



HAL
open science

Maximum a Posteriori Estimation for Linear Structural Dynamics Models Using Bayesian Optimization with Rational Polynomial Chaos Expansions

Felix Schneider, Iason Papaioannou, Bruno Sudret, Gerhard Müller

► To cite this version:

Felix Schneider, Iason Papaioannou, Bruno Sudret, Gerhard Müller. Maximum a Posteriori Estimation for Linear Structural Dynamics Models Using Bayesian Optimization with Rational Polynomial Chaos Expansions. 2024. hal-04726096

HAL Id: hal-04726096

<https://hal.science/hal-04726096v1>

Preprint submitted on 8 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MAXIMUM A POSTERIORI ESTIMATION FOR LINEAR STRUCTURAL DYNAMICS MODELS USING BAYESIAN OPTIMIZATION WITH RATIONAL POLYNOMIAL CHAOS EXPANSIONS

F. Schneider, I. Papaioannou, B. Sudret, G. Müller



Data Sheet

Journal:

Report Ref.: RSUQ-2024-005A

Arxiv Ref.: <https://arxiv.org/abs/2408.03569> [stat.ML]

DOI: -

Date submitted: August 7, 2024

Date accepted: -

Maximum a Posteriori Estimation for Linear Structural Dynamics Models Using Bayesian Optimization with Rational Polynomial Chaos Expansions

Felix Schneider^b, Iason Papaioannou^a, Bruno Sudret^c, Gerhard Müller^b

^aEngineering Risk Analysis Group, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

^bChair of Structural Mechanics, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

^cChair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

Abstract

Bayesian analysis enables combining prior knowledge with measurement data to learn model parameters. Commonly, one resorts to computing the maximum a posteriori (MAP) estimate, when only a point estimate of the parameters is of interest. We apply MAP estimation in the context of structural dynamic models, where the system response can be described by the frequency response function. To alleviate high computational demands from repeated expensive model calls, we utilize a rational polynomial chaos expansion (RPCE) surrogate model that expresses the system frequency response as a rational of two polynomials with complex coefficients. We propose an extension to an existing sparse Bayesian learning approach for RPCE based on Laplace's approximation for the posterior distribution of the denominator coefficients. Furthermore, we introduce a Bayesian optimization approach, which allows to adaptively enrich the experimental design throughout the optimization process of MAP estimation. Thereby, we utilize the expected improvement acquisition function as a means to identify sample points in the input space that are possibly associated with large objective function values. The acquisition function is estimated through Monte Carlo sampling based on the posterior distribution of the expansion coefficients identified in the sparse Bayesian learning process. By combining the sparsity-inducing learning procedure with the sequential experimental design, we effectively reduce the number of model evaluations in the MAP estimation problem. We demonstrate the applicability of the presented methods on the parameter updating problem of an algebraic two-degree-of-freedom system and the finite element model of a cross-laminated timber plate.

Keywords: Bayesian Optimization, Rational Polynomial Chaos Expansion, Bayesian Model Updating, Sparse Bayesian Learning, Surrogate Model, Maximum a-posteriori Estimation

1. Introduction

A fundamental task in almost all engineering and applied sciences fields is the assessment of the reliability, serviceability or comfortability of technical systems. This assessment can be done through various approaches, which we broadly categorize into model- and data-driven approaches. Within the field of data-driven approaches, data of a

system's performance is gathered and processed in order to make predictions of a system's performance in the sense outlined above. Data-driven approaches are also commonly referred to as non-parametric. On the other hand, in model-based approaches one tries to develop a fundamental understanding of the underlying natural laws that govern the behavior of the system under consideration. Based on this understanding, a mathematical model can be formulated that allows one to predict the system's performance. Typically, the model is defined by a set of differential equations, whose parameters define the characteristics of the system. In many cases, both of the above, i.e., data and a model, are available. In these cases it is of great interest to combine the two in order to improve the assessment of the system.

This task is often referred to as parametric system identification or model updating [1, 2, 3, 4]. The field of parametric system identification can further be split into probabilistic and non-probabilistic approaches. Non-probabilistic (deterministic) approaches to model updating aim at finding an optimal set of model parameters that minimize the misfit between model prediction and measurement data [5]. Probabilistic approaches to parametric model updating treat the unknown parameters in the model as random variables and express the discrepancy between the measurement data and the model as uncertain as well. Often, this discrepancy is expressed through uncertain measurement and modeling errors, whereby the formulation and assumptions regarding the involved errors are highly problem-specific. A common approach to solving the probabilistic parametric model updating problem is through the application of the Bayesian approach. Within Bayesian updating, the goal is to compute the probabilistic description of the system parameters conditional on the observed measurement data.

Within the Bayesian approach, one states the updating problem through the application of Bayes' rule. The goal is to find the probability distribution of the model parameters conditional on the observed data [6]. This so-called posterior distribution is expressed as the normalized product of the likelihood function and the prior distribution. The likelihood function is commonly derived based on an assumption on the misfit between the model and the observations, while the prior distribution encodes belief about the model parameters before the observations become available. The inclusion of the prior distribution effectively tackles problems of ill-posedness or non-identifiability that are common to non-Bayesian paradigms [7]. A common issue in Bayesian inference is the intractability of the normalizing integral, which requires performing integration over high-dimensional parameter spaces. To circumvent this problem, approximate numerical or semi-analytical approaches have been proposed, including Laplace approximation [8], Markov Chain Monte Carlo (MCMC) sampling [9, 10, 11, 7, 12, 13, 14] or Bayesian Updating with Structural Reliability Methods [15, 16, 17].

Whenever one is only interested in a point-estimate of the posterior distribution, however, the maximum-a-posteriori (MAP) estimate is often considered. The MAP estimate is defined as the mode of the posterior probability density function (PDF). The problem of finding the MAP estimate can thus be stated as an optimization problem.

Nevertheless, finding the optimum with numerical schemes requires evaluating the likelihood, and thus the model function, multiple times. For computationally complex models this renders the solution of the optimization problem an expensive task. In order to lower the computational burden, one can resort to surrogate modeling, whereby the original model is replaced by sufficiently accurate approximations that are fast to evaluate. Popular choices include polynomial chaos expansions (PCE) [18, 19], Neumann series expansions [20] and machine learning techniques such as neural networks [21] or Gaussian process models [22]. The surrogate modeling approach commonly involves an offline and an online phase. During the offline phase, the model is trained by means of a suitable training technique. Here, we consider non-intrusive approaches that rely on discrete evaluations of the model outcome, i.e., the training data, at a set of realizations of the model input parameters, i.e., the experimental design. After the successful training of the surrogate model to the training data, the surrogate model is utilized in the application at hand.

A number of methods have been proposed for the generation of the experimental design. Popular choices include Latin hypercube sampling (LHS) [23] or quasi-random sampling [24]. In their standard settings, these strategies generate a fixed specified number of samples that can be subsequently used in the training procedure. Using a fixed and static experimental design that follows the prior joint distribution of the input parameters can become sub-optimal for inverse problems. This is due to the fact that for inverse problems, the bulk of the posterior probability mass may lie in a region of the input space that is only sparsely represented by a set of samples following the prior distribution. Thereby, the accuracy of the surrogate models in this relevant part of the input space might be poor. To circumvent this problem, active learning or sequential experimental design strategies have been proposed, in which the experimental design is enriched throughout the training procedure.

Subsequently, we consider sequential design strategies in the context of solving optimization problems. Within this field, Bayesian optimization has emerged as a powerful tool for solving global optimization problems for black-box functions. The methodology is based on replacing the objective function in the problem through a random process surrogate model over the space of input parameters. Subsequently, based on an assessment of the predictive uncertainty of the random process model, suitable experimental design points are chosen for which an increase in the objective function value is expected. More specifically, the sample locations are commonly chosen based on maximizing a so called acquisition function. Popular acquisition functions are the probability of improvement (PI), upper confidence bound (UCB), or the expected improvement (EI). These acquisition functions have in common that they are *myopic* in the sense that only the improvement gained through a single next observation is considered, which is considered a one-step lookahead approach [25]. For an extensive review and further literature we refer the reader to [26, 27, 25, 28, 29]. Bayesian optimization has found widespread use since the publication of [30]. Therein, the authors propose to use a Gaussian process model, specifically, ordinary Kriging, to approximate the expensive-to-

evaluate objective function. Their framework, termed Efficient Global Optimization, sequentially chooses sample locations through maximizing the expected improvement acquisition function. In recent years, various variants and modifications, including constrained and multi-objective versions, have been proposed. We refer the reader to [27, 25] and the references therein. A sequential surrogate modeling approach for finite element model updating has been proposed in [31]. Therein, a Gaussian process surrogate model is used to approximate the objective function and the expected improvement acquisition function is employed. The use of non-stationary Gaussian processes in conjunction with the expected information gain criterion has been proposed in the context of Bayesian experimental design [32] for efficient sample allocation. An active learning approach for sequential surrogate modeling in the context of parametric model order reduction is proposed in [33].

Within the scope of this paper, we are particularly interested in Bayesian updating problems that utilize dynamic data and formulate the likelihood in terms of the misfit of the data in the frequency domain as proposed in, e.g., [34, 12, 13, 14, 35]. In such problems, due to the formulation of the likelihood function based on frequency domain data, the log-likelihood function will show a rational dependency on the model input parameters. In initial investigations, it was found that standard Bayesian optimization techniques utilizing Gaussian process regression with commonly used kernel functions do not lead to satisfactory MAP estimates. Recently, a kernel-based interpolation for complex-valued functions in the context of approximating frequency response functions has been proposed in [36]. While this approach enables efficient Gaussian process regression by utilizing specifically suitable kernels for frequency responses that could be utilized in the Bayesian optimization context, it is currently restricted to one-dimensional input spaces. The extension to multidimensional inputs still remains an open research question. Furthermore, the authors in [37] propose a multi-element PCE approach for Bayesian inversion to tackle locally complex dependencies of the likelihood function on the input parameters. Therein, locally refined spectral expansions are used to surrogate the likelihood function.

This paper proposes a sequential design-based MAP estimation procedure for parameter identification that extends the methodologies in [34] and [38]. The likelihood function is derived on the basis of an assumption on the logarithm of the complex-valued model error between the measured and simulated frequency response of the system as proposed in [34]. The model misfit is considered for a set of spatial and frequency domain observation points. To improve computational efficiency, we replace the original models that enter the likelihood function by a set of rational polynomial chaos expansion (RPCE) models that are specifically suitable for approximating frequency response functions. The RPCE expresses the model output as a ratio of two PCEs with complex-valued coefficients and is chosen due to its improved accuracy over standard polynomial chaos expansion (PCE) models in the context of approximating frequency responses [39, 40, 41, 42]. In [34], based on [43], a least-squares regression approach and static experimental design

have been applied to find the coefficients in the RPCE for subsequent use in the Bayesian model updating procedure. However, this still requires a significant number of model evaluations to compute the surrogate model coefficients to avoid poor approximation due to overfitting. Recently, building upon the sparse Bayesian regression methodology in [44], a Bayesian regression approach to learning the coefficients of the RPCE has been proposed in [38]. In this paper, we introduce an alternative Bayesian regression approach to the one in [38] that is based on a linearized error model formulation. Through the linearized formulation, we have access to a closed-form expression of the Hessian of the log-posterior of the denominator coefficients. This allows us to approximate the posterior distribution of the denominator coefficients through a proper complex normal distribution, i.e., a complex variant of Laplace's approximation. The proposed method is able to accurately represent the model output with a higher degree of sparsity compared to the approach in [38]. In addition, the Bayesian paradigm allows us to obtain a measure of uncertainty of the surrogate model prediction. Based on the developed Bayesian RPCE models, we tackle the MAP estimation problem in a Bayesian optimization framework. Under the above derivation, the objective function is expressed as a random process over the space of input parameters that is further parameterized by the random surrogate model coefficients as well as the surrogate model errors. We then sequentially add sample points to the experimental design based on maximizing the expected improvement acquisition function. Since we do not have direct access to the distribution of the random process model, the expected improvement is not available in closed-form and is therefore approximated through Monte Carlo sampling. Two numerical examples are presented that apply the proposed MAP estimation methodology to the parameter updating problem of a two-degree-of-freedom vibratory system and a cross-laminated timber (CLT) plate model.

The outline of the paper is as follows. First, the methodology is introduced in Section 2. In Section 2.1, the Bayesian updating problem, as presented in [34] is comprehensively summarized. Subsequently, Section 2.2 presents the novel sparse Bayesian learning strategy for RPCE surrogate models. Finally, based on the presentation in Sections 2.1 and 2.2, a novel Bayesian optimization strategy for MAP estimation utilizing RPCE models is introduced in Section 2.3. The proposed method is then applied to the parameter updating problem for an algebraic two-degree-of-freedom model as well as a finite element model in 3. In Section 4, open questions and critical issues are discussed. The paper closes with the conclusions in Section 5.

2. Methodology

In this section, we introduce the novel Bayesian model parameter updating methodology. The considered frequency domain dynamic model is comprehensively introduced in Section 2.1.1. Subsequently, the Bayesian approach to model updating is outlined in Section 2.1.2. An modified version of a recently introduced sparse Bayesian learn-

ing algorithm for RPCE is introduced in Section 2.2. This will be the basis for the Bayesian optimization approach introduced in Section 2.3.

2.1. Bayesian model parameter updating

2.1.1. Linear dynamic model with parameter uncertainty

Within the proposed framework, Bayesian updating makes use of measurement data to update the parameters of an engineering model of a given physical system. We assume that the physical system is modeled by a space-discretized, linear dynamic system with n_{dof} degrees of freedom (DOF). Let \mathbf{X} be a random vector with outcome space \mathbb{R}^d and joint PDF $f_{\mathbf{X}}$. \mathbf{X} models a set of uncertain parameters that influence the state of the dynamic system. The matrices $\mathbf{K}(\mathbf{X})$, $\mathbf{C}(\mathbf{X})$ and $\mathbf{M}(\mathbf{X})$ denote the $n_{\text{dof}} \times n_{\text{dof}}$ sized stiffness, damping and mass matrix with parametric uncertainty. The equation of motion describing the system state in the frequency domain is given as

$$\mathbf{K}(\mathbf{X}) \tilde{\mathbf{u}}(\omega, \mathbf{X}) + i\omega \mathbf{C}(\mathbf{X}) \tilde{\mathbf{u}}(\omega, \mathbf{X}) - \omega^2 \mathbf{M}(\mathbf{X}) \tilde{\mathbf{u}}(\omega, \mathbf{X}) = \tilde{\mathbf{f}}(\omega). \quad (1)$$

Here, $\tilde{\mathbf{f}}(\omega)$ and $\tilde{\mathbf{u}}(\omega, \mathbf{X})$ are the deterministic force and the uncertain displacement vector in the frequency domain and $i = \sqrt{-1}$ denotes the imaginary number.

From the above, it is evident that the outcome space of the solution $\tilde{\mathbf{u}}$ is the N -dimensional complex set $\mathbb{C}^{n_{\text{dof}}}$. The frequency response function (FRF) $\tilde{h}_{ij} : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{C}$, defining the acceleration at DOF i due to a force \tilde{f}_j at DOF j in terms of the circular frequency ω is then found by the ratio

$$\tilde{h}_{ij}(\omega, \mathbf{X}) = \mathbf{e}_i^T \left(\mathbf{K}(\mathbf{X}) + i\omega \mathbf{C}(\mathbf{X}) - \omega^2 \mathbf{M}(\mathbf{X}) \right)^{-1} \mathbf{e}_j, \quad (2)$$

where \mathbf{e}_i is the standard unit vector with unit entry at the i -th position and zero otherwise, and $(\cdot)^T$ denotes the transpose of a vector.

2.1.2. Bayesian model updating utilizing frequency domain data

In this section, we introduce the Bayesian formulation of the model updating problem. We follow the previously proposed approach in [34] and assume that the considered system is equipped with a set of n_s sensors, which measure the system's acceleration. Under controlled force excitation, the measured system excitation and response can be processed to compute the system's frequency response functions. Commonly this includes a discrete Fourier transform, which yields frequency response information on a fine frequency grid. The available data is further reduced to a set of frequencies of interest $\{\omega_{0,j} | j = 1, \dots, n_\omega\}$. We thus obtain a set of frequency response function measurements,

denoted by $\mathcal{D}_O = \{y_{O,i} | i = 1, \dots, n_O\}$, where $n_O = n_\omega n_s$ denotes the overall number of data points. This is illustrated on the left hand side of Fig. 1.

Similarly, the system response at the n_s spatial points is predicted at the n_ω frequency points through the model in Eq. (2). Motivated by the proposed surrogate modeling approach in Section 2.2, which allows the approximation of scalar-valued model responses, we denote by $\{\mathcal{M}_i(\mathbf{X}) | i = 1, \dots, n_O\}$ a set of models that predict the response at the n_O spatial and frequency points. Furthermore, $\mathbf{Y}_M = [Y_{M,1}; \dots; Y_{M,n_O}]$, where $Y_{M,i} = \mathcal{M}_i(\mathbf{X})$, denotes the corresponding output random vector. Each of the models \mathcal{M}_i depends on the uncertain input parameters \mathbf{X} . This is illustrated on the right hand side of Fig. 1.

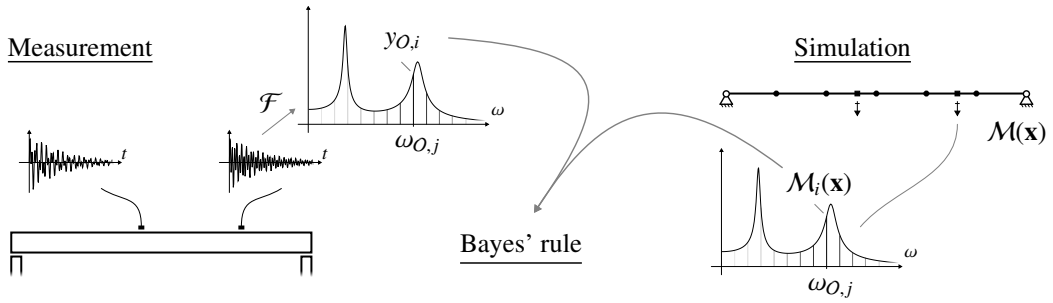


Figure 1: Illustration of data and model fusion through Bayes' rule. Data is gathered on a structure at specific locations. We consider this data in the frequency domain through application of the Fourier transform (denoted by \mathcal{F}) to the time domain data. The system response is also predicted through a simulation model. The information from observations as well as the models can then be combined through Bayes' rule to update the model parameters in \mathbf{X} .

We are now interested in updating our belief about the system parameters. To this end we apply Bayes' rule and obtain the posterior distribution of the system parameters as

$$f(\mathbf{x}|\mathcal{D}_O) = c_E^{-1} L(\mathbf{x}|\mathcal{D}_O) f_{\mathbf{X}}(\mathbf{x}), \quad (3)$$

where $L(\mathbf{x}|\mathcal{D}_O) \propto f(\mathcal{D}_O|\mathbf{x})$ denotes the likelihood function and $f_{\mathbf{X}}(\mathbf{x})$ denotes the prior distribution of the parameters. The normalization constant c_E , commonly referred to as marginal likelihood or evidence, is defined as $c_E = \int L(\mathbf{x}|\mathcal{D}_O) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$ and ensures that the posterior distribution is properly normalized. The likelihood is derived based on the assumption of the following error model,

$$y_{O,i} + \varepsilon_{O,i} = \mathcal{M}_i(\mathbf{x}) \cdot \varepsilon_{M,i}, \quad (4)$$

where $\varepsilon_{O,i}$ and $\varepsilon_{M,i}$ denote the observation noise and the multiplicative model error that jointly explain the misfit between the model output $\mathcal{M}_i(\mathbf{x})$ and the observation $y_{O,i}$. Subsequently, we assume that $\varepsilon_{O,i} \ll \varepsilon_{M,i}$ and hence can be disregarded and denote $\varepsilon_i = \varepsilon_{M,i}$ for notational convenience. We denote by $\boldsymbol{\varepsilon} = [\varepsilon_1; \dots; \varepsilon_{n_O}]$ the vector collecting

all errors; $\boldsymbol{\varepsilon}$ can be expressed as $\boldsymbol{\varepsilon} = |\boldsymbol{\varepsilon}| \cdot e^{i\boldsymbol{\varphi}_\varepsilon}$ with $|\boldsymbol{\varepsilon}|$ and $\boldsymbol{\varphi}_\varepsilon = \angle \boldsymbol{\varepsilon}$ denoting the element-wise absolute value and phase of the complex-valued vector $\boldsymbol{\varepsilon}$. Furthermore, we choose to model the distribution of $|\boldsymbol{\varepsilon}|$ using a multivariate lognormal distribution and the distribution of $\boldsymbol{\varphi}_\varepsilon$ through a multivariate normal distribution and consider them to be independent. Then the element-wise logarithm of the error vector

$$\ln \boldsymbol{\varepsilon} = \underbrace{\ln |\boldsymbol{\varepsilon}|}_{=: \mathbf{w}_\varepsilon} + i\boldsymbol{\varphi}_\varepsilon, \quad (5)$$

follows a complex normal distribution. Taking the logarithm of Eq. (4), and by introducing $\mathbf{y}_O = [y_{O,1}; \dots; y_{O,n_O}]$ and $\mathbf{y}_M(\mathbf{x}) = [\mathcal{M}_1(\mathbf{x}); \dots; \mathcal{M}_{n_O}(\mathbf{x})]$ we obtain

$$\ln \mathbf{y}_O = \ln \mathbf{y}_M(\mathbf{x}) + \ln \boldsymbol{\varepsilon}. \quad (6)$$

Based on Eq. (6), we can write the likelihood as

$$L(\mathbf{x}|\mathcal{D}_O) = f_{\ln \boldsymbol{\varepsilon}}(\ln \mathbf{y}_O - \ln \mathbf{y}_M(\mathbf{x})). \quad (7)$$

We proceed by defining the mean vector and covariance matrix of the logarithm of the error in order to fully define the likelihood function. Firstly, we assume that the logarithm of the absolute value and the phase of the error have zero mean, i.e., $\mathbb{E}[\mathbf{w}_\varepsilon] = \mathbf{0}_{n_O}$ and $\mathbb{E}[\boldsymbol{\varphi}_\varepsilon] = \mathbf{0}_{n_O}$. This will render the median of $|\boldsymbol{\varepsilon}|$ to be $|\boldsymbol{\varepsilon}|_{0.5} = \mathbf{1}_{n_O}$. Here, $\mathbf{0}_n$ and $\mathbf{1}_n$ denote the all zero and all ones vector with n entries. Under the above assumptions the real composite vector $\mathbf{r} = [\mathbf{w}_\varepsilon; \boldsymbol{\varphi}_\varepsilon]$ will follow a zero-mean normal distribution with covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{r}\mathbf{r}} = \mathbb{E}[\mathbf{r}\mathbf{r}^T] = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{w}_\varepsilon\mathbf{w}_\varepsilon} & \mathbf{0}_{n_O \times n_O} \\ \mathbf{0}_{n_O \times n_O} & \boldsymbol{\Sigma}_{\boldsymbol{\varphi}_\varepsilon\boldsymbol{\varphi}_\varepsilon} \end{bmatrix}, \quad (8)$$

where $\boldsymbol{\Sigma}_{\mathbf{w}_\varepsilon\mathbf{w}_\varepsilon} = \mathbb{E}[\mathbf{w}_\varepsilon\mathbf{w}_\varepsilon^T]$, $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}_\varepsilon\boldsymbol{\varphi}_\varepsilon} = \mathbb{E}[\boldsymbol{\varphi}_\varepsilon\boldsymbol{\varphi}_\varepsilon^T]$ and $\mathbf{0}_{n \times n}$ denotes the $n \times n$ matrix of zeros. The choice of a specific model for the above covariance matrices is problem-dependent and will be further discussed in the specific application cases in Section 3. The same holds for the definition of the prior distribution. Having fully specified the likelihood function and joint prior distribution, we proceed to the computation of the posterior distribution. Except for few cases, e.g., in the case of conjugate priors, a closed-form solution for the posterior distribution is not available. Full inference on the model parameters is then often done through sampling-based approaches. If the engineering analysis does not require computation of the full posterior distribution, one can resort to computing posterior point descriptors of the model parameters. A commonly chosen descriptor is the mode of the posterior PDF, the maximum a posteriori value.

Maximum a posteriori estimation. The maximum a posteriori (MAP) value \mathbf{x}^* can be found as the solution of the following optimization problem

$$\mathbf{x}^* \in \arg \max_{\mathbf{x} \in \mathbb{R}^d} [\ln L(\mathbf{x}|\mathcal{D}_O) + \ln f(\mathbf{x})] . \quad (9)$$

Under consideration of the derivation in the preceding section, the objective function $h(\mathbf{x}) \propto \ln L(\mathbf{x}|\mathcal{D}_O) + \ln f(\mathbf{x})$ can be written as

$$h(\mathbf{x}) = \frac{1}{2} \left[-\ln \det \Sigma_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon} - \ln \det \Sigma_{\boldsymbol{\varphi}_\varepsilon \boldsymbol{\varphi}_\varepsilon} - \|\mathbf{w}_O - \mathbf{w}_M(\mathbf{x})\|_{\Sigma_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon}^{-1}}^2 - \|\boldsymbol{\varphi}_O - \boldsymbol{\varphi}_M(\mathbf{x})\|_{\Sigma_{\boldsymbol{\varphi}_\varepsilon \boldsymbol{\varphi}_\varepsilon}^{-1}}^2 \right] + \ln f(\mathbf{x}) , \quad (10)$$

where $\|\mathbf{x}\|_A^2 = \mathbf{x}^H \mathbf{A} \mathbf{x}$, $\mathbf{w}_O = [\ln |y_{O,1}|; \dots; \ln |y_{O,n_O}|]$, $\mathbf{w}_M(\mathbf{x}) = [\ln |\mathcal{M}_1(\mathbf{x})|, \dots, \ln |\mathcal{M}_{n_O}(\mathbf{x})|]$, $\boldsymbol{\varphi}_O = [\angle y_{O,1}, \dots, \angle y_{O,n_O}]$ and $\boldsymbol{\varphi}_M(\mathbf{x}) = [\angle \mathcal{M}_1(\mathbf{x}), \dots, \angle \mathcal{M}_{n_O}(\mathbf{x})]$. For i.i.d. proper complex errors with $\Sigma_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon} = \Sigma_{\boldsymbol{\varphi}_\varepsilon \boldsymbol{\varphi}_\varepsilon} = \frac{\beta_O^{-1}}{2} \mathbf{I}_{n_O}$, we obtain

$$h(\mathbf{x}) = n_O \ln \beta_O - \beta_O \sum_{i=1}^{n_O} |\ln y_{O,i} - \ln y_{M,i}|^2 + \ln f(\mathbf{x}) . \quad (11)$$

We observe that the evaluation of the objective function $h(\mathbf{x})$ of Eqs. (10) or (11) requires computing the model prediction of the n_O models. Whenever this requires significant computing resources, solving the optimization problem in Eq. (9) using standard techniques becomes challenging. Therefore, we subsequently propose a Bayesian optimization strategy that utilizes RPCE surrogate models. To this end, we introduce the RPCE model and propose a novel Bayesian learning method for estimating its coefficients in the next section.

2.2. Rational polynomial chaos expansion

The proposed updating strategy aims to utilize surrogate models for the individual model responses that enter the likelihood function of Eq. (7). This likelihood depends on a numerical model of the systems' frequency response, which exhibit a rational type of dependency on the input parameters \mathbf{X} [45]. It was shown in [39, 43] that standard PCE surrogate models converge slowly and suffer from poor generalization when applied to frequency response function data. A natural modification to standard PCE is to consider rational PCE models, which are constructed through taking the ratio of two PCEs. To this end, we approximate the set of models $\{\mathcal{M}_i(\mathbf{X}) | i = 1, \dots, n_O\}$ through a set of surrogate models $\{\mathcal{S}_i(\mathbf{X}) | i = 1, \dots, n_O\}$, for which, ideally, $\mathcal{S}_i(\mathbf{X}) \approx \mathcal{M}_i(\mathbf{X})$.

2.2.1. Definition of the Rational Polynomial Chaos Expansion

As discussed above, assume that \mathbf{X} is a random vector with outcome space \mathbb{R}^d and given joint probability density function that models the uncertain input parameters of the numerical model $Y_{M,i}(\mathbf{X})$. Without loss of generality, we assume that the random vector \mathbf{X} follows the independent standard Gaussian distribution. If \mathbf{X} follows a non-Gaussian

distribution, it is possible to express $Y_{\mathcal{M},i}$ as a function of an underlying independent standard Gaussian vector through an isoprobabilistic transformation [46]. For the sake of simplicity, we drop the dependency on the index i and treat scalar valued models in the following. Vector-valued models, as introduced above, are treated by surrogating each entry in the output random vector individually.

Let $P(\mathbf{X})$ and $Q(\mathbf{X})$ be truncated polynomial chaos representations, such that

$$P(\mathbf{X}; \mathbf{p}) = \sum_{i=0}^{n_p-1} p_i \Psi_{p,i}(\mathbf{X}), \quad Q(\mathbf{X}; \mathbf{q}) = \sum_{i=0}^{n_q-1} q_i \Psi_{q,i}(\mathbf{X}). \quad (12)$$

Here $\{p_i \in \mathbb{C}, i = 0, \dots, n_p - 1\}$ and $\{q_i \in \mathbb{C}, i = 0, \dots, n_q - 1\}$ are complex coefficients and $\Psi_{p,i}$ and $\Psi_{q,i}$ are the multivariate orthonormal (probabilist) Hermite polynomials. The sets $\{\Psi_{p,i}, i = 0, \dots, n_p\}$ and $\{\Psi_{q,i}, i = 0, \dots, n_q\}$ are constructed through the d -fold tensorization of the univariate normalized Hermite polynomials, i.e.,

$$\Psi_{\mathbf{a}_p} = \prod_{i=1}^d \psi_{a_{p,i}}(X_i), \quad \Psi_{\mathbf{a}_q} = \prod_{i=1}^d \psi_{a_{q,i}}(X_i). \quad (13)$$

In here, $\mathbf{a}_p \in \mathbb{N}^d$ and $\mathbf{a}_q \in \mathbb{N}^d$ denote the index sets of the corresponding multivariate polynomials. Different strategies can be applied to obtain the index sets, out of which the total degree and the hyperbolic truncation scheme [47] are popular choices. In the total degree (TD) truncation, we retain all polynomials with a total polynomial degree less than or equal to m_p or m_q , i.e.,

$$\sum_{i=1}^d a_{p,i} \leq m_p, \quad \sum_{i=1}^d a_{q,i} \leq m_q. \quad (14)$$

The resulting number of polynomial terms in the total degree truncation are $n_p = \binom{d+m_p}{m_p}$ and $n_q = \binom{d+m_q}{m_q}$. The truncated set of multivariate polynomials is finally sorted in the lexicographic order [48]. The truncation rules are separately applied to both, numerator and denominator polynomial, with maximum polynomial degrees m_p and m_q .

We define the RPCE $R(\mathbf{X})$ obtained by taking the ratio of the two PCE representations of Eq. (12):

$$R(\mathbf{X}; \mathbf{p}, \mathbf{q}) := \frac{P(\mathbf{X}; \mathbf{p})}{Q(\mathbf{X}; \mathbf{q})} = \frac{\sum_{i=0}^{n_p-1} p_i \Psi_{p,i}(\mathbf{X})}{\sum_{i=0}^{n_q-1} q_i \Psi_{q,i}(\mathbf{X})}. \quad (15)$$

Stochastic collocation [49, 43] and Galerkin [42] methods to determine the coefficients \mathbf{p} and \mathbf{q} in the expansions in Eq. (15) have been presented in the literature. Recently, a sparse Bayesian learning approach was presented in [38] that allows to learn a sparse and probabilistic representation of the coefficient vectors. Therein, a two-stage strategy is proposed that utilizes the fact that the RPCE is linear in terms of the numerator coefficients. This allows to express the posterior distribution of the numerator coefficients, conditional on the denominator coefficients, in closed-form.

The posterior distribution of the denominator coefficients is then approximated through a Dirac at its MAP value. The method efficiently tackles the overfitting problem and significantly reduces the number of model evaluations compared to the least-squares regression approach in [43]. In [38], the likelihood of the expansion coefficients given a set of samples from \mathbf{X} and corresponding model evaluations is derived based on an assumption on the model misfit ε_S , defined as

$$\varepsilon_S = \mathcal{M}(\mathbf{X}) - R(\mathbf{X}; \mathbf{p}, \mathbf{q}). \quad (16)$$

It can be observed that the error in Eq. (16) is non-linear with respect to the denominator coefficients in \mathbf{q} . In the context of least-squares regression, minimizing the error requires an iterative solution procedure. We can, however, adjust the error ε_S , by multiplying Eq. (16) with the denominator polynomial, which gives the augmented error

$$\tilde{\varepsilon}_S = Q(\mathbf{X})\varepsilon_S = Q(\mathbf{X}; \mathbf{q}) \mathcal{M}(\mathbf{X}) - P(\mathbf{X}; \mathbf{p}). \quad (17)$$

The augmented error $\tilde{\varepsilon}_S$ is linear with respect to the denominator coefficients. In the following we will develop a Bayesian regression approach for learning the coefficients in the RPCE based on assuming a normal distribution for ε_S .

2.2.2. Training the RPCE through sparse Bayesian learning

For training the model, we assume that a data set $\mathcal{D}_M = \{(\mathbf{x}^{(k)}, m^{(k)}) | k = 1, \dots, n_{\text{tr}}\}$ of input samples $\mathbf{x}^{(k)}$ and corresponding model evaluations $m^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$ is available. We apply a Bayesian approach to learn the coefficients \mathbf{p} and \mathbf{q} , i.e., we treat them as random variables and estimate their posterior distribution, expressed through Bayes' rule as

$$f(\mathbf{p}, \mathbf{q} | \mathbf{m}) = c_S^{-1} f(\mathbf{m} | \mathbf{p}, \mathbf{q}) f(\mathbf{p}, \mathbf{q}), \quad (18)$$

where $f(\mathbf{m} | \mathbf{p}, \mathbf{q})$ is the likelihood function and $f(\mathbf{p}, \mathbf{q})$ is the prior distribution. The likelihood function is derived based on a Gaussian assumption on the augmented error of Eq. (17). We denote by $\mathbf{m} = [m^{(1)}; \dots; m^{(n_{\text{tr}})}] \in \mathbb{C}^{n_{\text{tr}} \times 1}$ the vector of model responses. Then, $\tilde{\varepsilon}_{S,k}$ denotes the augmented error at the k -th sample location, i.e., $\tilde{\varepsilon}_{S,k} = Q(\mathbf{x}^{(k)}; \mathbf{q})m^{(k)} - P(\mathbf{x}^{(k)}; \mathbf{p})$ and $\tilde{\varepsilon}_S = [\tilde{\varepsilon}_{S,1}; \dots; \tilde{\varepsilon}_{S,n_{\text{tr}}}]$ is a random vector with outcome space $\mathbb{C}^{n_{\text{tr}} \times 1}$ that collects all errors:

$$\tilde{\varepsilon}_S = \text{diag}(\Psi_q \mathbf{q}) \mathbf{m} - \Psi_p \mathbf{p}, \quad (19)$$

where $\Psi_p \in \mathbb{R}^{n_{\text{tr}} \times n_p}$ and $\Psi_q \in \mathbb{R}^{n_{\text{tr}} \times n_q}$ have as (k, j) -elements $\Psi_{p,j}(\mathbf{x}^{(k)})$ and $\Psi_{q,j}(\mathbf{x}^{(k)})$. We assume that the augmented errors $\tilde{\varepsilon}_S$ are jointly complex Gaussian with zero mean and covariance matrix $\Sigma_{\tilde{\varepsilon}_S \tilde{\varepsilon}_S} = \beta_S^{-1} \mathbf{I}_{n_{\text{tr}}}$, where \mathbf{I}_n denotes the

$n \times n$ identity matrix. Note that, since $\varepsilon_{S,k} = \tilde{\varepsilon}_{S,k}/Q(\mathbf{x}^{(k)}; \mathbf{q})$, the resulting covariance matrix for the vector of standard errors is

$$\boldsymbol{\Sigma}_{\varepsilon_S \varepsilon_S} = \beta_S^{-1} \mathbf{Q}^{-1} \mathbf{Q}^{-H}, \quad (20)$$

where $\mathbf{Q} := \text{diag}(\boldsymbol{\Psi}_p \mathbf{q})$. The conditional expectation and covariance matrix for the data vector \mathbf{m} follow as

$$\mathbb{E}[\mathbf{m} | \mathbf{p}, \mathbf{q}] = \mathbf{Q}^{-1} \boldsymbol{\Psi}_p \mathbf{p}, \quad (21)$$

$$\text{Cov}[\mathbf{m} | \mathbf{p}, \mathbf{q}] = \boldsymbol{\Sigma}_{\varepsilon \varepsilon} = \beta_S^{-1} \mathbf{Q}^{-1} \mathbf{Q}^{-H} = \beta_S^{-1} \text{diag}\left(\left|Q(\mathbf{x}^{(1)}; \mathbf{q})\right|^{-2}, \dots, \left|Q(\mathbf{x}^{(n_{tr})}; \mathbf{q})\right|^{-2}\right). \quad (22)$$

We note that the mean of the data is the same as the one of the standard residual formulation used in the sparse Bayesian regression approach of [38], whereas the variance of the data is weighted with the inverse of the square-magnitude of the denominator polynomial at the sample locations. Since \mathbf{m} depends linearly on $\tilde{\varepsilon}_S$, the likelihood

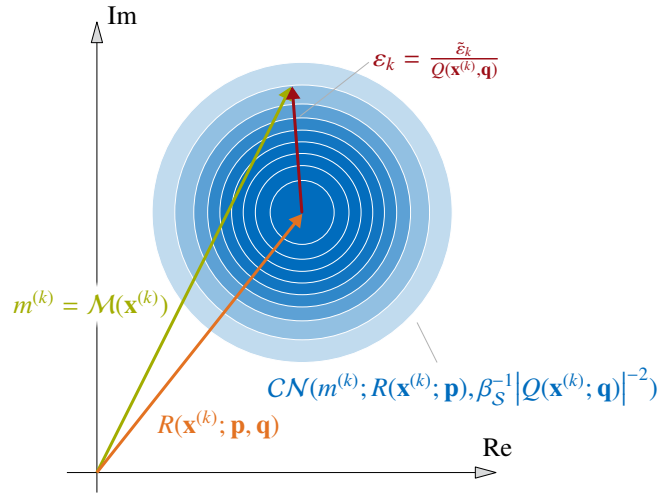


Figure 2: Illustration of the model error in the complex plane that is defined to derive the likelihood function. We assume an additive error $\tilde{\varepsilon}_k$ between the product of original model response and denominator polynomial $Q(\mathbf{x}^{(k)})M(\mathbf{x}^{(k)})$ and the numerator polynomial $P(\mathbf{x}^{(k)})$.

will be complex Gaussian with moments as in Eqs. (21) and (22). Then,

$$f(\mathbf{m} | \mathbf{p}, \mathbf{q}) = \left(\frac{\beta_S}{\pi}\right)^{n_{tr}} \det(\mathbf{Q}^H \mathbf{Q}) \exp\left\{-\beta_S (\mathbf{m} - \mathbf{Q}^{-1} \boldsymbol{\Psi}_p \mathbf{p})^H \mathbf{Q}^H \mathbf{Q} (\mathbf{m} - \mathbf{Q}^{-1} \boldsymbol{\Psi}_p \mathbf{p})\right\} \quad (23)$$

$$= \left(\frac{\beta_S}{\pi}\right)^{n_{tr}} \det(\mathbf{Q}^H \mathbf{Q}) \exp\left\{-\beta_S (\mathbf{Q} \mathbf{m} - \boldsymbol{\Psi}_p \mathbf{p})^H (\mathbf{Q} \mathbf{m} - \boldsymbol{\Psi}_p \mathbf{p})\right\}. \quad (24)$$

The error model and the involved relationship between the model and the surrogate are illustrated in Fig. 2. We employ independent priors for the numerator and denominator coefficients. Following the approach in [44, 38], the

prior distribution for the numerator coefficients is modeled as a zero mean complex proper Gaussian distribution, i.e.,

$$f(\mathbf{p}|\alpha_p) = \mathcal{CN}(\mathbf{p}|\mathbf{0}, \mathbf{\Lambda}_{\mathbf{pp}}^{-1}, \mathbf{0}), \quad (25)$$

where $\mathbf{\Lambda}_{\mathbf{pp}} = \text{diag } \alpha_p$ constitutes the numerator coefficients precision matrix and $\alpha_p = [\alpha_{p,1}; \dots; \alpha_{p,n_p}]$ is a vector containing the n_p hyperparameters (precisions). Since \mathbf{p} and \mathbf{m} given \mathbf{q} are jointly proper complex Gaussian, the conditional distribution of the numerator coefficients \mathbf{p} given \mathbf{m} and \mathbf{q} will also be proper complex Gaussian. Through combining Eqs. (24) and (25), after a few transformations and application of the Woodbury identity, we obtain the posterior mean and covariance matrix of the numerator coefficients as

$$\boldsymbol{\mu}_{\mathbf{p}|\mathbf{m},\mathbf{q}} = \beta_S \boldsymbol{\Sigma}_{\mathbf{pp}|\mathbf{m}} \boldsymbol{\Psi}_p^H \underbrace{\mathbf{Q}\mathbf{m}}_{:=\tilde{\mathbf{m}}}, \quad (26)$$

$$\boldsymbol{\Sigma}_{\mathbf{pp}|\mathbf{m}} = (\mathbf{\Lambda}_{\mathbf{pp}} + \beta_S \boldsymbol{\Psi}_p^H \boldsymbol{\Psi}_p)^{-1}, \quad (27)$$

where by $\tilde{\mathbf{m}} := \mathbf{Q}\mathbf{m} = [Q(\mathbf{x}^{(1)})_{y^{(1)}}; \dots; Q(\mathbf{x}^{(n_o)})_{y^{(n_o)}}]$ we denote the augmented data vector. In contrast to the formulation in [38], the posterior conditional covariance matrix of the numerator coefficients given \mathbf{q} does not depend on \mathbf{q} . Thus, the posterior distribution of the numerator coefficients, conditional on the denominator coefficients, is proper complex Gaussian with

$$f(\mathbf{p}|\mathbf{m}, \mathbf{q}, \alpha_p, \beta_S) = \frac{1}{\pi^{n_p} \det \boldsymbol{\Sigma}_{\mathbf{pp}|\mathbf{m}}} \exp\left(-(\mathbf{p} - \boldsymbol{\mu}_{\mathbf{p}|\mathbf{m},\mathbf{q}})^H \boldsymbol{\Sigma}_{\mathbf{pp}|\mathbf{m}}^{-1} (\mathbf{p} - \boldsymbol{\mu}_{\mathbf{p}|\mathbf{m},\mathbf{q}})\right). \quad (28)$$

The marginal likelihood of the data given \mathbf{q} is then obtained as

$$f(\mathbf{m}|\mathbf{q}, \alpha_p, \beta_S) = \frac{\det(\mathbf{Q}\mathbf{Q}^H)}{\pi^N \det(\mathbf{C})} \exp\{-\mathbf{m}^H \mathbf{Q}^H \mathbf{C}^{-1} \mathbf{Q}\mathbf{m}\}, \quad (29)$$

where

$$\mathbf{C} = \beta_S^{-1} \mathbf{I}_{n_r} + \boldsymbol{\Psi}_p \mathbf{\Lambda}_{\mathbf{pp}}^{-1} \boldsymbol{\Psi}_p^T. \quad (30)$$

In analogy to Eq. (25), the prior distribution of the denominator coefficients is assumed to be a zero mean proper complex Gaussian distribution, i.e.,

$$f(\mathbf{q}|\alpha_q) = \mathcal{CN}(\mathbf{q}|\mathbf{0}, \mathbf{\Lambda}_{\mathbf{qq}}^{-1}, \mathbf{0}). \quad (31)$$

where $\Lambda_{\mathbf{q}\mathbf{q}} = \text{diag } \alpha_q$ constitutes the denominator coefficients precision matrix and $\alpha_q = [\alpha_{q,1}; \dots; \alpha_{q,n_q}]$ is a vector containing the n_q hyperparameters (precisions).

The outlined formulation does not permit a closed-form solution for the posterior distribution of \mathbf{q} since the computation of the model evidence

$$f(\mathbf{m}|\alpha_p, \alpha_q, \beta_S) = \int_{\mathbb{C}^{n_q}} f(\mathbf{m}|\mathbf{q}, \alpha_p, \beta_S) f(\mathbf{q}|\alpha_q) d\mathbf{q}, \quad (32)$$

cannot be performed in closed-form. We therefore choose to approximate the posterior distribution of \mathbf{q} through a proper complex Gaussian distribution (cf. Appendix B.1 in [38] and references therein) centered at the maximum a posteriori (MAP) estimate of \mathbf{q} . In the analysis of real-valued random variables, this Gaussian approximation is commonly referred to as *Laplace's* approximation. Note that choosing a proper Gaussian distribution as an approximation neglects the off-diagonal elements in the full Hessian matrix [50]. This will restrain the posterior distribution of the denominator coefficients such that the real and imaginary part of a coefficient q_i will share the same variance. Subsequently, we find a set of hyperparameters that maximizes the approximate model evidence, which is known as type-II-maximum likelihood estimation. This strategy leads to a sequential two-stage approach for the approximation of the posterior distribution of the denominator coefficients as well as the estimation of the hyperparameters.

In a first step, the MAP estimate of the denominator coefficients \mathbf{q}^* is found through solving the following optimization problem

$$\mathbf{q}^* = \arg \max_{\mathbf{q} \in \mathbb{C}^{n_q}} f(\mathbf{m}|\mathbf{q}, \alpha_p, \beta_S) f(\mathbf{q}|\alpha_q). \quad (33)$$

Then, the posterior distribution of the denominator coefficients is approximated through

$$f(\mathbf{q}|\mathbf{m}, \alpha_p, \alpha_q, \beta_S) \approx \frac{1}{\pi^{n_q} \det\left(-(\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1}\right)} \exp\left\{- (\mathbf{q} - \mathbf{q}^*)^H \left(-\mathbf{H}_{\mathbf{q}\mathbf{q}}\right) (\mathbf{q} - \mathbf{q}^*)\right\}, \quad (34)$$

where $\mathbf{H}_{\mathbf{q}\mathbf{q}}$ denotes the \mathbb{C} -complex Hessian [50] of the log-posterior function of the denominator coefficients. We note that $\mathbf{H}_{\mathbf{q}\mathbf{q}}$ does not depend on the denominator coefficients and, hence, we do not explicitly state the evaluation point $\mathbf{q} = \mathbf{q}^*$. Utilizing the approximation in Eq. (34), the model evidence can be approximated through

$$f(\mathbf{m}|\alpha_p, \alpha_q, \beta_S) \approx f(\mathbf{m}|\mathbf{q}^*, \alpha_p, \beta_S) f(\mathbf{q}^*|\alpha_q) \pi^{n_q} \det\left(-(\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1}\right). \quad (35)$$

An optimal set of hyperparameters is then found through type-II-maximum likelihood estimation. To this end, we

maximize the approximate model evidence in Eq. (35) over all hyperparameters, such that

$$[\alpha_p^*, \alpha_q^*, \beta_S^*] = \arg \max_{\substack{[\alpha_p, \alpha_q, \beta_S] \\ \in \mathbb{R}^{p \times n_q \times 1}}} f(\mathbf{m}|\mathbf{q}^*, \alpha_p, \beta_S) f(\mathbf{q}^*|\alpha_q) \det\left(-(\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1}\right). \quad (36)$$

In the following, we summarize the resulting expressions in each of the above steps. Further detailed derivations are given in Appendix A to Appendix D.

In order to find the MAP estimate of \mathbf{q} , we maximize the log of the objective function in Eq. (33), which inserting Eqs. (29) and (31) into Eq. (33) and taking the logarithm results in

$$\mathbf{q}^* = \arg \max_{\mathbf{q} \in \mathbb{C}^{n_q}} \left[\ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H (\Upsilon^H \mathbf{C}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right], \quad (37)$$

where $\Upsilon = \text{diag}(\mathbf{m}) \Psi_q \in \mathbb{C}^{n_{\text{tr}} \times n_q}$. Eq. (37) is a nonlinear optimization problem in complex variables. In order to solve Eq. (37), we resort to a gradient based maximization technique. Since $f(\mathbf{m}|\mathbf{q}, \alpha_p, \beta_S) f(\mathbf{q}|\alpha_q)$ is the product of two PDFs, it is real-valued and thus a necessary condition for the objective function in Eq. (37) to take a maximum is given by

$$\frac{\partial}{\partial \bar{\mathbf{q}}} \left[\ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H (\Upsilon^H \mathbf{C}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] = \mathbf{0}_{n_q}, \quad (38)$$

where $\frac{\partial}{\partial \bar{\mathbf{q}}}$ denotes the generalized (or Wirtinger) derivative with respect to the complex conjugate of the denominator coefficients \mathbf{q} , denoted as the conjugate cogradient. The definition of the generalized derivatives can be found in [50, 51]. Since the objective function is real-valued, it is sufficient to consider only the conjugate cogradient, since in this case it holds $\frac{\partial f(\mathbf{q})}{\partial \mathbf{q}} = \overline{\frac{\partial f(\mathbf{q})}{\partial \bar{\mathbf{q}}}}$. In order to solve Eq. (37), we employ a quasi-Newton method and use a limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, provided by [52]. Details about the algorithm can be found in [53]. The algorithm uses a quasi-Newton step to update an approximation of the Hessian matrix of the problem in each iteration. We use the available line-search algorithm in the implementation by [52].

The quasi-Newton method requires the derivatives of the objective function with respect to the conjugate denominator coefficients. The derivative in Eq. (38) can be found analytically and reads

$$\frac{\partial}{\partial \bar{\mathbf{q}}} \left[\ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H (\Upsilon^H \mathbf{C}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] = \Psi_q^T (\Psi_q \bar{\mathbf{q}})^{\circ -1} - (\Upsilon^H \mathbf{C}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q}, \quad (39)$$

where $(\cdot)^{\circ -1}$ denotes the element-wise inverse. The next step in the outlined sequential approach requires the compu-

tation of the Hessian $\mathbf{H}_{\mathbf{q}\mathbf{q}}$ of the log-posterior. We find

$$\mathbf{H}_{\mathbf{q}\mathbf{q}} = \frac{\partial}{\partial \mathbf{q}} \left(\frac{\partial}{\partial \bar{\mathbf{q}}} \left[\ln \det \mathbf{Q}\mathbf{Q}^{\text{H}} - \mathbf{q}^{\text{H}} (\boldsymbol{\Upsilon}^{\text{H}} \mathbf{C}^{-1} \boldsymbol{\Upsilon} + \boldsymbol{\Lambda}_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] \right)^{\text{T}} = -(\boldsymbol{\Upsilon}^{\text{H}} \mathbf{C}^{-1} \boldsymbol{\Upsilon} + \boldsymbol{\Lambda}_{\mathbf{q}\mathbf{q}}). \quad (40)$$

We derive the partial derivatives and Hessian matrices in Appendix A.

After having computed Laplace's approximation for the denominator coefficients, we can express Eq. (36) using Eqs. (29), (31) and (40) as

$$\begin{aligned} [\boldsymbol{\alpha}_p^*, \boldsymbol{\alpha}_q^*, \beta_S^*] = \arg \max_{\substack{[\boldsymbol{\alpha}_p, \boldsymbol{\alpha}_q, \beta_S] \\ \in \mathbb{R}^{n_p \times n_q \times 1}}} & \left[n_{\text{tr}} \ln \beta_S + \ln \det \boldsymbol{\Sigma}_{\text{pp|m}} + \ln \det \boldsymbol{\Lambda}_{\text{pp}} + \ln \det \boldsymbol{\Lambda}_{\mathbf{q}\mathbf{q}} - \right. \\ & \left. - \beta_S \mathbf{m}^{\text{H}} \bar{\mathbf{Q}}^* (\mathbf{Q}^* \mathbf{m} - \boldsymbol{\Psi}_p \boldsymbol{\mu}_{\text{p|m}, \mathbf{q}^*}) - \mathbf{q}^{*\text{H}} \boldsymbol{\Lambda}_{\mathbf{q}\mathbf{q}} \mathbf{q}^* - \ln \det (\boldsymbol{\Upsilon}^{\text{H}} \mathbf{C}^{-1} \boldsymbol{\Upsilon} + \boldsymbol{\Lambda}_{\mathbf{q}\mathbf{q}}) \right], \quad (41) \end{aligned}$$

where $\mathbf{Q}^* := \text{diag}(\boldsymbol{\Psi}_q \mathbf{q}^*)$ denotes the diagonal matrix of denominator predictions evaluated for the MAP of \mathbf{q} . Here, we have used the Woodbury identity to express the inverse and determinant of \mathbf{C} as

$$\mathbf{C}^{-1} = \beta_S \mathbf{I}_{n_{\text{tr}}} - \beta_S^2 \boldsymbol{\Psi}_p \boldsymbol{\Sigma}_{\text{pp|m}} \boldsymbol{\Psi}_p^{\text{T}}, \quad (42)$$

$$\det \mathbf{C} = \det \boldsymbol{\Sigma}_{\text{pp|m}}^{-1} \det \boldsymbol{\Lambda}_{\text{pp}}^{-1} \det \beta_S^{-1} \mathbf{I}_{n_{\text{tr}}}. \quad (43)$$

The derivatives of the objective function in Eq. (41) are derived in Appendix B, Appendix C, and Appendix D. Based on these, we obtain the following update rules for the hyperparameters:

$$\alpha_{p,i} = \frac{1}{\left[\boldsymbol{\Sigma}_{\text{pp|m}} \right]_{ii} + \left[\boldsymbol{\mu}_{\text{p|m}, \mathbf{q}^*} \right]_i^2 + \left[\boldsymbol{\Delta}^{\text{H}} (-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \boldsymbol{\Delta} \right]_{ii}}, \quad (44)$$

$$\alpha_{q,i} = \frac{1}{|q_i^*|^2 + \left[(-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \right]_{ii}}, \quad (45)$$

$$\beta_S = \frac{N}{\| \mathbf{Q}\mathbf{m} - \boldsymbol{\Psi}_p \boldsymbol{\mu}_{\text{p|m}, \mathbf{q}^*} \|^2 + \text{tr}(\boldsymbol{\Sigma}_{\text{pp|m}} \boldsymbol{\Psi}_p^{\text{T}} \boldsymbol{\Psi}_p) + \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \mathbf{O}^{\text{H}} \mathbf{O} \right)}, \quad (46)$$

where $\boldsymbol{\Delta} = \beta_S \boldsymbol{\Upsilon}^{\text{H}} \boldsymbol{\Psi}_p \boldsymbol{\Sigma}_{\text{pp|m}}$ and $\mathbf{O} = (\mathbf{I}_{n_{\text{M}}} - \beta_S \boldsymbol{\Psi}_p \boldsymbol{\Sigma}_{\text{pp|m}} \boldsymbol{\Psi}_p^{\text{T}}) \boldsymbol{\Upsilon}$. The model coefficients are sequentially updated in an iterative manner until convergence is reached. Convergence is controlled through observation of the change of the hyperparameters. Whenever their change in log-scale drops below a certain, user-specified threshold, the algorithm is

terminated and the current state of coefficients and hyperparameters is utilized further. The algorithm is summarized in 1 and straightforwardly follows the one for the standard residual. We refer to [38] for further details.

Algorithm 1 Sparse Bayesian RPCE training

Require: The maximum number of iterations, i_{\max} , the tolerances for the convergence criteria ε_α and ε_β , the pruning thresholds $\alpha_{p,\max}$ and $\alpha_{q,\max}$, the initial coefficients \mathbf{p}_{init} and \mathbf{q}_{init} , k used in k -normalization of \mathbf{q}

while $i < i_{\max} \wedge (\max(\Delta \log \alpha) > \varepsilon_\alpha \vee \Delta \log \beta > \varepsilon_\beta)$ **do**

Find useful numerator weights p_i with $\alpha_{p,i}$ lower than threshold $\alpha_{p,\max}$, prune all other basis functions.

Update denominator coefficients \mathbf{q} with normalized MAP-estimate, $\mathbf{q} \leftarrow \mathbf{q}^* \cdot \|\mathbf{q}^*\|_k^{-1}$, by solving Eq. (33), and compute Hessian matrix $\mathbf{H}_{\mathbf{q}\mathbf{q}}$ using Eq. (40).

Find useful denominator coefficients q_i with $\alpha_{q,i}$ lower than threshold $\alpha_{q,\max}$, prune all other basis functions.

Update posterior mean and covariance for numerator coefficients, $\mathbf{p} \leftarrow \boldsymbol{\mu}_{\mathbf{p}|\mathbf{m},\mathbf{q}^*}$, $\boldsymbol{\Sigma} \leftarrow \boldsymbol{\Sigma}_{\mathbf{p}|\mathbf{m}}$ using Eqs. (27) and (26).

Update α_p , α_q and β using the update rules in Eqs. (44), (45) and (46)

Update maximum change in hyperparameters $\Delta \log \alpha_{\max,i} \leftarrow \Delta \log \alpha_{\max,i+1}$ and $\Delta \log \beta_{S,i} \leftarrow \Delta \log \beta_{S,i+1}$

end while

return Retained coefficients \mathbf{p} and \mathbf{q}

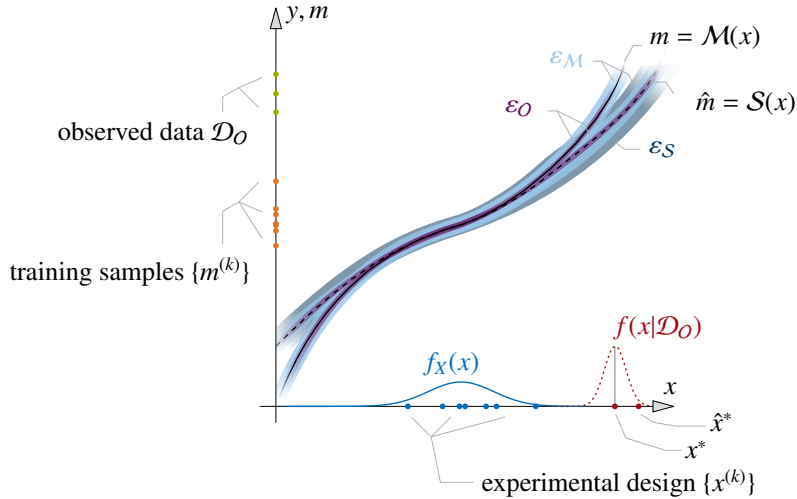


Figure 3: Illustration of the inference procedure using a surrogate model. The solid line represents the original model $\mathcal{M}(x)$. Based on the prior distribution $f(x)$, a set of input samples $\{x^{(k)}\}$, the experimental design, is generated. The corresponding model output samples, the training set, is computed as $\mathcal{M}(x^{(k)})$. The trained surrogate model $\hat{\mathcal{S}}(x^{(k)})$ can usually be expected to show good accuracy in the bulk of the prior distribution. For the inverse relationship, the model output and the observed data are related via the model error ε_M and the additive observation noise ε_O . If the surrogate model is used in the inverse problem, an additional random error ε_S can be considered. If now data \mathcal{D}_O is observed significantly distant from the training samples, the surrogate model accuracy in the corresponding region in the input space might be poor. Thus, the inferred samples, or the MAP estimate, might be prone to significant error. This motivates the development of the active learning procedure to place experimental design samples towards the bulk of the posterior distribution. Remark: We disregard the probabilistic nature of the Bayesian regression surrogate model in this illustration for reasons of simplicity. The surrogate model response is really a random process over x .

Until now we have assumed that the training procedure is performed using a fixed-size experimental design. Such design can be obtained through samples following the prior probability distribution of the input parameters. This approach works well in settings where the rational surrogate model is subsequently used for forward uncertainty

propagation applications based on the prior distributions, where one is interested in computing the probability distribution of a model output or function thereof based on a given distribution of the input random variables. In this case, the training samples will align with the sought-after distribution and the surrogate will possess good predictability in the relevant part of the domain of the output space. In the inverse setting, introduced in Sec. 2.1, however, the location of the bulk of the posterior probability mass is typically unknown a priori. Whenever the generalization capability of the surrogate model is poor or the bulk of the posterior probability mass lies far from the bulk of the prior probability mass, this approach might introduce significant surrogate model errors. A graphical illustration of this situation is given in Fig. 3. This problem can be tackled through active, sequential experimental design strategies, where the experimental design is sequentially enriched based on exploiting information in the already observed sample data. A powerful framework that builds towards this goal is Bayesian Optimization. We subsequently propose an active learning approach for training the introduced Bayesian rational surrogate model in the context of MAP estimation that draws upon Bayesian Optimization.

2.3. Active, sequential design through Bayesian optimization

Our goal is the computation of the MAP value as outlined in Section 2.1. In order to reduce the computational cost of the evaluation of the objective function, defined in Eq. (10), we replace the set of original models $\{\mathcal{M}_i\}$ with their surrogate model counterparts $\{\mathcal{S}_i\}$. Due to the fact that we are working with frequency domain models, the RPCE surrogate model, as introduced in Section 2.2, is a specifically well-suited approximation. To this end, each of the individual models that enter the objective function is approximated through an RPCE, i.e., $S_i(\mathbf{x}) = R(\mathbf{x}; \mathbf{p}_i, \mathbf{q}_i)$. For the n_O surrogate models, we collect the coefficients in the matrix $\mathbf{R} = [\mathbf{p}_1, \dots, \mathbf{p}_{n_O}, \mathbf{q}_1, \dots, \mathbf{q}_{n_O}]$, where \mathbf{p}_i and \mathbf{q}_i are the vectors of numerator and denominator coefficients of the i -th surrogate model. Under consideration of the surrogate model error, as defined in Eq. (16), we can replace the objective function in Eq. (10) by

$$\hat{h}(\mathbf{x}; \mathbf{R}, \boldsymbol{\varepsilon}_S) = \frac{1}{2} \left[-\ln \det \boldsymbol{\Sigma}_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon} - \ln \det \boldsymbol{\Sigma}_{\boldsymbol{\varphi}_\varepsilon \boldsymbol{\varphi}_\varepsilon} - \left\| \mathbf{w}_O - (\mathbf{w}_S(\mathbf{x}; \mathbf{R}) + \ln |\boldsymbol{\varepsilon}_S|) \right\|_{\boldsymbol{\Sigma}_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon}^{-1}}^2 - \left\| \boldsymbol{\varphi}_O - (\boldsymbol{\varphi}_S(\mathbf{x}; \mathbf{R}) + \angle \boldsymbol{\varepsilon}_S) \right\|_{\boldsymbol{\Sigma}_{\boldsymbol{\varphi}_\varepsilon \boldsymbol{\varphi}_\varepsilon}^{-1}}^2 \right] + \ln f(\mathbf{x}), \quad (47)$$

where

$$\mathbf{w}_S(\mathbf{x}; \mathbf{R}) = [\ln |y_{S,1}(\mathbf{x}; \mathbf{p}_1, \mathbf{q}_1)|; \dots; \ln |y_{S,n_O}(\mathbf{x}; \mathbf{p}_{n_O}, \mathbf{q}_{n_O})|], \quad (48)$$

$$\boldsymbol{\varphi}_S(\mathbf{x}; \mathbf{R}) = [\angle y_{S,1}(\mathbf{x}; \mathbf{p}_1, \mathbf{q}_1); \dots; \angle y_{S,n_O}(\mathbf{x}; \mathbf{p}_{n_O}, \mathbf{q}_{n_O})], \quad (49)$$

$$\boldsymbol{\varepsilon}_S = [\varepsilon_{S,1}; \dots; \varepsilon_{S,n_O}], \quad (50)$$

and $\varepsilon_{S,i}$ is the surrogate model error of the i -th RPCE model as defined in Eq. (16) that has covariance matrix as given in Eq. (20) with precision parameter $\beta_{S,i}$. We note that the objective function in Eq. (47), $\hat{h}(\mathbf{x}; \mathbf{R}, \varepsilon_S)$, now becomes a random process over the model input parameters that is expressed in terms of the random variables in \mathbf{R} and ε_S . Therefore, we treat the problem as a Bayesian optimization problem. In the context of Bayesian optimization the key idea is now to formulate the optimization problem as a Bayesian decision problem, which, based on a user-specified acquisition function, allows to adaptively sample the experimental design. The acquisition function can be understood through an underlying utility function in the Bayesian decision making framework [25, 27]. Subsequently, we comprehensively summarize the framework.

We denote by $\mathcal{D}_{\mathcal{M}}^1 = \{(\mathbf{x}^{(k)}, \mathbf{y}_{\mathcal{M}}(\mathbf{x}^{(k)})) | k = 1, \dots, n_{\text{init}}\}$ the initial training set with n_{init} number of samples. Therein, it holds $\mathbf{x}^{(k)} \in \mathbb{R}^d$ and $\mathbf{y}_{\mathcal{M}}(\mathbf{x}^{(k)}) \in \mathbb{C}^{n_o}$. Throughout the active learning procedure the data set will be enriched. In iteration step n , the observed data set up to this step is denoted $\mathcal{D}_{\mathcal{M}}^n$. The expected improvement acquisition is then defined as

$$\alpha_{\text{EI}}(\mathbf{x}; \mathcal{D}_{\mathcal{M}}^n) = \mathbb{E}_{\hat{h}|\mathcal{D}_{\mathcal{M}}^n} \left[\left(\hat{h}(\mathbf{x}; \mathbf{R}, \varepsilon_S) - h_{\text{max}}, 0 \right) \right], \quad (51)$$

where h_{max} denotes the currently observed maximum objective value in iteration step n , i.e.,

$$h_{\text{max}} = \max_{\mathbf{x} \in \{\mathbf{x}^{(k)} | k=1, \dots, n_{\text{init}}+n\}} h(\mathbf{x}). \quad (52)$$

The expectation in Eq. (51) is stated with respect to the distribution of the approximate objective \hat{h} , which is unknown. The randomness in \hat{h} for a given \mathbf{x} is due to its dependence on random variables \mathbf{R} and ε_S , hence, the expectation can be equivalently stated with respect to these random variables. Since the expectation in Eq. (51) cannot be solved in closed form, we resort to a sampling-based approximation. To this end, we approximate the expected improvement by the arithmetic mean as follows

$$\alpha_{\text{EI}}(\mathbf{x}; \mathcal{D}_{\mathcal{M}}^n) \approx \hat{\alpha}_{\text{EI}}(\mathbf{x}; \mathcal{D}_{\mathcal{M}}^n) = \frac{1}{n_{\alpha}} \sum_{k=1}^{n_{\alpha}} \max \left(\hat{h}(\mathbf{x}; \mathbf{R}^{(k)}, \varepsilon_S^{(k)}) - h_{\text{max}}, 0 \right), \quad (53)$$

where $\mathbf{R}^{(k)}$ contains samples of the RPCE coefficients and $\varepsilon_S^{(k)}$ denotes a sample of the surrogate model error vector. All samples are conditional on the available training points in the data set $\mathcal{D}_{\mathcal{M}}^n$ and the hyperparameters identified in the training process described in Section 2.2.2. We generate samples of the denominator coefficients of each surrogate model based on the identified hyperparameters after termination of the training using a transitional Markov Chain Monte Carlo [54, 55] sampler. Conditional on the denominator coefficients, the numerator coefficients will follow a complex normal distribution with mean vector and covariance matrix as defined in Eqs. (26) and (27). The surrogate

model error, conditional on the denominator coefficients, also follows a complex normal distribution with covariance matrix as defined in Eq. (20). This hierarchical dependency is illustrated in Fig. 4.

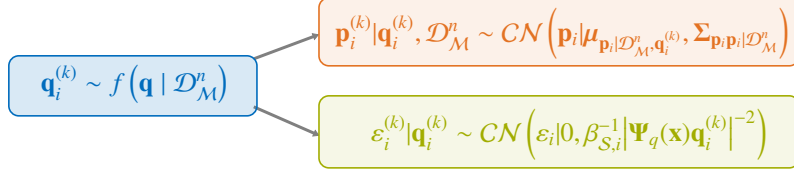


Figure 4: Illustration of the posterior sampling and the hierarchical dependency.

For numerical purposes, we approximate the maximum function in the expected improvement criterion through the softplus function, i.e., $\max(x, 0) \approx \text{softplus}(x)$. The softplus function is defined through

$$\text{softplus}(x) = \gamma^{-1} \log(1 + \exp(\gamma x)) . \quad (54)$$

This approximation is often encountered in machine learning algorithms to remove the discontinuity in the derivative of the maximum function [56]. Since the proposed approach approximates the expected value in Eq. (51) through the sample mean, the acquisition function will yield zeros for great portions of the input space. Relaxing the maximum function to the softplus function introduces continuous gradients in the acquisition function, which leads to a more robust identification of the optimum of the acquisition function.

Finally, the data set \mathcal{D}_M^{n+1} is enriched by the data point that maximizes the acquisition function, i.e.,

$$\mathbf{x}^+ = \arg \max_{\mathbf{x} \in \mathbb{R}^d} \hat{\alpha}_{\text{EI}}(\mathbf{x}; \mathcal{D}_M^n) . \quad (55)$$

The data set in the next iteration step is then $\mathcal{D}_M^{n+1} = \mathcal{D}_M^n \cup \{(\mathbf{x}^+, \mathbf{y}_M(\mathbf{x}^+))\}$. Subsequently, the set of surrogate models is retrained using the updated data set \mathcal{D}_M^{n+1} . The algorithm is terminated once a preset number of model function evaluations n_{budget} is reached. Subsequently, we consider two possibilities to determine the final optimizer \mathbf{x}^* after the algorithm has terminated, following the presentation in [25]. The first estimator, termed the *simple reward estimator*, selects as \mathbf{x}^* the point \mathbf{x} with highest objective function evaluation among the observed data points, i.e.,

$$\mathbf{x}_{\text{SR}}^* = \arg \max_{\mathbf{x} \in \{\mathbf{x}^{(k)} | k=1, \dots, n_{\text{budget}}\}} h(\mathbf{x}) . \quad (56)$$

This estimator thus only considers points for which the original model has been evaluated. The second estimator, termed the *global reward estimator*, is based on optimizing the objective function utilizing the random process surro-

gate model. To this end, a deterministic estimate of the random process model $\hat{h}(\mathbf{x})$ has to be chosen as a proxy in the optimization. A common choice is to use the mean of $\hat{h}(\mathbf{x})$. This estimator thereby utilizes the additional information contained in the surrogate model and should be chosen when there is trust in the surrogate model prediction away from the yet observed data points. Since evaluating the expected value of $\hat{h}(\mathbf{x})$ requires sampling in our case, we use the conditional mean of the numerator coefficients and the MAP estimate of the denominator coefficients as plug-in estimates and solve

$$\mathbf{x}_{\text{GR}}^* = \arg \max_{\mathbf{x} \in \mathbb{R}^d} \hat{h}(\mathbf{x}; \mathbf{R}^*, \mathbf{0}_{n_O}) . \quad (57)$$

where $\mathbf{R}^* = \left[\boldsymbol{\mu}_{\mathbf{p}_1 | \mathcal{D}_M^{n_{\text{budget}}}, \mathbf{q}_1^*}, \dots, \boldsymbol{\mu}_{\mathbf{p}_{n_O} | \mathcal{D}_M^{n_{\text{budget}}}, \mathbf{q}_{n_O}^*}, \mathbf{q}_1^*, \dots, \mathbf{q}_{n_O}^* \right]$. Furthermore, the surrogate model error mean $\mathbb{E}[\boldsymbol{\varepsilon}_S] = \mathbf{0}_{n_O}$ is chosen in the estimation of the global reward estimator. The full algorithm is summarized in Alg. 2. In the following

Algorithm 2 Active, sequential design through Bayesian optimization

Require: The maximum number of model evaluations, n_{budget} , the initial number of model evaluation n_{init} , the initial training data \mathcal{D}_M^1 , the observation data,

Set $n = 1$

while $n_{\text{init}} + n \leq n_{\text{budget}}$ **do**

Find the next sample point \mathbf{x}^+ as the optimizer of the expected improvement acquisition function in Eq. (53) using the surrogate models in $\{\mathcal{S}_i\}$. For this, generate samples of the posterior distribution of the surrogate models as illustrated in Fig. 4.

Evaluate the n_O models in $\{\mathcal{M}_i\}$ to yield $\mathbf{y}_M(\mathbf{x}^+)$.

Enrich the data set: $\mathcal{D}_M^{n+1} = \mathcal{D}_M^n \cup \{(\mathbf{x}^+, \mathbf{y}_M(\mathbf{x}^+))\}$.

Retrain the n_O surrogate models in $\{\mathcal{S}_i\}$ according to Algorithm 1 using the enriched data set \mathcal{D}_M^{n+1} .

end while

return MAP estimate \mathbf{x}^* , according to Eqs. (56) or (57).

section, we will apply the proposed Bayesian updating strategy to the problem of Bayesian parameter updating for two structural dynamic models.

3. Numerical examples

We apply the methodology outlined in Section 2 to two examples. For illustrative purposes, we investigate updating the parameters of a two degree of freedom system with varying numbers of input parameters in Section 3.1. The second example, presented in Section 3.2, includes updating the parameters of the finite element model of a cross-laminated timber plate.

In order to assess the performance of the proposed approach, we compare the results obtained from the active learning procedure with a reference solution obtained using the original model as well as a surrogate model based solution, where a fixed experimental design with size equal to the number of samples in the active learning data set

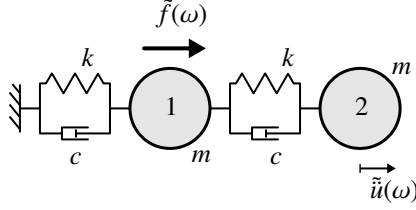


Figure 5: Illustration of two-degree-of-freedom system with fully dependent spring, mass and damping coefficient. The acceleration \ddot{u} is measured at the second mass, while the frequency response is considered with respect to the first mass with forcing f . The frequency response is then $h_{21}(\omega) = \frac{\ddot{u}(\omega)}{\tilde{f}(\omega)} = -\omega^2 \frac{\ddot{u}(\omega)}{\tilde{f}(\omega)}$ (compare with Eq. (58)).

in each iteration step is used. Throughout all application examples we use the particle swarm algorithm [57] for the acquisition function maximization.

3.1. Algebraic model: two-degree-of-freedom system

In this example we consider the Bayesian parameter updating problem for a two-degree-of-freedom system. It consists of two lumped masses that are interconnected by a spring and damper element. We assume that the properties of the system can be modelled through a single set of stiffness, damping and mass values. This way, the individual spring, mass and damping coefficients become fully dependent. The system is illustrated in Fig. 5. In the following, we investigate the proposed approach for a varying number of random input parameters:

1. Stiffness k random, mass m and damping c deterministic and equal to their prior mean value; i.e., $\mathbf{X} = k$,
2. Stiffness k , mass m and damping c random; i.e., $\mathbf{X} = [k, m, c]$.

The system parameters' prior distributions are assumed as summarized in Table 1. The damping ratio obtained for

Table 1: Prior distribution assumptions for two-degree-of-freedom model. The parameters are assumed to be independent.

Parameter	Distribution	Mean value $\mu_{(\cdot)}$	Coefficient of variation $\delta_{(\cdot)}$
Stiffness	k Lognormal	$4 \cdot 10^6 \frac{\text{N}}{\text{m}}$	0.2
Mass	m Lognormal	$3 \cdot 10^2 \text{ kg}$	0.2
Damping coefficient	c Lognormal	$2 \cdot 10^3 \frac{\text{N}}{\text{sm}}$	0.2

the mean parameter values follows to $\zeta = \frac{1}{2} \frac{\mu_c}{\sqrt{\mu_k \mu_m}} = 0.0289$. The system can thus be considered as lightly damped. We consider acceleration measurements of the frequency response function at the right mass 2 due to force excitation at the left mass 1 (cf. Fig. 5). The data is considered for the frequency set {10 Hz, 11 Hz, 12 Hz, 28 Hz, 30 Hz, 32 Hz}, thus, $n_O = 6$. The frequency response function is modeled through

$$h_{21}(\omega, \mathbf{x}) = -\omega^2 \frac{\ddot{u}(\omega)}{\tilde{f}(\omega)} = -\omega^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 2k & -k \\ -k & k \end{bmatrix} + i\omega \begin{bmatrix} 2c & -c \\ -c & c \end{bmatrix} - \omega^2 \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad (58)$$

where $\tilde{u}(\omega)$ and $\tilde{f}(\omega)$ are illustrated in Fig. 5. Throughout this section, the polynomial orders are chosen as $m_p = m_q = 2$ and total degree truncation is applied.

3.1.1. One Random Parameter

For illustrative purposes, we first consider a single random parameter. To this end, we choose the stiffness parameter k , synthetically generate measurement data using Eq. (58) for the value $k_{\text{true}} = 0.5\mu_k = 2 \cdot 10^6 \frac{\text{N}}{\text{m}}$ and add samples according to the noise model outlined in Section 2.1 with i.i.d. proper complex errors using $\beta_O = 10^2$.

We illustrate the proposed procedure in Fig. 6. Therein the figures in the upper row (Figs. 6a to 6c) depict the objective function (—) in the first three iteration steps as well as the approximation using the surrogate models (---). The approximation is computed using the mean value estimates of the surrogate model coefficients conditional on the adaptive experimental design in the corresponding iteration step and can thus be considered a first-order approximation to the mean of the approximation of the objective function $\mu_{\tilde{h}}(\mathbf{x})$. The figures in the lower row (Figs. 6d to 6f) depict the evaluation of the acquisition function, namely the expected improvement, for each iteration step. We start from an initial experimental design with $n_{\text{init}} = 3$ samples (\circ) in Figs. 6a and 6d. The maxima of the acquisition function in each iteration step are highlighted (*). The corresponding sample point is added to the experimental design in the subsequent iteration step (*). One can observe how the algorithm generates samples that steadily move towards the global maximum of the log-posterior. The first sample is placed closely to an initial sample point. Subsequently, the approximation of the log-posterior is highly accurate and the third added sample point is found in close proximity to the global optimum. Upon termination, the algorithm recovers the MAP estimate accurately as $k^* = 2.8 \cdot 10^6 \frac{\text{N}}{\text{m}}$, which is identical to the reference solution based on the original model.

3.1.2. Three Random Parameters

In this section, we consider the two-degree-of-freedom system of Section 3.1.1 and include all three random parameters in the MAP estimation. The measurements are synthesized for $k_{\text{true}} = 0.7\mu_k = 2.8 \cdot 10^6 \frac{\text{N}}{\text{m}}$, $m_{\text{true}} = 1.5\mu_m = 450 \text{ kg}$ and $c_{\text{true}} = 1.5\mu_c = 3 \cdot 10^3 \frac{\text{N}}{\text{sm}}$ and samples are added according to the noise model outlined in Section 2.1 with i.i.d. proper complex errors using $\beta_O = 10^2$. Subsequently, we perform MAP-estimation through the proposed active

Table 2: Parameter estimates for two-degree-of-freedom model with three random parameters.

Parameter	Units	Posterior		MAP estimate				
		mean	c.o.v.	reference	global reward	simple reward	fixed design	
Stiffness	k	$10^6 \frac{\text{N}}{\text{m}}$	2.77	0.083	2.7209	2.7209	2.7205	2.7024
Mass	m	10^2 kg	4.45	0.051	4.4066	4.4066	4.4086	4.3827
Damping	c	$10^3 \frac{\text{N}}{\text{sm}}$	2.61	0.111	2.5922	2.5922	2.6062	2.5835

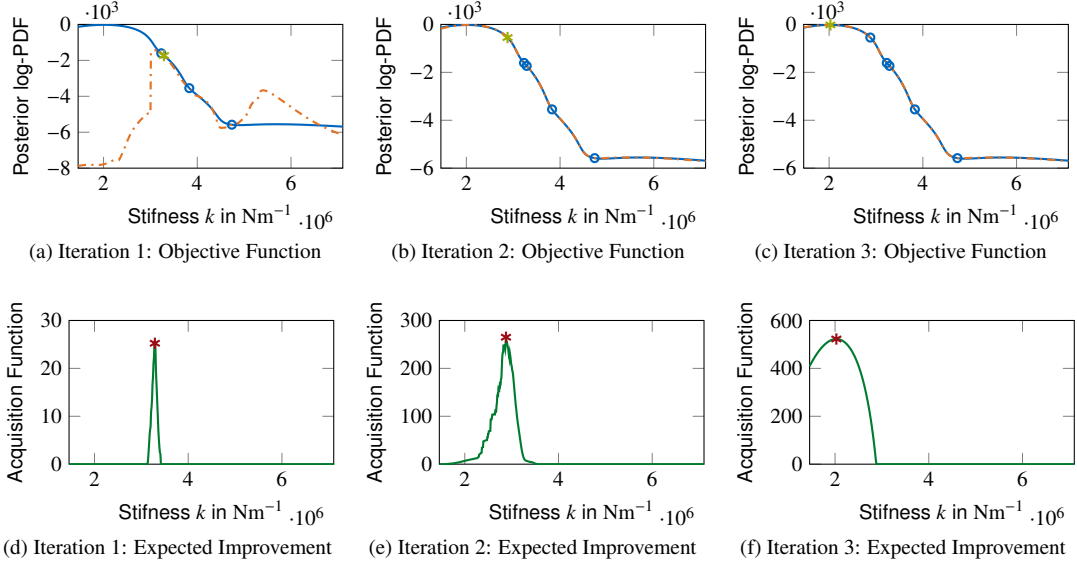


Figure 6: Objective value and expected improvement over three iterations for MAP estimation of two-degree-of-freedom system with random stiffness parameter k . The left column shows the log-posterior $f(x)$ that is computed using the original model (—) and the surrogate model based approximation (— · —). The latter is based on the adaptive experimental design obtained in the corresponding iteration step and the mean model coefficients. In each iteration step, a new sample point ($*$) is added to the experimental design (\circ) based on maximizing the expected improvement criterion. The right column shows the expected improvement criterion (—) and its maximum value ($*$). The identified parameter point will be added to the experimental design in the subsequent iteration step.

learning approach. In detail, starting from an initial experimental design generated through LHS, we sequentially add samples that maximize the expected improvement criterion. Within each iteration step, the simple and global reward estimators of Eqs. (56) and (57) are computed. Furthermore, in each iteration step, we generate an experimental design with sample size n_{tr} equal to the current number of samples in the active learning data set. We then identify the MAP value through maximization of the log-posterior, where we use the surrogate models that are trained on the fixed experimental design sample anew using LHS in each iteration step. The fixed design solution is computed to investigate whether the improvement in the MAP estimate is due to an increasing number of samples in the training set or the design adaptivity. To this end, we set $n_{init} = 15$ and $n_{budget} = 50$ for this example. The analysis is repeated $n_{rep} = 5$ times. The expected improvement is estimated using $n_{\alpha} = 10^2$ samples. The optimal expected improvement is found through particle swarm optimization [57] using the built-in algorithm in MATLAB[®]. The number of particles is chosen as $n_{sw} = 500$. Smaller numbers of particles have been found to yield reasonable results as well. A reference value for the MAP estimate is computed using particle swarm optimization. In addition to the MAP estimation, we generate $n_{post} = 10^5$ samples from the posterior distribution using aBUS-SuS [17]. In order to quantitatively measure the performance of the proposed method, we compute the relative Euclidean distance of the estimates in the standard normal space $\hat{\mathbf{u}}_{MAP}$ with respect to the reference MAP value \mathbf{u}_{MAP} that is obtained through a single optimization run

using the original model, i.e.,

$$\varepsilon_{\text{MAP}} = \frac{\|\hat{\mathbf{u}}_{\text{MAP}} - \mathbf{u}_{\text{MAP}}\|}{\|\mathbf{u}_{\text{MAP}}\|}. \quad (59)$$

We choose to compare the distance in the standard normal space due to the fact that the parameters have significantly different orders of magnitude. The posterior statistics and the median MAP estimates at $n_{\text{tr}} = 50$ are summarized in Tab. 2. It can be observed that the active learning based estimate is accurate to four digits when compared to the reference MAP value. The resulting median relative error for $n_{\text{tr}} = 50$ is $2.8 \cdot 10^{-5}$. In contrast, the fixed design based median estimate differs slightly, resulting in a median error of $2.2 \cdot 10^{-2}$ for $n_{\text{tr}} = 50$.

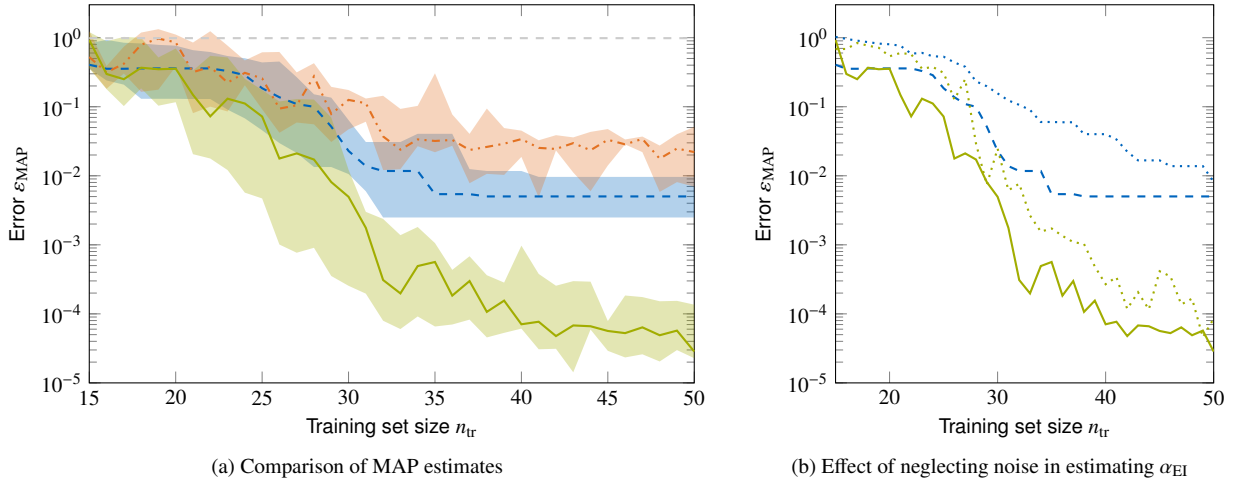


Figure 7: Convergence of the MAP estimate for the three parameter updating problem of the two-degree-of-freedom system. We depict the relative error ε_{MAP} in the standard normal space for the simple reward estimator ($- -$), the global reward estimator ($-$), and a fixed experimental design ($- \cdot -$) based estimator. The relative distance of the prior mean in the standard normal space is also given ($- -$). In Fig. 7b, the corresponding simple ($\cdot \cdot \cdot$) and global ($\cdot \cdot \cdot$) reward estimator without the inclusion of the noise in estimating α_{EI} are compared to their counterparts from Fig. 7a.

In Fig. 7a, we depict the median values of ε_{MAP} as well as min-to-max range over increasing number of experimental design samples. We observe that the global reward MAP estimator ($-$) consistently converges towards the MAP value from $n_{\text{tr}} \approx 20$ on. From around $n_{\text{tr}} \approx 35$, the error of the MAP estimate is in the order of 10^{-4} indicating a very good agreement to the MAP value obtained using the original model. It outperforms the fixed design RPCE based estimator ($- \cdot -$), highlighting the improvement that is gained through the adaptive sampling. The simple reward estimator ($- -$) converges to the reference MAP value as well, however, it plateaus at a higher level in comparison the global reward MAP estimator.

In Fig. 7b, we depict both active learning based MAP estimators, namely the global and simple reward estimator, for the case, where the randomness in the surrogate model error ε_{S} is neglected in the computation of the expected improvement. We depict the corresponding results from the previous study in Fig. 7a for comparison purposes. It can be observed that the estimates also converge to the reference value, however, their accuracy is slightly worse. Since

the inclusion of the error samples comes at almost no cost, we conclude that it is worthwhile to include the surrogate model error ϵ_S in the random process surrogate model.

In total, it can be seen that the proposed method is able to effectively guide the sampling process for efficient MAP estimation. The resulting estimates are more accurate than the ones based on a fixed experimental design, and thus require less model evaluations for a fixed target estimation error. Subsequently, we apply the methodology to the parameter updating problem of a more complex finite element model.

3.2. Finite element model: cross-laminated timber plate

In the following example, we consider the finite element model parameter updating for a cross-laminated timber plate. The plate and the corresponding measurement setup are illustrated in Fig. 8.

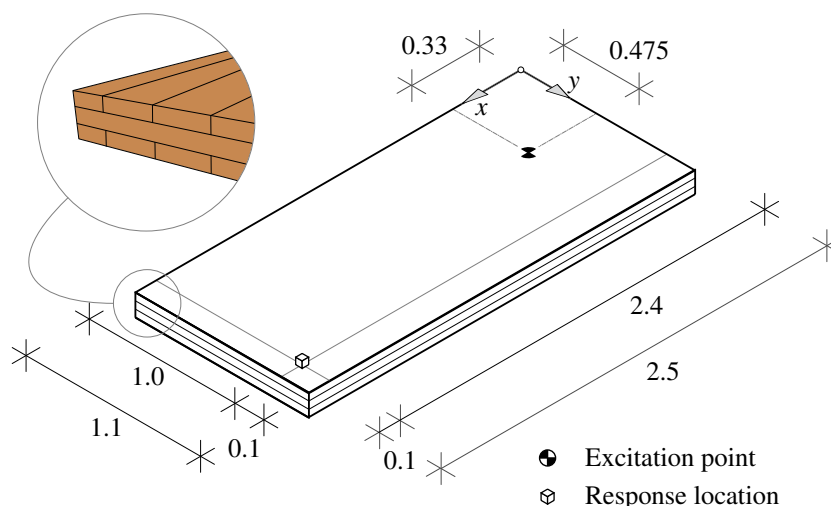


Figure 8: Illustration of cross-laminated timber specimen with excitation and response location.

The considered system is a plate of dimensions $l \times b \times h = 2.5 \text{ m} \times 1.1 \text{ m} \times 0.081 \text{ m}$ that consists of three timber plank layers that are laminated on top of each other in a cross-wise fashion. Forced input excitation measurement data is available for a plate suspended freely from the ceiling (see [58]). A finite element model of the structure is available in the commercial software package ANSYS®. The model uses solid elements based on three-dimensional linear elasticity for orthotropic materials. Details about the model are available in [34].

We choose to perform the updating based on the information from the sensor with location $(x_{\text{sensor}}, y_{\text{sensor}}) = (2.4 \text{ m}, 1.0 \text{ m})$. The frequency transfer function is estimated through the H1-estimator [4] and sub-sampled within the intervals [25 Hz, 35 Hz] and [65 Hz, 75 Hz] with a frequency step size of $\Delta f = 1 \text{ Hz}$. Within both intervals, resonance peaks can be observed. This sub-selection is made to reduce the overall number of observation points.

We model the error model covariance matrices $\Sigma_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon}$ and $\Sigma_{\varphi_\varepsilon \varphi_\varepsilon}$ under the assumption of a parametric correlation function over the frequency space. To this end, we utilize a homoscedastic model with variances σ_w^2 and σ_φ^2 that expresses the errors as random fields over the frequency domain and defines their correlation coefficient function as

$$\rho(f_1, f_2; l_{\text{co},f}, r) = r \exp\left(l_{\text{co},f}^{-1} |f_1 - f_2|\right) + (1 - r) \quad (60)$$

where f_1 and f_2 denote two frequency points, $l_{\text{co},f}$ is the correlation length of the exponential correlation model and r models the split between the frequency dependent correlation and a baseline model correlation that is constant for all data points. ρ is equally applied for both $\Sigma_{\mathbf{w}_\varepsilon \mathbf{w}_\varepsilon}$ and $\Sigma_{\varphi_\varepsilon \varphi_\varepsilon}$. A detailed discussion of the error model can be found in [34]. In general, the error model hyperparameters σ_w^2 , σ_φ^2 , $l_{\text{co},f}$ and r are unknown and need to be chosen based on assumptions or determined from the measurement data. Here, we use the values that were identified in [34], which are $r = 0.8$, $l_{\text{co},f} = 5$ Hz and $\sigma_w = \sigma_\varphi = 0.33$. The system parameters' prior distributions are assumed as summarized in Table 3.

Table 3: Prior distribution assumptions for the cross-laminated timber plate model. The parameters are assumed to be independent.

Parameter		Distribution	Mean value $\mu_{(\cdot)}$	Coefficient of variation $\delta_{(\cdot)}$
Young's modulus	E_x	Lognormal	$1.1 \cdot 10^{10} \frac{\text{N}}{\text{m}^2}$	0.1
Young's modulus	E_y	Lognormal	$0.85 \cdot 3.667 \cdot 10^8 \frac{\text{N}}{\text{m}^2}$	0.1
Shear modulus	G_{xy}	Lognormal	$0.7 \cdot 6.9 \cdot 10^8 \frac{\text{N}}{\text{m}^2}$	0.1
Damping ratio	ζ	Lognormal	$2 \cdot 10^{-2}$	0.3

The same analyses as for the two degree of freedom model are performed. The expected improvement is estimated using $n_\alpha = 50$ samples. The optimal expected improvement is found through particle swarm optimization [57] using the built-in algorithm in MATLAB[®]. The number of particles is chosen as $n_{\text{sw}} = 40$. Smaller numbers of particles have been found to yield reasonable results as well. A reference value for the MAP estimate is computed using particle swarm optimization. Here, we generate $n_{\text{post}} = 10^3$ samples from the posterior distribution using aBUS-SuS. The posterior statistics and the median MAP estimates at $n_{\text{tr}} = 50$ are summarized in Tab. 4.

In Fig. 9, we depict the median values of ε_{MAP} as well as min-to-max range over increasing number of experimental design samples. Again, the active learning based MAP estimators exhibit better convergence performance in comparison to the fixed design based estimator. Towards reaching the budget of $n_{\text{budget}} = 35$, the error ε_{MAP} lies in the order of 10^{-4} in median for the global reward estimator. The fixed design based strategy leads to an estimator with ε_{MAP} in the order of 10^{-2} in median. This highlights the improvement gained through the sequential experimental design strategy in estimating the MAP also for this more complex updating problem.

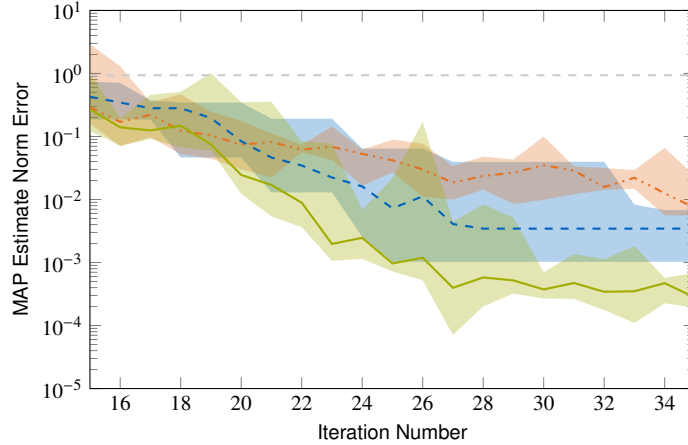


Figure 9: Convergence of the MAP estimate for the four parameter updating problem of the cross-laminated timber plate model using solid elements. We depict the relative error ε_{MAP} in the standard normal space for the simple reward estimator (—), the global reward estimator (—), and a fixed experimental design (---) based estimator. The relative distance of the prior mean in the standard normal space is also given (---).

Table 4: Reference parameter estimates for the finite element model of the CLT plate.

Parameter	Units	Units	Posterior		MAP estimate			
			mean	c.o.v.	reference	global reward	simple reward	fixed design
Young's modulus	E_x	$10^{10} \frac{\text{N}}{\text{m}^2}$	1.1843	0.023	1.1875	1.1875	1.1874	1.1875
Young's modulus	E_y	$10^8 \frac{\text{N}}{\text{m}^2}$	3.1339	0.089	3.0375	3.0375	3.0368	3.0372
Shear modulus	G_{xy}	$10^8 \frac{\text{N}}{\text{m}^2}$	4.5633	0.028	4.5923	4.5924	4.5925	4.5901
Damping ratio	ζ	10^{-2}	3.4512	0.126	3.2888	3.2887	3.2907	3.2993

4. Discussion

While the presented results indicate that the proposed active learning approach for MAP estimation in conjunction with RPCE models can be applied successfully, there remain a number of open questions that could be addressed in future research. Currently, the approach requires substituting each scalar model response in spatial and frequency domain that enters the likelihood by a separate surrogate model. For a large number of observations this causes a substantial computational effort due to the iterative training for each model as well as the individual sampling of the coefficients. A possible remedy for this is to formulate an RPCE model that allows one to surrogate the frequency response over the parameter space and the frequency domain simultaneously. Another approach is to apply a dimensionality reduction step to the discretized model response. Principal component analysis (PCA) has been applied to the frequency response analysis in conjunction with RPCE in [59]. It has been found, however, that the application of PCA to the investigated lightly damped systems in this paper is not straightforward since a large number of principal components is necessary and significant errors might be introduced in the procedure.

Furthermore, the current proposal requires training the Bayesian RPCE model in each active learning iteration step

anew, starting from a full set of basis vectors. Ideally, one would start from the set of basis vectors identified in the previous step and add or delete basis functions as necessary. While this approach is formulated and well-investigated for linear Bayesian models (cf. [60], [61]), an analogous variant for RPCE is not yet available. Without the ability to reactivate basis functions to the active set of basis functions, starting the sequential learning procedure from a truncated set of basis functions might be too restrictive.

The proposed procedure utilizes the well known expected improvement acquisition function. The design and investigation of alternative acquisition functions is an active research question and different proposals could also improve the approach presented in this work. It should be noted, however, that due to the iterative training procedure required for Bayesian RPCE, non-myopic acquisition functions, such as the knowledge gradient acquisition function, might be overly computationally expensive, since they require re-training for each sample in the evaluation of the expected value of the acquisition function. Another interesting extension is motivated by the observation that the acquisition function becomes flat and equal to zero for large parts of the input domain. This is due to the approximation of the expected value through a finite number of samples and the limitation of numerical computations to machine precision. Since a flat objective will exhibit vanishing gradients, this issue complicates the optimization of the expected improvement acquisition function. Similar issues have been reported and discussed in [56]. Therein, the authors propose an alternative formulation of the expected improvement acquisition function, namely the LogEI acquisition, that remedies the identified issue. The proposed modification could potentially also be applied to the methodology outlined in this paper.

Finally, in all examples, knowledge about the error model hyperparameters has been assumed. In common Bayesian updating settings, one often lacks this knowledge. For such cases, robust estimation of the hyperparameters would be necessary. As an extension to the presented methodology, we propose to sequentially update the model and hyperparameters within each active learning step. After having enriched the experimental design, an optimal set of hyperparameters can be obtained through solving a separate optimization problem. The updated set of hyperparameters can then be used in the subsequent iteration step for evaluating the acquisition function. The applicability of this approach and the influence of an initially possibly poor estimate of the hyperparameters on the procedure could be investigated in future research.

5. Conclusion

This work presents a novel maximum a-posteriori estimation technique for Bayesian inverse problems that utilizes frequency response function data to update parameters in linear, time-invariant mechanical models. The proposed method integrates a sequential experimental design strategy based on Bayesian optimization into the maximum a-

posteriori estimation procedure through utilizing rational polynomial chaos expansion surrogate models. Therein, the surrogate models alleviate the computational burden of solving the optimization problem and provide a highly accurate representation for the considered parametric frequency response function models. To train the surrogate model, we extend a previously introduced sparse Bayesian regression approach to determine the surrogate model parameters through introducing an alternative assumption on the distribution of the surrogate modeling error that enables a tractable Laplace approximation of the denominator coefficients. The training set is sequentially enriched through maximizing the expected improvement acquisition function. Since the required expectation operation cannot be carried out in closed-form, we resort to a sampling-based approximation based on samples from the posterior distribution of the surrogate model coefficients. The method is successfully applied to identify the MAP estimates for the parameters of a two-degree-of-freedom system as well as a finite element model of a cross laminated timber plate. In the former case, synthetic data is generated, while for the latter, actual measurement data is used. The active learning procedure is carried out until a prescribed budget of model evaluations is reached. We compare the proposed methodology to a fixed design based RPCE solution and evaluate the relative difference of the MAP estimates to a reference solution found by global optimization. For both models, the MAP estimates identified through the active learning based method show improved accuracy compared to a fixed design approach. This allows to significantly reduce the overall number of model evaluations in the estimation procedure. This reduction is due to the adaptive sequential experimental design strategy, which places samples in the relevant region of the input space and thus leads to better accuracy of the surrogate models in the region of high posterior probability density.

Acknowledgement

The first author thanks Xujia Zhu for his helpful comments and insightful discussions during the first authors stay at the Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich.

Appendix A. Partial Derivatives and Hessian Matrix with Respect to the Denominator Coefficients \mathbf{q}

In this section we derive the partial derivatives and Hessian matrix of the objective function as defined in the optimization problem in Eq. (37). To this end, we define

$$h_q(\mathbf{q}) = \ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H \left(\mathbf{Y}^H \tilde{\mathbf{C}}^{-1} \mathbf{Y} + \Lambda_{\mathbf{q}\mathbf{q}} \right) \mathbf{q}. \quad (\text{A.1})$$

Our goal is to compute the partial derivatives with respect to the conjugate denominator coefficients, i.e., $\frac{\partial h_q(\mathbf{q})}{\partial \mathbf{q}}$. We consider the first part in Eq. (A.1) and derive the partial derivative with respect to the i -th conjugate denominator

coefficient \bar{q}_i . To this end, we write $\ln \det \mathbf{Q}\mathbf{Q}^H = \ln \det \mathbf{Q} + \ln \det \mathbf{Q}^H$ and note that $\frac{\partial}{\partial \bar{q}_i} \ln \det \mathbf{Q} = \mathbf{0}$. Then

$$\frac{\partial}{\partial \bar{q}_i} \ln \det \mathbf{Q}\mathbf{Q}^H = \frac{\partial}{\partial \bar{q}_i} \ln \det \mathbf{Q}^H = \text{tr} \left(\mathbf{Q}^{-H} \frac{\partial \mathbf{Q}^H}{\partial \bar{q}_i} \right) = \text{tr} \left(\mathbf{Q}^{-H} \text{diag} \left((\Psi_q)_{:,i} \right) \right), \quad (\text{A.2})$$

where $(\cdot)_{:,i}$ denotes the i -th column of the matrix (\cdot) . The full vector of partial derivatives can then be conveniently computed through a matrix-vector product as

$$\frac{\partial}{\partial \bar{\mathbf{q}}} \ln \det \mathbf{Q}\mathbf{Q}^H = \Psi_q^T (\Psi_q \bar{\mathbf{q}})^{\circ-1} \quad (\text{A.3})$$

Furthermore, it holds

$$\frac{\partial}{\partial \mathbf{q}} \left[\mathbf{q}^H (\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] = (\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q}. \quad (\text{A.4})$$

Thus,

$$\frac{\partial}{\partial \bar{\mathbf{q}}} \left[\ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H (\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] = \Psi_q^T (\Psi_q \bar{\mathbf{q}})^{\circ-1} - (\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q}. \quad (\text{A.5})$$

Next, based on the result in Eq. (A.5), the Hessian matrix can be expressed as

$$\mathbf{H}_{\mathbf{q}\mathbf{q}} = \frac{\partial}{\partial \bar{\mathbf{q}}} \left(\frac{\partial}{\partial \mathbf{q}} \left[\ln \det \mathbf{Q}\mathbf{Q}^H - \mathbf{q}^H (\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}) \mathbf{q} \right] \right)^T = -(\Upsilon^H \tilde{\mathbf{C}}^{-1} \Upsilon + \Lambda_{\mathbf{q}\mathbf{q}}). \quad (\text{A.6})$$

Here, we have made use of the fact that the first summand in Eq. (A.5) does not depend on \mathbf{q} , but only on $\bar{\mathbf{q}}$.

Appendix B. Partial derivatives with respect to the hyperparameters α_p

In this section we derive the partial derivatives of the objective function as defined in the optimization problem in Eq. (41) with respect to the numerator coefficient precisions. To this end, we define

$$h_{\alpha_p}(\alpha_p) = \ln \det \Sigma_{\mathbf{p}\mathbf{p}|\mathbf{m}} + \ln \det \Lambda_{\mathbf{p}\mathbf{p}} - \beta_S \mathbf{m}^H \bar{\mathbf{Q}}^* (\mathbf{Q}^* \mathbf{m} - \Psi_p \tilde{\boldsymbol{\mu}}_{\mathbf{p}|\mathbf{m}}) - \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}}). \quad (\text{B.1})$$

We observe that the objective function in Eq. (B.1) possesses a similar structure compared to the one presented in [38] for the standard residual formulation. While in [38], the denominator coefficients enter in an augmented design matrix, here, they can be understood as augmenting the data vector. However, the dependency of the expressions on the precision parameters α_p is the same, so the results can be transferred. In analogy to the expressions in Appendix

D in [38], we obtain

$$\frac{\partial}{\partial \alpha_{p,i}} \left[\ln \det \mathbf{\Sigma}_{\mathbf{pp|m}} + \ln \det \mathbf{\Lambda}_{\mathbf{pp}} - \beta_S \mathbf{m}^H \mathbf{Q}^* (\mathbf{Q}^* \mathbf{m} - \mathbf{\Psi}_p \boldsymbol{\mu}_{\mathbf{p|m}, \mathbf{q}^*}) \right] = - \left[\mathbf{\Sigma}_{\mathbf{pp|m}, \mathbf{q}^*} \right]_{ii} + \alpha_{p,i}^{-1} - \left| \left[\boldsymbol{\mu}_{\mathbf{p|m}, \mathbf{q}^*} \right]_i \right|^2 \quad (\text{B.2})$$

Additionally, the derivative with respect to the last term in Eq. (B.1) is required. We find

$$\frac{\partial \ln \det (-\mathbf{H}_{\mathbf{qq}})}{\partial \alpha_{p,i}} = \text{tr} \left((-\mathbf{H}_{\mathbf{qq}})^{-1} \frac{\partial (-\mathbf{H}_{\mathbf{qq}})}{\partial \alpha_{p,i}} \right). \quad (\text{B.3})$$

It holds

$$\frac{\partial (-\mathbf{H}_{\mathbf{qq}})}{\partial \alpha_{p,i}} = \mathbf{\Upsilon}^H \frac{\partial \mathbf{C}^{-1}}{\partial \alpha_{p,i}} \mathbf{\Upsilon}, \quad (\text{B.4})$$

and through application of the following identity,

$$\mathbf{C}^{-1} = \beta_S - \beta_S^2 \mathbf{\Psi}_p \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbf{\Psi}_p^T, \quad (\text{B.5})$$

we can state

$$\frac{\partial \mathbf{C}^{-1}}{\partial \alpha_{p,i}} = -\beta_S^2 \mathbf{\Psi}_p \frac{\partial \mathbf{\Sigma}_{\mathbf{pp|m}}}{\partial \alpha_{p,i}} \mathbf{\Psi}_p^T = \beta_S^2 \mathbf{\Psi}_p \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbb{I}_{ii} \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbf{\Psi}_p^T, \quad (\text{B.6})$$

since $\frac{\partial \mathbf{\Sigma}_{\mathbf{pp|m}}}{\partial \alpha_{p,i}} = -\mathbf{\Sigma}_{\mathbf{pp|m}} \frac{\partial \mathbf{\Sigma}_{\mathbf{pp|m}}^{-1}}{\partial \alpha_{p,i}} \mathbf{\Sigma}_{\mathbf{pp|m}}$. \mathbb{I}_{ij} denotes the single-entry matrix with unit-entry at position (i, j) . Thereby

$$\frac{\partial \ln \det (-\mathbf{H}_{\mathbf{qq}})}{\partial \alpha_{p,i}} = \text{tr} \left(\beta_S^2 (-\mathbf{H}_{\mathbf{qq}})^{-1} \mathbf{\Upsilon}^H \mathbf{\Psi}_p \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbb{I}_{ii} \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbf{\Psi}_p^T \mathbf{\Upsilon} \right). \quad (\text{B.7})$$

Through the cyclic permutation property of the trace operator, we find

$$\begin{aligned} \frac{\partial \ln \det (-\mathbf{H}_{\mathbf{qq}})}{\partial \alpha_{p,i}} &= \text{tr} \left(\mathbb{I}_{ii} (\beta_S \mathbf{\Sigma}_{\mathbf{pp|m}} \mathbf{\Psi}_p^T \mathbf{\Upsilon}) (-\mathbf{H}_{\mathbf{qq}})^{-1} (\beta_S \mathbf{\Upsilon}^H \mathbf{\Psi}_p \mathbf{\Sigma}_{\mathbf{pp|m}}) \right) \\ &= \left[\mathbf{\Delta}^H (-\mathbf{H}_{\mathbf{qq}})^{-1} \mathbf{\Delta} \right]_{ii}. \end{aligned} \quad (\text{B.8})$$

where $\mathbf{\Delta} \in \mathbb{C}^{n_q \times n_p}$ is defined as

$$\mathbf{\Delta} = \beta_S \mathbf{\Upsilon}^H \mathbf{\Psi}_p \mathbf{\Sigma}_{\mathbf{pp|m}} = \mathbf{\Upsilon}^H \mathbf{C}^{-1} \mathbf{\Psi}_p \mathbf{\Lambda}_{\mathbf{pp}}^{-1}. \quad (\text{B.9})$$

Combining the results in Eq. (C.2) with Eq. (C.3), we obtain

$$\frac{\partial h_{\alpha_p}(\alpha_p)}{\partial \alpha_{p,i}} = - \left[\mathbf{\Sigma}_{\mathbf{pp|m}, \mathbf{q}^*} \right]_{ii} + \alpha_{p,i}^{-1} - \left| \left[\boldsymbol{\mu}_{\mathbf{p|m}, \mathbf{q}^*} \right]_i \right|^2 - \left[\mathbf{\Delta}^H (-\mathbf{H}_{\mathbf{qq}})^{-1} \mathbf{\Delta} \right]_{ii}. \quad (\text{B.10})$$

Setting the derivative to zero, i.e., $\frac{\partial h_{\alpha_p}(\alpha_p)}{\partial \alpha_{p,i}} = 0$, we can find the optimal value in closed form:

$$\alpha_{p,i} = \frac{1}{\left[\Sigma_{\mathbf{p}\mathbf{p}|\mathbf{m},\mathbf{q}} \right]_{ii} + \left| \left[\mu_{\mathbf{p}|\mathbf{m},\mathbf{q}^*} \right]_i \right|^2 + \left[\Delta^{\mathbf{H}} (-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \Delta \right]_{ii}}. \quad (\text{B.11})$$

Alternatively, we can define a fixed-point update as

$$\underbrace{1 - \alpha_{p,i} \left[\Sigma_{\mathbf{p}\mathbf{p}|\mathbf{m},\mathbf{q}^*} \right]_{ii}}_{=: \gamma_{p,i}} - \alpha_{p,i} \left| \left[\mu_{\mathbf{p}|\mathbf{m},\mathbf{q}^*} \right]_i \right|^2 - \alpha_{p,i} \left[\Delta^{\mathbf{H}} (-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \Delta \right]_{ii} = 0 \quad (\text{B.12})$$

$$\rightarrow \alpha_{p,i} = \frac{\gamma_{p,i}}{\left| \left[\mu_{\mathbf{p}|\mathbf{m},\mathbf{q}^*} \right]_i \right|^2 + \left[\Delta^{\mathbf{H}} (-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \Delta \right]_{ii}}. \quad (\text{B.13})$$

Appendix C. Partial derivatives with respect to the hyperparameters α_q

In this section we derive the partial derivatives of the objective function as defined in the optimization problem in Eq. (41) with respect to the denominator coefficient precisions. To this end, we define

$$h_{\alpha_q}(\alpha_q) = \ln \det \Lambda_{\mathbf{q}\mathbf{q}} - \mathbf{q}^{*\mathbf{H}} \Lambda_{\mathbf{q}\mathbf{q}} \mathbf{q}^* - \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}}). \quad (\text{C.1})$$

We start by computing

$$\frac{\partial}{\partial \alpha_{q,i}} \left[\ln \det \Lambda_{\mathbf{q}\mathbf{q}} - \mathbf{q}^{*\mathbf{H}} \Lambda_{\mathbf{q}\mathbf{q}} \mathbf{q}^* \right] = \frac{\partial}{\partial \alpha_{q,i}} \left[\sum_{j=0}^{n_q-1} \ln \alpha_{q,j} - \alpha_{q,j} |q_j^*|^2 \right] = \alpha_{q,i}^{-1} - |q_i^*|^2. \quad (\text{C.2})$$

Furthermore,

$$\frac{\partial \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \alpha_{q,i}} = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \frac{\partial (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \alpha_{q,i}} \right) = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \mathbb{I}_{ii} \right) = \left[(-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \right]_{ii}, \quad (\text{C.3})$$

since

$$\frac{\partial (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \alpha_{q,i}} = \frac{\partial \Lambda_{\mathbf{q}\mathbf{q}}}{\partial \alpha_{q,i}} = \mathbb{I}_{ii}. \quad (\text{C.4})$$

Combining the results in Eqs. (C.2) and (C.3), we obtain

$$\frac{\partial h_{\alpha_q}(\alpha_q)}{\partial \alpha_{q,i}} = \alpha_{q,i}^{-1} - |q_i^*|^2 - \left[(-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \right]_{ii}. \quad (\text{C.5})$$

Setting the derivative to zero, i.e., $\frac{\partial h_{\alpha_q}(\alpha_q)}{\partial \alpha_{q,i}} = 0$, we can find the optimal value in closed form:

$$\alpha_{q,i} = \frac{1}{|q_i^*|^2 + \left[(-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1}\right]_{ii}}. \quad (\text{C.6})$$

Alternatively, we can define a fix-point update as

$$\underbrace{1 - \alpha_{q,i} \left[(-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1}\right]_{ii}}_{=:\gamma_{q,i}} - \alpha_{q,i} |q_i^*|^2 = 0 \quad (\text{C.7})$$

$$\rightarrow \alpha_{q,i} = \frac{\gamma_{q,i}}{|q_i^*|^2}. \quad (\text{C.8})$$

Appendix D. Partial derivatives with respect to the hyperparameter β_S

In this section we derive the partial derivatives of the objective function as defined in the optimization problem in Eq. (41) with respect to the error model precision β_S . To this end, we define

$$h_{\beta_S}(\beta_S) = n_{\text{tr}} \ln \beta_S + \ln \det \Sigma_{\text{pp|m}} - \beta_S \mathbf{m}^H \overline{\mathbf{Q}^*} (\mathbf{Q}^* \mathbf{m} - \Psi_p \mu_{\text{p|m}}) - \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}}). \quad (\text{D.1})$$

Analogously to the argument made in Appendix B, we observe that we can transfer results from Appendix F in [38] and find

$$\frac{\partial}{\partial \beta_S} \left[n_{\text{tr}} \ln \beta_S + \ln \det \Sigma_{\text{pp|m}} - \beta_S \mathbf{m}^H \overline{\mathbf{Q}^*} (\mathbf{Q}^* \mathbf{m} - \Psi_p \mu_{\text{p|m}}) \right] = \frac{n_{\text{tr}}}{\beta_S} - \text{tr} \left(\Sigma_{\text{pp|m}} \Psi_p^T \Psi_p \right) - \|\mathbf{Q}\mathbf{m} - \Psi_p \mu_{\text{p|m},\mathbf{q}^*}\|^2. \quad (\text{D.2})$$

The remaining element is the computation of the derivative of the log-determinant of the Hessian matrix in Eq. (D.1).

It holds

$$\frac{\partial \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \beta_S} = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \frac{\partial (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \beta_S} \right) = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \Upsilon^H \frac{\partial \mathbf{C}^{-1}}{\partial \beta_S} \Upsilon \right), \quad (\text{D.3})$$

Under consideration of Eq. (30), we find

$$\frac{\partial \mathbf{C}^{-1}}{\partial \beta_S} = -\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \beta_S} \mathbf{C}^{-1} = -\beta_S^{-2} \mathbf{C}^{-1} \mathbf{C}^{-1}. \quad (\text{D.4})$$

Thus,

$$\frac{\partial \ln \det (-\mathbf{H}_{\mathbf{q}\mathbf{q}})}{\partial \beta_S} = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \Upsilon^H \mathbf{C}^{-1} \frac{\partial \Sigma_{\tilde{\epsilon}\tilde{\epsilon}}}{\partial \theta} \mathbf{C}^{-1} \Upsilon \right) = \text{tr} \left((-\mathbf{H}_{\mathbf{q}\mathbf{q}})^{-1} \mathbf{O}^H \mathbf{O} \right), \quad (\text{D.5})$$

where

$$\mathbf{O} = \beta_S^{-1} \mathbf{C}^{-1} \Upsilon. \quad (\text{D.6})$$

For numerical purposes it can be beneficial to rewrite the inverse \mathbf{C}^{-1} using the Woodbury inversion theorem. Then, Eq. (D.6) reads

$$\mathbf{O} = \beta_S^{-1} \left(\beta_S \mathbf{I}_{n_{\text{tr}}} - \beta_S^2 \Psi_p \Sigma_{\text{pp|m}} \Psi_p^T \right) \Upsilon = \left(\mathbf{I}_{n_{\text{tr}}} - \beta_S \Psi_p \Sigma_{\text{pp|m}} \Psi_p^T \right) \Upsilon. \quad (\text{D.7})$$

Combining the results in Eqs. (D.5), (D.2) and (D.7), we obtain

$$\frac{\partial h_{\beta_S}}{\partial \beta_S} = \frac{n_{\text{tr}}}{\beta_S} - \text{tr} \left(\Sigma_{\text{pp|m}} \Psi_p^T \Psi_p \right) - \left\| \mathbf{Qm} - \Psi_p \mu_{\text{p|m}} \right\|^2 - \text{tr} \left(\left(-\mathbf{H}_{\text{qq}} \right)^{-1} \mathbf{O}^H \mathbf{O} \right). \quad (\text{D.8})$$

Setting the derivative to zero, we can find the optimal value analytically,

$$\beta_S = \frac{n_{\text{tr}}}{\left\| \mathbf{Qm} - \Psi_p \mu_{\text{p|m,q}^*} \right\|^2 + \text{tr} \left(\Sigma_{\text{pp|m}} \Psi_p^T \Psi_p \right) + \text{tr} \left(\left(-\mathbf{H}_{\text{qq}} \right)^{-1} \mathbf{O}^H \mathbf{O} \right)}. \quad (\text{D.9})$$

Alternatively, we can define the following fixed-point update as follows

$$\beta_S = \frac{n_{\mathcal{M}} - \sum_{i=0}^{n_p-1} \gamma_{p,i}}{\left\| \mathbf{Qm} - \Psi_p \mu_{\text{p|m,q}^*} \right\|^2 + \text{tr} \left(\left(-\mathbf{H}_{\text{qq}} \right)^{-1} \mathbf{O}^H \mathbf{O} \right)}. \quad (\text{D.10})$$

References

- [1] K. J. Åström, P. Eykhoff, System identification—A survey, *Automatica* 7 (2) (1971) 123–162. doi:10.1016/0005-1098(71)90059-8.
- [2] T. Soederstroem, P. Stoica, *System Identification*, Prentice Hall International Series in Systems and Control Engineering, Prentice Hall, New York and London, 1989.
- [3] J. E. Mottershead, M. I. Friswell, Model updating in structural dynamics: A survey, *Journal of Sound and Vibration* 167 (2) (1993) 347–375. doi:10.1006/jsvi.1993.1340.
- [4] R. Pintelon, J. Schoukens, *System identification: a frequency domain approach*, John Wiley & Sons, 2012.
- [5] M. Friswell, J. E. Mottershead, *Finite element model updating in structural dynamics*, Vol. 38, Springer Science & Business Media, 2013.
- [6] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, D. B. Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 2013.
- [7] J. L. Beck, Bayesian system identification based on probability logic, *Structural Control and Health Monitoring* 17 (7) (2010) 825–847. doi:10.1002/stc.424.
- [8] D. J. MacKay, D. J. Mac Kay, *Information theory, inference and learning algorithms*, Cambridge University Press, 2003.
- [9] W. R. Gilks, S. Richardson, D. Spiegelhalter, *Markov chain Monte Carlo in practice*, Chapman and Hall/CRC, 1995.
- [10] S. H. Cheung, S. Bansal, A new gibbs sampling based algorithm for bayesian model updating with incomplete complex modal data, *Mechanical Systems and Signal Processing* 92 (2017) 156–172.
- [11] M. Muto, J. L. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, *Journal of Vibration and Control* 14 (1-2) (2008) 7–34. doi:10.1177/1077546307079400.
- [12] E. Zhang, P. Feissel, J. Antoni, Bayesian model updating with consideration of modeling error, *European Journal of Computational Mechanics* 19 (1-3) (2010) 255–266. doi:10.3166/EJCM.19.255-266.
- [13] E. L. Zhang, P. Feissel, J. Antoni, A comprehensive bayesian approach for model updating and quantification of modeling errors, *Probabilistic Engineering Mechanics* 26 (4) (2011) 550–560. doi:10.1016/j.probenmech.2011.07.001.
- [14] E. Zhang, J. D. Chazot, J. Antoni, M. Hamdi, Bayesian characterization of young’s modulus of viscoelastic materials in laminated structures, *Journal of Sound and Vibration* 332 (16) (2013) 3654–3666. doi:10.1016/j.jsv.2013.02.032.
- [15] D. Straub, I. Papaioannou, Bayesian updating with structural reliability methods, *Journal of Engineering Mechanics* 141 (3) (2015) 04014134. doi:10.1061/(ASCE)EM.1943-7889.0000839.
- [16] F. DiazDelaO, A. Garbuno-Inigo, S. Au, I. Yoshida, Bayesian updating and model class selection with subset simulation, *Computer Methods in Applied Mechanics and Engineering* 317 (2017) 1102–1121.
- [17] W. Betz, I. Papaioannou, J. L. Beck, D. Straub, Bayesian inference with subset simulation: Strategies and improvements, *Computer Methods in Applied Mechanics and Engineering* 331 (2018) 72–93. doi:10.1016/j.cma.2017.11.021.
- [18] R. G. Ghanem, P. D. Spanos, *Stochastic finite element method: Response statistics*, in: *Stochastic Finite Elements: A Spectral Approach*, Springer, 1991, pp. 101–119.
- [19] D. Xiu, G. E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM Journal on Scientific Computing* 24 (2) (2002) 619–644.
- [20] F. Yamazaki, M. Shinozuka, G. Dasgupta, Neumann expansion for stochastic finite element analysis, *Journal of Engineering Mechanics* 114 (8) (1988) 1335–1354.
- [21] M. Papadrakakis, V. Papadopoulos, N. D. Lagaros, Structural reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation, *Computer Methods in Applied Mechanics and Engineering* 136 (1-2) (1996) 145–163.
- [22] C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*, 3rd Edition, Adaptive computation and machine learning, MIT Press, Cambridge, Mass., 2008.

- [23] M. D. McKay, R. J. Beckman, W. J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (2) (1979) 239–245.
- [24] A. B. Owen, Quasi-Monte Carlo sampling, *Monte Carlo Ray Tracing: Siggraph 1* (2003) 69–88.
- [25] R. Garnett, *Bayesian Optimization*, Cambridge University Press, 2023.
- [26] P. I. Frazier, A tutorial on bayesian optimization. doi:10.48550/arXiv.1807.02811.
- [27] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. de Freitas, Taking the human out of the loop: A review of bayesian optimization, *Proceedings of the IEEE* 104 (1) (2016) 148–175. doi:10.1109/jproc.2015.2494218.
- [28] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, S. Venkatesh, Bayesian optimization for adaptive experimental design: A review, *IEEE Access* 8 (2020) 13937–13948. doi:10.1109/ACCESS.2020.2966228.
- [29] X. Wang, Y. Jin, S. Schmitt, M. Olhofer, Recent advances in bayesian optimization, *ACM Computing Surveys* 55 (13s) (2023) 1–36. doi:10.1145/3582078.
- [30] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization* 13 (4) (1998) 455–492. doi:10.1023/A:1008306431147.
URL <https://link.springer.com/article/10.1023/a:1008306431147>
- [31] S.-S. Jin, H.-J. Jung, Sequential surrogate modeling for efficient finite element model updating, *Computers & Structures* 168 (2016) 30–45. doi:10.1016/j.compstruc.2016.02.005.
- [32] P. Pandita, P. Tsilifis, N. M. Awalgaonkar, I. Bilionis, J. Panchal, Surrogate-based sequential bayesian experimental design using non-stationary gaussian processes, *Computer Methods in Applied Mechanics and Engineering* 385 (2021) 114007. doi:10.1016/j.cma.2021.114007.
- [33] H. Kapadia, L. Feng, P. Benner, Active-learning-driven surrogate modeling for efficient simulation of parametric nonlinear systems, *Computer Methods in Applied Mechanics and Engineering* 419 (2024) 116657. doi:10.1016/j.cma.2023.116657.
- [34] F. Schneider, I. Papaioannou, D. Straub, C. Winter, G. Müller, Bayesian parameter updating in linear structural dynamics with frequency transformed data using rational surrogate models, *Mechanical Systems and Signal Processing* 166 (2022) 108407. doi:10.1016/j.ymsp.2021.108407.
URL <https://www.sciencedirect.com/science/article/pii/S0888327021007573>
- [35] C. Mares, B. Dratz, J. Mottershead, M. Friswell, Model updating using Bayesian estimation, in: *International Conference on Noise and Vibration Engineering, ISMA2006*, Katholieke Universiteit Leuven, 2006, pp. 18–20.
- [36] J. Bect, N. Georg, U. Römer, S. Schöps, Rational kernel-based interpolation for complex-valued frequency response functions.
URL <http://arxiv.org/pdf/2307.13484v2>
- [37] P.-R. Wagner, S. Marelli, B. Sudret, Bayesian model inversion using stochastic spectral embedding, *Journal of Computational Physics* 436 (2021) 110141.
- [38] F. Schneider, I. Papaioannou, G. Müller, Sparse bayesian learning for complex-valued rational approximations, *International Journal for Numerical Methods in Engineering* 124 (8) (2023) 1721–1747. doi:10.1002/nme.7182.
- [39] E. Jacquelin, S. Adhikari, J.-J. Sinou, M. I. Friswell, Polynomial chaos expansion and steady-state response of a class of random dynamical systems, *Journal of Engineering Mechanics* 141 (4) (2015) 04014145. doi:10.1061/(ASCE)EM.1943-7889.0000856.
- [40] E. Jacquelin, S. Adhikari, J.-J. Sinou, M. I. Friswell, Polynomial chaos expansion in structural dynamics: Accelerating the convergence of the first two statistical moment sequences, *Journal of Sound and Vibration* 356 (2015) 144–154. doi:10.1016/j.jsv.2015.06.039.
- [41] E. Jacquelin, S. Adhikari, M. I. Friswell, J.-J. Sinou, Role of roots of orthogonal polynomials in the dynamic response of stochastic systems, *Journal of Engineering Mechanics* 142 (8) (2016) 06016004. doi:10.1061/(ASCE)EM.1943-7889.0001102.

- [42] Jacquelin E., Dessombz O., Sinou J.-J., Adhikari S., Friswell M. I., Polynomial chaos–based extended padé expansion in structural dynamics, *International Journal for Numerical Methods in Engineering* 111 (12) (2016) 1170–1191. doi:10.1002/nme.5497.
- [43] F. Schneider, I. Papaioannou, M. Ehre, D. Straub, Polynomial chaos based rational approximation in linear structural dynamics with parameter uncertainties, *Computers & Structures* 233 (2020) 106223. doi:https://doi.org/10.1016/j.compstruc.2020.106223.
URL <http://www.sciencedirect.com/science/article/pii/S0045794920300262>
- [44] M. E. Tipping, Sparse bayesian learning and the relevance vector machine, *Journal of machine learning research* 1 (Jun) (2001) 211–244.
- [45] M. Géradin, D. Rixen, *Mechanical vibrations: Theory and application to structural dynamics*, 3rd Edition, Wiley, Hoboken, New Jersey, 1997.
- [46] M. Rosenblatt, Remarks on a multivariate transformation, *The Annals of Mathematical Statistics* 23 (3) (1952) 470–472.
- [47] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics* 230 (6) (2011) 2345–2367. doi:10.1016/j.jcp.2010.12.021.
- [48] D. Cox, J. Little, D. O’shea, *Ideals, varieties, and algorithms*, Vol. 3, Springer, 2007.
- [49] T. Chantrasmı, A. Doostan, G. Iaccarino, Padé–legendre approximants for uncertainty analysis with discontinuous response surfaces, *Journal of Computational Physics* 228 (19) (2009) 7159–7180. doi:10.1016/j.jcp.2009.06.024.
- [50] P. J. Schreier, L. L. Scharf, *Statistical signal processing of complex-valued data: The theory of improper and noncircular signals*, Cambridge University Press, Cambridge, 2010. doi:10.1017/CBO9780511815911.
URL <https://doi.org/10.1017/CBO9780511815911>
- [51] K. Kreutz-Delgado, The complex gradient operator and the cr-calculus, arXiv preprint arXiv:0906.4835.
- [52] L. Sorber, M. van Barel, L. D. Lathauwer, *Complex optimization toolbox v1.03* (2013).
URL <http://esat.kuleuven.be/stadius/cot/>
- [53] L. Sorber, M. van Barel, L. D. Lathauwer, Unconstrained optimization of real functions in complex variables, *SIAM Journal on Optimization* 22 (3) (2012) 879–898. doi:10.1137/110832124.
- [54] J. Ching, Y.-C. Chen, Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging, *Journal of Engineering Mechanics* 133 (7) (2007) 816–832. doi:10.1061/(ASCE)0733-9399(2007)133:7(816).
- [55] W. Betz, I. Papaioannou, D. Straub, Transitional markov chain monte carlo: Observations and improvements, *Journal of Engineering Mechanics* 142 (5). doi:10.1061/(ASCE)EM.1943-7889.0001066.
- [56] S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected improvements to expected improvement for bayesian optimization, in: A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), *Advances in Neural Information Processing Systems*, Vol. 36, Curran Associates, Inc., 2023, pp. 20577–20612.
URL https://proceedings.neurips.cc/paper_files/paper/2023/file/419f72cbd568ad62183f8132a3605a2a-Paper-Confer
- [57] J. Kennedy, R. Eberhart, Particle swarm optimization, in: *IEEE International Conference on Neural Networks*, 1995. Proceedings, IEEE / Institute of Electrical and Electronics Engineers Incorporated, 1995, pp. 1942–1948. doi:10.1109/ICNN.1995.488968.
- [58] S. Mecking, T. Kruse, C. Winter, U. Schanda, *Schlussbericht: Vibroakustik im Planungsprozess für Holzbauten: Teilprojekt 3: Parameterentwicklung und SEA-Modellierung: Research Report*, 2017.
- [59] P. Manfredi, S. Grivet-Talocia, Fast stochastic surrogate modeling via rational polynomial chaos expansions and principal component analysis, *IEEE Access* 9 (2021) 102732–102745. doi:10.1109/ACCESS.2021.3097543.
- [60] M. E. Tipping, A. C. Faul, Fast marginal likelihood maximisation for sparse bayesian models, in: *International workshop on artificial intelligence and statistics*, PMLR, 2003, pp. 276–283.
- [61] S. D. Babacan, R. Molina, A. K. Katsaggelos, Bayesian compressive sensing using laplace priors, *IEEE transactions on image processing* : a

publication of the IEEE Signal Processing Society 19 (1) (2010) 53–63. doi:10.1109/TIP.2009.2032894.