



Identification and sampling of Bayesian posteriors of high-dimensional symmetric positive-definite matrices for data-driven updating of computational models

Maarten Arnst, Christian Soize

► To cite this version:

Maarten Arnst, Christian Soize. Identification and sampling of Bayesian posteriors of high-dimensional symmetric positive-definite matrices for data-driven updating of computational models. *Computer Methods in Applied Mechanics and Engineering*, 2019, 352, pp.300-323. 10.1016/j.cma.2019.04.025 . hal-02146333

HAL Id: hal-02146333

<https://hal.science/hal-02146333>

Submitted on 3 Jun 2019

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.

Identification and sampling of Bayesian posteriors of high-dimensional symmetric positive-definite matrices for data-driven updating of computational models

M. Arnst^{a,*}, C. Soize^b

^aUniversité de Liège, Aéronautique et Mécanique, Quartier Polytech 1, allée de la Découverte 9, 4000 Liège, Belgium

^bUniversité Paris-Est Marne-la-Vallée, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France

Abstract

We propose a probabilistic methodology for data-driven updating of non-Gaussian high-dimensional symmetric positive-definite matrices involved in computational models. We cast the data-driven updating as a Bayesian identification of the symmetric positive-definite matrices. The posterior thus obtained exhibits several hyperparameters that control the dispersion of the prior and the weight of the weighted distance that represents the model-data misfit in the likelihood function. Using an identification criterion that quantifies the agreement between the predictions and the data, we identify these hyperparameters so as to obtain not only improved predictions but also a probabilistic representation of model uncertainties. The numerical implementation of the Bayesian inversion by using a Markov chain Monte Carlo (MCMC) method is computationally challenging because the support of the posterior is restricted to a set of symmetric positive-definite matrices and the dimensionality of the problem grows with the square of the matrix dimension and hence can be high. We thus use a transformation of measure to set up the Markov chain in terms of real-valued state variables whose distribution is Gaussian under the prior but non-Gaussian under the posterior. This transformation of measure allows us to sample the posterior using an Itô-SDE-based MCMC method that inherits computational efficiency in high dimension from leveraging the gradient of the posterior. We apply

*Corresponding author

Email address: `maarten.arnst@uliege.be` (M. Arnst)

this methodology to a problem of a data-driven updating of a reduced-order model in computational structural dynamics.

Keywords: data-driven modeling, Bayesian inversion, symmetric positive-definite matrix, identification criterion, nonparametric probabilistic model, Markov Chain Monte Carlo

1. Introduction

We propose a probabilistic methodology for data-driven updating of non-Gaussian high-dimensional symmetric positive-definite matrices involved in computational models. For instance, the computational model could be a reduced-order model parameterized by symmetric positive-definite reduced matrices, or it could involve constitutive models with material tensors represented by symmetric positive-definite matrices, among other possibilities. In this methodology, we cast the updating as a Bayesian identification of the symmetric positive-definite matrices. Our aim is not only to improve the accuracy of the predictions but also to learn from the data about the impact of parametric uncertainties and modeling errors on the predictions.

Whereas an inverse problem formulated within a deterministic framework typically takes the form of an optimization problem, the Bayesian approach to inverse problems takes the form of an application of the Bayes formula to update a prior probability distribution by means of a likelihood function to obtain a posterior probability distribution [1, 2, 3]. The numerical implementation of an inverse problem within this Bayesian approach typically involves a Markov chain Monte Carlo (MCMC) method to sample from the posterior, and descriptors of the posterior are then estimated from the samples thus obtained by means of methods from mathematical statistics [4].

Casting the data-driven updating as a Bayesian identification of the symmetric positive-definite matrices raises challenges, both from the point of view of the formulation of the Bayesian inversion as from the point of view of the numerical implementation of the Bayesian inversion. First, the prior must be a matrix-variate probability distribution whose support is limited to a set of symmetric positive-definite matrices. The sets of symmetric positive-definite matrices are not linear spaces (vector spaces) and this probability distribution cannot be Gaussian. Second, the prior and the likelihood function must be defined—and their hyperparameters identified—in a manner that is consistent with our aim of obtaining not only improved predictions

but also a probabilistic representation of the model uncertainties. Third, an MCMC method is required that is able to explore efficiently the set of symmetric positive-definite matrices in accordance with the posterior. The MCMC method must ensure the symmetry and the positive definiteness of the matrix samples. And the dimensionality of the problem grows with the square of the matrix dimension and hence can be high, so that conventional MCMC methods can be inefficient. For example, Metropolis–Hastings variants of MCMC methods require a kernel to accept and reject proposals from, and choosing an adequate kernel can be difficult in high dimension.

We address these challenges as follows. First, we select the prior from the family of probability distributions for symmetric positive-definite random matrices introduced with a constructive approach based on the maximum-entropy principle in [5, 6]. This family of maximum-entropy probability distributions lends itself well to representing uncertainty in symmetric positive-definite matrices of computational models [5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Second, we define the likelihood function as an exponential of minus a weighted distance that represents the model-data misfit. The posterior thus obtained exhibits several hyperparameters that control the dispersion of the prior and the weight of the weighted distance that represents the model-data misfit involved in the likelihood function. Using an identification criterion that quantifies the agreement between the predictions and the data, we identify these hyperparameters so as to obtain not only improved predictions but also a probabilistic representation of the model uncertainties. Third, we introduce a transformation of measure from the set of symmetric positive-definite matrices to a linear space. This transformation of measure allows us to set up an MCMC method that ensures the symmetry and the positive definiteness of the matrix samples while being defined in terms of state variables with values in a linear space. We use the Itô-SDE-based MCMC method introduced in [15], which does not require a kernel to be chosen and inherits computational efficiency in high dimension from leveraging the gradient of the target to know “where to go” in the high-dimensional space [15, 16].

We apply the proposed methodology to a problem of a data-driven updating of a reduced-order model in computational structural dynamics. In this domain of application, the data-driven updating with a maximum-entropy probability distribution used as the prior can also be seen as a data-driven updating of a nonparametric probabilistic model constructed as introduced in [5]. However, the end result is *not* an identification of hyperparameters of a nonparametric probabilistic model as in [7, 8, 9, 17, 18]; indeed, the posterior

is in general no longer a member of the family of maximum-entropy probability distributions. Rather, we obtain an updated probabilistic reduced-order model with a posterior for the reduced matrices that is specifically tailored to provide improved predictions of the dynamical behavior of the structure as well as a probabilistic representation of the model uncertainties.

The proposed methodology relates to the broader context of MCMC methods [4, 19] as follows. MCMC methods for sampling from probability distributions whose support is not a linear space may be obtained by using a rejection mechanism to discard samples that lie outside the support, by using barrier functions to repel the Markov chain from the boundary of the support, by using reparameterizations, as well as in other ways. In the proposed methodology, the use of the transformation of measure may be viewed as the use of a reparameterization. MCMC methods for sampling from complex probability distributions may be accelerated by exploiting local or global structure of the target probability distribution. Such local or global structure may be deduced from analyses of previous samples, from gradients and Hessians of the posterior, as well as from other sources of information; and it may be exploited by adaptivity mechanisms; by multiscale methods; by surrogate modeling, for instance, in low dimension, the method in [20]; by transformations such as transport maps, for instance, in low dimension, the method in [21]; by deducing transition kernels from stochastic dynamics in a manner that exploits derivative information. In the proposed methodology, the use of the Itô-SDE-based MCMC method may be viewed as a method that may be classified in the latter class of methods.

This paper is organized as follows. In Sec. 2, we describe in a general setting a formulation and a numerical implementation of a Bayesian identification of symmetric positive-definite matrices. In Sec. 3, we provide details relevant to the use of a maximum-entropy probability distribution as the prior. In Sec. 4, we cast a problem of a data-driven updating of a reduced-order model in structural dynamics as a Bayesian identification of symmetric positive-definite matrices. In Sec. 5, we provide numerical results.

2. Bayesian identification of symmetric positive-definite matrices

We now describe in a general setting a formulation and a numerical implementation of a Bayesian identification of a symmetric positive-definite matrix. Specifically, we consider a Bayesian inversion of a model

$$\mathbf{y} = \mathbf{g}([X]), \quad \mathbf{g} : \mathbf{M}_n^+(\mathbb{R}) \rightarrow \mathbb{R}^m, \quad (1)$$

that solves a forward problem, that is, that associates to a value $[X]$ of the symmetric positive-definite matrix, which we assume to be of dimension n , a corresponding value \mathbf{y} of observables, which we assume to be collected in a vector of dimension m ; here and in the remainder of this paper, $M_n^+(\mathbb{R})$ denotes the set of n -dimensional symmetric positive-definite matrices with real entries. We denote by \mathbf{y}_{obs} the actually observed value of the observables.

2.1. Formulation within the Bayesian approach to inverse problems

The Bayesian approach to inverse problems leads to the Bayes formula that updates a prior with a likelihood to obtain a posterior as follows [1, 2, 3]:

$$\underbrace{\rho_{[\mathbf{X}]}([X]|\mathbf{y}_{\text{obs}})}_{\text{posterior}} \propto \underbrace{f(\mathbf{y}_{\text{obs}}|[X])}_{\text{likelihood}} \times \underbrace{\rho_0([X])}_{\text{prior}}. \quad (2)$$

In this Bayesian approach to inverse problems, the prior probability density function (PDF) ρ_0 is construed to represent any information available about the symmetric positive-definite matrix other than that contained in the data; the likelihood function f is construed to represent the information extracted about the symmetric positive-definite matrix from the data; and the posterior PDF $\rho_{[\mathbf{X}]}$ is construed to represent the plausibility of values of the symmetric positive-definite matrix as this plausibility is implied by all the available information. In (2) and in the remainder of this paper, the notation \propto expresses that the PDF in the left-hand side is proportional to the right-hand side, whereby the proportionality constant is implied by the requirement that the PDF be normalized: it must integrate to 1 over its domain of definition.

The determination of a suitable prior is where the symmetry and the positive definiteness of the matrix pose a first difficulty. Indeed, the prior must assign positive weight only to symmetric positive-definite matrices, that is, it must satisfy $\pi_0(M_n^+(\mathbb{R})) = 1$, where π_0 is the probability distribution that admits ρ_0 as a PDF, written as $\pi_0(dX) = \rho_0([X])dX$; for details about the definition of the volume element, dX , on $M_n^+(\mathbb{R})$, we refer to [5]. Through the Bayes formula, the posterior then inherits the symmetry and positive-definiteness properties encoded in the prior, so that samples from the posterior are symmetric positive-definite matrices. We address this difficulty by using a prior from the family of PDFs for symmetric positive-definite random matrices constructed by means of the maximum-entropy principle in [5, 6]. However, as the proposed methodology is applicable even for other types of prior, we proceed with presenting it in a general setting with the prior

denoted generically by ρ_0 ; in Sec. 3, we describe the use of the proposed methodology with a maximum-entropy PDF used as the prior.

We use for the likelihood function an expression as an exponential of minus a weighted least-squares model-data misfit represented by a distance:

$$f(\mathbf{y}_{\text{obs}}|[X]) \propto \exp\left(-\frac{1}{2}\|\mathbf{y}_{\text{obs}} - \mathbf{g}([X])\|_{[\Gamma]}^2\right); \quad (3)$$

here, $\|\epsilon\|_{[\Gamma]}^2 = \epsilon^T[\Gamma]\epsilon$ with $\epsilon = \mathbf{y}_{\text{obs}} - \mathbf{g}([X])$. If the data are perturbed by noise, this expression can be interpreted as a representation of the impact of the noise as a centered multivariate Gaussian random variable added to the observables predicted by the forward problem; thus, $[\Gamma]^{-1}$ is the covariance matrix of the Gaussian random variable. If the data are free of noise, this expression can be interpreted as a Gaussian regularization of a Dirac impulse concentrated at the data.

2.2. Transformation of measure

The posterior does not in general possess a special structure that can be exploited to set up a direct simulation method to sample from it. Thus, to sample from the posterior, an MCMC method must be used.

The construction of such an MCMC method is where the symmetry and the positive definiteness of the matrix pose a second difficulty: the samples produced by the MCMC method must satisfy the symmetry and positive-definiteness constraint. We address this difficulty by using a transformation of measure. We assume that even if the posterior does not possess a special structure, the prior (is chosen such that it) does possess a special structure that can be exploited to represent the prior on the set of symmetric positive-definite matrices in a relatively simple way as a deterministic transformation of another PDF on a linear space. Specifically, we assume that there exists a relatively simple deterministic transformation \mathcal{X} from the linear space $\mathbb{R}^{n(n+1)/2}$ to the set of symmetric positive-definite matrices $M_n^+(\mathbb{R})$,

$$\mathcal{X} : \mathbb{R}^{n(n+1)/2} \rightarrow M_n^+(\mathbb{R}) : \boldsymbol{\xi} \mapsto [X] = [\mathcal{X}(\boldsymbol{\xi})], \quad (4)$$

where the dimension $n(n+1)/2$ reflects the matrix symmetry, such that the image of the standard Gaussian on $\mathbb{R}^{n(n+1)/2}$ under \mathcal{X} is the prior on $M_n^+(\mathbb{R})$:

$$\pi_0 = \pi_N \circ \mathcal{X}^{-1}, \quad \pi_0(dX) = \rho_0([X])dX, \quad \pi_N(d\boldsymbol{\xi}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\xi}\|^2\right)d\boldsymbol{\xi}. \quad (5)$$

For the maximum-entropy PDF, such a transformation can be deduced immediately from the direct simulation method available to obtain samples from it, as we show in Sec. 3. Given such a transformation, we define the PDF ρ_{Ξ} on $\mathbb{R}^{n(n+1)/2}$ whose image under \mathcal{X} is the posterior on $M_n^+(\mathbb{R})$:

$$\rho_{\Xi}(\xi|\mathbf{y}_{\text{obs}}) \propto \exp\left(-\frac{1}{2}\|\mathbf{y}_{\text{obs}} - \mathbf{g}([\mathcal{X}(\xi)])\|_{[\Gamma]}^2\right) \times \exp\left(-\frac{1}{2}\|\xi\|^2\right). \quad (6)$$

Thus, owing to this transformation of measure, an MCMC method can be set up to sample from ρ_{Ξ} on $\mathbb{R}^{n(n+1)/2}$ where it does not face the symmetry and positive-definiteness constraint.

2.3. Change of variables

The first factor in the expression of ρ_{Ξ} represents the information extracted from the data and the second factor represents the prior information. The transformation of measure is such that the second factor is standard Gaussian with a covariance matrix equal to the identity matrix. However, as a consequence of the typical ill-posedness of inverse problems, the data can be expected to provide significant information only on certain components of the symmetric positive-definite matrix and leave other components rather uninformed. Correspondingly, the first factor in the expression of ρ_{Ξ} can be expected to concentrate its weight on narrower ranges of values only in certain directions in $\mathbb{R}^{n(n+1)/2}$ and distribute its weight over wider ranges of values in other directions.

To ensure that an MCMC method works well in such a situation, we introduce an additional change of variables defined as a linear transformation involving a Cholesky factor of an approximation to the covariance matrix. This change of variables is a standard approach related to principal component analysis. Specifically, we describe in Sec. 2.5 how an approximation $[\widehat{C}]$ to the covariance matrix of ρ_{Ξ} can be obtained via a linearization of the inverse problem. Given such an approximation $[\widehat{C}]$ to the covariance matrix of ρ_{Ξ} , we use its Cholesky factorization into the product of a lower-triangular matrix and its transpose,

$$[\widehat{C}] = [L_{\widehat{C}}][L_{\widehat{C}}]^T, \quad (7)$$

to define a transformation from $\mathbb{R}^{n(n+1)/2}$ to $\mathbb{R}^{n(n+1)/2}$ given by

$$\mathbf{q} = [L_{\widehat{C}}]^{-1}\xi. \quad (8)$$

Given this transformation, we define the PDF $\rho_{\mathbf{Q}}$ on $\mathbb{R}^{n(n+1)/2}$ whose image under this transformation is ρ_{Ξ} on $\mathbb{R}^{n(n+1)/2}$:

$$\rho_{\mathbf{Q}}(\mathbf{q}|\mathbf{y}_{\text{obs}}) \propto \exp\left(-\frac{1}{2}\|\mathbf{y}_{\text{obs}} - \mathbf{g}([\mathcal{X}([L_{\hat{C}}]\mathbf{q})])\|_{[\Gamma]}^2\right) \times \exp\left(-\frac{1}{2}\|[L_{\hat{C}}]\mathbf{q}\|^2\right). \quad (9)$$

Provided that $[\hat{C}]$ is a good approximation to the covariance matrix of ρ_{Ξ} , the covariance matrix of $\rho_{\mathbf{Q}}$ is approximately equal to the identity matrix.

2.4. Itô-SDE-based MCMC method

The dimension of $\mathbb{R}^{n(n+1)/2}$ grows with the square of the matrix dimension and hence can be high, so that conventional MCMC methods can be inefficient. We address this difficulty by using the Itô-SDE-based MCMC method introduced in [15]. This MCMC method is based on the Itô SDE

$$\begin{cases} d\mathbf{Q} = \mathbf{P}dt, \\ d\mathbf{P} = -\nabla_{\mathbf{q}}\psi(\mathbf{Q})dt - \frac{1}{2}\zeta_0\mathbf{P}dt + \sqrt{\zeta_0}d\mathbf{W}, \end{cases} \quad (10)$$

where $\zeta_0 > 0$ is a free parameter and $\{\mathbf{W}(t), t \geq 0\}$ is a normalized Wiener stochastic process. Under certain regularity conditions, this Itô SDE is such that when completed with an initial condition, its solution $\{(\mathbf{Q}(t), \mathbf{P}(t)), t \geq 0\}$ is ergodic for the so-called Gibbs PDF

$$\rho_{(\mathbf{Q}, \mathbf{P})}(\mathbf{q}, \mathbf{p}) \propto \exp\left(-\psi(\mathbf{q}) - \frac{\|\mathbf{p}\|^2}{2}\right), \quad (11)$$

where the ergodicity property means that time averages over trajectories of the solution can be used to approximate probabilistic averages with respect to the Gibbs PDF. With reference to this ergodicity property, the Itô-SDE-based MCMC method is obtained by setting the so-called potential ψ in the Itô SDE in (10) equal to the negative logarithm of the target PDF, that is, the PDF that must be sampled from, and then discretizing this Itô SDE in time. Thus, in the Itô-SDE-based MCMC method, it is the time discretization of an Itô SDE that provides the transition kernel of the Markov chain. Through the gradient of the potential involved in the Itô SDE, the transition kernel is able to leverage the gradient of the target PDF to guide the Markov chain to regions to which the target PDF attaches larger weight.

We thus set ψ equal to the negative logarithm of PDF $\rho_{\mathbf{Q}}$ defined in (9):

$$\psi(\mathbf{q}) = \frac{1}{2}\|\mathbf{y}_{\text{obs}} - \mathbf{g}([\mathcal{X}([L_{\hat{C}}]\mathbf{q})])\|_{[\Gamma]}^2 + \frac{1}{2}\|[L_{\hat{C}}]\mathbf{q}\|^2, \quad (12)$$

where we can use an equality sign because the multiplicative normalization constant in (9) need not be known, because the logarithm transforms it into an additive constant that the gradient in (10) then absorbs.

Before discretizing the Itô SDE in time, we introduce in it the transformation in (8) as well as the transformation

$$\mathbf{p} = [L_{\hat{C}}]^T \boldsymbol{\eta} \quad (13)$$

to obtain

$$\begin{cases} d\boldsymbol{\Xi} = [\hat{C}] \mathbf{H} dt, \\ d\mathbf{H} = -\nabla_{\boldsymbol{\xi}} \phi(\boldsymbol{\Xi}) dt - \frac{1}{2} \zeta_0 \mathbf{H} dt + \sqrt{\zeta_0} [L_{\hat{C}}]^{-T} d\mathbf{W}, \end{cases} \quad (14)$$

with ϕ equal to the negative logarithm of PDF $\rho_{\boldsymbol{\Xi}}$ defined in (6):

$$\phi(\boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{y}_{\text{obs}} - \mathbf{g}([\mathcal{X}(\boldsymbol{\xi})])\|_{[\Gamma]}^2 + \frac{1}{2} \|\boldsymbol{\xi}\|^2. \quad (15)$$

The purpose of (13) is to arrive in (14) at a system of equations that is a Hamiltonian equation for the Hamiltonian $h(\boldsymbol{\xi}, \boldsymbol{\eta}) = \phi(\boldsymbol{\xi}) + \frac{1}{2} \boldsymbol{\eta}^T [\hat{C}] \boldsymbol{\eta}$, in which there is still the additional friction term and the additional stochastic forcing term. Owing to the transformations in (8) and (13), the Itô-SDE-based MCMC method will sample directly from $\rho_{\boldsymbol{\Xi}}$, while still benefitting from the change of variables, which still manifests itself in (16) by the presence of the matrices $[\hat{C}]$ and $[L_{\hat{C}}]^{-T}$. The effect of the presence of these matrices is that of a normalization of the dynamics of the Itô SDE. One way of seeing this is as follows: injecting the second equation into the first equation in (14) reveals that the square roots of the eigenvalues of $[\hat{C}] D_{\boldsymbol{\xi}} \nabla_{\boldsymbol{\xi}} \phi$ are indicative of the time scales involved in the dynamics of the Itô SDE; considering that the inverse of the Hessian of the negative logarithm of a PDF provides an approximation to the covariance matrix of this PDF—an approximation that is exact for a Gaussian PDF—the effect of the presence of the matrices $[\hat{C}]$ and $[L_{\hat{C}}]^{-T}$ is that of a clustering of the eigenvalues of $[\hat{C}] D_{\boldsymbol{\xi}} \nabla_{\boldsymbol{\xi}} \phi$ close to 1 and thus that of a normalization of the time scales involved in the dynamics of the Itô SDE.

To discretize in time, we use a Störmer–Verlet method [22] that had

already been used for the Itô-SDE-based MCMC method in [23], leading to

$$\begin{cases} \Xi^{(\ell+\frac{1}{2})} = \Xi^{(\ell)} + \frac{\Delta t}{2}[\widehat{C}]\mathbf{H}^{(\ell)}, \\ \mathbf{H}^{(\ell+1)} = \frac{1-b}{1+b}\mathbf{H}^{(\ell)} - \frac{\Delta t}{1+b}\nabla_{\xi}\phi(\Xi^{(\ell+\frac{1}{2})}) + \frac{\sqrt{\zeta_0}}{1+b}[L_{\widehat{C}}]^{-\text{T}}\Delta\mathbf{W}^{(\ell+1)}, \\ \Xi^{(\ell+1)} = \Xi^{(\ell+\frac{1}{2})} + \frac{\Delta t}{2}[\widehat{C}]\mathbf{H}^{(\ell+1)}, \end{cases} \quad (16)$$

where

$$b = \Delta t \zeta_0/4, \quad (17)$$

in which Δt denotes the time step and $\Delta\mathbf{W}^{(1)}, \Delta\mathbf{W}^{(2)}, \dots$ are mutually independent centered multivariate Gaussian random variables with covariance matrix $\Delta t[I]$, in which $[I]$ is here the $n(n+1)/2$ -dimensional identity matrix. The change of variables and its effect of a normalization of the dynamics of the Itô SDE facilitate the choice of the time step and the control of the error entailed by the time discretization. The end result of the change of variables is that it allows the Markov chain to make faster progress in directions that are less informed by the data, without being constrained by a stability-related or error-related requirement of slower progress in directions that are more informed by the data.

When completed with an initial condition, the time-discrete equations in (16) define a Markov chain $\{(\Xi^{(\ell)}, \mathbf{H}^{(\ell)}), \ell \in \mathbb{N}\}$. The transformation of measure associates the chain of random matrices $\{[\mathbf{X}^{(\ell)}], \ell \in \mathbb{N}\}$ with the component chain of random variables $\{\Xi^{(\ell)}, \ell \in \mathbb{N}\}$:

$$[\mathbf{X}^{(\ell)}] = [\mathcal{X}(\Xi^{(\ell)})], \quad (18)$$

Owing to the transformation of measure, owing to the ergodicity property of the solution to the Itô SDE, and owing to the choice of the potential, but up to an error entailed by the time discretization, trajectories of this chain of random matrices can be construed as samples drawn from the posterior.

The gradient of the potential in the transition kernel is the sum of two terms, with the first term stemming from the weighted distance that represents the model-data misfit and the second term stemming from the prior:

$$\nabla_{\xi}\phi(\xi) = [J(\xi)]^{\text{T}}[\Gamma]\left(\mathbf{g}([\mathcal{X}(\xi)]) - \mathbf{y}_{\text{obs}}\right) + \xi; \quad (19)$$

here, $[J(\boldsymbol{\xi})]$ denotes the gradient of the composition of \mathcal{X} and \mathbf{g} , which, with the chain rule, can be written, in terms of entries, as follows:

$$J_{ij}(\boldsymbol{\xi}) = \frac{\partial g_i}{\partial \xi_j}([\mathcal{X}(\boldsymbol{\xi})]) = \sum_{\ell=1}^n \sum_{\ell'=1}^n \frac{\partial g_i}{\partial X_{\ell\ell'}}([\mathcal{X}(\boldsymbol{\xi})]) \frac{\partial X_{\ell\ell'}}{\partial \xi_j}(\boldsymbol{\xi}). \quad (20)$$

Thus, for the proposed methodology to work well, the forward problem must afford an efficient computation of not only the observables but also their gradient with respect to the matrix. For the maximum-entropy PDF, the expression of the gradient of the transformation can be readily obtained, as we show in Sec. 3. In the problem of a data-driven updating of a reduced-order model in computational structural dynamics, we use an adjoint-based computation of the gradient, as we discuss in Sec. 4.

2.5. Obtaining $[\hat{C}]$ from a linearization of the inverse problem

We will conclude the section with discussing how a linearization of the inverse problem allows an approximation $[\hat{C}]$ to the covariance matrix of ρ_{Ξ} to be obtained. Let the composition of the transformation \mathcal{X} and the forward problem \mathbf{g} be linearized about a reference value $\boldsymbol{\xi}_0$:

$$\mathbf{g}([\mathcal{X}(\boldsymbol{\xi})]) \approx \mathbf{g}([\mathcal{X}(\boldsymbol{\xi}_0)]) + [J(\boldsymbol{\xi}_0)](\boldsymbol{\xi} - \boldsymbol{\xi}_0); \quad (21)$$

here, among other choices, $\boldsymbol{\xi}_0$ can be chosen equal to a mode of the prior or a mode of the posterior, such as a value that maximizes ρ_{Ξ} , which can be computed with an optimization method such as a quasi-Newton method:

$$\begin{aligned} & ([J(\boldsymbol{\xi}^{(\ell)})]^T [\Gamma] [J(\boldsymbol{\xi}^{(\ell)})] + [I]) \boldsymbol{\xi}^{(\ell+1)} \\ &= -[J(\boldsymbol{\xi}^{(\ell)})]^T [\Gamma] \left(\mathbf{g}([\mathcal{X}(\boldsymbol{\xi}^{(\ell)})]) - [J(\boldsymbol{\xi}^{(\ell)})] \boldsymbol{\xi}^{(\ell)} - \mathbf{y}_{\text{obs}} \right). \end{aligned} \quad (22)$$

It can be seen by injecting (21) into (6) that the approximation of the composition of the transformation \mathcal{X} and the forward problem \mathbf{g} by its linearized expression leads to the approximation of the weighted distance that represents the model-data misfit by a quadratic expression and ultimately the approximation of ρ_{Ξ} by a Gaussian PDF as follows:

$$\rho_{\Xi}(\boldsymbol{\xi} | \mathbf{y}_{\text{obs}}) \approx \hat{\rho}_{\Xi}(\boldsymbol{\xi} | \mathbf{y}_{\text{obs}}) \propto \exp \left(-\frac{1}{2} \|\boldsymbol{\xi} - \underline{\boldsymbol{\xi}}\|_{[\hat{C}]^{-1}}^2 \right), \quad (23)$$

with mean value given by

$$\underline{\boldsymbol{\xi}} = ([J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)] + [I])^{-1} [J(\boldsymbol{\xi}_0)]^T [\Gamma] (\mathbf{y}_{\text{obs}} - \mathbf{g}([\mathcal{X}(\boldsymbol{\xi}_0)]) + [J(\boldsymbol{\xi}_0)] \boldsymbol{\xi}_0) \quad (24)$$

and covariance matrix given by

$$[\hat{C}] = ([J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)] + [I])^{-1}. \quad (25)$$

It is this matrix that can be used in the change of variables as the approximation to the covariance matrix of ρ_{Ξ} .

With reference to the change of variables, it is interesting to note that as in [2], insight into the information provided by the data can be gained from an eigendecomposition. Indeed, with $[\Lambda_r]$ the diagonal matrix with the r dominant eigenvalues and $[V_r]$ collecting in its columns the corresponding eigenvectors of $[J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)]$,

$$[J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)] \approx [V_r] [\Lambda_r] [V_r]^T, \quad [V_r] [V_r]^T = [I], \quad (26)$$

the Sherman–Morrison–Woodbury formula provides the approximation

$$[\hat{C}] = ([J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)] + [I])^{-1} \approx [I] - [V_r] [D_r] [V_r]^T, \quad (27)$$

with

$$[D_r] = ([\Lambda_r] + [I])^{-1} [\Lambda_r], \quad (28)$$

thus suggesting that the eigenvectors corresponding to the dominant eigenvalues of $[J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)]$ are indicative of the directions in $\mathbb{R}^{n(n+1)/2}$ that the data provide significant information on. As a consequence of the typical ill-posedness of inverse problems, the spectrum of $[J(\boldsymbol{\xi}_0)]^T [\Gamma] [J(\boldsymbol{\xi}_0)]$ can be expected to decay rapidly. Later in this paper, we will also find this result helpful for the identification of hyperparameters of the posterior.

3. Using a maximum-entropy PDF as the prior

We now describe the use of the proposed methodology with a prior taken from the family of PDFs for symmetric positive-definite random matrices constructed by means of the maximum-entropy principle in [5, 6].

3.1. Concise review of the family of maximum-entropy PDFs

The family of PDFs for symmetric positive-definite random matrices constructed by means of the maximum-entropy principle in [5, 6] has two hyperparameters: the mean value and a dispersion parameter. A symmetric positive-definite random matrix, say, $[\mathbf{X}]$, has such a maximum-entropy PDF with mean value $[\underline{X}]$, which must be a symmetric positive definite matrix,

and dispersion parameter δ if given the Cholesky factorization of $[\underline{X}]$ into the product of a lower-triangular matrix and its transpose,

$$[\underline{X}] = [L_{\underline{X}}][L_{\underline{X}}]^T, \quad (29)$$

the normalized symmetric positive-definite random matrix $[\mathbf{G}]$ in

$$[\mathbf{X}] = [L_{\underline{X}}][\mathbf{G}][L_{\underline{X}}]^T \quad (30)$$

has a PDF $\rho_{[\mathbf{G}]}$ that is obtained by maximizing entropy,

$$\max_{\rho_{[\mathbf{G}]}} - \int_{M_n^+(\mathbb{R})} \log \rho_{[\mathbf{G}]}([G]) \rho_{[\mathbf{G}]}([G]) dG, \quad (31)$$

under the following constraints:

$$\int_{M_n^+(\mathbb{R})} \rho_{[\mathbf{G}]}([G]) dG = 1, \quad (32)$$

$$\int_{M_n^+(\mathbb{R})} [G] \rho_{[\mathbf{G}]}([G]) dG = [I], \quad (33)$$

$$\int_{M_n^+(\mathbb{R})} \log(\det[G]) \rho_{[\mathbf{G}]}([G]) dG < +\infty. \quad (34)$$

The solution of the constrained optimization problem (31) with (32)–(34) by means of the method of Lagrange multipliers leads to

$$\rho_{[\mathbf{G}]}([G]) \propto 1_{M_n^+(\mathbb{R})}([G]) (\det[G])^{\frac{n+1}{2\delta^2} + \frac{1-n}{2}} \exp\left(-\frac{n+1}{2\delta^2} \text{tr}[G]\right), \quad (35)$$

where $1_{M_n^+(\mathbb{R})}([G]) = 1$ if $[G]$ is an n -dimensional symmetric positive-definite matrix with real entries and $1_{M_n^+(\mathbb{R})}([G]) = 0$ otherwise, $\det[G]$ is the determinant of $[G]$, and $\text{tr}[G]$ is the trace of $[G]$. The Lagrange multiplier associated with the constraint in (34) results in the presence of the hyperparameter in (35), namely, the dispersion parameter, δ , defined such that

$$\delta^2 = \frac{\int_{M_n^+(\mathbb{R})} \|[G] - [I]\|_F^2 \rho_{[\mathbf{G}]}([G]) dG}{\|[I]\|_F^2}, \quad (36)$$

in which $\|\cdot\|_F$ denotes the Frobenius norm; thus, δ ascertains the dispersion of $[\mathbf{G}]$ and therefore also of $[\mathbf{X}]$. The constraint in (32) imposes that $\rho_{[\mathbf{G}]}$

integrate to 1 over $M_n^+(\mathbb{R})$ so that samples of $[\mathbf{G}]$ and therefore also samples of $[\mathbf{X}]$ are symmetric and positive definite. The constraint in (33) imposes that the mean value of $[\mathbf{G}]$ be $[I]$ and therefore the mean value of $[\mathbf{X}]$ be $[\underline{X}]$. The constraint in (34) serves to ensure the existence of the second-order moment of the inverse of $[\mathbf{G}]$ and therefore of $[\mathbf{X}]$, as required when such symmetric positive-definite random matrices are used to represent uncertain symmetric positive-definite matrices of computational models, see, for instance, [5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

The PDF $\rho_{[\mathbf{G}]}$ possesses a special structure that can be exploited to set up a direct simulation method based on the representation of $[\mathbf{G}]$ as the product of a lower-triangular random matrix and its transpose,

$$[\mathbf{G}] = [\mathbf{L}][\mathbf{L}]^T, \quad (37)$$

with the entries $\{L_{\ell\ell'}, n \geq \ell \geq \ell' \geq 1\}$ of $[\mathbf{L}]$ mutually statistically independent random variables distributed as follows:

for $\ell = \ell'$:

$$L_{\ell\ell} = \sqrt{2} \frac{\delta}{\sqrt{n+1}} \sqrt{Y_{\ell\ell}}, \text{ with } Y_{\ell\ell} \sim \Gamma(\alpha_\ell, 1), \alpha_\ell = \frac{n+1}{2\delta^2} + \frac{1-\ell}{2}; \quad (38)$$

for $\ell > \ell'$:

$$L_{\ell\ell'} \sim \mathcal{N}(0, \sigma^2), \sigma = \frac{\delta}{\sqrt{n+1}}; \quad (39)$$

here, we denote by $Y_{\ell\ell} \sim \Gamma(\alpha_\ell, 1)$ that $Y_{\ell\ell}$ is a gamma random variable with shape parameter α_ℓ and scale parameter 1, and we denote by $L_{\ell\ell'} \sim \mathcal{N}(0, \sigma^2)$ that $L_{\ell\ell'}$ is a Gaussian random variable with mean 0 and variance σ^2 .

3.2. Using a maximum-entropy PDF as the prior in the proposed methodology

The use of a maximum-entropy PDF with mean value $[\underline{X}]$ and dispersion parameter δ as the prior in the proposed methodology amounts to setting

$$\rho_0([X]) \propto \rho_{[\mathbf{G}]}([L_{\underline{X}}]^{-1}[X][L_{\underline{X}}]^{-T}). \quad (40)$$

The transformation of measure requires a representation of the prior on the set of symmetric positive-definite matrices $M_n^+(\mathbb{R})$ as the image under a transformation \mathcal{X} of a standard Gaussian PDF on the linear space $\mathbb{R}^{n(n+1)/2}$.

Such a transformation \mathcal{X} immediately follows from the representation used in the direct simulation method in (37) with (38)–(39) by setting

$$[\mathcal{X}(\boldsymbol{\xi})] = [L_{\underline{X}}][\mathcal{L}(\boldsymbol{\xi})][\mathcal{L}(\boldsymbol{\xi})]^T[L_{\underline{X}}]^T, \quad (41)$$

where \mathcal{L} is the transformation from the linear space $\mathbb{R}^{n(n+1)/2}$ to the set of lower-triangular matrices with positive diagonal entries \mathcal{L}_L ,

$$\mathcal{L} : \mathbb{R}^{n(n+1)/2} \rightarrow \mathcal{L}_L : \boldsymbol{\xi} \mapsto [L] = [\mathcal{L}(\boldsymbol{\xi})], \quad (42)$$

such that, with the so-called half-vectorization

$$\boldsymbol{\xi} = (\xi_{11}, \dots, \xi_{n1}, \xi_{22}, \dots, \xi_{n2}, \dots, \xi_{(n-1)(n-1)}, \xi_{n(n-1)}, \xi_{nn}), \quad (43)$$

the entries $\{\mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}), n \geq \ell \geq \ell' \geq 1\}$ of $[\mathcal{L}(\boldsymbol{\xi})]$ are as follows:

for $\ell = \ell'$:

$$\mathcal{L}_{\ell\ell}(\boldsymbol{\xi}) = \sqrt{2} \frac{\delta}{\sqrt{n+1}} \sqrt{y_{\ell\ell}}, \quad y_{\ell\ell} = c_{\Gamma}^{-1}(c_{\mathcal{N}}(\xi_{\ell\ell}; 0, 1); \alpha_{\ell}, 1); \quad (44)$$

for $\ell > \ell'$:

$$\mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}) = \frac{\delta}{\sqrt{n+1}} \xi_{\ell\ell'}; \quad (45)$$

here, $c_{\Gamma}(\cdot; \alpha_{\ell}, 1)$ is the cumulative distribution function of a gamma random variable with shape parameter α_{ℓ} and scale parameter 1, and $c_{\mathcal{N}}(\cdot; 0, 1)$ is the cumulative distribution function of a Gaussian random variable with mean 0 and variance 1. It is owing to the isoprobability transform that the gamma PDF with shape parameter α_{ℓ} and scale parameter 1 is the image under the transformation $c_{\Gamma}^{-1}(c_{\mathcal{N}}(\cdot; 0, 1); \alpha_{\ell}, 1)$ of a standard Gaussian PDF.

The approximation of the covariance matrix in the change of variables and the computation of the gradient in the Itô-SDE-based MCMC method require the gradient of the transformation \mathcal{X} . From (41) with (44)–(45), it immediately follows that the components of the gradient of the transformation \mathcal{X} are given by

$$\partial_{\xi_{\ell\ell'}}[\mathcal{X}(\boldsymbol{\xi})] = [L_{\underline{X}}](\partial_{\xi_{\ell\ell'}}[\mathcal{L}(\boldsymbol{\xi})][\mathcal{L}(\boldsymbol{\xi})]^T + [\mathcal{L}(\boldsymbol{\xi})]\partial_{\xi_{\ell\ell'}}[\mathcal{L}(\boldsymbol{\xi})]^T)[L_{\underline{X}}]^T, \quad (46)$$

where

$$\partial_{\xi_{\ell\ell'}}[\mathcal{L}(\boldsymbol{\xi})] = \partial_{\xi_{\ell\ell'}}\mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}) \mathbf{e}_{\ell} \otimes \mathbf{e}_{\ell'}, \quad (47)$$

with

for $\ell = \ell'$:

$$\partial_{\xi_{\ell\ell}} \mathcal{L}_{\ell\ell}(\boldsymbol{\xi}) = \sqrt{2} \frac{\delta}{\sqrt{n+1}} \frac{1}{2} \frac{1}{\sqrt{y_{\ell\ell}}} \frac{\rho_{\mathcal{N}}(\xi_{\ell\ell}; 0, 1)}{\rho_{\Gamma}(y_{\ell\ell}; \alpha_{\ell}, 1)}; \quad (48)$$

for $\ell > \ell'$:

$$\partial_{\xi_{\ell\ell'}} \mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}) = \frac{\delta}{\sqrt{n+1}}; \quad (49)$$

here, $\rho_{\Gamma}(\cdot; \alpha_{\ell}, 1)$ is the PDF of a gamma random variable with shape parameter α_{ℓ} and scale parameter 1, and $\rho_{\mathcal{N}}(\cdot; 0, 1)$ is the standard Gaussian PDF.

4. Data-driven update of a reduced-order model

We now describe in the domain of computational structural dynamics a problem of a data-driven updating of a reduced-order model.

4.1. Reduced-order model

We consider a deterministic finite element model for the linear dynamical behavior of a dissipative structure about a static equilibrium in the frequency domain. Let this model be projected onto a reduced-order basis to obtain a reduced-order model of the following form:

$$\begin{cases} (-\omega^2 [\underline{M}] + i\omega [\underline{D}] + [\underline{K}]) \mathbf{z}(\omega) = [\underline{\Phi}]^T \mathbf{f}(\omega), \\ \mathbf{u}(\omega) = [\underline{\Phi}] \mathbf{z}(\omega); \end{cases} \quad (50)$$

here, ω denotes the (circular) frequency; $[\underline{\Phi}]$ the matrix whose columns collect the reduced-order basis; $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ the reduced mass, damping, and stiffness matrices; $\mathbf{f}(\omega)$ the vector of external forces; $\mathbf{z}(\omega)$ the vector of n reduced degrees of freedom; and $\mathbf{u}(\omega)$ the vector of displacement degrees of freedom. Although extensions of the proposed methodology to other types of dissipation and boundary conditions can be readily obtained, we assume that the dissipation and the boundary conditions are such that $[\underline{D}]$ and $[\underline{K}]$ are symmetric and positive definite.

4.2. Data-driven update of the reduced-order model

In computational structural dynamics, data-driven updating most often involves adjusting values of mechanical properties, such as stiffness, dissipation, and inertial properties, so that predictions better fit data relevant to the dynamical behavior of the structure. However, it is *not* such an adjustment of values of mechanical properties that we pursue. Rather, we seek to adjust *directly* the reduced matrices of the reduced-order model, and we seek to do so with a probabilistic approach that allows us to obtain not only improved predictions but also a probabilistic representation of the model uncertainties.

Thus, we embed the reduced-order model in (50) in a class of reduced-order models indexed by the values assigned to the reduced matrices, now treated as (generalized) parameters of the reduced-order model:

$$\begin{cases} (-\omega^2[M] + i\omega[D] + [K])\mathbf{z}(\omega; [M], [D], [K]) = [\underline{\Phi}]^T \mathbf{f}(\omega), \\ \mathbf{u}(\omega; [M], [D], [K]) = [\underline{\Phi}]\mathbf{z}(\omega; [M], [D], [K]), \end{cases} \quad (51)$$

and we cast the data-driven update as a Bayesian identification of the (generalized) parameters of the reduced-order model, that is, the reduced matrices. The specification of the posterior on the set of possible values of the reduced matrices then provides the updated probabilistic reduced-order model.

4.3. Using a maximum-entropy PDF as the prior

We use the product of maximum-entropy PDFs with mean values $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ and dispersion parameters δ_M , δ_D , and δ_K as the prior:

$$\rho_0([M], [D], [K]) \propto \prod_{A=M,D,K} \rho_{[\mathbf{G}_A]}([\underline{L}_A]^{-1}[A][\underline{L}_A]^{-T}); \quad (52)$$

here, $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ are the reduced mass, damping, and stiffness matrices of the reduced-order model in (50), and the PDFs $\rho_{[\mathbf{G}_M]}$, $\rho_{[\mathbf{G}_D]}$, and $\rho_{[\mathbf{G}_K]}$ are PDFs as in (35) with dispersion parameters δ_M , δ_D , and δ_K , respectively.

4.4. Interpretation of the prior as a nonparametric probabilistic model

The probabilistic reduced-order model obtained by specifying the prior on the set of possible values of the reduced matrices can be viewed as a nonparametric probabilistic model constructed as in [5]. Indeed, within the present problem setting, the nonparametric probabilistic approach of [5] involves representing the reduced matrices in the reduced-order model in (50)

with random matrices $[\mathbf{M}]$, $[\mathbf{D}]$, and $[\mathbf{K}]$ that are statistically independent and have maximum-entropy PDFs with mean values $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ and dispersion parameters δ_M , δ_D , and δ_K , respectively, thus leading to a probabilistic reduced-order model of the following form:

$$\begin{cases} (-\omega^2[\mathbf{M}] + i\omega[\mathbf{D}] + [\mathbf{K}])\mathbf{Z}(\omega) = [\Phi]^T \mathbf{f}(\omega), \\ \mathbf{U}(\omega) = [\Phi]\mathbf{Z}(\omega). \end{cases} \quad (53)$$

Here, and with reference to Sec. 3, by $[\mathbf{M}]$, $[\mathbf{D}]$, and $[\mathbf{K}]$ having maximum-entropy PDFs with mean values $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ and dispersion parameters δ_M , δ_D , and δ_K , we mean that given Cholesky factorizations of $[\underline{M}]$, $[\underline{D}]$, and $[\underline{K}]$ into products of lower-triangular matrices and their transposes,

$$[\underline{A}] = [L_A][L_A]^T, \quad A = M, D, K, \quad (54)$$

the normalized symmetric positive-definite matrices $[\mathbf{G}_M]$, $[\mathbf{G}_D]$, and $[\mathbf{G}_K]$ in

$$[\mathbf{A}] = [L_A][\mathbf{G}_A][L_A]^T, \quad A = M, D, K, \quad (55)$$

have PDFs $\rho_{[\mathbf{G}_M]}$, $\rho_{[\mathbf{G}_D]}$, and $\rho_{[\mathbf{G}_K]}$ that are obtained by maximizing entropy as in (31) with (32)–(34) and thus have a form as in (35) with dispersion parameters δ_M , δ_D , and δ_K , respectively. Thus, δ_M , δ_D , and δ_K ascertain the dispersion of $[\mathbf{G}_M]$, $[\mathbf{G}_D]$, and $[\mathbf{G}_K]$ and therefore also of $[\mathbf{M}]$, $[\mathbf{D}]$, and $[\mathbf{K}]$:

$$\delta_A^2 = \frac{\int_{\mathbf{M}_n^+(\mathbb{R})} \| [G_A] - [I] \|_F^2 \rho_{[\mathbf{G}_A]}([G_A]) dG_A}{\| [I] \|_F^2}, \quad A = M, D, K. \quad (56)$$

In the entropy maximization, the constraint in (34) serves to ensure the existence of the second-order moments of the inverses of $[\mathbf{G}_M]$, $[\mathbf{G}_D]$, and $[\mathbf{G}_K]$ and therefore of $[\mathbf{M}]$, $[\mathbf{D}]$, and $[\mathbf{K}]$, thus ensuring that the probabilistic reduced-order model has a second-order solution. In conclusion, the joint PDF of the random matrices $[\mathbf{M}]$, $[\mathbf{D}]$, and $[\mathbf{K}]$ is the PDF in (52), thus confirming that the probabilistic reduced-order model obtained by specifying the prior on the set of possible values of the reduced matrices is a nonparametric probabilistic model.

In the context of the nonparametric probabilistic approach, several methods have been proposed for the identification of the dispersion parameters. In [7, 8, 9], a method was proposed based on an identification criterion that gauges the agreement between data and the predictions of the nonparametric

probabilistic model. In this method, the data are plotted on the same axes as high-probability confidence regions for the predictions, and a combination of ideally small values of the dispersion parameters is sought for which the data are mostly contained within the high-probability confidence regions for the predictions. In [17, 18], approaches were proposed which allow the identification to be cast as an optimization problem.

4.5. Dynamic test data

Although extensions to other types of data can be readily obtained, we present a data-driven updating of the reduced-order model based on using frequency response functions (FRFs) sampled at frequencies in a frequency band of interest. Specifically, we assume that the data consist of a finite number, say, d , of FRFs sampled at a finite number, say, m , of frequencies $\omega_1, \dots, \omega_m$ in the frequency band of interest:

$$\mathbf{y}_{\text{obs}} = (\mathbf{h}^{\text{obs}}(\omega_1), \dots, \mathbf{h}^{\text{obs}}(\omega_m)), \quad (57)$$

with

$$\mathbf{h}^{\text{obs}}(\omega_k) = (h_1^{\text{obs}}(\omega_k), \dots, h_d^{\text{obs}}(\omega_k)). \quad (58)$$

4.6. Forward problem

We thus define the forward problem as the mapping that represents the use of the reduced-order model to associate to a value of the reduced matrices a corresponding value of the observables:

$$\mathbf{y} = \mathbf{g}([M], [D], [K]) = (\mathbf{h}(\omega_1; [M], [D], [K]), \dots, \mathbf{h}(\omega_m; [M], [D], [K])), \quad (59)$$

where

$$\mathbf{h}(\omega_k; [M], [D], [K]) = (h_1(\omega_k; [M], [D], [K]), \dots, h_d(\omega_k; [M], [D], [K])), \quad (60)$$

with

$$h_j(\omega_k; [M], [D], [K]) = \mathbf{c}_j^T [\underline{\Phi}] (-\omega_k^2 [M] + i\omega_k [D] + [K])^{-1} [\underline{\Phi}]^T \mathbf{b}_j, \quad (61)$$

in which for the j -th FRF, the so-called input-shape vector \mathbf{b}_j and the so-called output-shape vector \mathbf{c}_j link the finite element degrees of freedom to the position and the orientation of the actuator and the sensor, respectively.

4.7. Bayesian identification of the reduced matrices

Casting the data-driven update as a Bayesian identification of the reduced matrices with the maximum-entropy PDF as the prior leads to the posterior

$$\begin{aligned} & \rho_{([M],[D],[K])}([M], [D], [K] | \mathbf{y}_{\text{obs}}) \\ & \propto f(\mathbf{y}_{\text{obs}} | [M], [D], [K]) \times \rho_0([M], [D], [K]), \end{aligned} \quad (62)$$

in which f is the likelihood function. With reference to Sec. 2, we use for f an expression of the form of an exponential of minus a weighted distance that represents the model-data misfit. Although extensions to other types of weighted distance can be readily obtained, we use a logarithmic least-squares model-data misfit, which has the benefits of being as sensitive to resonances as anti-resonances and being insensitive to phase errors:

$$\begin{aligned} & f(\mathbf{y}_{\text{obs}} | [M], [D], [K]) \\ & \propto \exp \left(-\lambda \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \left| \log |h_j^{\text{obs}}(\omega_k)| - \log |h_j(\omega_k; [M], [D], [K])| \right|^2 \right). \end{aligned} \quad (63)$$

A comparison with the expression in (3) indicates that the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$, can be interpreted as the use of a diagonal covariance matrix (the matrix $[\Gamma]^{-1}$ in (3)) whose diagonal elements are the inverses of the products $\lambda \mu_{jk}$. Because f attaches greater weight to values of the reduced matrices that correspond to predictions that fit the data better, the updated probabilistic reduced-order model with the posterior $\rho_{([M],[D],[K])}$ specified on the set of possible values of the reduced matrices can be expected to be specifically tailored to provide improved predictions.

The posterior is parameterized by the dispersion parameters δ_M , δ_D , and δ_K and the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$, and we comment on the identification of these hyperparameters later in Sec. 4.9.

4.8. Numerical implementation of the Bayesian inversion

Because the prior is a product PDF, it suffices to construct separate representations of the maximum-entropy PDFs with mean values $[M]$, $[D]$, and $[K]$ and dispersion parameters δ_M , δ_D , and δ_K as images under transformations of Gaussian PDFs. We construct these transformations as in (41) with (42)–(45) and denote the transformations thus obtained by \mathcal{M} , \mathcal{D} , and \mathcal{K} :

$$\mathcal{A} : \mathbb{R}^{n(n+1)/2} \rightarrow \mathbb{M}_n^+(\mathbb{R}) : \boldsymbol{\xi}_A \mapsto [A] = [\mathcal{A}(\boldsymbol{\xi}_A)], \quad A = M, D, K. \quad (64)$$

Given these transformations, we define the PDF ρ_{Ξ} on $\mathbb{R}^{n(n+1)/2} \times \mathbb{R}^{n(n+1)/2} \times \mathbb{R}^{n(n+1)/2}$ whose image under the compound transformation is the posterior:

$$\rho_{\Xi}(\boldsymbol{\xi} | \mathbf{y}_{\text{obs}}) \propto \exp \left(-\lambda \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \left| \log |h_j^{\text{obs}}(\omega_k)| - \log |h_j(\omega_k; \boldsymbol{\xi})| \right|^2 \right) \times \exp \left(-\frac{1}{2} \|\boldsymbol{\xi}\|^2 \right), \quad (65)$$

in which

$$h_j(\omega_k; \boldsymbol{\xi}) = h_j(\omega_k; [\mathcal{M}(\boldsymbol{\xi}_M)], [\mathcal{D}(\boldsymbol{\xi}_D)], [\mathcal{K}(\boldsymbol{\xi}_K)]), \quad (66)$$

with

$$\boldsymbol{\xi} = (\boldsymbol{\xi}_M, \boldsymbol{\xi}_D, \boldsymbol{\xi}_K). \quad (67)$$

The linearization of the composition of the transformations \mathcal{M} , \mathcal{D} , and \mathcal{K} and the forward problem \mathbf{g} about a reference value $\boldsymbol{\xi}_0$ leads as in (21) and (24) with (25) to an approximation $[\hat{C}]$ to the covariance matrix of ρ_{Ξ} :

$$[\hat{C}] = \left(\lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \boldsymbol{\xi}_0) \otimes \mathbf{j}_j(\omega_k; \boldsymbol{\xi}_0) + [I] \right)^{-1}, \quad (68)$$

in which

$$\mathbf{j}_j(\omega_k; \boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}} \log |h_j(\omega_k; \boldsymbol{\xi})|. \quad (69)$$

We use for $\boldsymbol{\xi}_0$ a mode of the posterior, namely, a value that maximizes ρ_{Ξ} , which we compute with a quasi-Newton method:

$$\begin{aligned} & \left(\lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \boldsymbol{\xi}^{(\ell)}) \otimes \mathbf{j}_j(\omega_k; \boldsymbol{\xi}^{(\ell)}) + [I] \right) \boldsymbol{\xi}^{(\ell+1)} \\ &= -\lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \boldsymbol{\xi}^{(\ell)}) \left(\log |h_j(\omega_k; \boldsymbol{\xi}^{(\ell)})| - \mathbf{j}_j(\omega_k; \boldsymbol{\xi}^{(\ell)}) \cdot \boldsymbol{\xi}^{(\ell)} - \log |h_j^{\text{obs}}(\omega_k)| \right). \end{aligned} \quad (70)$$

The Itô-SDE-based MCMC method then follows from setting the potential ϕ equal to the negative logarithm of the PDF ρ_{Ξ} defined in (65), injecting this potential ϕ into the Itô SDE in (14), and discretizing this Itô SDE in time, thus leading to the time-discrete equations in (16) with

$$\phi(\boldsymbol{\xi}) = \lambda \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \left| \log |h_j^{\text{obs}}(\omega_k)| - \log |h_j(\omega_k; \boldsymbol{\xi})| \right|^2 + \frac{1}{2} \|\boldsymbol{\xi}\|^2 \quad (71)$$

and

$$\nabla_{\boldsymbol{\xi}} \phi(\boldsymbol{\xi}) = \lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \boldsymbol{\xi}) \left(\log |h_j(\omega_k; \boldsymbol{\xi})| - \log |h_j^{\text{obs}}(\omega_k)| \right)^2 + \boldsymbol{\xi}. \quad (72)$$

When completed with an initial condition, these time-discrete equations define a Markov chain $\{(\boldsymbol{\Xi}^{(\ell)}, \mathbf{H}^{(\ell)}), \ell \in \mathbb{N}\}$. The transformation of measure associates the chain of random matrices $\{[\mathbf{M}^{(\ell)}], [\mathbf{D}^{(\ell)}], [\mathbf{K}^{(\ell)}], \ell \in \mathbb{N}\}$ with the component chain of random variables $\{\boldsymbol{\Xi}^{(\ell)}, \ell \in \mathbb{N}\}$:

$$[\mathbf{M}^{(\ell)}] = [\mathcal{M}(\boldsymbol{\Xi}_M^{(\ell)})], \quad [\mathbf{D}^{(\ell)}] = [\mathcal{D}(\boldsymbol{\Xi}_D^{(\ell)})], \quad [\mathbf{K}^{(\ell)}] = [\mathcal{K}(\boldsymbol{\Xi}_K^{(\ell)})]. \quad (73)$$

As for (18), trajectories of this chain of random matrices can be construed as samples drawn from the posterior and hence used as samples from the random matrices in the Monte Carlo simulation with the updated probabilistic reduced-order model.

4.9. Identification of the hyperparameters

The posterior is parameterized by the dispersion parameters δ_M , δ_D , and δ_K and the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$. These hyperparameters control the dispersion of the prior and the weights of the weighted distance that represents the model-data misfit in the likelihood function. The smaller the values of δ_M , δ_D , and δ_K , the more the posterior concentrates its weight predominantly in the vicinity of the values of the reduced matrices in the reduced-order model in (50). The larger the values of the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$, the more the posterior concentrates its weight predominantly in the vicinity of values of the reduced matrices that minimize the weighted distance that represents the model-data misfit. Thus, especially if there are no values of the reduced matrices that result in a perfect representation of the dynamical behavior of the structure and some amount of dispersion should persist to serve as a representation of the model uncertainties, the dispersion parameters δ_M , δ_D , and δ_K cannot be too small and the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$ cannot be too large.

If the data are perturbed by noise, estimates of the significance of the noise can play a role in the identification of the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$. For instance, smaller weights μ_{jk} can be given to FRF samples that are perturbed more significantly by noise.

In a manner that builds on the method that had already been proposed for the identification of the dispersion parameters in the context of the non-parametric probabilistic approach, we use for the identification of the hyper-parameters of the posterior a method based on an identification criterion. In this method, we plot the data on the same axes as high-probability confidence regions for the predictions of the updated probabilistic reduced-order model, and we seek a combination of ideally small values of the dispersion parameters δ_M , δ_D , and δ_K and ideally large values of the weights λ and μ_{jk} , $1 \leq j \leq d$, $1 \leq k \leq m$ for which the data are mostly contained within the high-probability confidence regions for the predictions.

4.10. Adjoint-based computation of the requisite gradients

The approximation of the covariance matrix in (68) in the change of variables requires the computation of gradients as in (69), which are also required for the computation of the gradient of the potential in (72) in the Itô-SDE-based MCMC method. We use for these computations of these gradients an adjoint-based approach that requires for each frequency and for each FRF one direct solve of the reduced-order model and three adjoint solves associated with the reduced-order model. While the computations of the partial derivatives with respect to components of $\boldsymbol{\xi}_M$ and $\boldsymbol{\xi}_D$ can be carried out analogously, we provide below details about the computation of the partial derivatives with respect to components $\xi_{\ell\ell'}^K$ of $\boldsymbol{\xi}_K$.

One way of obtaining the adjoint-based computation starts from writing the partial derivative of $\log |h_j(\omega_k; [M], [D], [K])|$ with respect to $\xi_{\ell\ell'}^K$ as

$$\begin{aligned} & \partial_{\xi_{\ell\ell'}^K} \log |h_j(\omega_k; [M], [D], [K])| \\ &= \frac{1}{|h_j(\omega_k; [M], [D], [K])|} \partial_{\xi_{\ell\ell'}^K} |h_j(\omega_k; [M], [D], [K])| \\ &= \frac{1}{|h_j(\omega_k; [M], [D], [K])|^2} \text{Re}(\overline{\partial_{\xi_{\ell\ell'}^K} h_j(\omega_k; [M], [D], [K])} h_j(\omega_k; [M], [D], [K])), \end{aligned} \quad (74)$$

in which we denote by $\text{Re}(a + bi) = a$ the real part and by $\overline{a + bi} = a - bi$ the complex conjugate of a complex number $a + bi$. The partial derivative of $h_j(\omega_k; [M], [D], [K])$ with respect to $\xi_{\ell\ell'}^K$ reads as

$$\begin{aligned} & \partial_{\xi_{\ell\ell'}^K} h_j(\omega_k; [M], [D], [K]) \\ &= -\mathbf{c}_j^T [\Phi] (-\omega_k^2 [M] + i\omega_k [D] + [K])^{-1} (\partial_{\xi_{\ell\ell'}^K} [K]) (-\omega_k^2 [M] + i\omega_k [D] + [K])^{-1} [\Phi]^T \mathbf{b}_j. \end{aligned} \quad (75)$$

Hence, with the chain rule and with (46) with (47)–(49), the partial deriva-

tives of $\log |h_j(\omega_k; \boldsymbol{\xi})|$ with respect to the components $\xi_{\ell\ell'}^K$ of $\boldsymbol{\xi}_K$ read as

$$\begin{aligned} & \partial_{\xi_{\ell\ell'}^K} \log |h_j(\omega_k; \boldsymbol{\xi})| \\ &= \frac{-1}{|h_j(\omega_k; \boldsymbol{\xi})|^2} \operatorname{Re}(b_{j\ell}(\omega_k; \boldsymbol{\xi}) \partial_{\xi_{\ell\ell'}^K} \mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}_K) f_{j\ell'}(\omega_k; \boldsymbol{\xi}) + c_{j\ell'}(\omega_k; \boldsymbol{\xi}) \partial_{\xi_{\ell\ell'}^K} \mathcal{L}_{\ell\ell'}(\boldsymbol{\xi}_K) e_{j\ell}(\omega_k; \boldsymbol{\xi})), \end{aligned} \quad (76)$$

in which

$$\mathbf{q}_j(\omega_k; \boldsymbol{\xi}) = (-\omega_k^2 [\mathcal{M}(\boldsymbol{\xi}_M)] + i\omega_k [\mathcal{D}(\boldsymbol{\xi}_D)] + [\mathcal{K}(\boldsymbol{\xi}_K)])^{-1} [\underline{\Phi}]^T \mathbf{b}_j, \quad (\text{direct}), \quad (77)$$

$$\mathbf{b}_j(\omega_k; \boldsymbol{\xi}) = [L_K]^T \overline{\mathbf{q}_j(\omega_k; \boldsymbol{\xi})}, \quad (78)$$

$$\mathbf{c}_j(\omega_k; \boldsymbol{\xi}) = [\mathcal{L}(\boldsymbol{\xi}_K)]^T \mathbf{b}_j(\omega_k; \boldsymbol{\xi}), \quad (79)$$

$$\mathbf{d}_j(\omega_k; \boldsymbol{\xi}) = (-\omega_k^2 [\mathcal{M}(\boldsymbol{\xi}_M)] + i\omega_k [\mathcal{D}(\boldsymbol{\xi}_D)] + [\mathcal{K}(\boldsymbol{\xi}_K)])^{-1} [\underline{\Phi}]^T \mathbf{c}_j h_j(\omega_k; \boldsymbol{\xi}), \quad (\text{adjoint}), \quad (80)$$

$$\mathbf{e}_j(\omega_k; \boldsymbol{\xi}) = [L_K]^T \mathbf{d}_j(\omega_k; \boldsymbol{\xi}), \quad (81)$$

$$\mathbf{f}_j(\omega_k; \boldsymbol{\xi}) = [\mathcal{L}(\boldsymbol{\xi}_K)]^T \mathbf{e}_j(\omega_k; \boldsymbol{\xi}). \quad (82)$$

5. Numerical illustration

We now present a numerical illustration of a data-driven updating of a reduced-order model as described in Sec. 4.

5.1. Problem setting and data

We considered the dynamical behavior of a clamped-clamped beam with a length of 10 m, a width of 1.2 m, and a height of 1.6 m. We assumed its constitutive behavior to be isotropic, linear, and dissipative. We assumed a Young modulus of 10 GPa, a Poisson ratio of 0.2, a mass density of 1600 kg m^{-3} , and a modal damping ratio of 0.01. The frequency band of analysis is $[0, 1200] \text{ Hz}$.

We generated (synthetic) data with a three-dimensional finite element model with a mesh of $100 \times 12 \times 16$ three-dimensional 8-node elements of equal size. For the purpose of the data-driven updating of the reduced-order model, we used the three-dimensional finite element model to compute the 15 FRFs that link the force driven by an actuator oriented along the transverse direction (along the height of the beam) and located at 4.75 m (from the leftmost cross section) to the displacements gauged by sensors oriented along the transverse direction and located at equally spaced locations from 0.588 m to 4.116 m with a step of 0.588 m and from 5.294 m to 9.412 m with a step of 0.588 m. This sensor grid corresponds to setting up 16 equally spaced sensors and then leaving out the one closest to the actuator. In addition,

for the purpose of enabling a cross-validation, that is, to evaluate the predictive capabilities of the updated probabilistic reduced-order model, we used the three-dimensional finite element model to compute the FRF that links the force driven by an actuator oriented along the transverse direction and located at 5.6 m to the displacement at a sensor oriented along the transverse direction and located at 8.2 m. We sampled all these FRFs with a step of 0.1 Hz from 0 to 200 Hz and with a step of 1 Hz from 201 Hz to 1200 Hz.

5.2. *Reduced-order model*

We set up a one-dimensional finite element model based on the Timoshenko beam theory with a mesh of 200 elements of equal size. We used a shear correction factor of 5/6. We purposely set up such a simple deterministic model in order for there to clearly be modeling errors. Figure 2 compares the data with the corresponding FRFs predicted by the deterministic model. We can observe that at lower frequencies, the FRFs predicted by the deterministic model fit the data well; however, at higher frequencies, there are significant differences, which can be explained as consequences of modeling errors entailed by the use of the Timoshenko beam theory, whose representation of the mechanical behavior involves simplifying hypotheses, such as the rigidity of the cross section.

[Figure 1 about here.]

We projected the deterministic model onto a reduced-order basis constituted of the dynamical eigenmodes associated with the $n = 50$ lowest eigenfrequencies of the deterministic model. The 50-th eigenpair corresponds to an eigenfrequency of about 2355 Hz (Fig. 1).

5.3. *Using a maximum-entropy PDF as the prior*

[Figure 2 about here.]

[Figure 3 about here.]

We built a nonparametric probabilistic model by representing the reduced stiffness matrix of the reduced-order model with a random matrix and leaving the reduced mass and damping matrices deterministic. The nonparametric probabilistic model thus obtained has one hyperparameter, namely, the dispersion parameter, δ_K , which must be identified. To do so, we applied the

aforementioned method based on an identification criterion that quantifies the agreement between the data and the predictions of the nonparametric probabilistic model. We set the value of the dispersion parameter equal to the smallest value that suffices for the data to be mostly contained within high-probability confidence regions for the predictions of the nonparametric probabilistic model. This approach led us to set the dispersion parameter equal to $\hat{\delta}_K = 0.25$. Figures 2 and 3 compare the data with the predictions of the nonparametric probabilistic model for $\delta_K = 0.05$ and $\hat{\delta}_K = 0.25$, respectively. All results obtained with the nonparametric probabilistic model were obtained with 5000 Monte Carlo samples.

5.4. *Dynamic test data*

The predictions of the deterministic model and those of the nonparametric probabilistic model fit the data more closely at low frequencies than at higher frequencies (Fig. 3). On the one hand, the closer fit at lower frequencies suggests the feasibility of adjusting the reduced-order model so as to improve the predictions at lower frequencies. On the other hand, the looser fit at higher frequencies suggests that the identification of the value of $\hat{\delta}_K = 0.25$ followed mostly from the need to increase this value sufficiently for the data at higher frequencies to be mostly contained within the high-probability confidence regions for the predictions of the nonparametric probabilistic model, with the data at lower frequencies playing only a lesser role. These considerations motivated us to seek to update the reduced-order model based on using as data the FRFs sampled at lower frequencies so as to improve the predictions at lower frequencies. Specifically, we used as data the 15 FRFs in the frequency range between 0 and 900 Hz, sampled with the aforementioned steps of 0.1 Hz below 200 Hz and 1 Hz above 200 Hz.

Because the reduced stiffness matrix is defined via a projection onto dynamical eigenmodes, the data must be sufficiently rich to allow modal contributions to be distinguished. We used the modal-analysis criterion (MAC) to verify that the 15 sensor locations suffice to distinguish modes that contribute significantly to predictions in the frequency range between 0 and 900 Hz.

5.5. *Bayesian identification of the reduced stiffness matrix*

Because the dimension of the reduced stiffness matrix is $n = 50$, casting the data-driven updating as a Bayesian identification of the reduced stiffness matrix is a Bayesian inverse problem of dimension $n(n + 1)/2 = 1275$.

In the formulation of the likelihood function, we set $\mu_{jk} = 2\pi \times 0.1$ for $1 \leq k \leq 2000$ and $\mu_{jk} = 2\pi \times 1$ for $2001 \leq k \leq m = 3001$ for $1 \leq j \leq \mu = 15$, thus assigning equal weight to all FRFs but compensating for the finer sampling below 200 Hz than above 200 Hz. Hence, the posterior has one hyperparameter left, namely, λ . To gain some insight into the range of values of λ of relevance in this particular problem, we carried out a preliminary study in which we linearized the inverse problem about the mode $\boldsymbol{\xi} = \mathbf{0}$ of the prior and then computed as in (68) for several values of λ approximations to the covariance matrix of ρ_{Ξ} . This preliminary study indicated that for the model-data misfit to result in a meaningful reduction in the variance of certain components, the relevant range of values of λ should be between about 1 and 100. In the following, we first provide results for $\lambda = 12.5$ and then address the identification of the hyperparameter λ .

5.6. Numerical implementation of the Bayesian inversion

[Figure 4 about here.]

For the change of variables, we used the quasi-Newton method as in (70) to determine a mode of the posterior to serve as reference value $\boldsymbol{\xi}_0$ (Fig. 4(a)). We linearized the inverse problem about $\boldsymbol{\xi}_0$ to obtain first the sensitivity matrix $\sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \dot{\mathbf{j}}_j(\omega_k; \boldsymbol{\xi}_0) \otimes \dot{\mathbf{j}}_j(\omega_k; \boldsymbol{\xi}_0)$ and then as in (68) an approximation $[\hat{C}]$ to the covariance matrix of ρ_{Ξ} . The rapid decay of the spectrum of the sensitivity matrix (Fig. 4(b)), as well as certain diagonal entries of $[\hat{C}]$ being significantly reduced as compared with their unit values under the prior and other diagonal entries of $[\hat{C}]$ being rather unchanged (Fig. 4(c)), indicate that the data provide information on certain components but little or no information on others, thus highlighting the importance of the change of variables. The components that undergo the strongest reduction in variance are those with indices such as 1, 51, 100, 148, and so forth (Fig. 4(c)); the transformation of measure links these components to diagonal entries of the Cholesky factor of the reduced stiffness matrix that contribute significantly to diagonal entries of the reduced stiffness matrix and hence have a significant effect on the locations of resonances and anti-resonances of the FRFs. Components with indices larger than about 750 undergo no change in variance (Fig. 4(c)); these components are linked to entries of the Cholesky factor of the reduced stiffness matrix that have an effect mostly on the behavior of the FRFs at frequencies higher than 900 Hz and hence have only little or no impact on the model-data misfit.

[Figure 5 about here.]

For the Itô-SDE-based MCMC method, we obtained all results to follow with $\zeta_0 = 4$, $\Delta t = 0.1$, and 5000 steps in the Markov chain. Figure 5 shows the trajectories of the 51-st component and the 1000-th component of the component chain of random variables $\{\Xi_K^{(\ell)}, 1 \leq \ell \leq 5000\}$. We can observe that the trajectory of the 51-st component, which Fig. 4(c) indicated the data to provide information on, evolves about a central value that is larger than that under the prior and with a level of variability that is reduced as compared with that under the prior; by contrast, the trajectory of the 1000-th component, which Fig. 4(c) indicated the data to provide little or no information on, evolves about a central value and with a level of variability that are unchanged as compared with those under the prior.

[Figure 6 about here.]

We determined as in (73) the trajectory of the chain of random reduced stiffness matrices $\{\mathbf{K}^{(\ell)}, 1 \leq \ell \leq 5000\}$ corresponding to the component chain of random variables $\{\Xi_K^{(\ell)}, 1 \leq \ell \leq 5000\}$ and used it as samples from the random reduced stiffness matrix in the Monte Carlo simulation with the updated probabilistic reduced-order model. Figure 6 compares the predictions of the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$ with the data. We can observe that as compared with the predictions of the nonparametric probabilistic model in Fig. 3, the predictions of the updated probabilistic reduced-order model in Fig. 6 fit the data significantly more closely at the frequencies in the frequency range between 0 and 900 Hz that we addressed in the updating.

5.7. Identification of the hyperparameter

[Figure 7 about here.]

To identify the hyperparameter λ , we applied the aforementioned method based on an identification criterion that quantifies the agreement between the data and the predictions of the updated probabilistic reduced-order model. Whereas increasing δ_K can be expected to widen the high-probability confidence regions for the predictions, increasing λ can be expected to narrow the high-probability confidence regions for the predictions. We thus sought to set λ equal to the largest value for which the data remain mostly contained

within high-probability confidence regions for the predictions. This approach led us to set $\hat{\lambda} = 37.5$. Figure 7 compares the data with the predictions of the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\hat{\lambda} = 37.5$.

5.8. Cross-validation

[Figure 8 about here.]

As a cross-validation, we used the nonparametric probabilistic model and the updated probabilistic reduced-order model to predict the FRF that links the force driven by the actuator oriented along the transverse direction and located at 5.6 m to the displacement gauged by the sensor oriented along the transverse direction and located at 8.2 m, which we hadn't used yet (Fig. 8). We can observe that the predictions of both probabilistic models agree with the data and that as compared with the predictions of the nonparametric probabilistic model, the predictions of the updated probabilistic reduced-order model fit the data significantly more closely at the frequencies in the frequency range between 0 and 900 Hz that we addressed in the updating.

5.9. Random reduced mass and stiffness matrices

[Figure 9 about here.]

We repeated the previous study, representing, this time, however, not only the reduced stiffness matrix but also the reduced mass matrix with a random matrix. Figure 9 presents the cross-validation result of this study. The comparison with Fig. 8 indicates that the representation of the reduced mass matrix with a random matrix does not have a significant benefit in terms of further improving the predictions in this particular problem.

6. Conclusion

We proposed a probabilistic methodology for data-driven updating of non-Gaussian high-dimensional symmetric positive-definite matrices involved in parameterized computational models. This methodology is based on the application of the Bayes formula to update the prior probability model by using data in order to obtain the posterior for the high-dimensional symmetric positive-definite random matrices. The hyperparameters of the posterior are identified by using an identification criterion that quantifies the agreement between the predictions and the data. Samples from the posterior are

generated with an Itô-SDE-based MCMC method that leverages the gradient of the posterior for computational efficiency.

We applied the proposed methodology to a problem of a data-driven updating of a reduced-order model in computational structural dynamics. The proposed methodology allowed a direct identification of the reduced matrices from data with an informative prior. The data-driven update resulted in an updated probabilistic reduced-order model that provided improved predictions as well as a probabilistic representation of model uncertainties. In this application, the proposed methodology could also be seen as a data-driven updating of a nonparametric probabilistic model. Taking into account the difficulty of the problem that has been solved, the proposed method yields very good results.

There are many interesting directions for future work. The proposed methodology could be extended to other types of computational model, to other types of informative prior, to other types of likelihood function by changing the distance that represents the model-data misfit and/or by changing the Gaussian model of the noise on the observations, as well as to other approaches for the identification of the hyperparameters. In particular, it would be interesting to revisit the approaches that have already been proposed for the identification of the dispersion parameters of the prior probability distributions in the nonparametric probabilistic approach. In addition, whereas the proposed methodology has been applied in computational structural dynamics with synthetic data, it would be interesting to explore the application to more complex computational models with experimental data.

Acknowledgements

Part of this work was carried out while the first author was visiting the second author at the Laboratoire Modélisation et Simulation Multi Echelle of the Université Paris-Est Marne-la-Vallée. Financial support of the Fund for Scientific Research (FNRS) is gratefully acknowledged.

References

- [1] A. Tarantola, Inverse Problem Theory, SIAM, Philadelphia, United States, 2005.
- [2] H. Flath, L. Wilcox, A. Akçelik, J. Hill, B. van Bloemen Waanders, O. Ghattas, Fast algorithms for Bayesian uncertainty quantification in

- large-scale linear inverse problems based on low-rank partial Hessian approximations, *SIAM Journal on Scientific Computing* 33 (2015) 407–432, DOI: 10.1137/090780717.
- [3] M. Dashti, A. Stuart, The Bayesian approach to inverse problems, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), *Handbook of Uncertainty Quantification*, Springer, Cham, Switzerland, 2017, Ch. 10, pp. 311–428.
 - [4] C. Robert, G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2010.
 - [5] C. Soize, Nonparametric model of random uncertainties for reduced matrix models in structural dynamics, *Probabilistic Engineering Mechanics* 15 (2000) 277–294, DOI: 10.1016/S0266-8920(99)00028-4.
 - [6] C. Soize, Non-Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators, *Computer Methods in Applied Mechanics and Engineering* 195 (2006) 26–64, DOI: 10.1016/j.cma.2004.12.014.
 - [7] C. Soize, A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics, *Journal of Sound and Vibration* 288 (2006) 623–652, DOI: 10.1016/j.jsv.2005.07.009.
 - [8] H. Chebli, C. Soize, Experimental validation of a nonparametric probabilistic model of nonhomogeneous uncertainties for dynamical systems, *Journal of the Acoustical Society of America* 115 (2004) 697–705, DOI: 10.1121/1.1639335.
 - [9] M. Arnst, D. Clouteau, H. Cheli, R. Othman, G. Degrande, A non-parametric probabilistic model for ground-borne vibrations in buildings, *Probabilistic Engineering Mechanics* 21 (2006) 18–34, DOI: 10.1016/j.probengmech.2005.06.004.
 - [10] R. Cotteneau, D. Clouteau, C. Soize, Construction of a probabilistic model for impedance matrices, *Computer Methods in Applied Mechanics and Engineering* 196 (2007) 2252–2268, DOI: 10.1016/j.cma.2006.12.001.

- [11] M. Mignolet, C. Soize, Nonparametric stochastic modeling of linear systems with prescribed variance of several natural frequencies, *Probabilistic Engineering Mechanics* 23 (2008) 267–278, DOI: 10.1016/j.probengmech.2007.12.027.
- [12] C. Soize, I. Poloskov, Time-domain formulation in computational dynamics for linear viscoelastic media with model uncertainties and stochastic excitation, *Computer and Mathematics with Applications* 64 (2012) 3594–3612, DOI: 10.1016/j.camwa.2012.09.010.
- [13] A. Batou, C. Soize, Rigid multibody system dynamics with uncertain rigid bodies, *Multibody System Dynamics* 27 (2012) 285–319, DOI: 10.1007/s11044-011-9279-2.
- [14] J. Guillemot, C. Soize, Stochastic model and generator for random fields with symmetry properties: application to the mesoscopic modeling of elastic random media, *Multiscale Modeling & Simulation* 11 (2013) 840–870, DOI: 10.1137/120898346.
- [15] C. Soize, Construction of probability distributions in high dimension using the maximum entropy principle: Applications to stochastic processes, random fields and random matrices, *International Journal for Numerical Methods in Engineering* 76 (2008) 1583–1611, 10.1002/nme.2385.
- [16] M. Arnst, B. Abello Álvarez, J. Ponthot, R. Boman, Itô-SDE MCMC method for Bayesian characterization of errors associated with data limitations in stochastic expansion methods for uncertainty quantification, *Journal of Computational Physics* 349 (2017) 59–79, 10.1016/j.jcp.2017.08.005.
- [17] C. Soize, E. Capiez-Lernout, J.-F. Durand, C. Fernandez, L. Gagliardini, Probabilistic model identification of uncertainties in computational models for dynamical systems and experimental validation, *Computer Methods in Applied Mechanics and Engineering* 198 (2008) 150–163, DOI: 10.1016/j.cma.2008.04.007.
- [18] M. Arnst, D. Clouteau, M. Bonnet, Inversion of probabilistic structural models using measured transfer functions, *Computer Methods*

- in *Applied Mechanics and Engineering* 197 (2008) 589–608, DOI: 10.1016/j.cma.2007.08.011.
- [19] J. Liu, *Monte Carlo strategies in scientific computing*, Springer, New York, New York, 2008.
 - [20] Y. Marzouk, H. Najm, L. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, *Journal of Computational Physics* 224 (2007) 560–586, DOI: 10.1016/j.jcp.2006.10.010.
 - [21] M. Parno, Y. Marzouk, Transport map accelerated Markov Chain Monte Carlo, *SIAM Journal on Uncertainty Quantification* 6 (2018) 645–682, DOI: 10.1137/17M1134640.
 - [22] E. Hairer, C. Lubich, G. Wanner, Geometric numerical integration illustrated by the Störmer–Verlet method, *Acta Numerica* 12 (2003) 399–450, DOI: 10.1017/S0962492902000144.
 - [23] C. Soize, R. Ghanem, Data-driven probability concentration and sampling on manifold, *Journal of Computational Physics* 321 (2016) 242–258, DOI: 10.1016/j.jcp.2016.05.044.

List of Figures

1	Reduced-order model: 50 lowest eigenfrequencies of the deterministic model.	36
2	Using a maximum-entropy PDF as the prior: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the nonparametric probabilistic model for $\delta_K = 0.05$ (light blue filled region). Please refer to the online version for color figures.	37
3	Using a maximum-entropy PDF as the prior: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ (light blue filled region). Please refer to the online version for color figures.	38
4	Numerical implementation of the Bayesian inversion: (a) reference value ξ_0 , (b) spectrum of the sensitivity matrix given by $\sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \dot{\mathbf{j}}_j(\omega_k; \xi_0) \otimes \dot{\mathbf{j}}_j(\omega_k; \xi_0)$, and (c) diagonal entries of $[\hat{C}] = (\lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \dot{\mathbf{j}}_j(\omega_k; \xi_0) \otimes \dot{\mathbf{j}}_j(\omega_k; \xi_0) + [I])^{-1}$ for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$	39
5	Numerical implementation of the Bayesian inversion: trajectories of (a) the 51-st component and (b) the 1000-th component of the component chain of random variables $\{\Xi_K^{(\ell)}, 1 \leq \ell \leq 5000\}$ for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$	40
6	Numerical implementation of the Bayesian inversion: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$ (blue filled region). Please refer to the online version for color figures.	41
7	Identification of the hyperparameter: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\hat{\lambda} = 37.5$ (blue filled region). Please refer to the online version for color figures.	42

8	Cross-validation: data (green solid line) and 98%-confidence region for the corresponding FRF predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ (light blue filled region) and the updated probabilistic reduced-order model with $\hat{\delta}_K = 0.25$ and $\hat{\lambda} = 37.5$ (blue filled region). Please refer to the online version for color figures.	43
9	Random reduced mass and stiffness matrices: data (green solid line) and 98%-confidence region for the corresponding FRF predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ and $\hat{\delta}_M = 0.10$ (light blue filled region) and the updated probabilistic reduced-order model with $\hat{\delta}_K = 0.25$ and $\hat{\delta}_M = 0.10$ and $\hat{\lambda} = 50$ (blue filled region). Please refer to the online version for color figures.	44

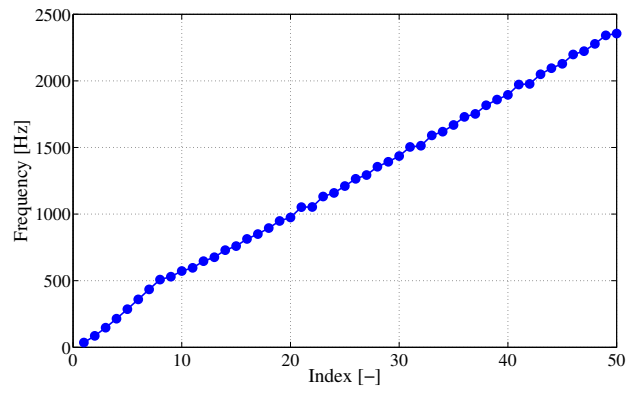
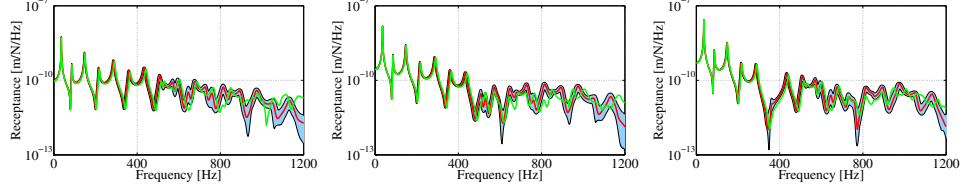


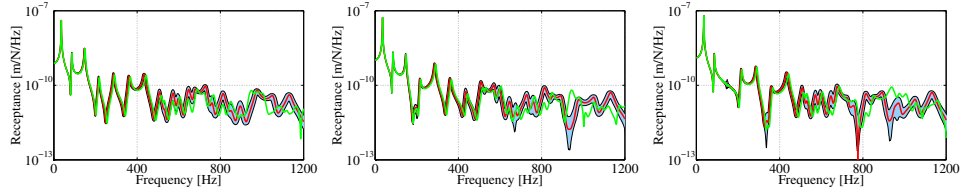
Figure 1: Reduced-order model: 50 lowest eigenfrequencies of the deterministic model.



(a) Sensor at 0.588 m.

(b) Sensor at 1.177 m.

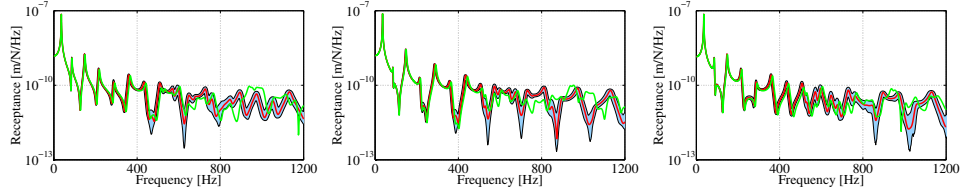
(c) Sensor at 1.765 m.



(d) Sensor at 2.353 m.

(e) Sensor at 2.941 m.

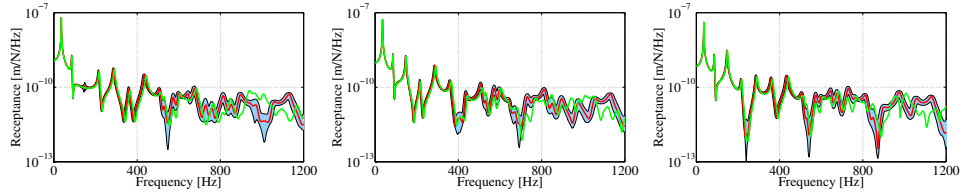
(f) Sensor at 3.529 m.



(g) Sensor at 4.118 m.

(h) Sensor at 5.294 m.

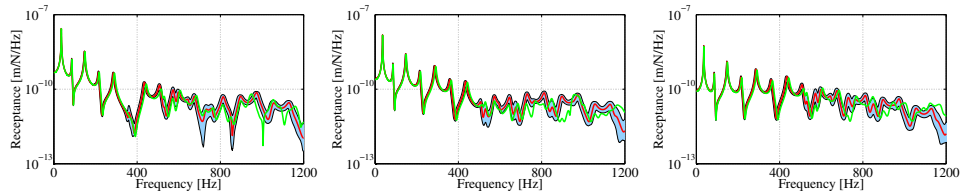
(i) Sensor at 5.882 m.



(j) Sensor at 6.471 m.

(k) Sensor at 7.059 m.

(l) Sensor at 7.647 m.

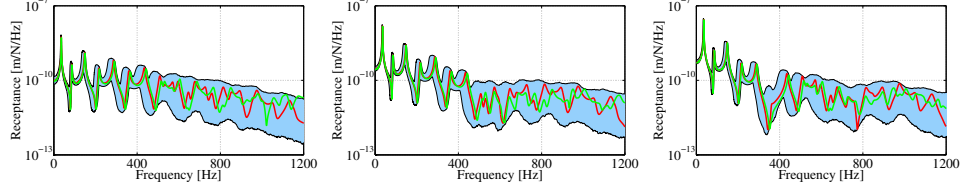


(m) Sensor at 8.253 m.

(n) Sensor at 8.824 m.

(o) Sensor at 9.412 m.

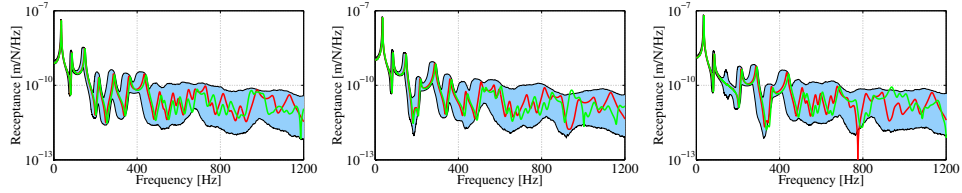
Figure 2: Using a maximum-entropy PDF as the prior: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the nonparametric probabilistic model for $\delta_K = 0.05$ (light blue filled region). Please refer to the online version for color figures.



(a) Sensor at 0.588 m.

(b) Sensor at 1.177 m.

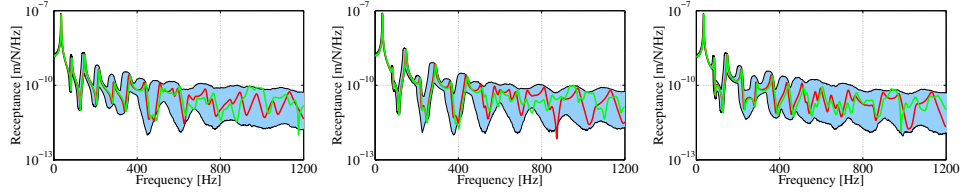
(c) Sensor at 1.765 m.



(d) Sensor at 2.353 m.

(e) Sensor at 2.941 m.

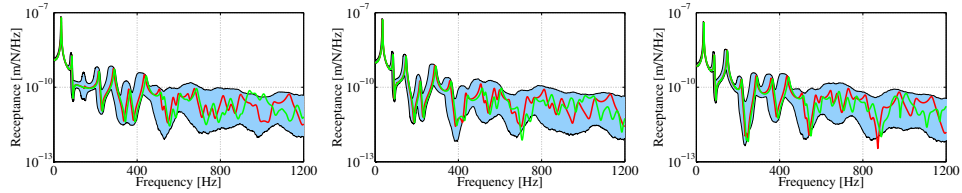
(f) Sensor at 3.529 m.



(g) Sensor at 4.118 m.

(h) Sensor at 5.294 m.

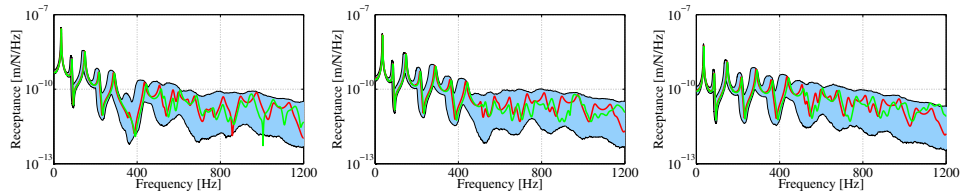
(i) Sensor at 5.882 m.



(j) Sensor at 6.471 m.

(k) Sensor at 7.059 m.

(l) Sensor at 7.647 m.

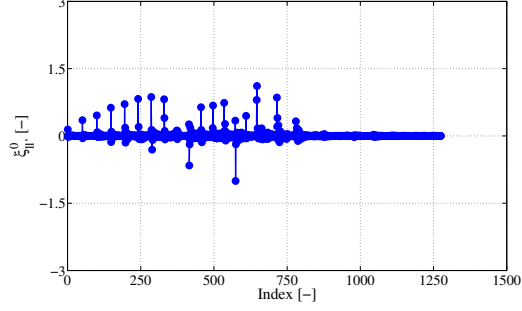


(m) Sensor at 8.253 m.

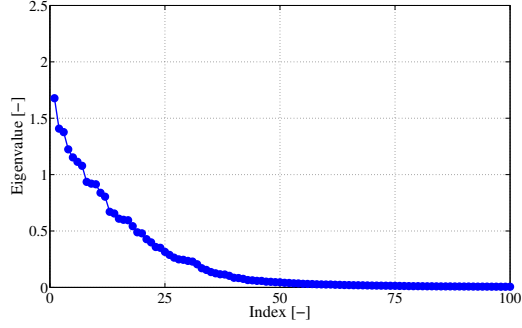
(n) Sensor at 8.824 m.

(o) Sensor at 9.412 m.

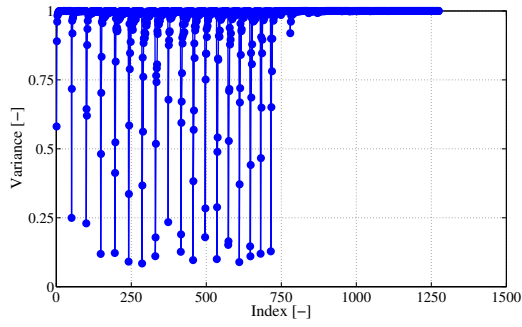
Figure 3: Using a maximum-entropy PDF as the prior: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ (light blue filled region). Please refer to the online version for color figures.



(a) Reference value ξ_0 .

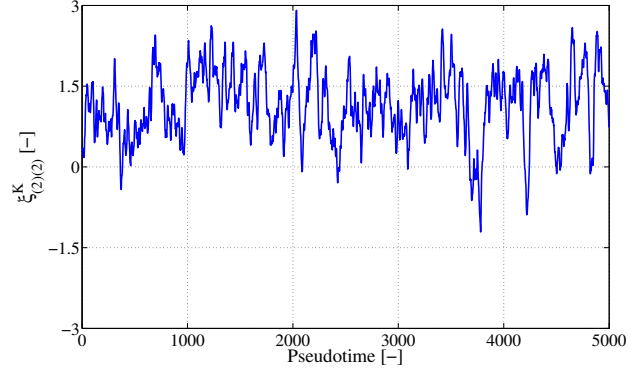


(b) Largest eigenvalues of sensitivity matrix.

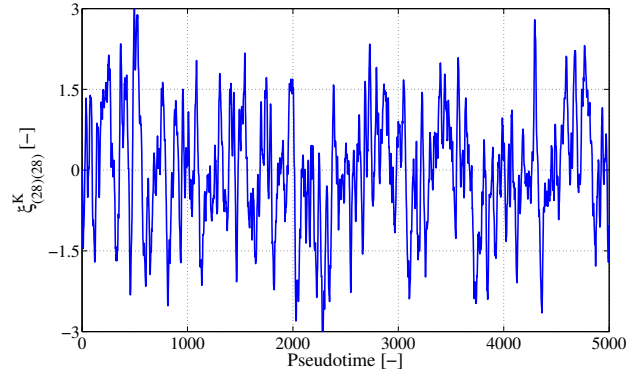


(c) Diagonal entries of $[\hat{C}]$.

Figure 4: Numerical implementation of the Bayesian inversion: (a) reference value ξ_0 , (b) spectrum of the sensitivity matrix given by $\sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \xi_0) \otimes \mathbf{j}_j(\omega_k; \xi_0)$, and (c) diagonal entries of $[\hat{C}] = (\lambda \sum_{j=1}^d \sum_{k=1}^m \mu_{jk} \mathbf{j}_j(\omega_k; \xi_0) \otimes \mathbf{j}_j(\omega_k; \xi_0) + [I])^{-1}$ for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$.

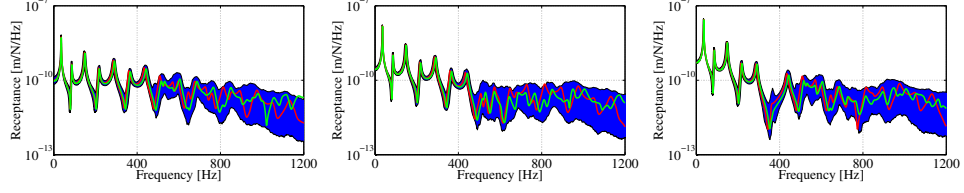


(a) Trajectory of 51-st component.

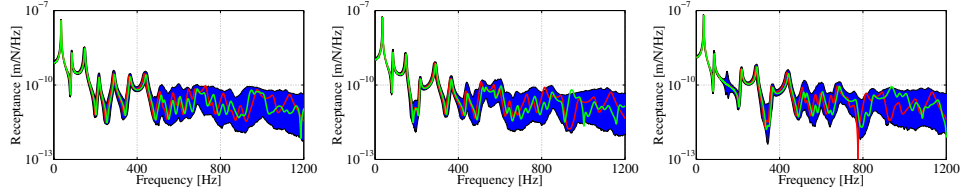


(b) Trajectory 1000-th component.

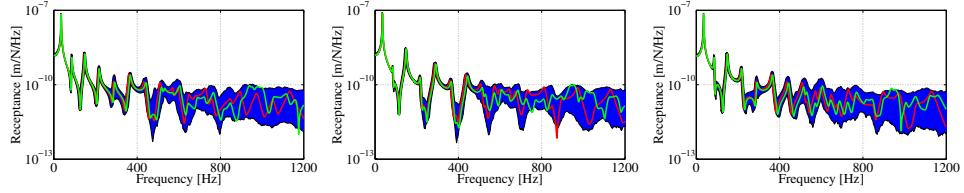
Figure 5: Numerical implementation of the Bayesian inversion: trajectories of (a) the 51-st component and (b) the 1000-th component of the component chain of random variables $\{\Xi_K^{(\ell)}, 1 \leq \ell \leq 5000\}$ for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$.



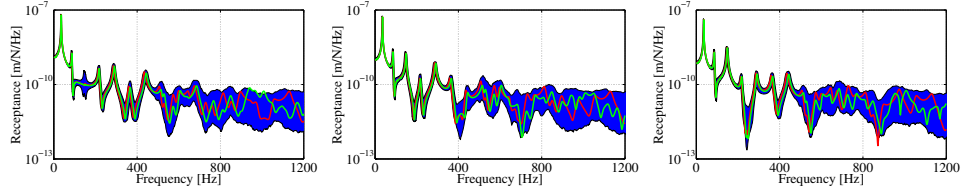
(a) Sensor at 0.588 m. (b) Sensor at 1.177 m. (c) Sensor at 1.765 m.



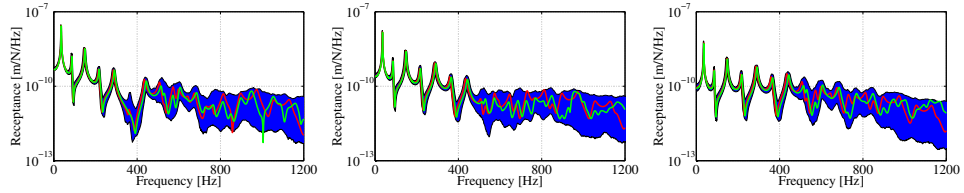
(d) Sensor at 2.353 m. (e) Sensor at 2.941 m. (f) Sensor at 3.529 m.



(g) Sensor at 4.118 m. (h) Sensor at 5.294 m. (i) Sensor at 5.882 m.

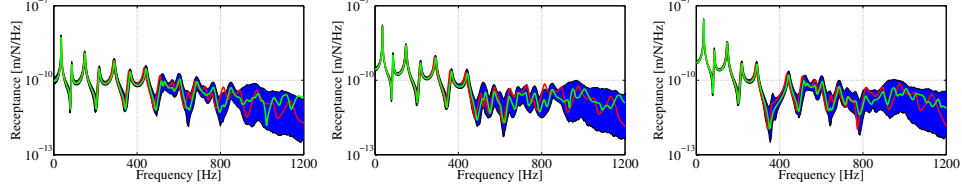


(j) Sensor at 6.471 m. (k) Sensor at 7.059 m. (l) Sensor at 7.647 m.



(m) Sensor at 8.253 m. (n) Sensor at 8.824 m. (o) Sensor at 9.412 m.

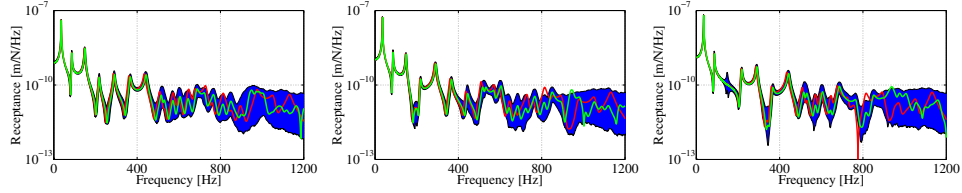
Figure 6: Numerical implementation of the Bayesian inversion: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\lambda = 12.5$ (blue filled region). Please refer to the online version for color figures.



(a) Sensor at 0.588 m.

(b) Sensor at 1.177 m.

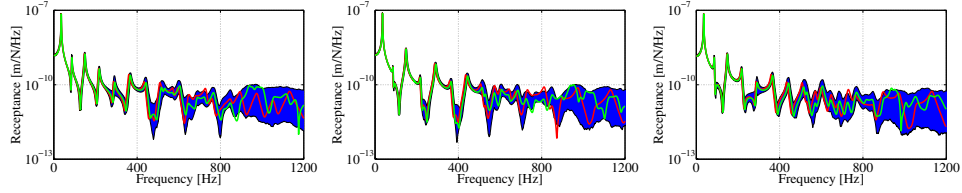
(c) Sensor at 1.765 m.



(d) Sensor at 2.353 m.

(e) Sensor at 2.941 m.

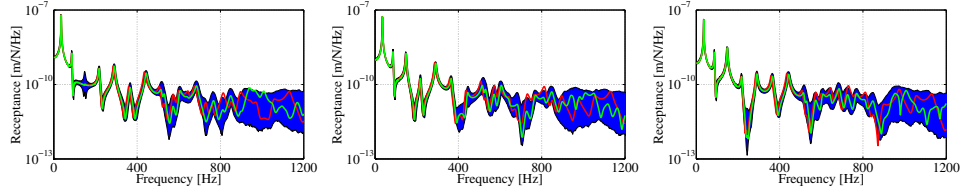
(f) Sensor at 3.529 m.



(g) Sensor at 4.118 m.

(h) Sensor at 5.294 m.

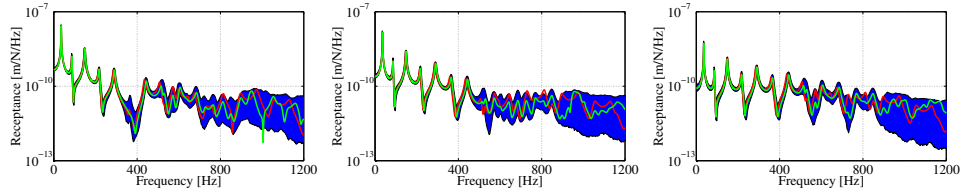
(i) Sensor at 5.882 m.



(j) Sensor at 6.471 m.

(k) Sensor at 7.059 m.

(l) Sensor at 7.647 m.



(m) Sensor at 8.253 m.

(n) Sensor at 8.824 m.

(o) Sensor at 9.412 m.

Figure 7: Identification of the hyperparameter: data (green solid line), corresponding FRFs predicted by the deterministic model (red solid line), and 98%-confidence regions for the corresponding FRFs predicted by the updated probabilistic reduced-order model for $\hat{\delta}_K = 0.25$ and $\hat{\lambda} = 37.5$ (blue filled region). Please refer to the online version for color figures.

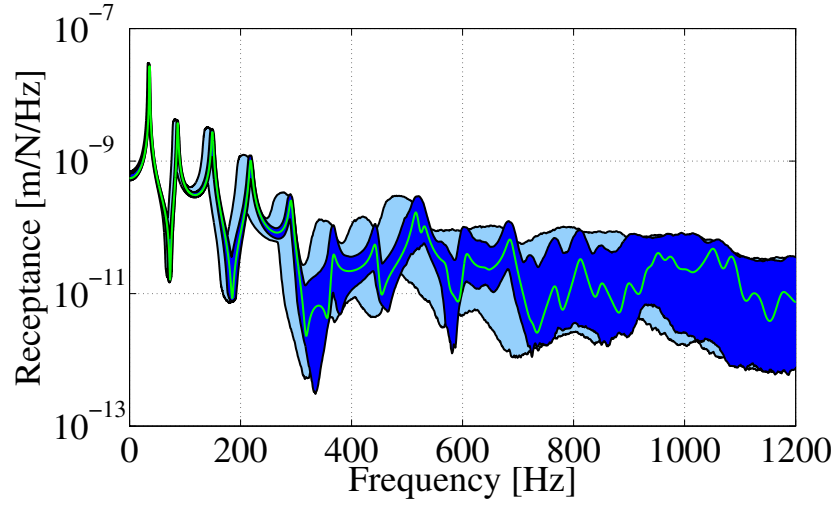


Figure 8: Cross-validation: data (green solid line) and 98%-confidence region for the corresponding FRF predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ (light blue filled region) and the updated probabilistic reduced-order model with $\hat{\delta}_K = 0.25$ and $\hat{\lambda} = 37.5$ (blue filled region). Please refer to the online version for color figures.

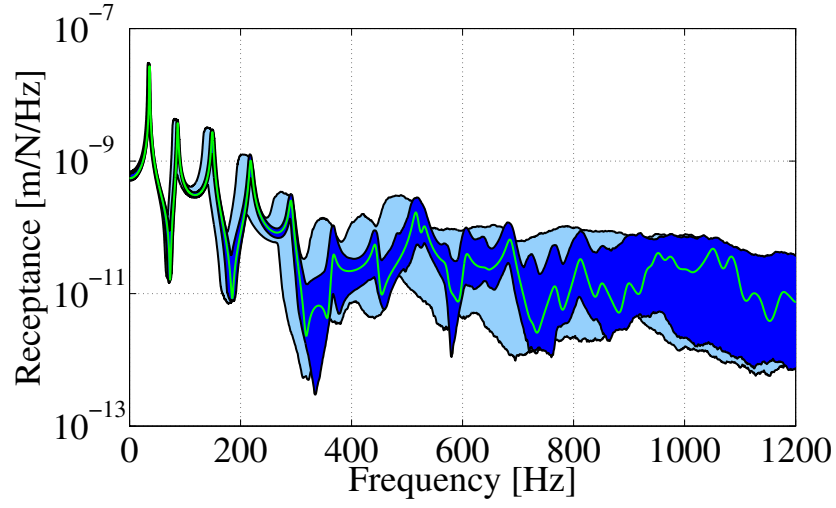


Figure 9: Random reduced mass and stiffness matrices: data (green solid line) and 98%-confidence region for the corresponding FRF predicted by the nonparametric probabilistic model for $\hat{\delta}_K = 0.25$ and $\hat{\delta}_M = 0.10$ (light blue filled region) and the updated probabilistic reduced-order model with $\hat{\delta}_K = 0.25$ and $\hat{\delta}_M = 0.10$ and $\hat{\lambda} = 50$ (blue filled region). Please refer to the online version for color figures.