Darmstadt, Darmstadt, Germany (sebastian.schoeps@tu-darmstadt.de).

where s denotes the complex frequency variable, also known as the Laplace variable. Assuming a suitably normalized excitation $\hat{\mathbf{g}}(s)$, the frequency response function is defined as the value $\omega \mapsto \hat{f}(\mathrm{i}\omega)$ of $\hat{f}$ on the imaginary axis, where ω is called the angular frequency, and we are typically interested more specifically in its value on a certain interval $\Omega = [\omega_{\min}, \omega_{\max}] \subseteq [0, +\infty)$. In the following, we omit explicitly indicating frequency domain variables to simplify the notation.

The location of the poles of $\hat{f}$ strongly depends on the properties of $\mathbf{K}, \mathbf{D}, \mathbf{M}$, see [38]. We assume, in particular, that no pole is placed on the frequency axis $\mathrm{i}\mathbb{R}$ and that the frequency response function is holomorphic on the shifted right half-plane $\Gamma_\alpha = \{s \in \mathbb{C} \mid \Re[s] > -\alpha\}$, $\alpha > 0$. The real parts of all poles are strictly negative for instance if $\mathbf{K}, \mathbf{D}, \mathbf{M}$ are symmetric positive definite, see Section 3 of [38]. The same holds true if the homogeneous version of (1.1) is stable, in the sense that all solutions decay exponentially to zero as $t \to \infty$. The holomorphy of response functions has recently been studied also in the context of partial differential equations, see [7, Proposition 5.3] for instance. There, the frequency response map for an acoustic scattering problem was studied and appropriate damping terms ensured a locally holomorphic response function, with a negative real part for all poles[1].

Adopting a data-driven approach, (1.3) must be solved repeatedly on a set of interpolation/training points $\omega_i \in \Omega$, with $s_i = \mathrm{i}\omega_i$. Numerical efficiency demands a small training set

$$\{\omega_i, f(\mathrm{i}\omega_i)\}_{i=1}^n, \quad \text{where } \omega_i \in \Omega, f(\mathrm{i}\omega_i) \in \mathbb{C}, i = 1, \dots, n, \tag{1.4}$$

hence, there is a need for accurate interpolation in the frequency domain.

The data-driven approximation of frequency response functions has attracted considerable interest in the literature, see for instance [16, 23, 28] and the references therein. Among the numerous available approaches we mention vector fitting [16] and the adaptive Antoulas-Anderson method [28] in particular, which are widely used, state-of-the-art approximation methods.

Vector Fitting (VF) is a rational approximation technique, specifically tailored to functions in the frequency domain. It is based on a representation in terms of partial fractions as

$$f(\mathrm{i}\omega) \approx \sum_{m=1}^{M} \frac{r_m}{\mathrm{i}\omega - p_m} + d + \mathrm{i}\omega h, \tag{1.5}$$

where the M poles p_m are relocated in each iteration by solving a linear least-square problem, see [15, 16] for details. The implementation guarantees that all poles are stable, i.e. $\mathcal{R}[p_m] < 0$, and are either real or come in complex-conjugate pairs.

The adaptive Antoulas-Anderson (AAA) method [28] employs the barycentric interpolation

$$f(\mathrm{i}\omega) \approx r(\omega) = \frac{n(\omega)}{d(\omega)} = \frac{\sum_{j \in J} \frac{w_j f(\mathrm{i}\omega_j)}{\omega - \omega_j}}{\sum_{j \in J} \frac{w_j}{\omega - \omega_j}}, \tag{1.6}$$

where $J \subseteq \{1, \dots, n\}$ has cardinality m. The rational function in (1.6) is of type $(m-1, m-1)$, which can be seen by multiplying both numerator and denominator by $\prod_{j \in J}(\omega - \omega_j)$. Moreover, $r(\omega_j) = f(\mathrm{i}\omega_j)$ for all $j \in J$. The weights w_j and

[1] Because of a different convention [7] establishes a negative *imaginary* part of the eigenvalues

nodes ω_j, $j \in J$, are determined adaptively in a two-step procedure, based on linear least squares problems and a greedy strategy [28].

Other data-driven approaches, related to rational interpolation and model order reduction are the Loewner framework [1] and the recent contribution [29], which employs the Heaviside representation. A Bayesian rational Polynomial Chaos-type model has been put forth in [37] to capture the effect of uncertain parameters, e.g., on frequency response functions. A complex-valued version of support vector machine regression has been presented in [39], which is restricted to the so-called circular case with a single kernel only. Complex interpolation with a pair of kernels has been addressed in [6, 33] and also from a Gaussian process regression perspective in [5, 17].

Despite recent progress with complex kernel methods, a general framework with a complete mathematical background on the underlying reproducing kernel Hilbert spaces is missing. In this paper, we introduce a new kernel-based interpolation method which is well adapted to frequency responses. We will put special emphasis on the complex-valued setting and show that the data are used more efficiently if a dedicated kernel method is constructed and interpolation of the real- and imaginary part individually is avoided. To address problems with a few dominant poles we include a low-order rational basis into the kernel method and present a new model selection scheme. We compare our rational kernel-based interpolation method against both AAA and vector fitting and observe an improved or at least comparable performance for a variety of test cases. Finally, the paper develops the required notions of reproducing spaces and minimum norm interpolation for complex-valued kernel methods in general.

The material is structured in the following way. In Section 2 we introduce the concept of a complex/real kernel Hilbert space and consider the special case of frequency response functions as well as the connections to complex-valued Gaussian process regression. Section 3 introduces our new method, which employs a kernel, a pseudo-kernel and an additional rational basis for capturing dominant poles. Finally, Section 4 reports several examples from PDE-based applications, comparing our method to AAA and vector fitting before conclusions are drawn.

Nota bene: A method sharing some similarities with the one proposed in Section 3 has been published recently in the automatic control literature [17]. We became aware of it at very late stage in the writing of the present article. After introducing our new method in Section 3, we discuss similarities and differences in Remark 3.3.

2. Complex/Real RKHS interpolation.

In order to address kernel-based interpolation of the frequency response function, we start by recalling basic facts on reproducing kernel Hilbert spaces (RKHSs); see, e.g., [30] for a comprehensive introduction to this topic.

DEFINITION 2.1 (Complex RKHS). *A complex RKHS H over a non-empty set $\mathbb{S}$ is a complex Hilbert space of functions $\mathbb{S} \to \mathbb{C}$ such that, for all $s \in \mathbb{S}$, the evaluation functional $\delta_s : H \to \mathbb{C}$, $f \mapsto f(s)$, is continuous.*

The Riesz representation theorem implies that there exists a unique function $k : \mathbb{S} \times \mathbb{S} \to \mathbb{C}$, called the reproducing kernel of H, such that $k(\cdot, s) \in H$ and

$$f(s) = \delta_s(f) = \langle f, k(\cdot, s) \rangle_H \tag{2.1}$$

for all $s \in \mathbb{S}$ and $f \in H$, where $\langle \cdot, \cdot \rangle_H$ denotes the Hermitian inner product of H. Equation (2.1) is called the reproduction property, and it is easily seen that the

kernel k is *Hermitian* (i.e., $k(s,s_0) = k(s_0,s)^*$ for all $s, s_0 \in \mathbb{S}$) and *positive definite*: for all $n \in \mathbb{N}^*$ and all (s_1,α_1), …, $(s_n,\alpha_n) \in \mathbb{S} \times \mathbb{C}$,

$$\sum_{1 \leq i,j \leq n} \alpha_i^* \alpha_j k(s_i,s_j) \;\geq\; 0. \tag{2.2}$$

THEOREM 2.2 (Moore-Aronszajn). *For any positive definite Hermitian kernel $k : \mathbb{S} \times \mathbb{S} \to \mathbb{C}$, there exists a unique complex Hilbert space H of functions on $\mathbb{S}$ such that the reproduction property holds with reproducing kernel k.*

Real RKHSs are defined similarly, replacing $\mathbb{C}$ by $\mathbb{R}$ in Definition 2.1: in this case H is a real Hilbert space, the reproducing kernel is symmetric positive definite, and a suitably modified statement of the Moore-Aronszajn theorem holds as well.

THEOREM 2.3 (Interpolation). *Let H be a real or complex RKHS over $\mathbb{S}$ with kernel $k : \mathbb{S} \times \mathbb{S} \to \mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ depending on the type of RKHS. Let $n \in \mathbb{N}^*$, $s_1, \ldots, s_n \in \mathbb{S}$ and $y_1, \ldots, y_n \in \mathbb{K}$. Then there exists a function $g \in H$ such that $g(s_i) = y_i$ for all $i \in \{1, \ldots, n\}$ if, and only if, the system*

$$\begin{bmatrix} k(s_1,s_1) & \ldots & k(s_1,s_n) \\ \vdots & \ddots & \vdots \\ k(s_n,s_1) & \ldots & k(s_n,s_n) \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{2.3}$$

admits a solution. Furthermore, for any solution of (2.3), $g = \sum_{i=1}^n \gamma_i \, k(\cdot, s_i)$ is the unique interpolant of the data (s_1,y_1), …, (s_n,y_n) with minimal norm in H.

A positive definite kernel is called *strictly* positive definite if the kernel matrix $K_n = (k(s_i,s_j))_{1 \leq i,j \leq n}$ is invertible (equivalently, if (2.2) is strict for all $(\alpha_1, \ldots, \alpha_n) \neq 0$) whenever $s_1, \ldots, s_n$ are distinct points. This ensures that (2.3) has a unique solution.

We will proceed by introducing several complex RKHS and their kernels. For $s \in \mathbb{C}$, let $\Re[s]$ and $\Im[s]$ denote the real and imaginary part, respectively. An important example is the Hardy space $H^2(D)$ on the unit disc, where $D = \{s \in \mathbb{C} : |s| < 1\}$. This space plays a role in the analysis of the stability of discrete dynamical systems, see [2], for instance. Here, in the context of continuous-time dynamical systems, we are more interested in the corresponding Hardy space

$$H^2(\Gamma_\alpha) = \left\{ f \in \mathrm{Hol}(\Gamma_\alpha) : \|f\|_{H^2(\Gamma_\alpha)} = \sup_{x > -\alpha} \left(\int_{-\infty}^{\infty} \left| f(x+\mathrm{i}y)^2 \right| \, \mathrm{d}y \right)^{\frac{1}{2}} < \infty \right\}, \tag{2.4}$$

where $\mathrm{Hol}(\Gamma_\alpha)$ denotes the space of holomorphic functions on Γ_α. Note, that there is a Banach space isometry between the H^2 spaces on disc and half-plane, see [18, Chapter 8] for details.

THEOREM 2.4. *The space $H^2(\Gamma_\alpha)$ is a complex RKHS, with strictly positive definite reproducing kernel k given by*

$$k_\alpha(s,s_0) = \frac{1}{2\pi \, (2\alpha + s + s_0^*)}, \quad s, s_0 \in \Gamma_\alpha. \tag{2.5}$$

A proof is given in Appendix A.1. Following standard terminology in complex analysis (see, e.g., [22]), we will refer to k_α as the *Szegö kernel* for the domain Γ_α. Evaluating (2.5) only on the imaginary axis $s = \mathrm{i}\omega$, the expression simplifies to

$$k_\alpha(\mathrm{i}\omega, \mathrm{i}\omega_0) = \frac{1}{2\pi \, (2\alpha + \mathrm{i}(\omega - \omega_0))}, \quad \omega, \omega_0 \in \Omega. \tag{2.6}$$

We consider the stable spline kernel [34, 23] as another example. This kernel has been proposed in the time domain to model functions with a certain smoothness, which additionally incorporate impulse response stability [34]. The corresponding kernel for the frequency domain transfer function has been obtained in [23] and reads

$$k_\alpha(\mathrm{i}\omega, \mathrm{i}\omega_0) = \frac{1}{2}\frac{1}{3\alpha + \mathrm{i}(\omega - \omega_0)} \times \left(\frac{1}{2\alpha + \mathrm{i}\omega} + \frac{1}{2\alpha - \mathrm{i}\omega_0} - \frac{1}{3(3\alpha + \mathrm{i}\omega)} - \frac{1}{3(3\alpha - \mathrm{i}\omega_0)}\right). \tag{2.7}$$

Other related kernels can be found in the control literature, see [23, 17].

2.1. Complex/real RKHS interpolation. The frequency response function fulfills the symmetry property $f^*(s) = f(s^*)$ for all $s \in \Gamma_\alpha$, since it is the Laplace transform of a real-valued function. We are thus naturally led to cast our interpolation problem not in $H^2(\Gamma_\alpha)$ but in the subset

$$H^2_{\mathrm{sym}}(\Gamma_\alpha) \;=\; \big\{f \in H^2(\Gamma_\alpha) :\; \forall s \in \Gamma_\alpha,\; f^*(s) = f(s^*)\big\}. \tag{2.8}$$

This set of complex-valued functions, however, cannot by endowed with the structure of a complex RKHS. In fact, it is not even a vector space over $\mathbb{C}$: indeed, for any $f \in H^2_{\mathrm{sym}}(\Gamma_\alpha)$ and $s \in \Gamma_\alpha$, we would have $(\mathrm{i}f)^*(s) = -\mathrm{i}f^*(s) = -\mathrm{i}f(s^*)$ and $(\mathrm{i}f)^*(s) = (\mathrm{i}f)(s^*) = \mathrm{i}f(s^*)$, which is a contradiction if $f(s^*) \neq 0$.

Observing that the subset of $H^2(\Gamma_\alpha)$ defined by (2.8) is a real vector space of complex-valued functions, we define in the following a new type of function space, which we call a complex/real RKHS.

DEFINITION 2.5 (Complex/real RKHS). *Let $\mathbb{S}$ denote a non-empty set and let H denote a real Hilbert space of complex-valued functions on $\mathbb{S}$. We say that H is a complex/real RKHS if the evaluation functionals are continuous (i.e., for all $s \in \mathbb{S}$, the function $\delta_s : H \to \mathbb{C}$, $f \mapsto f(s)$, is continuous).*

In the remaining part of this section we will establish general results related to these spaces. Section 2.2 will then present consequences for the RKHS with the symmetry property $f^*(s) = f(s^*)$.

Remark 2.6. Any complex RKHS H (such as $H^2(\Gamma_\alpha)$) can be seen as a complex/real RKHS by forgetting the complex structure, i.e., by considering H as a real vector space, endowed with the real inner product $\langle f, g\rangle \mapsto \Re\left(\langle f, g\rangle_H\right)$. More generally, any real subspace of H (such as $H^2_{\mathrm{sym}}(\Gamma_\alpha)$), endowed with this inner product, is clearly a complex/real RKHS. The converse statement is false, however.

PROPOSITION 2.7. *There exists a complex/real RKHS of dimension two over the reals that is not a real subspace of a complex RKHS.*

The elements of a complex/real RKHS are complex-valued functions over $\mathbb{S}$, but can be conveniently represented as real-valued functions over $\tilde{\mathbb{S}} = \mathbb{S} \times \{\mathrm{R}, \mathrm{I}\}$ through the mapping $\mathcal{A} : \mathbb{C}^{\mathbb{S}} \to \mathbb{R}^{\tilde{\mathbb{S}}}$ defined by

$$(\mathcal{A}f)(s, a) = G_a(f(s)), \tag{2.9}$$

where $G_{\mathrm{R}}(s) = \Re(s)$ and $G_{\mathrm{I}}(s) = \Im(s)$. This mapping defines an isometric isomorphism of real Hilbert spaces between H and the real vector space $\tilde{H} = \mathcal{A}H \subseteq \mathbb{R}^{\tilde{\mathbb{S}}}$, endowed with the image inner product. The image space $\tilde{H}$ is easily seen to be a real

RKHS if and only if H is a complex/real RKHS: this observation will be useful both from a theoretical point of view, to establish properties of complex/real RKHSs, and from a practical point of view (see Section 4).

Remark 2.8. Complex/real RKHSs can also been seen a special case of vector-valued RKHSs [9, 26], through the usual identification of $\mathbb{C}$ with $\mathbb{R}^2$.

The term "functional" is used in a loose sense in Definition 2.5, since H is a real vector space while δ_s is a complex-valued function. Therefore, in contrast with the usual case of complex RKHSs, the continuous functionals δ_s, $s \in \mathbb{S}$, do not belong to the topological dual of H. The real and imaginary evaluation functions however—namely, $\Re \circ \delta_s$ and $\Im \circ \delta_s$—do belong to the topological dual, and can thus be expressed through inner products.

PROPOSITION 2.9. *Let H be a complex/real RKHS on a set $\mathbb{S}$, and set*

$$k_{a a_0}(s, s_0) \;=\; \tilde{k}\left((s, a), (s_0, a_0)\right), \qquad s, s_0 \in \mathbb{S}, \quad a, a_0 \in \{\mathrm{R}, \mathrm{I}\}, \tag{2.10}$$

where $\tilde{k}$ denotes the reproducing kernel of $\tilde{H} = \mathcal{A}H$. Then, for all $s \in \mathbb{S}$, we have

$$\delta_s \;=\; \underbrace{\langle \,\cdot\,, \varphi_{\mathrm{R}}(\cdot, s)\rangle_H}_{\Re \circ \delta_s} + \mathrm{i}\, \underbrace{\langle \,\cdot\,, \varphi_{\mathrm{I}}(\cdot, s)\rangle_H}_{\Im \circ \delta_s}, \tag{2.11}$$

where $\varphi_{\mathrm{R}} = k_{\mathrm{RR}} + \mathrm{i}\, k_{\mathrm{IR}}$ and $\varphi_{\mathrm{I}} = k_{\mathrm{RI}} + \mathrm{i}\, k_{\mathrm{II}}$.

This result associates to each complex/real RKHS a pair $(\varphi_{\mathrm{R}}, \varphi_{\mathrm{I}})$ of kernels $\varphi_a : \mathbb{S} \times \mathbb{S} \to \mathbb{C}$, $a \in \{\mathrm{R}, \mathrm{I}\}$. Characterizing admissible choices for this pair of kernels, in the spirit of Theorem 2.2 for complex RKHSs, is possible but not convenient. Instead, motivated by the connection between complex/real RKHSs and complex Gaussian processes (to be discussed in Section 2.3), and in particular the work of Picinbono [32], we introduce another pair of kernels as follows.

DEFINITION 2.10. *Let H denote a complex/real RKHS and let k_{RR}, k_{II}, k_{RI}, k_{IR}, φ_{R} and φ_{I} be defined as in Proposition 2.9. Then we define the* complex kernel k *of the complex/real RKHS as*

$$k \;=\; (k_{\mathrm{RR}} + k_{\mathrm{II}}) + \mathrm{i}\,(k_{\mathrm{IR}} - k_{\mathrm{RI}}) \;=\; \varphi_{\mathrm{R}} - \mathrm{i}\varphi_{\mathrm{I}}, \tag{2.12}$$

and its pseudo-kernel c *as:*

$$c \;=\; (k_{\mathrm{RR}} - k_{\mathrm{II}}) + \mathrm{i}\,(k_{\mathrm{IR}} + k_{\mathrm{RI}}) \;=\; \varphi_{\mathrm{R}} + \mathrm{i}\varphi_{\mathrm{I}}. \tag{2.13}$$

PROPOSITION 2.11. *The functions of the form $\gamma\, k(\cdot, s_0) + \gamma^*\, c(\cdot, s_0)$, with $\gamma \in \mathbb{C}$ and $s_0 \in \mathbb{S}$, span a dense subset of H.*

Remark 2.12. Proposition 2.11 suggests that the concept of a complex/real RKHS, introduced in this article, provides a rigorous formalization of the idea of a "wide-linear complex-valued RKHS" (WL-RKHS) proposed in [6] (see Definition 3.1).

It can be shown that the complex/real RKHS obtained by forgetting the complex structure of a complex RKHS with reproducing kernel k_0, as described in Remark 2.6, is the complex/real RKHS with complex kernel $k = 2k_0$ and vanishing pseudo-kernel—which, borrowing terminology from the signal processing literature [32], can be called *circular*. The factor 2 in the relation between k and k_0 is the price to pay for the consistency of Definition 2.10 with the concepts of covariance and pseudo-covariance functions for complex Gaussian processes (see Section 2.3). More generally, we have the following characterization of the set of admissible (k, c) pairs.

THEOREM 2.13. *For a given complex/real RKHS H, the kernels k and c introduced in Definition 2.10 satisfy the following:*

i) k is complex-valued, Hermitian and positive definite.
ii) c is complex-valued and symmetric.

Moreover, for all $n \geq 1$ and all $s_1, \dots s_n \in \mathbb{S}$:

iii) $\ker K_n \subseteq \ker C_n^*$ *and,*
iv) if K_n is positive definite, $K_n^* - C_n^* K_n^{-1} C_n$ is positive semi-definite,

where $K_n = (k(s_i, s_j))_{1 \leq i,j \leq n}$ and $C_n = (c(s_i, s_j))_{1 \leq i,j \leq n}$.

Conversely, for any pair of functions $k, c : \mathbb{S} \times \mathbb{S} \to \mathbb{C}$ that satisfies these four properties, there exists a unique complex/real RKHS on $\mathbb{S}$ with complex kernel k and pseudo-kernel c.

THEOREM 2.14 (Interpolation in a complex/real RKHS). *Let H denote a complex/real RKHS over $\mathbb{S}$ with complex kernel k and pseudo-kernel c. Let $n \in \mathbb{N}^*$, $s_1, \dots, s_n \in \mathbb{S}$ and $y_1, \dots, y_n \in \mathbb{C}$. Then there exists a function $g \in H$ such that $g(s_i) = y_i$ for all $i \in \{1, \dots, n\}$ if, and only if, the system*

$$K_n \gamma + C_n \gamma^* = y \tag{2.14}$$

admits a solution $\gamma \in \mathbb{C}^n$, where $K_n = (k(s_i, s_j))_{1 \leq i,j \leq n}$, $C_n = (c(s_i, s_j))_{1 \leq i,j \leq n}$, and $y = (y_1, \dots, y_n)^{\mathrm{T}}$. Furthermore, for any solution of (2.14),

$$g = \sum_{i=1}^{n} \gamma_i \, k(\cdot, s_i) + \sum_{i=1}^{n} \gamma_i^* \, c(\cdot, s_i) \tag{2.15}$$

is the unique interpolant of the data (s_1, y_1), …, (s_n, y_n) with minimal norm in H.

For the usual setting of real or complex RKHSs, strictly positive definite kernels guarantee that the interpolation system (2.3) has a solution for any data $y_1, \dots, y_n$. This remains true for the system (2.14) in the case of a complex/real RKHS if the associated real kernel $\tilde{k}$ is strictly positive definite on $\tilde{\mathbb{S}} = \mathbb{S} \times \{\mathrm{R}, \mathrm{I}\}$.

2.2. Complex/real RKHS with symmetry condition. We now characterize, in full generality, the complex/real RKHSs where a symmetry condition of the form $f^*(s) = f(s^*)$ holds for all $f \in H$ and $s \in \mathbb{S}$. The following theorem provides a necessary and sufficient condition on k for such a space to exist and gives the expression of the corresponding pseudo-kernel. The expression appeared previously in [23, Equations (48)–(49)] for a special type of kernel.

THEOREM 2.15. *Let $\mathbb{S}$ denote a non-empty set, equipped with an involution $s \mapsto s^*$ and $k : \mathbb{S} \times \mathbb{S} \to \mathbb{C}$ denote a Hermitian positive definite kernel on $\mathbb{S}$. Then the following assertions are equivalent:*

i) There exists a complex/real RKHS H on $\mathbb{S}$, with complex kernel k, such that

$$\forall f \in H, \ \forall s \in \mathbb{S}, \quad f^*(s) = f(s^*). \tag{2.16}$$

ii) There exists a complex/real RKHS H on $\mathbb{S}$, with complex kernel k and pseudo-kernel c defined by

$$\forall s, s_0 \in \mathbb{S}, \quad c(s, s_0) = k(s, s_0^*). \tag{2.17}$$

iii) $\forall s, s_0 \in \mathbb{S}$, $k(s, s_0^*) = k(s_0, s^*)$.

If any (and consequently all) of these assertions holds, then the complex/real RKHS H with complex covariance k and pseudo kernel (2.17) is the unique RKHS on $\mathbb{S}$ with complex covariance k such that (2.16) holds. Moreover, denoting by $H_{\mathbb{C}}$ the complex RKHS with kernel k, we have $H_{\mathbb{C}} = H \oplus \mathrm{i}H$, $H = \{f \in H_{\mathbb{C}} \mid (2.16) \text{ holds}\}$ and $\langle f, g\rangle = \Re \langle f, g\rangle_{H_{\mathbb{C}}}$ for all $f, g \in H$.

It follows from this theorem that $H^2_{\mathrm{sym}}(\Gamma_\alpha)$ can be characterized as the complex/real RKHS over Γ_α with complex kernel (2.5) and pseudo-kernel:

$$c_\alpha(s, s_0) = \frac{1}{2\pi(2\alpha + s + s_0)}, \quad s,\, s_0 \in \Gamma_\alpha. \tag{2.18}$$

More generally, Theorem 2.15 shows that the problem of minimum-norm interpolation in a complex RKHS, with a symmetry constraint of the form (2.16), can be solved by considering the equivalent problem of minimal-norm interpolation in the complex/real RKHS with the same complex kernel and the pseudo-kernel given by (2.17). In presence of the symmetry condition, even if the complex kernel k is strictly positive definite, $\tilde{k}$ is not and an additional condition on the data is required to ensure that (2.3) has a solution.

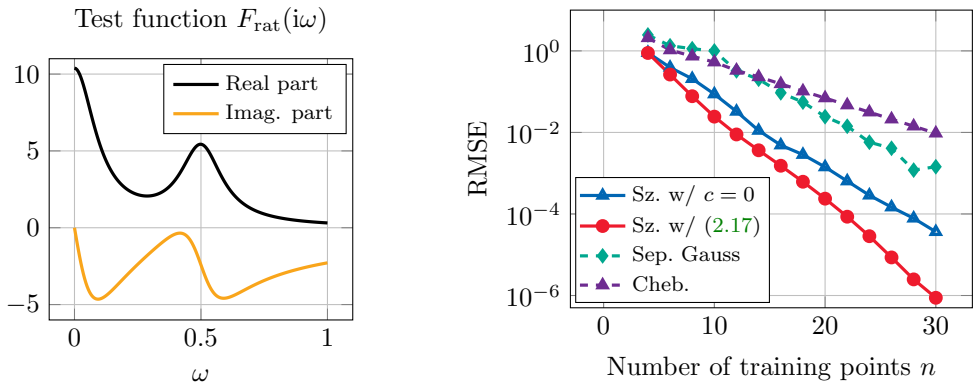
THEOREM 2.16. *In the setting of Theorem 2.15, assume that k is strictly positive definite, c is given by (2.17), and $s_1, \dots, s_n \in \mathbb{S}$ are distinct. Then (2.14) has a solution if, and only if, $y_j = y_i^*$ for all i, j such that $s_j = s_i^*$. When this holds, there is a unique solution such that $\gamma_i = \gamma_j^*$ for all i, j such that $s_j = s_i^*$.*

For illustration, we consider the third order rational function

$$F_{\mathrm{rat}}(\mathrm{i}\omega) = \frac{1}{\mathrm{i}\omega - (-0.1)} + \frac{0.5}{\mathrm{i}\omega - (-0.1 - 0.5\mathrm{i})} + \frac{0.5}{\mathrm{i}\omega - (-0.1 + 0.5\mathrm{i})},\ \omega \in [0, 1], \tag{2.19}$$

which is the Laplace transform of the real-valued function $t \mapsto e^{-0.1t}\big(1 + \cos(0.5t)\big)$ and thus belongs to $H^2_{\mathrm{sym}}(\Gamma_{0.1+\epsilon}) \subseteq H^2(\Gamma_{0.1+\epsilon})$ for all $\epsilon > 0$. To illustrate the importance of the choice of pseudo-kernel, we conduct a convergence study in terms of the root-mean-square error (RMSE) of the approximations, using equidistant training points (details on the implementation and selection of hyper-parameters will be given in the following sections). In Figure 1 we demonstrate that choosing a suitable pseudo-kernel might have a significant impact on the convergence properties of the (complex/real) RKHS interpolation. For the test function (2.19), the pseudo-kernel (2.17) improves the convergence significantly. Note that the test function is a low order rational function which is here only used to illustrate the impact of the pseudo-kernel. Accordingly, rational interpolation techniques as AAA or VF reach machine accuracy already with ≈ 8 training points and are hence excluded in the convergence plot for clarity. However, it can already be observed that complex/real RKHS interpolation with the Szegö kernel outperforms the alternative approach of separate kernel approximations for real and imaginary part with a Gaussian kernel, as well as polynomial interpolation on Chebyshev nodes.

2.3. Relation to Gaussian process interpolation. This section draws connections between minimum norm interpolation in a RKHS and the posterior mean prediction of a Gaussian process (GP), for both the complex and complex/real case. GPs are widely used, but to the authors knowledge this is the first time that the RKHS associated to any complex GP prediction is characterized. Another intention of this section is to make results from the GP literature available for interpolation



(a) Function $F_{\mathrm{rat}}(\mathrm{i}\omega)$ (with real-valued inverse Laplace transform).

(b) Convergence study of the RMSE for different approximations of $F_{\mathrm{rat}}(\mathrm{i}\omega)$.

Fig. 1: Left: Illustrations with the test function $F_{\mathrm{rat}} \in H^2_{\mathrm{sym}}(\Gamma_{0.1+\epsilon}), \epsilon > 0$ defined in (2.19). Right: Convergence of the RMSE as a function of the number of (equidistant) training points. Solid lines: complex/real interpolation with the Szegö kernel for $H^2(\Gamma_\alpha)$, combined with the zero pseudo-kernel (blue) and the pseudo-kernel (2.17) (red). Dashed lines: interpolation with a Gaussian kernel for the real and imaginary part separately (green) and polynomial interpolation on Chebyshev nodes (purple).

with a complex/real RKHS. In particular, we are interested in employing statistical methods for model selection (see, e.g., [31] and references therein)—this will be further developed in Section 3.2. We consider zero-mean processes in this section, for simplicity see Remark 2.19 below.

Complex GPs are covered for instance in [27]. A complex GP is a complex process, where the real and imaginary part considered jointly are a real GP. We consider a zero-mean complex-valued random process ξ on $\mathbb{S}$, with covariance function $\mathfrak{k}$ and pseudo-covariance function $\mathfrak{c}$:

$$\mathsf{E}\,(\xi(s)\xi(s_0)^*) = \mathfrak{k}(s, s_0), \tag{2.20}$$

$$\mathsf{E}\,(\xi(s)\xi(s_0)) = \mathfrak{c}(s, s_0). \tag{2.21}$$

Relying on the mapping $\mathcal{A}$, we can work in a real-valued setting, i.e., with a real-valued GP $\tilde{\xi}$ indexed on $\tilde{\mathbb{S}}$. In the real-valued case, it is well-known that the conditional mean of a GP is identical to the minimum-norm interpolant in the RKHS associated to its covariance function. Hence, using $\mathcal{A}$, the conditional mean of a complex GP ξ is also identical to a minimum-norm interpolant, but this time in a complex/real RKHS, the complex kernel k and pseudo-kernel c of which are equal to $\mathfrak{k}$ and $\mathfrak{c}$ respectively (this follows from Equations (2.12)–(2.13)). It is given by Equation (2.15) in general, which simplifies to

$$\mathsf{E}\,(\xi(s)|y) = \sum_{i=1}^{n} \gamma_i k(s, s_i), \quad \text{with } K_n\gamma = y, \tag{2.22}$$

if the pseudo-covariance is zero (i.e., in the circular case).

Remark 2.17. A common approach to deal with complex data is to use GP interpolation for the real and imaginary part separately (see, e.g., [14]). This corresponds, using notations from Proposition 2.9, to $k_{RI} = k_{IR} = 0$, and therefore to a complex GP with covariance $k = k_{RR} + k_{II}$ and pseudo-covariance $c = k_{RR} - k_{II}$.

Remark 2.18. GP regression with both covariance and pseudo-covariance function has also been considered under the name widely linear posterior mean. In [32] it is first shown that the posterior mean is widely linear [33], which leads to

$$\mathsf{E}\left(\xi(s)|y\right) = (k_{s,n} - c_{s,n}K_n^{-*}C_n^H)P_n^{-*}y + (c_{s,n} - k_{s,n}K_n^{-1}C_n)P_n^{-1}y^*, \tag{2.23}$$

where $P_n = K_n^* - C_n^H K_n^{-1} C_n$ and P_n^{-*} denotes the complex conjugate of the inverse of P_n. The formulas for the circular and non-circular case can also be found in [5].

Remark 2.19. In practice, GP models often include a non-zero mean function m, usually written as a linear combination $m(x) = \sum_{\ell=1}^{L} \beta_\ell h_\ell(x)$ of known basis functions h_ℓ, with unknown coefficients β_ℓ. If the coefficients are estimated by maximum likelihood (as in Section 3), the posterior mean of the GP is then equal to the interpolant with minimal *semi-norm* in $G = V + H$, where $V = \mathrm{span}\{h_1, \dots h_L\}$ and the semi-norm is defined by $|g|_G = \inf_{v \in V} \|g - v\|_H$.

3. Hybrid algorithm.

We focus from now, unless otherwise specified, on functions satisfying the property $f^*(s) = f(s^*)$, and we employ the Szegö kernel (2.5), together with the pseudo-kernel (2.18), for complex/real interpolation. In practice, the convergence of complex/real RKHS interpolation can be significantly slower than that of rational approximations techniques (such as AAA or VF) when the function has a few dominant poles p_i, i.e., poles with small attenuation $\Re[p_i] \approx 0$. In this section, we discuss how complex/real RKHS interpolation with the Szegö kernel and associated pseudo-kernel can be combined with a small number of rational basis functions for the approximation of such frequency response functions.

3.1. Gaussian process model.

We propose to use a complex GP model with rational mean function $m = \sum_{\ell=1}^{L} \beta_\ell h_\ell$ (cf. Remark 2.19), covariance function $\sigma^2 k_\alpha$ and pseudo-covariance function $\sigma^2 c_\alpha$, where k_α denotes the Szegö kernel (2.5), c_α the associated pseudo-kernel (2.18), and σ^2, α, β_1, …, β_L are real parameters with $\sigma^2 > 0$ and $\alpha > 0$. For the mean function m we assume a rational function satisfying the property $m^*(s) = m(s^*)$, of the form

$$m(s) = \sum_{i=1}^{K} \left\{ \frac{1}{s - p_i}\, r_i \,+\, \frac{1}{s - p_i^*}\, r_i^* \right\}, \tag{3.1}$$

with residues $r_1, \dots, r_K \in \mathbb{C}$ and (stable) complex conjugate poles $p_1, p_1^*, \dots, p_K, p_K^* \in \mathbb{C}$ such that $\Re(p_i) < 0$ and $\Im(p_i) > 0$ for all i. This representation is similar to the one used in VF [15, 16]. Equation (3.1) can be rewritten as $m = \sum_{\ell=1}^{L} \beta_\ell h_\ell$ with $L = 2K$,

$$\beta_\ell = \begin{cases} \Re(r_i) & \text{if } \ell = 2i-1, \\ \Im(r_i) & \text{if } \ell = 2i, \end{cases} \quad \text{and} \quad h_\ell(s) = \begin{cases} \frac{1}{s-p_i} + \frac{1}{s-p_i^*} & \text{if } \ell = 2i-1, \\ \frac{\mathrm{i}}{s-p_i} - \frac{\mathrm{i}}{s-p_i^*} & \text{if } \ell = 2i. \end{cases}$$

Note that m is an element of $H^2_{\mathrm{sym}}(\Gamma_{\alpha'})$ with $\alpha' = \min_{1 \le i \le K} |\Re(p_i)| + \epsilon, \epsilon > 0$. For simplicity we only consider complex conjugate poles in (3.1), but real poles could be included as well, as in VF. In the context of the present work, we typically consider a small number K of pole pairs ($K \le K_{\max} = \min(5, \lfloor n/4 \rfloor)$ in the examples).

For a given number K of pole pairs, we select the hyper-parameters σ^2, α, $\mathbf{p} = (p_1, \ldots, p_K)$ and $\mathbf{r} = (r_1, \ldots, r_K)$ by maximization of a penalized log-likelihood function, where the penalty stems from a vague log-normal prior on α; see Supplementary Material for details. An original procedure for the selection of an appropriate number K of pole pairs will be presented in the next section.

Remark 3.1. Note that we do not include a constant basis function, as is usually done in Gaussian process modeling, to ensure that the interpolant satisfies the desired property (namely, goes to zero) when $\omega \to \pm\infty$.

Remark 3.2. Alternatively, the residues $r_1, \ldots, r_K$ could be integrated out analytically using a Gaussian prior, resulting in additional terms in the covariance and pseudo-covariance functions of the GP; see, e.g., [17]. This would allow the uncertainty on the residues, for a given set of poles, to be reflected in the uncertainty quantification (posterior variances) produced by the GP model. We do not pursue this idea further in this article, since our focus is on interpolation rather than uncertainty quantification.

3.2. Adaptive pole selection. Selecting a suitable number K of pole pairs to be included in the mean function (3.1) is a crucial step to ensure good accuracy of the proposed hybrid method. In this section we propose a model selection procedure to select this number automatically, in a data-driven manner. While this procedure relies on the well-established idea of (leave-one-out) cross-validation, it contains an original ingredient in the form an "instability penality", which will be described below.

First we build $K_{\max} + 1$ interpolants $f_n^{(K)}$, where the superscript K indicates the number of pole pairs, ranging from 0 (zero-mean Gaussian process model) to $K_{\max}$. Following standard VF practice [16], we begin with the maximum number of poles, $K = K_{\max}$, using an equidistant distribution of poles close to the frequency axis as a starting point for optimization. The other interpolants are then constructed iteratively, going backwards: at each step optimization is initialized using K of the $K + 1$ poles selected at the previous step, by removing the least relevant pole according to the (penalized) log-likelihood function.

Model selection is then based on leave-one-out (LOO) cross-validation, i.e., on the error indicators

$$\epsilon_{\mathrm{LOO}}^K = \frac{1}{n} \sum_{i=1}^{n} \left| f(\mathrm{i}\omega_i) - \hat{f}_{n-1,i}^{(K)}(\mathrm{i}\omega_i) \right|^2, \quad K = 0, 1, \ldots, K_{\max}, \tag{3.2}$$

where $\hat{f}_{n-1,i}^{(K)}$ denotes a model constructed without the i-th data point. Keeping the poles and kernel hyper-parameters fixed, when removing points, makes it possible to reduce the computational effort, but was found to introduce an undesired preference for models with a larger number of poles. Hence, we employ the LOO criterion with re-tuning, using the poles and hyper-parameters of $f_n^{(K)}$ as an initial guess when constructing $\hat{f}_{n-1,i}^{(K)}$, $1 \leq i \leq n$.

Furthermore, we introduce an additional penalty term, which also takes *global* model variations into account. This approach can be motivated by the example illustrated in Figure 2 (top). The corresponding vibro-acoustic benchmark model will be described in Section 4, however, here we simply consider the approximation of the dashed function, based on interpolation of the training points (black dots), as a general example. At the top, it can be observed that the LOO criterion (3.2) leads to the selection of a model (solid lines) $\hat{f}_n^{(5)}$ which wrongly identifies a pole

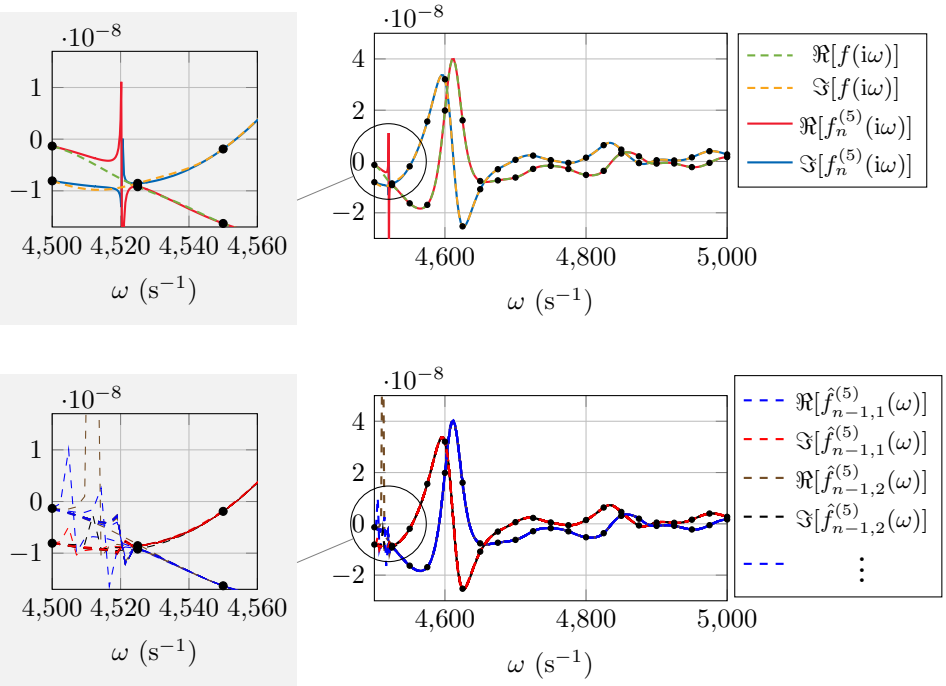


Fig. 2: Top: Dashed lines show the function to approximate. Black dots indicate the training data. Solid lines represent a *bad* approximation model which, however, is selected by the LOO criterion. Zoomed plot (gray background) highlights the influence of a wrongly identified pole. Bottom: Leave-on-out predictions, which show strong local variations between $4500\,\mathrm{s}^{-1}$ and $4520\,\mathrm{s}^{-1}$. However, these variations do not significantly affect the values at the respective training points.

at $\approx 4520\,\mathrm{s}^{-1}$. However, this effect is rather local, it mainly takes place between two training points (illustrated by black dots). At the bottom, we show the models $\hat{f}^{(K)}_{n-1,i}(\omega)$, $i = 1, \dots, n$, which show strong variations close to $\approx 4510\,\mathrm{s}^{-1}$ but rather small errors at the training points ω_i. To take this into account, we introduce an instability penalty term, which leads to the criterion

$$\epsilon^K_{\mathrm{LOO,stab}} = \epsilon^K_{\mathrm{LOO}} + \lambda \frac{1}{n} \frac{1}{M} \sum_{i=1}^{n} \sum_{j=1}^{M} \left| f^{(K)}_n(\mathrm{i}\hat{\omega}_j) - \hat{f}^{(K)}_{n-1,i}(\mathrm{i}\hat{\omega}_j) \right|^2, \tag{3.3}$$

where $\{\hat{\omega}_j\}_{j=1}^M$ denotes a fine grid on Ω (more precisely, an equidistant grid with $M = 10n + 1$ points). The weighting factor λ is chosen as

$$\lambda = 0.2 \, \frac{\epsilon^0_{\mathrm{LOO}}}{\frac{1}{n}\frac{1}{M} \sum_{i=1}^{n} \sum_{j=1}^{M} \left| f^{(0)}_n(\mathrm{i}\hat{\omega}_j) - \hat{f}^{(0)}_{n-1,i}(\mathrm{i}\hat{\omega}_j) \right|^2}, \tag{3.4}$$

i.e., 0.2 after normalizing both terms w.r.t. the respective values of the purely kernel-based interpolation model. To our knowledge, this approach for model selection has not been considered before, although it is related to the continuously-defined LOO

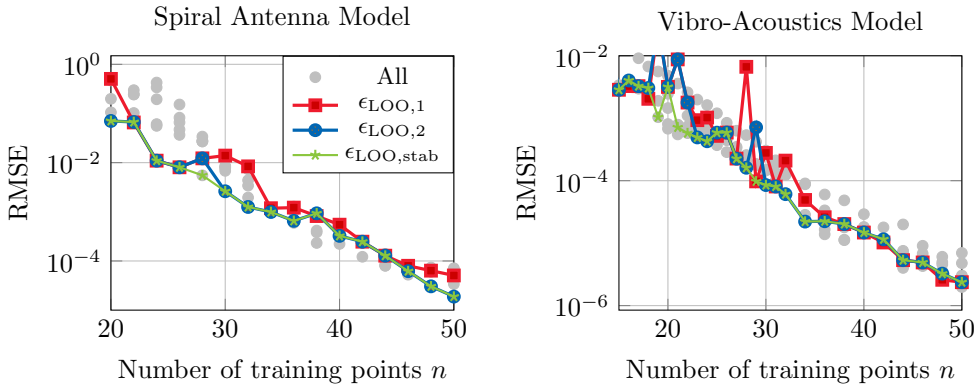


Fig. 3: Comparison of different model selection criteria for two benchmark problems. $\epsilon_{\mathrm{LOO},1}$ and $\epsilon_{\mathrm{LOO},2}$ denote the leave-on-out residual without and with retuning of hyper-parameters, respectively. The stabilized criterion $\epsilon_{\mathrm{LOO,stab}}$ (with retuning) defined in (3.3) gives the best results.

error [20, 21, 13]. The continuously-defined LOO error was employed for sequential sampling, while we propose to use it to construct an instability penalty for model selection. Stability selection [24, 25] is another related approach, which is also based on resampling of the data, but usually employed for variable selection.

Employing the stabilized criterion (3.3) for model selection gives satisfactory results for the benchmark examples considered in this work. For illustration, we consider the convergence studies for two models, which will be described in Section 4. Figure 3 shows the root-mean-square-errors (RMSEs) of the available models with gray dots and the accuracy of the selected models by the different criterions. It can be observed that the stabilized criterion $\epsilon^K_{\mathrm{LOO,stab}}$ gives the best results, while LOO residuals with retuning is superior to the approach without retuning.

Remark 3.3. The combination of kernel methods with a small number of rational basis functions has also been considered in [17] for data-driven modeling of frequency response functions. Therein, the authors employ first order stable spline kernels, which encode stability, causality and smoothness and add a rational basis for capturing the resonsant poles of the transfer function. A prior is formed over the impulse responses linked to the resonant poles, which allows to derive additional kernels (one for each resonant pole) via the Fourier transform.

Our approach proceeds in a similar way, as our VF-inspired rational basis could also be transformed into additional kernels through a prior over β. Differences can be found in the model selection strategies, which are based on the local rational method in [17], whereas our approach is based on statistical model selection. Additionaly, our focus here is on providing a complete background on the RKHS concepts of complex/real interpolation, whereas [17] is additionally targeting uncertainty quantification for the data-driven modeling procedure.

4. Numerical results. We apply the presented approximation techniques to a number of benchmark functions from different fields. We always employ n training points $(\omega_i, f(\mathrm{i}\omega_i))$, where the ω_i are equidistant frequency points in $[\omega_{\min}, \omega_{\max}]$, for simplicity. The accuracy of different approximations is then quantified in terms of

the root-mean-square error (RMSE), which is evaluated on a refined equidistant grid with 201 points for all numerical examples.

In the following we give a few details on the implementation. For AAA [28], we rely on the implementation of the `chebfun` toolbox [12]. For VF, we employ the `VectFit3` toolbox [16, 15, 11], where we use complex equidistant starting poles distributed according to the general recommendation, and always run 30 iterations. We apply the "relaxed non-triviality constraint" [15], include the constant but not the linear term, and enforce stable poles. The number of complex starting pole pairs is set to the maximum number of $2\lfloor\frac{n-1}{2}\rfloor$, which leads to the best results for the smooth test functions considered. For kernel interpolation we consider a separate interpolation of the real and imaginary part with the squared exponential kernel (SE) and complex/real interpolation with the Szegö kernel. The latter is also considered in combination with an adaptive rational basis (Sz.-Rat.) as described in Section 3. The implementation is done in `Matlab` as well, based on the `STK` toolbox [4]. To this end, we employ the mapping $\mathcal{A}$ defined in (2.9) for the complex/real RKHS interpolation, which allows to realize the implementation based on real RKHS interpolation on an augmented input space $\Omega \times \{0, 1\}$. Note that this approach could be employed with any toolbox for real RKHS interpolation that provides the option to specify custom kernel functions. The tuning of the hyper-parameters and poles based on the likelihood function (see Section 3) is carried out using `fmincon` in Matlab, i.e., gradient-based optimization (more precisely an interior point algorithm), which we combine with a multistart procedure; see Supplementary Material for more details.

Remark 4.1. By investigating the shape of the likelihood function for a number of benchmark problems, we have found that the logarithmic reparameterization, discussed in [3] for instance, is not beneficial for the parameter α. Hence, it is only applied to the scaling parameter σ.

4.1. Electric circuit (high order rational function). We consider in the following a parallel connection of N underdamped series RLC circuits, as illustrated on the left side in Figure 4. The admittance is given as

$$Y(s) = \sum_{i=1}^{N} \frac{s}{s^2 L_i + s R_i + C_i^{-1}} = \sum_{i=1}^{N} \frac{c_i}{s - a_i} + \frac{c_i^*}{s - a_i^*}, \tag{4.1}$$

where $\Re[a_i] = -\frac{R_i}{2L_i}$ (an explicit representation of the poles a_i and residues c_i is given in the Supplementary Material) and we consider the frequency range $[10\,\mathrm{kHz}, 25\,\mathrm{kHz}]$. First, we assume $N_1 = 1000$ random series RLC elements, where $C_i \sim \mathcal{U}(1, 20)\,\mu\mathrm{F}$ and $L_i \sim \mathcal{U}(0.1, 2)\,\mathrm{mH}$, and we assume the resistance R_i to be roughly proportional to the inductance, with random variations of $\pm 20\%$: $R_i = L_i(1 + \Delta)\,\Omega(\mathrm{mH})^{-1}$, where $\Delta \sim \mathcal{U}(-0.2, 0.2)$.

Note that for any combination of those parameters, the corresponding series RLC circuits are underdamped. For one particular realization, the distribution of the $2N = 2000$ poles is illustrated in Figure 4. The corresponding admittance $Y_1(\mathrm{i}\omega)$ is shown in Figure 5 with dashed black lines. We then conduct a convergence study for the particular realization of the electric circuit, which is shown in Figure 6 (top, left). We repeat the convergence study for 100 random realizations and depict the median RMSE at each point in Figure 6 (top, right). It can be observed that for the considered range of the number of training points (where $n \leq 60 \ll N$) the complex/real Szegö kernel-based interpolation outperforms AAA and VF. Employing the hybrid algorithm (Sz.-Rat.) does not yield an improvement, but leads to similarly

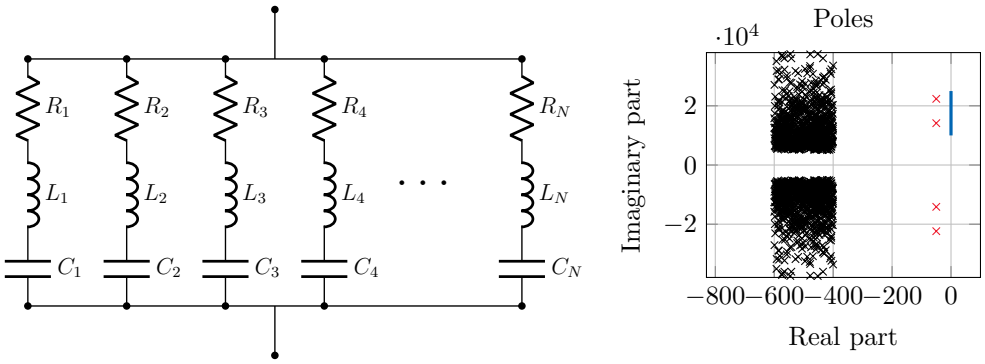


Fig. 4: Left: Parallel connection of (underdamped) series RLC circuits. Right: Black crosses indicate the distribution of $2N_1 = 2000$ poles of the circuit admittance Y_1 in the complex plane. Red crosses indicate the two additional poles considered for the circuit admittance Y_2 with $2N_2 = 2004$ poles. Blue line indicates the considered frequency range.

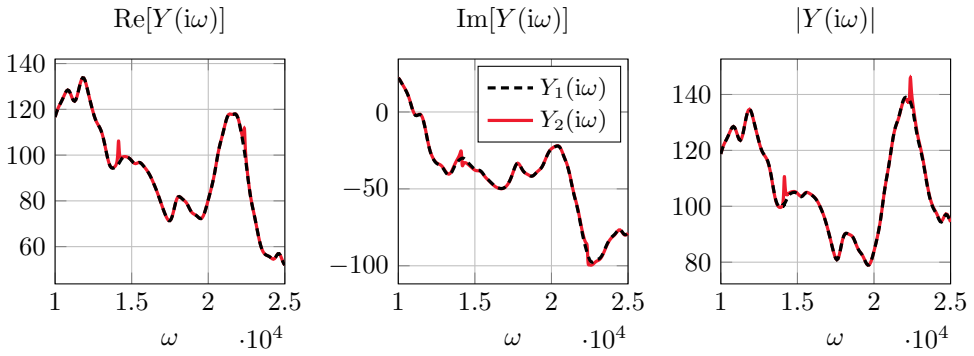


Fig. 5: Complex admittances Y_1 and Y_2 of the electric circuits versus frequency for a particular random parameter realization and $N_1 = 1000$ and $N_2 = 1002$, respectively.

good results.

In our second experiment, we introduce two additional circuit elements with a very small damping, i.e. we now consider $N_2 = 1002$ and

$$C_{1001} = 5\,\mathrm{pF}, \qquad L_{1001} = 1\,\mathrm{mH}, \qquad R_{1001} = 0.1\,\Omega,$$
$$C_{1002} = 2\,\mathrm{pF}, \qquad L_{1002} = 1\,\mathrm{mH}, \qquad R_{1002} = 0.1\,\Omega.$$

This leads to two additional poles which are closer to the input domain, as illustrated by the red crosses in Figure 4. The corresponding admittance $Y_2(\mathrm{i}\omega)$ differs very little from $Y_1(\mathrm{i}\omega)$, except for two sharp peaks, as can be seen in Figure 5. However, the accuracy of the respective RKHS interpolation is significantly affected. In particular, at the bottom of Figure 6, it can be observed that the convergence order of Szegö kernel interpolation is significantly reduced. By adding the rational basis we are able to mitigate the impact of the two dominant poles: it exhibits fast convergence and an improvement w.r.t. AAA and VF can again be observed.

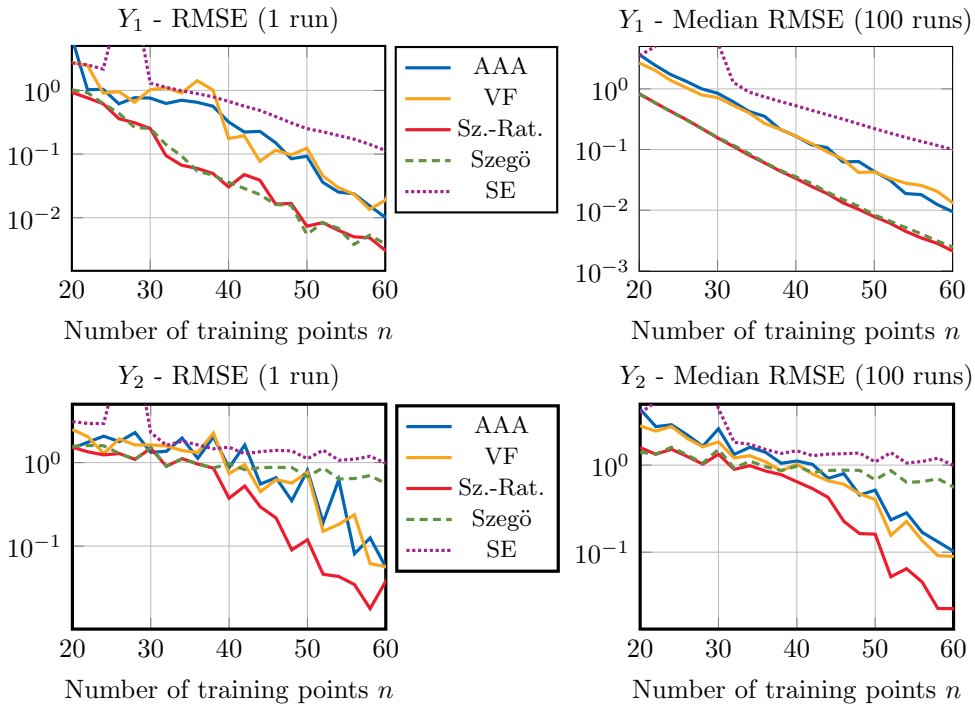


Fig. 6: Convergence study for admittances Y_1 (top) and Y_2 (bottom). Left: RMSE for one particular realization. Right: Median for 100 random realizations.

4.2. PDE-based examples. In the following, we investigate a number of PDE-based examples. We start with the acoustic Helmholtz equation, in particular, the PAC-MAN benchmark example, introduced in [40] which is also included in the platform for benchmark cases in computational acoustics from the European Acoustics Association [19]. The model, shown in Figure 7, has the PAC-MAN shape with an opening angle of 30° and radius of 1 m. As in [40, Section 6.1], we consider as excitation a vibration of the surface of the PAC-MAN with cylindrical modes and observe the radiated field p_i at a point in 2 m distance at an angle of 10°. As in [19], the computation was done based on the implementation of the analytical solution provided in [40] by replacing the python module `scipy` by `mpmath` for the computation of higher order Bessel functions. In particular, we set the truncation order to 300. The complex acoustic pressure field phasor p_i of the total sound-field versus the frequency $f \in [2000\,\text{Hz}, 4000\,\text{Hz}]$ is shown in Figure 7 (top, right). We then conduct a convergence study w.r.t. the number of training points, which is depicted in Figure 7 (bottom, left). It can be observed that the complex/real Szegö kernel-based interpolation outperforms the alternative approaches in the range up to about 40 training points. Adding the rational mean function does not further improve the accuracy, but does not harm the accuracy either.

Next, we consider an electromagnetic model problem, which is a demonstration example of CST Microwave Studio [10], solving the full set of Maxwell equations in the frequency domain. The model consists of a waveguide junction with 4 ports, which contains a small metallic disk and is connected to an external cavity resonator (see

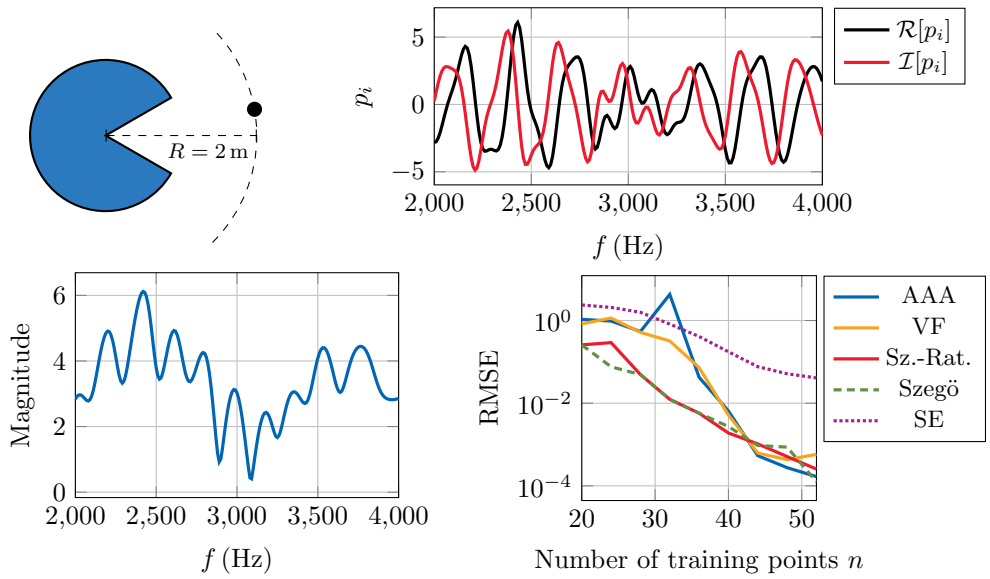


Fig. 7: Top left: We consider a surface vibration of the PAC-MAN model and evaluate the radiated acoustic field p_i at a point (black dot) in 2 m distance to the center. Top right: Complex frequency response function. Bottom left: Magnitude of frequency response function. Bottom right: Convergence study w.r.t. the number of training points.

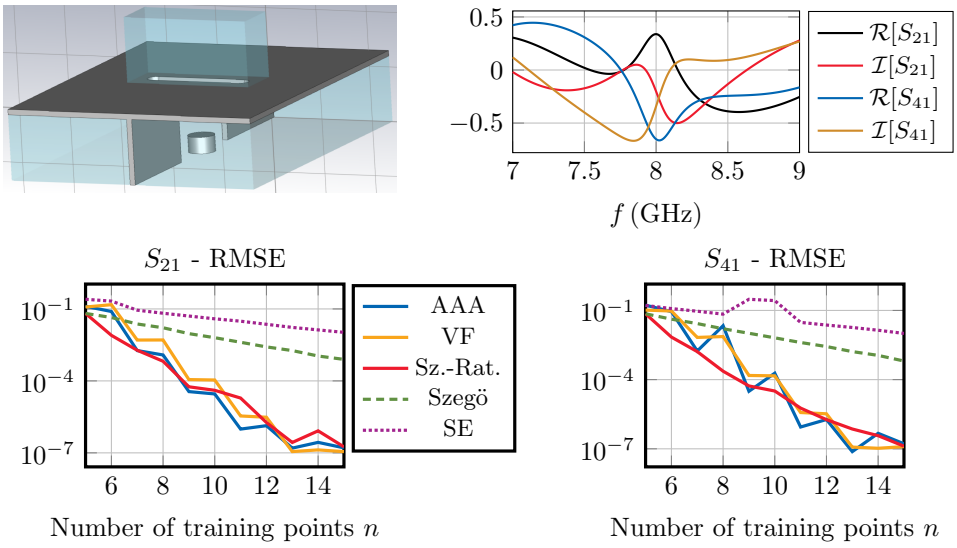


Fig. 8: Top left: Waveguide junction model, taken from CST Microwave Studio [10]. Top right: Complex frequency response functions. Bottom: Convergence studies w.r.t. the number of training points.

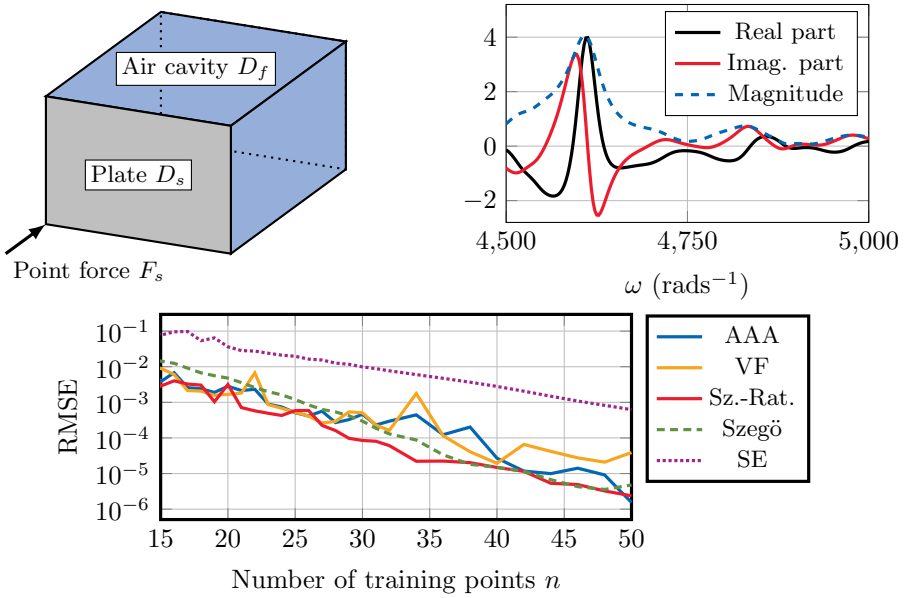


Fig. 9: Top left: Vibro-acoustic benchmark problem, based on [36]. Top right: Complex frequency response function. Bottom: Convergence study w.r.t. the number of training points.

Figure 8). The structure is excited at the first port and simulated using the finite element method in the frequency domain. In particular, we set the solver accuracy of the 3rd order solver to 10^{-6} and use a curved mesh with standard settings. We employ an initial adaptive mesh refinement at 9 GHz, where we set the scattering parameter criterion threshold with 2 subsequent checks to 10^{-4}. As quantity of interest we consider the scattering parameters on a frequency range of $[7\,\text{GHz}, 9\,\text{GHz}]$ using equidistant sample points, where we restrict ourself to S_{21} and S_{41} for brevity, however, the results are qualitatively similar for all four scattering parameters. It can be seen that, the QoIs have a dominant pole at around 8 GHz. This causes the purely kernel-based interpolations to be inferior compared to the rational approximations. However, the proposed combination of kernel-based interpolation and rational approximations leads to satisfactory results, with an accuracy comparable to that of AAA and VF.

The final test case is a vibroacoustic finite element model, taken from [36] and depicted in Figure 9. A 2D Mindlin plate (vibrating structure D_s) is excited by a point force and strongly coupled to a 3D acoustic domain (air cavity D_f). Then, the response at a point in the fluid is evaluated. See [36] for more details on the model. We consider the frequency response on a frequency interval $\omega \in [4500\,\text{s}^{-1}, 5000\,\text{s}^{-1}]$, shown in Figure 9 (top, right). The convergence study, given in Figure 9 (bottom), indicates that the proposed approach usually achieves an accuracy at least comparable to that of AAA and VF, with at certain points an improvement by about an order of magnitude can be observed. It can also be seen that the rational mean function improves the accuracy at the majority of points compared to the pure Szegö kernel-based interpolation.

5. Conclusion. We have presented a comprehensive framework for kernel-based interpolation of complex-valued functions and frequency response functions. In the complex-valued case, the pseudo-kernel is an additional ingredient, which can be used to improve the interpolation accuracy. We have introduced the concept of complex/real reproducing kernel Hibert spaces to reveal the role of the pseudo-kernel and to establish results on minimum norm interpolation. Furthermore, we have proposed a hybrid method, which complements the kernel-interpolant with a low-order rational function and a new model selection criterion: this extension is crucial to account for dominant poles in applications.

The capabilities of the rational-kernel method have been illustrated with several examples, from circuits to frequency response functions originating from PDE problems. In all examples the performance was at least comparable, in some cases improved, compared to AAA and vector fitting on the same set of training data.

The kernel method was further linked to complex-valued Gaussian process regression, which can be used in future work to include noise and adaptive sampling. A generalization to the multivariate case, where, e.g., uncertain parameters are considered as well, and comparisons against multivariate AAA [35] or rational Polynomial Chaos [37], would also be of interest.

Appendix A. Proofs.

A.1. Proof of Theorem 2.4. We assume without loss of generality that $\alpha = 0$ in this proof—i.e., we consider the case of the Hardy space $H^2(\Gamma_0)$ on the right half-plane $\Gamma_0 = \{s \in \mathbb{C} \mid \Re[s] > 0\}$. The general case follows by translation.

The fact that $H^2(\Gamma_0)$ is an RKHS is well known. Indeed, recall the one-sided Paley-Wiener theorem (see, e.g., Chapter 8 of [18]): for all $f \in H^2(\Gamma_0)$, there exists a unique $\widehat{f} \in L^2(\mathbb{R}_+)$ such that

$$f(s) \;=\; \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \widehat{f}(t)\, e^{-st}\, \mathrm{d}t, \qquad \forall s \in \Gamma_0, \tag{A.1}$$

and the mapping $f \mapsto \widehat{f}$ is a surjective isometry: $\|f\|_{H^2(\Gamma_0)} = \|\widehat{f}\|_{L^2(\mathbb{R}_+)}$. This proves that $H^2(\Gamma_0)$ is a Hilbert space, and a simple application of the Cauchy-Schwartz inequality for $s = x + \mathrm{i}y \in \Gamma_0$ yields:

$$|f(s)| \;\leq\; \frac{1}{2\sqrt{\pi x}} \cdot \|\widehat{f}\|_{L^2(\mathbb{R}_+)},$$

which proves that the evaluation functionals are continuous on $H^2(\Gamma_0)$.

Let us now determine the kernel k of this RKHS. Let $s_0 \in \Gamma_0$ and set $h = k(\cdot, s_0)$. Then, for any $f \in H^2(\Gamma_0)$, the reproduction property combined with (A.1) yields:

$$\langle f,\, h \rangle_{H^2(\Gamma_0)} \;=\; f(s_0) \;=\; \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \widehat{f}(t)\, e^{-s_0 t}\, \mathrm{d}t \;=\; \left\langle \widehat{f},\, \frac{1}{\sqrt{2\pi}}\, e^{-s_0^*(\cdot)} \right\rangle_{L^2(\mathbb{R}_+)},$$

which implies that $\widehat{h} = \frac{1}{\sqrt{2\pi}}\, e^{-s_0^*(\cdot)}$ since $f \mapsto \widehat{f}$ is an isometric isomorphism. The expression of the kernel follows:

$$k(s, s_0) \;=\; h(s) \;=\; \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \widehat{h}(t)\, e^{-st}\, \mathrm{d}t \;=\; \frac{1}{2\pi\,(s + s_0^*)}. \tag{A.2}$$

It remains to show that k is strictly positive definite. For any $m \geq 1$ and $s_1, \ldots, s_m \in \Gamma_0$, the kernel matrix $K_m = (k(s_i, s_j))_{1 \leq i,j \leq m}$ can be seen as the conju-

gate Gram matrix of $h_1, \dots, h_m$ in $L^2(\mathbb{R}_+)$, where $h_j(t) = \frac{1}{\sqrt{2\pi}} e^{-s_j^* t}$, $t \geq 0$. Assume that $s_1, \dots, s_m$ are distinct. Then it is well known that the complex exponentials $e^{-s_1^*(\cdot)}, \dots, e^{-s_m^*(\cdot)}$ are linearly independent entire functions on $\mathbb{C}$. It follows, using the identity theorem, that $h_1, \dots, h_m$ are linearly independent as well. The kernel matrix K_m is thus invertible and, consequently, positive definite. Therefore k is strictly positive definite.

Remark A.1. The expression of the reproducing kernel is also derived in [8, Theorem 2.12] (for the upper half-plane instead of Γ_0) using a different approach involving the kernel of the Hardy space of the unit disk. Note, however, that the factor 2π in the denominator of (A.2) is missing in [8, Equation (2.9)]; the discrepancy comes from a missing factor $\frac{1}{2\pi}$ in the definition of the norm on $H^p(\mathbb{D})$ on page 14.

A.2. Proof of Proposition 2.7. Take $H = \{\alpha f_0, \alpha \in \mathbb{C}\}$, where $f_0 : \mathbb{X} \to \mathbb{C}$ is some fixed function, and define a real inner product over H by $\langle \alpha f_0, \beta f_0 \rangle := \Re\alpha \cdot \Re\beta + 4\,\Im\alpha \cdot \Im\beta$. Assuming that $f_0 \not\equiv 0$, the resulting space is complex/real RKHS of dimension two, spanned by $\{f_0, \mathrm{i}f_0\}$. (H is also a complex vector space of dimension 1.)

It not possible to embed H as a subspace of a complex Hilbert space $H_{\mathbb{C}}$ with inner product $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ such that $\langle f, g \rangle = \Re \langle f, g \rangle_{\mathbb{C}}$ for all $f, g \in H$. To see it, note for instance that $\|f_0\| = 1$ while $\|\mathrm{i}f_0\| = 2$.

A.3. Proof of Proposition 2.9. Let $f \in H$, $s_0 \in \mathbb{S}$ and $a_0 \in \{\mathrm{R}, \mathrm{I}\}$. Then

$$G_{a_0}(f(s_0)) = (\mathcal{A}f)(s_0, a_0) = \left\langle \mathcal{A}f,\, \tilde{k}(\cdot, (s_0, a_0)) \right\rangle_{\tilde{H}} \tag{A.3}$$

$$= \left\langle f,\, \mathcal{A}^{-1}\left(\tilde{k}(\cdot, (s_0, a_0))\right) \right\rangle_{H}. \tag{A.4}$$

Taking $a_0 = \mathrm{R}$, we have thus proved that $\Re \circ \delta_{s_0} = \langle \bullet, \varphi_{\mathrm{R}}(\cdot, s_0) \rangle_H$, where

$$\varphi_{\mathrm{R}}(\cdot, s_0) = \mathcal{A}^{-1}\left(\tilde{k}(\cdot, (s_0, \mathrm{R}))\right) \in H \tag{A.5}$$

can be computed as follows:

$$\Re[\varphi_{\mathrm{R}}(s, s_0)] = (\mathcal{A}[\varphi_{\mathrm{R}}(\cdot, s_0)])(s, \mathrm{R}) = \tilde{k}((s, \mathrm{R}), (s_0, \mathrm{R})) = k_{\mathrm{RR}}(s, s_0), \tag{A.6}$$

$$\Im[\varphi_{\mathrm{R}}(s, s_0)] = (\mathcal{A}[\varphi_{\mathrm{R}}(\cdot, s_0)])(s, \mathrm{I}) = \tilde{k}((s, \mathrm{I}), (s_0, \mathrm{R})) = k_{\mathrm{IR}}(s, s_0). \tag{A.7}$$

The expression of $\varphi_{\mathrm{I}}(\cdot, s_0)$ is derived similarly by taking $a_0 = \mathrm{I}$ in (A.4).

A.4. Proof of Proposition 2.11. In a real or complex RKHS, it is well known that the partial kernel functions $k(\cdot, s_0)$, $s \in \mathbb{S}$, span a dense subset of the Hilbert space. Moreover, recall that the bijection $\mathcal{A}$ defined in Section 2.1 is an isometric isomorphism between H and a real RKHS $\tilde{H}$ on $\tilde{\mathbb{S}} = \mathbb{S} \times \{\mathrm{R}, \mathrm{I}\}$, whose kernel $\tilde{k}$ can be recovered from k and c by inverting (2.9)–(2.10). The claim then follows from the observation that any function on $\tilde{\mathbb{S}}$ of the form

$$\tilde{g} = \sum_{i=1}^{n} \alpha_i \tilde{k}(\cdot, (s_i, \mathrm{R})) + \sum_{i=1}^{n} \beta_i \tilde{k}(\cdot, (s_i, \mathrm{I})),$$

where $\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n \in \mathbb{R}$, corresponds to the image by $\mathcal{A}$ of

$$\begin{aligned} g &= \sum_{i=1}^{n} \alpha_i \, \mathcal{A}^{-1}\left(\tilde{k}\left(\cdot, (s_i, \mathrm{R})\right)\right) + \sum_{i=1}^{n} \beta_i \, \mathcal{A}^{-1}\left(\tilde{k}\left(\cdot, (s_i, \mathrm{I})\right)\right) \\ &= \sum_{i=1}^{n} \alpha_i \, \varphi_{\mathrm{R}}\left(\cdot, s_i\right) + \sum_{i=1}^{n} \beta_i \, \varphi_{\mathrm{I}}\left(\cdot, s_i\right) \\ &= \sum_{i=1}^{n} \gamma_i \, k\left(\cdot, s_i\right) + \sum_{i=1}^{n} \gamma_i^* \, c\left(\cdot, s_i\right), \qquad \text{with } \gamma_i = \frac{1}{2}\left(\alpha_i + \mathrm{i}\beta_i\right). \end{aligned}$$

A.5. Proof of Theorem 2.13. Assume first that k and c are the complex kernel and pseudo-kernel associated to a given complex/real RKHS H. Let $\tilde{\xi}$ denote a zero-mean (e.g., Gaussian) real-valued random process indexed by $\mathbb{S}$ with covariance function equal to the kernel $\tilde{k}$ of the real RKHS $\tilde{H} = \mathcal{A}H$, and set $\xi = \tilde{\xi}(\cdot, \mathrm{R}) + \mathrm{i}\,\tilde{\xi}(\cdot, \mathrm{I})$. Then ξ is a complex-valued random process on $\mathbb{S}$, with covariance function k and pseudo-covariance function c; indeed, for all $s, s_0 \in \mathbb{S}$,

$$\begin{aligned} \mathsf{E}\left(\xi(s)\,\xi(s_0)^*\right) &= \left(\tilde{k}\left((s, \mathrm{R}), (s_0, \mathrm{R})\right) + \tilde{k}\left((s, \mathrm{I}), (s_0, \mathrm{I})\right)\right) \\ &\quad + \mathrm{i}\left(\tilde{k}\left((s, \mathrm{I}), (s_0, \mathrm{R})\right) - \tilde{k}\left((s, \mathrm{R}), (s_0, \mathrm{I})\right)\right) = k(s, s_0), \end{aligned}$$

and similarly $\mathsf{E}\left(\xi(s)\,\xi(s_0)\right) = c(s, s_0)$. It follows readily that k is Hermitian and positive definite, and that c is symmetric, which proves i) and ii).

Pick $s_1, \ldots s_n \in \mathbb{S}$, and set $K_n = \left(k(s_i, s_j)\right)_{1 \le i,j \le n}$ and $C_n = \left(c(s_i, s_j)\right)_{1 \le i,j \le n}$. Then K_n and C_n are respectively the covariance and pseudo-covariance matrix of the random vector $Z = \left(\xi(s_1), \ldots, \xi(s_n)\right)^{\mathrm{T}}$, and thus iv) is precisely the "only if" part the following result, due to [32].

PROPOSITION A.2. *Let $n \in \mathbb{N}^*$. Let K be a complex, Hermitian, positive definite matrix of order n, and let C be a complex, symmetric matrix of the same size. Then there exists a complex random vector Z with covariance matrix K and pseudo-covariance matrix C if, and only if, $K^* - C^{\mathrm{H}} K^{-1} C$ is positive semi-definite.*

It remains to prove iii): let $u \in \ker K_n$. Then $u^{\mathrm{H}} K_n u = \mathsf{E}\left(\left|u^{\mathrm{H}} Z\right|^2\right) = 0$, therefore $u^{\mathrm{H}} Z = 0$ almost surely, and as a consequence:

$$C_n^* u = \mathsf{E}\left(Z Z^{\mathrm{T}}\right)^* u = \mathsf{E}\left(Z^* Z^{\mathrm{H}} u\right) = \mathsf{E}\left(Z^* (u^{\mathrm{H}} Z)^{\mathrm{H}}\right) = 0.$$

This completes the proof of i)–iv).

Conversely, assume now that k and c are two functions from $\mathbb{S} \times \mathbb{S}$ to $\mathbb{C}$, such that i)–iv) hold. Then it is easy to see that there is a unique function $\tilde{k} : \mathbb{S} \times \{\mathrm{R}, \mathrm{I}\} \to \mathbb{R}$ such that (2.12)–(2.13) hold, given by

$$\begin{aligned} k_{\mathrm{RR}}(s, s_0) &= \frac{1}{2} \Re\left(k(s, s_0) + c(s, s_0)\right) \\ k_{\mathrm{II}}(s, s_0) &= \frac{1}{2} \Re\left(k(s, s_0) - c(s, s_0)\right) \\ k_{\mathrm{IR}}(s, s_0) &= \frac{1}{2} \Im\left(k(s, s_0) + c(s, s_0)\right) = k_{\mathrm{RI}}(s_0, s). \end{aligned}$$

It remains to prove that $\tilde{k}$ is positive definite. It is easy to see that this is true if, and only if, the matrices K_n and C_n defined above are the covariance and

pseudo-covariance matrices of a complex random vector Z, for any choice of the points $s_1, \dots, s_n \in \mathbb{S}$. Pick such a set of points, and let r denote the rank of K_n. Assume without loss of generality that

$$K_n \;=\; \begin{pmatrix} K_{11} & K_{12} \\ K_{12}^{\mathrm{H}} & K_{22} \end{pmatrix}, \tag{A.8}$$

with K_{11} a positive definite $r \times r$ matrix. Then $K_{22} = K_{12}^{\mathrm{H}} K_{11}^{-1} K_{12}$ and

$$K_n \;=\; M \begin{pmatrix} K_{11} & 0 \\ 0 & 0 \end{pmatrix} M^{\mathrm{H}}, \qquad \text{where } M \;=\; \begin{pmatrix} \mathrm{I}_r & 0 \\ K_{12}^{\mathrm{H}} K_{11}^{-1} & \mathrm{I}_{n-r} \end{pmatrix} \tag{A.9}$$

Denote by C_{11} the upper-left $r \times r$ block in C_n. Then it follows from iv) that $K_{11}^* - C_{11}^{\mathrm{H}} K_{11}^{-1} C_{11}$ is positive semi-definite, and thus by Proposition A.2 there exists a complex random vector Z_1 of size r with covariance matrix K_{11} and pseudo-covariance matrix C_{11}. It is then clear from (A.9) that K_n is the covariance matrix of

$$Z = M \begin{pmatrix} Z_1 \\ 0 \end{pmatrix}.$$

To complete the proof, it remains to observe that C_n is the pseudo-covariance matrix of Z:

$$C_n \;=\; M \begin{pmatrix} C_{11} & 0 \\ 0 & 0 \end{pmatrix} M^{\mathrm{T}} \;=\; \mathsf{E}\left(Z Z^{\mathrm{T}}\right), \tag{A.10}$$

which follows from the facts that C_n is symmetric and that $\ker K_n \subseteq \ker C_n^*$, respectively by ii) and iii).

A.6. Proof of Theorem 2.14. Using the bijection $\mathcal{A}$ defined in Section 2.1, the interpolation problem on $\mathbb{S}$ with complex-valued data (s_1, y_1), …, (s_n, y_n) can be reformulated as an interpolation problem on $\tilde{\mathbb{S}} = \mathbb{S} \times \{\mathrm{R}, \mathrm{I}\}$ with real-valued data $((s_1, \mathrm{R}), \Re(y_1))$, $((s_1, \mathrm{I}), \Im(y_1))$, …, $((s_n, \mathrm{R}), \Re(y_n))$, $((s_n, \mathrm{I}), \Im(y_n))$. The claim then follows from Theorem 2.3 using, as in the proof of Proposition 2.11, the fact that $\mathcal{A}$ is an isometric isomorphism between H and the real RKHS $\tilde{H} = \mathcal{A}(H)$.

A.7. Proof of Theorem 2.15. *i)* $\Rightarrow$ *ii)*. Let H denote a complex/real RKHS on $\mathbb{S}$ with complex kernel k, such that (2.16) holds. Let c denote the pseudo-covariance of H. Let $s_0 \in \mathbb{S}$. It follows from Proposition 2.11 that

$$f_\gamma \;=\; \gamma\, k(\cdot, s_0) \;+\; \gamma^*\, c(\cdot, s_0)$$

is in H for all $\gamma \in \mathbb{C}$. Using (2.16), we see then that

$$\begin{aligned} f_\gamma(s^*) \;&=\; \gamma\, k(s^*, s_0) \;+\; \gamma^*\, c(s^*, s_0) \\ &=\; \gamma\, c(s, s_0)^* \;+\; \gamma^*\, k(s, s_0)^* \;=\; f_\gamma(s)^* \end{aligned}$$

holds for all $\gamma \in \mathbb{C}$. This yields in particular that $c(s, s_0) = k(s^*, s_0)^* = k(s_0, s^*)$, and the claim follows from the symmetry of c:

$$c(s, s_0) = c(s_0, s) = k(s, s_0^*).$$

Note that we have actually proved a little more than *ii)*: if *i)* holds, then *ii)* holds for the *same* complex/real RKHS H. Since we will now prove that *ii)* $\Rightarrow$ *iii)* $\Rightarrow$ *i)*,

it follows that the complex/real RKHS with complex kernel k and pseudo-kernel c defined by (2.17), if it exists, is the only complex/real RKHS with complex kernel k such that (2.16) holds.

ii) $\Rightarrow$ *iii)*. Let H denote a complex/real RKHS on $\mathbb{S}$ with complex kernel k. Assume that the pseudo-kernel c satisfies (2.17). Then, for all $s, s_0 \in \mathbb{S}$,

$$k(s, s_0^*) \;=\; c(s, s_0) \;=\; c(s_0, s) \;=\; k(s_0, s^*).$$

iii) $\Rightarrow$ *i)*. Let k denote a Hermitian positive definite kernel on $\mathbb{S}$ such that

$$\forall s, s_0 \in \mathbb{S}, \quad k(s, s_0^*) = k(s_0, s^*). \tag{A.11}$$

Let $(H_{\mathbb{C}}, \langle \cdot, \cdot \rangle_{\mathbb{C}})$ denote the complex RKHS with kernel k and let $\langle \cdot, \cdot \rangle_{\mathbb{R}} = \Re \langle \cdot, \cdot \rangle_{\mathbb{C}}$. Then, as observed in Remark 2.6, $(H_{\mathbb{C}}, \langle \cdot, \cdot \rangle_{\mathbb{R}})$ is a complex/real RKHS. The associated real and imaginary evaluation kernels, which we denote by $\varphi_{\mathrm{R}}^{\diamond}$ and $\varphi_{\mathrm{I}}^{\diamond}$ respectively, are easily seen to be given by $\varphi_{\mathrm{R}}^{\diamond} = k$ and $\varphi_{\mathrm{I}}^{\diamond} = \mathrm{i}\, k$, and the complex kernel and pseudo-kernel follow:

$$k^{\diamond} = \varphi_{\mathrm{R}}^{\diamond} - \mathrm{i}\varphi_{\mathrm{I}}^{\diamond} = 2k \quad \text{and} \quad c^{\diamond} = \varphi_{\mathrm{R}}^{\diamond} + \mathrm{i}\varphi_{\mathrm{I}}^{\diamond} = 0.$$

Now let H denote the subset of all the functions $f \in H_{\mathbb{C}}$ that satisfy (2.16): H is clearly a real subspace of $H_{\mathbb{C}}$, and thus $(H, \langle \cdot, \cdot \rangle_{\mathbb{R}})$ is a complex/real RKHS as well. Moreover, for any $f \in H$,

$$\begin{aligned}
\Re f(s) \;&=\; \Re \left\{ \frac{1}{2} \left(f(s) + f(s^*)^* \right) \right\} \\
&=\; \frac{1}{2} \left\{ \langle f,\, \varphi_{\mathrm{R}}^{\diamond}(\cdot, s) \rangle_{\mathbb{R}} + \langle f,\, \varphi_{\mathrm{R}}^{\diamond}(\cdot, s^*) \rangle_{\mathbb{R}} \right\} \\
&=\; \left\langle f,\, \frac{1}{2} \left(\varphi_{\mathrm{R}}^{\diamond}(\cdot, s) + \varphi_{\mathrm{R}}^{\diamond}(\cdot, s^*) \right) \right\rangle_{\mathbb{R}}.
\end{aligned}$$

As a consequence of (A.11), the function $s \mapsto \frac{1}{2} \left(\varphi_{\mathrm{R}}^{\diamond}(\cdot, s) + \varphi_{\mathrm{R}}^{\diamond}(\cdot, s^*) \right)$ in this inner product satisfies

$$\begin{aligned}
&\frac{1}{2} \left(\varphi_{\mathrm{R}}^{\diamond}(s_0, s)^* + \varphi_{\mathrm{R}}^{\diamond}(s_0, s^*)^* \right) \\
&= \frac{1}{2} \left(\varphi_{\mathrm{R}}^{\diamond}(s, s_0) + \varphi_{\mathrm{R}}^{\diamond}(s^*, s_0) \right)
\end{aligned}$$

is an element of H, which proves that the real evaluation functional φ_{R} of $(H, \langle \cdot, \cdot \rangle_{\mathbb{R}})$ is given by

$$\varphi_{\mathrm{R}}(s, s_0) \;=\; \frac{1}{2} \left(\varphi_{\mathrm{R}}^{\diamond}(s, s_0) + \varphi_{\mathrm{R}}^{\diamond}(s, s_0^*) \right) \;=\; \frac{1}{2} \left(k(s, s_0) + k(s, s_0^*) \right).$$

Similarly for the imaginary evaluation functional φ_{I}:

$$\varphi_{\mathrm{I}}(s, s_0) \;=\; \frac{1}{2} \left(\varphi_{\mathrm{I}}^{\diamond}(s, s_0) - \varphi_{\mathrm{I}}^{\diamond}(s, s_0^*) \right) \;=\; \frac{\mathrm{i}}{2} \left(k(s, s_0) - k(s, s_0^*) \right).$$

Therefore $\varphi_{\mathrm{R}} - \mathrm{i}\varphi_{\mathrm{I}} = k$ is the complex kernel of $(H, \langle \cdot, \cdot \rangle_{\mathbb{R}})$, which proves *i)*.

To prove the remaining assertions, assume that *i–iii)* hold. Let G denote the closed linear span of $\{k(\cdot, s_0);\ s_0 \in \mathbb{S}\}$ over $\mathbb{R}$. Then we have $G + \mathrm{i}G = H_{\mathbb{C}}$, and it

follows from (A.11) that $G \subseteq H$. Observing that

$$\mathrm{i}H = \{f \in H_{\mathbb{C}} \mid \forall s \in \mathbb{S},\, f(s^*) = -f(s)^*\},$$

we conclude that $H \cap \mathrm{i}H = \{0\}$, therefore $G = H$ and $H \oplus \mathrm{i}H = H_{\mathbb{C}}$.

A.8. Proof of Theorem 2.16. Observe first that, without loss of generality, we can add m extra data points (s_i, y_i), for some $m \leq n$, in such way that 1) the points $s_i \in \mathbb{S}$ $(1 \leq i \leq n+m)$ are still distinct, and 2) for each i we have $s_j = s_i^*$ and $y_j = y_i^*$ for some j.

Existence. Since k is strictly positive definite, we can find $\alpha_1, \ldots, \alpha_{n+m} \in \mathbb{C}$ such that $h = \sum_{i=1}^{n+m} \alpha_i k(\cdot, s_i)$ interpolates the extended data $(s_1, y_1), \ldots, (s_{n+m}, y_{n+m})$. This function h belongs to $H_{\mathbb{C}}$ but not in general to H. Set $g(s) = \frac{1}{2}\left(h(s) + h(s^*)^*\right)$. Then g clearly satisfies the symmetry condition $(g(s^*) = g(s)^*$ for all $s \in \mathbb{S})$ and still interpolates the extended data $(s_1, y_1), \ldots, (s_{n+m}, y_{n+m})$. Moreover, using iii) from Theorem 2.15, we obtain that

$$g(s) = \frac{1}{2} \sum_{i=1}^{n+m} \left(\alpha_i k(s, s_i) + \alpha_i^* k(s, s_i^*)\right),$$

which shows that $g \in H_{\mathbb{C}}$, and thus $g \in H$. Besides, we easily see using (2.17) that: if $s_i = s_i^*$ then

$$\frac{1}{2}\Big\{\alpha_i k(s, s_i) + \alpha_i^* k(s, s_i^*)\Big\} = \gamma_i k(s, s_i) + \gamma_i^* c(s, s_i) \tag{A.12}$$

with $\gamma_i = \frac{1}{2}\alpha_i$, and if $s_j = s_i^*$ with $i \neq j$ then

$$\begin{aligned} &\frac{1}{2}\Big\{\big(\alpha_i k(s, s_i) + \alpha_i^* k(s, s_i^*)\big) + \big(\alpha_j k(s, s_j) + \alpha_j^* k(s, s_j^*)\big)\Big\} \\ &\quad = \big(\gamma_i k(s, s_i) + \gamma_i^* c(s, s_i)\big) + \big(\gamma_j k(s, s_j) + \gamma_j^* c(s, s_j)\big) \end{aligned} \tag{A.13}$$

with $\gamma_i = \frac{1}{2}\left(\alpha_i + \alpha_j^*\right)$ and $\gamma_j = 0$. It follows that g can be rewritten under the form (2.15), using the fact that $\gamma_j = 0$ in (A.13) to get rid of the m extra terms. Thus $\gamma = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$ solves (2.14), which proves the "existence" part of the theorem.

Uniqueness. Let $g \in H$ denote a function of the form (2.15), where the coefficients γ_i are such that (2.14) holds. Using the property that $c(s, s_i) = k(s, s_i^*)$, any such function can be rewritten as $g = \sum_{i=1}^{n+m} \alpha_i k(\cdot, s_i)$. Moreover, since the s_i's are $n+m$ distinct points in $\mathbb{S}$ and k is strictly positive definite, the coefficients $\alpha_i \in \mathbb{C}$ are uniquely determined by the interpolation conditions: $g(s_i) = y_i$, $1 \leq i \leq n+m$. The first n conditions come directly from (2.14), and the m additional conditions must hold as well by symmetry, since $g \in H$.

For each i such that $s_i = s_i^*$, it is easily seen that $\alpha_i = \gamma_i + \gamma_i^*$ is real, and thus the value of γ_i is uniquely determined by α_i and the additional condition that $\gamma_i = \gamma_i^*$. Similarly, if $s_i = s_j^*$ for some $i, j \leq n$, $i \neq j$, then $\alpha_i = \gamma_i + \gamma_j^*$, $\alpha_j = \gamma_i^* + \gamma_j$, and therefore γ_i, γ_j are uniquely determined by α_i, α_j and the condition $\gamma_i = \gamma_j^*$. Finally, if $s_i = s_j^*$ for some $i \leq n$ and $j > n$, then $\alpha_i = \gamma_i$. We have thus proved that there is a unique $\gamma = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$, with the property that $\gamma_i = \gamma_j^*$ when $s_i = s_j^*$, such that (2.14) holds.

Acknowledgments. We thank Christopher Blech, Harikrishnan Sreekumar and Sabine Langer for suggesting acoustic benchmarks and for providing the implementation of the finite element solver used to compute the vibroacoustics data set.

REFERENCES

[1] A. C. Antoulas, S. Lefteriu, A. C. Ionita, P. Benner, and A. Cohen, *A tutorial introduction to the loewner framework for model reduction*, in Model Reduction and Approximation: Theory and Algorithms, vol. 15, SIAM, 2017, pp. 335–376.

[2] L. Baratchart, M. Cardelli, and M. Olivi, *Identification and rational L^2 approximation: a gradient algorithm*, Automatica, 27 (1991), pp. 413–417.

[3] S. Basak, S. Petit, J. Bect, and E. Vazquez, *Numerical issues in maximum likelihood parameter estimation for Gaussian process regression*, arXiv:2101.09747, (2021).

[4] J. Bect, E. Vazquez, et al., *STK: a Small (Matlab/Octave) Toolbox for Kriging. Release 2.8*, 2023, https://github.com/stk-kriging/stk/.

[5] R. Boloix-Tortosa, J. J. Murillo-Fuentes, F. J. Payán-Somet, and F. Pérez-Cruz, *Complex Gaussian processes for regression*, IEEE Trans. Neural Netw. Learn. Syst., 29 (2018), pp. 5499–5511.

[6] R. Boloix-Tortosa, J. J. Murillo-Fuentes, I. Santos, and F. Pérez-Cruz, *Widely linear complex-valued kernel methods for regression*, IEEE Trans. Signal. Process., 65 (2017), pp. 5240–5248.

[7] F. Bonizzoni, F. Nobile, I. Perugia, and D. Pradovera, *Least-squares Padé approximation of parametric and stochastic Helmholtz maps*, Adv. Comput. Math., 46 (2020), pp. 1–28.

[8] J. Bonyo, *Reproducing kernels for Hardy and Bergman spaces of the upper half plane*, Commun. Adv. Math. Sci., 3 (2020), pp. 13–23.

[9] J. Burbea and P. R. Masani, *Banach and Hilbert spaces of vector-valued functions: their general theory and applications to holomorphy*, Pitman Publishing, 1984.

[10] Dassault Systèmes, *CST STUDIO SUITE 2019*, https://www.cst.com.

[11] D. Deschrijver, M. Mrozowski, T. Dhaene, and D. De Zutter, *Macromodeling of multiport systems using a fast implementation of the vector fitting method*, IEEE Microw. Wireless Compon. Lett., 18 (2008), pp. 383–385.

[12] T. A. Driscoll, N. Hale, and L. N. Trefethen, *Chebfun guide*, 2014.

[13] J. N. Fuhg, A. Fau, and U. Nackenhorst, *State-of-the-art and comparative review of adaptive sampling methods for Kriging*, Arch. Comput. Methods Eng., (2020), pp. 1–59.

[14] M. Fuhrländer and S. Schöps, *A blackbox yield estimation workflow with Gaussian process regression for industrial problems*, J. Math. Ind., 10 (2020), p. 25.

[15] B. Gustavsen, *Improving the pole relocating properties of vector fitting*, IEEE Trans. Power Deliv., 21 (2006), pp. 1587–1592.

[16] B. Gustavsen and A. Semlyen, *Rational approximation of frequency domain responses by vector fitting*, IEEE Trans. Power Deliv., 14 (1999), pp. 1052–1061.

[17] N. Hallemans, R. Pintelon, B. Joukovsky, D. Peumans, and J. Lataire, *FRF estimation using multiple kernel-based regularisation*, Automatica, 136 (2022), p. 110056.

[18] K. Hoffman, *Banach spaces of analytic functions*, Prentice Hall, 1962.

[19] M. Hornikx, M. Kaltenbacher, and S. Marburg, *A platform for benchmark cases in computational acoustics*, Acta Acust. united Ac., 101 (2015), pp. 811–820.

[20] R. Jin, W. Chen, and A. Sudjianto, *On sequential sampling for global metamodeling in engineering design*, in International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 36223, 2002, pp. 539–548.

[21] B. Kim, Y. Lee, and D.-H. Choi, *Construction of the radial basis function based on a sequential sampling approach using cross-validation*, J. Mech. Sci. Technol., 23 (2009), pp. 3357–3365.

[22] S. G. Krantz, *Function theory of several complex variables. Second Edition*, AMS Chelsea Publishing, 2001.

[23] J. Lataire and T. Chen, *Transfer function and transient estimation by Gaussian process regression in the frequency domain*, Automatica, 72 (2016), pp. 217–229.

[24] Z. Liu, D. Lesselier, B. Sudret, and J. Wiart, *Surrogate modeling based on resampled polynomial chaos expansions*, Reliab. Eng. Syst. Saf., 202 (2020), p. 107008.

[25] N. Meinshausen and P. Bühlmann, *Stability selection*, J. R. Stat. Soc., B: Stat. Methodol., 72 (2010), pp. 417–473.

[26] C. A. Micchelli and M. Pontil, *On learning vector-valued functions*, Neural Comput., 17 (2005), pp. 177–204.

[27] K. S. Miller, *Complex Gaussian processes*, SIAM Rev., 11 (1969), pp. 544–567.

[28] Y. Nakatsukasa, O. Sète, and L. N. Trefethen, *The AAA algorithm for rational approximation*, SIAM J. Sci. Comput., 40 (2018), pp. A1494–A1522.

[29] F. Nobile and D. Pradovera, *Non-intrusive double-greedy parametric model reduction by interpolation of frequency-domain rational surrogates*, arXiv:2008.10864, (2020).

[30] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*, vol. 152, Cambridge University Press, 2016.
[31] S. Petit, J. Bect, P. Feliot, and E. Vazquez, *Parameter selection in Gaussian process interpolation: an empirical study of selection criteria.* arXiv:2107.06006v4, 2022.
[32] B. Picinbono, *Second-order complex random vectors and normal distributions*, IEEE Trans. Signal Process, 44 (1996), pp. 2637–2640.
[33] B. Picinbono and P. Chevalier, *Widely linear estimation with complex data*, IEEE Trans. Signal Process., 43 (1995), pp. 2030–2033.
[34] G. Pillonetto and G. De Nicolao, *A new kernel-based approach for linear system identification*, Automatica, 46 (2010), pp. 81–93.
[35] A. C. Rodriguez and S. Gugercin, *The p-AAA algorithm for data driven modeling of parametric dynamical systems*, arXiv:2003.06536, (2020).
[36] U. Römer, M. Bollhöfer, H. Sreekumar, C. Blech, and S. Langer, *An adaptive sparse grid rational arnoldi method for uncertainty quantification of dynamical systems in the frequency domain*, Int. J. Numer. Meth., (2021), pp. 5487–5511.
[37] F. Schneider, I. Papaioannou, and G. Müller, *Sparse bayesian learning for complex-valued rational approximations*, International Journal for Numerical Methods in Engineering, 124 (2023), pp. 1721–1747.
[38] F. Tisseur and K. Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.
[39] F. Treviso, R. Trinchero, and F. G. Canavero, *Multiple delay identification in long interconnects via LS-SVM regression*, IEEE Access, 9 (2021), pp. 39028–39042.
[40] H. Ziegelwanger and P. Reiter, *The PAC-MAN model: Benchmark case for linear acoustics in computational physics*, J. Comput. Phys., 346 (2017), pp. 152–171.

SUPPLEMENTARY MATERIAL

SM1. Introduction. This supplementary material is structured in the following way. We first present additional material on the new method presented in the paper: Section SM2.1 provides details on a non-intrusive implementation, and Section SM2.2 contains specifics on parameter optimization. We then collect material relevant to the examples. In particular, a partial fraction representation of the circuit model in Section SM3.1 and an additional numerical example in Section SM3.2. The algorithm applied to this example, a spiral antenna, yields results which are comparable to the pacman model shown in the main paper. Finally, some theoretical results regarding circular complex/real RKHS are collected in Section SM4.1.

SM2. Implementation details.

SM2.1. Non-intrusive implementation.

SM2.1.1. Zero-mean case. The main idea is to construct an isomorphic real-valued GP $\tilde{g}(\tilde{\mathbf{x}}) \sim \mathrm{GP}(0, \tilde{k})$ on an *augmented input space* $\tilde{\mathbf{x}} \in \left(\mathbb{R}^n \times \{0,1\}\right)$, s.t.,

$$\begin{aligned} \tilde{g}(\begin{bmatrix}\mathbf{x} & 0\end{bmatrix}) &= \mathrm{Re}[g(\mathbf{x})], \\ \tilde{g}(\begin{bmatrix}\mathbf{x} & 1\end{bmatrix}) &= \mathrm{Im}[g(\mathbf{x})]. \end{aligned} \tag{SM2.1}$$

The augmented training data $\left(\tilde{\mathbf{x}}, \tilde{y}\right) \in \left(\mathbb{R}^n \times \{0,1\}\right) \times \mathbb{R}$ is for each observation $\left(\mathbf{x}^{(i)}, y^{(i)}\right) \in \mathbb{R}^n \times \mathbb{C}$ obtained as:

$$\begin{aligned} \tilde{\mathbf{x}}^{(i,1)} &= \begin{bmatrix}\mathbf{x}^{(i)} & 0\end{bmatrix}, \quad \tilde{y}^{(i,1)} = \mathrm{Re}[y^{(i)}], \\ \tilde{\mathbf{x}}^{(i,2)} &= \begin{bmatrix}\mathbf{x}^{(i)} & 1\end{bmatrix}, \quad \tilde{y}^{(i,2)} = \mathrm{Im}[y^{(i)}]. \end{aligned}$$

The *new* covariance function $\tilde{k}$ can be derived by enforcing (SM2.1),

$$\tilde{k}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}') = \tilde{k}([\mathbf{x}\ j], [\mathbf{x}'\ j']) = \begin{cases} \frac{1}{2}\,\mathrm{Re}[k(\mathbf{x},\mathbf{x}') + c(\mathbf{x},\mathbf{x}')] & j = j' = 0 \\ \frac{1}{2}\,\mathrm{Re}[k(\mathbf{x},\mathbf{x}') - c(\mathbf{x},\mathbf{x}')] & j = j' = 1 \\ \frac{1}{2}\,\mathrm{Im}[-k(\mathbf{x},\mathbf{x}') + c(\mathbf{x},\mathbf{x}')] & j = 0, j' = 1 \\ \frac{1}{2}\,\mathrm{Im}[k(\mathbf{x},\mathbf{x}') + c(\mathbf{x},\mathbf{x}')] & j = 1, j' = 0 \end{cases},$$

Note that this approach requires to define the modified covariance function $\tilde{k}$, however, no (other) internal functions of existing GP implementations need to be able to cope with complex numbers, which is why we refer to the implementation as *non-intrusive*.

SM2.1.2. Linear model in the mean function. Consider now the superposition

$$\underline{g}(\mathbf{x}) = g(\mathbf{x}) + \mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{b}$$

of a mean-free (real/)complex Gaussian Process $g(\mathbf{x}) \sim \mathrm{CGP}(0, k, c)$ and a complex linear model, where $h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{C}^m$ denote explicit basis functions and $\mathbf{b} \in \mathbb{C}^m$ the corresponding coefficients. Define the augmented process

$$\underline{\tilde{g}}(\tilde{\mathbf{x}}) = \tilde{g}(\tilde{\mathbf{x}}) + \tilde{\mathbf{h}}(\tilde{\mathbf{x}})^{\mathrm{T}}\tilde{\mathbf{b}},$$

where $\tilde{g}(\tilde{\mathbf{x}}) \sim \mathrm{GP}(0, \tilde{k})$, $\tilde{\mathbf{b}} \in \mathbb{R}^{2m}$, $\tilde{\mathbf{h}}(\tilde{\mathbf{x}}) : \mathbb{R}^n \times \{0,1\} \to \mathbb{R}^{2m}$ and we require, similarly as in the last subsection, that

$$\underline{\tilde{g}}([\mathbf{x}\ 0]) = \mathrm{Re}[\underline{g}(\mathbf{x})],$$
$$\underline{\tilde{g}}([\mathbf{x}\ 1]) = \mathrm{Im}[\underline{g}(\mathbf{x})].$$

Incorporating (SM2.1), we can conclude that

$$\tilde{\mathbf{h}}([\mathbf{x}\ 0])^{\mathrm{T}}\tilde{\mathbf{b}} = \mathrm{Re}[\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{b}] = \mathrm{Re}[\mathbf{h}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{Re}[\mathbf{b}] - \mathrm{Im}[\mathbf{h}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{Im}[\mathbf{b}]$$
$$\tilde{\mathbf{h}}([\mathbf{x}\ 1])^{\mathrm{T}}\tilde{\mathbf{b}} = \mathrm{Im}[\mathbf{h}(\mathbf{x})^{\mathrm{T}}\mathbf{b}] = \mathrm{Im}[\mathbf{h}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{Re}[\mathbf{b}] + \mathrm{Re}[\mathbf{h}(\mathbf{x})]^{\mathrm{T}}\,\mathrm{Im}[\mathbf{b}].$$

needs to be fulfilled. This can be achieved by setting

$$\tilde{\mathbf{h}}(\tilde{\mathbf{x}}) = \tilde{\mathbf{h}}([\mathbf{x}\ j]) = \begin{bmatrix} \begin{cases} +\mathrm{Re}[\mathbf{h}(\mathbf{x})] & j=0 \\ +\mathrm{Im}[\mathbf{h}(\mathbf{x})] & j=1 \end{cases} & \begin{cases} -\mathrm{Im}[\mathbf{h}(\mathbf{x})] & j=0 \\ +\mathrm{Re}[\mathbf{h}(\mathbf{x})] & j=1 \end{cases} \end{bmatrix}^{\mathrm{T}},$$

which leads to the coefficients vector

$$\tilde{\mathbf{b}} = \begin{bmatrix} \mathrm{Re}[\mathbf{b}] \\ \mathrm{Im}[\mathbf{b}] \end{bmatrix}.$$

SM2.2. Details on parameter optimization.

SM2.2.1. Penalized log-likelihood criterion. We select the parameters α and σ^2 of the the Szegö kernel $\sigma^2 k_\alpha$ (cf. Section 3 in the article), together with the dominant poles $\mathbf{p} = (p_1, \ldots, p_K)$ in the case of the hybrid method, by maximizing a penalized log-likelihood criterion:

$$J\left(\alpha, \sigma^2, \mathbf{p}\right) = \max_{\mathbf{r} \in \mathbb{C}^K} \ln\left(p\left(\mathbf{y} | \alpha, \sigma^2, \mathbf{p}, \mathbf{r}\right)\right) + \ln(\rho(\alpha)), \tag{SM2.2}$$

where the first term is the log-likelihood of the model, maximized (profiled) analytically with respect to the residues $\mathbf{r}$ of the rational mean function model m, and the second term is a penalty term, designed to pull α away from 0.

More precisely, we take for $\rho(\alpha)$ the probability density function (pdf) of a log-normal random variable with parameters μ_α (to be specified) and $\sigma_\alpha = 3$; in other words, we use a "vague" prior distribution on α, such that $\log(\alpha)$ is Gaussian with mean μ_α and variance σ_α^2. The parameter μ_α is chosen in such a way that the log-normal density for α has its mode at $|\Omega| = \omega_{\max} - \omega_{\min}$, or equivalently that the prior density for $\alpha / |\Omega|$ has its mode at 1. Using that the mode of the lognormal density is at $e^{\mu_\alpha - \sigma_\alpha^2}$, we deduce that $\mu_\alpha = \sigma_\alpha^2 + \ln|\Omega|$. The resulting pdf

$$\rho(\alpha) = \frac{1}{\alpha \sigma_\alpha \sqrt{2\pi}} \exp\Big(\frac{-(\ln(\alpha) - \mu_\alpha)^2}{2\sigma_\alpha^2}\Big) \tag{SM2.3}$$

is shown in Figure SM1 for $|\Omega| = 1$. It can be seen that the chosen parameters allow for the choice of α within a range of several orders of magnitude.

The penalized log-likelihood criterion (SM2.2) is maximized numerically using bound-constrained gradient-based optimization—more precisely, interior point algorithm available from Matlab's `fmincon` function—with a multistart procedure. Details about the bounds for the search domain and the initial points for the local search are provided in Sections SM2.2.2–SM2.2.3.

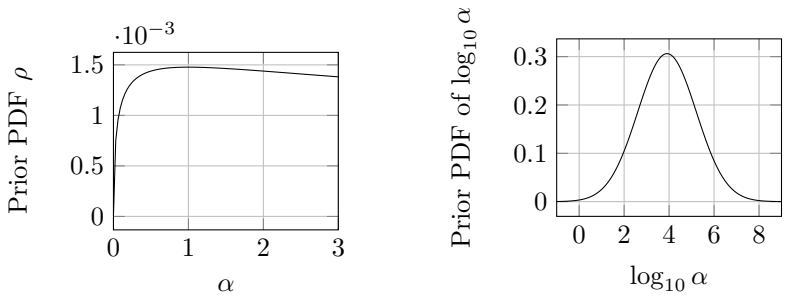


Fig. SM1: Log-normal prior on hyper-parameter α for $|\Omega| = 1$. Left: log-normal prior density of α. Note that the mode of the density is indeed at $\alpha = |\Omega| = 1$. Right: prior density of $\log_{10}(\alpha)$. This is a Gaussian density with mean $\mu_\alpha/\ln(10) \approx 3.91$ and standard deviation $\sigma_\alpha/\ln(10) \approx 1.30$.

Remark SM2.1. This parameter selection procedure can be considered as *maximum a posteriori* estimate in the Bayesian sense. Indeed, the penalized log-likelihood criterion (SM2.2) can be seen as the log-posterior density, up to a constant, assuming a lognormal prior for α and an improper uniform prior for all the other parameters.

Remark SM2.2. Even when the complex kernel k is strictly positive definite, the distribution of data under the GP model does not always admit a probability density function with respect to Lebesgue's measure on $\mathbb{R}^{2n}$ (cf. related discussion regarding the strict positive definiteness of $\tilde{k}$ in Section 2.2 of the article). When this happens, a suitable reference measure has to be used in order to define the likelihood function. For instance, when the pseudo-kernel $c(s, s_0) = k(s, s_0^*)$ is used to enforce the symmetry condition, the value at $\omega = 0$ must be real, which yields a degenerate distribution if the response is evaluated at $\omega = 0$: the solution is simply to remove the imaginary part of the response at this point from the vector of observed variables. See Section 2 of [23] for related considerations.

SM2.2.2. Bounds for the search domain. We optimize with respect to the transformed kernel parameters

$$\begin{aligned} \theta_1 &= \ln\left(\frac{\sigma^2}{2\pi}\right), \\ \theta_2 &= \alpha, \end{aligned}$$

within the optimization bounds

$$\begin{aligned} -15 \;\le\; \theta_1 \;&\le\; 15, \\ 10^{-6}|\Omega| \;\le\; \theta_2 \;&\le\; |\Omega|, \end{aligned}$$

where $|\Omega| = \omega_{\max} - \omega_{\min} = \max\{\omega_i,\, 1 \le i \le n) - \min\{\omega_i,\, 1 \le i \le n)$.

For the hybrid model, the poles are optimized simultaneously with the kernel hyper-parameters, within the bounds

$$\begin{aligned} -|\Omega| \;\le\; \Re(p_i) \;&\le\; -10^{-6}|\Omega|, \\ \max\left\{10^{-6}|\Omega|,\; \omega_{\min} - \frac{|\Omega|}{3}\right\} \;\le\; \Im(p_i) \;&\le\; \omega_{\max} + \frac{|\Omega|}{3}, \end{aligned}$$

where we set the maximum number of poles pairs to $K_{\max} = \min\{5, \lfloor n/4 \rfloor\}$. The bounds are enlarged, if needed, in such a way that all the poles in the starting point of the optimization are contained within them.

SM2.2.3. Starting point(s). We use a multistart procedure to optimize the kernel parameters α and σ^2. More precisely, we start $N_{\text{ms}} = 20$ local optimizations, with the initial value of α uniformly distributed between $10^{-6}|\Omega|$ and $|\Omega|$. For a given value of α, and a given choice of poles in the case of the hybrid algorithm, the GLS (generalized least squares) estimate is used as a starting point for σ^2.

For the poles in the hybrid algorithm, we start from equidistant poles $p_1, \ldots, p_K$ close to the frequency axis:

$$p_k \;=\; -\delta_{\Re}\,|\Omega| \;+\; \mathrm{i}\left(\omega_{\min} + \left(k - \frac{1}{2}\right)\,\delta_{\Im}\right), \qquad 1 \leq k \leq K,$$

where $\delta_{\Re} = 10^{-3}$ (weak attenuation) and $\delta_{\Im} = |\Omega|\,/K_{\max}$. The kernel parameters are initialized a described previously (with the GLS estimate for σ^2 and a multi-start procedure for α).

SM3. Complements for the results section.

SM3.1. Partial fraction representation of RLC circuit. The residues c_i, c_i^* and poles a_i, a_i^* of the partial fraction representation of the electric circuit admittance

$$Y(s) = \sum_{i=1}^{N} \frac{c_i}{s - a_i} + \frac{c_i^*}{s - a_i^*}, \tag{SM3.1}$$

are given as

$$a_i = \frac{-R_i}{2L_i} + \mathrm{i}\sqrt{\frac{1}{L_i C_i} - \left(\frac{R_i}{2L_i}\right)^2}, \tag{SM3.2}$$

$$c_i = \frac{a_i}{L(a_i - a_i^*)} = \frac{\sqrt{\frac{1}{L_i C_i} - \left(\frac{R_i}{2L_i}\right)^2} + \frac{R_i}{2L_i}\mathrm{i}}{2L_i\sqrt{\frac{1}{L_i C_i} - \left(\frac{R_i}{2L_i}\right)^2}}, \tag{SM3.3}$$

where we assumed an underdamped system, i.e.

$$\frac{R_i}{2}\sqrt{\frac{C_i}{L_i}} < 1, \tag{SM3.4}$$

which implies that the argument of the square roots is positive.

SM3.2. Additional numerical example. The model is a spiral antenna, depicted in Figure SM2, where we consider the reflection coefficient S_{11} on a frequency range of $[4\,\text{GHz}, 6\,\text{GHz}]$ as quantity of interest. The data sets are obtained using the boundary element method in CST Microwave Studio [10]. The results are qualitatively the same as for the PAC-MAN model, see Fig. SM2 (bottom).

SM4. Complements for the theoretical section.

SM4.1. Circular complex/real RKHSs.

Definition SM4.1. *We say that a complex/real RKHS is* circular *if it has a vanishing pseudo-kernel.*

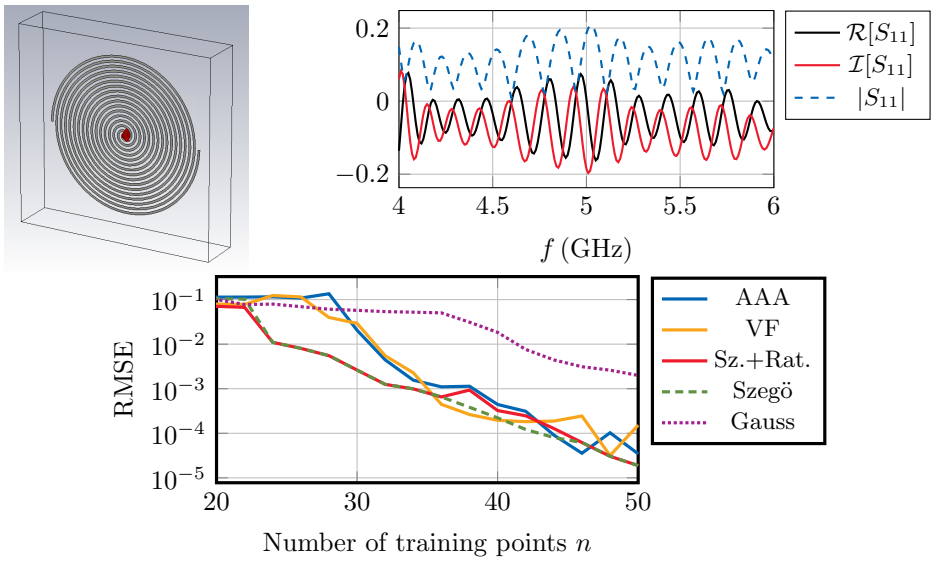


Fig. SM2: Top left: Spiral antenna model, taken from CST Microwave Studio [10]. Top right: Complex frequency response function S_{11}. Bottom: Convergence study w.r.t. the number of training points.

The terms "proper" or "strictly complex" are also sometimes used instead of "circular", in the statistics and signal processing literature, for the case where the pseudo-covariance of a complex Gaussian random vector or function vanishes (see, e.g., [32, 5]).

THEOREM SM4.2. *Let $H_{\mathbb{C}}$ denote a complex RKHS with kernel k_0, and let H denote the complex/real RKHS obtained by considering $H_{\mathbb{C}}$ as a real vector space, endowed with the inner product: $\langle f, g\rangle \mapsto \Re\left(\langle f, g\rangle_{\mathbb{C}}\right)$. Then H is the circular complex/real RKHS with complex kernel $k = 2k_0$.*

Since a complex/real RKHS is uniquely characterized by its (k, c) pair, the converse holds as well: given a circular complex/real RKHS H with complex kernel k, there is a unique complex RKHS $H_{\mathbb{C}}$ (namely, the complex RKHS with kernel $k_0 = \frac{1}{2}k$) such that H is obtained from $H_{\mathbb{C}}$ as in Theorem SM4.2.

Proof. The main idea is already included in the proof of Theorem 2.14 of the article, but we give here a slightly more detailed version. Let φ_{R} and φ_{I} denote the real and imaginary evaluation kernels of H. Then, for all $f \in H$ and $s \in \mathbb{S}$,

$$\langle f,\, k_0(\cdot, s)\rangle \;=\; \Re\left(\langle f,\, k_0(\cdot, s)\rangle\right) \;=\; \Re\left(f(s)\right)$$

and

$$\begin{aligned}\langle f,\, \mathrm{i}k_0(\cdot, s)\rangle \;&=\; \Re\left(\langle f,\, \mathrm{i}k_0(\cdot, s)\rangle\right) \;=\; \Re\left(-\mathrm{i}\,\langle f,\, k_0(\cdot, s)\rangle\right)\\ &=\; \Re\left(-i\; f(s)\right) \;=\; \Im\left(f(s)\right),\end{aligned}$$

which proves that $\varphi_{\mathrm{R}} = k_0$ and $\varphi_{\mathrm{I}} = \mathrm{i}k_0$. The complex kernel and pseudo-kernel of H

are thus given by

$$\begin{aligned} k &= \varphi_{\mathrm{R}} - \mathrm{i}\varphi_{\mathrm{I}} = 2k_0, \\ c &= \varphi_{\mathrm{R}} + \mathrm{i}\varphi_{\mathrm{I}} = 0. \end{aligned}$$ ∎

SM4.2. A relation between the Szegö and rational quadratic kernels. Consider the Szegö kernel on $H^2(\Gamma_\alpha)$:

$$\begin{aligned} k_\alpha(s, s_0) &= \frac{1}{2\pi\,(2\alpha + s + s_0^*)} \\ &= \frac{1}{2\pi}\,\frac{(2\alpha + x + x_0) - \mathrm{i}\,(y - y_0)}{(2\alpha + x + x_0)^2 + (y - y_0)^2}, \end{aligned}$$

where $s = x + \mathrm{i}y$, $s_0 = x_0 + \mathrm{i}y_0 \in \Gamma_\alpha$. In the circular case, the corresponding kernels for the real and imaginary parts are given by:

$$k_{\mathrm{R}}(s, s_0) = k_{\mathrm{I}}(s, s_0) = \frac{1}{4\pi}\,\frac{2\alpha + x + x_0}{(2\alpha + x + x_0)^2 + (y - y_0)^2},$$

For a fixed value of $x = x_0 > -\alpha$, this is of the form

$$(y, y_0) \mapsto \frac{1}{4\pi}\,\frac{A}{A^2 + (y - y_0)^2}, \qquad \text{with } A = 2\alpha + x + x_0 > 0,$$

which is a special case of the so-called *rational quadratic* kernel (see, e.g., [43, 44]), also called *generalized inverse multiquadric* kernel (see, e.g., [42, 41]).

ADDITIONAL REFERENCES

[41] M. Bozzini, M. Rossini, and R. Schaback, *Generalized Whittle–Matérn and polyharmonic kernels*, Advances in Computational Mathematics, 39 (2013), pp. 129–141.
[42] X.-G. Hu, T.-S. Ho, and H. Rabitz, *The collocation method based on a generalized inverse multiquadric basis for bound-state problems*, Computer physics communications, 113 (1998), pp. 168–179.
[43] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
[44] P. Sollich and C. Williams, *Using the equivalent kernel to understand Gaussian process regression*, Advances in Neural Information Processing Systems, 17 (2004).