Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints

Christian Soize

Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints

Christian Soize[a,*]

[a] *Université Gustave Eiffel, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France*

Abstract

This paper deals with a probabilistic learning inference that allows for integrating data (target set) into predictive models for which the target set is constituted of statistical moments of the quantity of interest (QoI) and for which the training set is constituted of a small number of points. In a first part, we present a mathematical analysis of a general methodology that allows for estimating a posterior probability model for a stochastic boundary value problem from a prior probability model of a control parameter. The given targets are statistical moments of the QoI for which the underlying realizations are not available. Under these conditions, the Kullback-Leibler minimum principle is used for estimating the posterior probability measure. The constraints are represented by an implicit nonlinear mapping for which a statistical surrogate model is introduced. The MCMC generator and the necessary numerical elements are given to facilitate the implementation of the methodology. In a second part, an application is presented to illustrate the proposed theory and is also, as such, a contribution to the three-dimensional stochastic homogenization of heterogeneous linear elastic media in the case of a non-separation of the microscale and macroscale. Consequently, the macroscale is another mesoscale at larger scale with random effective/apparent elastic properties. For the construction of the posterior probability measure by using the probabilistic learning inference, in addition to the constraints defined by given statistical moments of the random effective/apparent elasticity tensor, the second-order moment of the random normalized residue of the stochastic partial differential equation has been added as a constraint. This constraint guarantees that the algorithm seeks to bring the statistical moments closer to their targets while preserving a small residue.

Keywords: Probabilistic learning, statistical inverse problem, Kullback-Leibler divergence, implicit constraints, stochastic homogenization, uncertainty quantification

1. Introduction

(i) Objective of the paper. This paper deals with a very important question related to probabilistic learning algorithms that allow for integrating data (target set) into predictive models for which the training set is constituted of a small number of points and for which the target set is made up of statistical moments of some quantities of interest (QoI). The considered constraints are thus implicit. It is assumed that these statistical moments such as mean values, second-order moments, have been estimated with realizations (samples) that are no longer available. This situation occurs quite frequently when the data (the realizations/samples) have been lost, or deteriorated, or not commented on, or no longer exist, or are not accessible, but for which the values of some statistical moments have been published or are given in technical reports. As the experimental measurements remain very expensive, and many have been done in the past decades in many fields of physics and engineering sciences, it is very interesting to integrate into the models these experiments to which we know only statistical moments. Additional physics-based constraints can simultaneously considered. This is the case if we want the learning process to be also controlled by the model, that is to say, that the learned probability measure minimizes, for instance, the mean-square norm of the random normalized residue

*Corresponding author: C. Soize, christian.soize@univ-eiffel.fr
Email address: christian.soize@univ-eiffel.fr (C. Soize)

of the stochastic partial differential equation of the boundary value problem (BVP); this type of constraint is thus also implicit. We therefore consider a model driven by a stochastic BVP for which there are a vector-valued control parameter and a vector-valued quantity of interest (QoI) that is defined by a given transformation of the solution of the stochastic BVP. The graph of the functional dependence between the control parameter and the QoI defines a manifold that is not explicitly described. We consider the statistical inverse problem consisting in identifying the control parameter by giving the target set related to the QoI. A prior probabilistic model of the control parameter is introduced, which induces a random QoI. A training set, constituted of independent realizations belonging to the manifold, is constructed by solving the BVP. The framework of the developments presented is that for which the probabilistic learning inference is made using a computational model of which a single evaluation is assumed very high. Thus it is assumed that the training set consists of a small number of points as opposed to big data. The prior probability measure on the manifold is estimated with the points of the training set. The probabilistic learning inference consists in estimating the posterior probability model on the manifold, which is closest to the prior probability measure while satisfying the constraints defined by an implicit function $\boldsymbol{h}^c$ for which an algebraic representation is not available.

(ii) Ways for addressing the integration of data into the predictive models and methodology proposed in the paper. The literature related to the field treated is so vast that it would not be reasonable to attempt a review of the state of the art. We will therefore limit ourselves to mentioning the main statistical methods, which make it possible to integrate data into the predictive models in order to allow the positioning of the method that is proposed in this paper.

(ii-1) The Bayesian method provides a rational and efficient framework for integrating data into the predictive models. This statistical tool allows for estimating the posterior probability measure of the control parameter giving its prior probability measure and a target set of realizations for the QoI. Consequently, the Bayesian inference is certainly one of the most popular statistical tools to solve the statistical inverse problems in the context of parametric or nonparametric statistics (see, for instance, [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] for general aspects, [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22] for specific aspects related to statistical inverse problems, [23, 24] for variational Bayesian methods, [25] for Bayesian sequential inference, [26] for Bayesian inference for changepoint problems or [27] for Bayesian numerical homogenization). Bayesian inferences have also been considered in the framework of machine learning [28, 29] and probabilistic learning for small data sets and in high dimension [11]. Bayesian inference is therefore a powerful statistical tool for integrating raw data but requires that targets be given in the form of realizations, which is not the hypothesis introduced in this paper. Note also that Bayesian inference can remains tricky to use [30], in particular for the high dimension.

The maximum-likelihood method (see [4, 5, 3]) is also used for estimating the hyperparameters of probability measure from a target set of realizations (see also [31, 32, 33, 34, 35] for improvements of the basis methodology). This statistical tool also allows for solving statistical inverse problem in high dimension for non-Gaussian random field (see [36, 37]). Similarly to the Bayesian inference, the maximum-likelihood method is also a very effective statistical method for estimating hyperparameters of probability measures, but requires that targets be given in the form of realizations.

All the parameterized functional representations of random quantities in finite or in infinite dimension, such as the spectral approaches, allow for integrating data into predictive models by using the Bayesian inference, the maximum likelihood, or the least-square approach for identifying the parameters of the representations. In this context, it should be noted that a general parameterized representation of non-Gaussian second-order random quantities (vector-valued random variables, stochastic processes, and random fields) is given by the polynomial chaos expansions (see [38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52]).

(ii-2) In Artificial Intelligence (see [53] for a remarkable synthesis of this field), Machine learning has taken a considerable place in all fields of sciences and engineering and today provides an essential and very effective tool for many applications. This tool is mainly based on the use of Artificial Neural Networks (ANNs), which allow building an algebraic representation from data for both unsupervised and supervised cases. By construction, this method makes it possible to integrate data. In this context, physics-informed machine learning has seen a growing interest in recent years. Numerous works have been carried out in computational science and engineering [54], for instance in the field of computational fluid dynamics [55, 56, 57, 58], for projection-based model reduction [59], for hierarchical materials and their performance using multiresolution analysis and clustering discretization [60, 61]. Likewise, physics-informed machine learning has also been used in the context of linear [62], quasilinear [63], and nonlinear partial differential equations [64, 65, 66]. It should be noted that the use of ANNs requires the availability

of big data in order to be able to identify the network parameters by learning algorithms such as the deep learning [67] and therefore, to obtain a predictive ANN. As we have previously explained, for the probabilistic learning that is presented in this paper, it is assumed that the training set is made up of a small number of points and consequently, these types of approaches are, in general, not available.

The statistical and probabilistic learning approaches deals with the problem of inferring a function based on data. In particular for supervised learning, it consists in training from a set of data consisting of input-output pairs. These methods have also been extensively developed (see for instance, [68, 69, 70, 71, 72, 73, 74, 75, 76, 77]) and play an increasingly important role in computational physics and engineering science, in particular for problems using large scale computational models. In this framework, statistical methods for supervised learning have been developed in the form of surrogate models from which approximations of the expensive functions can be easily evaluated [78, 79, 80, 81, 82, 83, 84, 85, 86]. The probabilistic learning is the framework of this paper as we have explained in paragraph (i).

(ii-3) In the context of probabilistic learning and Information Theory, the Kullback-Leibler minimum principle [87, 88, 89] is a method that allows for estimating the posterior probability measure on the manifold given its prior probability measure and the constraints related to the statistical moments for which targets are given. This method is the one that is used in the methodology presented in this paper. It should be noted that this principle has widely been used in many fields (see for instance, [88, 90, 91, 92, 93]) in particular, for reinforcement learning [94] and for probabilistic learning [95, 96]). The posterior probability measure is represented by an algebraic expression of the prior probability measure and of a vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ associated with the vector-valued function $\boldsymbol{h}^c$. The optimal value $\boldsymbol{\lambda}^{\text{sol}}$ of the Lagrange multipliers is obtained as the limit of a sequence $\{\boldsymbol{\lambda}^i\}_i$ of Lagrange multipliers allowing for constructing a sequence of probability measures whose limit, for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$ is the searched posterior probability measure. In this paper, we will call "constrained learned set", the set of realizations that results from the learning process in taking into account the constraints. As the framework of the developments presented is that of the high dimension with a small training set, for each value $\boldsymbol{\lambda}^i$ of $\boldsymbol{\lambda}$, the constrained learned set must be generated with a MCMC algorithm. The MCMC generator is based on a nonlinear Itô stochastic differential equation (ISDE) associated with a nonlinear stochastic dissipative Hamiltonian dynamical system. The presence of a dissipative term makes it possible to delete the transient part in order to quickly reach the stationary response associated with the invariant measure. The evaluation of the drift vector of the ISDE requires to evaluate the gradient of function $\boldsymbol{h}^c$ a large number of times. In the framework of this work, there is not an available algebraic expression of this gradient. Furthermore, it is not possible to do a direct numerical calculation of it, taking into account the high dimension (function $\boldsymbol{h}^c$ is implicit and for each evaluation of it, the BVP must be used). For instance, components of $\boldsymbol{h}^c$ can be related to the norm of the random normalized residue of the partial differential equation (including the boundary conditions) of the BVP and the statistical moments of random observations. The construction of a surrogate model of implicit function $\boldsymbol{h}^c$ by using a deterministic approach, such as the meshless methods [97, 98, 99, 100, 101, 102], is not adapted taking into account a possible high dimension of the space on which $\boldsymbol{h}^c$ is defined. To circumvent this difficulty, we generalize the approach proposed in [96], which consists in constructing a statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$, depending on the number N of points generated in the constrained learned set, for which its gradient has an explicit algebraic representation.

(iii) Organization and novelties of the paper. First of all, let us point out that a similar problem has been tackled in [95] devoted to take into account constraints in the PLoM (probabilistic learning on manifolds) method [103, 104, 105]. However, in [95], function $\boldsymbol{h}^c$ is explicit. In this work, $\boldsymbol{h}^c$ is implicit (that is to say the use of the BVP is required for evaluating its value in any given point). The presented methodology is novel and general (note that PLoM is not used, but could be implemented if necessary, but this would be prejudicial to the clarity of the developments). In addition, a mathematical analysis of the methodology is presented. This analysis is necessary because, due to the use of the statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$, the constrained learned realizations of the posterior probability measure on the manifold are generated with the MCMC generator that is the limit of a sequence of MCMC generators, each generator of the sequence depending on $\hat{\boldsymbol{h}}^N$ instead of $\boldsymbol{h}^c$. This means that convergence properties with respect to N must be studied. The organization of the paper is the following.

In Section 2, we set the problem and summarize the methodology of the probabilistic learning inference that is proposed. The used hypotheses are discussed in order to expose the difficulties involved by the choice of a general framework for the developments: small number of points in the training set, targets defined by statistical moments of the observations, implicit description of function $\boldsymbol{h}^c$ and the necessity to construct a statistical surrogate model.

Section 3 deals with the mathematical analysis of the proposed methodology. Lemma 1 proves the convexity of the admissible set $C_{\text{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{n_c}$ of the Lagrange multipliers $\boldsymbol{\lambda}$ and the integrability properties related to the sequence of posterior probability measures indexed by $\boldsymbol{\lambda}$. Proposition 1 gives the construction of $\boldsymbol{\lambda}^{\text{sol}}$ as the unique solution of a convex optimization problem posed in $C_{\text{ad},\boldsymbol{\lambda}}$. After building the statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$ (Definition 3), Proposition 2 proves the convergence of the sequence $\{\hat{\boldsymbol{h}}^N\}_N$ towards $\boldsymbol{h}^c$ and of its gradient for $\boldsymbol{\lambda}$ fixed in $C_{\text{ad},\boldsymbol{\lambda}}$. Proposition 3 proves the existence and uniqueness of the invariant measure of the ISDE for each fixed value of $\boldsymbol{\lambda}$ and gives the generator of the constrained learned set depending on $\boldsymbol{\lambda}$. Proposition 4 proves the mean-square convergence, as N goes to infinity, of the stationary stochastic solution of the ISDE calculated with $\hat{\boldsymbol{h}}^N$ towards the one calculated with $\boldsymbol{h}^c$. After having given the construction of the iteration algorithm for calculating the sequence $\{\boldsymbol{\lambda}^i\}_i$ that converges to $\boldsymbol{\lambda}^{\text{sol}}$, Proposition 5 gives the rate of convergence of the sequence of posterior probability measures indexed by $\boldsymbol{\lambda}$.

In Section 4, a few numerical elements are given for implementing the methodology. The Störmer-Verlet scheme that allows for solving the ISDE is detailed and the explicit expression of the gradient of $\hat{\boldsymbol{h}}^N$ is given. We present the algorithm for calculating $\boldsymbol{\lambda}^{\text{sol}}$ and for generating the constrained learned set $\mathcal{D}_{\boldsymbol{H}^c}$.

Section 5 is devoted to stochastic homogenization that has given rise to a large number of works (see for instance [106, 107, 108, 109, 110, 111, 112, 113]) and for which the analysis of the representative volume element (RVE) size has received a particular attention (see [114, 115, 116, 117, 118, 119, 120, 121, 122]). In the context of the stochastic homogenization of the three-dimensional elastic microstructures of materials, the effective quantities are sought at the macroscopic scale by making a spatial average over a representative volume element (RVE). The effective quantities are those of the macroscopic scale when the dimensions of the RVE in the three directions are taken sufficiently large so that the statistical fluctuations of the effective quantities are negligible compared to the statistical mean value. In this paper, we present an application related to the stochastic homogenization of a random linear elastic medium at mesoscale posed on the domain $\Omega \subset \mathbb{R}^3$ of the microstructure, which is not a RVE, which means that there is no scale separation between the mesoscale and the "macroscale". Since there is no scale separation, the "macroscale" is in fact another mesoscale at a larger scale than the initial mesoscale, which is described by random effective/apparent quantities. Consequently, the effective/apparent elasticity tensor at macroscale is not deterministic and has statistical fluctuations. This case is obtained when the spatial correlation lengths of the random apparent elasticity field at mesoscale are not sufficiently small with respect to the size of domain Ω. For instance, such a situation is encountered when the size of an experimental specimen, which is measured with the objective to perform an inverse identification of its effective properties, is not sufficiently large compared to the size of the heterogeneities. It is then interesting to identify the probabilistic model of the random apparent elasticity field, without having a scale separation. One can then estimate the effective properties using the identified random apparent elasticity field with a stochastic computational model for which domain Ω is an RVE, that is to say, has a size that is largest than the one of the specimen. In order to analyze the scale separation, we will consider several values of the spatial correlation lengths to cover three cases:

SC1: partial separation (separated for two directions but not in the third one).

SC2: not separated in the three directions.

SC3: strongly separated in the three directions.

Notations

x, η: lower-case Latin or Greek letters are deterministic real variables.
$\boldsymbol{x}, \boldsymbol{\eta}$: boldface lower-case Latin or Greek letters are deterministic vectors.
X: upper-case Latin letters are real-valued random variables.
$\boldsymbol{X}$: boldface upper-case Latin letters are vector-valued random variables.
$[x]$: lower-case Latin letters between brackets are deterministic matrices.
$[\boldsymbol{X}]$: boldface upper-case letters between brackets are matrix-valued random variables.
$C_{\text{ad},\boldsymbol{\lambda}}$: admissible set of $\boldsymbol{\lambda} \in \mathbb{R}^{n_c}$.
$\mathbb{C}$: fourth-order tensor-valued random field.
$\mathcal{D}_d$: training set.
$\mathcal{D}_{\boldsymbol{H}^c}$: constrained learned set at convergence for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$.
$\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}}$: constrained learned set for $\boldsymbol{\lambda}^i$.
N: number of points in the constrained learned set.
N_d: number of points in the training set.
$\mathbb{N}$, $\mathbb{R}$: set of all the integers $\{0, 1, 2, \ldots\}$, set of all the real numbers.

$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$\mathbb{M}_{n,m}$: set of all the $(n \times m)$ real matrices.
$\mathbb{M}_n$: set of all the square $(n \times n)$ real matrices.
$\mathbb{M}_n^+$: set of all the positive-definite symmetric $(n \times n)$ real matrices.
$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\boldsymbol{x} = (x_1, \ldots, x_n)$: point in $\mathbb{R}^n$.
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1 y_1 + \ldots + x_n y_n$: inner product in $\mathbb{R}^n$.
$\| \boldsymbol{x} \|$: norm in $\mathbb{R}^n$ such that $\| \boldsymbol{x} \| = \langle \boldsymbol{x}, \boldsymbol{x} \rangle$.
$[x]^T$: transpose of matrix $[x]$.
tr$\{[x]\}$: trace of the square matrix $[x]$.
$\| [x] \|_F$: Frobenius norm of matrix $[x]$.
$\delta_{kk'}$: Kronecker's symbol.
$\delta_{\boldsymbol{x}_0}$: Dirac measure at point $\boldsymbol{x}_0$.
a.s.: almost surely.
BVP: boundary value problem.
det: determinant.
dof: degree of freedom.
E: mathematical expectation operator.
err: error function.
ISDE: Itô stochastic differential equation.
KDE: kernel density estimation.
pdf: probability density function.
PDE: partial differential equation.

2. Setting the problem and summarizing the methodology

In this paper, all the random variables are defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ in which Θ is the sample set, $\mathcal{T}$ is the σ-field of Θ, and where $\mathcal{P}$ is a probability measure on the measurable space $(\Theta, \mathcal{T})$. A "sample" $\boldsymbol{X}(\theta), \theta \in \Theta$ of a random variable $\boldsymbol{X}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ will also be called a "realization" of $\boldsymbol{X}$.

2.1. Framework of the developments presented in the paper

For instance, we consider a stochastic elliptic BVP on an open bounded domain $\Omega \subset \mathbb{R}^d$ with $d \geq 1$, whose partial differential equation (PDE) is written as $\boldsymbol{\mathcal{N}}(\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{W}) = \boldsymbol{0}$ *a.s.* The unknown is the non-Gaussian vector-valued field $\{\boldsymbol{Y}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ and which satisfies the boundary conditions. The coefficients of the stochastic elliptic operator depend on a non-Gaussian second-order vector-valued random field $\boldsymbol{G}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ and on a random vector-valued control parameter $\boldsymbol{W}$ also defined on $(\Theta, \mathcal{T}, \mathcal{P})$. It is assumed that the weak formulation of this stochastic BVP admits a unique strong stochastic solution $\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{G}, \boldsymbol{W})$, which is a second-order random field. The observation (quantity of interest) is, for instance, a second-order vector-valued random variable $\boldsymbol{Q} = \boldsymbol{O}(\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{W})$ in which $\boldsymbol{O}$ is a given measurable mapping.

The problem under consideration belongs to the class of the statistical inverse problems. A prior probability model of $\{\boldsymbol{G}, \boldsymbol{W}\}$ is given and we are interested in estimating a posterior model $\{\boldsymbol{G}^c, \boldsymbol{W}^c\}$ of $\{\boldsymbol{G}, \boldsymbol{W}\}$ in order that some statistical moments of the posterior observations $\boldsymbol{Q}^c = \boldsymbol{O}(\boldsymbol{Y}^c, \boldsymbol{G}^c, \boldsymbol{W}^c)$ with $\boldsymbol{Y}^c = \boldsymbol{f}(\boldsymbol{G}^c, \boldsymbol{W}^c)$, are equal to some given targets (the superscript "c" is introduced to designate the solution with the constraints, which corresponds to the posterior model). The statistical moments of $\boldsymbol{Q}^c$ are globally written as $E\{\boldsymbol{\mathcal{M}}^c(\boldsymbol{Q}^c)\} = \boldsymbol{b}^c$ in which $\boldsymbol{b}^c \in \mathbb{R}^{n_c}$ is the target, $\boldsymbol{q} \mapsto \boldsymbol{\mathcal{M}}^c(\boldsymbol{q})$ is a given measurable mapping, and E is the mathematical expectation operator. As explained in Section 1, the realizations that have allowed $\boldsymbol{b}^c$ to be estimated are not available. Only $\boldsymbol{b}^c$ is known (it can be the case when one or several components of $\boldsymbol{b}^c$ have been estimated with experimental realizations that are not available). However, as we will see in Section 5, a component of the vector of statistical moments can also be related to the random normalized residue of the PDE of the stochastic boundary value problem.

2.2. Hypotheses concerning the problem to be solved

(i) Small dimension N_d of the training set. The subscript "d" is introduced to designate the quantities related to the training set that is constructed by using the Monte Carlo numerical simulation method. Let $\{\boldsymbol{g}_d^1, \ldots, \boldsymbol{g}_d^{N_d}\}$ and $\{\boldsymbol{w}_d^1, \ldots, \boldsymbol{w}_d^{N_d}\}$ be N_d independent realizations of random variable $\{\boldsymbol{G}, \boldsymbol{W}\}$, generated by using the prior probability model of $\{\boldsymbol{G}, \boldsymbol{W}\}$. Each realization of the BVP defined as the PDE $\boldsymbol{\mathcal{N}}(\boldsymbol{y}_d^j, \boldsymbol{g}_d^j, \boldsymbol{w}_d^j) = \boldsymbol{0}$ with its boundary conditions, is solved. Consequently, N_d independent realizations $\{\boldsymbol{y}_d^j, j = 1, \ldots, N_d\}$ of random field $\boldsymbol{Y}$ are computed and are such that $\boldsymbol{y}_d^j = \boldsymbol{f}(\boldsymbol{g}_d^j, \boldsymbol{w}_d^j)$. The N_d independent realizations $\{\boldsymbol{q}_d^j, j = 1, \ldots, N_d\}$ of random observation $\boldsymbol{Q}$ are thus deduced such that $\boldsymbol{q}_d^j = \boldsymbol{O}(\boldsymbol{y}_d^j, \boldsymbol{g}_d^j, \boldsymbol{w}_d^j)$. The training set is then made up of a small number N_d of points $\boldsymbol{x}_d^j = \{\boldsymbol{y}_d^j, \boldsymbol{g}_d^j, \boldsymbol{w}_d^j\}$ for $j = 1, \ldots, N_d$, which are N_d independent realizations of $\boldsymbol{X} = \{\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{W}\}$. Note that a strong hypothesis in the present work is that the BVP can only be solved a small number of times. This means that the training set is a small data set (as opposed to a big data set). Consequently, for constructing the posterior probability measure on the manifold, a learning tool must be used for generating the constrained learned realizations of $\boldsymbol{X}$ without solving the BVP, but using only the training set.

(ii) Statistical inverse problem. The available information consists of the targets of statistical moments related to observation $\boldsymbol{Q}^c$ and represented by $\boldsymbol{b}^c$. As we have explained in Section 1, the realizations that have been used for estimating $\boldsymbol{b}^c$ are not available. This is the second strong hypothesis used in this work. Consequently the classical statistical tools such as the Bayesian inference or the maximum likelihood method cannot easily be used for solving the statistical inverse problem under consideration.

(iii) Finite reduced-order representation. The second-order random variable $\boldsymbol{X} = \{\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{W}\}$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, is assumed to be with values in a real Hilbert space $\mathbb{X}$ equipped with the inner product $\langle \boldsymbol{X}, \boldsymbol{X}' \rangle_{\mathbb{X}}$ and its associated norm $\| \boldsymbol{X} \|_{\mathbb{X}} = \langle \boldsymbol{X}, \boldsymbol{X} \rangle_{\mathbb{X}}^{1/2}$. Consequently, $\boldsymbol{X}$ belongs the Hilbert space $L^2(\Theta, \mathbb{X})$ of the equivalent class of all the second-order random variables with values in $\mathbb{X}$, equipped with the inner product $\langle\langle \boldsymbol{X}, \boldsymbol{X}' \rangle\rangle = E\{\langle \boldsymbol{X}, \boldsymbol{X}' \rangle_{\mathbb{X}}\}$ for which the square of the associated norm is $|||\, \boldsymbol{X} \,|||^2 = E\{\| \boldsymbol{X} \|_{\mathbb{X}}^2\} = \int_\Theta \| \boldsymbol{X}(\theta) \|_{\mathbb{X}}^2 \, d\mathcal{P}(\theta)$. Since the problem is in infinite dimension, in order to implement the probabilistic learning inference that we propose, we need to introduce a finite representation $\boldsymbol{X}^{(\nu)}$ of dimension ν of random variable $\boldsymbol{X}$ in $L^2(\Theta, \mathbb{X})$. Assuming that the covariance operator is a Hilbert-Schmidt [123], symmetric, positive operator in $\mathbb{X}$, $\boldsymbol{X}^{(\nu)}$ can be represented using the truncated Karhunen-Loève expansion [124, 125] of $\boldsymbol{X}$,

$$\boldsymbol{X}^{(\nu)} = \underline{\boldsymbol{x}} + \sum_{\alpha=1}^{\nu} \sqrt{\kappa_\alpha}\, \boldsymbol{\varphi}^\alpha \, H_\alpha \,, \tag{2.1}$$

in which the eigenvalues of the covariance operator are $\kappa_1 \geq \ldots \geq \kappa_\nu \geq \ldots = 0$ with $\sum_{\alpha=1}^{+\infty} \kappa_\alpha^2 < +\infty$, where the family of the eigenfunctions $\{\boldsymbol{\varphi}^\alpha\}_\alpha$ is a Hilbertian basis of $\mathbb{X}$, where $\underline{\boldsymbol{x}} = E\{\boldsymbol{X}\}$, and where $\boldsymbol{H} = (H_1, \ldots, H_\nu)$ is a second-order, centered, $\mathbb{R}^\nu$-valued random variable whose covariance matrix is the identity matrix $[I_\nu]$ in $\mathbb{M}_\nu$. It should be noted that the covariance operator of $\boldsymbol{X}$ is not positive definite but only positive (its kernel is not reduced to zero) and therefore, there is a zero eigenvalue with a finite multiplicity, which is not taken into account in the truncated representation of $\boldsymbol{X}$, defined by Eq. (2.1), in which $\kappa_\nu > 0$. The following heuristic argument can be given for explaining why the covariance operator of $\boldsymbol{X}$ is only positive. Since the deterministic mapping $\boldsymbol{\mathcal{N}}$ relates the random quantities $\boldsymbol{Y}$, $\boldsymbol{G}$, and $\boldsymbol{W}$, the "functional components" of $\boldsymbol{X}$ are not "independent", and so the covariance operator is not positive definite. For $\alpha \in \{1, \ldots, \nu\}$, the component H_α is written as $H_\alpha = \kappa_\alpha^{-1/2} \langle \boldsymbol{X} - \underline{\boldsymbol{x}}, \boldsymbol{\varphi}^\alpha \rangle_{\mathbb{X}}$. The training set $\mathcal{D}_d$ related to $\boldsymbol{H}$ is made up of the N_d independent realizations $\{\boldsymbol{\eta}_d^j, j = 1, \ldots, N_d\}$ such that $\boldsymbol{\eta}_d^j = \kappa_\alpha^{-1/2} \langle \boldsymbol{x}_d^j - \underline{\boldsymbol{x}}, \boldsymbol{\varphi}^\alpha \rangle_{\mathbb{X}}$. If the kernel of the covariance operator was explicitly known, then ν would be chosen in order that $|||\, \boldsymbol{X} - \boldsymbol{X}^{(\nu)} \,||| \leq \varepsilon$ for a sufficiently small value of ε. In this paper, we assume that the kernel of the covariance operator is unknown. Therefore, we can only obtain an approximation of the covariance operator using an empirical estimator built with the N_d points $\{\boldsymbol{x}_d^j, j = 1, \ldots, N_d\}$ (for instance, see Section 5.6). Under these conditions the largest value of ν will be $N_d - 1$ and the discretization of Eq. (2.1) will simply correspond to a normalization of the points that constitute the training set $\mathcal{D}_d$.

2.3. Formulation using the Kullback-Leibler divergence minimum principle

Taking into account Section 2.2-(ii), we use the Kullback-Leibler divergence minimum principle [87, 88, 89] for estimating the posterior probability measure $P_{\boldsymbol{H}^c}(d\boldsymbol{\eta}) = p_{\boldsymbol{H}^c}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}^c = (H_1^c, \ldots, H_\nu^c)$. This estimation of $P_{\boldsymbol{H}^c}$ is performed (1) using the prior probability measure $P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$

in which $p_{\boldsymbol{H}}$ is estimated with the Gaussian KDE method using the points $\{\boldsymbol{\eta}_d^1, \ldots, \boldsymbol{\eta}_d^{N_d}\}$ of the training set $\mathcal{D}_d$, and (2) using the constraint defined by the given statistical moments,

$$E\{\boldsymbol{h}^c(\boldsymbol{H}^c)\} = \boldsymbol{b}^c \in \mathbb{R}^{n_c} . \tag{2.2}$$

In Eq. (2.2), $\boldsymbol{\eta} \mapsto \boldsymbol{h}^c(\boldsymbol{\eta})$ is a function from $\mathbb{R}^\nu$ into $\mathbb{R}^{n_c}$ with $n_c > 1$, such that $\boldsymbol{h}^c(\boldsymbol{H}^c)$ is a random variable equal to $\boldsymbol{\mathcal{M}}^c(\boldsymbol{Q}^c)$ that is expressed as a function of $\boldsymbol{H}^c$. We recall that the Kullback-Leibler divergence between two probability measures $p(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ and $p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ is defined by $D(p, p_{\boldsymbol{H}}) = \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) \log(p(\boldsymbol{\eta})/p_{\boldsymbol{H}}(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$, and is such that $D(p, p_{\boldsymbol{H}}) \geq 0$ (that can be proven by applying the Jensen inequality [126, 127]) and $D(p, p_{\boldsymbol{H}}) = 0$ if and only if $p = p_{\boldsymbol{H}}$. Note that $(p, p_{\boldsymbol{H}}) \mapsto D(p, p_{\boldsymbol{H}})$ is not a distance because the symmetry property and the triangle inequality are not verified. It can easily be seen that the cross entropy $S(p, p_{\boldsymbol{H}}) = -\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) \log(p_{\boldsymbol{H}}(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$ and the entropy $S(p) = -\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) \log(p(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$ are related to $D(p, p_{\boldsymbol{H}})$ by $S(p, p_{\boldsymbol{H}}) = S(p) + D(p, p_{\boldsymbol{H}})$. Finally, it can be proven (see for instance [89]) that $(p, p_{\boldsymbol{H}}) \mapsto D(p, p_{\boldsymbol{H}})$ is a convex function in the pair $(p, p_{\boldsymbol{H}})$. The probability density function $p_{\boldsymbol{H}^c}$ on $\mathbb{R}^\nu$, which satisfies the constraint defined by Eq. (2.2) and which is closest to $p_{\boldsymbol{H}}$, is the solution of the optimization problem (see for instance [88, 89, 95]),

$$p_{\boldsymbol{H}^c} = \arg \min_{p \in C_{\mathrm{ad},p}} \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta}) \log\left(\frac{p(\boldsymbol{\eta})}{p_{\boldsymbol{H}}(\boldsymbol{\eta})}\right) d\boldsymbol{\eta} , \tag{2.3}$$

in which the admissible set $C_{\mathrm{ad},p}$ is defined by

$$C_{\mathrm{ad},p} = \left\{ \boldsymbol{\eta} \mapsto p(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+ , \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = 1 , \int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \boldsymbol{b}^c \right\} . \tag{2.4}$$

It should be noted that, since we are interested in divergence (from $p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$) of probability measure $p(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ that satisfies the constraints expressed in $C_{\mathrm{ad},p}$, we are in fact interested in the divergence and therefore, the symmetry condition is irrelevant in this case.

2.4. Methodology used for solving the optimization problem and MCMC generator

The constraints, which are defined in admissible set $C_{\mathrm{ad},p}$, are taken into account by introducing the Lagrange multipliers $\lambda_0 - 1$ with $\lambda_0 \in \mathbb{R}^+$, which is associated with the normalization condition, and $\boldsymbol{\lambda} \in C_{\mathrm{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{n_c}$, which is associated with the moments constraints. The admissible set $C_{\mathrm{ad},\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is a subset of $\mathbb{R}^{n_c}$, which is completely defined by Definition 2 in Section 3. The Lagrange multiplier λ_0 is eliminated as a function of $\boldsymbol{\lambda}$. In Eq. (2.3), the posterior pdf $p_{\boldsymbol{H}^c}$ is constructed as the limit of a sequence $\{p_{\boldsymbol{H}_\lambda}\}_\lambda$ of probability density functions of a $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}_\lambda = (H_{\lambda,1}, \ldots, H_{\lambda,\nu})$ that depends on $\boldsymbol{\lambda}$. The construction of $\{p_{\boldsymbol{H}_\lambda}\}_\lambda$ requires to generate with a MCMC algorithm a constrained learned set $\mathcal{D}_{\boldsymbol{H}_\lambda} = \{\boldsymbol{\eta}_\lambda^1, \ldots \boldsymbol{\eta}_\lambda^N\}$ constituted of $N \gg N_d$ independent realizations $\{\boldsymbol{\eta}_\lambda^\ell, \ell = 1, \ldots, N\}$ of $\boldsymbol{H}_\lambda$. When the convergence is reached with respect to $\boldsymbol{\lambda}$, the constrained learned set $\mathcal{D}_{\boldsymbol{H}^c} = \{\boldsymbol{\eta}_c^1, \ldots, \boldsymbol{\eta}_c^N\}$ is generated. This set is made up of N independent realizations $\{\boldsymbol{\eta}_c^\ell, \ell = 1, \ldots, N\}$ of $\boldsymbol{H}^c$ whose probability measure is $p_{\boldsymbol{H}^c}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ (subscript or superscript "c" is introduced to designate the quantities related to the constrained learned set (posterior model)). For $\boldsymbol{\lambda}$ fixed in $C_{\mathrm{ad},\boldsymbol{\lambda}}$, $\mathcal{D}_{\boldsymbol{H}_\lambda}$ is generated using a MCMC generator based on a nonlinear Itô stochastic differential equation (ISDE) associated with the nonlinear stochastic dissipative Hamiltonian dynamical system proposed in [128] and based on [129]. This MCMC generator allows for deleting the transient part to rapidly reach the stationary response associated with the invariant measure for which measure $p_{\boldsymbol{H}^c}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ is a marginal measure of this invariant measure. The ISDE is solved by using the Störmer-Verlet algorithm, which yields an efficient and accurate MCMC algorithm. Using an ergodicity property, obtaining a portion of the realization of the stationary process of the ISDE, which corresponds to its invariant measure, is computed by integrating the ISDE over a long period of time. Snapshots are taken on this portion. The successive snapshots are taken sufficiently far apart to guarantee a quasi statistical independence. This approach is efficient but is numerically "recursive" and therefore cannot be parallelized. The proposed MCMC algorithm, which is detailed in Section 4.1, makes it possible to reach the stationary solution very quickly thanks to the introduction of the dissipation term. For each realization of the initial condition and of the Wiener process, we very quickly obtain a realization associated with the invariant measure. By generating independent realizations of the initial condition and of the Wiener process, independent realizations are thus calculated and these calculations can therefore be parallelized for strongly decreasing the elapsed time on a multicore computer. The independent realizations of the initial condition are generated by scanning all the points of the training set. The parallel loops are explicitly defined in Algorithm 1, which is detailed in Section 4.4).

2.5. Formal construction of the optimal solution using the sequence $\{p_{\boldsymbol{H}_\lambda}\}_\lambda$

Let us assumed that the optimization problem defined by Eq. (2.3) has almost one solution $p_{\boldsymbol{H}^c}$ and that $p = p_{\boldsymbol{H}^c}$ is a regular point of the continuously differentiable functional $p \mapsto \int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - \boldsymbol{b}^c$. For $\lambda_0 \in \mathbb{R}^+$ and $\boldsymbol{\lambda} \in C_{\text{ad},\lambda}$, we define the Lagrangian,

$$\mathcal{L}ag(p, \lambda_0, \boldsymbol{\lambda}) = \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log\left(\frac{p(\boldsymbol{\eta})}{p_{\boldsymbol{H}}(\boldsymbol{\eta})}\right) d\boldsymbol{\eta} + (\lambda_0 - 1)\,(\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - 1) + \langle \boldsymbol{\lambda}\,, \int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - \boldsymbol{b}^c \rangle\,.$$

For all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, the pdf of $\boldsymbol{H}$ is written as $p_{\boldsymbol{H}}(\boldsymbol{\eta}) = c_\nu\, \zeta(\boldsymbol{\eta})$ in which the positive-valued function ζ is integrable on $\mathbb{R}^\nu$, is such that $\text{supp}\, \zeta = \mathbb{R}^\nu$, and where the positive constant c_ν is such that $c_\nu^{-1} = \int_{\mathbb{R}^\nu} \zeta(\boldsymbol{\eta})\, d\boldsymbol{\eta}$. We define the sequence $\{p_{\boldsymbol{H}_\lambda}\}_\lambda$ of pdf $\boldsymbol{\eta} \mapsto p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})$ on $\mathbb{R}^\nu$, indexed by $\boldsymbol{\lambda}$, such that $p_{\boldsymbol{H}_\lambda}(.\,; \boldsymbol{\lambda})$ is an extremum of functional $p \mapsto \mathcal{L}ag(p, \lambda_0, \boldsymbol{\lambda})$. Using the calculus of variations yields

$$p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,; \boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda})\, \zeta(\boldsymbol{\eta})\, \exp\{-\langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\} \quad , \quad \forall\, \boldsymbol{\eta} \in \mathbb{R}^\nu\,, \tag{2.5}$$

in which $c_0(\boldsymbol{\lambda})$ is the constant of normalization that depends on $\boldsymbol{\lambda}$ (note that λ_0 is eliminated and we have $c_0(\boldsymbol{\lambda}) = c_\nu\, \exp\{-\lambda_0\}$). The existence of a unique solution requires that the constraints be algebraically independent in the following sense: given any bounded positive measure $P(d\boldsymbol{\eta})$ on $\mathbb{R}^\nu$ with support $\mathbb{R}^\nu$, there exists a bounded set $\mathcal{B}$ in $\mathbb{R}^\nu$ with $P(\mathcal{B}) > 0$ such that

$$\forall\, \boldsymbol{\nu} \in \mathbb{R}^{n_c}\ ,\ \|\,\boldsymbol{\nu}\,\| \neq 0\ ,\ \int_{\mathcal{B}} \langle \boldsymbol{h}^c(\boldsymbol{\eta})\,, \boldsymbol{\nu}\rangle^2 P(d\boldsymbol{\eta}) > 0\,. \tag{2.6}$$

Under the condition defined by Eq. (2.6), there exists (see [130]) $\boldsymbol{\lambda}^{\text{sol}}$ in $C_{ad,\boldsymbol{\lambda}}$ such that the functional $(p, \lambda_0, \boldsymbol{\lambda}) \mapsto \mathcal{L}ag(p, \lambda_0, \boldsymbol{\lambda})$ is stationary at point $p = p_{\boldsymbol{H}^c}$ for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$ and $\lambda_0 = -\log(c_0(\boldsymbol{\lambda}^{\text{sol}})/c_\nu)$. Consequently, $p_{\boldsymbol{H}^c} = p_{\boldsymbol{H}_{\boldsymbol{\lambda}^{\text{sol}}}}(.\,; \boldsymbol{\lambda}^{\text{sol}})$ and Eq. (2.5) yield

$$p_{\boldsymbol{H}^c}(\boldsymbol{\eta}) = c_0(\boldsymbol{\lambda}^{\text{sol}})\, \zeta(\boldsymbol{\eta})\, \exp\{-\langle \boldsymbol{\lambda}^{\text{sol}}, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\} \quad , \quad \forall\, \boldsymbol{\eta} \in \mathbb{R}^\nu\,. \tag{2.7}$$

Taking into account the introduced hypotheses, $p_{\boldsymbol{H}^c}$ is the unique solution of the optimization problem defined by Eq. (2.3), in which $\boldsymbol{\lambda}^{\text{sol}}$ is the unique solution of a convex optimization problem that will be defined by Proposition 1 in Section 3) and which will be, under required hypotheses, the solution of the following nonlinear algebraic equation in $\boldsymbol{\lambda}$, $\int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta})\, p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})\, d\boldsymbol{\eta} = \boldsymbol{b}^c$.

2.6. Implicit definition of function $\boldsymbol{h}^c$*, resulting difficulties, and necessity to construct a statistical surrogate model*

Using the methodology presented in Section 2.4, the drift vector of the ISDE will involve the matrix $[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu, n_c}$ (the transpose of the Jacobian matrix of $\boldsymbol{h}^c$). As we have explained in Section 1, an additional strong assumption used in this paper is that function $\boldsymbol{\eta} \mapsto \boldsymbol{h}^c(\boldsymbol{\eta})$ from $\mathbb{R}^\nu$ into $\mathbb{R}^{n_c}$ is not explicitly defined by an algebraic expression. It is assumed that we can only compute $\boldsymbol{a}^\ell = \boldsymbol{h}^c(\boldsymbol{\eta}^\ell) \in \mathbb{R}^{n_c}$ for any point $\boldsymbol{\eta}^\ell$ given in $\mathbb{R}^\nu$ (for instance and as previously underlined, a component of $\boldsymbol{h}^c$ can be related to the square of a norm of the random normalized residue of the stochastic PDE). The MCMC generator requires the evaluation of $[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})]$ for a large number of values of $\boldsymbol{\eta}$. Consequently, a statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$ is constructed and allows for deducing an algebraic representation $\boldsymbol{\eta} \mapsto [\nabla_{\boldsymbol{\eta}} \hat{\boldsymbol{h}}^N \boldsymbol{\eta})]$ of function $\boldsymbol{\eta} \mapsto [\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})]$ from $\mathbb{R}^\nu$ into $\mathbb{M}_{\nu, n_c}$. Such statistical surrogate model is an approximation whose convergence with respect to N will be given by Proposition 2 in Section 3.

3. Mathematical analysis of the proposed methodology

In this section, there are some repetitions with respect to Section 2, but we have preferred to do them so that Section 3 be mathematically coherent and self contained.

Definition 1 (Training set $\mathcal{D}_d$ and probability measure of $\boldsymbol{H}$). Let ν and N_d be integers such that $N_d > \nu$. Let $\mathcal{D}_d$ be the set of N_d points $\boldsymbol{\eta}_d^1, \ldots, \boldsymbol{\eta}_d^{N_d}$ given in $\mathbb{R}^\nu$ such that

$$\widehat{\underline{\boldsymbol{\eta}}} = \frac{1}{N_d} \sum_{j=1}^{N_d} \boldsymbol{\eta}_d^j = \boldsymbol{0}_\nu \quad , \quad [\widehat{C}_{\boldsymbol{H}}] = \frac{1}{N_d - 1} \sum_{j=1}^{N_d} (\boldsymbol{\eta}_d^j - \widehat{\underline{\boldsymbol{\eta}}}) \otimes (\boldsymbol{\eta}_d^j - \widehat{\underline{\boldsymbol{\eta}}}) = [I_\nu]\,. \tag{3.1}$$

Let $\boldsymbol{H} = (H_1, \ldots, H_\nu)$ be the $\mathbb{R}^\nu$-valued random variable defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$ whose probability measure $P_{\boldsymbol{H}}(d\boldsymbol{\eta}) = p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ is defined by the probability density function $\boldsymbol{\eta} \mapsto p_{\boldsymbol{H}}(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+$ with respect to the Lebesgue measure $d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$,

$$p_{\boldsymbol{H}}(\boldsymbol{\eta}) = c_\nu\, \zeta(\boldsymbol{\eta}) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^\nu \quad , \quad c_\nu = (\sqrt{2\pi}\, \hat{s})^{-\nu}\,, \tag{3.2}$$

in which $\hat{s} = s \left(s^2 + (N_d - 1)/N_d\right)^{-1/2}$ with $s = (4/(N_d(2+\nu)))^{1/(\nu+4)}$, and where $\boldsymbol{\eta} \mapsto \zeta(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+$ is written as

$$\zeta(\boldsymbol{\eta}) = \frac{1}{N_d} \sum_{j=1}^{N_d} \exp\left\{ -\frac{1}{2\hat{s}^2} \parallel \frac{\hat{s}}{s}\, \boldsymbol{\eta}_d^j - \boldsymbol{\eta} \parallel^2 \right\}. \tag{3.3}$$

We define the potential function $\boldsymbol{\eta} \mapsto \phi(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}$, related to $p_{\boldsymbol{H}}$, such that

$$\zeta(\boldsymbol{\eta}) = \exp\{-\phi(\boldsymbol{\eta})\}\,. \tag{3.4}$$

Remark 1. Definition 1 of $p_{\boldsymbol{H}}$ corresponds to a Gaussian kernel-density estimation (KDE) using the training set $\mathcal{D}_d$, involving the modification proposed in [47] of the classical formulation [131] for which s is the Silverman bandwidth. With such a modification, the normalization of $\boldsymbol{H}$ is preserved for any value of N_d,

$$E\{\boldsymbol{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\, p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \frac{1}{2\hat{s}^2}\, \widehat{\underline{\boldsymbol{\eta}}} = \mathbf{0}_\nu\,, \tag{3.5}$$

$$E\{\boldsymbol{H} \otimes \boldsymbol{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta} \otimes \boldsymbol{\eta}\, p_{\boldsymbol{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \hat{s}^2\, [I_\nu] + \frac{\hat{s}^2}{s^2} \frac{(N_d - 1)}{N_d}\, [\widehat{C}_{\boldsymbol{H}}] = [I_\nu]\,. \tag{3.6}$$

Theorem 3.1 in [104] proves that, for all $\boldsymbol{\eta}$ fixed in $\mathbb{R}^\nu$, Eq. (3.2) with Eq. (3.3) is a consistent estimation of the sequence $\{p_{\boldsymbol{H}}\}_{N_d}$ for $N_d \to +\infty$.

Hypothesis 1 (Concerning function $\boldsymbol{h}^c$). *Let n_c be the integer such that $1 \leq n_c \leq \nu$. It is assumed that $\boldsymbol{\eta} \mapsto \boldsymbol{h}^c(\boldsymbol{\eta})$ verifies the property defined by Eq.* (2.6)*, is continuously differentiable from $\mathbb{R}^\nu$ into $\mathbb{R}^{n_c}$,*

$$\boldsymbol{h}^c \in C^1(\mathbb{R}^\nu, \mathbb{R}^{n_c})\,, \tag{3.7}$$

and there exist constants $\alpha > 0$, $\beta > 0$, $c_\alpha > 0$, and $c_\beta > 0$, independent of $\boldsymbol{\eta}$, such that for $\parallel \boldsymbol{\eta} \parallel \to +\infty$,

$$\parallel \boldsymbol{h}^c(\boldsymbol{\eta}) \parallel \leq c_\alpha \parallel \boldsymbol{\eta} \parallel^\alpha \quad , \quad \parallel [\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})] \parallel_F \leq c_\beta \parallel \boldsymbol{\eta} \parallel^\beta, \tag{3.8}$$

in which $[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu, n_c}$ with $[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})]_{\alpha k} = \partial h_k^c(\boldsymbol{\eta})/\partial \eta_\alpha$, and where $\parallel . \parallel_F$ is the Frobenius norm.

Definition 2 (Admissible subset $C_{\mathrm{ad},\boldsymbol{\lambda}}$ of $\mathbb{R}^{n_c}$). Under Hypothesis 1, the admissible set $C_{\mathrm{ad},\boldsymbol{\lambda}}$ of Lagrange multiplier $\boldsymbol{\lambda}$ is defined as the open subset of $\mathbb{R}^{n_c}$ such that

$$C_{\mathrm{ad},\boldsymbol{\lambda}} = \{\boldsymbol{\lambda} \in \mathbb{R}^{n_c} \mid 0 < E\{\exp\{-\langle \boldsymbol{\lambda}, \boldsymbol{h}^c(\boldsymbol{H})\rangle\}\} < +\infty\,, \tag{3.9}$$

in which the pdf of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}$ is defined by Eq. (3.2). It is also assumed that $\boldsymbol{h}^c$ is such that $C_{\mathrm{ad},\boldsymbol{\lambda}}$ is not reduced to the empty set,

$$C_{\mathrm{ad},\boldsymbol{\lambda}} \neq \emptyset\,. \tag{3.10}$$

Lemma 1 (Convexity of $C_{\mathrm{ad},\boldsymbol{\lambda}}$ and integrability properties). *Under Hypothesis 1 and with Definition 2,*
(a) $C_{\mathrm{ad},\boldsymbol{\lambda}}$ defined by Eq. (3.9) *is a convex open subset of $\mathbb{R}^{n_c}$.*
(b) $\forall \boldsymbol{\lambda} \in C_{\mathrm{ad},\boldsymbol{\lambda}}$, let $\boldsymbol{\eta} \mapsto \mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}$ be the function defined by

$$\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = \phi(\boldsymbol{\eta}) + \langle \boldsymbol{\lambda}, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\,, \tag{3.11}$$

in which $\phi(\boldsymbol{\eta}) = -\log \boldsymbol{\zeta}(\boldsymbol{\eta})$ (see Eq. (3.4)*). One then has*

$$0 < \int_{\mathbb{R}^\nu} \exp\{-\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\}\, d\boldsymbol{\eta} < +\infty\,. \tag{3.12}$$

(c) The pdf $\boldsymbol{\eta} \mapsto p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ *with respect to* $d\boldsymbol{\eta}$*, defined by Eq.* (2.5)*, which can be written as*

$$p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda})\,\exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\} \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^\nu\,, \tag{3.13}$$

is such that the constant $c_0(\boldsymbol{\lambda})$ *of normalization verifies*

$$0 < c_0(\boldsymbol{\lambda}) < +\infty \quad , \quad \forall \boldsymbol{\lambda} \in \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}\,. \tag{3.14}$$

(d) $\forall \boldsymbol{\lambda} \in \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$*, we have* $\mathcal{V}_\lambda(\boldsymbol{\eta}) \to +\infty$ *if* $\|\,\boldsymbol{\eta}\,\| \to +\infty$*, and we have,*

$$\int_{\mathbb{R}^\nu} \|\,\boldsymbol{h}^c(\boldsymbol{\eta})\,\|^2 \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}\,d\boldsymbol{\eta} \;<\; +\infty \quad , \quad \int_{\mathbb{R}^\nu} \|\,[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})]\,\|_F \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}\,d\boldsymbol{\eta} \;<\; +\infty\,. \tag{3.15}$$

PROOF. (Lemma 1).
(a) $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ is a convex subset if $\forall \mu \in [0,1]$, $\forall \boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ in $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, $\mu\boldsymbol{\lambda} + (1-\mu)\boldsymbol{\lambda}' \in \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$. Let $A = \exp\{-\langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{H})\rangle\}$ and $B = \exp\{-\langle \boldsymbol{\lambda}'\,, \boldsymbol{h}^c(\boldsymbol{H})\rangle\}$ be $\mathbb{R}^+$-valued random variables. Since $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are in $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, we have $E\{A\} < +\infty$ and $E\{B\} < +\infty$. We have to prove that $E\{\exp\{-\langle \mu\boldsymbol{\lambda} + (1-\mu)\boldsymbol{\lambda}'\,, \boldsymbol{h}^c(\boldsymbol{H})\rangle\}\} < +\infty$, that is to say, $E\{A^\mu\,B^{1-\mu}\} < +\infty$. For $\mu = 0$ and $\mu = 1$, it is verified. For $\mu \in]0,1[$, since A and B are almost-surely positive and using the Hölder inequality yield $E\{A^\mu\,B^{1-\mu}\} \leq (E\{A\})^\mu \times (E\{B\})^{1-\mu} < +\infty$, which finishes the proof of the convexity of $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$.
(b) Using Eqs. (3.2), (3.4), and (3.11), yields $\int_{\mathbb{R}^\nu} \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}\,d\boldsymbol{\eta} = \frac{1}{c_\nu} \int_{\mathbb{R}^\nu} \exp\{-\langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\}\,p_{\boldsymbol{H}}(\boldsymbol{\eta})\,d\boldsymbol{\eta} = \frac{1}{c_\nu} E\{\exp\{-\langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\}\}$, which is positive and finite due to Eq. (3.9) and to $0 < c_\nu < +\infty$. We have thus proven Eq. (3.12).
(c) Using Eqs. (3.12) and (3.13), and since we must have $\int_{\mathbb{R}^\nu} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta})\,d\boldsymbol{\eta} = 1$, we deduce Eq. (3.14).
(d) Since $\boldsymbol{h}^c$ is continuous on $\mathbb{R}^\nu$, (see Eq. (3.7)), $\forall \boldsymbol{\lambda} \in \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, $\boldsymbol{\eta} \mapsto \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}$ is continuous on $\mathbb{R}^\nu$ and then is locally integrable on $\mathbb{R}^\nu$. Eq. (3.12) implies the integrability at infinity of $\boldsymbol{\eta} \mapsto \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}$. Since $\boldsymbol{\eta} \mapsto \mathcal{V}_\lambda(\boldsymbol{\eta})$ is continuous on $\mathbb{R}^\nu$, it can be deduced that $\mathcal{V}_\lambda(\boldsymbol{\eta}) \to +\infty$ if $\|\,\boldsymbol{\eta}\,\| \to +\infty$. Using Eq. (3.8), for $\|\,\boldsymbol{\eta}\,\| \to +\infty$, one has $\|\,\boldsymbol{h}^c(\boldsymbol{\eta})\,\|^2 \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\} \leq c_\alpha^2\,\|\,\boldsymbol{\eta}\,\|^{2\alpha} \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}$ and $\|\,[\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})]\,\|_F \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\} \leq c_\beta\,\|\,\boldsymbol{\eta}\,\|^\beta \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\}$, which allow for proving the integrability at infinity and then proving Eq. (3.15).

Proposition 1 (Construction of the probability measure of $\boldsymbol{H}_\lambda$**).** *We consider Hypothesis 1 and Definition 2. For all* $\boldsymbol{\lambda}$ *in* $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$*, let*

$$p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda})\,\zeta(\boldsymbol{\lambda})\exp\{-\langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{\eta})\rangle\} \tag{3.16}$$

be the pdf of $\boldsymbol{H}_\lambda$ *(see Eq.* (2.5)*) with* $c_0(\boldsymbol{\lambda})$ *satisfying Eq.* (3.14)*).*
(a) The $\mathbb{R}^{n_c}$*-valued random variable* $\boldsymbol{h}^c(\boldsymbol{H}_\lambda)$ *is of second-order,*

$$E\{\|\,\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\,\|^2\} < +\infty\,. \tag{3.17}$$

(b) Let $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda}) : \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}} \to \mathbb{R}$ *be defined by*

$$\Gamma(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}\,, \boldsymbol{b}^c\rangle - \log c_0(\boldsymbol{\lambda})\,, \tag{3.18}$$

in which $\boldsymbol{b}^c$ *is given in* $\mathbb{R}^{n_c}$*. For all* $\boldsymbol{\lambda}$ *in* $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$*, we have*

$$\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}) = \boldsymbol{b}^c - E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\} \in \mathbb{R}^{n_c}\,, \tag{3.19}$$

$$[\Gamma''(\boldsymbol{\lambda})] = [cov\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\}] \in \mathbb{M}^+_{n_c}\,, \tag{3.20}$$

where the positive-definite covariance matrix $[\Gamma''(\boldsymbol{\lambda})]$ *of* $\boldsymbol{h}^c(\boldsymbol{H}_\lambda)$ *is such that* $[\Gamma''(\boldsymbol{\lambda})]_{kk'} = \partial^2\Gamma(\boldsymbol{\lambda})/\partial\lambda_k\partial\lambda_{k'}$*.*
(c) Γ *is a strictly convex function on* $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$*. There is a unique solution* $\boldsymbol{\lambda}^{\mathrm{sol}}$ *in* $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ *of the convex optimization problem,*

$$\boldsymbol{\lambda}^{\mathrm{sol}} = \arg \min_{\boldsymbol{\lambda} \in \mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}} \Gamma(\boldsymbol{\lambda})\,. \tag{3.21}$$

If the following equation in $\boldsymbol{\lambda}$*,*

$$\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}) = \mathbf{0}_{n_c}\,, \tag{3.22}$$

has a solution $\tilde{\boldsymbol{\lambda}}$ *that belongs to* $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$*, then this solution is unique and we have* $\boldsymbol{\lambda}^{\mathrm{sol}} = \tilde{\boldsymbol{\lambda}}$*. The pdf* $p_{\boldsymbol{H}^c}$ *of* $\boldsymbol{H}^c$*, which satisfies the constraint* $E\{\boldsymbol{h}^c(\boldsymbol{H}^c)\} = \boldsymbol{b}^c$ *is written (see Eq.* (3.13) *or* (3.16)*) as*

$$p_{\boldsymbol{H}^c}(\boldsymbol{\eta}) = p_{\boldsymbol{H}_{\boldsymbol{\lambda}^{\mathrm{sol}}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}^{\mathrm{sol}}) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^\nu\,. \tag{3.23}$$

PROOF. (Proposition 1).
(a) Using Eq. (3.13), Eq. (3.14), and the first equation Eq. (3.15) yield

$$E\{\| \boldsymbol{h}^c(\boldsymbol{H}_\lambda) \|^2\} = \int_{\mathbb{R}^\nu} \| \boldsymbol{h}^c(\boldsymbol{\eta}) \|^2 \, c_0(\boldsymbol{\lambda}) \, \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\} \, d\boldsymbol{\eta} < +\infty \, .$$

(b) The definition of function Γ given by Eq. (3.18) is similar to the one introduced in the discrete case for finding the probability measure of maximal entropy [132, 88]. Let us prove Eqs. (3.19) and (3.20). Eq. (3.11) yields $\nabla_\lambda \mathcal{V}_\lambda(\boldsymbol{\eta}) = \boldsymbol{h}^c(\boldsymbol{\eta})$ and from Eq. (3.13), it can be deduced that

$$\nabla_\lambda p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \left(c_0(\boldsymbol{\lambda})^{-1} \, \nabla_\lambda c_0(\boldsymbol{\lambda}) - \boldsymbol{h}^c(\boldsymbol{\eta}) \right) \, p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, . \tag{3.24}$$

Integrating Eq. (3.13) on $\mathbb{R}^\nu$ and taking the logarithm yields $\log \, c_0(\boldsymbol{\lambda}) = -\log \int_{\mathbb{R}^\nu} \exp\{-\mathcal{V}_\lambda(\boldsymbol{\eta})\} \, d\boldsymbol{\eta}$ and consequently,

$$c_0(\boldsymbol{\lambda})^{-1} \, \nabla_\lambda c_0(\boldsymbol{\lambda}) = \int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta}) \, p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta} = E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\} \, . \tag{3.25}$$

Eq. (3.18) yields $\nabla_\lambda \Gamma(\boldsymbol{\lambda}) = \boldsymbol{b}^c - c_0(\boldsymbol{\lambda})^{-1} \, \nabla_\lambda c_0(\boldsymbol{\lambda})$, which proves Eq. (3.19) by using Eq. (3.25). It should be noted that Eq. (3.17) implies the existence of the mean value $E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\}$. Taking the derivative of Eq. (3.19) with respect to $\boldsymbol{\lambda}$ yields

$$[\Gamma\,''(\boldsymbol{\lambda})] = -\int_{\mathbb{R}^\nu} \boldsymbol{h}^c(\boldsymbol{\eta}) \otimes \nabla_\lambda p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta} \, . \tag{3.26}$$

Substituting Eq. (3.25) into Eq. (3.24) yields $\nabla_\lambda p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = (E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\} - \boldsymbol{h}^c(\boldsymbol{\eta}))\, p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$, which with Eq. (3.26), gives $[\Gamma\,''(\boldsymbol{\lambda})] = E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda) \otimes \boldsymbol{h}^c(\boldsymbol{H}_\lambda)\} - (E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\}) \otimes (E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda)\})$ that is the covariance matrix of the $\mathbb{R}^{n_c}$-valued random variable $\boldsymbol{h}^c(\boldsymbol{H}_\lambda)$. Again Eq. (3.17) proves the existence of matrix $[\Gamma\,''(\boldsymbol{\lambda})]$ as a covariance matrix, which is semi-positive definite. We have to prove that this matrix is positive definite, which will be true if the matrix $[M_\lambda] = E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda) \otimes \boldsymbol{h}^c(\boldsymbol{H}_\lambda)\}$ belongs to $\mathbb{M}^+_{n_c}$, that is to say if $\langle [M_\lambda] \, \boldsymbol{v} \, , \boldsymbol{v} \rangle^2 > 0$ for all $\boldsymbol{v}$ in $\mathbb{R}^{n_c}$ with $\| \boldsymbol{v} \| \neq 0$. Since $p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta}$ is a probability measure, this will be true if $\boldsymbol{\eta} \mapsto \langle \boldsymbol{h}^c(\boldsymbol{\eta}) \, , \boldsymbol{v} \rangle^2$ is not zero on a set $\mathcal{B}$ such that $\int_{\mathcal{B}} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta} > 0$ that is the case due to the hypothesis defined by Eq. (2.6).
(c) From Lemma 1-(a), $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ is a convex set and from Eq. (3.20), $[\Gamma\,''(\boldsymbol{\lambda})]$ is a positive-definite matrix for all $\boldsymbol{\lambda}$ in $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$. It can then be deduced that $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ is strictly convex on $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ and therefore, Eq. (3.21) holds and $\boldsymbol{\lambda}^{\mathrm{sol}}$ is unique. Note that the existence of a solution of Eq. (3.22), which would then be a global minimum, could be not in $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$. However, if the equation $\nabla_\lambda \Gamma(\boldsymbol{\lambda}) = \mathbf{0}_{n_c}$ admits a solution $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^{n_c}$ that belongs to $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, then this solution is unique and we have $\boldsymbol{\lambda}^{\mathrm{sol}} = \tilde{\boldsymbol{\lambda}}$, which is the solution of the convex optimization problem defined by Eq. (3.21), and then Eq. (3.22) holds for $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{\mathrm{sol}}$. Finally, under the condition that $\tilde{\boldsymbol{\lambda}}$ belongs to $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, Eq. (3.19) shows that $E\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}^{\mathrm{sol}}})\} = \boldsymbol{b}^c$. Taking into account Eq. (2.7), the solution is given by Eq. (3.23) and is unique due to the uniqueness of solution $\boldsymbol{\lambda}^{\mathrm{sol}}$ of $\nabla_\lambda \Gamma(\boldsymbol{\lambda}) = \mathbf{0}_{n_c}$.

Definition 3 (Surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$). Let $\boldsymbol{\lambda}$ be fixed in $\mathcal{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ and let $\mathcal{D}_{\boldsymbol{H}_\lambda} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}}, \ldots, \boldsymbol{\eta}^N_{\boldsymbol{\lambda}}\}$ be the constrained learned set whose points are $N \gg N_d$ independent realizations of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}_{\boldsymbol{\lambda}}$ for which the pdf $\boldsymbol{\eta} \mapsto p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ is defined by Eq. (3.13). Let $\boldsymbol{A}_{\boldsymbol{\lambda}} = \boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})$ be the $\mathbb{R}^{n_c}$-valued random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ whose N independent realizations $\boldsymbol{a}^1_{\boldsymbol{\lambda}}, \ldots, \boldsymbol{a}^N_{\boldsymbol{\lambda}}$ are such that $\boldsymbol{a}^\ell_{\boldsymbol{\lambda}} = \boldsymbol{h}^c(\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}}) \in \mathbb{R}^{n_c}$ for $\ell = 1, \ldots, N$. The surrogate model $\boldsymbol{\eta} \mapsto \hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) : \mathbb{R}^\nu \to \mathbb{R}^{n_c}$ of $\boldsymbol{h}^c$ is defined, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, by

$$\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \sum_{\ell=1}^{N} \boldsymbol{a}^\ell_{\boldsymbol{\lambda}} \, \frac{\beta^N_{\boldsymbol{\eta}}(\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}})}{\sum_{\ell'=1}^N \beta^N_{\boldsymbol{\eta}}(\boldsymbol{\eta}^{\ell'}_{\boldsymbol{\lambda}})} \, , \tag{3.27}$$

in which for all $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{\eta}}$ in $\mathbb{R}^\nu$,

$$\beta^N_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}) = \exp\{-\frac{1}{2 s^2_{\mathrm{SB}}} \| \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \|^2_H\} \, , \tag{3.28}$$

$$\| \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \|^2_H = \langle [\sigma_{\boldsymbol{H}_\lambda}]^{-2} (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \, , \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \rangle \, , \tag{3.29}$$

in which $[\sigma_{\boldsymbol{H}_\lambda}]$ is the diagonal positive-definite matrix in $\mathbb{M}_\nu^+$ such that $[\sigma_{\boldsymbol{H}_\lambda}]_{\alpha\alpha}$ is the standard deviation of the real-valued random variable $H_{\lambda,\alpha}$, estimated using $\mathcal{D}_{\boldsymbol{H}_\lambda}$, and where s_{SB} is the Silverman bandwidth that depends on N and written as

$$s_{\text{SB}} = \left(\frac{4}{N(2+n_c+\nu)} \right)^{1/(n_c+\nu+4)} . \tag{3.30}$$

Remark 2 (Rationale of Definition 3). Let us assume that $\boldsymbol{\lambda}$ is fixed in $C_{\text{ad},\lambda}$. For all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, we have the following identity,

$$\boldsymbol{h}^c(\boldsymbol{\eta}) = E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda) \,|\, \boldsymbol{H}_\lambda = \boldsymbol{\eta}\}\,, \tag{3.31}$$

in which $E\{\boldsymbol{h}^c(\boldsymbol{H}_\lambda) \,|\, \boldsymbol{H}_\lambda = \boldsymbol{\eta}\}$ is the conditional mathematical expectation of the $\mathbb{R}^{n_c}$-valued second-order random variable $\boldsymbol{h}^c(\boldsymbol{H}_\lambda)$ given $\boldsymbol{H}_\lambda = \boldsymbol{\eta}$. Let $P_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(d\boldsymbol{a}, d\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ be the probability measure on $\mathbb{R}^{n_c} \times \mathbb{R}^\nu$ of the $\mathbb{R}^{n_c} \times \mathbb{R}^\nu$-valued random variable $(\boldsymbol{A}_\lambda, \boldsymbol{H}_\lambda)$. Since $\boldsymbol{A}_\lambda = \boldsymbol{h}^c(\boldsymbol{H}_\lambda)$, the support of $P_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}$ is the manifold defined by the graph $\{(\boldsymbol{a},\boldsymbol{\eta}) \in \mathbb{R}^{n_c} \times \mathbb{R}^\nu \,;\, \boldsymbol{a} = \boldsymbol{h}^c(\boldsymbol{\eta})\}$. Let $P_{\boldsymbol{H}_\lambda}(d\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{\eta}$ be the probability measure of $\boldsymbol{H}_\lambda$, for which the density $p_{\boldsymbol{H}_\lambda}$ is defined by Eq. (3.13) with $\text{supp}\, p_{\boldsymbol{H}_\lambda} = \mathbb{R}^\nu$. Eq. (3.31) can then be rewritten, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, as

$$\boldsymbol{h}^c(\boldsymbol{\eta}) = \int_{\boldsymbol{a}\in\mathbb{R}^{n_c}} \boldsymbol{a}\, P_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(d\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})\,, \tag{3.32}$$

in which $P_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(d\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})$ is the conditional probability measure of $\boldsymbol{A}_\lambda$ given $\boldsymbol{H}_\lambda = \boldsymbol{\eta}$ in $\mathbb{R}^\nu$, which could also be written as $\delta_0(\boldsymbol{a} - \boldsymbol{h}^c(\boldsymbol{\eta}))$ in which δ_0 is the Dirac measure on $\mathbb{R}^{n_c}$ at point $\boldsymbol{a} = \boldsymbol{0}_{n_c}$. Note that if the joint probability measure $P_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(d\boldsymbol{a}, d\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ of $(\boldsymbol{A}_\lambda, \boldsymbol{H}_\lambda)$ had a density $p_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ with respect to $d\boldsymbol{a} \otimes d\boldsymbol{\eta}$ (that is not the case), then Eq. (3.32) could be written as $\boldsymbol{h}^c(\boldsymbol{\eta}) = \int_{\boldsymbol{a}\in\mathbb{R}^{n_c}} \boldsymbol{a}\, p_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{a}$ with $p_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda}) = p_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda}) / p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$. The statistical surrogate model $\boldsymbol{\eta} \mapsto \hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) : \mathbb{R}^\nu \to \mathbb{R}^{n_c}$ of $\boldsymbol{h}^c$ is then defined by approximating (regularizing) $P_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(d\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})$ by the conditional probability measure,

$$\hat{p}^N_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{a} = (\hat{p}^N_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}))^{-1}\, \hat{p}^N_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{a}\,, \tag{3.33}$$

in which the pdf $\hat{p}^N_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ on $\mathbb{R}^{n_c} \times \mathbb{R}^\nu$ with respect to $d\boldsymbol{a} \otimes d\boldsymbol{\eta}$ is defined by the following Gaussian kernel density representation, based on the N independent realizations $\{(\boldsymbol{a}^\ell_\lambda, \boldsymbol{\eta}^\ell_\lambda), \ell = 1, \ldots, N\}$ of $(\boldsymbol{A}_\lambda, \boldsymbol{H}_\lambda)$, and where $\hat{p}^N_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \int_{\mathbb{R}^{n_c}} \hat{p}^N_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{a}$. Therefore, we have

$$\hat{p}^N_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \frac{1}{N} \sum_{\ell=1}^{N} \Big((\sqrt{2\pi}\, s_{\text{SB}})^{n_c+\nu} \det[\sigma_{\boldsymbol{A}_\lambda}]\, \det[\sigma_{\boldsymbol{H}_\lambda}] \Big)^{-1} \exp\left\{ -\frac{1}{2 s^2_{\text{SB}}} \left(\| \boldsymbol{a}^\ell_\lambda - \boldsymbol{a} \|^2_A + \| \boldsymbol{\eta}^\ell_\lambda - \boldsymbol{\eta} \|^2_H \right) \right\}, \tag{3.34}$$

in which s_{SB} is defined by Eq. (3.30), where $[\sigma_{\boldsymbol{A}_\lambda}]$ is the diagonal positive-definite matrix in $\mathbb{M}^+_{n_c}$ such that $[\sigma_{\boldsymbol{A}_\lambda}]_{kk}$ is the standard deviation of the real-valued random variable $A_{\lambda,k}$, estimated using the realizations $\{\boldsymbol{a}^\ell_\lambda, \ell = 1, \ldots, N\}$, and where for all $\tilde{\boldsymbol{a}}$ and $\boldsymbol{a}$ in $\mathbb{R}^{n_c}$, $\| \tilde{\boldsymbol{a}} - \boldsymbol{a} \|^2_A = \langle [\sigma_{\boldsymbol{A}_\lambda}]^{-2} (\tilde{\boldsymbol{a}} - \boldsymbol{a})\,, \tilde{\boldsymbol{a}} - \boldsymbol{a} \rangle$. From Eq. (3.32) and using the approximation of $P_{\boldsymbol{A}_\lambda \,|\, \boldsymbol{H}_\lambda}(d\boldsymbol{a} \,|\, \boldsymbol{\eta}\,;\boldsymbol{\lambda})$ defined by Eq. (3.33), we have $\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = (\hat{p}^N_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}))^{-1} \int_{\boldsymbol{a}\in\mathbb{R}^{n_c}} \boldsymbol{a}\, \hat{p}^N_{\boldsymbol{A}_\lambda,\boldsymbol{H}_\lambda}(\boldsymbol{a},\boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{a}$, which using Eq. (3.34) yields Eq. (3.27).

Proposition 2 (Convergence of sequences $\{\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\}_N$ and $\{[\nabla_{\boldsymbol{\eta}} \hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})]\}_N$). *Let $\boldsymbol{\lambda}$ be fixed in $C_{\text{ad},\lambda}$ and let $\boldsymbol{\eta}$ be fixed in $\mathbb{R}^\nu$. Under the hypothesis $\boldsymbol{h}^c \in C^1(\mathbb{R}^\nu, \mathbb{R}^{n_c})$ (see Eq. (3.7)), $\forall \varepsilon > 0$, there exists a finite integer $N_\varepsilon(\boldsymbol{\eta},\boldsymbol{\lambda})$ depending on ε, $\boldsymbol{\eta}$, and $\boldsymbol{\lambda}$, such that $\forall N \geq N_\varepsilon(\boldsymbol{\eta},\boldsymbol{\lambda})$,*

$$\| \hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) - \boldsymbol{h}^c(\boldsymbol{\eta}) \| \;\leq\; \varepsilon \quad , \quad \| [\nabla_{\boldsymbol{\eta}} \hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})] - [\nabla_{\boldsymbol{\eta}} \boldsymbol{h}^c(\boldsymbol{\eta})] \|_F \;\leq\; \varepsilon\,. \tag{3.35}$$

Proof. (Proposition 2). The probability measure $P_{\boldsymbol{H}_\lambda}(d\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ admits a continuous density $p_{\boldsymbol{H}_\lambda}(.\,;\boldsymbol{\lambda})$ with respect to $d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ (see Eq. (3.13)). Using the notation of Definition 3, for $s > 0$, let $\tilde{\boldsymbol{\eta}} \mapsto \kappa_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}\,; s, \boldsymbol{\lambda}) = (\sqrt{2\pi}\, s)^{-\nu} (det[\sigma_{\boldsymbol{H}_\lambda}])^{-1}$ $\exp\{-\frac{1}{2s^2} \| \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \|^2_H\}$ be the function defined on $\mathbb{R}^\nu$ with values in $\mathbb{R}^+$. Since $\forall \alpha \in \{1, \ldots, \nu\}$, $\lim_{s\to 0_+} \{s\, [\sigma_{\boldsymbol{H}_\lambda}]_{\alpha\alpha}\} = 0$, it can be seen that

$$\lim_{s\to 0_+} \kappa_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}\,; s, \boldsymbol{\lambda})\, d\tilde{\boldsymbol{\eta}} = \delta_0(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta})\,, \tag{3.36}$$

in the vector space of bounded measures, in which $\delta_0(\tilde{\boldsymbol{\eta}})$ is the Dirac measure on $\mathbb{R}^\nu$ at point $\tilde{\boldsymbol{\eta}} = \mathbf{0}_\nu$. Eq. (3.27) with Eqs. (3.28) and (3.29) can be rewritten as $\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s_{\text{SB}},\boldsymbol{\lambda})$ in which for all $s > 0$, $\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s,\boldsymbol{\lambda}) = \Big((1/N)\sum_{\ell=1}^N \boldsymbol{h}^c(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^\ell)\,\kappa_{\boldsymbol{\eta}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^\ell\,;s,\boldsymbol{\lambda})\Big)\Big((1/N)\sum_{\ell'=1}^N \kappa_{\boldsymbol{\eta}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^{\ell'}\,;s,\boldsymbol{\lambda})\Big)^{-1}$. Since $\boldsymbol{h}^c$, $p_{\boldsymbol{H}_\lambda}(.\,;\boldsymbol{\lambda})$, and $\kappa_{\boldsymbol{\eta}}(.\,;s,\boldsymbol{\lambda})$ are continuous on $\mathbb{R}^\nu$, and since for any value of N, $\boldsymbol{\eta}_{\boldsymbol{\lambda}}^1,\ldots,\boldsymbol{\eta}_{\boldsymbol{\lambda}}^N$ are N independent realizations of $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}_{\boldsymbol{\lambda}}$, for $s > 0$ fixed, $\lim_{N\to+\infty}\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s,\boldsymbol{\lambda}) = \Big(E\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})\,\kappa_{\boldsymbol{\eta}}(\boldsymbol{H}_{\boldsymbol{\lambda}}\,;s,\boldsymbol{\lambda})\}\Big)\Big(E\{\kappa_{\boldsymbol{\eta}}(\boldsymbol{H}_{\boldsymbol{\lambda}}\,;s,\boldsymbol{\lambda})\}\Big)^{-1}$, which can be rewritten as $\lim_{N\to+\infty}\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s,\boldsymbol{\lambda}) = \Big(\int_{\mathbb{R}^\nu}\boldsymbol{h}^c(\tilde{\boldsymbol{\eta}})\,p_{\boldsymbol{H}_\lambda}(\tilde{\boldsymbol{\eta}}\,;\boldsymbol{\lambda})\,\kappa_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}\,;s,\boldsymbol{\lambda})\Big)\Big(\int_{\mathbb{R}^\nu}p_{\boldsymbol{H}_\lambda}(\tilde{\boldsymbol{\eta}}\,;\boldsymbol{\lambda})\,\kappa_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}\,;s,\boldsymbol{\lambda})\Big)^{-1}$. Since $\boldsymbol{h}^c$ belongs to $C^1(\mathbb{R}^\nu,\mathbb{R}^{n_c})$ and $p_{\boldsymbol{H}_\lambda}(.\,;\boldsymbol{\lambda})$ to $C^0(\mathbb{R}^\nu,\mathbb{R}^+)$, using Eq. (3.36) yields, for $\boldsymbol{\eta}$ fixed in $\mathbb{R}^\nu$, $\lim_{s\to0_+}\lim_{N\to+\infty}\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s,\boldsymbol{\lambda}) = \boldsymbol{h}^c(\boldsymbol{\eta}) \in \mathbb{R}^\nu$ and $\lim_{s\to0_+}\lim_{N\to+\infty}[\nabla_{\boldsymbol{\eta}}\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s,\boldsymbol{\lambda})] = [\nabla_{\boldsymbol{\eta}}\boldsymbol{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu,n_c}$. Consequently, $\forall\boldsymbol{\lambda}\in\mathcal{C}_{\text{ad},\boldsymbol{\lambda}}$, $\forall\boldsymbol{\eta}\in\mathbb{R}^\nu$, $\forall\varepsilon>0$, there exists $s_\varepsilon > 0$ and a finite integer $N_\varepsilon(\boldsymbol{\eta},\boldsymbol{\lambda})$ depending on ε, $\boldsymbol{\eta}$, and $\boldsymbol{\lambda}$, such that $\|\widehat{\mathbb{h}}^{N_\varepsilon}(\boldsymbol{\eta}\,;s_\varepsilon,\boldsymbol{\lambda}) - \boldsymbol{h}^c(\boldsymbol{\eta})\| \le \varepsilon$ and $\|[\nabla_{\boldsymbol{\eta}}\widehat{\mathbb{h}}^{N_\varepsilon}(\boldsymbol{\eta}\,;s_\varepsilon,\boldsymbol{\lambda})] - [\nabla_{\boldsymbol{\eta}}\boldsymbol{h}^c(\boldsymbol{\eta})]\| \le \varepsilon$. The Silverman bandwidth $s_{\text{SB}} = s_{\text{SB}}(N)$ defined by Eq. (3.30), goes to 0 when $N \to +\infty$. Therefore, choosing $N \ge N_\varepsilon(\boldsymbol{\eta},\boldsymbol{\lambda})$ such that $s_{\text{SB}}(N) < s_\varepsilon$ (that is always possible) yields $\|\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s_{\text{SB}}(N),\boldsymbol{\lambda}) - \boldsymbol{h}^c(\boldsymbol{\eta})\| \le \varepsilon$ and $\|[\nabla_{\boldsymbol{\eta}}\widehat{\mathbb{h}}^N(\boldsymbol{\eta}\,;s_{\text{SB}}(N),\boldsymbol{\lambda})] - [\nabla_{\boldsymbol{\eta}}\boldsymbol{h}^c(\boldsymbol{\eta})]\| \le \varepsilon$, which proves Eq. (3.35).

Notation (*Normalized $\mathbb{R}^\nu$-valued Wiener stochastic process*). Let $\{\boldsymbol{W}^{\text{wien}}(t) = (W_1^{\text{wien}}(t),\ldots,W_\nu^{\text{wien}}(t)), t \ge 0\}$ be the Wiener process, defined on $(\Theta,\mathcal{T},\mathcal{P})$, indexed by $\mathbb{R}^+$, with values in $\mathbb{R}^\nu$, such that $W_1^{\text{wien}},\ldots,W_\nu^{\text{wien}}$ are mutually independent, $\boldsymbol{W}^{\text{wien}}(0) = \mathbf{0}_\nu$ a.s., $\boldsymbol{W}^{\text{wien}}$ is a process with independent increments such that, for all $0 \le t' < t < +\infty$, the increment $\boldsymbol{W}^{\text{wien}}(t) - \boldsymbol{W}^{\text{wien}}(t')$ is a $\mathbb{R}^\nu$-valued second-order, Gaussian, centered random variable whose covariance matrix is $(t-t')\,[I_\nu]$.

Proposition 3 (MCMC generator of $\boldsymbol{H}_{\boldsymbol{\lambda}}$). *Let $\boldsymbol{h}^c$ be the function satisfying Hypothesis 1. Let $\boldsymbol{\lambda}$ be fixed in $\mathcal{C}_{\text{ad},\boldsymbol{\lambda}}$. Consequently, Lemma 1-(d) holds. Let $\{(\boldsymbol{U}_{\boldsymbol{\lambda}}(t),\boldsymbol{V}_{\boldsymbol{\lambda}}(t)), t \ge 0\}$ be the stochastic process, defined on $(\Theta,\mathcal{T},\mathcal{P})$, indexed by $\mathbb{R}^+$, with values in $\mathbb{R}^\nu\times\mathbb{R}^\nu$, which verifies the following ISDE for $t > 0$, with the initial condition $(\boldsymbol{u}_0,\boldsymbol{v}_0)$ given in $\mathbb{R}^\nu\times\mathbb{R}^\nu$,*

$$d\boldsymbol{U}_{\boldsymbol{\lambda}}(t) = \boldsymbol{V}_{\boldsymbol{\lambda}}(t)\,dt\,, \tag{3.37}$$

$$d\boldsymbol{V}_{\boldsymbol{\lambda}}(t) = \boldsymbol{L}_{\boldsymbol{\lambda}}(\boldsymbol{U}_{\boldsymbol{\lambda}}(t))\,dt - \frac{1}{2}f_0\,\boldsymbol{V}_{\boldsymbol{\lambda}}(t)\,dt + \sqrt{f_0}\,d\boldsymbol{W}^{\text{wien}}(t)\,, \tag{3.38}$$

$$\boldsymbol{U}_{\boldsymbol{\lambda}}(0) = \boldsymbol{u}_0\ ,\ \boldsymbol{V}_{\boldsymbol{\lambda}}(0) = \boldsymbol{v}_0\ a.s. \tag{3.39}$$

(a) The initial condition $\boldsymbol{u}_0 \in \mathbb{R}^\nu$ is chosen from the points of the training set $\mathcal{D}_d = \{\boldsymbol{\eta}_d^1,\ldots,\boldsymbol{\eta}_d^{N_d}\}$ (see Section 4.2) while the initial condition $\boldsymbol{v}_0$ is chosen as any realization of a normalized Gaussian $\mathbb{R}^\nu$-valued random variable $\boldsymbol{V}_G$, independent of $\boldsymbol{W}^{\text{wien}}$, whose probability density function with respect to $d\boldsymbol{v}$ is $p_{\boldsymbol{V}_G}(\boldsymbol{v}) = (2\pi)^{-\nu/2}\,\exp\{-\|\boldsymbol{v}\|^2/2\}$.
(b) The parameter $f_0 > 0$ allows the dissipation term in the dissipative Hamiltonian system to be controlled and to rapidly reach the stationary response associated with the invariant measure (the value $f_0 = 4$ is generally a good choice).
(c) For all $\boldsymbol{u} = (u_1,\ldots,u_\nu)$ in $\mathbb{R}^\nu$, the vector $\boldsymbol{L}_{\boldsymbol{\lambda}}(\boldsymbol{u})$ in $\mathbb{R}^\nu$ is defined by $\boldsymbol{L}_{\boldsymbol{\lambda}}(\boldsymbol{u}) = -\nabla_{\boldsymbol{u}}\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})$ that can be written as

$$\boldsymbol{L}_{\boldsymbol{\lambda}}(\boldsymbol{u}) = \frac{1}{\zeta(\boldsymbol{u})}\,\nabla_{\boldsymbol{u}}\zeta(\boldsymbol{u}) - [\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\,\boldsymbol{\lambda}\,. \tag{3.40}$$

(d) The stochastic solution $\{(\boldsymbol{U}_{\boldsymbol{\lambda}}(t),\boldsymbol{V}_{\boldsymbol{\lambda}}(t)), t \ge 0\}$ of the ISDE defined by Eqs. (3.37) to (3.39) is unique, has almost-surely continuous trajectories, and is a second-order diffusion stochastic process. For $t \to +\infty$, this diffusion process converges to a stationary second-order diffusion stochastic process $\{(\boldsymbol{U}_{\boldsymbol{\lambda}}^{\text{st}}(\tau),\boldsymbol{V}_{\boldsymbol{\lambda}}^{\text{st}}(\tau)), \tau \ge 0\}$ associated with the unique invariant probability measure on $\mathbb{R}^\nu\times\mathbb{R}^\nu$,

$$p_{\boldsymbol{H}_\lambda,\boldsymbol{V}_G}(\boldsymbol{\eta},\boldsymbol{v}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}\otimes d\boldsymbol{v} = (p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta})\otimes(p_{\boldsymbol{V}_G}(\boldsymbol{v})\,d\boldsymbol{v})\,, \tag{3.41}$$

in which $p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ is the pdf defined by Eq. (3.13).
(e) For t_s sufficiently large, we can choose $\boldsymbol{H}_{\boldsymbol{\lambda}}$ as $\boldsymbol{U}_{\boldsymbol{\lambda}}(t_s)$. The generation of the constrained learned set $\mathcal{D}_{\boldsymbol{H}_\lambda} = \{\boldsymbol{\eta}_{\boldsymbol{\lambda}}^1,\ldots,\boldsymbol{\eta}_{\boldsymbol{\lambda}}^N\}$, made up of $N \gg N_d$ independent realizations of $\boldsymbol{H}_{\boldsymbol{\lambda}}$ whose probability measure is $p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}$, consists in solving Eqs. (3.37) to (3.39) for $t \in [0\,,t_s]$ and then using the realizations of $\boldsymbol{U}_{\boldsymbol{\lambda}}(t_s)$ (see the numerical aspects in Section 4).

PROOF. (Proposition 3). Since $\boldsymbol{h}^c \in C^1(\mathbb{R}^\nu, \mathbb{R}^{n_c})$ and $\phi(\boldsymbol{u}) = -\log\zeta(\boldsymbol{u})$ with $\zeta(\boldsymbol{u})$ given by Eq. (3.3), function $\boldsymbol{u} \mapsto \mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})$ defined by Eq. (3.11) belongs to $C^1(\mathbb{R}^\nu, \mathbb{R})$. Therefore, $\boldsymbol{u} \mapsto \|\nabla_{\boldsymbol{u}}\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})\|$ is locally bounded on $\mathbb{R}^\nu$. Using Eqs. (3.11) and (3.12), it can be seen that, for all $\boldsymbol{\lambda} \in C_{\text{ad},\boldsymbol{\lambda}}$, $\inf_{\|\boldsymbol{u}\|>R}\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u}) \to +\infty$ if $R \to +\infty$, and $\inf_{\boldsymbol{u}\in\mathbb{R}^\nu}\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})$ is a finite real number. Using Eqs. (3.3), (3.4), and (3.11) yields

$$\int_{\mathbb{R}^\nu} \|\nabla_{\boldsymbol{u}}\mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})\|\, p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{u}\,;\boldsymbol{\lambda})\, d\boldsymbol{u} \le \int_{\mathbb{R}^\nu} \frac{1}{\zeta(\boldsymbol{u})}\|\nabla_{\boldsymbol{u}}\zeta(\boldsymbol{u})\|\, p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{u}\,;\boldsymbol{\lambda})\, d\boldsymbol{u} + \int_{\mathbb{R}^\nu} \|[\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\|_F\, \|\boldsymbol{\lambda}\|\, p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{u}\,;\boldsymbol{\lambda})\, d\boldsymbol{u}\,, \tag{3.42}$$

because $\|[\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\,\boldsymbol{\lambda}\| \le \|[\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\|\,\|\boldsymbol{\lambda}\|$ and $\|[\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\| \le \|[\nabla_{\boldsymbol{u}}\boldsymbol{h}^c(\boldsymbol{u})]\|_F$. From Eqs. (3.12) and (3.3), the first term in the right-hand side member of Eq. (3.42) is finite, while from the second equation (3.15), the second term is also finite. It can then be deduced that the left-hand side member of Eq. (3.42) is finite. Consequently, Theorems 6, 7, and 9 in Pages 214 to 216 of [129], and the expression of the invariant measure given by Theorem 4 in Page 211 of the same reference, for which the Hamiltonian is $\mathcal{H}(\boldsymbol{u},\boldsymbol{v}) = \|\boldsymbol{v}\|^2/2 + \mathcal{V}_{\boldsymbol{\lambda}}(\boldsymbol{u})$, prove that the solution of Eqs. (3.37) to (3.39) is unique and is a second-order diffusion stochastic process with almost-surely continuous trajectories, which converges for $t \to +\infty$ to a second-order stationary diffusion process with almost surely continuous trajectories $\{(\boldsymbol{U}^{\text{st}}_{\boldsymbol{\lambda}}(\tau), \boldsymbol{V}^{\text{st}}_{\boldsymbol{\lambda}}(\tau)), \tau \ge 0\}$ associated with the invariant probability measure given by Eq. (3.41). For any $\tau > 0$, $\boldsymbol{U}^{\text{st}}_{\boldsymbol{\lambda}}(\tau) = \lim_{t\to+\infty}\boldsymbol{U}_{\boldsymbol{\lambda}}(t+\tau)$ in probability measure.

Proposition 4 (Convergence of the sequence of MCMC generator using the statistical surrogate model). *Let $\boldsymbol{\lambda}$ be fixed in $C_{\text{ad},\boldsymbol{\lambda}}$ and let us use Proposition 3. For all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, let $\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ be the approximation of $\boldsymbol{h}^c(\boldsymbol{\eta})$ defined by Eq.* (3.27) *and let $\boldsymbol{u} \mapsto \hat{\boldsymbol{L}}_{\boldsymbol{\lambda}}(\boldsymbol{u})$ be the twice continuously differentiable function on $\mathbb{R}^\nu$ with values in $\mathbb{R}^\nu$ such that, for all $\boldsymbol{u}$ in $\mathbb{R}^\nu$,*

$$\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}}(\boldsymbol{u}) = \frac{1}{\zeta(\boldsymbol{u})}\,\nabla_{\boldsymbol{u}}\zeta(\boldsymbol{u}) - [\nabla_{\boldsymbol{u}}\hat{\boldsymbol{h}}^N(\boldsymbol{u}\,;\boldsymbol{\lambda})]\,\boldsymbol{\lambda}\,, \tag{3.43}$$

in which $\boldsymbol{\zeta}$ is defined by Eq. (3.3) *and where $[\nabla_{\boldsymbol{u}}\hat{\boldsymbol{h}}^N(\boldsymbol{u}\,;\boldsymbol{\lambda})]$ is explicitly given (see Eq.* (4.8)*) by differentiating function $\boldsymbol{u} \mapsto \hat{\boldsymbol{h}}^N(\boldsymbol{u}\,;\boldsymbol{\lambda})$ defined by Eq.* (3.27)*. Let $\{(\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t), \boldsymbol{V}^N_{\boldsymbol{\lambda}}(t)), t \ge 0\}$ be the stochastic process solution of the ISDE defined by Eqs.* (3.37) *to* (3.39) *in which $\boldsymbol{L}_{\boldsymbol{\lambda}}$ is replaced by $\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}}$,*

$$d\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t) = \boldsymbol{V}^N_{\boldsymbol{\lambda}}(t)\,dt\,, \tag{3.44}$$

$$d\boldsymbol{V}^N_{\boldsymbol{\lambda}}(t) = \hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}}(\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t))\,dt - \frac{1}{2}f_0\,\boldsymbol{V}^N_{\boldsymbol{\lambda}}(t)\,dt + \sqrt{f_0}\,d\boldsymbol{W}^{\text{wien}}(t)\,, \tag{3.45}$$

$$\boldsymbol{U}^N_{\boldsymbol{\lambda}}(0) = \boldsymbol{u}_0\ ,\ \boldsymbol{V}^N_{\boldsymbol{\lambda}}(0) = \boldsymbol{v}_0\ a.s.\,, \tag{3.46}$$

and where $\boldsymbol{u}_0$, $\boldsymbol{v}_0$, f_0, and $\boldsymbol{W}^{\text{wien}}$ are the quantities defined in Proposition 3. Then the stochastic solution $\{(\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t), \boldsymbol{V}^N_{\boldsymbol{\lambda}}(t)), t \ge 0\}$ of Eqs. (3.44) *to* (3.46) *is unique, has almost-surely continuous trajectories, and is a second-order diffusion stochastic process, which converges to a stationary second-order diffusion stochastic process for $t \to +\infty$, associated with the unique invariant probability measure on $\mathbb{R}^\nu \times \mathbb{R}^\nu$, $\hat{p}^N_{\boldsymbol{H}_{\boldsymbol{\lambda}},V_G}(\boldsymbol{\eta},\boldsymbol{v}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}\otimes d\boldsymbol{v} = (\hat{p}^N_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta})\otimes(p_{V_G}(\boldsymbol{v})\,d\boldsymbol{v})$, in which $\hat{p}^N_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \hat{c}^N_0(\boldsymbol{\lambda})\,\exp\{-\widehat{\mathcal{V}}^N_{\boldsymbol{\lambda}}(\boldsymbol{\eta})\}$ with $\widehat{\mathcal{V}}^N_{\boldsymbol{\lambda}}(\boldsymbol{\eta}) = -\log\zeta(\boldsymbol{\zeta}) + \langle\boldsymbol{\lambda}\,,\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\rangle$. Then for all $t \in [0\,,t_s]$ with $t_s < +\infty$, the sequence $\{\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t)\}_N$ of second-order $\mathbb{R}^\nu$-valued random variables converges in mean-square to the second-order $\mathbb{R}^\nu$-valued random variable $\boldsymbol{U}_{\boldsymbol{\lambda}}(t)$ of Proposition 3,*

$$\lim_{N\to+\infty} E\{\|\boldsymbol{U}^N_{\boldsymbol{\lambda}}(t) - \boldsymbol{U}_{\boldsymbol{\lambda}}(t)\|^2\} = 0 \quad , \quad \forall t \in [0\,,t_s]\,. \tag{3.47}$$

PROOF. (Proposition 4). The classical theorem, such as Theorem 5.1 Page 118 of [133], cannot directly be used because, for $\|\boldsymbol{\eta}\|\to+\infty$, the growth of the drift vector must be at most linear, which is not verified taking into account Eq. (3.40) and the fact that $\|[\nabla_{\boldsymbol{\eta}}\boldsymbol{h}^c(\boldsymbol{\eta})]\|_F$ can have any polynomial growth at infinity (see Eq. (3.8)). Consequently, an adapted proof of this Proposition 4 must be done. For $\boldsymbol{\lambda}$ be fixed in $C_{\text{ad},\boldsymbol{\lambda}}$, the unique second-order stochastic process with almost-surely continuous trajectories $\{(\boldsymbol{U}_{\boldsymbol{\lambda}}(t), \boldsymbol{V}_{\boldsymbol{\lambda}}(t)), t \ge 0\}$ of Proposition 3 can be written as

$$\boldsymbol{Z}_{\boldsymbol{\lambda}}(t) = \boldsymbol{z}_0 + \int_0^t \mathbb{o}_{\boldsymbol{\lambda}}(\boldsymbol{Z}_{\boldsymbol{\lambda}}(\tau))\,d\tau + \int_0^t [\mathbb{b}]\,d\boldsymbol{W}^{\text{wien}}(\tau)\,, \tag{3.48}$$

in which $\boldsymbol{z}_0 = (\boldsymbol{u}_0, \boldsymbol{v}_0)$, $\mathbb{o}_{\boldsymbol{\lambda}}(\boldsymbol{z}) = (\boldsymbol{v}\,, \boldsymbol{L}_{\boldsymbol{\lambda}}(\boldsymbol{u}) - (1/2)\,f_0\,\boldsymbol{v})$ and where $\boldsymbol{z} = (\boldsymbol{u},\boldsymbol{v})$ with $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^\nu$, where $\boldsymbol{z}_0$, $\boldsymbol{z}$, and $\mathbb{o}_{\boldsymbol{\lambda}}(\boldsymbol{z})$ are in $\mathbb{R}^{2\nu} = \mathbb{R}^\nu\times\mathbb{R}^\nu$, and where $[\mathbb{b}] = [\,[0_\nu]\ \sqrt{f_0}\,[I_\nu]\,]^T \in \mathbb{M}_{2\nu,\nu}$. Reusing the proof of Proposition 3, it can be seen

that Eqs. (3.44) to (3.46) admits a unique solution $\{(\boldsymbol{U}_\lambda^N(t), \boldsymbol{V}_\lambda^N(t)), t \geq 0\}$ (with the properties given in Proposition 3), which can be written as,

$$\boldsymbol{Z}_\lambda^N(t) = z_0 + \int_0^t \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda^N(\tau))\, d\tau + \int_0^t [\mathbb{b}]\, d\boldsymbol{W}^{\text{wien}}(\tau)\,, \tag{3.49}$$

in which $\widehat{\mathbb{o}}_\lambda^N(z) = (\boldsymbol{v}\,, \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{u}) - (1/2)\, f_0\, \boldsymbol{v})$. Subtracting the two equations Eqs. (3.48) and (3.49) yields

$$\boldsymbol{Z}_\lambda^N(t) - \boldsymbol{Z}_\lambda(t) = \boldsymbol{\chi}_\lambda^N(t) + \int_0^t \left(\widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda^N(\tau)) - \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda(\tau))\right) d\tau\,,$$

$$\boldsymbol{\chi}_\lambda^N(t) = \int_0^t \left(\widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda(\tau)) - \mathbb{o}_\lambda(\boldsymbol{Z}_\lambda(\tau))\right) d\tau\,. \tag{3.50}$$

Let $L^2(\Theta, \mathbb{R}^m)$ be the Hilbert space of the equivalent classes of second-order $\mathbb{R}^m$-valued random variables defined on $(\Theta, \mathcal{T}, \mathcal{P})$, equipped with the inner product $\langle\langle \boldsymbol{A}\,, \boldsymbol{A}'\rangle\rangle = E\{\langle \boldsymbol{A}\,, \boldsymbol{A}'\rangle\}$ and the associated norm $|||\, \boldsymbol{A}\, |||= (E\{\|\, \boldsymbol{A}\, \|^2\})^{1/2}$. Let us define $\boldsymbol{G}(\tau) = \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda^N(\tau)) - \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda(\tau))$ and $\mathbb{1}(\tau) = 1$. The Schwarz inequality $\int_0^t \mathbb{1}(\tau) \times |||\, \boldsymbol{G}(\tau)\, ||| \, d\tau \leq (\int_0^t \mathbb{1}(\tau)^2 d\tau)^{1/2} (\int_0^t ||| \, \boldsymbol{G}(\tau)\, |||^2 \, d\tau)^{1/2}$ yields $\left(\int_0^t ||| \, \boldsymbol{G}(\tau)\, ||| \, d\tau\right)^2 \leq t \int_0^t ||| \, \boldsymbol{G}(\tau)\, |||^2 \, d\tau$. Using the inequality $2\alpha\beta \leq 2\alpha^2 + \beta^2/2$ for all $\alpha > 0$ and $\beta > 0$ and since $|||\, \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda^N(\tau)) - \widehat{\mathbb{o}}_\lambda^N(\boldsymbol{Z}_\lambda(\tau))\, |||^2 = (1 + f_0^2/4) ||| \, \boldsymbol{V}_\lambda^N(\tau) - \boldsymbol{V}_\lambda(\tau)\, |||^2 + ||| \, \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda^N(\tau)) - \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda(\tau))\, |||^2$, we obtain

$$||| \, \boldsymbol{Z}_\lambda^N(t) - \boldsymbol{Z}_\lambda(t)\, |||^2 \leq 3\, ||| \boldsymbol{\chi}_\lambda^N(t)\, |||^2 + \frac{3}{2} t \int_0^t \left((1 + \frac{f_0^2}{4}) ||| \, \boldsymbol{V}_\lambda^N(\tau) - \boldsymbol{V}_\lambda(\tau)\, |||^2 + ||| \, \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda^N(\tau)) - \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda(\tau))\, |||^2 \right) d\tau. \tag{3.51}$$

Let t be fixed such that $0 < t \leq t_s < +\infty$. Since $\{\boldsymbol{U}_\lambda^N(\tau), \tau \geq 0\}$ and $\{\boldsymbol{U}_\lambda(\tau), \tau \geq 0\}$ are dependent second-order $\mathbb{R}^\nu$-valued stochastic process with almost-surely continuous trajectories, there exists a finite positive constant $r_\lambda(t_s)$ depending on $\boldsymbol{\lambda}$ and t_s, $0 < r_\lambda(t_s) < +\infty$, such that $\sup_{0 \leq \tau \leq t_s} ||| \, \boldsymbol{U}_\lambda^N(\tau)\, ||| < r_\lambda(t_s)$ and $\sup_{0 \leq \tau \leq t_s} ||| \, \boldsymbol{U}_\lambda(\tau)\, ||| < r_\lambda(t_s)$. Let $\mathcal{U}_{t_s}$ be the open ball of $L^2(\Theta, \mathbb{R}^\nu)$ such that $\mathcal{U}_{t_s} = \left\{\boldsymbol{U} \in L^2(\Theta, \mathbb{R}^\nu)\,; ||| \, \boldsymbol{U}\, ||| < r_\lambda(t_s)\right\}$. Due to the convexity of the open ball in a normed vector space, $\mathcal{U}_{t_s}$ is a convex open set of $L^2(\Theta, \mathbb{R}^\nu)$ ($\forall \boldsymbol{U}, \boldsymbol{U}' \in \mathcal{U}_{t_s}$, $\forall \mu \in [0\,, 1]$, we have $||| \, (1-\mu)\, \boldsymbol{U} + \mu\, \boldsymbol{U}'\, ||| \leq (1-\mu) ||| \, \boldsymbol{U}\, ||| + \mu\, ||| \boldsymbol{U}'\, ||| \leq (1-\mu)\, r_\lambda(t_s) + \mu\, r_\lambda(t_s) = r_\lambda(t_s)$, which shows that $(1-\mu)\, \boldsymbol{U} + \mu\, \boldsymbol{U}' \in \mathcal{U}_{t_s}$). Let $\boldsymbol{U} \mapsto \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U})$ be the mapping from $L^2(\Theta, \mathbb{R}^\nu)$ into $L^2(\Theta, \mathbb{R}^\nu)$, in which $\hat{\boldsymbol{L}}_\lambda^N$ is defined by Eq. (3.43). Since $\boldsymbol{u} \mapsto \hat{\boldsymbol{h}}^N(\boldsymbol{u}\,; \boldsymbol{\lambda})$ is twice continuously differentiable on $\mathbb{R}^\nu$, then function $\boldsymbol{u} \mapsto \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{u})$ is continuously differentiable on $\mathbb{R}^\nu$. It can easily be verified that $\forall \boldsymbol{U} \in \mathcal{U}_{t_s}$, $\left(E\{\|\, [\nabla_{\boldsymbol{u}} \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U})]\, \|_F^2\}\right)^{1/2} \leq k_\lambda(t_s)$, in which $k_\lambda(t_s)$ is a finite positive constant depending on $\boldsymbol{\lambda}$ and t_s. Consequently, using Theorem 3.3.2 Page 45 of [134] for Banach spaces, for all $0 \leq \tau \leq t \leq t_s$, we have

$$||| \, \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda^N(\tau)) - \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda(\tau))\, ||| \leq k_\lambda(t_s) ||| \, \boldsymbol{U}_\lambda^N(\tau) - \boldsymbol{U}_\lambda(\tau)\, |||\,. \tag{3.52}$$

From Eqs. (3.51) and (3.52), it can be deduced using the Gronwall Lemma (see [135] or [136], Page 362) that, for all $0 \leq t \leq t_s < +\infty$,

$$||| \, \boldsymbol{Z}_\lambda^N(t) - \boldsymbol{Z}_\lambda(t)\, |||^2 \leq c_\lambda(t_s) ||| \boldsymbol{\chi}_\lambda^N(t)\, |||^2\,, \tag{3.53}$$

with $c_\lambda(t_s) = (9/2)\, t_s^2 \, \max\{(1 + f_0^2/4)\,, k_\lambda(t_s)^2\} < +\infty$. Eq. (3.50) yields $||| \boldsymbol{\chi}_\lambda^N(t)\, ||| \leq \int_0^t ||| \, \hat{\boldsymbol{L}}_\lambda^N(\boldsymbol{U}_\lambda(\tau)) - \boldsymbol{L}_\lambda(\boldsymbol{U}_\lambda(\tau))\, ||| \, d\tau$. Using Eqs. (3.40) and (3.43) yields

$$||| \boldsymbol{\chi}_\lambda^N(t)\, ||| \leq \|\, \boldsymbol{\lambda}\, \| \int_0^t \left(E\{\|\, [\nabla_{\boldsymbol{u}} \hat{\boldsymbol{h}}^N(\boldsymbol{U}_\lambda(\tau)\,; \boldsymbol{\lambda})] - [\nabla_{\boldsymbol{u}} \boldsymbol{h}^c(\boldsymbol{U}_\lambda(\tau))]\, \|_F^2\right)^{1/2} d\tau\,. \tag{3.54}$$

From Proposition 2, it can be deduced that, for $N \to +\infty$, the right-hand side member of Eq. (3.54) goes to 0 and consequently, Eq. (3.53) shows that $\boldsymbol{Z}_\lambda^N(t) \to \boldsymbol{Z}_\lambda(t)$ for the mean-square convergence.

Iterative algorithm for calculating $\boldsymbol{\lambda}^{\text{sol}}$. Under Proposition 1, for $\boldsymbol{\lambda} \in \mathcal{C}_{\text{ad},\lambda}$, since $\Gamma(\boldsymbol{\lambda})$ cannot be evaluated in high dimension using Eq. (3.18) due to the presence of constant $c_0(\boldsymbol{\lambda})$ (the normalization constant), $\boldsymbol{\lambda}^{\text{sol}}$ cannot directly be estimated using the gradient descent algorithm applied to the convex optimization problem defined by Eq. (3.21). We

will then assumed that $\boldsymbol{\lambda}^{\rm sol}$ can be calculated as the unique solution in $\boldsymbol{\lambda}$ of equation $\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}) = \mathbf{0}_{n_c}$ (see Proposition 1-(c)) and in particular Eq. (3.22)), that is to say (see Eq. (3.19)), solving the following equation in $\boldsymbol{\lambda}$ on $\mathbb{R}^{n_c}$,

$$E\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})\} - \boldsymbol{b}^c = \mathbf{0}_{n_c} \,. \tag{3.55}$$

This equation is solved using the Newton iterative method [137] applied to function $\boldsymbol{\lambda} \mapsto \nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda})$ as proposed in [138, 139], that is to say,

$$\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - [\Gamma''(\boldsymbol{\lambda}^i)]^{-1}\, \nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}^i) \quad , \quad i = 0, 1, \ldots, i_{\max} \,, \tag{3.56}$$

with $\boldsymbol{\lambda}^0 = \mathbf{0}_{n_c}$, in which $\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda})$ and $[\Gamma''(\boldsymbol{\lambda})]$ are defined by Eqs. (3.19) and (3.20), and where $i_{\max}$ is a given integer sufficiently large. An estimation of $\boldsymbol{\lambda}^{\rm sol}$ is chosen as

$$\boldsymbol{\lambda}^{\rm sol} = \boldsymbol{\lambda}^{i_{\rm sol}} \quad , \quad i_{\rm sol} = \arg \min_{i=1,\ldots,i_{\max}} {\rm err}(i) \,, \tag{3.57}$$

in which the error function $i \mapsto {\rm err}(i) : \{1, \ldots, i_{\max}\} \to \mathbb{R}^+$ is defined as follows for taking into account the possible types of algebraic quantities (scalar, vectors, tensors) that are used for defining function $\boldsymbol{h}^c$. Therefore, let M be an integer such that $1 \leq M \leq n_c$ and for which $\boldsymbol{h}^c$ and $\boldsymbol{b}^c$ are written as $\boldsymbol{h}^c(\boldsymbol{\eta}) = (\boldsymbol{h}^{c,k_1}(\boldsymbol{\eta}), \ldots, \boldsymbol{h}^{c,k_M}(\boldsymbol{\eta}))$ and $\boldsymbol{b}^c = (\boldsymbol{b}^{c,k_1}, \ldots, \boldsymbol{b}^{c,k_M})$ with $\sum_{m=1}^M k_m = n_c$.The error function is then defined by

$${\rm err}(i) = \left(\sum_{m=1}^M w_m \left(\frac{{\rm err}_m(i)}{{\rm err}_m(1)} \right)^2 \right)^{1/2} \quad , \tag{3.58}$$

in which $\{w_m \geq 0\,, m = 1, \ldots M\}$ are given real numbers and where

$${\rm err}_m(i) = \frac{1}{\| \boldsymbol{b}^{c,k_m} \|} \| \boldsymbol{b}^{c,k_m} - E\{\boldsymbol{h}^{c,k_m}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\} \| \,. \tag{3.59}$$

Proposition 5 (Rate of convergence of the sequence $\{p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}\}_{\boldsymbol{\lambda}}$). *Let $\boldsymbol{\lambda}^{i+1}$ and $\boldsymbol{\lambda}^i$ be given values of $\boldsymbol{\lambda}$ in $C_{{\rm ad},\boldsymbol{\lambda}}$ and let $p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\,.\,;\boldsymbol{\lambda})$ be the pdf of $\boldsymbol{H}_{\boldsymbol{\lambda}}$ defined by Eq.* (3.13)*. For $\| \boldsymbol{\lambda}^{i+1} - \boldsymbol{\lambda}^i \|$ sufficiently small, we have*

$$\| p_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i+1}}}(\,.\,;\boldsymbol{\lambda}^{i+1}) - p_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}}(\,.\,;\boldsymbol{\lambda}^i) \|_{L^1(\mathbb{R}^\nu,\mathbb{R})} \;\leq\; \| \boldsymbol{\lambda}^{i+1} - \boldsymbol{\lambda}^i \| \left(tr\,[\Gamma''(\boldsymbol{\lambda}^i)] \right)^{1/2} + o\left(\| \boldsymbol{\lambda}^{i+1} - \boldsymbol{\lambda}^i \| \right), \tag{3.60}$$

in which $[\Gamma''(\boldsymbol{\lambda}^i)] \in \mathbb{M}^+_{n_c}$ is defined by Eq. (3.20) *for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^i$.*

PROOF. (Proposition 5). In this proof, for simplifying the writing, $\boldsymbol{\lambda}^i$ is simply written as $\boldsymbol{\lambda}$. Proposition 1 shows that Γ is twice differentiable in $C_{{\rm ad},\boldsymbol{\lambda}}$. For all $\boldsymbol{\eta}$ fixed in $\mathbb{R}^\nu$, the Taylor expansion of $\boldsymbol{\lambda}^{i+1} \mapsto p_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i+1}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}^{i+1})$ around $\boldsymbol{\lambda}$, truncated at the first order, is written as,

$$p_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i+1}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}^{i+1}) = p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) + \langle \nabla_{\boldsymbol{\lambda}} p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,, \boldsymbol{\lambda}^{i+1} - \boldsymbol{\lambda} \rangle + \ldots \tag{3.61}$$

Eq. (3.18) is written as $c_0(\boldsymbol{\lambda}) = \exp\{\langle \boldsymbol{\lambda}\,, \boldsymbol{b}^c \rangle - \Gamma(\boldsymbol{\lambda})\}$ and is substituted in Eq. (3.13) (or equivalently, in Eq. (2.5)) of $p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ yielding $p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \zeta(\boldsymbol{\eta})\, \exp\{-\Gamma(\boldsymbol{\lambda}) - \langle \boldsymbol{\lambda}\,, \boldsymbol{h}^c(\boldsymbol{\eta}) - \boldsymbol{b}^c \rangle\}$. The gradient with respect to $\boldsymbol{\lambda}$ can be written as

$$\nabla_{\boldsymbol{\lambda}} p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = -\left(\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}) + \boldsymbol{h}^c(\boldsymbol{\eta}) - \boldsymbol{b}^c \right)\, p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \,. \tag{3.62}$$

Let us introduce the score variable

$$\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \nabla_{\boldsymbol{\lambda}} \log p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})^{-1}\, \nabla_{\boldsymbol{\lambda}} p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \,, \tag{3.63}$$

which can be rewritten, using Eq. (3.62), as $\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = -\left(\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}) + \boldsymbol{h}^c(\boldsymbol{\eta}) - \boldsymbol{b}^c \right)$ yielding with the use of Eq. (3.19),

$$\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = -\left(\boldsymbol{h}^c(\boldsymbol{\eta}) - E\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})\} \right) \,. \tag{3.64}$$

Eq. (3.64) shows that $E\{\| \boldsymbol{v}(\boldsymbol{H}_{\boldsymbol{\lambda}};\boldsymbol{\lambda}) \|^2\} = E\left\{ \| \boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}}) - E\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})\} \|^2 \right\} = {\rm tr}\,[{\rm cov}\{\boldsymbol{h}^c(\boldsymbol{H}_{\boldsymbol{\lambda}})\}]$ and using Eq. (3.20) yields,

$$E\{\| \boldsymbol{v}(\boldsymbol{H}_{\boldsymbol{\lambda}};\boldsymbol{\lambda}) \|^2\} = \int_{\mathbb{R}^\nu} \| \boldsymbol{v}(\boldsymbol{\eta};\boldsymbol{\lambda}) \|^2\, p_{\boldsymbol{H}_{\boldsymbol{\lambda}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\, d\boldsymbol{\eta} = {\rm tr}\,[\Gamma''(\boldsymbol{\lambda})] \,. \tag{3.65}$$

Eq. (3.63) can be written as $\nabla_{\lambda} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$. Consequently,

$$\int_{\mathbb{R}^\nu} |\,\langle\,\nabla_{\lambda} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\rangle\,|\,d\boldsymbol{\eta} = \int_{\mathbb{R}^\nu} |\,\langle\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\rangle\,|\,p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}$$
$$\leq \|\,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\| \int_{\mathbb{R}^\nu} \|\,\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,\|\,p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta} \;\leq\; \|\,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\| \left(\int_{\mathbb{R}^\nu} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}\right)^{1/2} \left(\int_{\mathbb{R}^\nu} \|\,\boldsymbol{v}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,\|^2 p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta}\right)^{1/2}. \tag{3.66}$$

Since $\int_{\mathbb{R}^\nu} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,d\boldsymbol{\eta} = 1$ and using Eq. (3.65), we obtain

$$\int_{\mathbb{R}^\nu} |\,\langle\,\nabla_{\lambda} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\rangle\,|\,d\boldsymbol{\eta} \;\leq\; (\mathrm{tr}\,[\Gamma''(\boldsymbol{\lambda})])^{1/2}\;\|\,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\|\,. \tag{3.67}$$

From Eq. (3.61), it can be deduced that $\int_{\mathbb{R}^\nu} |\,p_{\boldsymbol{H}_{\lambda^{i+1}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}^{i+1}) - p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,|\,d\boldsymbol{\eta} \;\leq\; \int_{\mathbb{R}^\nu} |\,\langle\,\nabla_{\lambda} p_{\boldsymbol{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})\,,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}\,\rangle\,|\,d\boldsymbol{\eta} + o(\|\,\boldsymbol{\lambda}^{i+1}-\boldsymbol{\lambda}^{i}\,\|)$ that yields Eq. (3.60) by using Eq. (3.67).

4. A few numerical elements for implementation of the methodology

4.1. Choice of the integration scheme for solving the ISDE introduced in Propositions 3 and 4

As we have previously explained, for $\boldsymbol{\lambda} \in C_{\mathrm{ad},\boldsymbol{\lambda}}$, the ISDE defined by Eqs. (3.44) to (3.46), must be solved for $t \in [0\,,t_s]$ (see Proposition 3-(e)) with the initial condition at $t = 0$ defined in Proposition 3-(a), in order to generate the constrained learned set $\mathcal{D}_{\boldsymbol{H}_\lambda} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}},\ldots,\boldsymbol{\eta}^N_{\boldsymbol{\lambda}}\}$ with $N \gg N_d$. It is assumed that N is written as $N = N_d \times n_{\mathrm{MC}}$ with $n_{\mathrm{MC}} \ll 1$. Therefore, a discretization scheme [140, 141] must be used. The case of Hamiltonian dynamical systems has been analyzed in [142] by using an implicit Euler scheme. The Störmer-Verlet scheme (see [143] for the deterministic case and [144] for the stochastic case) is a very efficient scheme that allows for having a long-time energy conservation for non-dissipative Hamiltonian dynamical systems. In [145], we have proposed to use an extension of the Störmer-Verlet scheme for stochastic dissipative Hamiltonian systems, that we have also used in [146, 47, 103, 95].

4.2. Störmer-Verlet scheme and computation of the constrained learned set $\mathcal{D}_{\boldsymbol{H}_{\lambda^i}}$

Let i be the index of the sequence $\{\boldsymbol{\lambda}^i, i = 0,1,\ldots,i_{\max}\}$ of Lagrange multipliers computed using Eq. (3.56) with $\boldsymbol{\lambda}^0 = \boldsymbol{0}_{n_c}$. Let $t_m = m\,\Delta t$ for $m = 0,1,\ldots,M_s$ (with $M_s > 1$ an integer) be the time sampling in which $t_s = M_s\,\Delta t$ (and thus $t_{M_s} = t_s$). Let $\Delta\boldsymbol{W}^{\mathrm{wien}}_{m+1} = \boldsymbol{W}^{\mathrm{wien}}(t_{m+1}) - \boldsymbol{W}^{\mathrm{wien}}(t_m)$ be the Gaussian, second-order, centered, $\mathbb{R}^\nu$-valued random variable such that $E\{\Delta\boldsymbol{W}^{\mathrm{wien}}_{m+1} \otimes \Delta\boldsymbol{W}^{\mathrm{wien}}_{m+1}\} = \Delta t\,[I_\nu]$. Let $\{\theta_\ell, \ell = 1,\ldots,N\}$ be N independent realizations in Θ. For $m = 0,1,\ldots,M_s-1$, let $\Delta\mathbb{W}^\ell_{m+1} = \Delta\boldsymbol{W}^{\mathrm{wien}}_{m+1}(\theta_\ell)$ be the realization θ_ℓ of $\Delta\boldsymbol{W}^{\mathrm{wien}}_{m+1}$. Following the choice of $(\boldsymbol{u}_0, \boldsymbol{v}_0)$ defined in Proposition 3-(a), let $\boldsymbol{u}^1_0,\ldots,\boldsymbol{u}^N_0$ in $\mathbb{R}^\nu$ be such that for $k = 1,\ldots,n_{\mathrm{MC}}$ and for $j = 1,\ldots,N_d$, we take $\boldsymbol{u}^\ell_0 = \boldsymbol{\eta}^j_d$ with $\ell = j + (k-1)\times N_d$. Let $\boldsymbol{v}^1_0,\ldots,\boldsymbol{v}^N_0$ in $\mathbb{R}^\nu$ be N independent realizations of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{V}_G$ also defined in Proposition 3-(a). Note that the realizations $\Delta\mathbb{W}^\ell_{m+1}$, $\boldsymbol{u}^\ell_0$, and $\boldsymbol{v}^\ell_0$, for $\ell = 1,\ldots,N$ are independent of $\{\boldsymbol{\lambda}^i\}_i$. For $i \in \{0,1,\ldots,i_{\max}\}$ and for $\ell \in \{1,\ldots,N\}$, we introduce the realizations $\mathbb{U}^{i,\ell}_m = \boldsymbol{U}^N_{\boldsymbol{\lambda}^i}(t_m\,;\theta_\ell)$ and $\mathbb{V}^{i,\ell}_m = \boldsymbol{V}^N_{\boldsymbol{\lambda}^i}(t_m\,;\theta_\ell)$. For $m \in \{0,1,\ldots,M_s-1\}$, the Störmer-Verlet scheme applied to realization θ_ℓ of Eqs. (3.44) to (3.46) yields the following recurrence,

$$\mathbb{U}^{i,\ell}_{m+1/2} = \mathbb{U}^{i,\ell}_m + \frac{\Delta t}{2}\,\mathbb{V}^{i,\ell}_m\,, \tag{4.1}$$

$$\mathbb{V}^{i,\ell}_{m+1} = \frac{1-\gamma}{1+\gamma}\,\mathbb{V}^{i,\ell}_m + \frac{\Delta t}{1+\gamma}\,\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}^{i-1}}(\mathbb{U}^{i,\ell}_{m+1/2}) + \frac{\sqrt{f_0}}{1+\gamma}\,\Delta\mathbb{W}^\ell_{m+1}\,, \tag{4.2}$$

$$\mathbb{U}^{i,\ell}_{m+1} = \mathbb{U}^{i,\ell}_{m+1/2} + \frac{\Delta t}{2}\,\mathbb{V}^{i,\ell}_{m+1}\,, \tag{4.3}$$

with the initial condition

$$\mathbb{U}^{i,\ell}_0 = \boldsymbol{u}^\ell_0 \quad,\quad \mathbb{V}^{i,\ell}_0 = \boldsymbol{v}^\ell_0\,, \tag{4.4}$$

in which $\gamma = f_0\,\Delta t/4$ and where, using Eq. (3.43),

$$\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}^{i-1}}(\boldsymbol{u}) = \frac{1}{\zeta(\boldsymbol{u})}\,\nabla_{\boldsymbol{u}}\zeta(\boldsymbol{u}) - [\nabla_{\boldsymbol{u}}\hat{\boldsymbol{h}}^N(\boldsymbol{u}\,;\boldsymbol{\lambda}^{i-1})]\,\boldsymbol{\lambda}^{i-1} \tag{4.5}$$

The matrix $[\nabla_{\boldsymbol{u}}\hat{\boldsymbol{h}}^N(\boldsymbol{u}\,;\boldsymbol{\lambda}^{i-1})]$ is given by Eq. (4.8), which depends on

$$\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i-1}}} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}^{i-1}}, \ldots, \boldsymbol{\eta}^N_{\boldsymbol{\lambda}^{i-1}}\}\,. \tag{4.6}$$

It should be noted that, in Eq. (4.2), $\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}^{i-1}}$ has been used instead of $\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}^{i}}$ because $\hat{\boldsymbol{L}}^N_{\boldsymbol{\lambda}^{i}}$ depends on $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i}}}$ that is unknown, recalling that the aim of the recurrence defined by Eqs. (4.1) to (4.4) is precisely to calculate $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i}}}$ that is written as

$$\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i}}} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}^{i}}, \ldots, \boldsymbol{\eta}^N_{\boldsymbol{\lambda}^{i}}\} \quad , \quad \boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}^{i}} = \boldsymbol{U}^N_{\boldsymbol{\lambda}^{i}}(t_s\,;\theta_\ell) = \mathbb{U}^{i,\ell}_{M_s}\,. \tag{4.7}$$

4.3. Explicit expression of the gradient of the statistical surrogate model of $\boldsymbol{h}^c$

Using Definition 3, for fixed value of $\boldsymbol{\lambda} \in C_{\text{ad},\boldsymbol{\lambda}}$, the gradient $[\nabla_{\boldsymbol{\eta}}\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})] \in \mathbb{M}_{\nu,n_c}$ at point $\boldsymbol{\eta} \in \mathbb{R}^\nu$ of the statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$ can be written as

$$[\nabla_{\boldsymbol{\eta}}\hat{\boldsymbol{h}}^N(\boldsymbol{\eta}\,;\boldsymbol{\lambda})] = \sum_{\ell=1}^{N} \boldsymbol{\gamma}^\ell_{\boldsymbol{\lambda}} \otimes \boldsymbol{a}^\ell_{\boldsymbol{\lambda}} \quad , \quad \boldsymbol{\gamma}^\ell_{\boldsymbol{\lambda}} = \nabla_{\boldsymbol{\eta}}\left(\frac{\beta^N_{\boldsymbol{\eta}}(\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}})}{\sum_{\ell'=1}^N \beta^N_{\boldsymbol{\eta}}(\boldsymbol{\eta}^{\ell'}_{\boldsymbol{\lambda}})}\right), \tag{4.8}$$

in which $\boldsymbol{a}^\ell_{\boldsymbol{\lambda}} = \boldsymbol{h}^c(\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}}) \in \mathbb{R}^{n_c}$ and where $\boldsymbol{\gamma}^\ell_{\boldsymbol{\lambda}} \in \mathbb{R}^\nu$ is explicitly calculated using Eqs. (3.28) and (3.29).

4.4. Summary of the complete algorithm

The algorithm for calculating $\boldsymbol{\lambda}^{\text{sol}}$ and $\mathcal{D}_{\boldsymbol{H}^c} = \{\boldsymbol{\eta}^1_c, \ldots, \boldsymbol{\eta}^N_c\}$ with $\boldsymbol{\eta}^\ell_c = \boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}^{\text{sol}}}$ for $\ell = 1, \ldots, N$ is summarized in Algorithm 1.

Algorithm 1 Algorithm for calculating $\boldsymbol{\lambda}^{\text{sol}}$ and $\mathcal{D}_{\boldsymbol{H}^c} = \{\boldsymbol{\eta}^1_c, \ldots, \boldsymbol{\eta}^N_c\}$.

Data: N_d, $\mathcal{D}_d = \{\boldsymbol{\eta}^1_d, \ldots, \boldsymbol{\eta}^{N_d}_d\}$, N, $i_{\max}$, M_s, t_s, Δt, f_0, $\gamma = f_0\,\Delta t/4$
Init: $\Delta\mathbb{W}^\ell_{m+1}, \ell \in \{1, \ldots, N\}, m \in \{1, \ldots, M_s - 1\}$, $\boldsymbol{u}^\ell_0$ and $\boldsymbol{v}^\ell_0$ for $\ell \in \{1, \ldots, N\}$, $\boldsymbol{\lambda}^0 = \boldsymbol{0}_{n_c}$
for $i = 1 : i_{\max}$ **do**
  for $\ell = 1 : N$ (*loop in parallel computation*) **do**
    $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}^i}, \ldots, \boldsymbol{\eta}^N_{\boldsymbol{\lambda}^i}\}$ from Eq. (4.7), using Eqs. (4.1) to (4.4) and $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i-1}}}$ ($\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^0}}$ not used for $i = 1$)
  end for
  for $\ell = 1 : N$ (*loop in parallel computation*) **do**
    $\boldsymbol{h}^c(\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}^i}), \ell = 1, \ldots, N$ using the BVP
  end for
  $\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}^i)$ and $[\Gamma''(\boldsymbol{\lambda}^i)]$ using Eqs. (3.19) to (3.20) and $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}}$
  $\text{err}(i)$ using Eqs. (3.58) with (3.59)
  $\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - \alpha_{\text{relax}}\,[\Gamma''(\boldsymbol{\lambda}^i)]^{-1}\,\nabla_{\boldsymbol{\lambda}}\Gamma(\boldsymbol{\lambda}^i)$ using Eq. (3.56) with a relaxation factor $\alpha_{\text{relax}} \in]0\,,1]$
  $\boldsymbol{\lambda}^i \leftarrow \boldsymbol{\lambda}^{i+1}$
  $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{i-1}}} \leftarrow \mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}}$
end for
$\boldsymbol{\lambda}^{\text{sol}} = \boldsymbol{\lambda}^{i_{\text{sol}}}$, $i_{\text{sol}} = \arg\min_i \text{err}(i)$ from Eq. (3.57)
$\mathcal{D}_{\boldsymbol{H}^c} \leftarrow \mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^{\text{sol}}}}$

5. Application to stochastic homogenization without scale separation

In this section, we consider the stochastic boundary value problem associated with the stochastic homogenization of a random elastic medium without scale separation, which has been presented in Section 1. The physical space $\mathbb{R}^3$ is referred to a Cartesian reference system whose the generic point is $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$. We consider the stochastic homogenization of a heterogeneous linear elastic microstructure occupying the 3D bounded open domain $\Omega =]\,0\,,1\,[\times]\,0\,,1\,[\times]\,0\,,0.1\,[\subset \mathbb{R}^3$ (square thick plate) with boundary $\partial\Omega$. The homogenization method on Ω in the one proposed in [147] that we have already used in [120, 122]. In this section, we use the convention for summation over repeated Latin indices taking values in $\{1, 2, 3\}$.

5.1. Stochastic elliptic boundary value problem

For all m and r in $\{1, 2, 3\}$ the unknown field is the $\mathbb{R}^3$-valued random field $\{\boldsymbol{Y}(\boldsymbol{\xi}) = (Y_1(\boldsymbol{\xi}), Y_2(\boldsymbol{\xi}), Y_3(\boldsymbol{\xi})), \boldsymbol{\xi} \in \Omega\}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by Ω, such that for $i = 1, 2, 3$, and almost surely,

$$-\frac{\partial}{\partial \xi_j}\Big(\mathbb{C}_{ijpq}(\boldsymbol{\xi})\,\varepsilon_{pq}(\boldsymbol{Y}^{mr}(\boldsymbol{\xi}))\Big) = \mathbf{0}_3 \quad , \quad \forall \boldsymbol{\xi} \in \Omega\,, \tag{5.1}$$

$$\boldsymbol{Y}^{mr}(\boldsymbol{\xi}) = \boldsymbol{y}_0^{mr} \quad , \quad \forall \boldsymbol{\xi} \in \partial\Omega\,, \tag{5.2}$$

in which the strain tensor is $\varepsilon_{pq}(\boldsymbol{y}) = (\partial y_p/\partial \xi_q + \partial y_q/\partial \xi_p)/2$ for all $\boldsymbol{y} = (y_1, y_2, y_3)$. For all $\boldsymbol{\xi} \in \partial\Omega$, $\boldsymbol{y}_0^{mr} = (y_{0,1}^{mr}, y_{0,2}^{mr}, y_{0,3}^{mr})$ is defined by

$$y_{0,j}^{mr} = (\delta_{jm}\,\xi_r + \delta_{jr}\,\xi_m)/2 \quad , \quad j \in \{1, 2, 3\}\,, \tag{5.3}$$

in which δ_{jm} is the Kronecker symbol. At mesoscale, the linear elastic heterogeneous medium is described by the random apparent elasticity field $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$, which is a non-Gaussian fourth-order tensor-valued random field $\mathbb{C} = \{\mathbb{C}_{ijpq}\}_{ijpq}$ with i, j, p, and q in $\{1, 2, 3\}$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$. The stochastic homogenization consists, for i, j, m, and r in $\{1, 2, 3\}$, in analyzing at macroscale the component $\mathbb{C}_{ijmr}^{\text{eff}}$ of the random effective/apparent elasticity tensor $\{\mathbb{C}_{ijmr}^{\text{eff}}\}_{ijmr}$, which is defined by

$$\mathbb{C}_{ijmr}^{\text{eff}} = \frac{1}{|\Omega|}\int_\Omega \mathbb{C}_{ijpq}(\boldsymbol{\xi})\,\varepsilon_{pq}(\boldsymbol{Y}^{mr}(\boldsymbol{\xi}))\,d\boldsymbol{\xi}\,, \tag{5.4}$$

in which $\boldsymbol{Y}^{mr}$ is the $\mathbb{R}^3$-valued random field that satisfies Eqs. (5.1) to (5.3) and where $|\Omega| = \int_\Omega d\boldsymbol{\xi}$. The random effective/apparent elasticity tensor $\mathbb{C}^{\text{eff}}$ is symmetric and positive definite almost surely. If there was a scale separation, then the statistical fluctuations of this tensor would be negligible.

5.2. Prior probability model of random field $\mathbb{C}$

The prior probability model of $\mathbb{C}$ used for generating the training set is the one presented in [122, 148]. This is a second-order, non-Gaussian, positive-definite fourth-order tensor-valued homogeneous random field, indexed by $\mathbb{R}^3$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with a spectral random measure. This random field is parameterized as

$$\mathbb{C}(\boldsymbol{\xi}) = \mathbb{c}(\boldsymbol{G}(\boldsymbol{\xi}), \underline{z}) \quad , \quad \boldsymbol{\xi} \in \Omega\,, \tag{5.5}$$

in which $\{\boldsymbol{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ is a non-Gaussian second-order, homogeneous, $\mathbb{R}^{21}$-valued random field indexed by $\mathbb{R}^3$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with random spectral measure, where $\underline{z} = (\underline{z}_1, \underline{z}_2, \underline{z}_3)$ is the nominal value of a $\mathbb{R}^3$-valued control parameter, and where $\mathbb{c}$ is a given mapping from $\mathbb{R}^{21} \times \mathbb{R}^3$ into the fourth-order tensor on $\mathbb{R}^3$.

(i) *Isotropic nominal model at mesoscale*. At mesoscale, the nominal model is a linear, elastic, homogeneous, isotropic medium whose elasticity tensor $\underline{\mathbb{C}}$ depends only on the Young modulus $\underline{E} = 1.7 \times 10^{11}\,N/m^2$ and Poisson coefficient $\underline{\nu}_P = 0.24$. The corresponding bulk modulus $\underline{C}_{\text{bulk}} = \underline{E}\,/(3(1 - 2\underline{\nu}))$ and shear modulus $\underline{C}_{\text{shear}} = \underline{E}\,/(2(1 + \underline{\nu}))$ are $1.08974 \times 10^{11}\,N/m^2$ and $6.85484 \times 10^{10}\,N/m^2$.

(ii) *Anisotropic statistical fluctuations at mesoscale*. The statistical fluctuations of the random medium are assumed to be anisotropic, which means that, for all $\boldsymbol{\xi}$ fixed in $\mathbb{R}^3$, the random apparent elasticity tensor $\mathbb{C}(\boldsymbol{\xi})$ is a full anisotropic

tensor. The hyperparameters that control the anisotropic statistical fluctuations of the random apparent elasticity field (see [122]) are:
(1) the dispersion coefficient $\underline{\delta}_{\mathbb{C}} = 0.3$ that controls the level of statistical fluctuations of the random medium.
(2) the spatial correlation lengths $\underline{L}_{c1}$, $\underline{L}_{c2}$, and $\underline{L}_{c3}$ (for directions ξ_1, ξ_2, and ξ_3) of the random field $\{\boldsymbol{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ and the dispersion coefficient $\underline{\delta}_s = 0.1$ that controls the level of uncertainties of its spectral measure (these spectral-measure uncertainties will not be controlled, which means that $\underline{\delta}_s$ will not be a control parameter and its value is fixed). As explained at the end of Section 1, three cases, SC1, SC2, and SC3, of the correlation lengths are considered for analyzing the level of scale separation and are defined in Table 1. Taking into account the definition of domain Ω and the values of the spatial correlation lengths, there will not have a scale separation and consequently, the effective/apparent elasticity tensor will exhibit statistical fluctuations.

Table 1: Values of the spatial correlation lengths $\underline{L}_{c1}$, $\underline{L}_{c2}$, and $\underline{L}_{c3}$ for cases SC1, SC2, and SC3 of scale separation.

	$\underline{L}_{c1}$	$\underline{L}_{c2}$	$\underline{L}_{c3}$
SC1	0.1	0.1	0.1
SC2	0.3	0.3	0.1
SC3	0.5	0.5	0.2

(iii) *Nominal value $\underline{z}$ of the control parameter*. It is defined by $\underline{z}_1 = \underline{C}_{\text{bulk}}$, $\underline{z}_2 = \underline{C}_{\text{shear}}$, and $\underline{z}_3 = \underline{\delta}_{\mathbb{C}}$.

(iv) *Random control parameter for the probabilistic learning inference*. For estimating the posterior model using the probabilistic learning inference methodology, in addition to the prior probability model of random field $\boldsymbol{G}$, we introduce a $\mathbb{R}^{n_w}$-valued random control parameter $\boldsymbol{W} = (W_1, \ldots, W_{n_w})$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ and independent of $\boldsymbol{G}$ such that, $n_w = 3$ and

$$W_1 = \log C_{\text{bulk}} \quad , \quad W_2 = \log C_{\text{shear}} \quad , \quad W_3 = \log \delta_{\mathbb{C}} \, , \tag{5.6}$$

in which (a) C_{bulk} and C_{shear} are Gamma independent random variables (see [149]) whose mean values are the nominal values $\underline{C}_{\text{bulk}}$ and $\underline{C}_{\text{shear}}$ previously defined and for which the coefficient of variation of C_{bulk} is chosen as $\delta_{\text{bulk}} = 0.5$ yielding $\delta_{\text{shear}} = 0.25$ (note that, with the model proposed in [149], δ_{shear} is deduced from δ_{bulk} and cannot arbitrarily be chosen); (b) $\delta_{\mathbb{C}}$ is chosen as a uniform random variable on $[0.1\,, 0.5]$ whose mean values is $\underline{\delta}_{\mathbb{C}}$.

It should be noted that the statistical fluctuations of C_{bulk}, C_{shear}, and $\delta_{\mathbb{C}}$ are chosen sufficiently large in order that the range of fluctuations of the random effective/apparent elasticity tensor covers the experimental target (see Section 5.5) in order to be able to improve the prior probabilistic model with the posterior probabilistic model by solving the inverse statistical problem with the proposed probabilistic learning inference approach.

5.3. Solution of the stochastic BVP

Under the hypotheses introduced for constructing random field $\mathbb{C}$, Proposition 5.1 of [148] proves that for $1 \leq m \leq r \leq 3$, the strong stochastic solution $\{\boldsymbol{Y}^{mr}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ of the weak formulation of the stochastic elliptic BVP defined by Eqs. (5.1) to (5.3) exists, is unique, and is a second-order random field,

$$E\{\, \| \boldsymbol{Y}^{mr}(\boldsymbol{\xi}) \|^2\} < +\infty \quad , \quad \forall \boldsymbol{\xi} \in \Omega \, . \tag{5.7}$$

Due to Corollary 5.1 of [148] and its proof, the random effective/apparent elasticity tensor $\mathbb{C}^{\text{eff}}$ is a second-order random variable,

$$\sum_{1 \leq i \leq j \leq 3} \sum_{1 \leq m \leq r \leq 3} E\{(\mathbb{C}^{\text{eff}}_{ijmr})^2\} < +\infty \, . \tag{5.8}$$

5.4. Stochastic computational model and random effective/apparent elasticity matrix

The finite element method is used for discretizing the weak formulation of the stochastic BVP. The finite element mesh of domain $\overline{\Omega}$ is made up of $60 \times 60 \times 6 = 21\,600$ eight-nodes solid elements, 26 047 nodes, and 78 141 dofs (25 926 Dirichlet conditions on $\partial\Omega$ and $n_y = 52\,215$ the remaining dofs). There are 2^3 integration points in each finite element, which yields $n_p = 172\,800$ integration points for the spatial discretization of the random fourth-order tensor-valued elasticity field $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$. The discretization of random field $\mathbb{C}$ is expressed as a function of a $\mathbb{R}^{n_g}$-valued random variable $\boldsymbol{\mathcal{G}}$ corresponding to the spatial discretization of random field $\boldsymbol{G}$ with $n_g = 21 \times n_p = 3\,628\,800$. For $1 \leq m \leq r \leq 3$, let $\boldsymbol{\mathcal{Y}}^{mr}$ be the $\mathbb{R}^{n_y}$-valued random variable of the free dofs (that corresponds to the dofs of the nodes inside Ω of the finite element discretization of random field $\{\boldsymbol{Y}^{mr}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \overline{\Omega}\}$). Therefore, the stochastic computational model can be written as the $\mathbb{R}^{n_y}$-valued stochastic equation,

$$\boldsymbol{\mathcal{N}}^{mr}(\boldsymbol{\mathcal{Y}}^{mr}, \boldsymbol{\mathcal{G}}, \boldsymbol{W}) = \mathbf{0}_{n_y} \ a.s.\,, \tag{5.9}$$

which is a stochastic linear equation that can be rewritten as $[a^{mr}(\boldsymbol{\mathcal{G}}, \boldsymbol{W})]\, \boldsymbol{\mathcal{Y}}^{mr} - \boldsymbol{b}^{mr}(\boldsymbol{\mathcal{G}}, \boldsymbol{W}) = \mathbf{0}_{n_y}$ in which, for $\boldsymbol{g}_d \in \mathbb{R}^{n_g}$ and $\boldsymbol{w}_d \in \mathbb{R}^{n_w}$, $[a^{mr}(\boldsymbol{g}_d, \boldsymbol{w}_d)]$ is a matrix in $\mathbb{M}^+_{n_y}$ (thus, invertible) and where $\boldsymbol{b}^{mr}(\boldsymbol{g}_d, \boldsymbol{w}_d)$ is a vector in $\mathbb{R}^{n_y}$, which depends on the Dirichlet condition defined by Eq. (5.2). For $1 \leq m \leq r \leq 3$, the finite element discretization of the right-hand side member of Eq. (5.4) yields the $\mathbb{M}^+_6$-valued random effective/apparent elasticity matrix $[\mathbb{C}^{\text{eff}}]$ such that $[\mathbb{C}^{\text{eff}}]_{\mathbf{ij}} = \mathbb{C}^{\text{eff}}_{ijmr}$ in which the indices $\mathbf{i} = (i, j)$ with $1 \leq i \leq j \leq 3$ and $\mathbf{j} = (m, r)$ with $1 \leq m \leq r \leq 3$ are with values in $\{1, \ldots, 6\}$. This random matrix can be written as

$$[\mathbb{C}^{\text{eff}}] = [\boldsymbol{\mathcal{O}}(\{\boldsymbol{\mathcal{Y}}^{mr}, 1 \leq m \leq r \leq 3\}, \boldsymbol{\mathcal{G}}, \boldsymbol{W})]\,, \tag{5.10}$$

in which $(\{\boldsymbol{y}^{mr}, 1 \leq m \leq r \leq 3\}, \boldsymbol{g}, \boldsymbol{w}) \mapsto [\boldsymbol{\mathcal{O}}(\{\boldsymbol{y}^{mr}, 1 \leq m \leq r \leq 3\}, \boldsymbol{g}, \boldsymbol{w})]$ is a measurable mapping from $\mathbb{R}^{6\times n_y} \times \mathbb{R}^{n_g} \times \mathbb{R}^{n_w}$ into $\mathbb{M}^+_6$. Note that, for $1 \leq m \leq r \leq 3$, $\boldsymbol{\mathcal{Y}}^{mr}$ satisfying Eq. (5.9) corresponds to the strong stochastic solution of the finite element discretization of the weak formulation of the stochastic BVP (see Section 5.3). Due to Eqs. (5.7) and (5.8), we have,

$$E\{\, \|\boldsymbol{\mathcal{Y}}^{mr}\|^2 \} < +\infty \quad , \quad E\{\, \|\, [\mathbb{C}^{\text{eff}}]\, \|_F^2 \} < +\infty\,, \tag{5.11}$$

which proves that $\boldsymbol{\mathcal{Y}}^{mr}$ and $[\mathbb{C}^{\text{eff}}]$ are second-order random variables.

5.5. Definition of the statistical moments and their targets

From Eq. (5.11), the $\mathbb{M}^+_6$-valued random variable $[\mathbb{C}^{\text{eff}}]$ is of second-order. We can then define its first two moments. (i) The first statistical moment of interest is the mean value $[\,\underline{\mathbb{C}}^{\text{eff}}] = E\{\,[\mathbb{C}^{\text{eff}}]\,\} \in \mathbb{M}^+_6$ of random matrix $[\mathbb{C}^{\text{eff}}]$ while its target counterpart is the given matrix $[\,\underline{\mathbb{C}}^{\text{exp}}] \in \mathbb{M}^+_6$. Let $\mu_{\text{exp}} = \|\,[\,\underline{\mathbb{C}}^{\text{exp}}]\,\|_F$ be the Frobenius norm of $[\,\underline{\mathbb{C}}^{\text{exp}}]$. Introducing the subscript "n" to designate a normalization, we define the normalized quantities with respect to μ_{exp} as,

$$[\mathbb{C}^{\text{eff}}_n] = \frac{1}{\mu_{\text{exp}}}\,[\mathbb{C}^{\text{eff}}] \quad , \quad [\,\underline{\mathbb{C}}^{\text{eff}}_n] = \frac{1}{\mu_{\text{exp}}}\,[\,\underline{\mathbb{C}}^{\text{eff}}] \quad , \quad [\,\underline{\mathbb{C}}^{\text{exp}}_n] = \frac{1}{\mu_{\text{exp}}}\,[\,\underline{\mathbb{C}}^{\text{exp}}]\,. \tag{5.12}$$

The corresponding constraint equation will then be written as

$$E\{\,[\mathbb{C}^{\text{eff}}_n]\,\} = [\,\underline{\mathbb{C}}^{\text{exp}}_n]\,. \tag{5.13}$$

(ii) The second statistical moment of interest is the coefficient of dispersion δ^{eff} of random matrix $[\mathbb{C}^{\text{eff}}]$ and its target counterpart δ^{exp}. Let Δ^{eff}_2 be the positive-valued random variable defined by

$$\Delta^{\text{eff}}_2 = \frac{1}{\|\,[\,\underline{\mathbb{C}}^{\text{eff}}]\,\|_F^2}\, \|\,[\mathbb{C}^{\text{eff}}] - [\,\underline{\mathbb{C}}^{\text{eff}}]\,\|_F^2\,. \tag{5.14}$$

Consequently, δ^{eff} that is defined by $\delta^{\text{eff}} = \Big(E\{\,\|\,[\mathbb{C}^{\text{eff}}] - [\,\underline{\mathbb{C}}^{\text{eff}}]\,\|_F^2 / \|\,[\,\underline{\mathbb{C}}^{\text{eff}}]\,\|_F^2\,\}\Big)^{1/2}$ can be rewritten as

$$\delta^{\text{eff}} = \sqrt{E\{\Delta^{\text{eff}}_2\}}\,. \tag{5.15}$$

Defining $\mu_{\text{eff}} = \| [\underline{\mathbb{C}}^{\text{eff}}] \|_F$, the constraint equation that is defined by

$$\delta^{\text{eff}} = \delta^{\text{exp}} , \tag{5.16}$$

can be rewritten, using Eqs. (5.12) to (5.15), as

$$E\{ \| [\mathbb{C}_n^{\text{eff}}] - [\underline{\mathbb{C}}_n^{\text{eff}}] \|_F^2 \} = \left(\frac{\mu_{\text{eff}}}{\mu_{\text{exp}}} \, \delta^{\text{exp}} \right)^2 . \tag{5.17}$$

while Eq. (5.14) yields

$$\Delta_2^{\text{eff}} = \left(\frac{\mu_{\text{exp}}}{\mu_{\text{eff}}} \right)^2 \| [\mathbb{C}_n^{\text{eff}}] - [\underline{\mathbb{C}}_n^{\text{eff}}] \|_F^2 \} . \tag{5.18}$$

It should be noted that, if δ^{eff} goes to zero, then the statistical fluctuations of $[\mathbb{C}^{\text{eff}}]$ goes to zero because, due to the Tchebychev inequality, $[\mathbb{C}^{\text{eff}}]$ goes in probability to its mean value $[\underline{\mathbb{C}}^{\text{eff}}]$ (this would be the case of a scale separation).

(iii) As we have explained and as we will see in Sections 5.7 and 5.8, the constrained learned set is constructed by also taking as a constraint the second-order moment of the normalized random residue of the stochastic PDE. On the other hand, as shown by equations Eqs. (5.19) and (5.29), the spatial discretization $\boldsymbol{\mathcal{G}}$ of the random field $\{\boldsymbol{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ is obviously a fundamental quantity that intervenes in the calculation of this normalized random residue during the constrained learning process. Furthermore, it is essential that the mathematical properties of $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ (random field with values in the positive-definite symmetric fourth-order tensors, lower bound, statistical homogeneity, control of the random spectral measure), defined by Eq. (5.5), remain verified. This implies strong mathematical properties on the random field $\{\boldsymbol{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$. This class of random fields with a random spectral measure, which was constructed and analyzed in [148] and that we are using in this paper, already corresponds to a large class, given all the algebraic and probabilistic properties that this class must verify. The targets have been simulated by using the stochastic computational model while keeping the same class of random fields for $\{\mathbb{C}(\boldsymbol{\xi}) \in \Omega\}$ but for which the values of its hyperparameters are different of those used for the prior probabilistic model. It should be noted that changing the class by keeping this type of algebraic representation would be very difficult. A way for performing an extension of the class would then be to use a polynomial Chaos representations as proposed in [150, 139] to properly take into account all the required properties. As the objective of this paper is not to build an even larger class for generating the targets, we have preferred to keep the same class. Finally, note that this choice does not facilitate the analysis, because the methodology and its proven mathematical properties are independent of the choice of the random fields class provide that the class be correctly constructed for the stochastic elliptic operator. Nevertheless, the proximity of the constructed posterior probability measure to the prior probability measure is certainly dependent on the chosen class.

(iv) For the three considered cases of scale separation, the numerical values of μ_{eff} for the training set (and thus denoted by $\mu_{\text{eff},d}$) and μ_{exp} for the target are given in Table 2.

Table 2: For cases SC1, SC2, and SC3, values of $\mu_{\text{eff},d}$ and μ_{exp}.

	$\underline{L}_{c1}$	$\underline{L}_{c2}$	$\underline{L}_{c3}$	$\mu_{\text{eff},d} \times 10^{11}$	$\mu_{\text{exp}} \times 10^{11}$
SC1	0.1	0.1	0.1	4.2106	4.6317
SC2	0.3	0.3	0.1	4.1925	4.6549
SC3	0.5	0.5	0.2	4.1943	4.6706

5.6. Training set computed with the prior probability model and its normalization

The stochastic computational model defined in Section 5.4 is used for generating the training set related to the random variable $\boldsymbol{X} = (\{\boldsymbol{\mathcal{Y}}^{mr}, 1 \leq m \leq r \leq 3\}, \boldsymbol{\mathcal{G}}, \boldsymbol{W})$ with values in $\mathbb{R}^{n_x} = \mathbb{R}^{6 \times n_y} \times \mathbb{R}^{n_g} \times \mathbb{R}^{n_w}$ with $n_x = 6\, n_y + n_g + n_w =$

3 942 093. The Monte Carlo numerical simulation method is used with $N_d = 50$ independent realizations and the prior probability model of $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{W}$. We then obtain the training set $\{\boldsymbol{x}_d^j, j = 1, \ldots, N_d\}$ with $\boldsymbol{x}_d^j = (\{\boldsymbol{y}_d^{mr,j}, 1 \leq m \leq r \leq 3\}, \boldsymbol{g}_d^j, \boldsymbol{w}_d^j)$ in which $\boldsymbol{y}_d^{mr,j} \in \mathbb{R}^{n_y}$ is the solution (see Eq. (5.9)) of

$$\boldsymbol{\mathcal{N}}^{mr}(\boldsymbol{y}_d^{mr,j}, \boldsymbol{g}_d^j, \boldsymbol{w}_d^j) = \boldsymbol{0}_{n_y}\,, \tag{5.19}$$

that is to say of the linear equation $[a^{mr}(\boldsymbol{g}_d^j, \boldsymbol{w}_d^j)]\, \boldsymbol{y}_d^{mr,j} = \boldsymbol{b}^{mr}(\boldsymbol{g}_d^j, \boldsymbol{w}_d^j)$ (see Section 5.4). The statistical moments defined in Section 5.5 can then be computed using the N_d independent realizations $\{[\mathbb{C}_d^{\mathrm{eff},j}], j = 1, \ldots, N_d\}$ of the $\mathbb{M}_6^+$-valued random matrix $[\mathbb{C}_d^{\mathrm{eff}}]$ defined by Eq. (5.10) (a subscript "d" is introduced to designate the computation done with the training set, based on the prior probability model). Since $N_d \ll n_x$, the normalization of $\boldsymbol{X}$ is performed using the principal component analysis as follows. Let $\boldsymbol{x}_{\mathrm{ctr}}^j = \boldsymbol{x}_d^j - \underline{\boldsymbol{x}}$ with $\underline{\boldsymbol{x}} = (1/N_d) \sum_{j=1}^{N_d} \boldsymbol{x}_d^j \in \mathbb{R}^{n_x}$. Let $[x_{\mathrm{ctr}}] = [\boldsymbol{x}_{\mathrm{ctr}}^1 \ldots \boldsymbol{x}_{\mathrm{ctr}}^{N_d}]$ be the matrix in $\mathbb{M}_{n_x, N_d}$ and let $[\Phi]\,[S]\,[\Psi]^T = [x_{\mathrm{ctr}}]$ be the thin SVD [151] (economy size SVD) of matrix $[x_{\mathrm{ctr}}]$. The diagonal entries of diagonal matrix $[S]$ are the singular values $S_1 \geq \ldots \geq S_{N_d-1} > S_{N_d} = 0$ that are in decreasing order and we have $S_{N_d} = 0$. The matrix $[\Phi]$ is in $\mathbb{M}_{n_x,\nu}$ with $\nu = N_d - 1$ and $[\Phi]^T\,[\Phi] = [I_\nu]$. Then random vector $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \underline{\boldsymbol{x}} + [\Phi]\,[\kappa]^{1/2}\,\boldsymbol{H}\,, \tag{5.20}$$

in which $[\kappa]$ is the diagonal matrix such that $\kappa_\alpha = [\kappa]_{\alpha\alpha} = S_\alpha^2/(N_d - 1)$, and where $\boldsymbol{H}$ is the $\mathbb{R}^\nu$-valued random variable whose N_d independent realizations are computed by

$$\boldsymbol{\eta}_d^j = [\kappa]^{-1/2}\,[\Phi]^T(\boldsymbol{x}_d^j - \underline{\boldsymbol{x}}) \quad , \quad j = 1, \ldots, N_d\,. \tag{5.21}$$

Note that $\{\kappa_\alpha\}_\alpha$ are the eigenvalues of the covariance matrix of $\boldsymbol{X}$ estimated with $\{\boldsymbol{x}_d^1, \ldots, \boldsymbol{x}_d^{N_d}\}$. Random vector $\boldsymbol{H}$ is then normalized. The empirical estimation of its mean value and its covariance matrix are given by Eq. (3.1) (centered and identity matrix). From Eq. (5.20), it can be deduced that, for $1 \leq m \leq r \leq 3$, we have

$$\boldsymbol{\mathcal{Y}}^{mr} = \underline{\boldsymbol{y}}^{mr} + [\Phi_y^{mr}]\,[\kappa]^{1/2}\,\boldsymbol{H} \quad , \quad \underline{\boldsymbol{y}}^{mr} \in \mathbb{R}^{n_y} \quad , \quad [\Phi_y^{mr}] \in \mathbb{M}_{n_y,\nu}\,, \tag{5.22}$$

$$\boldsymbol{\mathcal{G}} = \underline{\boldsymbol{g}} + [\Phi_g]\,[\kappa]^{1/2}\,\boldsymbol{H} \quad , \quad \underline{\boldsymbol{g}} \in \mathbb{R}^{n_g} \quad , \quad [\Phi_g] \in \mathbb{M}_{n_g,\nu}\,, \tag{5.23}$$

$$\boldsymbol{W} = \underline{\boldsymbol{w}} + [\Phi_w]\,[\kappa]^{1/2}\,\boldsymbol{H} \quad , \quad \underline{\boldsymbol{w}} \in \mathbb{R}^{n_w} \quad , \quad [\Phi_w] \in \mathbb{M}_{n_w,\nu}\,. \tag{5.24}$$

The training set $\mathcal{D}_d$ introduced in Definition 1 is written as

$$\mathcal{D}_d = \{\boldsymbol{\eta}_d^1, \ldots, \boldsymbol{\eta}_d^{N_d}\} \quad , \quad \boldsymbol{\eta}_d^j \in \mathbb{R}^\nu\,. \tag{5.25}$$

Figure 1 displays the distribution of the eigenvalues κ_α for the 3 cases, SC1, SC2, and SC3 of scale separation, computed with the training set and the prior probability model.

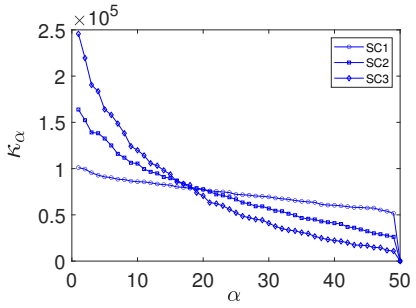

Figure 1: For cases, SC1, SC2, and SC3, distribution of the eigenvalues κ_α.

5.7. Definition of the random normalized residue induced by the use of the constrained learned set

The construction of the random normalized residue is based on a similar approach of the one that has been used in [96]. Let us consider the constrained learned set $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}} = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}^i}, \ldots, \boldsymbol{\eta}^N_{\boldsymbol{\lambda}^i}\}$ generated with Algorithm 1 for iteration i. Using Eqs. (5.22) to (5.24), for $\ell \in \{1, \ldots, N\}$, the corresponding realizations $\boldsymbol{y}^{mr,\ell}_{\boldsymbol{\lambda}^i} \in \mathbb{R}^{n_y}$ for $1 \leq m \leq r \leq 3$, $\boldsymbol{g}^{\ell}_{\boldsymbol{\lambda}^i} \in \mathbb{R}^{n_g}$, and $\boldsymbol{w}^{\ell}_{\boldsymbol{\lambda}^i} \in \mathbb{R}^{n_w}$, are computed by

$$\boldsymbol{y}^{mr,\ell}_{\boldsymbol{\lambda}^i} = \underline{\boldsymbol{y}}^{mr} + [\Phi^{mr}_y]\,[\kappa]^{1/2}\,\boldsymbol{\eta}^{\ell}_{\boldsymbol{\lambda}^i}\,, \tag{5.26}$$

$$\boldsymbol{g}^{\ell}_{\boldsymbol{\lambda}^i} = \underline{\boldsymbol{g}} + [\Phi_g]\;[\kappa]^{1/2}\,\boldsymbol{\eta}^{\ell}_{\boldsymbol{\lambda}^i}\,, \tag{5.27}$$

$$\boldsymbol{w}^{\ell}_{\boldsymbol{\lambda}^i} = \underline{\boldsymbol{w}} + [\Phi_w]\;[\kappa]^{1/2}\,\boldsymbol{\eta}^{\ell}_{\boldsymbol{\lambda}^i}\,. \tag{5.28}$$

For $1 \leq m \leq r \leq 3$ and for $\ell \in \{1, \ldots, N\}$ the realization $\boldsymbol{\mathcal{R}}^{mr,\ell}_{\boldsymbol{\lambda}^i}$ of the $\mathbb{R}^{n_y}$-valued random residue are computed using Eq. (5.9) and Eqs. (5.26) to (5.28), $\boldsymbol{\mathcal{R}}^{mr,\ell}_{\boldsymbol{\lambda}^i} = \boldsymbol{\mathcal{N}}^{mr}(\boldsymbol{y}^{mr,\ell}_{\boldsymbol{\lambda}^i}, \boldsymbol{g}^{\ell}_{\boldsymbol{\lambda}^i}, \boldsymbol{w}^{\ell}_{\boldsymbol{\lambda}^i})$. We define the realization $\hat{\rho}^{\ell}_{\boldsymbol{\lambda}^i}$ of the random residue $\hat{\rho}_{\boldsymbol{\lambda}^i}$ by

$$\hat{\rho}^{\ell}_{\boldsymbol{\lambda}^i} = \frac{1}{\sqrt{6\,n_y}} \left(\sum_{1 \leq m \leq r \leq 3} \| \boldsymbol{\mathcal{R}}^{mr,\ell}_{\boldsymbol{\lambda}^i} \|^2 \right)^{1/2} . \tag{5.29}$$

Finally, we define the realization $\rho^{\ell}_{\boldsymbol{\lambda}^i}$ of the random normalized residue $\rho_{\boldsymbol{\lambda}^i}$ by

$$\rho^{\ell}_{\boldsymbol{\lambda}^i} = \frac{\hat{\rho}^{\ell}_{\boldsymbol{\lambda}^i}}{\underline{\hat{\rho}}_0}\,, \tag{5.30}$$

in which $\underline{\hat{\rho}}_0$ is the estimation of $E\{\rho_{\boldsymbol{\lambda}^1}\}$ using the constrained learned set $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^1}}$ of the first iteration $i = 1$. Since for $i = 1$, $\boldsymbol{\lambda}^{i-1} = \boldsymbol{\lambda}^0 = \boldsymbol{0}_{n_c}$, $\rho_{\boldsymbol{\lambda}^1}$ is the random normalized residue of the constrained learned set without taking into account the constraints and we have, for $i = 1$, $E\{\rho_{\boldsymbol{\lambda}^1}\} = E\{\hat{\rho}_{\boldsymbol{\lambda}^1}\}/\underline{\hat{\rho}}_0 = 1$.

5.8. Defining function $\boldsymbol{h}^c$ related to the constraints and defining the targets represented by $\boldsymbol{b}^c$

We define the function $\boldsymbol{\eta} \mapsto \boldsymbol{h}^c(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{n_c}$ related to the constraints (see Eq. (2.2)) and we define the target represented by vector $\boldsymbol{b}^c$ given in $\mathbb{R}^{n_c}$. Three constraints are introduced, the second-order moment of the random normalized residue (see Section 5.7) and two statistical moments: the normalized mean value of the random effective/apparent elasticity matrix $[\mathbb{C}^{\text{eff}}]$ and its coefficient of dispersion (see Section 5.5). Below we consider iteration i, and then $\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^{i-1}, \boldsymbol{\lambda}^i$ are known.

(i) The random normalized residue $\rho_{\boldsymbol{\lambda}^i}$ whose realizations are defined by Eq. (5.30), is an implicit function of $\boldsymbol{H}_{\boldsymbol{\lambda}^i}$, that we can rewrite as $\rho_{\boldsymbol{\lambda}^i}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})$. The second-order moment of the normalized random residue is then written as $E\{(\rho_{\boldsymbol{\lambda}^i}(\boldsymbol{H}_{\boldsymbol{\lambda}^i}))^2\}$ yielding $h^c_{\rho}(\boldsymbol{H}_{\boldsymbol{\lambda}^i}) = (\rho_{\boldsymbol{\lambda}^i}(\boldsymbol{H}_{\boldsymbol{\lambda}^i}))^2$. Therefore, using the notation of Eq. (3.55), we have $E\{\,h^c_{\rho}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\} = b^c_{\rho}$, in which we choose $b^c_{\rho} = 1$ (this value is close to the value of the second-order moment of the random normalized residue without constraint) and where $\boldsymbol{\eta} \mapsto h^c_{\rho}(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}$ is a positive-valued implicit function.

(ii) The second constraint is given by Eq. (5.13). Transforming the upper triangular matrix of $[\mathbb{C}^{\text{eff}}_n] \in \mathbb{M}^+_6$ in a $\mathbb{R}^{21}$-vector, Eq. (5.13) is rewritten, similarly to Eq. (3.55), as $E\{\,\boldsymbol{h}^c_{\mathbb{C}}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\} = \boldsymbol{b}^c_{\mathbb{C}}$, in which $\boldsymbol{\eta} \mapsto \boldsymbol{h}^c_{\mathbb{C}}(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{21}$ is an implicit function and where $\boldsymbol{b}^c_{\mathbb{C}} \in \mathbb{R}^{21}$ is the reshaping of the upper triangular matrix of $[\,\underline{\mathbb{C}}^{\text{exp}}_n] \in \mathbb{M}^+_6$ defined by Eq. (5.12).

(iii) The last constraint is given by Eq. (5.17) that is rewritten, using Eq. (3.55), as $E\{\,h^c_{\delta}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\} = b^c_{\delta}$, in which $b^c_{\delta} = (\mu_{\text{eff}}\,\delta^{\text{exp}}/\mu_{\text{exp}})^2$ and where $\boldsymbol{\eta} \mapsto h^c_{\delta}(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}$ is a positive-valued implicit function.

(iv) Finally, for given $\boldsymbol{\lambda}$, and in particular for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$ yielding $\boldsymbol{H}^c = \boldsymbol{H}_{\boldsymbol{\lambda}}$, the constraint is defined by Eq. (3.55) with

$$\boldsymbol{\eta} \mapsto \boldsymbol{h}^c(\boldsymbol{\eta}) = (h^c_{\rho}(\boldsymbol{\eta}), \boldsymbol{h}^c_{\mathbb{C}}(\boldsymbol{\eta}), h^c_{\delta}(\boldsymbol{\eta})) : \mathbb{R}^{\nu} \to \mathbb{R}^{n_c} = \mathbb{R} \times \mathbb{R}^{21} \times \mathbb{R}\,, \tag{5.31}$$

$$b^c = (b^c_\rho, \boldsymbol{b}^c_{\mathbb{C}}, b^c_\delta) \in \mathbb{R}^{n_c} = \mathbb{R} \times \mathbb{R}^{21} \times \mathbb{R}\,, \tag{5.32}$$

in which $n_c = 23$. Using the mathematical developments presented in [148], it can be verified that $\boldsymbol{h}^c$ satisfies Hypothesis 1. We will also assume that Eq. (2.6) holds. Indeed, verifying *a priori* that the components of $\boldsymbol{h}^c$ are algebraically independent is very difficult for the considered problem. This hypothesis will indirectly be verified by checking that the solution $\boldsymbol{\lambda}^{\rm sol}$ is well identified (see Proposition 1-(c) and Section 5.9).

5.9. Error function and convergence analysis of the sequence of MCMC generator

The error function is defined by Eqs. (3.58) and (3.59). The constraint on the random normalized residue (see Section 5.8-(i)) is introduced with $b^c_\rho = 1$ in order to avoid the increasing of $E\{\, h^c_\rho(\boldsymbol{H}^c)\}$ with respect to $E\{\, h^c_\rho(\boldsymbol{H}_{\boldsymbol{\lambda}^1})\}$ (first iteration without constraints effects). This constraint, which is taken into account in Algorithm 1 for computing $\{\boldsymbol{\lambda}^i, i = 1, \ldots, i_{\max}\}$, is not taken into account in the error function (see Eq. (3.58)) to identify the index $i_{\rm sol}$ (see Eq. (3.57)) of the optimal value $\boldsymbol{\lambda}^{\rm sol} = \boldsymbol{\lambda}^{i_{\rm sol}}$ of $\boldsymbol{\lambda}^i$. Consequently, Eq. (3.58) is written, for $i \in \{1, \ldots, i_{\max}\}$,

$$\mathrm{err}(i) = \sqrt{(\mathrm{err}_{\mathbb{C}}(i) \,/\, \mathrm{err}_{\mathbb{C}}(1))^2 + (\mathrm{err}_\delta(i) \,/\, \mathrm{err}_\delta(1))^2}\,, \tag{5.33}$$

$$\mathrm{err}_{\mathbb{C}}(i) = \| \boldsymbol{b}^c_{\mathbb{C}} - E\{\, \boldsymbol{h}^c_{\mathbb{C}}(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}\| \,/\, \| \boldsymbol{b}^c_{\mathbb{C}}\| \quad , \quad \mathrm{err}_\delta(i) = |\, b^c_\delta - E\{\, h^c_\delta(\boldsymbol{H}_{\boldsymbol{\lambda}^i})\}\,| \,/\, b^c_\delta\,. \tag{5.34}$$

For the three cases, SC1, SC2,and SC3, Fig. 2 displays the error function $i \mapsto \mathrm{err}(i)$ defined by Eq. (5.33), computed with the constrained learned set $\mathcal{D}_{\boldsymbol{H}_{\boldsymbol{\lambda}^i}}$ for $N = 1000$, 2000, 6000, and 10 000. It can be seen that convergence is reached for $N = 10\,000$ and that, at convergence, function $i \mapsto \mathrm{err}(i)$ is relatively smooth (that is not the case for $N = 1000$). These graphs show a good illustration of the convergence of the sequence in N of the MCMC generator using the statistical surrogate model $\hat{\boldsymbol{h}}^N$ of $\boldsymbol{h}^c$ (see Proposition 4). When the convergence is reached for $N = 10\,000$, Table 3 gives the value of $i_{\rm sol}$ such that $\boldsymbol{\lambda}^{\rm sol} = \boldsymbol{\lambda}^{i_{\rm sol}}$ (see Eq. (3.57)) and the corresponding value $\mathrm{err}(i_{\rm sol})$ of the error.

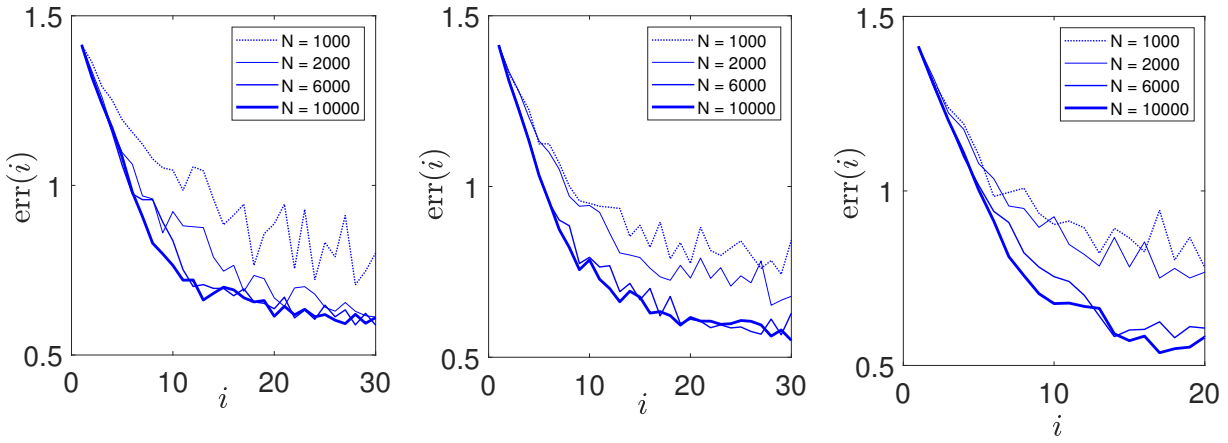

Figure 2: For scale separation, SC1 (left figure), SC2 (central figure), and SC3 (right figure), graph of error function $i \mapsto \mathrm{err}(i)$ for $N = 1000$, 2000, 6000, and 10 000.

Table 3: For cases SC1, SC2, and SC3 of scale separation, value $\mathrm{err}(i_{\rm sol})$ of the error function for the solution $\boldsymbol{\lambda}^{\rm sol} = \boldsymbol{\lambda}_{i_{\rm sol}}$ computed with the constrained learned set for $N = 10\,000$.

Case	SC1	SC2	SC3
$i_{\max}$	30	30	20
$i_{\rm sol}$	27	30	17
$\mathrm{err}(i_{\rm sol})$	0.5924	0.5495	0.5364

5.10. Residue and posterior second-order moments of the random effective/apparent elasticity matrix estimated with the constrained learned set

The posterior statistics of the random normalized residue and the random effective/apparent elasticity matrix are estimated with the constrained learned set $\mathcal{D}_{\boldsymbol{H}^c} = \mathcal{D}_{\boldsymbol{H}_{\lambda^{\text{sol}}}}$ as a function of N for the three cases, SC1, SC2, and SC3.
(i) The posterior second-order moment of the random normalized residue is $E\{\rho_c^2\}$ and compared to 1, see Section 5.8-(i).
(ii) The posterior mean value is $[\,\underline{\mathbb{C}}_c^{\text{eff}}\,] = E\{[\mathbb{C}_c^{\text{eff}}]\}$ and is compared to $[\,\underline{\mathbb{C}}^{\text{exp}}\,]$ (see Eq. (5.13) for the normalized version). We also consider the Frobenius norm $\|\,[\,\underline{\mathbb{C}}_c^{\text{eff}}\,]\,\|_F$ that is compared to $\|\,[\,\underline{\mathbb{C}}^{\text{exp}}\,]\,\|_F$.
(iii) For the posterior dispersion coefficient of $[\mathbb{C}_c^{\text{eff}}]$, instead of comparing δ_c^{eff} to δ^{exp} (see Eq. (5.16)), we use the maximum likelihood consisting in comparing $\delta_{c,\text{ML}}^{\text{eff}}$ to δ^{exp} in which

$$\delta_{c,\text{ML}}^{\text{eff}} = \sqrt{\delta_{2,\text{ML}}^{\text{eff}}} \quad , \quad \delta_{2,\text{ML}}^{\text{eff}} = \max_{\delta_2} p_{\Delta_{2,C}^{\text{eff}}}(\delta_2)\,, \tag{5.35}$$

where $p_{\Delta_{2,C}^{\text{eff}}}$ is the pdf of the random variable $\Delta_{2,c}^{\text{eff}}$ defined by Eq. (5.18), at convergence $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$ (introduction of subscript "c"). For the three cases, SC1, SC2, and SC3, Fig. 3 displays the graph of the posterior pdf $\delta_2 \mapsto p_{\Delta_{2,C}^{\text{eff}}}(\delta 2)$ of random variable $\Delta_{2,c}^{\text{eff}}$, estimated with the constrained learned set for $N = 10\,000$ and its prior counterpart estimated with the training set (constructed using the prior probability model). Fig. 4 (left figure) displays the graph of the Frobenius norm $N \mapsto \|\,[\,\underline{\mathbb{C}}_c^{\text{eff}}(N)]\,\|_F$ of the posterior mean value $[\,\underline{\mathbb{C}}_c^{\text{eff}}(N)] = E\{\,[\mathbb{C}_c^{\text{eff}}(N)]\}$ of the random effective/apparent elasticity matrix $[\mathbb{C}_c^{\text{eff}}(N)]$ as a function of N and estimated using the constrained learned set, while Fig. 4 (right figure) shows the graph of the maximum likelihood $N \mapsto \delta_{c,\text{ML}}^{\text{eff}}(N)$ of the coefficient of dispersion of $[\mathbb{C}_c^{\text{eff}}(N)]$ defined by Eq. (5.35). Fig. 5 shows the graph of the posterior pdf $r \mapsto p_{\rho_c}(r)$ of the random normalized residue ρ_c, estimated with the constrained learned set for $N = 10\,000$ and its counterpart for the estimation performed with the training set (constructed using the prior probability model).
(iv) For cases SC1, SC2, and SC3, Table 4 gives the posterior statistics computed with the constrained learned set for $N = 10\,000$ (subscript "c"), the prior statistics computed with the training set (subscript "d"), and the targets (superscript "exp"), for the second-order moment of the random normalized residue, for the Frobenius norm of the mean value of the random effective/apparent elasticity matrix, and for the coefficient of dispersion of this random matrix. For the same three cases, Table 5 gives the values of the entries of the mean matrices $[\,\underline{\mathbb{C}}_d^{\text{eff}}\,]$ computed with the training set, $[\,\underline{\mathbb{C}}_c^{\text{eff}}\,]$ computed with the constrained learned set for $N = 10\,000$, and $[\,\underline{\mathbb{C}}^{\text{exp}}\,]$ for the targets. Note that entries $(4,5)$, $(4,6)$, and $(5,6)$, which are small with respect to the other entries, are not given.

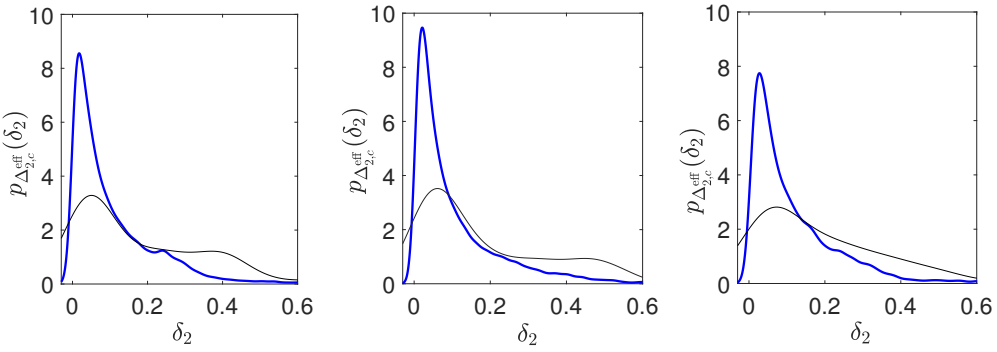

Figure 3: For scale separation, SC1 (left figure), SC2 (central figure), and SC3 (right figure), graph of the posterior pdf of random variable $\Delta_{2,c}^{\text{eff}}$ (thick blue line) estimated with the constrained learned set for $N = 10\,000$ and its prior counterpart (thin black line) corresponding to an estimation with the training set (constructed using the prior probability model).

5.11. Posterior probability model of parameters

The prior probability model concerns the $\mathbb{R}^{n_g}$-valued random variable $\boldsymbol{\mathcal{G}}$ that corresponds to the spatial discretization of the $\mathbb{R}^{21}$-valued random field $\boldsymbol{G}$, and the $\mathbb{R}^3$-valued random variable $\boldsymbol{W}$ that is related (see Eq. (5.6)) to the

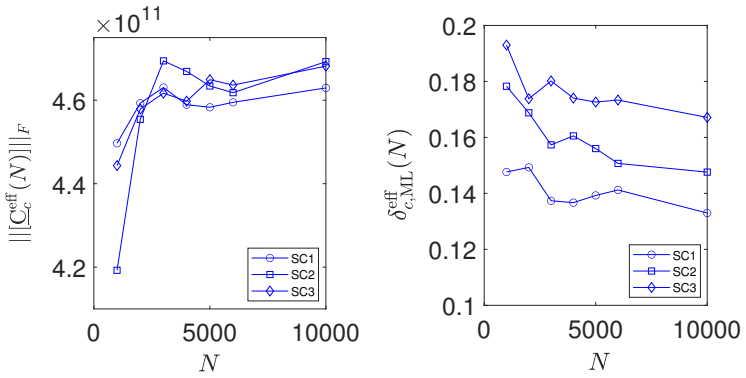

Figure 4: For scale separation, SC1, SC2, and SC3, graph of $N \mapsto \|[\underline{\mathbb{C}}_c^{\text{eff}}(N)]\|_F$ (left figure) and $N \mapsto \delta_{c,\text{ML}}^{\text{eff}}(N)$ (right figure), computed with the constrained learned set.

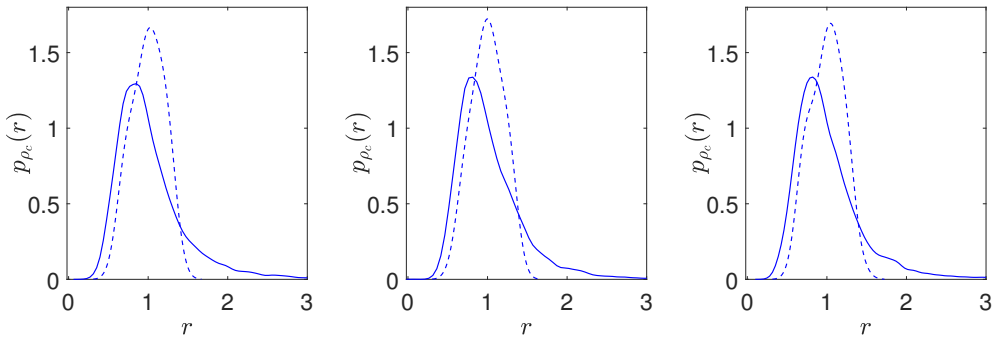

Figure 5: For scale separation, SC1 (left figure), SC2 (central figure), and SC3 (right figure), graph of the posterior pdf of random normalized residue ρ_c (solid line) estimated with the constrained learned set for $N = 10\,000$ and its prior counterpart (dashed line) corresponding to an estimation with the training set (constructed using the prior probability model).

random bulk modulus C_{bulk} and the random shear modulus C_{shear}, which control the elasticity tensor of the mean isotropic model at mesoscale, and to the dispersion coefficient $\delta_{\mathbb{C}}$ that controls the level of anisotropic statistical fluctuations of the random apparent elasticity field at mesoscale (see Section 5.2). For cases SC1, SC2, and SC3, Fig. 6 displays the posterior pdf $c \mapsto p_{C_{\text{bulk}}}(c)$ of C_{bulk} (left figure), $c \mapsto p_{C_{\text{shear}}}(c)$ of C_{shear} (central figure), and $c \mapsto p_{\delta_{\mathbb{C}}}(c)$ of $\delta_{\mathbb{C}}$ (right figure), estimated with the constrained learned set for $N = 10\,000$, and their prior counterparts estimated with the training set constructed using the prior probability model. It should be noted that for each one of the random variables C_{bulk}, C_{shear}, and $\delta_{\mathbb{C}}$, its prior probability model is the same for the three cases and consequently, does not depend on the case contrary to its posterior probability model that depends on it.

5.12. Discussion about the presented results

(i) The results obtained with the posterior model (see Tables 4 and 5) show that the constrained learned set significantly improves the prior probability model used for generating the training set. The comparison of the posterior statistics with the targets are good.

(ii) As explained in Section 5.9 the residue constraint is taken into account for the generation of the constrained learned set, but does not intervene in the estimation of the optimal value $\boldsymbol{\lambda}^{\text{sol}}$ of $\boldsymbol{\lambda}$ using the error function. Nevertheless and as expected, Fig. 5 and Table 4 show that the residue is controlled and stayed small with respect to the reference (the training) for the optimal solution.

(iii) The convergence of the sequence of MCMC generators with respect to the number of points generated in the constrained learned set is good as shown by Figs. 2 and 4 in accordance to Proposition 4.

Table 4: For cases SC1, SC2, and SC3, posterior statistics computed with the constrained learned set for $N = 10\,000$ (subscript "c"), prior statistics (subscript "d") computed with the training set, and targets.

	SC1	SC2	SC3
$E\{\rho_c^2\}$	1.2938	1.2687	1.2413
b_ρ^c	1	1	1
$\|[\underline{\mathbb{C}}_d^{\text{eff}}]\|_F \times 10^{11}$	4.2106	4.1925	4.1943
$\|[\underline{\mathbb{C}}_c^{\text{eff}}]\|_F \times 10^{11}$	4.6294	4.6923	4.6816
$\|[\underline{\mathbb{C}}^{\text{exp}}]\|_F \times 10^{11}$	4.6317	4.6549	4.6706
$\delta_{d,\text{ML}}^{\text{eff}}$	0.2257	0.2469	0.2701
$\delta_{c,\text{ML}}^{\text{eff}}$	0.1329	0.1476	0.1671
δ^{exp}	0.0946	0.1374	0.1825

Table 5: For cases SC1, SC2, and SC3, values of $[\underline{\mathbb{C}}_d^{\text{eff}}]_{ij}$ computed with the training set, $[\underline{\mathbb{C}}_c^{\text{eff}}]_{ij}$ computed with the constrained learned set for $N = 10\,000$, and $[\underline{\mathbb{C}}^{\text{exp}}]_{ij}$ for the targets.

	SC1			SC2			SC3		
Entries of (6×6) matrix	$[\underline{\mathbb{C}}_d^{\text{eff}}]$	$[\underline{\mathbb{C}}_c^{\text{eff}}]$	$[\underline{\mathbb{C}}^{\text{exp}}]$	$[\underline{\mathbb{C}}_d^{\text{eff}}]$	$[\underline{\mathbb{C}}_c^{\text{eff}}]$	$[\underline{\mathbb{C}}^{\text{exp}}]$	$[\underline{\mathbb{C}}_d^{\text{eff}}]$	$[\underline{\mathbb{C}}_c^{\text{eff}}]$	$[\underline{\mathbb{C}}^{\text{exp}}]$
(1, 1)	2.0904	2.2600	2.2652	2.0792	2.2914	2.2810	2.0465	2.2751	2.2946
(1, 2)	0.7427	0.8809	0.8753	0.7269	0.8982	0.8809	0.7140	0.8874	0.8824
(1, 3)	0.7458	0.8804	0.8745	0.7324	0.8983	0.8800	0.7381	0.8999	0.8826
(2, 2)	2.0832	2.2603	2.2668	2.0786	2.2946	2.2846	2.0917	2.2830	2.2822
(2, 3)	0.7451	0.8802	0.8734	0.7486	0.8950	0.8754	0.7471	0.9045	0.8808
(3, 3)	2.0839	2.2647	2.2680	2.0777	2.2841	2.2697	2.1038	2.2928	2.2812
(4, 4)	0.6702	0.6909	0.6958	0.6785	0.6985	0.7003	0.6835	0.6976	0.7027
(5, 5)	0.6714	0.6903	0.6949	0.6732	0.6950	0.6960	0.6726	0.6872	0.6980
(6, 6)	0.6713	0.6924	0.6960	0.6727	0.6958	0.6970	0.6749	0.6933	0.6991

(iv) The dispersion of the target, measured by the value of δ^{exp}, is smaller than the one of the prior probability model (with which the training set has been constructed). So, it was expected that the posterior random effective/apparent elasticity matrix be less dispersed than the one exhibited by the prior probability model. The results confirm this point as it can be seen in Tables 4 and 5 and also in Figs. 3 and 6.
(v) Comment (ii) is extended as follows. It can be seen that the target is "well" reached for the mean value of the random effective/apparent elasticity matrix (see Table 5), while it is "less well" reached for its coefficient of dispersion computed using the maximum likelihood. This can be explained by the fact that the second-order moment $E\{\rho_c^2\}$ of the random normalized residue of the equation is kept small during the probabilistic learning of the posterior probability measure, what prevents the dispersion coefficient from reaching its target. We have carried out numerical tests without imposing the constraint related to $E\{\rho_c^2\}$, which should stay close to 1. We have observed that the dispersion coefficient "reasonably" reached its target but that $E\{\rho_c^2\}$ did not stay close to 1 but took on significant values greater than 1. There is indeed a choice of objective between (1) correctly satisfying the constraint on the dispersion coefficient while degrading the value of $E\{\rho_c^2\}$ or (2) preserving a small value of $E\{\rho_c^2\}$ to the detriment of perfectly reaching the target for the dispersion coefficient. We have chosen to present the compromise consisting in taking into account the constraint on the residue during the probabilistic learning process, but the error function that we have chosen to identify the optimal value $\boldsymbol{\lambda}^{\text{sol}}$ of $\boldsymbol{\lambda}$ does not take into account the constraint on the residue.
(vi) A last comment concerns the effects of no scale separation. As expected, for the three cases SC1, SC2, and SC3, Table 4 shows that the coefficient of dispersion is significant and increases with the spatial correlation lengths of

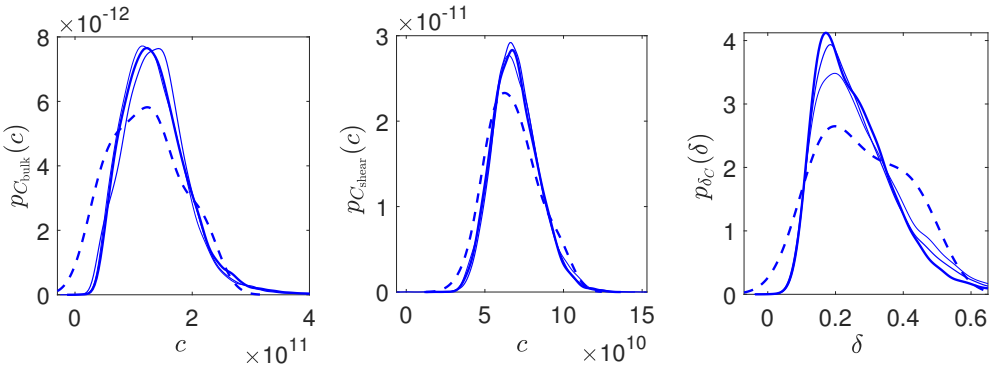

Figure 6: Graphs of the posterior pdf $c \mapsto p_{C_{\text{bulk}}}(c)$ (left figure), $c \mapsto p_{C_{\text{shear}}}(c)$ (central figure), and $\delta \mapsto p_{\delta_{\mathbb{C}}}(\delta)$ (right figure), estimated with the constrained learned set for $N = 10\,000$, for cases, SC1 (thin solid line), SC2 (med solid line), and SC3 (thick solid line), and the corresponding prior pdf estimated with the training set constructed using the prior probability model (dashed line).

the random apparent elasticity field at mesoscale, inducing statistical fluctuations of the effective/apparent elasticity tensor at macroscale. It should be noted that, even for the case SC1, for which homogenization in the plane of the plate (domain Ω) is guaranteed (the correlation lengths $\underline{L}_{C1}$ and $\underline{L}_{C2}$ being much lower than 1), this is not the case for the correlation length $\underline{L}_{C3}$ that is equal to the thickness of the plate. Consequently, there is no homogenization at the macroscopic scale and the effective/apparent elasticity tensor remains random and is not deterministic.

6. Conclusions

In this paper, we have presented a general methodology to estimate a posterior probability model for a stochastic boundary value problem from a prior probability model. The given targets are statistical moments for which the underlying realizations are not available. Under these conditions, it has been proposed to use the Kullback-Leibler divergence minimum principle for estimating the posterior probability measure, given the prior probability measure and the constraints related to the targets of the statistical moments. We have proposed the construction of a statistical surrogate model of the implicit mapping that represents the constraints. The constrained learned set, which defines the posterior model, is constructed using only a training set constituted of a small number of points. A mathematical analysis of the proposed methodology has been presented. We have defined the required mathematical hypotheses, which have allowed us to prove the convergence of introduced approximations. We have also given all the necessary numerical elements, which facilitate the implementation of the methodology in a parallel computing framework.

The application presented to illustrate the proposed theory is also, as such, a contribution to the three-dimensional stochastic homogenization of heterogeneous linear elastic media in the case of a non-separation of the microscale and macroscale, that is to say, when there are significant statistical fluctuations in the effective/apparent elasticity tensor at macroscale. The prior stochastic model of the elasticity tensor field at the mesoscopic scale is an advanced model, recently proposed, which takes into account uncertainties on its spectral measure. In addition to the statistical moments of the random effective/apparent elasticity tensor, for the construction of the posterior probability measure by probabilistic learning inference, the second-order moment of the random normalized residue of the stochastic partial differential equation has been added as a constraint. This constraint guarantees that the algorithm seeks to bring the statistical moments closer to their targets while preserving a small residue. The results obtained are those which were expected and give a very good illustration of the theory developed for a non-trivial application.

References

[1] J. M. Bernardo, A. F. M. Smith, Bayesian Theory, John Wiley & Sons, Chichester, 2000.
[2] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.
[3] J. C. Spall, Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Vol. 65, John Wiley & Sons, 2005.
[4] P. Congdon, Bayesian Statistical Modelling, Vol. 704, John Wiley & Sons, 2007.

[5] B. P. Carlin, T. A. Louis, Bayesian Methods for Data Analysis, Chapman and Hall/CRC, 2008.
[6] J. E. Gentle, Computational Statistics, Springer, New York, 2019. doi:10.1007/978-0-387-98144-4.
[7] J. Marin, P. Pudlo, C. Robert, R. Ryder, Approximate Bayesian computational methods, Statistics and Computing 22 (6) (2012) 1167–1180. doi:10.1007/s11222-011-9288-2.
[8] G. Givens, J. Hoeting, Computational Statistics, 2nd Edition, John Wiley and Sons, Hoboken, New Jersey, 2013.
[9] S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, R. E. McCulloch, Bayes and big data: The consensus monte carlo algorithm, International Journal of Management Science and Engineering Management 11 (2) (2016) 78–88. doi:10.1080/17509653.2016.1142191.
[10] R. Ghanem, D. Higdon, H. Owhadi, Handbook of Uncertainty Quantification, Vol. 1 to 3, Springer, Cham, Switzerland, 2017. doi:10.1007/978-3-319-12385-1.
[11] C. Soize, R. Ghanem, C. Desceliers, Sampling of Bayesian posteriors with a non-Gaussian probabilistic learning on manifolds from a small dataset, Statistics and Computing 30 (5) (2020) 1433–1457. doi:10.1007/s11222-020-09954-6.
[12] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Vol. 160, Springer Science & Business Media, 2005. doi:10.1007/b138659.
[13] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, Journal of Computational Physics 224 (2) (2007) 560–586. doi:10.1016/j.jcp.2006.10.010.
[14] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559. doi:10.1017/S0962492910000061.
[15] C. Soize, A computational inverse method for identification of non-Gaussian random fields using the Bayesian approach in very high dimension, Computer Methods in Applied Mechanics and Engineering 200 (45-46) (2011) 3083–3099. doi:10.1016/j.cma.2011.07.005.
[16] T. A. El Moselhy, Y. M. Marzouk, Bayesian inference with optimal maps, Journal of Computational Physics 231 (23) (2012) 7815–7850. doi:10.1016/j.jcp.2012.07.022.
[17] H. G. Matthies, E. Zander, B. V. Rosić, A. Litvinenko, O. Pajonk, Inverse problems in a Bayesian setting, in: Computational Methods for Solids and Fluids, Vol. 41, Springer, 2016, pp. 245–286. doi:10.1007/978-3-319-27996-1˙10.
[18] I. Bilionis, N. Zabaras, Bayesian uncertainty propagation using Gaussian processes, in: R. Ghanem, D. Higdon, O. Houman (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 15, pp. 555–600.
[19] M. Dashti, A. M. Stuart, The Bayesian approach to inverse problems, in: R. Ghanem, D. Higdon, O. Houman (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 10, pp. 311–428. doi:10.1007/978-3-319-12385-1˙7.
[20] M. Arnst, B. Abello Álvarez, J.-P. 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. doi:10.1016/j.jcp.2017.08.005.
[21] A. Spantini, T. Cui, K. Willcox, L. Tenorio, Y. Marzouk, Goal-oriented optimal approximations of Bayesian linear inverse problems, SIAM Journal on Scientific Computing 39 (5) (2017) S167–S196. doi:10.1137/16M1082123.
[22] G. Perrin, C. Soize, Adaptive method for indirect identification of the statistical properties of random fields in a Bayesian framework, Computational Statistics 35 (1) (2020) 111–133. doi:10.1007/s00180-019-00936-5.
[23] Y. Shen, D. Cornford, M. Opper, C. Archambeau, Variational Markov chain Monte Carlo for Bayesian smoothing of non-linear diffusions, Computational Statistics 27 (1) (2012) 149–176. doi:10.1007/s00180-011-0246-4.
[24] N. Depraetere, M. Vandebroek, A comparison of variational approximations for fast inference in mixed logit models, Computational Statistics 32 (1) (2017) 93–125. doi:10.1007/s00180-015-0638-y.
[25] A. Golightly, D. J. Wilkinson, Bayesian sequential inference for nonlinear multivariate diffusions, Statistics and Computing 16 (4) (2006) 323–338. doi:10.1007/s11222-006-9392-x.
[26] P. Fearnhead, Exact and efficient Bayesian inference for multiple changepoint problems, Statistics and Computing 16 (2) (2006) 203–213. doi:10.1007/s11222-006-8450-8.
[27] H. Owhadi, Bayesian numerical homogenization, Multiscale Modeling & Simulation 13 (3) (2015) 812–828. doi:10.1137/140974596.
[28] M. Neil, M. Tailor, D. Marquez, Inference in hybrid Bayesian networks using dynamic discretization, Statistics and Computing 17 (3) (2007) 219–233. doi:10.1007/s11222-007-9018-y.
[29] R. Sambasivan, S. Das, S. K. Sahu, A Bayesian perspective of statistical machine learning for big data, Computational Statistics 35 (3) (2020) 893–930. doi:10.1007/s00180-020-00970-8.
[30] H. Owhadi, C. Scovel, T. Sullivan, On the brittleness of Bayesian inference, SIAM Review 57 (4) (2015) 566–582. doi:10.1137/130938633.
[31] G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallographica Section D: Biological Crystallography 53 (3) (1997) 240–255. doi:10.1107/S0907444996012255.
[32] W. Zhang, F. Zhang, J. Zhang, J. Zhang, J. Zhang, Optimization of identification structure parameters based on recursive maximum likelihood iteration, in: 2018 International Computers, Signals and Systems Conference (ICOMSSC), IEEE, 2018, pp. 726–731. doi:10.1109/ICOMSSC45026.2018.8941760.
[33] R. K. Vinayak, W. Kong, G. Valiant, S. Kakade, Maximum likelihood estimation for learning populations of parameters, in: International Conference on Machine Learning, PMLR, 2019, pp. 6448–6457.
[34] P. Sur, E. J. Candès, A modern maximum-likelihood theory for high-dimensional logistic regression, Proceedings of the National Academy of Sciences 116 (29) (2019) 14516–14525. doi:10.1073/pnas.1810420116.
[35] W. Lio, B. Liu, Uncertain maximum likelihood estimation with application to uncertain regression analysis, Soft Computing 24 (13) (2020) 9351–9360. doi:10.1007/s00500-020-04951-3.
[36] C. Desceliers, R. Ghanem, C. Soize, Maximum likelihood estimation of stochastic chaos representations from experimental data, International Journal for Numerical Methods in Engineering 66 (6) (2006) 978–1001. doi:10.1002/nme.1576.
[37] C. Desceliers, C. Soize, R. Ghanem, Identification of chaos representations of elastic properties of random media using experimental vibration tests, Computational mechanics 39 (6) (2007) 831–838. doi:10.1007/s00466-006-0072-7.
[38] R. Ghanem, P. D. Spanos, Stochastic Finite Elements: a Spectral Approach, Springer-Verlag, New York, 1991.
[39] D. Xiu, G. E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing

24 (2) (2002) 619–644. doi:10.1137/S1064827501387826.
[40] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410. doi:10.1137/S1064827503424505.
[41] X. Wan, G. E. Karniadakis, Multi-element generalized polynomial chaos for arbitrary probability measures, SIAM Journal on Scientific Computing 28 (3) (2006) 901–928. doi:10.1137/050627630.
[42] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, Journal of Computational Physics 228 (6) (2009) 1862–1902. doi:10.1016/j.jcp.2008.11.024.
[43] C. Soize, C. Desceliers, Computational aspects for constructing realizations of polynomial chaos in high dimension, SIAM Journal on Scientific Computing 32 (5) (2010) 2820–2831. doi:10.1137/100787830.
[44] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367. doi:10.1016/j.jcp.2010.12.021.
[45] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Identification of polynomial chaos representations in high dimension from a set of realizations, SIAM Journal on Scientific Computing 34 (6) (2012) A2917–A2945. doi:10.1137/11084950X.
[46] R. Tipireddy, R. Ghanem, Basis adaptation in homogeneous chaos spaces, Journal of Computational Physics 259 (2014) 304–317. doi:10.1016/j.jcp.2013.12.009.
[47] C. Soize, Polynomial chaos expansion of a multimodal random vector, SIAM-ASA Journal on Uncertainty Quantification 3 (1) (2015) 34–60. doi:10.1137/140968495.
[48] S. Abraham, M. Raisee, G. Ghorbanias, F. Contino, C. Lacor, A robust and efficient stepwise regression method for building sparse polynomial chaos expansions, Journal of Computational Physics 332 (2017) 461–474. doi:10.1016/j.jcp.2016.12.015.
[49] P. Tsilifis, R. Ghanem, Reduced Wiener chaos representation of random fields via basis adaptation and projection, Journal of Computational Physics 341 (2017) 102–120. doi:10.1016/j.jcp.2017.04.009.
[50] P. Tsilifis, R. Ghanem, Bayesian adaptation of chaos representations using variational inference and sampling on geodesics, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 474 (2217) (2018) 20180285. doi:10.1098/rspa.2018.0285.
[51] M. Mignolet, C. Soize, Compressed principal component analysis of non-Gaussian vectors, SIAM/ASA Journal on Uncertainty Quantification 8 (4) (2020) 1261–1286. doi:10.1137/20M1322029.
[52] N. Luthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, SIAM/ASA Journal on Uncertainty Quantification 9 (2) (2021) 593–649. doi:10.1137/20M1315774.
[53] S. Russel, P. Norvig, Artifical Intelligence, A Modern Approach, Third Edition, Pearson, Harlow, 2016.
[54] S. Saha, Z. Gan, L. Cheng, J. Gao, O. L. Kafka, X. Xie, H. Li, M. Tajdari, H. A. Kim, W. K. Liu, Hierarchical deep learning neural network (HiDeNN): An artificial intelligence (AI) framework for computational science and engineering, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113452. doi:10.1016/j.cma.2020.113452.
[55] J.-X. Wang, J.-L. Wu, H. Xiao, Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data, Physical Review Fluids 2 (3) (2017) 034603. doi:10.1103/PhysRevFluids.2.034603.
[56] A. P. Singh, S. Medida, K. Duraisamy, Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils, AIAA Journal (2017) 2215–2227doi:10.2514/1.J055595.
[57] J.-L. Wu, H. Xiao, E. Paterson, Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework, Physical Review Fluids 3 (7) (2018) 074602. doi:10.1103/PhysRevFluids.3.074602.
[58] X. Yang, S. Zafar, J.-X. Wang, H. Xiao, Predictive large-eddy-simulation wall modeling via physics-informed neural networks, Physical Review Fluids 4 (3) (2019) 034602. doi:10.1103/PhysRevFluids.4.034602.
[59] R. Swischuk, L. Mainini, B. Peherstorfer, K. Willcox, Projection-based model reduction: Formulations for physics-based machine learning, Computers & Fluids 179 (2019) 704–717. doi:10.1016/j.compfluid.2018.07.021.
[60] H. Li, O. L. Kafka, J. Gao, C. Yu, Y. Nie, L. Zhang, M. Tajdari, S. Tang, X. Guo, G. Li, et al., Clustering discretization methods for generation of material performance databases in machine learning and design optimization, Computational Mechanics 64 (2) (2019) 281–305. doi:10.1007/s00466-019-01716-0.
[61] C. Yu, O. L. Kafka, W. K. Liu, Multiresolution clustering analysis for efficient modeling of hierarchical material systems, Computational Mechanics (2021) 1–14doi:10.1007/s00466-021-01982-x.
[62] M. Raissi, P. Perdikaris, G. E. Karniadakis, Machine learning of linear differential equations using Gaussian processes, Journal of Computational Physics 348 (2017) 683–693. doi:10.1016/j.jcp.2017.07.050.
[63] J. Sirignano, K. Spiliopoulos, DGM: A deep learning algorithm for solving partial differential equations, Journal of Computational Physics 375 (2018) 1339–1364. doi:10.1016/j.jcp.2018.08.029.
[64] M. Raissi, G. E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, Journal of Computational Physics 357 (2018) 125–141. doi:10.1016/j.jcp.2017.11.039.
[65] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics 378 (2019) 686–707. doi:10.1016/j.jcp.2018.10.045.
[66] D. Kats, Z. Wang, Z. Gan, W. K. Liu, G. J. Wagner, Y. Lian, A physics-informed machine learning method for predicting grain structure characteristics in directed energy deposition, Computational Materials Science 202 (2022) 110958. doi:10.1016/j.commatsci.2021.110958.
[67] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117. doi:10.1016/j.neunet.2014.09.003.
[68] B. Schölkopf, A. Smola, K. Müller, Kernel principal component analysis, in: W. Gerstner, A. Germond, M. Hasler, J. Nicoud (Eds.), Artificial Neural Networks ICANN 97, Vol. 1327 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 1997, pp. 583–588.
[69] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 2000. doi:10.1007/978-1-4757-3264-1.
[70] C. C. Aggarwal, C. Zhai, Mining Text Data, Springer Science & Business Media, 2012. doi:10.1007/978-1-4614-3223-4.
[71] A. S. Dalalyan, A. B. Tsybakov, Sparse regression learning by aggregation and Langevin Monte-Carlo, Journal of Computer and System Sciences 78 (5) (2012) 1423–1443. doi:10.1016/j.jcss.2011.12.023.
[72] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT press, 2012.

[73] M.-F. F. Balcan, V. Feldman, Statistical active learning algorithms, in: C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 26, Curran Associates, Inc., 2013, pp. 1295–1303.
URL http://papers.nips.cc/paper/5101-statistical-active-learning-algorithms.pdf
[74] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Vol. 112, Springer, 2013.
[75] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, W. Zhang, Knowledge vault: A web-scale approach to probabilistic knowledge fusion, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge discovery and data mining (2014) 601–610doi:10.1145/2623330.2623623.
[76] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, Nature 521 (7553) (2015) 452–459. doi:10.1038/nature14541.
[77] J. Taylor, R. J. Tibshirani, Statistical learning and selective inference, Proceedings of the National Academy of Sciences 112 (25) (2015) 7629–7634. doi:10.1073/pnas.1507583112.
[78] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (4) (1998) 455–492. doi:10.1023/A:1008306431147.
[79] X. Du, W. Chen, Sequential optimization and reliability assessment method for efficient probabilistic design, ASME Journal of Mechanical Design 126 (2) (2004) 225–233. doi:10.1115/1.1649968.
[80] N. Queipo, R. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, K. Tucker, Surrogate-based analysis and optimization, Progress in Aerospace Science 41 (1) (2005) 1–28. doi:10.1016/j.paerosci.2005.02.001.
[81] R. Byrd, G. Chin, W. Neveitt, J. Nocedal, On the use of stochastic Hessian information in optimization methods for machine learning, SIAM Journal of Optimization 21 (3) (2011) 977–995. doi:10.1137/10079923X.
[82] M. Eldred, Design under uncertainty employing stochastic expansion methods, International Journal for Uncertainty Quantification 1 (2) (2011) 119–146. doi:10.1615/Int.J.UncertaintyQuantification.v1.i2.20.
[83] W. Yao, X. Chen, W. Luo, M. vanTooren, J. Guo, Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles, Progress in Aerospace Sciences 47 (6) (2011) 450–479. doi:10.1016/j.paerosci.2011.05.001.
[84] T. Homem-de Mello, G. Bayraksan, Monte Carlo sampling-based methods for stochastic optimization, Surveys in Operations Researh and Management Science 19 (1) (2014) 56–85. doi:10.1016/j.sorms.2014.05.001.
[85] V. Papadopoulos, G. Soimiris, D. Giovanis, M. Papadrakakis, A neural network-based surrogate model for carbon nanotubes with geometric nonlinearities, Computer Methods in Applied Mechanics and Engineering 328 (2018) 411–430. doi:10.1016/j.cma.2017.09.010.
[86] D. G. Giovanis, M. D. Shields, Data-driven surrogates for high dimensional models using Gaussian process regression on the Grassmann manifold, Computer Methods in Applied Mechanics and Engineering 370 (2020) 113269. doi:10.1016/j.cma.2020.113269.
[87] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86. doi:10.1214/aoms/1177729694.
[88] J. N. Kapur, H. K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, San Diego, 1992.
[89] T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, Hoboken, 2006.
[90] N. Vasconcelos, P. Ho, P. Moreno, The Kullback-Leibler kernel as a framework for discriminant and localized representations for visual recognition, Proceedings of the European Conference on Computer Vision (2004) 430–441doi:10.1007/978-3-540-24672-5˙34.
[91] W. Zhang, S. Shan, X. Chen, W. Gao, Local Gabor binary patterns based on Kullback-Leibler divergence for partially occluded face recognition, IEEE Signal Processing Letters 14 (11) (2007) 875–878. doi:10.1109/LSP.2007.903260.
[92] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, G. Stoltz, et al., Kullback-Leibler upper confidence bounds for optimal sequential allocation, The Annals of Statistics 41 (3) (2013) 1516–1541. doi:10.1214/13.AOS1119.
[93] N. Saleem, G. Ijaz, Low rank sparse decomposition model based speech enhancement using gammatone filterbank and Kullback-Leibler divergence, International Journal of Speech Technology 21 (2) (2018) 217–231. doi:10.1007/s10772-018-9500-2.
[94] S. Filippi, O. Cappé, A. Garivier, Optimism in reinforcement learning and Kullback-Leibler divergence, Proceedings of the 48th Annual Allerton IEEE Conference on Communication, Control, and Computing (2010) 115–122doi:10.1109/ALLERTON.2010.5706896.
[95] C. Soize, R. Ghanem, Physics-constrained non-Gaussian probabilistic learning on manifolds, International Journal for Numerical Methods in Engineering 121 (1) (2020) 110–145. doi:10.1002/nme.6202.
[96] C. Soize, R. Ghanem, Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets, Computer Methods in Applied Mechanics and Engineering 380 (2021) 113777. doi:10.1016/j.cma.2021.113777.
[97] B. Nayroles, G. Touzot, P. Villon, Generalizing the finite element method: diffuse approximation and diffuse elements, Computational Mechanics 10 (5) (1992) 307–318. doi:10.1007/BF00364252.
[98] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, P. Krysl, Meshless methods: an overview and recent developments, Computer Methods in Applied Mechanics and Engineering 139 (1-4) (1996) 3–47. doi:10.1016/S0045-7825(96)01078-X.
[99] C. A. Duarte, J. T. Oden, H-p clouds, an h-p meshless method, Numerical Methods for Partial Differential Equations: An International Journal 12 (6) (1996) 673–705. doi:10.1002/(SICI)1098-2426(199611)12:6¡673::AID-NUM3¿3.0.CO;2-P.
[100] P. Breitkopf, A. Rassineux, G. Touzot, P. Villon, Explicit form and efficient computation of MLS shape functions and their derivatives, International Journal for Numerical Methods in Engineering 48 (3) (2000) 451–466. doi:10.1002/(SICI)1097-0207(20000530)48:3¡451::AID-NME892¿3.0.CO;2-1.
[101] A. Rassineux, P. Villon, J.-M. Savignat, O. Stab, Surface remeshing by local Hermite diffuse interpolation, International Journal for numerical methods in Engineering 49 (1-2) (2000) 31–49. doi:10.1002/1097-0207(20000910/20)49:1/23.0.CO;2-6.
[102] X. Zhang, K. Z. Song, M. W. Lu, X. Liu, Meshless methods based on collocation with radial basis functions, Computational mechanics 26 (4) (2000) 333–343. doi:10.1007/s004660000181.
[103] 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.
[104] C. Soize, R. Ghanem, Probabilistic learning on manifolds, Foundations of Data Science 2 (3) (2020) 279–307. doi:10.3934/fods.2020013.
[105] C. Soize, R. Ghanem, Probabilistic learning on manifolds with partition, International Journal for Numerical Methods in Engineering 123 (1) (2022) 268–290. doi:10.1002/nme.6856.
[106] G. Papanicolaou, S. Varadhan, Boundary Value Problems with Rapidly Oscillating Random Coefficients, North Holland, 1981.

[107] S. Torquato, G. Stell, Microstructure of two-phase random media. v. the n-point matrix probability functions for impenetrable spheres, The Journal of Chemical Physics 82 (2) (1985) 980–987. doi:10.1063/1.448475.
[108] G. Nguetseng, A general convergence result for a functional related to the theory of homogenization, SIAM Journal on Mathematical Analysis 20 (3) (1989) 608–623. doi:10.1137/0520043.
[109] G. Allaire, Homogenization and two-scale convergence, SIAM Journal on Mathematical Analysis 23 (6) (1992) 1482–1518. doi:10.1137/0523084.
[110] K. Sab, On the homogenization and the simulation of random materials, European Journal of Mechanics, A/Solids 11 (5) (1992) 585–607.
[111] K. T. Andrews, S. Wright, Stochastic homogenization of elliptic boundary-value problems with Lp-data, Asymptotic Analysis 17 (3) (1998) 165–184.
[112] M. Ostoja-Starzewski, Random field models of heterogeneous materials, International Journal of Solids and Structures 35 (19) (1998) 2429–2455. doi:10.1016/S0020-7683(97)00144-3.
[113] V. V. Jikov, S. M. Kozlov, O. A. Oleinik, Homogenization of Differential Operators and Integral Functionals, Springer-Verlag, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-84659-5.
[114] W. J. Drugan, J. R. Willis, A micromechanics-based nonlocal constitutive equation and estimates of representative volume element size for elastic composites, Journal of the Mechanics and Physics of Solids 44 (4) (1996) 497–524. doi:10.1016/0022-5096(96)00007-5.
[115] Z.-Y. Ren, Q.-S. Zheng, Effects of grain sizes, shapes, and distribution on minimum sizes of representative volume elements of cubic polycrystals, Mechanics of Materials 36 (12) (2004) 1217–1229. doi:10.1016/j.mechmat.2003.11.002.
[116] K. Sab, B. Nedjar, Periodization of random media and representative volume element size for linear composites, Comptes Rendus Mécanique 333 (2) (2005) 187–195. doi:10.1016/j.crme.2004.10.003.
[117] M. Ostoja-Starzewski, Material spatial randomness: From statistical to representative volume element, Probabilistic Engineering Mechanics 21 (2) (2006) 112–132. doi:10.1016/j.probengmech.2005.07.007.
[118] M. Ostoja-Starzewski, X. Du, Z. Khisaeva, W. Li, Comparisons of the size of the representative volume element in elastic, plastic, thermoelastic, and permeable random microstructures, International Journal for Multiscale Computational Engineering 5 (2) (2007) 73–82. doi:10.1615/IntJMultCompEng.v5.i2.10.
[119] X. Yin, W. Chen, A. To, C. McVeigh, W. K. Liu, Statistical volume element method for predicting microstructure–constitutive property relations, Computer Methods in Applied Mechanics and Engineering 197 (43-44) (2008) 3516–3529. doi:10.1016/j.cma.2008.01.008.
[120] C. Soize, Tensor-valued random fields for meso-scale stochastic model of anisotropic elastic microstructure and probabilistic analysis of representative volume element size, Probabilistic Engineering Mechanics 23 (2-3) (2008) 307–323. doi:10.1016/j.probengmech.2007.12.019.
[121] T. Zhang, F. Pled, C. Desceliers, Robust multiscale identification of apparent elastic properties at mesoscale for random heterogeneous materials with multiscale field measurements, Materials 13 (12) (2020) 2826. doi:10.3390/ma13122826.
[122] C. Soize, Computational stochastic homogenization of heterogeneous media from an elasticity random field having an uncertain spectral measure, Computational Mechanics 68 (2021) 1003–1021. doi:10.1007/s00466-021-02056-8.
[123] I. M. Gelfand, N. I. Vilenkin, Les Distributions. Tome 4. Application de l'Analyse Harmonique, Dunod, 1967. doi:10.1016/0375-9474(67)90547-7.
[124] K. Karhunen, Ueber lineare methoden in der wahrscheinlichkeitsrechnung, Amer. Acad. Sci., Fennicade, Ser. A, I 37 (1947) 3–79.
[125] M. Loève, Functions aleatoires du second ordre, in: P. Lévy (Ed.), Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris, 1948, pp. 366–420.
[126] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Mathematica 30 (1) (1906) 175–193. doi:10.1007/BF02418571.
[127] R. Durrett, Probability, Theory and Examples (5th ed.), Cambridge University Press, Cambridge, 2019.
[128] 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 (10) (2008) 1583–1611. doi:10.1002/nme.2385.
[129] C. Soize, The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions, Vol. Series on Advances in Mathematics for Applied Sciences: Vol 17, World Scientific, Singapore, 1994. doi:10.1142/2347.
[130] D. G. Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, New York, 2009.
[131] A. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations, Vol. 18, Oxford University Press, Oxford: Clarendon Press, New York, 1997. doi:10.1007/s001800000033.
[132] N. Agmon, Y. Alhassid, R. D. Levine, An algorithm for finding the distribution of maximal entropy, Journal of Computational Physics 30 (2) (1979) 250–258. doi:10.1016/0021-9991(79)90102-5.
[133] A. Friedman, Stochastic Differential Equations and Applications, Dover Publications, Inc., Mineola, New York, 2006.
[134] H. Cartan, Cours de Calcul Différentiel, Hermann, Paris, 1985.
[135] T. H. Gronwall, Note on the derivatives with respect to a parameter of the solutions of a system of differential equations, Annals of Mathematics 20 (4) (1919) 292–296. doi:10.2307/1967124.
[136] J. Dieudonné, Calcul Infinitésimal, Hermann, Paris, 1980.
[137] C. T. Kelley, Solving Nonlinear Equations With Newton's Method, SIAM, 2003. doi:10.1137/1.9780898718898.
[138] A. Batou, C. Soize, Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension, SIAM/ASA Journal on Uncertainty Quantification 1 (1) (2013) 431–451. doi:10.1137/120901386.
[139] C. Soize, Uncertainty Quantification. An Accelerated Course with Advanced Applications in Computational Engineering, Springer, New York, 2017. doi:10.1007/978-3-319-54339-0.
[140] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differentials Equations, Springer-Verlag, Heidelberg, 1992.
[141] D. Talay, L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, Stochastic analysis and applications 8 (4) (1990) 483–509. doi:10.1080/07362999008809220.
[142] D. Talay, Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme, Markov Process and Related Fields 8 (2) (2002) 163–198.

[143] 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.
[144] K. Burrage, I. Lenane, G. Lythe, Numerical methods for second-order stochastic differential equations, SIAM Journal on Scientific Computing 29 (1) (2007) 245–264. doi:10.1137/050646032.
[145] C. Soize, I. E. Poloskov, Time-domain formulation in computational dynamics for linear viscoelastic media with model uncertainties and stochastic excitation, Computers & Mathematics with Applications 64 (11) (2012) 3594–3612. doi:10.1016/j.camwa.2012.09.010.
[146] J. Guilleminot, C. Soize, Stochastic model and generator for random fields with symmetry properties: application to the mesoscopic modeling of elastic random media, Multiscale Modeling & Simulation (A SIAM Interdisciplinary Journal) 11 (3) (2013) 840–870. doi:10.1137/120898346.
[147] M. Bornert, T. Bretheau, P. Gilormini, Homogenization in Mechanics of Materials, ISTE Ltd and John Wiley and Sons, New York, 2008.
[148] C. Soize, Stochastic elliptic operators defined by non-Gaussian random fields with uncertain spectrum, The American Mathematical Society Journal Theory of Probability and Mathematical Statistics 105 (2021) 113–136. doi:10.1090/tpms/1159.
[149] J. Guilleminot, C. Soize, On the statistical dependence for the components of random elasticity tensors exhibiting material symmetry properties, Journal of Elasticity 111 (2) (2013) 109–130. doi:10.1007/s10659-012-9396-z.
[150] A. Nouy, C. Soize, Random field representations for stochastic elliptic boundary value problems and statistical inverse problems, European Journal of Applied Mathematics 25 (3) (2014) 339–373. doi:10.1017/S0956792514000072.
[151] G. H. Golub, C. F. Van Loan, Matrix Computations, Second Edition, Johns Hopkins University Press, Baltimore and London, 1993.