An overview on uncertainty quantification and probabilistic learning on manifolds in multiscale mechanics of materials

Christian Soize

An overview on uncertainty quantification and probabilistic learning on manifolds in multiscale mechanics of materials

Christian Soize[a,*]

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

Abstract

An overview of the author works, many of which were carried out in collaboration, is presented. The first part concerns the quantification of uncertainties for complex engineering science systems for which analyzes are now carried out using large numerical simulation models. More recently, machine learning methods have appeared in this field to address certain problems of nonconvex optimization under uncertainties and inverse identification, which are not affordable with standard computer resources. Thus the second part is relative to the presentation of a method of probabilistic learning on manifolds recently proposed for the case of small data and which makes it possible to build statistical surrogate models useful to perform probabilistic inferences. The illustrations are mainly focused on the multiscale analyzes of microstructures made up of heterogeneous continuous materials, which cannot be described in terms of constituents and which are modeled with stochastic apparent quantities at mesoscale.

Keywords: Uncertainty quantification, probabilistic learning, stochastic homogenization, heterogeneous material, multiscale mechanics

1. Introduction

Uncertainty quantification for computational models in science and engineering on the one hand, and probabilistic learning for the construction of metamodels (surrogate models) on the other hand, are very vast and active research fields, which remain very current, and which have generated and generate a very large number of published works for several decades. Consequently, this article cannot be and is not a review of the state of the art, but of course the author has remained attentive to refer the reader to numerous references, which will be given as developements are presented.

The author has contributed to the development and to the use of probability theory and mathematical statistics in Mechanics of complex systems since 1973, in particular to the development of methods for Uncertainty quantification (UQ) in computational stochastic mechanics since the end of the 1990s, and to the development of probabilistic learning on manifolds devoted to the construction of statistical surrogate models (metamodels) for large computational models in engineering science for the last 10 years. The author offers an overview of his work most often carried out in collaboration, concerning current prospective approaches in the field of uncertain computational models and their identification with applications to the mechanics of materials whose microstructures are complex mechanical systems.

This overview is mainly an introduction to this field of research, which provides an overview of the main concepts and methods. However, the use of probability theory and statistics remains a relatively difficult field of mathematics and it is not possible to give in such a paper the details, the mathematical elements, and even certain foundations, which would be necessary for that a reader uninitiated to this domain perfectly understands all the aspects. Inevitably deadlocks must be made and this lack will be overcome by referring to published articles and books. However, the author goal is to help readers to understand the concepts, methods of analysis, and fields of application. Concerning Uncertainty Quantification and the principal probability and statistical tools useful for probabilistic learning, we refer the reader to Section 2.4 devoted to bibliographical complements.

*Corresponding author: C. Soize, christian.soize@univ-eiffel.fr

Email address: christian.soize@univ-eiffel.fr (Christian Soize)

Published in Mathematics and Mechanics of Complex Systems, 11(1), 87-174 (2023). *October 24, 2023*

1.1. Organization of the paper

Sections 1 to 6 are devoted to the first aspect of proposed overview, namely the uncertainty quantification principally oriented towards multiscale mechanics of materials. Section 2 presents a very brief overview on the main concepts in stochastic modeling for Uncertainty Quantification (UQ) and ends with bibliographical complements. Section 3 is devoted to the construction of prior stochastic models of uncertainties. The types of representation for stochastic modeling are presented. We introduce the Maximum Entropy (MaxEnt) principle from Information Theory, which is a fundamental statistical tools to construct an informative prior probability distribution using the available information. We also show how MaxEnt can be used as a general numerical tool for constructing a probability distribution in any dimension. Section 4 deals with Random Matrix Theory that is an important probability tool for the probability modeling of uncertainties in computational mechanics. It is use for nonparametric probabilistic modeling of model errors in the computational models and also to construct prior probability models of tensor-valued random fields such as fourth-order elasticity fields of heterogeneous materials. We show how the ensembles of random matrices are constructed using MaxEnt for symmetric real random matrices. Three ensembles of random matrices that are used in uncertainty quantification are presented: ensemble SG_0^+ of positive-definite random matrices with a unit mean value; ensemble $\mathrm{SG}_\varepsilon^+$ of positive-definite random matrices with a unit mean value and a positive-definite lower bound; ensemble $\mathrm{SE}_\varepsilon^+$ of positive-definite random matrices with a given mean value and a positive-definite lower bound. Section 5 is central to construct informative prior probability models of elasticity random fields of elastic media. We present algebraic prior probability models for heterogeneous anisotropic elastic media and for heterogeneous elastic media with statistical fluctuations in a symmetry class and with anisotropic statistical fluctuations. In Section 6, two illustrations are presented. The first illustration concerns stochastic homogenization of heterogeneous materials at microscale in the context of a probabilistic analysis of the representative volume element size in stochastic homogenization of a heterogeneous complex microstructure. The second illustration deals with the stochastic continuum modeling of the random interphase of a polymer nanocomposite using atomistic simulations and a statistical inverse problem.

Sections 7 and 8 are devoted to the second aspect of the proposed overview and concerns the use of probabilistic learning in computational science and engineering, in particular in multiscale mechanics of materials. In Section 7, we present the methodology and the algorithms of the Probabilistic Learning on Manifolds (PLoM). This is a machine learning tool that has specifically been developed for small datasets. Some illustrations are given and show the capability of the PLoM algorithm. In particular we present the nonconvex optimization under uncertainties for which the number of function evaluations is limited. Next we summarize the methodology and the algorithms of the probabilistic learning under constraints based on the Kullback-Leibler divergence minimization principle. This methodology allows the probabilistic inference to be carried out to identify the probability model using targets. Finally, Section 8 is devoted to an illustration of the probabilistic learning inference for 3D stochastic homogenization of a heterogeneous linear elastic microstructure with random spectrum and without scale separation.

1.2. Convention for the variables, vectors, and matrices

x, η: lower-case Latin or Greek letters are deterministic real variables.
$\mathbf{x}, \boldsymbol{\eta}$: boldface lower-case Latin or Greek letters are deterministic vectors.
X: upper-case Latin letters are real-valued random variables.
$\mathbf{X}$: boldface upper-case Latin letters are vector-valued random variables.
$[x]$: lower-case Latin letters between brackets are deterministic matrices.
$[\mathbf{X}]$: boldface upper-case letters between brackets are matrix-valued random variables.

1.3. Notations for sets of matrices and their norm

$\mathbb{R}^n$: Euclidean vector space.
$\mathbb{M}_{n,m}$: set of the $(n \times m)$ real matrices.
$\mathbb{M}_n$: set of the square $(n \times n)$ real matrices.
$\mathbb{M}_n^S$: set of the symmetric square $(n \times n)$ real matrices.
$\mathbb{M}_n^+$: set of the positive-definite $(n \times n)$ real matrices.
$\mathbb{M}_n^{+0}$: set of the positive $(n \times n)$ real matrices.
$[I_n]$: identity matrix in $\mathbb{M}_n$.

$\mathbf{x}$: for $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} = (x_1, \ldots, x_m)$.
$\mathbf{x}^T$: transpose of the column matrix made up of the components of $\mathbf{x} \in \mathbb{R}^m$.
$[x]^T$: transpose of matrix $[x]$.
tr$[x]$: trace of matrix $[x]$.
$\langle \mathbf{x}, \mathbf{y} \rangle$: Euclidean inner product of $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, equal to $x_1 y_1 + \ldots x_m y_m$.
$\|\mathbf{x}\|$: Euclidean norm of $\mathbf{x}$ in $\mathbb{R}^n$, equal to $\langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$.
$\| [x] \| = \sup_{\| \mathbf{y} \| = 1} \| [x]\, \mathbf{y} \|$.
$\| [x] \|_F$: Frobenius norm of matrix $[x]$, equal to $(\mathrm{tr}\{[x]^T [x]\})^{1/2}$.
1_B: indicatrix function of set B.

1.4. Convention used for random variables and random fields

(i) Random variable. In this paper, for any finite integer $m \geq 1$, the Euclidean space $\mathbb{R}^m$ is equipped with the σ-algebra $\mathscr{B}_{\mathbb{R}^m}$. If $\mathbf{Y}$ is a $\mathbb{R}^m$-valued random variable defined on the probability space $(\Theta, \mathscr{T}, \mathscr{P})$, $\mathbf{Y}$ is a mapping $\theta \mapsto \mathbf{Y}(\theta)$ from Θ into $\mathbb{R}^m$, measurable from $(\Theta, \mathscr{T})$ into $(\mathbb{R}^m, \mathscr{B}_{\mathbb{R}^m})$, and $\mathbf{Y}(\theta)$ is a realization (sample) of $\mathbf{Y}$ for $\theta \in \Theta$. The probability distribution of $\mathbf{Y}$ is the probability measure $P_{\mathbf{Y}}(d\mathbf{y})$ on the measurable set $(\mathbb{R}^m, \mathscr{B}_{\mathbb{R}^m})$ (we will simply say on $\mathbb{R}^m$). The Lebesgue measure on $\mathbb{R}^m$ is noted $d\mathbf{y}$ and when $P_{\mathbf{Y}}(d\mathbf{y})$ is written as $p_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}$, $p_{\mathbf{Y}}$ is the probability density function (pdf) on $\mathbb{R}^m$ of $P_{\mathbf{Y}}(d\mathbf{y})$ with respect to $d\mathbf{y}$.

If $\mathbf{Y}$ and $\mathbf{Y}'$ are $\mathbb{R}^m$- and $\mathbb{R}^{m'}$-valued random variables defined on $(\Theta, \mathscr{T}, \mathscr{P})$ and if $\mathbf{f}$ is a measurable function on $\mathbb{R}^m \times \mathbb{R}^{m'}$ with values in $\mathbb{R}^n$, which means that $\mathbf{Z} = \mathbf{f}(\mathbf{Y}, \mathbf{Y}')$ is a $\mathbb{R}^n$-valued random variable, then we have

$$E\{\mathbf{Z}\} = \int_{\mathbb{R}^n} \mathbf{z}\, P_{\mathbf{Z}}(d\mathbf{z}) = E\{\mathbf{f}(\mathbf{Y}, \mathbf{Y}')\} = \int_{\mathbb{R}^m \times \mathbb{R}^{m'}} \mathbf{f}(\mathbf{y}, \mathbf{y}')\, P_{\mathbf{Y},\mathbf{Y}'}(d\mathbf{y}, d\mathbf{y}')\,,$$

in which E is the mathematical expectation, $P_{\mathbf{Z}}(d\mathbf{z})$ is probability distribution on $\mathbb{R}^n$ of $\mathbf{Z}$, where $P_{\mathbf{Y},\mathbf{Y}'}(d\mathbf{y}, d\mathbf{y}')$ is the joint probability distribution of $\mathbb{R}^m \times \mathbb{R}^{m'}$ of $\mathbf{Y}$ and $\mathbf{Y}'$, and where $P_{\mathbf{Z}}$ is the image of $P_{\mathbf{Y},\mathbf{Y}'}$ by $\mathbf{f}$.

We say that $\mathbf{Y}$ is a second-order random variable if $E\{\|\mathbf{Y}\|^2\} = \int_{\mathbb{R}^m} \|\mathbf{y}\|^2\, P_{\mathbf{Y}}(d\mathbf{y}) < +\infty$. The set of all the second-order $\mathbb{R}^m$-random variables defined on $(\Theta, \mathscr{T}, \mathscr{P})$ is denoted by $L^2(\Theta, \mathbb{R}^m)$ (quotient by the equivalence relation of almost-surely (a.s.) equal random variables). Equipped with the inner product and the associated norm

$$\langle\langle \mathbf{Y}, \mathbf{Y}' \rangle\rangle = E\{\langle \mathbf{Y}, \mathbf{Y}' \rangle\} \quad , \quad |||\mathbf{Y}||| = \langle\langle \mathbf{Y}, \mathbf{Y} \rangle\rangle^{1/2}\,,$$

$L^2(\Theta, \mathbb{R}^m)$ is a Hilbert space.

(ii) Second-order random field. A second-order random field $\{\mathbf{V}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by any uncountable subset Ω of $\mathbb{R}^d$ with $d \geq 1$ (possibly with $\Omega = \mathbb{R}^d$), with values in $\mathbb{R}^m$, is defined by a mapping $\boldsymbol{\xi} \mapsto \mathbf{V}(\boldsymbol{\xi})$ from Ω into $L^2(\Theta, \mathbb{R}^m)$.

For all $\boldsymbol{\xi}$ fixed in Ω, $\mathbf{V}(\boldsymbol{\xi}) \in L^2(\Theta, \mathbb{R}^m)$, that is to say is the second-order random variable $\theta \mapsto \mathbf{V}(\boldsymbol{\xi}, \theta)$ from Θ into $\mathbb{R}^m$. For all θ fixed in Θ, $\mathbf{V}(\boldsymbol{\xi}, \theta)$ is a realization of $\mathbb{R}^m$-valued random variable $\mathbf{V}(\boldsymbol{\xi})$ and $\boldsymbol{\xi} \mapsto \mathbf{V}(\boldsymbol{\xi}, \theta)$ is a trajectory (or sample path) of random field $\mathbf{V}$.

Let $o_\nu = \{\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^\nu\}$ be a finite subset of Ω and let $\mathbb{V}_{o_\nu} = (\mathbf{V}(\boldsymbol{\xi}^1), \ldots, \mathbf{V}(\boldsymbol{\xi}^\nu))$ be the random variable with values in $\mathbb{R}^n$ with $n = \nu \times m$. The probability distribution $P_n(d\mathbb{v})$ on $\mathbb{R}^n$ of random variable $\mathbb{V}_{o_\nu}$ is a marginal probability distribution of random field $\mathbf{V}$. The system of marginal distributions of random field $\mathbf{V}$ is made up of the uncountable family of all the finite probability distributions $P_n(d\mathbb{v})$ obtained for all the possible finite nonempty and nonordered subsets o_ν of Ω.

2. Main concepts in stochastic modeling for Uncertainty Quantification (UQ)

This very brief section is relatively important for the understanding of what is UQ and we refer the reader to [1] for more details.

2.1. Aleatory and epistemic uncertainties

The *aleatory uncertainties* concern physical phenomena, which are random by nature. This is the case, for instance, for the pressure field in a fully developed turbulent boundary layer or for the geometrical distribution of the inclusions inside the matrix of a biphasic microstructure. The *epistemic uncertainties* concern a lack of knowledge about the parameters of a computational model, which is the discretization of a boundary value problem (BVP). They also concern modeling errors, which cannot be described using the parameters of the computational model. This is the case, for example, of uncertainties due to the lack of knowledge of the mechanical description of a boundary condition, or to geometric tolerances induced by manufacturing processes, or even to the effective mechanical properties of materials resulting from a change of scale (without scale separation). These modeling errors are also due to the existence of hidden degrees of freedom, which are not taken into account and which are related to the use of reduced kinematics in the computational model (beam theory instead of 3D elasticity) or which are associated with secondary dynamic subsystems not modeled in the computational model.

In the framework of the *probability theory* and *mathematical statistics* of uncertainty quantification, there is no need to distinguish these two types of uncertainty (and we simply say "uncertainties") because the tools are exactly the same for the stochastic modeling of uncertainties, for the propagation of uncertainties in the computational model, and for the identification by solving a statistical inverse problem.

2.2. Sources of uncertainties and variabilities

We consider a *designed mechanical system* (conceptual system), its *computational model*, and the *real mechanical system*. The computational model is constructed by using the finite element approximation of the BVP that is the mathematical-mechanical model of the designed mechanical system. This BVP is posed on an open bounded domain Ω of $\mathbb{R}^3$, with boundary $\partial\Omega = \Gamma_0 \cup \Gamma_1$. The manufacturing process of the designed mechanical system yields the *real mechanical system*. The BVP depends on a control parameter $\mathbf{w}$ (such as mass density, geometrical and mechanical parameters), on an uncontrolled parameter $\mathbf{u}$ (such as the tensor-valued elasticity field of a material), and depends on the mathematical modeling of the environment of the real mechanical system yielding nonhomogeneous term in Ω (external body force field), Neuman boundary condition on Γ_1 (surface force field), and Dirichlet condition on Γ_0.

In general, the errors related to the construction of an approximate solution of the computational model have to be reduced and controlled, and should not be considered as uncertainties. There are three sources of uncertainties: (i) the uncertainties on $\mathbf{w}$ and $\mathbf{u}$ that are defined as the *model-parameter uncertainties*, (ii) the *model uncertainties* induced by modeling errors during the mathematical-mechanical modeling process (which cannot be taken into account by the model-parameter uncertainties), and (iii) the uncertainties in the real mechanical system due to its *variabilities* induced by the manufacturing process and due to small differences in the configurations (an experimental configuration of a complex real mechanical system differs from the designed mechanical system and is never perfectly known).

2.3. Challenges and role played by probability theory and mathematical statistics

The deterministic model of a complex real mechanical system is generally not sufficient in many cases. The robustness of the computational model must be improved in taking into account its own uncertainties, the variabilities of the real mechanical system, and the *experimental errors*. For complex real mechanical systems, the experimental errors are mainly due to the lack of knowledge of the experimental configuration that differs from the designed mechanical system and that is not perfectly known. These experimental errors are also due to measurement noise but today it is generally negligible compared to the previous one.

The probability theory is a powerful mathematical tool, which allows (i) to construct, in finite or infinite dimension, a *prior stochastic model* of uncertainties (random vectors, matrices or tensors, random fields, etc.), (ii) to analyze the *propagation* of uncertainties in the computational model using applied mathematics, (iii) to *identify* the *prior* and the *posterior stochastic models* of uncertainties, in finite or infinite dimension, using available data (experimental data) and mathematical statistics to solve *statistical inverse problems* with, if necessary, the help of *Machine Learning* tools in a probabilistic/statistics framework. The first challenge is related to the effective construction of stochastic models such as the construction of informative prior probability distributions using algebraic representations such as polynomial chaos expansions and random generators such as Markov Chain Monte Carlo (MCMC) generator for random vectors in high stochastic dimension, random matrices in any stochastic dimension, tensor-valued random fields in high stochastic dimension. This step is fundamental and is difficult enough. The second challenge is related to

model uncertainties induced by the modeling errors because this type of stochastic modeling is related to the operators of the BVP that are uncertain and that cannot be taken into account through the model-parameter uncertainties of the BVP. If partial and limited data are available, the third challenge is related to the identification and the updating of the stochastic models by solving statistical inverse problems using the *least-square method*, the *maximum likelihood* approach, or the *Bayesian inference*. In high stochastic dimension, what is generally the case for the identification of a random field that is an uncontrolled parameter of the BVP, this can be a very challenging problem. Finally, the fourth challenge is related to the algorithmic strategy for robust updating, robust optimization, and robust design. It is also a challenging problem with respect to the computational resources and consequently, the formulations require the use of reduced-order models, surrogate models (metamodels), and machine learning tools implemented in a statistical framework.

2.4. Bibliographical complements

Concerning the concepts, the methodologies, and the algorithms to implement Uncertainty Quantification (UQ) in science and engineering we refer the reader to [1, 2, 3, 4, 5, 6, 7]. The mathematical tools used for uncertainty quantification are principally based on the probability theory and stochastic processes [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18], and on mathematical statistics [19] with several important aspects, in computational statistics [20, 21], in multivariate statistics [22, 23], in nonparametric statistics with smoothing techniques [24, 25, 26, 27, 28, 29, 30, 31], and in approaches based on the maximum likelihood and the Bayesian methods to solve statistical inverse problems [32, 33]. It should be noted that the Markov Chain Monte Carlo (MCMC) methods and associated algorithms are important computational tools for sampling any probability distribution [34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44].

Uncertainty quantification requires to construct the probability models of uncertainties. Therefore, tools are necessary to construct informative prior probability distributions, to analyze statistical dependencies between random quantities, to estimate and formulate the "proximity" between probability distributions. These tools are given by Information Theory, including important notions such as entropy, cross-entropy, maximum entropy principle, Kullback-Leibler divergence [45, 46, 47, 48, 49, 50, 51, 52].

To implement UQ in stochastic boundary value problems and computational models, one needs to construct probability models for random quantities such that random matrices, stochastic processes, and random fields. In the area of engineering sciences, random matrix theory [53] was used to construct random finite representations of stochastic operators [1], in particular by introducing ensembles of positive-definite symmetric random matrices to construct nonparametric probabilistic models of the model errors in computational models, such as those initiated and proposed in [54, 55, 56, 57, 58, 59]. Random fields theory is necessary for the stochastic modeling of numerous boundary value problems with uncertainties [60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70] in particular, the construction of random fields with values in ensembles of positive symmetric tensors that allow the coefficients of elliptic partial differential operators to be modeled [71, 72, 73, 74]. The polynomial chaos expansion of stochastic processes was introduced in [75, 76] and an effective Karhunen-Loève-based construction for random fields was pioneered in [77, 78]. The Wiener-Askey polynomial chaos expansion was used by [79], the developments of random fields in polynomial chaos for arbitrary probability measure were introduced in [80], and a compressed principal component analysis of non-Gaussian vectors using symmetric polynomial chaos has been proposed in [81]. Polynomial chaos expansions in finite and in infinite dimension have been and is intensively used for uncertainties modeling and propagation of uncertainties [82, 83, 84, 85, 86, 87, 88] (see also hereinafter the stochastic solvers and the stochastic finite elements). It should be noted that the nonparametric modeling of model uncertainties induced by modeling errors in large computational models has notably been developed in linear dynamics [54, 55, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98], in vibroacoustics and fluid-structure interactions [92, 99, 100, 101, 102], in nonlinear dynamics [103, 104, 105, 106, 107, 108, 109, 110, 111], for models updating [112, 113], and for model errors with probabilistic learning [114, 115].

Once the probabilistic models of uncertainties are built, it is necessary to study the propagation of these uncertainties in the systems. It is therefore necessary to have methods for solving stochastic equations. The first set of methods is based on the Monte Carlo numerical simulation methods [34, 116, 117, 118, 119, 120, 121, 122]. The second set is based on spectral projection methods [123, 124, 125, 126, 127, 128], such as those based on polynomial chaos expansions [77, 78, 3, 129] and called stochastic finite element method when the discretization method of the boundary value problems are performed using the finite element method [78, 130, 131, 132, 133, 134, 2, 135, 136, 137, 138, 139, 140, 141].

Most often, the uncertainties probability model is a prior model. If targets are available for observations of the system, coming from experimental measurements or from more precise numerical simulations, a posterior probability model of uncertainties can be estimated by solving inverse statistical problems based on the maximum likelihood, the Bayesian inference, and machine learning. For general overviews on statistical inverse methods, see [142, 143, 144, 145, 146, 147, 148, 149], and for complements, see [150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160]. The statistical identification of the coefficients of polynomial chaos representations of random fields can be found in [161, 162, 163], in particular in [164, 165, 166] for high dimension, and for representations of random vectors in [167, 168, 169, 170, 171, 172, 173]. The inverse identification of random matrices have been proposed in [174, 175, 176]. Statistical inverse methods are also used to perform model updating [177, 178, 179, 180], model selection [181, 182], and to construct surrogate models (or metamodels) [183, 184, 185, 186, 187, 188, 189].

Within the framework of the methodologies described above to implement UQ, an important question concerns the optimization problems under uncertainties (OUU) such as the stochastic optimization [190], the computational methods devoted to optimization and robust design [191, 192, 193, 100], the multidisciplinary optimization [194, 195, 196, 197], the multicriteria and multiobjective optimization [198, 199, 200], and the multiscale optimization [201, 202]. The OUU problems often require the use of stochastic expansion [203], the use of surrogate models [204, 205, 206], and those of machine learning and probabilistic learning algorithms [207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220].

Finally, the machine learning tools and artificial intelligence [221, 222, 223, 224], such as the probabilistic and statistical learning [225, 226, 227, 228], are used in UQ for problems that would require computer resources that are not available with the most usual approaches.Thus methods have emerged in the field of engineering sciences, such as the learning on manifolds [229, 230, 231, 232, 233, 234, 235, 236, 237, 238] and the physics-informed probabilistic learning [239, 240, 241, 242, 243].

3. Prior stochastic modeling of uncertainties

3.1. Types of representation for stochastic modeling

Let us consider a computational model for which the control parameter $\mathbf{w}$ is a vector in $\mathbb{R}^{n_w}$. The quantity of interest (QoI), $\mathbf{q} \in \mathbb{R}^{n_q}$, is a system observation that is written as $\mathbf{q} = \mathbf{h}(\mathbf{w})$ in which $\mathbf{h}$ is a measurable mapping from $\mathbb{R}^{n_w}$ into $\mathbb{R}^{n_q}$. This mapping is implicit and is defined through the solution of the BVP whose approximation is constructed via the computational model. Presently, there is no uncontrolled parameter and therefore $\mathbf{h}$ is a deterministic mapping. Control parameter $\mathbf{w}$ is assumed to be uncertain and is modeled by a $\mathbb{R}^{n_w}$-random variable $\mathbf{W}$ defined on a probability space $(\Theta, \mathscr{T}, \mathscr{P})$, with probability distribution $P_{\mathbf{W}}(d\mathbf{w}; \mathbf{s})$ whose support is the set $\mathscr{C}_w \subset \mathbb{R}^{n_w}$ and where $\mathbf{s}$ is a hyperparameter that belongs to an admissible set $\mathscr{C}_s \subset \mathbb{R}^{n_s}$. Consequently, QoI $\mathbf{q}$ is modeled by a $\mathbb{R}^{n_q}$-valued random variable $\mathbf{Q} = \mathbf{h}(\mathbf{W})$ whose probability distribution $P_{\mathbf{Q}}(d\mathbf{q}; \mathbf{s})$ is the image (transport) of $P_{\mathbf{W}}$ by $\mathbf{h}$. It can then be seen the existence of two main steps. The first step is the stochastic modeling of uncertainties, that is to say, the construction of a *prior probability distribution* $P_{\mathbf{W}}(d\mathbf{w}; \mathbf{s})$ of uncertain parameter $\mathbf{W}$; this is the purpose of this Section 3. The second step is the analysis of the propagation of uncertainties, that is to say, the construction of $P_{\mathbf{Q}}(d\mathbf{q}; \mathbf{s})$ that is carried out by using an adapted stochastic solver (as explained in Section 2.4). This methodological description clearly shows the impact of the use of an arbitrary stochastic modeling of $\mathbf{W}$. Even if the stochastic solver is "perfect" (estimation of $P_{\mathbf{Q}}(d\mathbf{q}; \mathbf{s})$), if $P_{\mathbf{W}}(d\mathbf{w}; \mathbf{s})$ is arbitrarily chosen (thus probably wrong, such that a Gaussian distribution for a positive-valued random variable), its nonlinear transformation by $\mathbf{h}$ will yield a probability distribution $P_{\mathbf{Q}}(d\mathbf{q}; \mathbf{s})$ perfectly estimated, but which will be arbitrary (thus probably wrong). This means that the first step is absolutely fundamental in UQ.

For stochastic modeling of uncertainties, two cases have to be considered. The case for which there is no available experimental data and the one for which experimental data are available. However, even if experimental data are not available, it is important to consider UQ in order to make robust analyzes with respect to uncertainties. This enables robust predictions with the computational model as well as robust optimization and design. If *big data* are available, then *nonparametric statistics* can be used to estimate $P_{\mathbf{W}}(d\mathbf{w}; \mathbf{s})$. If only *small/limited data* are available, then an *informative prior* probability distribution $P^{\text{prior}}_{\mathbf{W}}(d\mathbf{w}; \mathbf{s})$ of $\mathbf{W}$ has to be constructed. *Parametric statistics* such as the least-square and maximum likelihood methods have to be used to estimate $\mathbf{s}$ or, the Bayesian inference can be used to construct a *posterior* probability distribution $P^{\text{post}}_{\mathbf{W}}(d\mathbf{w})$ of $\mathbf{W}$ (note that, for high stochastic dimension, a high quality of the informative prior probability distribution is needed to make the Bayesian inference approach feasible).

Two fundamental approaches can be considered to construct a probability distribution $P_{\mathbf{W}}$ of uncertain parameter $\mathbf{W}$: the direct and the indirect approaches.

Direct approach. (i) If a large amount of data is available such as n_d independent realizations of $\mathbf{W}$ with a large value of n_d, then the probability distribution $P_{\mathbf{W}}$ of $\mathbf{W}$ can directly be estimated using the Kernel Density Estimation (KDE) method from the nonparametric statistics. (ii) If there is no data or if only a small amount of data is available (small value of n_d), then the construction of an informative prior probability distribution $P_{\mathbf{W}}^{\text{prior}}(d\mathbf{w}; \mathbf{s}) = p_{\mathbf{W}}^{\text{prior}}(\mathbf{w}; \mathbf{s})\, d\mathbf{w}$, represented by a pdf $p_{\mathbf{W}}^{\text{prior}}$ on $\mathbb{R}^{n_w}$, can be constructed by using the *Maximum Entropy* principle (MaxEnt) from Information Theory. In such a case, $\mathbf{s}$ appears as a hyperparameter that is related to the available information introduced as constraints in the MaxEnt principle. Often, $\mathbf{s}$ describes some statistical properties of uncertainties, such as the level of uncertainties, and a sensitivity analysis can be performed with respect to the values of $\mathbf{s}$. If a small dataset is available then hyperparameter $\mathbf{s}$ can be estimated by solving a statistical inverse problem.

Indirect approach. It consists in introducing a representation $\mathbf{W} = \mathbf{f}(\mathbf{G}; \mathbf{s})$ in which $\mathbf{g} \mapsto \mathbf{f}(\mathbf{g}; \mathbf{s})$ is a deterministic nonlinear (measurable) mapping, where $\mathbf{G}$ is a given $\mathbb{R}^{n_g}$-valued random variable whose pdf $p_{\mathbf{G}}(\mathbf{g})$ with respect to $d\mathbf{g}$ is given on $\mathbb{R}^{n_g}$, and where $\mathbf{s}$ is the hyperparameter. Therefore, $P_{\mathbf{W}}^{\text{prior}}(d\mathbf{w}; \mathbf{s})$ is the image of $p_{\mathbf{G}}(\mathbf{g})\, d\mathbf{g}$ by mapping $\mathbf{f}(.; \mathbf{s})$. Two main representations can be used for construction mapping $\mathbf{f}$. The first one is the use of a truncated *Polynomial Chaos Expansion* (PCE) that is written as $\mathbf{W} = \sum_{k=0}^{K} \mathbf{a}^k\, \Psi_{\boldsymbol{\alpha}^{(k)}}(\mathbf{G})$ in which $\boldsymbol{\alpha}^{(k)}$ is a value of the multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{n_g}) \in \mathbb{N}^{n_g}$, where $\{\Psi_{\boldsymbol{\alpha}}(\mathbf{g}), \boldsymbol{\alpha} \in \mathbb{N}^{n_g}\}$ is the family of orthogonal polynomials with respect to the measure $p_{\mathbf{G}}(\mathbf{g})\, d\mathbf{g}$, that is to say $\int_{\mathbb{R}^{n_g}} \Psi_{\boldsymbol{\alpha}}(\mathbf{g})\, \Psi_{\boldsymbol{\beta}}(\mathbf{g})\, p_{\mathbf{G}}(\mathbf{g})\, d\mathbf{g} = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$, and where $\mathbf{s} = \{\mathbf{a}^0, \ldots, \mathbf{a}^K\}$. The second one consists in constructing a *prior algebraic representation*, $\mathbf{W} = \mathbf{f}(\mathbf{G}; \mathbf{s})$ of $\mathbf{W}$ in which hyperparameter $\mathbf{s}$ has a small dimension n_s and where $\mathbf{g} \mapsto \mathbf{f}(\mathbf{g}; \mathbf{s})$ is a given nonlinear mapping (such an approach will be used in Section 5 for constructing a prior probability model of tensor-valued random fields).

3.2. Maximum Entropy principle from Information Theory as a direct approach to construct an informative prior probability distribution

This statistical tool is presented for a random vector but can be applied to any random quantities (for instance to random matrices, see Section 4).

(i) Entropy as a measure of uncertainties for a vector-valued random variable. Let $\mathbf{W}$ be the $\mathbb{R}^{n_w}$-valued random variable, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, whose probability distribution is written as $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w}$ in which $p_{\mathbf{W}}$ is an unknown pdf on $\mathbb{R}^{n_w}$ but for which its support is known and such that *supp* $p_{\mathbf{W}} = \mathscr{C}_w \subset \mathbb{R}^{n_w}$. The Shannon entropy [45] (see also [49, 50, 51]) of $p_{\mathbf{W}}$, measuring uncertainty, is the real number defined by

$$\mathscr{E}(p_{\mathbf{W}}) = -\int_{\mathbb{R}^{n_w}} p_{\mathbf{W}}(\mathbf{w})\, \log(p_{\mathbf{W}}(\mathbf{w}))\, d\mathbf{w} = -E\{\log(p_{\mathbf{W}}(\mathbf{W}))\} \in \mathbb{R}\,. \tag{3.1}$$

For a uniform pdf on a compact set $\kappa \subseteq \mathscr{C}_w \subset \mathbb{R}^{n_w}$, that is to say $p_{\mathbf{W}}(\mathbf{w}) = (1/|\kappa|)\, 1_{\kappa}(\mathbf{w})$, we have $\mathscr{E}(p_{\mathbf{W}}) = \log|\kappa|$. More the Shannon entropy is large, and more uncertainty is high. The limit, $\mathscr{E} = -\infty$, corresponds to no uncertainty (deterministic case).

(ii) Maximum entropy principle (MaxEnt). Introduced by Jaynes [47, 48], the MaxEnt consists in constructing the pdf $p_{\mathbf{W}}$ that corresponds to the largest uncertainty on the set of all the possible pdfs on $\mathbb{R}^{n_w}$ with given support $\mathscr{C}_w$, which satisfy the constraints defined by the available information. In addition to the support property, the available information is made up of statistical properties on $\mathbf{W}$ defined as a mathematical expectation,

$$E\{\mathbf{g}^c(\mathbf{W})\} = \int_{\mathbb{R}^{n_w}} \mathbf{g}^c(\mathbf{w})\, p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w} \;=\; \mathbf{b}^c \;\in\; \mathbb{R}^{n_c}\,, \tag{3.2}$$

in which $n_c \geq 1$ is a finite integer, where $\mathbf{w} \mapsto \mathbf{g}^c(\mathbf{w})$ is a given (measurable) mapping from $\mathbb{R}^{n_w}$ into $\mathbb{R}^{n_c}$, and where $\mathbf{b}^c$ is a given vector in $\mathbb{R}^{n_c}$. The definition of MaxEnt allows the following optimization problem to be defined for constructing $p_{\mathbf{W}}$,

$$p_{\mathbf{W}} = \arg \max_{p \in \mathscr{C}_{\text{ad}}} \mathscr{E}(p)\,, \tag{3.3}$$

in which the admissible set $\mathscr{C}_{\text{ad}}$ is the subset of $\mathscr{C}_{\text{free}} = \{p \in L^1(\mathbb{R}^{n_w}, \mathbb{R}^+)\,,\ supp\ p = \mathscr{C}_w\}$ such that

$$\mathscr{C}_{\text{ad}} = \{p \in \mathscr{C}_{\text{free}}\,,\ \int_{\mathbb{R}^{n_w}} p(\mathbf{w})\,d\mathbf{w} = 1\,,\ \int_{\mathbb{R}^{n_w}} \mathbf{g}^c(\mathbf{w})\,p(\mathbf{w})\,d\mathbf{w} \ = \ \mathbf{b}^c\}\,. \tag{3.4}$$

Existence and uniqueness of this optimization problem has to be studied as a function of the defined information represented by $\mathbf{g}^c$ and $\mathbf{b}^c$ (see paragraph (iv)).

(iii) Reformulation of the optimization problem by using Lagrange's multipliers. To solve the optimization problem defined by Eq. (3.3), the problem is reformulated on set $\mathscr{C}_{\text{free}}$ instead of $\mathscr{C}_{\text{ad}}$ by introducing the Lagrange multipliers $\lambda_0 \in \mathbb{R}^+$ associated with the normalization condition $\int_{\mathbb{R}^{n_w}} p(\mathbf{w})\,d\mathbf{w} - 1 = 0$ and $\boldsymbol{\lambda} \in \mathscr{C}_\lambda \subset \mathbb{R}^{n_c}$ associated with the constraint $\int_{\mathbb{R}^{n_w}} \mathbf{g}^c(\mathbf{w})\,p(\mathbf{w})\,d\mathbf{w} - \mathbf{b}^c = \mathbf{0}$ in which $\mathscr{C}_\lambda$ is the admissible set for $\boldsymbol{\lambda}$, defined by

$$\mathscr{C}_\lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^{n_c}\ ,\ \int_{\mathbb{R}^{n_c}} \exp(-\langle \boldsymbol{\lambda}\,, \mathbf{g}^c(\mathbf{w})\rangle)\,d\mathbf{w} < +\infty\}\,. \tag{3.5}$$

The Lagrangian is defined, for all λ_0 in $\mathbb{R}^+$, $\boldsymbol{\lambda}$ in $\mathscr{C}_\lambda$, and p in $\mathscr{C}_{\text{free}}$, by

$$Lag(p; \lambda_0, \boldsymbol{\lambda}) = \mathscr{E}(p) - (\lambda_0 - 1)(\int_{\mathbb{R}^{n_w}} p(\mathbf{w})\,d\mathbf{w} - 1) - \langle \boldsymbol{\lambda}, \int_{\mathbb{R}^{n_w}} \mathbf{g}^c(\mathbf{w}) p(\mathbf{w})\,d\mathbf{w} - \mathbf{b}^c\rangle\,. \tag{3.6}$$

(iv) Existence and uniqueness of MaxEnt. If the constraints are algebraically independent, then there exists a unique solution [244, 1], which corresponds to the stationary point of the Lagrangian and which is written as

$$p_{\mathbf{W}}(\mathbf{w}) = 1_{\mathscr{C}_w}(\mathbf{w})\,c_0^{\text{sol}}\,\exp(-\langle \boldsymbol{\lambda}^{\text{sol}},\, \mathbf{g}^c(\mathbf{w})\rangle) \quad , \quad \forall \mathbf{w} \in \mathbb{R}^{n_w}\,, \tag{3.7}$$

in which the constant of normalization $c_0^{\text{sol}} = \exp(-\lambda_0^{\text{sol}})$ can be written as $c_0^{\text{sol}} = (\int_{\mathscr{C}_w} \exp(-\langle \boldsymbol{\lambda}^{\text{sol}}, \mathbf{g}^c(\mathbf{w})\rangle)\,d\mathbf{w})^{-1}$ and where $(\lambda_0^{\text{sol}}, \boldsymbol{\lambda}^{\text{sol}}) \in \mathbb{R}^+ \times \mathscr{C}_\lambda$ is the unique solution in $(\lambda_0, \boldsymbol{\lambda})$ of the following equations,

$$\int_{\mathscr{C}_w} \exp(-\lambda_0 - \langle \boldsymbol{\lambda}\,, \mathbf{g}^c(\mathbf{w})\rangle)\,d\mathbf{w} = 1 \quad , \quad \int_{\mathscr{C}_w} \mathbf{g}^c(\mathbf{w})\,\exp(-\lambda_0 - \langle \boldsymbol{\lambda}\,, \mathbf{g}^c(\mathbf{w})\rangle)\,d\mathbf{w} = \mathbf{b}^c\,. \tag{3.8}$$

(v) Analytical examples of classical probability distributions deduced from MaxEnt and a few properties. Appendix A gives explicit algebraic expressions of classical probability density functions, obtained using MaxEnt. These are Uniform, Gamma, and Gaussian distributions for a real-valued random variable, and multivariate Exponential and Gaussian distributions for a random vector. What is interesting is to see the available information that leads to each of these distributions and also to see some properties of these random variables.

3.3. MaxEnt as a numerical tool for constructing a probability distribution in any dimension

Such a numerical tool is necessary, because in general, Eq. (3.8) cannot explicitly be solved. An adapted numerical algorithm must be used to circumvent the difficulties that appear in high dimension. The constant of normalization $c_0 = \exp(-\lambda_0)$ goes to zero as R^{-n_w} in which R is a "radius" of $\mathscr{C}_w$, which is directly involved in the numerical calculations, what cannot be computed in high dimension. In addition, integrals in high dimension have to be computed for calculating the Lagrange multipliers (see Eq. (3.8)). An algorithm to compute the Lagrange multipliers and to construct a generator of realizations of $\mathbf{W}$ are given in Appendix B.

4. Random Matrix Theory for uncertainty quantification in computational mechanics

The random matrix theory was intensively studied by physicists and mathematicians in the context of nuclear physics. The developments began with Wigner in the 1950s, important efforts were performed in the 1960s by Wigner (1962), Dyson (1962), Mehta and others. In 1967 Mehta published an excellent book concerning a synthesis of the random matrix theory [53] (second edition in 1991). For physical applications, an important ensemble is the Gaussian Orthogonal Ensemble (GOE) such that any random matrix in GOE is a real symmetric random matrix whose entries

are mutually independent, and which is invariant under orthogonal linear transformations. Applying MaxEnt with this information yields a probability distribution of a $\mathbb{M}_n^S$-valued random matrix with respect to the volume element $d^S G$ on Euclidean space $\mathbb{M}_n^S(\mathbb{R})$ (see Section 4.1), which is written as

$$p_{[\mathbf{G}]}([G]) = c_G \, \exp(-\frac{n+1}{4\delta^2} \, \mathrm{tr}\{[G]^2\}) \quad , \quad G_{kj} = G_{jk} \ , \ 1 \le j \le k \le n \, , \tag{4.1}$$

in which c_G is the constant of normalization and where δ is a hyperparameter defined in [1] (pp.98-99).

The random matrix theory is an important tool to construct prior probability distributions of system-parameter uncertainties and model uncertainties induced by modeling errors in computational mechanics. The GOE can be viewed as a generalization of the Gaussian real random variables to the Gaussian symmetric real random matrices. A random matrix $\mathbf{G}$ whose pdf is given by Eq. (4.1) is not positive almost surely or negative almost surely (no signature), and in addition, $E\{\|[\mathbf{G}]^{-1}\|_F^2\} = +\infty$. Consequently, this ensemble cannot be used for stochastic modeling of a symmetric real matrix for which a positiveness property and an integrability of its inverse are required. Such a situation is similar to the one that we have presented in Section Appendix A.1-(iii) or Appendix A.2-(ii). Consequently, new ensembles of random matrices are necessary to develop uncertainty quantification in computational mechanics. A first ensemble of positive-definite random matrices has been introduced by the author [54] when he proposed the novel nonparametric probabilistic approach of model uncertainties induced by the modeling errors in computational dynamics, and later, other ensembles of random matrices derived from the first ensemble to take into account different types of random operators [55, 56, 90, 57, 104, 245, 99, 94] encountered in computational mechanics, computational fluid mechanics, fluid-structure interaction including vibroacoustics. For complete developments on this subject, we refer the reader to [1, 59]. Below, we will limit the presentation to the fundamental ensemble SG_0^+ of positive-definite random matrices, which are the "germs" of two other useful ensembles, $\mathrm{SG}_\varepsilon^+$ and $\mathrm{SE}_\varepsilon^+$, that we will be used for constructing the prior probability model of tensor-valued random elasticity fields (see Section 5).

4.1. Prerequisites to construct ensembles of random matrices by MaxEnt

The Euclidean space $\mathbb{M}_n$ is equipped with the inner product $\ll [G] \, , [H] \gg = \mathrm{tr}\{[G]^T [H]\}$ and with the associated Frobenius norm $\|G\|_F = \ll [G] \, , [G] \gg^{1/2}$. Induced by the inner product, the *volume element* on Euclidean space $\mathbb{M}_n$ is $dG = \prod_{j,k=1}^n dG_{jk}$ and on Euclidean space $\mathbb{M}_n^S$ is $d^S G = 2^{n(n-1)/4} \prod_{1 \le j \le k \le n} dG_{jk}$. Let $[\mathbf{G}]$ be a $\mathbb{M}_n^S$-valued random matrix defined on $(\Theta, \mathscr{T}, \mathscr{P})$ whose probability distribution $P_{[\mathbf{G}]} = p_{[\mathbf{G}]}([G]) \, d^S G$ is defined by the pdf $[G] \mapsto p_{[\mathbf{G}]}([G])$ from $\mathbb{M}_n^S(\mathbb{R})$ into $\mathbb{R}^+ = [0 \, , +\infty[$ with respect to $d^S G$ on $\mathbb{M}_n^S$. Then this pdf satisfies the normalization condition,

$$\int_{\mathbb{M}_n^S} p_{[\mathbf{G}]}([G]) \, d^S G = 1 \, . \tag{4.2}$$

The support of the probability density function, *supp* $p_{[\mathbf{G}]}$ of pdf $p_{[\mathbf{G}]}$ is any subset $\mathscr{S}_n$ of $\mathbb{M}_n^S$, possibly with $\mathscr{S}_n = \mathbb{M}_n^S$ (example: $\mathscr{S}_n = \mathbb{M}_n^+ \subset \mathbb{M}_n^S$).

4.2. The Shannon entropy as a measure of uncertainties for a symmetric real random matrix and MaxEnt principle

Similarly to the vector case presented in Section 3.2, the Shannon entropy of $p_{[\mathbf{G}]}$, which measures the level of uncertainty, is the real number,

$$\mathscr{E}(p_{[\mathbf{G}]}) = -\int_{\mathscr{S}_n} p_{[\mathbf{G}]}([G]) \, \log\big(p_{[\mathbf{G}]}([G])\big) \, d^S G = -E\{\log\big(p_{[\mathbf{G}]}([\mathbf{G}])\big)\} \, . \tag{4.3}$$

For this random matrix case, MaxEnt is then applied similarly to the random vector case.

4.3. Ensemble SG_0^+ of positive-definite random matrices with a unit mean value

This is a fundamental ensemble [1, 54] that is used to construct many derived ensembles of random matrices (see Sections 4.4 and 4.5) and which is important for the nonparametric probabilistic approach of model uncertainties and for the construction of prior probability model of elasticity random fields (see Section 5).

(i) Definition of SG_0^+ by the MaxEnt principle and pdf. The set SG_0^+ is the set of all the random matrices $[\mathbf{G}_0]$ with values in $\mathbb{M}_n^+$, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, and constructed using MaxEnt with the following available information,

$$E\{[\mathbf{G}_0]\} = [I_n] \quad , \quad E\{\log(\det[\mathbf{G}_0])\} = \nu_{G_0} \quad , \quad |\nu_{G_0}| < +\infty \, . \tag{4.4}$$

The support $\mathscr{S}_n$ of the pdf $[G] \mapsto p_{[\mathbf{G}_0]}([G]) : \mathbb{M}_n^S \to \mathbb{R}^+$ with respect to the $d^S G$ is $\mathscr{S}_n = \mathbb{M}_n^+ \subset \mathbb{M}_n^S$ and we have

$$p_{[\mathbf{G}_0]}([G]) = 1_{\mathscr{S}_n}([G])\, c_{G_0} \left(\det [G]\right)^{(n+1)\frac{(1-\delta^2)}{2\delta^2}} \exp(-\frac{n+1}{2\delta^2} \operatorname{tr}[G]) \, . \tag{4.5}$$

The hyperparameter δ, which is such that $0 < \delta < (n+1)^{1/2}(n+5)^{-1/2}$, allows for controlling the level of statistical fluctuations of $[\mathbf{G}_0]$, and is such that

$$\delta = \left\{ \frac{E\{\| [\mathbf{G}_0] - E\{[\mathbf{G}_0]\} \|_F^2\}}{\| E\{[\mathbf{G}_0]\} \|_F^2} \right\}^{1/2} = \left\{ \frac{1}{n} E\{\| [\mathbf{G}_0] - [I_n] \|_F^2\} \right\}^{1/2} \, . \tag{4.6}$$

It should be noted that constant ν_{G_0} has been eliminated for the benefit of δ in order to perform a reparameterization in δ instead of keeping ν_{G_0} that has no physical meaning.

(ii) Invertibility and convergence property when dimension goes to infinity. It is proven in [54] that

$$E\{\|[\mathbf{G}_0]^{-1}\|^2\} \leq E\{\|[\mathbf{G}_0]^{-1}\|_F^2\} < +\infty \, , \tag{4.7}$$

and in [55] that $\forall n \geq 2$, $E\{\|[\mathbf{G}_0]^{-1}\|^2\} \leq c_\delta < +\infty$ in which c_δ is a positive finite constant that is independent of n but that depends on hyperparameter δ. This invertibility property, which is due to the constraint $E\{\log(\det[\mathbf{G}_0])\} = \nu_{G_0}$ with $|\nu_{G_0}| < +\infty$, is important for the nonparametric probabilistic approach of model uncertainties and for the construction of prior probability model of elasticity random field (see Section 5). This is the reason why the truncated Gaussian distribution restricted to $\mathbb{M}_n^+$ and the GOE do not satisfy this invertibility condition that is absolutely required for stochastic modeling of uncertainties in many cases.

(iii) Algebraic representation and generator of realizations. Any random matrix $[\mathbf{G}_0]$ in SG_0^+ can be written as $[\mathbf{G}_0] = [\mathbf{L}]^T [\mathbf{L}]$ in which $[\mathbf{L}]$ is the upper triangular random matrix with values in $\mathbb{M}_n$ such that:

(a) random variables $\{[\mathbf{L}]_{jj'}, 1 \leq j \leq j' \leq n\}$ are statistically independent.

(b) for $1 \leq j < j' \leq n$, $[\mathbf{L}]_{jj'} = \sigma_n \, U_{jj'}$ in which $\sigma_n = \delta\,(n+1)^{-1/2}$ and where $U_{jj'}$ is a real-valued Gaussian random variable with zero mean and variance equal to 1.

(c) for $1 \leq j \leq n$, $[\mathbf{L}]_{jj} = \sigma_n \sqrt{2V_j}$ where V_j is a positive-valued Gamma random variable whose probability density function with respect to $d\mathrm{v}$ is

$$p_{V_j}(\mathrm{v}) = 1_{[0,+\infty]}(\mathrm{v}) \frac{1}{\Gamma\left(\frac{n+1}{2\delta^2} + \frac{1-j}{2}\right)} \, \mathrm{v}^{\frac{n+1}{2\delta^2} - \frac{1+j}{2}} \, e^{-\mathrm{v}} \, . \tag{4.8}$$

With realizations $\{U(\theta)_{jj'}, 1 \leq j \leq j' \leq n\}$ and $\{V_j(\theta), 1 \leq j \leq n\}$ for $\theta \in \Theta$ (that are very easy to generate), the corresponding realization $[\mathbf{G}_0(\theta)] \in \mathbb{M}_n^+$ of random matrix $[\mathbf{G}_0]$ is easily constructed using the above algebraic expressions.

4.4. Ensemble SG_ε^+ of positive-definite random matrices with a unit mean value and a positive-definite lower bound

This ensemble is derived from ensemble SG_0^+ (see [1]).

(i) Construction of ensemble SG_ε^+. A $\mathbb{M}_n^+$-valued random matrix $[\mathbf{G}]$ in ensemble $\mathrm{SG}_\varepsilon^+$ has a mean value that is the unit matrix and has a lower bound that is a positive-definite matrix controlled by an arbitrary positive number ε. Any random matrix in $\mathrm{SG}_\varepsilon^+$ is written as

$$[\mathbf{G}] = \frac{1}{1+\varepsilon}\{[\mathbf{G}_0] + \varepsilon\,[I_n]\} \quad , \quad [\mathbf{G}_0] \in \mathrm{SG}_0^+ \quad , \quad \varepsilon > 0 \, . \tag{4.9}$$

We have $0 < [G_\ell] < [\mathbf{G}]$ *a.s.*, in which the lower bound is the positive-definite matrix $[G_\ell] = c_\varepsilon[I_n] \in \mathbb{M}_n^+$ with $c_\varepsilon = \varepsilon/(1+\varepsilon)$.

(ii) Properties and hyperparameter for controlling the level of statistical fluctuations. For all $\varepsilon > 0$, we have

$$E\{[\mathbf{G}]\} = [I_n] \quad , \quad E\{\log(\det([\mathbf{G}] - [G_\ell]))\} = \nu_{G_\varepsilon} \,, \tag{4.10}$$

in which $\nu_{G_\varepsilon} = \nu_{G_0} - n\log(1+\varepsilon)$. The hyperparameter δ_G of a random matrix $[\mathbf{G}]$ belonging to $\mathrm{SG}_\varepsilon^+$in is defined as its coefficient of variation,

$$\delta_G = \left\{ \frac{E\{\| [\mathbf{G}] - E\{[\mathbf{G}]\} \|_F^2\}}{\| E\{[\mathbf{G}]\} \|_F^2} \right\}^{1/2} = \left\{ \frac{1}{n} E\{\| [\mathbf{G}] - [I_n] \|_F^2\} \right\}^{1/2} . \tag{4.11}$$

Consequently, $\delta_G = \delta/(1+\varepsilon)$ in which δ is the hyperparameter defined for ensemble SG_0^+.

4.5. Ensemble $\mathrm{SE}_\varepsilon^+$ of positive-definite random matrices with a given mean value and a positive-definite lower bound

This ensemble is derived from ensemble $\mathrm{SG}_\varepsilon^+$ (see [1] and [54]).

(i) Construction of ensemble $\mathrm{SE}_\varepsilon^+$. Let $[\underline{A}]$ be a given matrix in $\mathbb{M}_n^+$ for which the Cholesky factorization is written as $[\underline{A}] = [L_{\underline{A}}]^T [L_{\underline{A}}]$ with $[L_{\underline{A}}]$ an upper triangular matrix in $\mathbb{M}_n$. Below, $[L_{\underline{A}}]$ could be replaced by the square root $[\underline{A}]^{1/2}$ of $[\underline{A}]$. A $\mathbb{M}_n^+$-valued random matrix $[\mathbf{A}]$ in ensemble $\mathrm{SE}_\varepsilon^+$ has a mean value that is the matrix $[\underline{A}]$ given in $\mathbb{M}_n^+$ and has a lower bound $[A_\ell] = c_\varepsilon[\underline{A}] \in \mathbb{M}_n^+$ with $c_\varepsilon = \varepsilon/(1+\varepsilon)$, and is defined by

$$[\mathbf{A}] = [L_{\underline{A}}]^T [\mathbf{G}] [L_{\underline{A}}] \quad , \quad [\mathbf{G}] \in \mathrm{SG}_\varepsilon^+ \quad , \quad 0 < [A_\ell] < [\mathbf{A}] \ a.s. . \tag{4.12}$$

(ii) Second-order properties and invertibility. It can be seen that any $[\mathbf{A}]$ in $\mathrm{SE}_\varepsilon^+$ is such that

$$E\{[\mathbf{A}]\} = [\underline{A}] \in \mathbb{M}_n^+ \quad , \quad E\{\| [\mathbf{A}] \|^2\} \leq E\{\| [\mathbf{A}] \|_F^2\} < +\infty \,. \tag{4.13}$$

We have the following invertibility property, which shows that a random matrix in $\mathrm{SE}_\varepsilon^+$ can be used for modelling the coefficient of a random elliptic differential operator. For all $\mathbf{X} \in L^2(\Theta, \mathbb{R}^n)$, the quadratic form $b_{\mathbf{A}}(\mathbf{X}, \mathbf{X}) = E\{\langle [\mathbf{A}]\mathbf{X}, \mathbf{X} \rangle\}$ is such that $b_{\mathbf{A}}(\mathbf{X}, \mathbf{X}) \geq c_\varepsilon E\{\| [L_{\underline{A}}]\mathbf{X} \|^2\}$, random matrix $[\mathbf{A}]$ is thus invertible a.s., and $[\mathbf{A}]^{-1}$ is a second-order random variable,

$$E\{\| [\mathbf{A}]^{-1} \|^2\} \leq E\{\| [\mathbf{A}]^{-1} \|_F^2\} < +\infty \,. \tag{4.14}$$

5. Algebraic prior probability model for heterogeneous elastic media

In this section, the developments are devoted to the algebraic prior probability models of random fields, which are presented in the context of continuum mechanics of heterogeneous materials for 3D linear elasticity of complex microstructures. These results will be used to build the stochastic homogenized models that will be presented later. In continuum mechanics, a simple heterogeneous microstructure can be described in terms of constituents, for instance, a polymer matrix with long carbon fibers. A complex heterogeneous microstructure cannot be described in terms of constituents, for instance, a live tissue such as a cortical bone. In such a case the random elasticity field cannot be defined at microscale. Therefore, a random apparent elasticity field is defined at mesoscale and the stochastic homogenization consists in performing the change of scale from the mesoscale to the macroscale that is characterized by effective mechanical properties. These effective properties are quasi-deterministic if there is a scale separation, and are random if not. In this last case, the stochastic homogenization yields random apparent mechanical properties at macroscale. This section deals with the construction of a prior probability model of the mesoscale apparent random elasticity field for a 3D linear complex heterogeneous microstructure.

5.1. Voigt notation of the tensor-valued random elasticity field

Let Ω be bounded open domain of $\mathbb{R}^3$ with generic point $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$. For all $\boldsymbol{\xi}$ fixed in Ω, the random fourth-order elasticity tensor $\mathbb{C}(\boldsymbol{\xi}) = \{\mathbb{C}_{ijkh}(\boldsymbol{\xi})\}_{ijkh}$ is represented in Voigt notation by the random (6×6) matrix $[\mathbf{C}(\boldsymbol{\xi})]$ such that

$$[\mathbf{C}(\boldsymbol{\xi})]_{\mathbf{ij}} = \mathbb{C}_{ijkh}(\boldsymbol{\xi}) \quad \text{with} \quad \mathbf{i} = (i, j) \quad \text{and} \quad \mathbf{j} = (k, h)\,, \tag{5.1}$$

with i, j, k, and h in $\{1, 2, 3\}$, and $\mathbf{i}$ and $\mathbf{j}$ in $\{1, \ldots, 6\}$. Random elasticity field $[\mathbf{C}] = \{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is a non-Gaussian random field, defined on the probability space $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by Ω, with values in $\mathbb{M}_6^+$. As we have explained in Section 4, the $\mathbb{M}_6^+$-valued random field $[\mathbf{C}]$ cannot be Gaussian. Consequently, the only description of second-order quantities, that is to say, its mean function and its covariance function, are not sufficient. The probability distribution of the field, that is to say its system of marginal distributions, must be constructed. For this, we will use the algebraic prior representation of the indirect approach presented in Section 3.1, which will be based on the use of the random matrix theory presented in Section 4.

5.2. Algebraic prior probability model for heterogeneous anisotropic elastic media

This section deals with the construction of an algebraic prior model of the random elasticity field $\{[\mathbf{C}(\boldsymbol{\xi})]\,, \boldsymbol{\xi} \in \Omega\}$ defined in Section 5.1, for heterogeneous anisotropic elastic media that exhibit anisotropic statistical fluctuations. This model was initially introduced in [71, 72], was completed in [246] by introducing lower and upper bounds, is detailed/summarized in [1, 68], and for which an extension has recently been proposed in [74, 247] to take into account an uncertain spectrum. For this construction, the developments presented in Sections 4.3 to 4.5 are used. We have limited the presentation to an algebraic square-type representation. The reader will find also an algebraic exponential-type representation of the random elasticity field in [1, 73].

(i) Representation of random field $[\mathbf{C}]$ *with a lower bound and a random germ field with values in ensemble* SG_0^+ *of random matrices.* It is assumed that $\{[\mathbf{C}(\boldsymbol{\xi})]\,, \boldsymbol{\xi} \in \Omega\}$ ia a homogeneous and second-order random field. Let $[C_\ell]$ be the given deterministic $\mathbb{M}_6^+$-valued lower bound and let $[\underline{C}] \in \mathbb{M}_6^+$ the mean value, independent of $\boldsymbol{\xi}$, of random field $[\mathbf{C}]$. The matrices $[C_\ell]$ and $[\underline{C}]$ are assumed to be such that the matrix $[\underline{A}] = [\underline{C}] - [C_\ell]$ belongs to $\mathbb{M}_6^+$. The algebraic prior probability representation of $[\mathbf{C}]$ is defined by

$$[\mathbf{C}(\boldsymbol{\xi})] = [C_\ell] + [\underline{A}]^{1/2}\,[\mathbf{G}_0(\boldsymbol{\xi})]\,[\underline{A}]^{1/2} \quad , \quad \forall \boldsymbol{\xi} \in \Omega\,, \tag{5.2}$$

in which $[\underline{A}]^{1/2}$ is the square root of $[\underline{A},]$. In Eq. (5.2) $\{[\mathbf{G}_0(\boldsymbol{\xi})], \boldsymbol{\xi} \in \mathbb{R}^3\}$ is the non-Gaussian second-order random field, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by $\mathbb{R}^3$, with values in $\mathbb{M}_6^+$, homogeneous on $\mathbb{R}^3$, such that

$$E\{[\mathbf{G}_0(\boldsymbol{\xi})]\} = [I_6]\ ,\ E\{\log(\det[\mathbf{G}_0(\boldsymbol{\xi})])\} = \nu_{_{G_0}}\ ,\ |\nu_{_{G_0}}| < +\infty \quad , \quad \forall \boldsymbol{\xi} \in \mathbb{R}^3\,, \tag{5.3}$$

in which the real constant $\nu_{_{G_0}}$ is independent of $\boldsymbol{\xi}$ because random field $[\mathbf{G}_0]$ is homogeneous. For all $\boldsymbol{\xi}$ in Ω, we then have the properties

$$E\{[\mathbf{C}(\boldsymbol{\xi})]\} = [\underline{C}] \in \mathbb{M}_6^+ \quad , \quad [\mathbf{C}(\boldsymbol{\xi})] - [C_\ell] > 0 \quad a.s. \tag{5.4}$$

(ii) Construction of the random field $\{[\mathbf{G}_0(\boldsymbol{\xi})], \boldsymbol{\xi} \in \mathbb{R}^3\}$ *and its generator of realization.* The non-Gaussian random field $[\mathbf{G}_0]$ is constructed as a nonlinear mapping of $6 \times (6 + 1)/2$ independent, second-order, centered, homogeneous, Gaussian, and normalized random fields $\{\mathscr{U}_{jk}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}_{1 \le j \le k \le 6}$ defined on $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by $\mathbb{R}^3$, with values in $\mathbb{R}$. We then have

$$E\{\mathscr{U}_{jk}(\boldsymbol{\xi})\} = 0 \quad , \quad E\{\mathscr{U}_{jk}(\boldsymbol{\xi})^2\} = 1 \quad , \quad 1 \le j \le k \le 6\,. \tag{5.5}$$

The centered Gaussian random fields $\{\mathscr{U}_{jk}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}_{1 \le j \le k \le 6}$ are defined by $6 \times (6 + 1)/2$ autocorrelation functions $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3) \mapsto R_{jk}(\boldsymbol{\zeta}) = E\{\mathscr{U}_{jk}(\boldsymbol{\xi} + \boldsymbol{\zeta})\,\mathscr{U}_{jk}(\boldsymbol{\xi})\}$ from $\mathbb{R}^3$ into $\mathbb{R}$, such that $R_{jk}(0) = 1$. The spatial-correlation lengths of $\mathscr{U}_{jk}$ are defined by

$$L_\alpha^{jk} = \int_0^{+\infty} |R_{jk}(\boldsymbol{\zeta}^{(\alpha)})|\, d\zeta_\alpha \quad , \quad \zeta_\beta^{(\alpha)} = \zeta_\alpha \delta_{\alpha\beta} \quad , \quad \alpha = 1, 2, 3\,, \tag{5.6}$$

and are chosen as hyperparameters. For instance, a possible autocorrelation function is written as

$$R_{jk}(\boldsymbol{\zeta}) = \Pi_{\alpha=1}^{3} \rho_{\alpha}^{jk}(\zeta_\alpha) \quad , \quad \rho_{\alpha}^{jk}(\zeta_\alpha) = \frac{4(L_{\alpha}^{jk})^2}{\pi^2 \zeta_\alpha^2} \ \sin^2\!\left(\frac{\pi \zeta_\alpha}{2 L_{\alpha}^{jk}}\right) , \tag{5.7}$$

which shows that $\mathscr{U}_{jk}$ is a mean-square continuous random field on $\mathbb{R}^3$, whose power spectral density function on $\mathbb{R}^3$ has the compact support, $\Pi_{\alpha=1}^{3}[-\pi/L_{\alpha}^{jk}, \pi/L_{\alpha}^{jk}]$. Such a model has $3 \times 6 \times (6+1)/2$ real hyperparameters. Random field $\{[\mathbf{G}_0(\boldsymbol{\xi})], \boldsymbol{\xi} \in \mathbb{R}^3\}$ and its generator of realizations are constructed following Section 4.3-(iii). For all $\boldsymbol{\xi}$ in $\mathbb{R}^3$, we have $[\mathbf{G}_0(\boldsymbol{\xi})] = [\mathbf{L}(\boldsymbol{\xi})]^T [\mathbf{L}(\boldsymbol{\xi})]$ in which $[\mathbf{L}(\boldsymbol{\xi})]$ is an upper (6×6) real triangular random matrix such that:

(a) random fields $\{[\mathbf{L}(\boldsymbol{\xi})]_{jk}, \boldsymbol{\xi} \in \Omega\}$, $1 \le j \le k \le 6$, are statistically independent.
(b) for $1 \le j < k \le 6$, $[\mathbf{L}(\boldsymbol{\xi})]_{jk} = \sigma_6 \, \mathscr{U}_{jk}(\boldsymbol{\xi})$ in which $\sigma_6 = \delta/\sqrt{6+1}$.
(c) for $1 \le j \le 6$, $[\mathbf{L}(\boldsymbol{\xi})]_{jj} = \sigma_6 \sqrt{2\,\gamma(a_j, \mathscr{U}_{jj}(\boldsymbol{\xi}))}$ in which $a_j = (6+1)/(2\delta^2) + (1-j)/2$ and where function $u \mapsto \gamma(\alpha, u)$ is such that $\Gamma_\alpha = \gamma(\alpha, U)$ is a gamma random variable with parameter α and where U is a normalized Gaussian random variable.

The realizations of $\{\mathscr{U}_{jk}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ for $1 \le j \le k \le 6$ are constructed using a generator of Gaussian homogeneous random fields [60, 62, 71]. From the above algebraic expressions, it can then be deduced the realizations of random field $\{[\mathbf{G}_0(\boldsymbol{\xi})], \boldsymbol{\xi} \in \mathbb{R}^3\}$.

(iii) Hyperparameter $\mathbf{s}$ *of the algebraic prior probability model of random anisotropic elastic field* $[\mathbf{C}]$. The components of hyperparameter $\mathbf{s}$, which belongs to the admissible set $\mathscr{C}_s \subset \mathbb{R}^{n_s}$, of the algebraic prior probability model of random anisotropic elastic field $\{[\mathbf{C}(\boldsymbol{\xi}; \mathbf{s})], \boldsymbol{\xi} \in \Omega\}$ are made up of the reshaping of $[C_\ell] \in \mathbb{M}_6^+$ (lower bound) and $[\underline{C}] \in \mathbb{M}_6^+$ (mean value), the $3 \times 6 \times (6+1)/2$ spatial-correlation lengths $\{L_1^{jk}, L_2^{jk}, L_3^{jk}\}_{1 \le j \le k \le 6}$, and the dispersion parameter δ with $0 < \delta < \sqrt{(6+1)/(6+5)}$ that controls the level of anisotropic statistical fluctuations.

(iv) Case of an uncertain spectrum. The power spectral density functions on $\mathbb{R}^3$ associated with the autocorrelation functions of random fields $\{\mathscr{U}_{jk}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}_{1 \le j \le k \le 6}$ are deterministic. In [74], an extension is proposed for stochastic elliptic operators defined by non-Gaussian random fields with uncertain spectrum. An application is also presented, which is devoted to the computational stochastic homogenization of heterogeneous media for which the elasticity random field has an uncertain spectral measure.

5.3. Algebraic prior stochastic model for heterogeneous elastic media with statistical fluctuations in a symmetry class and with anisotropic statistical fluctuations

This section deals with the construction of an algebraic prior model of random elasticity field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ defined in Section 5.1, for a heterogeneous elastic medium, which exhibits both statistical fluctuations in a given symmetry class (corresponding to a material symmetry) and anisotropic statistical fluctuations around this symmetry class, the two statistical fluctuations being controlled independently. This model has been introduced in [248, 249] following a first version proposed in [250] on the base of [251]. For this construction, the developments presented in Section 5.2 are used. It should be noted that a survey on the second-order description of mean-square homogeneous random fields can be found in [70] (but this paper does not consider the construction of probability measures that are required for non-Gaussian random fields).

(i) Hypotheses on the statistical fluctuations. We consider a 3D heterogeneous elastic medium with random elasticity field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$. For all $\boldsymbol{\xi}$ fixed in Ω, elasticity random matrix $[\mathbf{C}(\boldsymbol{\xi})]$:

(a) is, in mean, close to a given symmetry class (independent of $\boldsymbol{\xi}$), corresponding to a material symmetry.
(b) exhibits more or less anisotropic statistical fluctuations around this symmetry class.
(c) exhibits a level of statistical fluctuations in the symmetry class, which must be controlled independently of the level of anisotropic statistical fluctuations.

(ii) Notation and properties for positive matrices with symmetry classes. A given symmetry class is defined by a subset $\mathbb{M}_6^{\text{sym}} \subset \mathbb{M}_6^+$ such that, any matrix $[M]$ in $\mathbb{M}_6^{\text{sym}}$ exhibits the given symmetry and can be written as

$$[M] = \sum_{j=1}^{n_{\text{sym}}} m_j \, [E_j^{\text{sym}}] \quad , \quad [E_j^{\text{sym}}] \in \mathbb{M}_6^S \, , \tag{5.8}$$

in which $\{[E_j^{\text{sym}}], j = 1, \ldots, n_{\text{sym}}\}$ is a matrix algebraic basis of $\mathbb{M}_6^{\text{sym}}$ (see the Walpole tensor basis in [252, 253]), where $n_{\text{sym}} \leq 6 \times (6+1)/2$, and where $\mathbf{m} = (m_1, \ldots, m_{n_{\text{sym}}}) \in \mathscr{C}_{\mathbf{m}} \subset \mathbb{R}^{n_{\text{sym}}}$, in which the admissible subset $\mathscr{C}_{\mathbf{m}}$ of $\mathbb{R}^{n_{\text{sym}}}$ is defined by

$$\mathscr{C}_{\mathbf{m}} = \{\mathbf{m} \in \mathbb{R}^{n_{\text{sym}}} \mid \sum_{j=1}^{n_{\text{sym}}} m_j [E_j^{\text{sym}}] \in \mathbb{M}_6^+\} \, . \tag{5.9}$$

The matrices $[E_j^{\text{sym}}]$ are symmetric (belong to $\mathbb{M}_6^S$), but are not positive definite (do not belong to $\mathbb{M}_6^+$). The dimension n_{sym} of the symmetry class is equal to: 2 for isotropic, 3 for cubic, 5 for transversely isotropic, 6 or 7 for trigonal, 6 or 7 for tetragonal, 9 for orthotropic, 13 for monoclinic, and 21 for anisotropic symmetry. If $[M]$ and $[M']$ belong to $\mathbb{M}_6^{\text{sym}}$, then it can be proven that

$$[M]\,[M'] \in \mathbb{M}_6^{\text{sym}} \quad , \quad [M]^{-1} \in \mathbb{M}_6^{\text{sym}} \quad , \quad [M]^{1/2} \in \mathbb{M}_6^{\text{sym}} \, . \tag{5.10}$$

Any matrix $[N]$ belonging to $\mathbb{M}_6^{\text{sym}}$ can be written as

$$[N] = \exp_{\mathbb{M}}([\mathscr{N}]) \quad , \quad [\mathscr{N}] = \sum_{j=1}^{n_{\text{sym}}} y_j \, [E_j^{\text{sym}}] \quad , \quad \mathbf{y} = (y_1, \ldots, y_{n_{\text{sym}}}) \in \mathbb{R}^{n_{\text{sym}}} \, , \tag{5.11}$$

in which $\exp_{\mathbb{M}}$ is the matrix exponential from $\mathbb{M}_n^S(\mathbb{R})$ into $\mathbb{M}_n^+(\mathbb{R})$. It should be noted that matrix $[\mathscr{N}]$ is a symmetric real matrix but does not belong to $\mathbb{M}_n^{\text{sym}}(\mathbb{R})$ (because $\mathbf{y}$ is in $\mathbb{R}^{n_{\text{sym}}}$ and therefore, $[\mathscr{N}]$ is not a positive-definite matrix).

(iii) Algebraic prior stochastic model for the random elasticity field with material symmetries. Let $\{[C_\ell(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ be the given $\mathbb{M}_6^+$-valued lower-bound field and let $\{[\underline{C}(\boldsymbol{\xi})] = E\{[\mathbf{C}(\boldsymbol{\xi})]\}, \boldsymbol{\xi} \in \Omega\}$ be the given $\mathbb{M}_6^+$-valued mean field. It is assumed that these two deterministic fields are such that, for all $\boldsymbol{\xi}$ in Ω, $[\underline{K}(\boldsymbol{\xi})] = [\underline{C}(\boldsymbol{\xi})] - [C_\ell(\boldsymbol{\xi})]$ belongs to $\mathbb{M}_6^+$. The prior stochastic model is then defined by

$$[\mathbf{C}(\boldsymbol{\xi})] = [C_\ell(\boldsymbol{\xi})] + [\mathbf{K}(\boldsymbol{\xi})] \quad , \quad \forall \boldsymbol{\xi} \in \Omega \, , \tag{5.12}$$

in which $\{[\mathbf{K}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is the $\mathbb{M}_6^+$-valued random field with $E\{[\mathbf{K}(\boldsymbol{\xi})]\} = [\underline{K}(\boldsymbol{\xi})]$, defined by

$$[\mathbf{K}(\boldsymbol{\xi})] = [\underline{S}(\boldsymbol{\xi})]^T \, [\mathbf{A}(\boldsymbol{\xi})]^{1/2} [\mathbf{G}_0(\boldsymbol{\xi})] \, [\mathbf{A}(\boldsymbol{\xi})]^{1/2} \, [\underline{S}(\boldsymbol{\xi})] \, . \tag{5.13}$$

In Eq. (5.13), the random field $\{[\mathbf{A}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ and the field $\{[\underline{S}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ are defined as follows.

(a) $\{[\mathbf{G}_0(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is the $\mathbb{M}_6^+$-valued random field defined in Section 5.2-(ii), which is such that $E\{[\mathbf{G}_0(\boldsymbol{\xi})]\} = [I_6]$ and for which the level of statistical fluctuations is controlled by hyperparameter δ.

(b) $\{[\mathbf{A}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is a $\mathbb{M}_6^{\text{sym}}$-valued random field, which allows for generating statistical fluctuations in the symmetry class, and is assumed to be statistically independent of random field $[\mathbf{G}_0]$. For all $\boldsymbol{\xi}$ in Ω, $E\{[\mathbf{A}(\boldsymbol{\xi})]\} = [\underline{A}(\boldsymbol{\xi})] = P^{\text{sym}}([\underline{K}(\boldsymbol{\xi})])$ in which P^{sym} is the projection operator in the symmetry class from $\mathbb{M}_6^+$ in $\mathbb{M}_6^{\text{sym}}$. The random matrix $[\mathbf{A}(\boldsymbol{\xi})]$ is defined by

$$[\mathbf{A}(\boldsymbol{\xi})] = [\underline{A}(\boldsymbol{\xi})]^{1/2} [\mathbf{N}(\boldsymbol{\xi})] \, [\underline{A}(\boldsymbol{\xi})]^{1/2} \, . \tag{5.14}$$

The non-Gaussian $\mathbb{M}_6^{\text{sym}}$-valued random field $\{[\mathbf{N}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \mathbb{R}^3\}$ is such that $E\{[\mathbf{N}(\boldsymbol{\xi})]\} = [I_6]$ and is constructed

such that

$$[\mathbf{N}(\boldsymbol{\xi})] = \exp_{\mathbb{M}}([\boldsymbol{\mathcal{N}}(\boldsymbol{\xi})]) \quad , \quad [\boldsymbol{\mathcal{N}}(\boldsymbol{\xi})] = \sum_{j=1}^{n_{\text{sym}}} Y_j(\boldsymbol{\xi})[E_j^{\text{sym}}] \, , \tag{5.15}$$

in which $\{\mathbf{Y}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ is a non-Gaussian, homogeneous, second-order, mean-square continuous random field indexed by $\mathbb{R}^3$, with values in $\mathbb{R}^{n_{\text{sym}}}$, which is constructed using MaxEnt under the available information $E\{[\mathbf{N}(\boldsymbol{\xi})]\} = [I_6]$ and $E\{\log(\det([\mathbf{N}(\boldsymbol{\xi})]))\} = \nu_{_N}(\boldsymbol{\xi})$ with $|\nu_{_N}(\boldsymbol{\xi})| < +\infty$, in which the field $\nu_{_N}$ is supposed to be known. The hyperparameters of random field $[\mathbf{A}]$ are $[\underline{A}]$, the $3 \times n_{\text{sym}}$ spatial correlation lengths, and δ_A that controls the level of statistical fluctuations in the symmetry class (the details of this construction can be found in [248]).

(c) The $\mathbb{M}_6$-valued deterministic field $\{[\underline{S}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is constructed as follows. Let us consider the Cholesky factorizations $[\underline{K}(\boldsymbol{\xi})] = [L_{\underline{K}}(\boldsymbol{\xi})]^T \, [L_{\underline{K}}(\boldsymbol{\xi})]$ of $[\underline{K}(\boldsymbol{\xi})] \in \mathbb{M}_6^+$ and $[\underline{A}(\boldsymbol{\xi})] = [L_{\underline{A}}(\boldsymbol{\xi})]^T \, [L_{\underline{A}}(\boldsymbol{\xi})]$ of $[\underline{A}(\boldsymbol{\xi})] = P^{\text{sym}}([\underline{K}(\boldsymbol{\xi})]) \in \mathbb{M}_6^{\text{sym}}$ in which $[L_{\underline{K}}(\boldsymbol{\xi})]$ and $[L_{\underline{A}}(\boldsymbol{\xi})]$ are upper triangular matrices. Then, we have $[\underline{S}(\boldsymbol{\xi})] = [L_{\underline{A}}(\boldsymbol{\xi})]^{-1} \, [L_{\underline{K}}(\boldsymbol{\xi})]$ yielding $[\underline{K}(\boldsymbol{\xi})] = [\underline{S}(\boldsymbol{\xi})]^T \, [\underline{A}(\boldsymbol{\xi})] \, [\underline{S}(\boldsymbol{\xi})]$.

6. Illustrations in stochastic homogenization of heterogeneous materials at microscale and nanoscale

Two illustrations are presented. The first one deals with a probabilistic analysis of the representative volume element size in stochastic homogenization of a heterogeneous complex microstructure for which the random medium at mesoscale has anisotropic statistical fluctuations [72]. The random elasticity field is thus modeled with the algebraic prior probability model presented in Section 5.2. The second one is devoted to the stochastic modeling of random interphases from atomistic simulations for a polymer nanocomposite and its identification solving a statistical inverse problem [254]. An equivalent model in continuum mechanics is considered in interphase region that has a finite thickness (it is not a perfect interface). In this region, the random elasticity field is modeled using the algebraic prior stochastic model with statistical fluctuations in a symmetry class, presented in Section 5.3.

6.1. Probabilistic analysis of the representative volume element size in stochastic homogenization of a heterogeneous complex microstructure

The homogenization of elastic materials with heterogeneous microstructures composed of several phases with well defined constituents (from a continuum mechanics point of view) and the calculation of the macroscopic properties (effective properties) have received considerable attention (see for instance [255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265]), for stochastic homogenization (see [266, 267, 268, 269, 270, 271, 272, 273, 274]). In the field of linear and nonlinear mechanics, it should also be noted the works devoted to computational multiresolution materials and multiscale method (see for instance [246, 254, 275, 276, 277, 278, 279, 280, 281]), and more recently, for data-driven and machine learning approaches applied to heterogeneous materials (see for instance [282, 283, 213, 284, 149, 217, 219, 242]). In linear elasticity, the random microstructure can be homogenized if there exists a Representative Volume Element (RVE) for which the random fluctuations of the random effective stiffness tensor around its statistical mean value are "negligible". The analysis of the RVE size has received a particular attention (see for instance [285, 286, 287, 288, 289, 290]). Often, the statistics-based bounding techniques only use the lower-order statistics (first- and second-order moments). The probability distributions, which give the detailed probabilistic information, are not taken into account. It should be noted that generalized continuum theories have been widely used for describing microstructure-dependent mechanical effects in granular materials [291, 292, 293] and in metamaterials [294, 295, 296, 297, 298, 299, 300]. Although deterministic models cannot appropriately describe the behavior of such real physical systems, very few works consider uncertainties in the framework of generalized continua such as [301, 302, 303, 304]. In [305], uncertainties are taken into account in granular second-gradient models.

For elastic heterogeneous complex microstructures that cannot be described in terms of microscale constituents (such as living tissues for which constituents/phases cannot be described at the microscale), a stochastic homogenization has been proposed in [72] on the base of a stochastic modeling at mesoscale of the apparent elasticity field. This approach is summarized in this section. It should be noted that the hyperparameters of the prior probabilistic model can be identified from experiments by solving an inverse statistical problem as proposed in [242, 149, 217, 306, 307, 175, 308, 309, 310, 73]. In this section, we consider a heterogeneous complex elastic

microstructure that occupies the open bounded domain $\Omega =]0,1[^3$ of $\mathbb{R}^3$. The notations introduced in Section 5.1 are reused. At mesoscale, the microstructure is modeled as a random continuum medium.

(i) Stochastic elliptic boundary value problem for computational stochastic homogenization. For all m and r in $\{1,2,3\}$ the unknown random displacement field is the $\mathbb{R}^3$-valued random field $\{\mathbf{Y}^{mr}(\boldsymbol{\xi}) = (Y_1^{mr}(\boldsymbol{\xi}), Y_2^{mr}(\boldsymbol{\xi}), Y_3^{mr}(\boldsymbol{\xi})), \boldsymbol{\xi} \in \Omega\}$ defined on $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by Ω, such that for $i = 1,2,3$, and almost surely,

$$-\frac{\partial}{\partial \xi_j}\left(\mathbb{C}_{ijpq}(\boldsymbol{\xi})\,\varepsilon_{pq}(\mathbf{Y}^{mr}(\boldsymbol{\xi}))\right) = \mathbf{0}_3 \quad , \quad \forall \boldsymbol{\xi} \in \Omega\,, \tag{6.1}$$

$$\mathbf{Y}^{mr}(\boldsymbol{\xi}) = \mathbf{y}_0^{mr} \quad , \quad \forall \boldsymbol{\xi} \in \partial\Omega\,, \tag{6.2}$$

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

$$y_{0,j}^{mr} = (\delta_{jm}\,\xi_r + \delta_{jr}\,\xi_m)/2 \quad , \quad j \in \{1,2,3\}\,, \tag{6.3}$$

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 $(\Theta, \mathscr{T}, \mathscr{P})$. 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 elasticity tensor $\{\mathbb{C}_{ijmr}^{\text{eff}}\}_{ijmr}$, which is defined [257, 259] by

$$\mathbb{C}_{ijmr}^{\text{eff}} = \frac{1}{|\Omega|}\int_\Omega \mathbb{C}_{ijpq}(\boldsymbol{\xi})\,\varepsilon_{pq}(\mathbf{Y}^{mr}(\boldsymbol{\xi}))\,d\boldsymbol{\xi}\,, \tag{6.4}$$

in which $\mathbf{Y}^{mr}$ is the $\mathbb{R}^3$-valued random field that satisfies Eqs. (6.1) to (6.3) and where $|\Omega| = \int_\Omega d\boldsymbol{\xi}$. The random effective elasticity tensor $\mathbb{C}^{\text{eff}}$ is symmetric and positive definite almost surely. Using the Voigt notation defined in Section 5.1, the tensor-valued random elasticity field $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ is represented by the $\mathbb{M}_6^+$-valued random field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ and the random effective tensor $\mathbb{C}^{\text{eff}}$ is represented by a $\mathbb{M}_6^+$-valued random matrix $[\mathbf{C}^{\text{eff}}]$. If there is a scale separation (which means that Ω is effectively a RVE, then $[\mathbf{C}^{\text{eff}}]$ is quasi-deterministic and is the effective elasticity matrix. If not, Ω is not a RVE and $[\mathbf{C}^{\text{eff}}]$ is a random matrix with significant statistical fluctuations, which is an apparent elasticity matrix at macroscale, independent of $\boldsymbol{\xi}$.

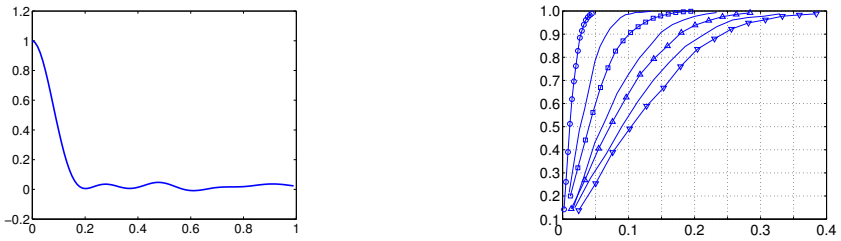
(ii) Prior probabilistic model, computational model, and its strong stochastic solution. The probability model of random field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is the algebraic prior probability model for heterogeneous anisotropic elastic media presented in Section 5.2 for which the hyperparameters (see Section 5.2-(iii)) are defined as follows. The mean value $[\underline{C}] \in \mathbb{M}_6^+$ is the elasticity matrix of a homogeneous anisotropic linear elastic microstructure, defined by equation (70) of [72]. The lower-bound matrix is chosen as $[C_\ell] = (\varepsilon/(1+\varepsilon))[I_6]$ with $\varepsilon = 10^{-6}$. The dispersion parameter δ, which allows the level of anisotropic statistical fluctuations to be controlled, is chosen as $\delta = 0.4$. The spatial-correlation lengths, which are defined by Eq. (5.6), are generated from a value L_d such that $L_1^{jk} = L_2^{jk} = L_3^{jk} = L_d$ for $1 \le j \le k \le 6$ and the autocorrelation functions are defined by Eq. (5.7). We then introduce the spatial-correlation length $L_d^c = \int_0^{+\infty} |r^c(\zeta)|\,d\zeta$, related to the random field $[\mathbf{C}]$, in which $\zeta \mapsto r^c(\zeta)$ is a correlation function on $\mathbb{R}$,

$$r^c(\zeta) = \frac{\text{tr}\,E\{([\mathbf{C}(\boldsymbol{\xi}+\boldsymbol{\kappa}(\zeta))] - [\underline{C}])([\mathbf{C}(\boldsymbol{\xi})] - [\underline{C}])\}}{E\{\|[\mathbf{C}(\boldsymbol{\xi})] - [\underline{C}]\|_F^2\}}\,, \tag{6.5}$$

in which $\boldsymbol{\kappa}(\zeta) = (\zeta, 0, 0)$ (note that the result would the same if $\boldsymbol{\kappa}(\zeta)$ was equal to $(0,\zeta,0)$ or to $(0,0,\zeta)$). In such a case, $L_d^c = 1.113\,L_d$. For $L_d = 0.1$ and $\delta = 0.4$, Fig. 1-(a) displays the graph of the correlation function $\zeta \mapsto r^c(\zeta)$ defined by Eq. (6.5). Under the introduced hypotheses, it is proven (see Proposition 5.1 in [74] using a deterministic spectrum) that the weak formulation of the stochastic elliptic boundary value problem admits a unique strong stochastic solution that is a second-order random field, $E\{\|\mathbf{Y}^{mr}(\boldsymbol{\xi})\|^2\} < +\infty\,, \forall \boldsymbol{\xi} \in \Omega$. The stochastic computational model is obtained by discretizing the weak formulation of the boundary value problem with the finite element method. The approximation of the strong stochastic solution is constructed using the Monte Carlo simulation method (see for instance, [311, 120]) for which the mean-square convergence of the operator norm of the random effective elasticity matrix is obtained for

900 realizations.

(iii) Probabilistic analysis of the RVE size. We introduce the normalized random variable $Z = \|[\mathbf{C}^{\text{eff}}]\|/E\{\|[\mathbf{C}^{\text{eff}}]\|\}$. For $\delta = 0.4$, Fig. 1-(b) displays the graph of the function $\beta \mapsto \mathbb{P}(\beta) = \text{Proba}\{1-\beta < Z \leq 1+\beta\}$ that shows the evolution of the probability distribution of Z as a function of L_d. For instance, for $L_d = 0.2$, if we consider $\beta = 0.02$, 0.04, and 0.08, this figure shows that $\text{Proba}\{0.98 < Z \leq 1.02\} = 0.36$, $\text{Proba}\{0.96 < Z \leq 1.04\} = 0.65$, and $\text{Proba}\{0.92 < Z \leq 1.08\} = 0.95$. Clearly, the scales are badly separated for the value 0.2 of L_d. It can also be seen that the scales are reasonably well separated for $L_d = 0.1$.



(a) Correlation function $\zeta \mapsto r^c(\zeta)$

(b) Graph of function $\beta \mapsto \mathbb{P}(\beta)$

Figure 1: For $\delta = 0.4$, figure (a) displays the graph of correlation function for $L_d = 0.1$ (horizontal axis ζ, vertical axis $r^c(\zeta)$); figure (b) shows the graph of function $\beta \mapsto \mathbb{P}(\beta) = \text{Proba}\{1-\beta < Z \leq 1+\beta\}$ for $L_d = 0.1$ (circle), 0.2 (no marker), 0.3 (square), 0.4 (no marker), 0.5 (triangle-up), 0.6 (no marker), 0.7 (triangle-down). [Figure from [72]].

6.2. *Stochastic continuum modeling of the random interphase of a polymer nanocomposite using atomistic simulations and a statistical inverse problem*

A specificity of nanoreinforced composites is the existence of an interphase region surrounding an inclusion, which has a finite thickness (not reduced to a perfect interface). Within this region the polymer chains exhibit conformational and geometrical properties that are different from those in the bulk polymer phase. A lot of experimental characterization of these properties have been carried out (see for instance [312, 313, 314, 315, 316]). These observations are well correlated by atomistic simulations and highlights a preferred orientation of the polymer chain segments tangentially to the particle surface [317, 318, 319, 320, 321, 322].

The application presented below deals with the continuum mechanics of a polymer system reinforced by a nanoscopic inclusion of silica [254] and is a complementary and fundamental aspect of model construction and inverse calibration based on atomistic simulations. The elastic properties of the interphase is modeled by a non-Gaussian tensor-valued random field. Molecular dynamics simulations are used to infer some basic properties and to construct a simulated database devoted to the model calibration. This identification step is subsequently addressed by solving a statistical inverse problem stating the equivalence of the apparent properties obtained from atomistic computations and those estimated from stochastic homogenization in a continuum mechanics formulation. Let Ω be the interphase region between the polymer matrix and a silicon nanoinclusion inserted in the polymer. It is assumed that

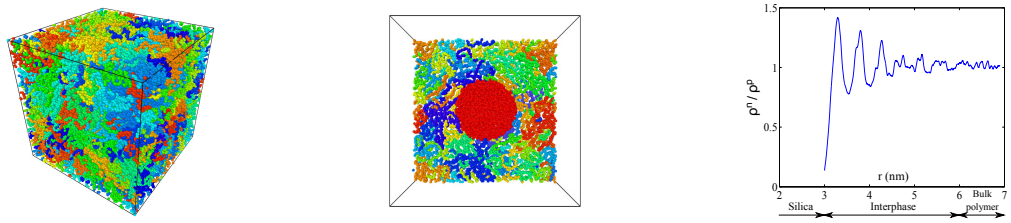
(a) the prior stochastic model $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ of the random apparent elasticity field in the interphase region Ω with a given symmetry class is an adaptation (see [254]) of the one presented in Section 5.3.

(b) the molecular dynamics is used to generate atomistic simulated data that are used for identifying random field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ in the context of continuum mechanics.

(c) the identification of the prior stochastic model is carried out by solving a statistical inverse problem.

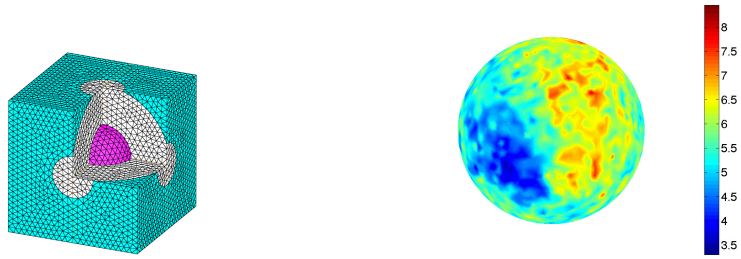
Molecular dynamics (MD) modeling and atomistic simulations. The amorphous polymer is made up of long chains with CH_2 sites, represented through a coarse graining with harmonic potentials and the Lennard Jones potential. The silicon nanoinclusion is made up of amorphous bulk of SiO_2 molecules described in terms of Si and O atoms, with Coulomb potential. The interaction CH_2 - SiO_2 is modeled by a Lennard Jones potential (see for instance [323]). The MD simulations [324] is performed with a target volume fraction of 4.7% in the atomistic domain Ω_a that is a cube of $6.8 \times 10^{-9}\, m$ side, which contains 10 polymer chains. Each polymer chain is made up of 1 000 CH_2 sites

yielding a total of 10 000 CH_2. The SiO_2 nanoinclusion is a sphere of $3 \times 10^{-9}\, m$ diameter with 275 Si atoms and 644 O atoms. The temperature is $100\ ^oK$ and the pressure is the control variable. Six mechanical tests in traction and shear are simulated. A time-spatial averaging is performed for estimating the apparent strain that allows for deducing realizations of the random apparent elasticity matrix $[\mathbf{C}^{\text{app,MD}}]$ related to domain Ω_a in the sense of the continuum mechanics. Figs. 2-(a) and (b) show a visualization of an instantaneous configuration of the polymer chains and of the atoms of the inclusion. Fig. 2-(c) displays the polymer density ρ^n in the nanocomposite divided by the pure polymer density ρ^p as a function of the distance r from the center of the sphere representing the silicon nanoinclusion. This figure shows that the interphase thickness, t, is between $2 \times 10^{-9}\, m$ and $3 \times 10^{-9}\, m$.



(a) Instantaneous configuration (b) Instantaneous configuration (c) Function $r \mapsto (\rho^n/\rho^p)(r)$

Figure 2: Figures (a) and (b): visualization of an instantaneous configuration of the polymer chains and of the atoms of the inclusion. Figure (c): polymer density ρ^n in the nanocomposite divided by the pure polymer density ρ^p as a function of the distance r (in nanometer) from the center of the sphere representing the silicon nanoinclusion. [Figure from [254]].



(a) 3D view of the meshed continuum model (b) Realization of identified random elasticity field $[\mathbf{C}(\boldsymbol{\xi})]_{11}$

Figure 3: Figure (a): 3D view of the meshed continuum model (the inclusion appears in magenta (or in dark gray), the interphase in white and the polymer matrix in turquoise (or in gray)). Figure (b): plot of a realization of component (1, 1) of the identified random elasticity field $[\mathbf{C}(\boldsymbol{\xi})]_{11}$ (in GPa) in the interphase domain. [Figure from [254]].

(ii) Continuum mechanics model and prior stochastic model of the elasticity field describing the interphase. A continuum mechanics model is constructed. The finite element method is used for solving the 6 stochastic BVP corresponding to the 6 mechanical tests. The stochastic computational model is made up of 190 310 finite elements. Fig. 3-(a) shows a 3D view of the meshed continuum model. In the interphase region Ω whose thickness is t, the prior stochastic model of the elasticity field $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ is constructed as explained in Section 5.3. There are only statistical fluctuations in the class of transversally isotropic material symmetries in the spherical coordinates (r, φ, ψ) in the orthonormal spherical frame (e_r, e_φ, e_ψ). The hyperparameters are the dispersion parameter δ related to the statistical fluctuations in the symmetry class, the spatial-correlation length L_r along the radial direction e_r, and the spatial-correlation length L_a along the two directions e_φ and e_ψ (assuming a same value). The Silica inclusion and the polymer bulk are linear elastic, isotropic, homogeneous, and deterministic media, for which the elastic properties (bulk and shear moduli) have been estimated from the MD simulations.

(iii) Statistical inverse problem and results. The maximum likelihood method is used for estimating the optimal values of the hyperparameters of $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$. The observed random quantity is the random apparent elasticity matrix $[\mathbf{C}^{\text{app}}]$ related to domain Ω_a. The probability density function of $[\mathbf{C}^{\text{app}}]$ is estimated by using realizations that are computed by stochastic homogenization, using the finite element solutions of the 6 stochastic BVP with

the prior stochastic model $\{[\mathbf{C}(\boldsymbol{\xi})], \boldsymbol{\xi} \in \Omega\}$ in the interphase region $\Omega \subset \Omega_a$. The likelihood function is estimated with the realizations of $[\mathbf{C}^{\text{app,MD}}]$ that has been computed with the MD simulations (considered as the experiments). The optimal values of the hyperparameters are $\delta = 0.2$, $L_r = t/4$, and $L_a = 3.5 \times 10^{-9}$ m. Fig. 3-(b) displays a realization of component $(1, 1)$ of the identified random elasticity field $[\mathbf{C}(\boldsymbol{\xi})]_{11}$ (in GPa) inside the interphase domain (relatively close to the exterior boundary surface of the interphase domain) and computed with the optimal values of the hyperparameters.

7. Probabilistic Learning on Manifolds (PLoM) as a machine learning tool for small dataset and probabilistic inference

The probabilistic learning is a very active domain of research for constructing surrogate models (see for instance, [325, 229, 326, 44, 233, 235, 239]). Probabilistic Learning on Manifolds (PLoM) is a tool in computational statistics, introduced in 2016 [230] and which can be viewed as a tool for scientific machine learning. The PLoM approach has specifically been developed for the small dataset cases [230, 232, 234, 149, 236]. The method avoids the scattering of learned realizations associated with the probability distribution in order to preserve its concentration in the neighborhood of the random manifold defined by the parameterized computational model. This method allows for solving unsupervised and supervised problems under uncertainty for which the training datasets are small. This situation is encountered in many problems of physics and engineering science with expensive function evaluations. The exploration of the admissible solution space in these situations is thus hampered by available computational resources.

Several extensions have been proposed to take into account implicit constraints induced by physics, computational models, and measurements [240, 241, 242], to reduce the stochastic dimension using a statistical partition approach [238], and to update the prior probability distribution by a target dataset whose points are, for instance, experimental realizations of the system observations [243]. Consequently, PLoM constrained by a stochastic computational model and statistical moments or samples/realizations allows performing probabilistic learning inference and constructing predictive statistical surrogate model for large parameterized stochastic computational models. This last capability of PLoM can also be viewed as an alternative method to the Bayesian inference for the high dimension [143, 327, 20, 145, 146, 147, 148, 7, 328, 160] and is a complementary approach to existing methods in machine learning for sampling distributions on manifolds under constraints (although a Bayesian inference methodology has also been developed using the probabilistic learning on manifolds for the high dimensions [149]).

PLoM has successfully been adapted to tackle these challenges for several related problems including nonconvex optimization under uncertainty [211, 213, 212, 114, 215, 329, 218, 330], fracture paths in random composites [331], ultrasonic transmission technique in cortical bone microstructures [149], updating digital twins under uncertainties [219], updating of under observed dynamical system [220], calculation of the Sobol indices [332], dynamic monitoring [333], surrogate modeling of structural seismic response [334].

In this section, we present a brief overview of this method and we will present an application to the homogenization of heterogeneous microstructures in continuum mechanics.

7.1. Setting the problem of the probabilistic learning on manifolds

The probabilistic learning on manifolds in presented in the context of the supervised learning, which is adapted to the framework of computational mechanics and engineering science.

(i) Statistical surrogate model of a parameterized solution of a stochastic computational model. We consider the following equation

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbf{U}) \,, \tag{7.1}$$

associated with a parameterized stochastic computational model, obtained by the discretization of a parameterized stochastic BVP. In this equation,

(a) $\mathbf{W}$ is a *control parameter* that is a $\mathbb{R}^{n_w}$-valued random variable defined on the probability space $(\Theta, \mathscr{T}, \mathscr{P})$ with a given prior probability distribution $P_{\mathbf{W}}(d\mathbf{w})$ whose support is $\mathscr{C}_w \subset \mathbb{R}^{n_w}$.

(b) $\mathbf{U}$ is an *uncontrolled parameter* that is a $\mathbb{R}^{n_u}$-valued random variable defined on $(\Theta, \mathscr{T}, \mathscr{P})$ with a given prior probability distribution $P_{\mathbf{U}}(d\mathbf{u})$, which is assumed statistically independent of $\mathbf{W}$.

(c) $\mathbf{Q}$ is the quantity of interest (QoI) related to system observations, which is a $\mathbb{R}^{n_q}$-valued random variable defined on $(\Theta, \mathscr{T}, \mathscr{P})$ and depending on $\mathbf{W}$ and $\mathbf{U}$, as formulated by Eq. (7.1).

(d) $(\mathbf{w}, \mathbf{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbf{u})$ is a measurable mapping from $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ into $\mathbb{R}^{n_q}$, which is not explicitly known and which is indirectly constructed by solving the parameterized stochastic computational model.

(e) We define the stochastic mapping $\mathbf{F}$ on $(\Theta, \mathscr{T}, \mathscr{P})$ such that, for all $\mathbf{w}$ in $\mathbb{R}^{n_w}$, $\mathbf{F}(\mathbf{w}) = \mathbf{f}(\mathbf{w}, \mathbf{U})$. The stochasticity of $\mathbf{F}$ comes from the randomness of the uncontrolled parameter $\mathbf{U}$. The random graph $\{(\mathbf{w}, \mathbf{F}(\mathbf{w})), \mathbf{w} \in \mathbb{R}^{n_w}\}$ defined a random manifold in $\mathbb{R}^{n_x}$ with $n_x = n_w + n_q$, which can be viewed as a statistical surrogate model associated with the parameterized stochastic computational model, and which is completely defined by the joint probability distribution $P_{\mathbf{W},\mathbf{Q}}(d\mathbf{w}, d\mathbf{q})$ on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_q}$. Consequently, $P_{\mathbf{W},\mathbf{Q}}(d\mathbf{w}, d\mathbf{q})$ is concentrated in a subset of $\mathbb{R}^{n_w} \times \mathbb{R}^{n_q}$ located in the region of the random manifold. It should be noted that, in this context, the surrogate model is defined in a statistical framework and not using any algebraic representation.

Let $\mathbf{X} = (\mathbf{W}, \mathbf{Q})$ be the random variable, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, with values in $\mathbb{R}^{n_x} = \mathbb{R}^{n_w} \times \mathbb{R}^{n_q}$, whose probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ on $\mathbb{R}^{n_x}$ is concentrated in a subset of $\mathbb{R}^{n_x}$ due to the random manifold defined by the random graph $\{(\mathbf{w}, \mathbf{F}(\mathbf{w})), \mathbf{w} \in \mathbb{R}^{n_w}\}$. In the following, we will assume that $P_{\mathbf{X}}(d\mathbf{x})$ is written as $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$ in which $p_{\mathbf{X}}$ is the pdf on $\mathbb{R}^{n_x}$, which is concentrated in a region of $\mathbb{R}^{n_x}$.

(ii) Training dataset for random vector $\mathbf{X}$. A generator of the prior probability distributions $P_{\mathbf{W}}(d\mathbf{w})$ and $P_{\mathbf{U}}(d\mathbf{u})$ has to be constructed (based on the direct or the indirect approach as explained in Section 3.1) and is used for constructing n_d independent realizations $\mathbf{w}_d^j = \mathbf{W}(\theta^j) \in \mathbb{R}^{n_w}$ and $\mathbf{u}_d^j = \mathbf{U}(\theta^j) \in \mathbb{R}^{n_q}$, $\theta^j \in \Theta$, of the random variables $\mathbf{W}$ and $\mathbf{U}$. For each realization $(\mathbf{w}_d^j, \mathbf{u}_d^j)$ of $(\mathbf{W}, \mathbf{U})$, the computational model is used for calculating the corresponding realization $\mathbf{q}_d^j = \mathbf{Q}(\theta^j) \in \mathbb{R}^{n_q}$ such that $\mathbf{q}_d^j = \mathbf{f}(\mathbf{w}_d^j, \mathbf{u}_d^j)$. The corresponding realization $\mathbf{x}_d^j = \mathbf{X}(\theta^j) \in \mathbb{R}^{n_x}$ of $\mathbf{X}$ is then written as $\mathbf{x}_d^j = (\mathbf{w}_d^j, \mathbf{q}_d^j)$. For each realization θ^j, a deterministic calculation is thus carried out with the computational model. Often, the n_d realizations are performed using parallel calculation. The training dataset $\mathscr{D}_{n_d}(\mathbf{x})$, made up of n_d independent realizations $\mathbf{x}_d^j \in \mathbb{R}^{n_x}$ of the $\mathbb{R}^{n_x}$-valued random variable $\mathbf{X}$ whose probability distribution is $P_{\mathbf{X}}(d\mathbf{x})$ on $\mathbb{R}^{n_x}$, is then defined by

$$\mathscr{D}_{n_d}(\mathbf{x}) = \{\mathbf{x}_d^j = (\mathbf{w}_d^j, \mathbf{q}_d^j), j = 1, \ldots, n_d\}\,. \tag{7.2}$$

It is assumed that the calculation of a single realization with the computational model is expensive and consequently, the number n_d is small, which means that $\mathscr{D}_{n_d}(\mathbf{x})$ is a small dataset and not a big dataset. Under this assumption, PLoM will be an adapted machine learning tool.

(iii) Role played by the learned dataset generated by the probabilistic learning on manifolds for the construction of statistical surrogate models. The construction of a statistical surrogate model can be defined as follows. From the n_d realizations of the training dataset, we have to estimate the joint probability distribution $P_{\mathbf{W},\mathbf{Q}}(d\mathbf{w}, d\mathbf{q})$ on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_q}$ of $(\mathbf{W}, \mathbf{Q})$ and more particularly, the conditional statistics such as the conditional probability distribution $\mathrm{P}(d\mathbf{q}; \mathbf{w})$ of $\mathbf{Q}$ given $\mathbf{W} = \mathbf{w}$, for any $\mathbf{w}$ given in $\mathscr{C}_w$,

$$\mathrm{P}(d\mathbf{q}; \mathbf{w}) = P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}\,|\,\mathbf{w}) \quad , \quad \mathbf{w} \in \mathscr{C}_w \tag{7.3}$$

or to estimate the conditional expectation $\mathbf{h}(\mathbf{w})$ of $\boldsymbol{\mathcal{H}}(\mathbf{Q})$ given $\mathbf{W} = \mathbf{w}$, in which $\mathbf{q} \mapsto \boldsymbol{\mathcal{H}}(\mathbf{q})$ a given measurable mapping from $\mathbb{R}^{n_q}$ in $\mathbb{R}^{n_h}$,

$$\mathbf{h}(\mathbf{w}) = E\{\boldsymbol{\mathcal{H}}(\mathbf{Q})|\mathbf{W} = \mathbf{w}\} = \int_{\mathbb{R}^{n_q}} \boldsymbol{\mathcal{H}}(\mathbf{q})\, P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}\,|\,\mathbf{w}) \quad , \quad \mathbf{w} \in \mathscr{C}_w\,. \tag{7.4}$$

Note that the non-Gaussian probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ of $\mathbf{X} = (\mathbf{W}, \mathbf{Q})$, which is concentrated in an unknown region of $\mathbb{R}^{n_x}$, is unknown but can be estimated with nonparametric statistics using the training dataset, $\mathscr{D}_{n_d}(\mathbf{x}) = \{\mathbf{x}_d^j, j = 1, \ldots, n_d\}$ of length n_d.

Under the introduced hypothesis, the training dataset is small due to a high numerical cost of a single evaluation $\mathbf{q}_d^j = \mathbf{f}(\mathbf{w}_d^j, \mathbf{u}_d^j)$ using the computational model. Therefore, n_d will be, in general, not sufficiently large for obtaining a good convergence of the estimate $\mathrm{P}^{(n_d)}(d\mathbf{q}; \mathbf{w})$ towards $\mathrm{P}(d\mathbf{q}; \mathbf{w})$ or the estimate $\mathbf{h}^{(n_d)}(\mathbf{w})$ towards $\mathbf{h}(\mathbf{w})$ for all $\mathbf{w}$ in $\mathscr{C}_w$. For circumventing this difficulty, one way is the use of the probabilistic learning on manifolds. It allows to generate a

learned dataset $\mathscr{D}_{\text{learn}}(\mathbf{x})$ constituted of $n_{\text{learn}} \gg n_d$ learned realizations (resampling) of $\mathbf{X}$,

$$\mathscr{D}_{\text{learn}}(\mathbf{x}) = \{\mathbf{x}^{\ell}_{\text{learn}}, \ell = 1, \ldots, n_{\text{learn}}\}\,, \tag{7.5}$$

keeping the concentration (for illustration, see Fig. 4-(b)), that is to say by avoiding the scattering of the learned realizations for which the concentration is lost (see Fig. 4-(c)). Then, the convergence of the conditional statistics are obtained in estimating those with $\mathscr{D}_{\text{learn}}(\mathbf{x})$ instead of $\mathscr{D}_{n_d}(\mathbf{x})$, that is to say, using the learned realizations $\mathbf{w}^{\ell}_{\text{learn}}$ and $\mathbf{q}^{\ell}_{\text{learn}}$ of random variables $\mathbf{W}$ and $\mathbf{Q}$, such that

$$(\mathbf{w}^{\ell}_{\text{learn}}, \mathbf{q}^{\ell}_{\text{learn}}) = \mathbf{x}^{\ell}_{\text{learn}} \quad , \quad \ell = 1, \ldots, n_{\text{learn}}\,. \tag{7.6}$$

7.2. Methodology and algorithm of the probabilistic learning on manifolds (PLoM)

In this section we summarized the methodology and the algorithm of the probabilistic learning on manifolds [230, 236]. We will present some illustrations showing the capability of the PLoM algorithm. In particular, we will show how a general nonconvex optimization under uncertainties (OUU) and under nonlinear constraints can be solved with a limited number of function evaluations of the parameterized stochastic computational model by using the PLoM approach. For certain challenging applications, we will also present how the normalization can be preserved by introducing constraints and we will also introduce the extension of PLoM with partition [238]. The algorithm with constraints that is useful for the probabilistic learning inference will be presented in Section 7.3.

7.2.1. Summary of the methodology and algorithm

The methodology is described by the following steps, each one participating to the PLoM algorithm.

(i) Defining the random matrix $[\mathbf{X}]$*, its realization* $[x_d]$ *from the training dataset, and the estimates-based second-order moments.* Let $[\mathbf{X}] = [\mathbf{X}^1, \ldots, \mathbf{X}^{n_d}]$ be the random matrix defined on $(\Theta, \mathscr{T}, \mathscr{P})$, with values in $\mathbb{M}_{n_x,n_d}$ such that $[\mathbf{X}] = [\mathbf{X}^1, \ldots, \mathbf{X}^{n_d}]$ whose columns are n_d independent copies of random vector $\mathbf{X}$. Let $[x_d]$ be the matrix defined by

$$[x_d] = [\mathbf{x}^1_d \ldots \mathbf{x}^{n_d}_d] \quad \in \quad \mathbb{M}_{n_x,n_d}\,, \tag{7.7}$$

whose columns are the n_d realizations of the training dataset $\mathscr{D}_{n_d}(\mathbf{x})$. Consequently, $[x_d]$ is a realization of $[\mathbf{X}]$. Let $[\underline{x}]$ be the matrix in $\mathbb{M}_{n_x,n_d}$, whose columns are identical and equal to the empirical estimate $\underline{\mathbf{x}} \in \mathbb{R}^{n_x}$ of the mean value $\mathbf{m}_{\mathbf{X}} = E\{\mathbf{X}\} \in \mathbb{R}^{n_x}$ of random vector $\mathbf{X}$. We then have,

$$[\underline{x}] = [\underline{\mathbf{x}} \ldots \underline{\mathbf{x}}] \in \mathbb{M}_{n_x,n_d} \quad , \quad \underline{\mathbf{x}} = \frac{1}{n_d} \sum_{j=1}^{n_d} \mathbf{x}^j_d \in \mathbb{R}^{n_x}\,. \tag{7.8}$$

The empirical estimate $[\widehat{C}_{\mathbf{X}}] \in \mathbb{M}^{+0}_{n_x}$ of the covariance matrix $[C_{\mathbf{X}}] = E\{(\mathbf{X} - \mathbf{m}_{\mathbf{X}})\,(\mathbf{X} - \mathbf{m}_{\mathbf{X}})^T\} \in \mathbb{M}^{+0}_{n_x}$ of $\mathbf{X}$ is,

$$[\widehat{C}_{\mathbf{X}}] = \frac{1}{n_d - 1}\,[x_c]\,[x_c]^T \in \mathbb{M}^{+0}_{n_x} \quad , \quad [x_c] = [x_d] - [\underline{x}] \in \mathbb{M}_{n_x,n_d}\,, \tag{7.9}$$

in which $[x_c]$ is the matrix of the centered realizations of the training dataset.

(ii) Reduced normalized random matrix $[\mathbf{H}]$ *and its realization* $[\eta_d]$ *constructed by using the principal component analysis (PCA).* Let $[\mu]$ be the $(\nu \times \nu)$ diagonal matrix of the ν positive eigenvalues $\mu_1 \geq \ldots \geq \mu_\nu > 0$ of matrix $[\widehat{C}_{\mathbf{X}}] \in \mathbb{M}^{+0}_{n_x}$ defined by Eq. (7.9) and let $[\varphi] \in \mathbb{M}_{n_x,\nu}$ be the matrix whose columns are the associated eigenvectors $\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^\nu$, which are such that $[\widehat{C}_{\mathbf{X}}]\,\boldsymbol{\varphi}^\alpha = \mu_\alpha\,\boldsymbol{\varphi}^\alpha$ with $[\varphi]^T\,[\varphi] = [I_\nu]$. The reduced normalized representation $[\mathbf{X}^{(\nu)}]$ of $[\mathbf{X}]$ of order $\nu \leq n_x - 1$, is constructed using the training-dataset-based PCA of $\mathbf{X}$ and is written as

$$[\mathbf{X}^{(\nu)}] = [\underline{x}] + [\varphi]\,[\mu]^{1/2}\,[\mathbf{H}]\,, \tag{7.10}$$

in which $[\mathbf{H}] = [\mathbf{H}^1, \ldots, \mathbf{H}^{n_d}]$ is a $\mathbb{M}_{\nu,n_d}$-valued random matrix,

$$[\mathbf{H}] = [\mu]^{-1/2}[\varphi]^T\,([\mathbf{X}] - [\underline{x}])\,. \tag{7.11}$$

whose columns are n_d independent copies of a random vector $\mathbf{H}$ with values in $\mathbb{R}^\nu$. The realization $[\eta_d] = [\boldsymbol{\eta}_d^1 \ldots \boldsymbol{\eta}_d^{n_d}] \in \mathbb{M}_{\nu,n_d}$ of $[\mathbf{H}]$ is then computed by,

$$[\eta_d] = [\mu]^{-1/2}[\varphi]^T\,[x_c]\,. \tag{7.12}$$

Consequently, the training dataset $\mathscr{D}_{n_d}(\boldsymbol{\eta})$ of random vector $\mathbf{H}$ is made up of the n_d independent realizations $\boldsymbol{\eta}_d^j \in \mathbb{R}^\nu$ and we have,

$$\mathscr{D}_{n_d}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}_d^1 \ldots \boldsymbol{\eta}_d^{n_d}\}\,. \tag{7.13}$$

The value ν is classically calculated in order that the L^2- error function $\nu \mapsto \mathrm{err}_{\mathbf{X}}(\nu)$ defined by

$$\mathrm{err}_{\mathbf{X}}(\nu) = 1 - \frac{\sum_{\alpha=1}^{\nu} \mu_\alpha}{\mathrm{tr}[\widehat{C}_{\mathbf{X}}]}\,, \tag{7.14}$$

be smaller than $\varepsilon_{\mathrm{PCA}}$, in which $\mathrm{tr}[\widehat{C}_{\mathbf{X}}] = \|\,[x_c]\,\|_F^2/(n_d-1)$. The empirical estimates $\underline{\boldsymbol{\eta}}_d \in \mathbb{R}^\nu$ and $[\widehat{C}_{\mathbf{H}}] \in \mathbb{M}_\nu^+$ of the mean value and the covariance matrix of $\mathbf{H}$, calculated with the independent realizations $\boldsymbol{\eta}_d^1, \ldots, \boldsymbol{\eta}_d^{n_d}$ are such that

$$\underline{\boldsymbol{\eta}}_d = \mathbf{0}_\nu \quad , \quad [\widehat{C}_{\mathbf{H}}] = [I_\nu]\,. \tag{7.15}$$

Consequently, the components of $\mathbf{H}$ are centered and orthogonal in $L^2(\Theta, \mathbb{R}^\nu)$ and therefore uncorrelated. If $\nu < n_x - 1$, then there is a normalization and a statistical reduction. If $\nu = n_x - 1$, there is only a normalization. It should be noted that, although the components of $\mathbf{H}$ are uncorrelated, they are not statistically independent because $\mathbf{H}$ is a non-Gaussian random variable.

Often, n_x is very large and $\nu < n_d$. In this condition, matrices $[\mu]$ and $[\varphi]$ are not computed by solving the eigenvalue problem related to $[\widehat{C}_{\mathbf{X}}]$ but are computed by extracting the ν largest singular values $s_1 \geq \ldots \geq s_\nu > 0$ and the associated left orthonormal vectors $\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^\nu$ of matrix $[x_c]$ (without computing matrix $[\widehat{C}_{\mathbf{X}}]$). The singular value decomposition of $[x_c]$ is written as $[x_c] = [\varphi]\,[s]\,[\psi]^T$ and consequently, yields the eigenvalue $\mu_\alpha = s_\alpha^2/(n_d-1)$ and the associated eigenvector $\boldsymbol{\varphi}^\alpha$ for $\alpha = 1, \ldots, \nu$.

(iii) Estimating the pdf of random matrix $[\mathbf{H}]$ *with the training dataset* $\mathscr{D}_{n_d}(\boldsymbol{\eta})$ *and the nonparametric statistics*. The probability density function $[\eta] \mapsto p_{[\mathbf{H}]}([\eta])$ defined on $\mathbb{M}_{\nu,n_d}$ with respect to the measure $d[\eta] = \Pi_{\alpha=1}^{\nu}\Pi_{j=1}^{n_d}\, d\eta_{\alpha j}$ of random matrix $[\mathbf{H}]$ is estimated using the multidimensional Gaussian kernel-density estimation (KDE) [26] and the training dataset $\mathscr{D}_{n_d}(\boldsymbol{\eta})$ defined by Eq. (7.13). The modification proposed in [88] is used for constructing the Gaussian KDE, which yields

$$p_{[\mathbf{H}]}([\eta]) = \Pi_{\ell=1}^{n_d}\, p_{\mathbf{H}^\ell}(\boldsymbol{\eta}^\ell) \quad , \quad [\eta] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^{n_d}]\,, \tag{7.16}$$

$$p_{\mathbf{H}^\ell}(\boldsymbol{\eta}^\ell) = \frac{1}{n_d}\sum_{j=1}^{n_d} \frac{1}{(\sqrt{2\pi}\,\hat{s})^\nu} \exp(-\frac{1}{2\hat{s}^2}\,\|\,\frac{\hat{s}}{s_{\mathrm{SB}}}\,\boldsymbol{\eta}_d^j - \boldsymbol{\eta}^\ell\,\|^2) \quad , \quad \forall \boldsymbol{\eta}^\ell \in \mathbb{R}^\nu\,, \tag{7.17}$$

in which $\hat{s} = s_{\mathrm{SB}}\,\left(s_{\mathrm{SB}}^2 + (n_d-1)/n_d\right)^{-1/2}$ where $s_{\mathrm{SB}} = (4/(n_d(2+\nu)))^{1/(\nu+4)}$ is the Silverman bandwidth. Since $\mathbf{H}^1, \ldots, \mathbf{H}^{n_d}$ are n_d independent copies of $\mathbf{H}$, we have $p_{\mathbf{H}} = p_{\mathbf{H}^1} = \ldots = p_{\mathbf{H}^{n_d}}$. Note that with this modification, the normalization of $\mathbf{H}$ is preserved for any value of n_d,

$$E\{\mathbf{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\, p_{\mathbf{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \frac{1}{2\hat{s}^2}\,\underline{\boldsymbol{\eta}}_d = \mathbf{0}_\nu\,, \tag{7.18}$$

$$E\{\mathbf{H}\,\mathbf{H}^T\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\,\boldsymbol{\eta}^T\, p_{\mathbf{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \hat{s}^2\,[I_\nu] + \frac{\hat{s}^2}{s^2}\,\frac{(n_d-1)}{n_d}\,[\widehat{C}_{\mathbf{H}}] = [I_\nu]\,. \tag{7.19}$$

(iv) Construction of a truncated diffusion-maps basis. For preserving the concentration of the learned realizations in the region in which the points of the training dataset are concentrated, PLoM introduces an algebraic basis of the

vector space $\mathbb{R}^{n_d}$, derived from the diffusion-maps method [335, 336] that consists in constructing a Markov chain between the pairs of points of the training dataset. This algebraic basis will allow for integrating a local geometry information in the learned probability distribution of $\mathbf{H}$ for preserving its concentration property. Let $[K]$ and $[b]$ be the matrices such that, for all i and j in $\{1, \ldots, n_d\}$,

$$[K]_{ij} = \exp\{-(4\,\varepsilon_{\text{DM}})^{-1} \|\boldsymbol{\eta}_d^i - \boldsymbol{\eta}_d^j\|^2\} \quad , \quad [b]_{ij} = \delta_{ij}\, b_i \quad , \quad b_i = \sum_{j=1}^{n_d} [K]_{ij}\,, \tag{7.20}$$

in which $\varepsilon_{\text{DM}} > 0$ is a smoothing parameter. Note that this kernel is well adapted to the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}$ that is centered and with a covariance matrix that is equal to $[I_\nu]$. The eigenvalues $\kappa_1, \ldots, \kappa_{n_d}$ and the associated eigenvectors $\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^{n_d}$ of the right-eigenvalue problem $[\mathbb{P}]\,\boldsymbol{\psi}^\alpha = \kappa_\alpha\,\boldsymbol{\psi}^\alpha$ are such that $1 = \kappa_1 > \kappa_2 \geq \ldots \geq \kappa_{n_d}$ and are computed by solving the generalized eigenvalue problem for symmetric matrices with $[b]$ a positive-definite diagonal matrix, $[K]\,\boldsymbol{\psi}^\alpha = \kappa_\alpha\,[b]\,\boldsymbol{\psi}^\alpha$ with $\langle [b]\,\boldsymbol{\psi}^\alpha, \boldsymbol{\psi}^\beta \rangle = \delta_{\alpha\beta}$. It can easily be seen that the largest eigenvalue is $\kappa_1 = 1$ and the associated eigenvector $\boldsymbol{\psi}^1$ is a vector whose all its components are equal. For a given integer $\zeta \geq 0$, the diffusion-maps basis $\{\mathbf{g}^1, \ldots, \mathbf{g}^\alpha, \ldots, \mathbf{g}^{n_d}\}$ is a vector basis of $\mathbb{R}^{n_d}$ defined by $\mathbf{g}^\alpha = \kappa_\alpha^\zeta\,\boldsymbol{\psi}^\alpha$. For a given integer m, the truncate diffusion-maps basis of order m is defined as the family $\{\mathbf{g}^1, \ldots, \mathbf{g}^m\}$ that is represented by the matrix $[g_m]$ such that $[g_m] = [\mathbf{g}^1 \ldots \mathbf{g}^m] \in \mathbb{M}_{n_d,m}$, with $\mathbf{g}^\alpha = (g_1^\alpha, \ldots, g_{n_d}^\alpha)$ and $[g_m]_{\ell\alpha} = g_\ell^\alpha$. This basis depends on two parameters, ε_{DM} and m, which have to be identified. It is proven in [236], that the PLoM method does not depend on ζ that can therefore be chosen to 0. We have to find the optimal value $m_{\text{opt}} \leq n_d$ of m and the smallest value $\varepsilon_{\text{opt}} > 0$ of ε_{DM} such that (see [238])

$$1 = \kappa_1 > \kappa_2(\varepsilon_{\text{opt}}) \simeq \ldots \simeq \kappa_{m_{\text{opt}}}(\varepsilon_{\text{opt}}) \gg \kappa_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}}) \geq \ldots \geq \kappa_{n_d}(\varepsilon_{\text{opt}}) > 0\,, \tag{7.21}$$

with an amplitude jump equal to an order of magnitude (a factor 10 as demonstrated in [236]) between $\kappa_{m_{\text{opt}}}(\varepsilon_{\text{opt}})$ and $\kappa_{m_{\text{opt}}+1}(\varepsilon_{\text{opt}})$. A further in-depth analysis makes it possible to state the following algorithm to estimate ε_{opt} and m_{opt}. Let $\varepsilon_{\text{DM}} \mapsto \text{Jump}(\varepsilon_{\text{DM}})$ be the function on $]0, +\infty[$ defined by $\text{Jump}(\varepsilon_{\text{DM}}) = \kappa_{m_{\text{opt}}+1}(\varepsilon_{\text{DM}})/\kappa_2(\varepsilon_{\text{DM}})$. Finally the last version of the algorithm proposed in [238] is the following:
- set the value of m to $m_{\text{opt}} = \nu + 1$;
- identify the smallest possible value ε_{opt} of ε_{DM} in order that $\text{Jump}(\varepsilon_{\text{opt}}) \leq 0.1$ and such that Equation (7.21) be verified.

(v) Generator of learned realizations of random matrix $[\mathbf{H}]$. The algebraically independent vectors $\mathbf{g}^1, \ldots, \mathbf{g}^m \in \mathbb{R}^{n_d}$ span a subspace of $\mathbb{R}^{n_d}$ that characterizes, for the optimal values m_{opt} and ε_{opt} of m and ε_{DM}, the local geometry structure of dataset $\{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\}$. So the PLoM method introduces the $\mathbb{M}_{\nu,n_d}$-valued random matrix $[\mathbf{H}_m] = [\mathbf{Z}_m]\,[g_m]^T$ with $m < n_d$, corresponding to a data-reduction representation of random matrix $[\mathbf{H}]$, in which $[\mathbf{Z}_m]$ is a $\mathbb{M}_{\nu,m}$-valued random matrix. First of all, a Markov chain Monte Carlo (MCMC) generator for random matrix $[\mathbf{H}]$ is constructed [337] in the class of Hamiltonian Monte Carlo methods [14, 337, 40], by solving a nonlinear Itô stochastic differential equation (ISDE) that corresponds to a stochastic nonlinear dissipative Hamiltonian dynamical system, for which the probability distribution $p_{[\mathbf{H}]}([\boldsymbol{\eta}])\,d[\eta]$ of random matrix $[\mathbf{H}]$, defined by Eqs. (7.16) and (7.17), is the marginal probability distribution of the unique invariant measure (see [230, 236]). Note that, due to the dissipation, this ISDE allows for removing the transient part to rapidly attain the stationary response associated with the invariant measure. Then, the generator of the learned realizations is obtained by constructing the reduced-order ISDE related to the random matrix $[\mathbf{Z}_m]$ by projecting the ISDE on the subspace generated by $[g_m]^T$. This generator is defined by the following Theorem 1 that is proven in [236].

Theorem 1 (MCMC generator for the learned realizations). *Let m be fixed to the value $m_{\text{opt}} < n_d$ and ε_{DM} to ε_{opt}. Let $\{([\mathcal{Z}(r)], [\mathcal{Y}(r)]), r \in \mathbb{R}^+\}$ be the $\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m}$-valued stochastic process, satisfying, for $r > 0$ and with initial conditions for $r = 0$, the reduced-order ISDE,*

$$d[\mathcal{Z}(r)] = [\mathcal{Y}(r)]\,dr\,, \tag{7.22}$$

$$d[\mathcal{Y}(r)] = [\mathscr{L}([\mathcal{Z}(r)])]\,dr - \frac{1}{2} f_0\,[\mathcal{Y}(r)]\,dr + \sqrt{f_0}\,[d\mathbf{W}^{\text{wien}}(r)]\,[a_m]\,, \tag{7.23}$$

$$[\mathcal{Y}(0)] = [y_0]\,[a_m] \quad , \quad [\mathcal{Z}(0)] = [\eta_d]\,[a_m] \quad , \quad a.s.\,, \tag{7.24}$$

in which $[a_m] = [g_m]\,([g_m]^T\,[g_m])^{-1} \in \mathbb{M}_{n_d,m}$ *and where the matrix* $[y_0] \in \mathbb{M}_{\nu,n_d}$ *is any realization of a normalized Gaussian random matrix* $[Y_0]$*, independent of* $\mathbf{W}^{\text{wien}}$*, whose entries* $\{[Y_0]_{\alpha j}\}_{\alpha j}$ *are* $\nu \times n_d$ *independent normalized Gaussian real-valued random variables. In Eq.* (7.23)*,* $[\mathscr{L}([\mathcal{Z}(r)])] = [L([\mathcal{Z}(r)]\,[g_m]^T)]\,[a_m]$ *such that,* $\forall\,[u] = [\mathbf{u}^1 \ldots \mathbf{u}^{n_d}]$ *in* $\mathbb{M}_{\nu,n_d}$ *with* $\mathbf{u}^\ell \in \mathbb{R}^\nu$*, matrix* $[L([u])] \in \mathbb{M}_{\nu,n_d}$ *is such that, for* $k \in \{1,\ldots,\nu\}$ *and* $\ell \in \{1,\ldots,n_d\}$*,*

$$[L([u])]_{k\ell} = \frac{\{\boldsymbol{\nabla}_{\mathbf{u}^\ell}\, p_{\mathbf{H}^\ell}(\mathbf{u}^\ell)\}_k}{p_{\mathbf{H}^\ell}(\mathbf{u}^\ell)} = \frac{1}{\hat{s}^2}\,\frac{\sum_{j=1}^{n_d}\{\frac{\hat{s}}{s_{\text{SB}}}\,\boldsymbol{\eta}_d^j - \mathbf{u}^\ell\}_k \exp(-\frac{1}{2\hat{s}^2}\,\|\,\frac{\hat{s}}{s_{\text{SB}}}\,\boldsymbol{\eta}_d^j - \mathbf{u}^\ell\,\|^2)}{\sum_{j=1}^{n_d}\exp(-\frac{1}{2\hat{s}^2}\,\|\,\frac{\hat{s}}{s_{\text{SB}}}\,\boldsymbol{\eta}_d^j - \mathbf{u}^\ell\,\|^2)}\,. \tag{7.25}$$

Parameter f_0*, such that* $0 < f_0 < 4/\hat{s}$*, allows the transient part to be controlled (a common value is* $f_0 = 4$ *knowing that* $\hat{s} < 1$*) and* $[\mathbf{W}^{\text{wien}}]$ *is the* $\mathbb{M}_{\nu,n_d}$*-valued normalized Wiener stochastic process. Then, there is a unique stochastic solution that is a second-order diffusion stochastic process, which is asymptotic, for* $r \to +\infty$*, to a stationary stochastic process and we have* $[\mathbf{Z}_{\text{learn}}] = \lim_{r\to+\infty}[\mathcal{Z}(r)]$ *in probability distribution, which allows for generating* $n_{\text{MC}} \gg n_d$ *learned realizations* $[z^1_{\text{learn}}],\ldots,[z^{n_{\text{MC}}}_{\text{learn}}]$*. The* n_{MC} *learned realizations generated by the probability distribution of* $[\mathbf{H}_{\text{learn}}] = [\mathbf{Z}_{\text{learn}}]\,[g_m]^T$ *are then computed by*

$$[\boldsymbol{\eta}^\ell_{\text{learn}}] = [z^\ell_{\text{learn}}]\,[g_m]^T \quad , \quad \ell = 1,\ldots,n_{\text{MC}}\}\,, \tag{7.26}$$

and the n_{MC} *learned realizations of the* $\mathbb{M}_{n_x,n_d}$*-valued random variable* $[\mathbf{X}_{\text{learn}}]$ *are then deduced from Eq.* (7.10)*,*

$$[\mathbf{x}^\ell_{\text{learn}}] = [\underline{x}] + [\varphi]\,[\mu]^{1/2}\,[\boldsymbol{\eta}^\ell_{\text{learn}}] \quad , \quad \ell = 1,\ldots,n_{\text{MC}}\} \tag{7.27}$$

The reshaping of matrix $[\boldsymbol{\eta}^\ell_{\text{learn}}]$ *(resp.* $[\mathbf{x}^\ell_{\text{learn}}] \in \mathbb{M}_{n_x,n_d}$*) allows for obtaining* $n_{\text{learn}} = n_{\text{MC}} \times n_d$ *additional realizations* $\{\boldsymbol{\eta}^{\ell'}_{\text{learn}}, \ell' = 1,\ldots,n_{\text{learn}}\}$ *of* $\mathbf{H}_{\text{learn}}$ *(resp.* $\{\mathbf{x}^{\ell'}_{\text{learn}}, \ell' = 1,\ldots,n_{\text{learn}}\}$ *of* $\mathbf{X}_{\text{learn}}$*).*

The reduced-order ISDE defined by Eqs. (7.22) to (7.24) is solved by using the Störmer-Verlet algorithm [338, 339] detailed in Appendix C, which yields an efficient and accurate MCMC algorithm. The stationary response of this reduced-order ISDE is solved in parallel computation, which allows for strongly decreasing the elapsed time on a multicore computer. Efficient values of the algorithm parameters are given at the end of Appendix C. The PLoM methodology allows the concentration of the learned probability distribution to be preserved as shown by the following theorem proven in [236].

Theorem 2 (Concentration of the learned probability distribution). *For all* m *in* $\{1,\ldots,n_d\}$*, random matrix* $[\mathbf{H}_{\text{learn}}]$ *is rewritten as* $[\mathbf{H}^{n_d}_m]$*. Let* $d^2_{n_d}(m)$ *be the square of the* $L^2(\Theta,\mathbb{M}_{\nu,n_d})$*-distance between random matrix* $[\mathbf{H}^{n_d}_m]$ *and matrix* $[\eta_d]$ *that is defined by the* n_d *points of the training dataset* $\mathscr{D}_{n_d}(\boldsymbol{\eta})$*, such that*

$$d^2_{n_d}(m) = E\{\|\,[\mathbf{H}^{n_d}_m] - [\eta_d]\,\|^2_F\}/E\{\|\,[\eta_d]\,\|^2_F\}\,. \tag{7.28}$$

Then there exists an optimal value m_{opt} *with* $1 < m_{\text{opt}} < n_d$ *such that*

$$\min_{m\in\mathscr{M}_{\text{opt}}} d^2_{n_d}(m) \;<\; d^2_{n_d}(n_d) = 1 + n_d/(n_d-1)\,, \tag{7.29}$$

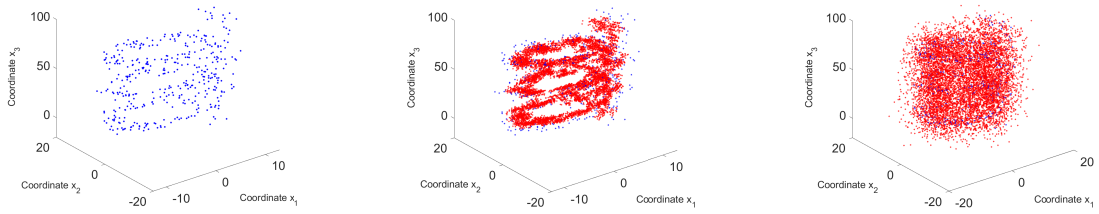
in which $\mathscr{M}_{\text{opt}} = \{m_{\text{opt}},\ldots,n_d-1\}$*. This result shows that the PLoM method is better than the usual one corresponding to* $d^2_{n_d}(n_d)$*.*

7.2.2. Illustrations showing the capability of the PLoM algorithm

Three illustrations are presented. The first two show how PLoM preserves the concentration of the learned probability distribution, on the one hand on a very simple example [230] in dimension 3 and on the other hand for experimental data [212] in dimension 35. The third illustration is devoted to a basic difficult constrained nonconvex optimization problem with uncertainties [211]. Similar optimization problems have been analyzed with PLoM for much more complicated engineering problems such as the design optimization under uncertainties of a mesoscale implant in biological tissues [213], the optimal well-placement [212], the design optimization of a scramjet under uncertainty [215], the optimization of wake steering in wind farms [329], the probabilistic learning and updating of a digital twin for composite material systems [219], the detuning optimization of detuned bladed disks with uncertain

mistuning [218].

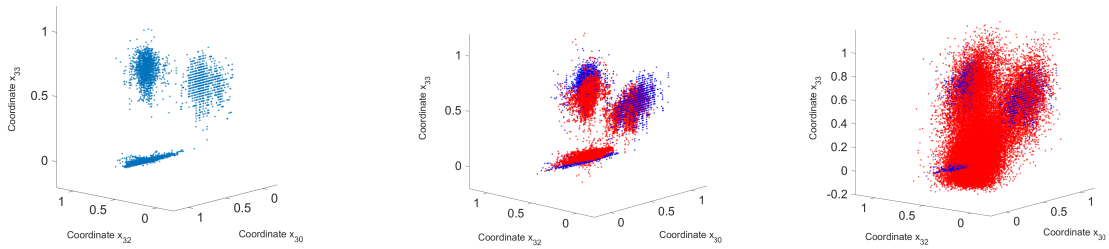
(i) Manifolds as a helical in 3D Euclidean space [230]. Using the notation introduced in Section 7.1, we have $n_w = 2$, $n_q = 1$, $n_x = n_w + n_q = 3$, $\nu = n_x = 3$, and the training dataset has $n_d = 400$ points shown in Fig. 4-(a). The PLoM algorithm is used with $m_{\text{opt}} = 4$ for generating 8 000 learned realizations that are shown in Fig. 4-(b). Fig. 4-(c) displays 8 000 points generated with a resampling from the training dataset, using a classical MCMC algorithm. It can be seen that the PLoM algorithm has preserved the concentration of the learned probability distribution while the concentration has been lost with a classical MCMC algorithm.



(a) Training dataset (b) PLoM algorithm (c) Classical MCMC algorithm

Figure 4: Concentration-loss analysis: 400 blue points of the training dataset (a), (b) and (c); 8 000 red points for the learned realizations with the PLoM algorithm (b), and for the resampling with a classical MCMC algorithm (c). [From [230]].

(ii) Analysis of a petro-physics experimental database [212]. We have $n_x = 35$, $\nu = 32$. Fig. 5 is relative to coordinates x_{30}, x_{32}, and x_{33}. The training dataset has $n_d = 13\,056$ points shown in Fig. 5-(a). The PLoM algorithm is used with $m_{\text{opt}} = 33$ for generating 39 168 learned realizations that are shown in Fig. 5-(b). Fig. 5-(c) displays 39 168 points generated with a resampling from the training dataset, using a classical MCMC algorithm. Again, it can be seen that the PLoM algorithm has preserved the concentration of the learned probability distribution while the concentration has been lost with the use of a classical MCMC algorithm.



(a) Training dataset (b) PLoM algorithm (c) Classical MCMC algorithm

Figure 5: Concentration-loss analysis: 13 056 blue points of the training dataset (a), (b) and (c); 39 168 red points for the learned realizations with the PLoM algorithm (b), and for the resampling with a classical MCMC algorithm (c). [From [230]].

(iii) Optimization under uncertainties (OUU) using a limited number of function evaluations of a stochastic computational model [211].

(iii-1) OUU formulation in a general framework. We consider the following optimization problem under uncertainties, related to a stochastic computational model whose parameterization is defined in Section 7.1,

$$\mathbf{w}^{\text{opt}} = \arg \min_{\mathbf{w} \in \mathscr{C}_w \,,\, \mathbf{b}(\mathbf{w}) < 0} g(\mathbf{w}) \,. \tag{7.30}$$

The design parameter, $\mathbf{w} \in \mathscr{C}_w \subset \mathbb{R}^{n_w}$, is the control parameter, which is modelled by a $\mathbb{R}^{n_w}$-valued random variable $\mathbf{W}$ for which the support of its prior probability distribution $P_{\mathbf{W}}(d\mathbf{w})$ is $\mathscr{C}_w$. The objective function is written as $\mathbf{w} \mapsto g(\mathbf{w}) = E\{\mathscr{G}(\mathbf{w}, \mathbf{U})\}$ from $\mathbb{R}^{n_w}$ into $\mathbb{R}$ in which $(\mathbf{w}, \mathbf{u}) \mapsto \mathscr{G}(\mathbf{w}, \mathbf{u})$ is a measurable function from $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ into $\mathbb{R}$ and where the uncontrolled parameter $\mathbf{U}$ is a $\mathbb{R}^{n_u}$-valued random variable, independent of $\mathbf{W}$, whose prior probability distribution $P_{\mathbf{U}}(d\mathbf{u})$ is given. The nonlinear constraints are defined by the function $\mathbf{w} \mapsto \mathbf{b}(\mathbf{w}) = E\{\mathscr{B}(\mathbf{w}, \mathbf{U})\}$ from $\mathbb{R}^{n_w}$ into $\mathbb{R}^{n_b}$ with $n_b \geq 1$, in which $(\mathbf{w}, \mathbf{u}) \mapsto \mathscr{B}(\mathbf{w}, \mathbf{u})$ is a measurable function from $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ into $\mathbb{R}^{n_b}$. Let G be the real-valued random variable defined by $G = \mathscr{G}(\mathbf{W}, \mathbf{U})$ and let $\mathbf{B}$ be the $\mathbb{R}^{n_b}$valued random variable such that $\mathbf{B} = \mathscr{B}(\mathbf{W}, \mathbf{U})$. The prior probability distributions of $\mathbf{W}$ and $\mathbf{U}$ are used for generating n_d independent realizations $\{(\mathbf{w}^j, \mathbf{u}^j), j = 1, \ldots, n_d\}$. Let n_x be $n_w + 1 + n_b$. The parameterized computational model is used for constructing the training dataset of the $\mathbb{R}^{n_x}$-valued random variable $\mathbf{X} = (\mathbf{W}, G, \mathbf{B})$. The n_d independent realizations of $\mathbf{X}$ are $\{\mathbf{x}^j = (\mathbf{w}^j, g^j, \mathbf{b}^j), j = 1, \ldots, n_d\}$ with $g^j = \mathscr{G}(\mathbf{w}^j, \mathbf{u}^j)$ and $\mathbf{b}^j = \mathscr{B}(\mathbf{w}^j, \mathbf{u}^j)$. The PLoM algorithm allows for generating the learned dataset $(\mathbf{w}^\ell_{\text{learn}}, g^\ell_{\text{learn}}, \mathbf{b}^\ell_{\text{learn}}) = \mathbf{x}^\ell_{\text{learn}}, \ell = 1, \ldots, N_{\text{learn}}$ with $N_{\text{learn}} \gg n_d$. For all $\mathbf{w}$ in $\mathscr{C}_w$, the conditional statistics

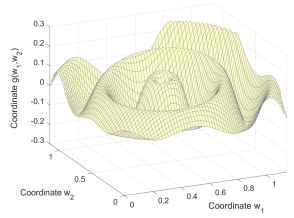
$$g(\mathbf{w}) \simeq E\{G_{\text{learn}} \,|\, \mathbf{W}_{\text{learn}} = \mathbf{w}\} \quad , \quad \mathbf{b}(\mathbf{w}) \simeq E\{\mathbf{B}_{\text{learn}} \,|\, \mathbf{W}_{\text{learn}} = \mathbf{w}\}\,, \tag{7.31}$$

are estimated using explicit algebraic formulas based on the joint probability density functions $p_{G_{\text{learn}}, \mathbf{W}_{\text{learn}}}$ and $p_{\mathbf{B}_{\text{learn}}, \mathbf{W}_{\text{learn}}}$ that are estimated using the Gaussian KDE and the learned realizations (see Section 5 of [211]). The nonconvex optimization problem defined by Eq. (7.30) can be solved using a random search algorithm (such as a genetic algorithm) and the conditional statistics defined by Eq. (7.31) are estimated for any $\mathbf{w}$ without using the parameterized stochastic computational model. Note that these conditional statistics can be viewed as a statistical surrogate model whose evaluations are very fast.

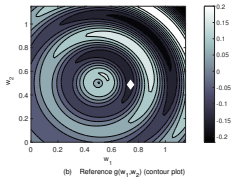
(iii-2) Application. The problem is defined by Fig. 6. We have $n_w = 2$, $n_b = 4$, $n_u = 4$ and $n_x = n_w + 1 + n_b = 7$. The 3D plot of the graph of the reference cost function $\mathbf{w} \mapsto g(\mathbf{w})$ is shown in Fig. 6-(a) and its 2D contour plot in Fig. 6-(b). The 3D plot of the graphs of the four components b_1, b_2, b_3 and b_4 of function $\mathbf{w} \mapsto \mathbf{b}(\mathbf{w})$ are shown in Figs. 6-(c) to (f). The additional details of the construction of the stochastic computational model can be found in [211]. Since the optimal solution of this OUU is not exactly known, a reference optimal solution $\mathbf{w}^{\text{opt}}_{\text{ref}}$ has been constructed by using the Monte Carlo simulation method with 10 000 realizations, and yields $\mathbf{w}^{\text{opt}}_{\text{ref}} = (0.74, 0.49)$, $g(\mathbf{w}^{\text{opt}}_{\text{ref}}) = -0.123$. For this reference optimal solution, the four constraints are active. In Fig. 6-(b), this reference optimal solution is located in the white diamond. For the probabilistic learning construction, the training set is generated with $n_d = 900$ independent realizations. The 2D contour plot of cost function g, estimated with this 900 points, are displayed in Fig. 7-(a). It can seen that the generated image is very different from the one corresponding to the reference (see Fig. 6-(b)). Moreover, the solution of the optimization problem constructed with this training set is located in the white square that is located at the lower edge of the image and which is therefore very far from the optimal reference solution represented by the white diamond. The image of the objective function synthesized using $n_{\text{learn}} = 9\,000$ learned realizations from the 900-points training dataset is shown in Fig. 7-(b). This result is quite remarkable and the quality obtained for the learned objective function as well as for the learned constraints functions makes it possible to build a very good approximation of the optimal solution $\mathbf{w}^{\text{opt}} = (0.70, 0.49)$, $g(\mathbf{w}^{\text{opt}}_{\text{ref}}) = -0.112$ (white disc). compared to the optimal reference solution $\mathbf{w}^{\text{opt}}_{\text{ref}}$ (white diamond). The results obtained for $n_{\text{learn}} = 90\,000$ learned realizations (see Fig. 7-(c)) confirms this analysis.

7.2.3. Normalization by constraining the second-order moments of the components of $\mathbf{H}$

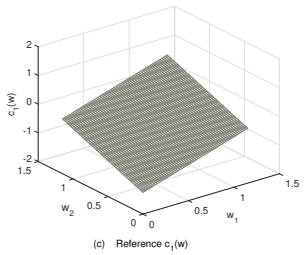
The normalization conditions defined by Eqs. (7.18) and (7.19) should be preserved for $\mathbf{H}_{\text{learn}}$ although these conditions have not been introduced as constraints in the PLoM algorithm presented in Section 7.2. Referring to the applications that have been addressed, in general, the mean value of $\mathbf{H}_{\text{learn}}$, which is estimated using the learned realizations $\{\boldsymbol{\eta}^\ell_{\text{learn}}, \ell = 1, \ldots, n_{\text{learn}}\}$, is sufficiently close to zero ($\ll 1$) and the estimate of the covariance matrix of $\mathbf{H}_{\text{learn}}$ is sufficiently close to $[I_\nu]$. However, sometimes, although this covariance matrix stays close to a diagonal matrix, its diagonal entries can be lower than 1. This situation has been encountered for small values of ν, a few units, even ten, but has never been encountered for larger values of ν, a few tens, even hundreds. In the case where the diagonal entries of this estimated covariance matrix become smaller than 1, some constraints can be imposed into the PLoM algorithm by using the Kullback-Leibler divergence minimum principle as proposed in [240], which we briefly summarize below.



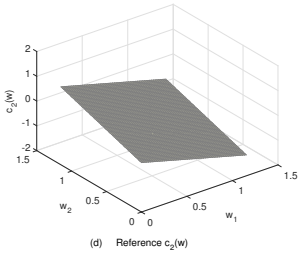
(a) 3D plot of the reference objective function $\mathbf{w} \mapsto g(\mathbf{w})$.



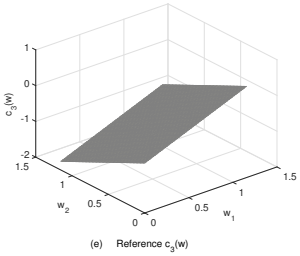
(b) 2D contour plot of the reference objective function $\mathbf{w} \mapsto g(\mathbf{w})$.



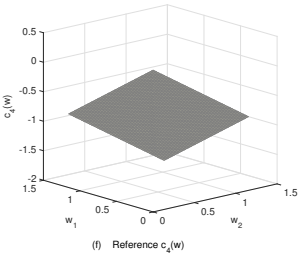
(c) Constraint function $\mathbf{w} \mapsto b_1(\mathbf{w})$



(d) Constraint function $\mathbf{w} \mapsto b_2(\mathbf{w})$

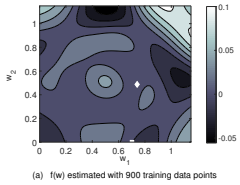


(e) Constraint function $\mathbf{w} \mapsto b_3(\mathbf{w})$

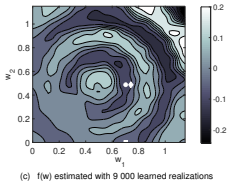


(f) Constraint function $\mathbf{w} \mapsto b_4(\mathbf{w})$

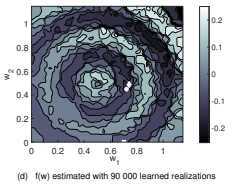
Figure 6: Reference: graph of the objective function $\mathbf{w} \mapsto g(\mathbf{w})$ in 3D (a) and in 2D contour plot (b). The reference optimal solution $\mathbf{w}_{\text{ref}}^{\text{opt}}$ is the white diamond plotted in (b). Graphs of the constraint functions $\mathbf{w} \mapsto b_1(\mathbf{w})$ in (c), $b_2(\mathbf{w})$ in (d), $b_3(\mathbf{w})$ in (e), and $b_4(\mathbf{w})$ in (f). [From [211]].



(a) Objective function g estimated with $n_d = 900$ points of the training dataset



(b) Objective function g estimated with $n_{\text{learn}} = 9\,000$ learned realizations



(c) Objective function g estimated with $n_{\text{learn}} = 90\,000$ learned realizations

Figure 7: 2D contour plot of the objective function $\mathbf{w} \mapsto g(\mathbf{w})$ estimated with $n_d = 900$ points of the training dataset (a), with $n_{\text{learn}} = 9\,000$ learned realizations (b), and with $n_{\text{learn}} = 90\,000$ learned realizations (c). The white diamond is the reference optimal solution $\mathbf{w}_{\text{ref}}^{\text{opt}}$, the white square is the optimal solution estimated with the training dataset, and, in figures (b) and (c), the white disc is the optimal solution $\mathbf{w}^{\text{opt}}$ estimated with the learned realizations. [From [211]].

The constraints $\{E\{(H_k)^2\} = 1, k = 1, \ldots, \nu\}$ are rewritten as $E\{\mathbf{g}^c(\mathbf{H})\} = \mathbf{b}^c$ on $\mathbb{R}^\nu$, in which $\mathbf{g}^c$ and $\mathbf{b}^c$ are such that $g_k^c(\mathbf{H}) = (H_k)^2$ and $b_k^c = 1$ for k in $\{1, \ldots, \nu\}$. To take into account these constraints in the PLoM algorithm presented in Section 7.2, Eq. (7.25) is replaced, for $k \in \{1, \ldots, \nu\}$ and $\ell \in \{1, \ldots, \nu\}$, by the following one,

$$[L([u])]_{k\ell} = \frac{\{\boldsymbol{\nabla}_{\mathbf{u}^\ell}\, p_{\mathbf{H}^\ell}(\mathbf{u}^\ell)\}_k}{p_{\mathbf{H}^\ell}(\mathbf{u}^\ell)} - 2\lambda_k\, u_k^\ell \quad , \quad [u] = [\mathbf{u}^1 \ldots \mathbf{u}^{n_d}] \in \mathbb{M}_{\nu,n_d}\,. \tag{7.32}$$

In Eq. (7.32), $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_\nu)$ is a Lagrange multiplier with values in an admissible set $\mathscr{C}_\lambda \subset \mathbb{R}^\nu$, which allows the constraint $E\{\mathbf{g}^c(\mathbf{H})\} - \mathbf{b}^c = \mathbf{0}$ to be imposed. This optimal value $\boldsymbol{\lambda}^{\text{sol}}$ of Lagrange multiplier $\boldsymbol{\lambda}$ is computed using the formulation defined in Section 3.2 and the iterative algorithm given in Appendix B (in which $\mathbf{w} \in \mathbb{R}^{n_w}$ and the $\mathbb{R}^{n_w}$-valued random variables $\mathbf{W}$ and $\mathbf{W}_\lambda$ must be replaced by $\boldsymbol{\eta} \in \mathbb{R}^\nu$ and the $\mathbb{R}^\nu$-valued random variables $\mathbf{H}$ and $\mathbf{H}_\lambda$, respectively). Applications using PloM under these constraints can be found in [240, 238].

7.2.4. Extension to PLoM with partition

Also as part of improving PLoM algorithm, there are some applications for which the number n_d of points in the training dataset is very small and for which the optimal dimension $m_{\rm opt}$ of the diffusion-map basis is very close to this number, that is to say, $m_{\rm opt} = \nu + 1 \lesssim n_d$. In this case, PLoM may not be more efficient than a standard MCMC algorithm yielding a loss of concentration of the learned probability distribution. For these challenging cases, one possible way to improve the PLoM algorithm is to perform a partition of the random vector $\mathbf{H}$, into statistically independent groups in a non-Gaussian framework (see [238]). In this manner, statistical knowledge about the dataset, beyond its localization on a manifold, is relied upon to enhance information extraction and representation.

(i) Construction of the optimal partition of $\mathbf{H}$. From the training set $\{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\}$, an optimal partition of the non-Gaussian normalized $\mathbb{R}^\nu$-valued random variable $\mathbf{H} = (H_1, \ldots, H_\nu)$ is performed using the algorithm proposed in [340]. Such a partition is composed of n_p groups consisting in n_p mutually independent random vectors $\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}$. Since $\mathbf{H}$ is a normalized random vector (zero mean vector and covariance matrix equal to $[I_\nu]$), for $i = 1, \ldots, n_p$, $\mathbf{Y}^i$ is a normalized $\mathbb{R}^{\nu_i}$-valued random variable $\mathbf{Y}^i = (Y_1^i, \ldots, Y_{\nu_i}^i) = (H_{r_1^i}, \ldots, H_{r_{\nu_i}^i})$ in which $1 \leq r_1^i < r_2^i < \ldots < r_{\nu_i}^i \leq \nu$, with $\nu = \nu_1 + \ldots + \nu_{n_p}$, and where $Y_k^i = H_{r_k^i}$. Random vector $\mathbf{Y}^i$ is non-Gaussian and such that the estimate of its mean vector and covariance matrix is $\mathbf{0}_{\nu_i}$ and $[I_{\nu_i}]$, respectively. We then have $\mathbf{H} = perm(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})$ in which *perm* is the permutation operator acting on the components of vector $\widetilde{\mathbf{H}} = (\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})$ in order to reconstitute $\mathbf{H} = perm(\widetilde{\mathbf{H}})$. For each group i, the training set is represented by the matrix $[\eta_d^i] \in \mathbb{M}_{\nu_i, n_d}$ whose columns are the n_d realizations $\{\boldsymbol{\eta}_d^{i,j}, j = 1, \ldots, n_d\}$ of the $\mathbb{R}^{\nu_i}$-valued random variable $\mathbf{Y}^i$, which are deduced from an adapted extraction (due to the permutations) of the components of vectors $\{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\}$. The partition is identified by constructing the function $i_{\rm ref} \mapsto \tau_\nu(i_{\rm ref})$ of the mutual information and then by deducing the optimal level $i_{\rm ref}^{\rm opt}$. The mutual information $i^\nu(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})$ between the random vectors $\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}$ is defined (see for instance [50, 51]) by

$$i^\nu(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}) = -E\left\{ \log\left(\frac{p_{\mathbf{Y}^1}(\mathbf{Y}^1) \times \ldots \times p_{\mathbf{Y}^{n_p}}(\mathbf{Y}^{n_p})}{p_{\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}}(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})}\right)\right\},$$

in which the conventions $0 \log(0/a) = 0$ for $a \geq 0$ and $b \log(b/0) = +\infty$ for $b > 0$ are used, where $p_{\mathbf{Y}^i}$ is the pdf of $\mathbf{Y}^i$, and where $p_{\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}}$ is the joint pdf of $\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}$. Let $\mathbf{G}$ be the Gaussian second-order centered $\mathbb{R}^\nu$-valued random vector for which its covariance matrix is $[I_\nu]$. Consequently, the components of $\mathbf{G}$ are mutually independent. Applying to $\mathbf{G}$ the same partition that the one defined for $\mathbf{Y}$, we can write $\mathbf{G} = (\mathbf{G}^1, \ldots, \mathbf{G}^{n_p})$, and its mutual information is $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^{n_p})$. Let $i_{\rm ref} \geq 0$ be any fixed real value of the mutual information for two real-valued random variables (see Section 3.5.3 of [340]). Let $\tau_\nu(i_{\rm ref})$ be the mutual information defined by

$$\tau_\nu(i_{\rm ref}) = 1 - \frac{i^\nu(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})}{i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^{n_p})}.$$

The optimal level $i_{\rm ref}^{\rm opt}$ is such that

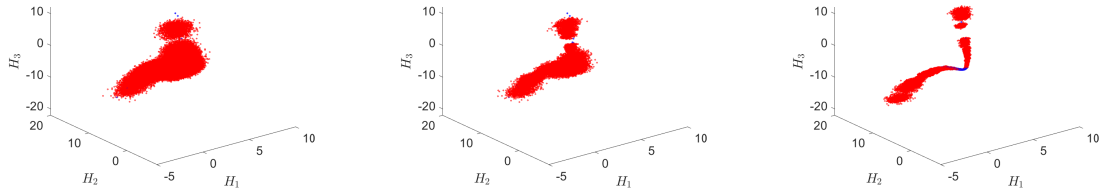
$$i_{\rm ref}^{\rm opt} = \inf_{i_{\rm ref}^*} \{\, i_{\rm ref}^* = \arg\max_{i_{\rm ref} \geq 0} \ \tau_\nu(i_{\rm ref}) \,\}.$$

For calculating the mutual information, the pdf are estimated by using the multidimensional Gaussian KDE method with the points of the training set. For each given $i_{\rm ref}$, the partition is constructed using a graph theory algorithm (see [340]).

(ii) Use of the PLoM for each independent group. Let i be fixed in $\{1, \ldots, n_p\}$. The PLoM algorithm presented in Section 7.2) is applied to the $\mathbb{R}^{\nu_i}$-valued random variable $\mathbf{Y}^i$ of the optimal partition $\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p}$ of $\mathbf{H} = perm(\mathbf{Y}^1, \ldots, \mathbf{Y}^{n_p})$ and allows for generating $n_{\rm MC}$ learned realizations $\{[\boldsymbol{\eta}_{\rm learn}^{i,\ell}], \ell = 1, \ldots, n_{\rm MC}\}$ for each random variable $[\mathbf{Y}_{\rm learn}^i]$ with $m_{i,\rm opt} = \nu_i + 1$ and $\varepsilon_{i,\rm opt}$ calculated as explained in Section 7.2.1-(iv). When ν_i is very small, it can be necessary to apply Section 7.2.3 to preserve the normalization of $\mathbf{Y}^i$ by constraining the second-order moments of its components.

(iii) Illustration. This illustration is Application 1 from [238]. We consider the normalized non-Gaussian $\mathbb{R}^\nu$-valued

random variable $\mathbf{H} = (H_1, \ldots, H_\nu)$ with $\nu = 60$. The training dataset his generated with $n_d = 1\,200$ independent realizations $\{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\}$. Using the training dataset, the partition algorithm presented in *(i)* above is applied and yields $n_p = 3$ groups with $\mathbf{Y}^1 = (H_1, \ldots, H_{10})$, $\mathbf{Y}^2 = (H_{11}, \ldots, H_{30})$, and $\mathbf{Y}^3 = (H_{31}, \ldots, H_{60})$ (consequently, $\nu_1 = 10$, $\nu_2 = 20$, and $\nu_3 = 30$). The learned dataset is generated (1) with a classical MCMC algorithm (no PLoM), (2) with the PLoM algorithm without partition, and (3) with the PLoM algorithm with partition. For these three cases, $n_{\text{MC}} = 1\,000$ yielding $n_{\text{learn}} = 1\,200\,000$ realizations $\{\boldsymbol{\eta}_{\text{learn}}^{\ell'}, \ell' = 1, \ldots, n_{\text{learn}}\}$ ($n_{\text{learn}} = n_{\text{MC}} \times n_d$). The concentration of the learned probability distribution is quantified by using Theorem 2 and yields $d^2_{n_d}(n_d) \simeq 2$ for case (1), $d^2_{n_d}(m_{\text{opt}}) = 0.094$ with $m_{\text{opt}} = 61$ for case (2), and $d^2_{wg,n_d}(\mathbf{m}_{\text{opt}}) = 0.016$ with $\mathbf{m}_{\text{opt}} = (11, 21, 31)$ for case (3). Figure 8 shows the clouds of the learned realizations $\{\eta_{\text{learn}}^{\ell'}, \ell' = 1, \ldots, n_{\text{learn}}\}$ for the first three components (H_1, H_2, H_3) of the 60 components of $\mathbf{H}$, which have been generated by cases (1), (2), and (3). It can be seen that the classical MCMC yields a scattering of the learned realizations with a loss of the concentration. The PLoM algorithm better preserves the concentration of the learned realizations for this challenging case. The PLoM algorithm with partition allows for retrieving the concentration that corresponds to the reference (which is known for this, see [238]).



(a) Classical MCMC algorithm (b) PLoM without partition (c) PLoM with partition

Figure 8: Clouds of the $n_{\text{learn}} = 1\,200\,000$ realizations of (H_1, H_2, H_3) generated with a classical MCMC algorithm (a), generated with PLoM algorithm without partition (b), and generated with PLoM algorithm with partition, coinciding with the reference (c). [From [238]].

7.3. Probabilistic learning under constraints

In Section 7.2, we have presented the PLoM algorithm that allows statistical surrogate models of complex systems to be constructed from small training datasets. When data are available, coming either from experimental measurements or from expensive numerical simulations carried out with large computational models implemented on very high performance computers, it becomes interesting, even necessary, to integrate these data into the learning process. In addition to data integration, we also want to be able to constrain the learned probability distribution by physics, for example, by constraining with nonlinear partial differential equations that underlie the stochastic computational model, in order to minimize the stochastic residual generated by the learned realizations. Such learning process is used to build statistical surrogate models for the parameterized stochastic computational models that are used for generating the small training datasets. Always placing ourselves in the framework for which the available training dataset is small, the objective of this section is to extend probabilistic learning on manifolds to the case of probabilistic learning inference on manifolds in order to perform the data integration, i.e. to extend PLoM to PLoM under constraints. From a general point of view, in the case of a small learning dataset and for high stochastic dimensions, this problem remains difficult. To cover these aspects, several methodologies based on theoretical developments have been proposed and validated through relatively complex applications. It is

(a) the physics-constrained probabilistic learning on manifolds [240] for which the constraints are defined by given statistical moments.

(b) the probabilistic learning that is constrained by nonlinear partial differential equations of stochastic boundary value problems [241], which has especially been developed for applications in stochastic nonlinear solid and fluid computational dynamics.

(c) the case for which the constraints are defined by realizations of system observations, the methodology being based on a weak formulation of Fourier transform of probability measures [243];

(d) the probabilistic learning inference under implicit constraints [242].

To validate these methodologies, several applications of PLoM under constraints have been performed. The application presented in [240] is related to a probabilistic learning inference consisting in a statistical inverse problem for a stochastic elliptic BVP related to linear elasticity of a heterogeneous anisotropic medium occupying a 3D bounded domain. The target dataset is constituted of experimental second-order statistical moments of the random displacement field given on the boundary. The inverse problem consists in identifying the tensor-valued random elasticity field using the target dataset. To construct the training dataset, the algebraic prior probability model of the random elasticity field for heterogeneous anisotropic elastic media, presented in Section 5.2, is used. The probabilistic learning on manifolds, constrained by given statistical moments is then used for integrating the experimental data. This application is revisited in [243] for which the target dataset of the second-order moments are replaced by a target dataset made up of experimental realizations of the displacement field on the boundary. In that case, the constraints are defined by the weak formulation of the Fourier transform of the learned probability distribution. If in addition to prescribed second-order statistical moments, the second-order moment of the random residue of the stochastic equations of the computational model has to be controlled during the learning process, then the constraints are defined by an implicit vector-valued mapping. This case has been analyzed in the application presented in [242], which deals with the 3D stochastic homogenization of heterogeneous elastic microstructure for which the prior probability model of the random tensor-valued elasticity field at mesoscale is the one presented in Section 5.2, which has been extended to the case of a random spectrum following the model presented in [74, 247]. A brief summary of this application will be presented in Section 8. In [241] devoted to PLoM constrained by nonlinear partial differential equations for small training datasets, two applications are presented for high-dimensional nonlinear stochastic dynamical systems: one is in computational fluid dynamics for unsteady 2D Navier-Stokes equation for incompressible fluid and another is in nonlinear elastodynamics for a 3D silicon MEMS accelerometer. In this section, the presented overview is limited to methodology (d).

7.3.1. Formulation of the probabilistic learning inference using the Kullback-Leibler divergence

The probabilistic learning inference based on a formulation of PLoM under constraints is an important question related to probabilistic learning algorithm that allows for integrating data (target dataset) into predictive models for which the training dataset is constituted of a small number of points and for which the target dataset 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 previously explained, physics-based constraints can simultaneously considered. This is the case if we want the learning process to be also controlled by the model, for instance, that the mean-square norm of the random normalized residue of the stochastic partial differential equation of the BVP be controlled. this type of constraint is thus also implicit. This section is mainly based on [242].

(i) Stochastic boundary value problem. In Sections 7.1-(i) and (ii), we have presented the PLoM formulation in finite dimension in order to facilitate the reading. In this section, we briefly introduce the formulation in infinite dimension. All the random quantities are defined on a probability space $(\Theta, \mathscr{T}, \mathscr{P})$. Let us consider a stochastic elliptic BVP on an open bounded domain $\Omega \subset \mathbb{R}^d$ (for instance, $d = 3$), whose partial differential equation (PDE) is written as $\boldsymbol{\mathcal{N}}(\mathbf{Y}, \mathbf{G}, \mathbf{W}) = \mathbf{0}$ *a.s.* The unknown is the non-Gaussian vector-valued field $\{\mathbf{Y}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ that satisfies the boundary conditions. The coefficients of the stochastic elliptic operator depend on a non-Gaussian second-order vector-valued random field $\mathbf{G}$ and on a random vector-valued control parameter $\mathbf{W}$. It is assumed that the weak formulation of this stochastic BVP admits a unique strong stochastic solution $\mathbf{Y} = \mathbf{f}(\mathbf{G}, \mathbf{W})$ that is a second-order random field. It is well known that the nonlinear operator $\mathbf{f}$ is not exactly known. Only an approximation can be constructed, for instance, using the finite element method to discretize the weak formulation and then using the Monte Carlo numerical method to approximate the strong stochastic solution. The observation (quantity of interest) is, for instance, a second-order vector-valued random variable $\mathbf{Q} = \boldsymbol{O}(\mathbf{Y}, \mathbf{G}, \mathbf{W})$ in which the observation operator, $\boldsymbol{O}$, is a given measurable mapping.

(ii) Probabilistic learning inference. The probabilistic learning inference belongs to the class of the statistical inverse problems. A prior probability model of $\{\mathbf{G}, \mathbf{W}\}$ is given. We are interested in estimating a posterior model $\{\mathbf{G}^c, \mathbf{W}^c\}$ of

$\{\mathbf{G}, \mathbf{W}\}$ in order that some statistical moments of the posterior observations $\mathbf{Q}^c = \boldsymbol{O}(\mathbf{Y}^c, \mathbf{G}^c, \mathbf{W}^c)$ with $\mathbf{Y}^c = \mathbf{f}(\mathbf{G}^c, \mathbf{W}^c)$ be 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 $\mathbf{Q}^c$ are globally written as $E\{\boldsymbol{\mathcal{M}}^c(\mathbf{Q}^c)\} = \mathbf{b}^c$ in which $\mathbf{b}^c \in \mathbb{R}^{n_c}$ is the target, $\mathbf{q} \mapsto \boldsymbol{\mathcal{M}}^c(\mathbf{q})$ is a given measurable mapping.

(iii) Small training dataset. As previously, subscript "d" designates the quantities related to the training dataset. Let $\{\mathbf{g}_d^1, \ldots, \mathbf{g}_d^{n_d}\}$ and $\{\mathbf{w}_d^1, \ldots, \mathbf{w}_d^{n_d}\}$ be n_d independent realizations of random variable $\{\mathbf{G}, \mathbf{W}\}$, generated by using the prior probability model of $\{\mathbf{G}, \mathbf{W}\}$. Each realization $\mathbf{y}_d^j$ is computed by solving the weak formulation of the PDE $\boldsymbol{\mathcal{N}}(\mathbf{y}_d^j, \mathbf{g}_d^j, \mathbf{w}_d^j) = \mathbf{0}$ with the boundary conditions. Consequently, n_d independent realizations $\{\mathbf{y}_d^j, j = 1, \ldots, n_d\}$ of random field $\mathbf{Y}$ are computed and are such that $\mathbf{y}_d^j = \mathbf{f}(\mathbf{g}_d^j, \mathbf{w}_d^j)$. The n_d independent realizations $\{\mathbf{q}_d^j, j = 1, \ldots, n_d\}$ of random observation $\mathbf{Q}$ are thus deduced and such that $\mathbf{q}_d^j = \boldsymbol{O}(\mathbf{y}_d^j, \mathbf{g}_d^j, \mathbf{w}_d^j)$. The training dataset is then made up of a small number n_d of points $\mathbf{x}_d^j = \{\mathbf{y}_d^j, \mathbf{g}_d^j, \mathbf{w}_d^j\}$ for $j = 1, \ldots, n_d$, which are n_d independent realizations of $\mathbf{X} = \{\mathbf{Y}, \mathbf{G}, \mathbf{W}\}$. It is assumed 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). Therefore, the posterior model is constructed using a learning tool for generating the constrained learned realizations of $\mathbf{X}$ without solving the BVP, but using only the training dataset.

(iv) Finite reduced-order representation and training dataset $\mathscr{D}_{n_d}(\boldsymbol{\eta})$. Similarly to the methodology of PLoM presented in Section 7.2.1, a finite reduced representation of $\mathbf{X}$ is constructed. The second-order random variable $\mathbf{X} = \{\mathbf{Y}, \mathbf{G}, \mathbf{W}\}$, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, is assumed to be with values in a real Hilbert space $\mathbb{X}$ equipped with the inner product $\langle \mathbf{X}, \mathbf{X}' \rangle_{\mathbb{X}}$ and its associated norm $\|\mathbf{X}\|_{\mathbb{X}} = \langle \mathbf{X}, \mathbf{X} \rangle_{\mathbb{X}}^{1/2}$. Consequently, $\mathbf{X}$ belongs the Hilbert space $L^2(\Theta, \mathbb{X})$, equipped with the inner product $\langle\langle \mathbf{X}, \mathbf{X}' \rangle\rangle = E\{\langle \mathbf{X}, \mathbf{X}' \rangle_{\mathbb{X}}\}$ for which the square of the associated norm is $|||\mathbf{X}|||^2 = E\{\|\mathbf{X}\|_{\mathbb{X}}^2\} = \int_\Theta \|\mathbf{X}(\theta)\|_{\mathbb{X}}^2 \, d\mathscr{P}(\theta)$. Since the problem is in infinite dimension, in order to implement the probabilistic learning inference, we need to introduce a finite representation $\mathbf{X}^{(\nu)}$ of dimension ν of random variable $\mathbf{X}$ in $L^2(\Theta, \mathbb{X})$. Assuming that the covariance operator is a symmetric, positive, trace operator in $\mathbb{X}$ [341], $\mathbf{X}^{(\nu)}$ can be represented using the truncated Karhunen-Loève expansion [342, 8] of $\mathbf{X}$,

$$\mathbf{X}^{(\nu)} = \underline{\mathbf{x}} + \sum_{\alpha=1}^{\nu} \sqrt{\kappa_\alpha} \, \boldsymbol{\varphi}^\alpha \, H_\alpha \,, \tag{7.33}$$

in which the eigenvalues of the covariance operator are $\kappa_1 \geq \ldots \geq \kappa_\nu \geq \ldots = 0$ with $\sum_{\alpha=1}^{+\infty} \kappa_\alpha < +\infty$, where the family of the eigenfunctions $\{\boldsymbol{\varphi}^\alpha\}_\alpha$ is a Hilbertian basis of $\mathbb{X}$, where $\underline{\mathbf{x}} = E\{\mathbf{X}\}$, and where $\mathbf{H} = (H_1, \ldots, H_\nu)$ is a second-order $\mathbb{R}^\nu$-valued random variable such that,

$$\forall \alpha \in \{1, \ldots, \nu\} \quad , \quad H_\alpha = \kappa_\alpha^{-1/2} \, \langle \mathbf{X} - \underline{\mathbf{x}}, \boldsymbol{\varphi}^\alpha \rangle_{\mathbb{X}} \quad \text{and} \quad E\{\mathbf{H}\} = \mathbf{0}_\nu \quad , \quad E\{\mathbf{H}\,\mathbf{H}^T\} = [I_\nu] \,. \tag{7.34}$$

The training set $\mathscr{D}_{n_d}(\boldsymbol{\eta})$ for $\mathbf{H}$ is then defined by n_d independent realizations $\boldsymbol{\eta}_d^j$ such that

$$\mathscr{D}_{n_d}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\} \quad , \quad \boldsymbol{\eta}_d^j = \kappa_\alpha^{-1/2} \, \langle \mathbf{x}_d^j - \underline{\mathbf{x}}, \boldsymbol{\varphi}^\alpha \rangle_{\mathbb{X}} \,. \tag{7.35}$$

If the context of the probabilistic learning inference, the kernel of the covariance operator of $\mathbf{X}$ is not explicitly known. Therefore, the value of ν can be computed in studying the graph of the function $\nu \mapsto |||\mathbf{X}^{(\nu)} - \underline{\mathbf{x}}||| = (\sum_{\alpha=1}^{\nu} \kappa_\alpha)^{1/2}$.

(v) Formulation using the Kullback-Leibler divergence minimum principle. Let $P_{\mathbf{H}}(d\boldsymbol{\eta}) = p_{\mathbf{H}}(\boldsymbol{\eta}) \, d\boldsymbol{\eta}$ be the prior probability distribution on $\mathbb{R}^\nu$ of $\mathbf{H}$, for which the pdf $\boldsymbol{\eta} \mapsto p_{\mathbf{H}}(\boldsymbol{\eta})$ on $\mathbb{R}^\nu$ is estimated using the multidimensional Gaussian KDE method and the training dataset $\mathscr{D}_{n_d}(\boldsymbol{\eta})$ defined by Eq. (7.35). Therefore, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, $p_{\mathbf{H}}(\boldsymbol{\eta})$ is written as

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = c_\nu \, \zeta(\boldsymbol{\eta}) \quad \text{with} \quad c_\nu = (\sqrt{2\pi}\,\hat{s})^{-\nu} \quad \text{and} \quad \zeta(\boldsymbol{\eta}) = \frac{1}{n_d} \sum_{j=1}^{n_d} \exp\left\{ -\frac{1}{2\hat{s}^2} \, \| \frac{\hat{s}}{s} \, \boldsymbol{\eta}_d^j - \boldsymbol{\eta} \|^2 \right\} , \tag{7.36}$$

in which $\hat{s}$ is the modified Silverman bandwidth defined in Section 7.2.1-(iii). Let $\mathbf{H}^c = (H_1^c, \ldots, H_\nu^c)$ be the $\mathbb{R}^\nu$-

valued random variable whose posterior probability distribution, written as $P_{\mathbf{H}^c}(d\boldsymbol{\eta}) = p_{\mathbf{H}^c}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$, satisfies (see Section 7.3.1-(ii)) the constraint $E\{\boldsymbol{\mathcal{M}}^c(\mathbf{Q}^c)\} = \mathbf{b}^c$ that is rewritten in function of $\mathbf{H}^c$ as

$$E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c \in \mathbb{R}^{n_c}\,, \tag{7.37}$$

in which $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ is a function from $\mathbb{R}^\nu$ into $\mathbb{R}^{n_c}$, which satisfies the following hypothesis.

Hypothesis 1 (Concerning function $\mathbf{h}^c$). Let n_c be the integer such that $1 \leq n_c \leq \nu$.
(H1) Given any bounded positive measure $P(d\boldsymbol{\eta})$ on $\mathbb{R}^\nu$ with support $\mathbb{R}^\nu$, there exists a bounded set $\mathscr{B}$ in $\mathbb{R}^\nu$ with $P(\mathscr{B}) > 0$ such that $\forall\, \mathbf{v} \in \mathbb{R}^{n_c}$ with $\|\,\mathbf{v}\,\| \neq 0$, we have $\int_{\mathscr{B}} \langle \mathbf{h}^c(\boldsymbol{\eta})\,, \mathbf{v}\rangle^2 P(d\boldsymbol{\eta}) > 0$, which means that the constraints are algebraically independent.
(H2) Function $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ is in $C^1(\mathbb{R}^\nu, \mathbb{R}^{n_c})$ (continuously differentiable function from $\mathbb{R}^\nu$ into $\mathbb{R}^{n_c}$).
(H3) There exist constants $\alpha > 0$, $\beta > 0$, $c_\alpha > 0$, and $c_\beta > 0$, independent of $\boldsymbol{\eta}$, such that for $\|\,\boldsymbol{\eta}\,\| \to +\infty$, we have $\|\,\mathbf{h}^c(\boldsymbol{\eta})\,\| \;\leq\; c_\alpha\, \|\,\boldsymbol{\eta}\,\|^\alpha$ and $\|\,[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})]\,\|_F \;\leq\; c_\beta\, \|\,\boldsymbol{\eta}\,\|^\beta$ in which $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu,n_c}$ with $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})]_{\alpha k} = \partial h^c_k(\boldsymbol{\eta})/\partial \eta_\alpha$.

The Kullback-Leibler divergence between two probability measures $p(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ and $p_{\mathbf{H}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ is defined [46, 49, 50] by $D(p, p_{\mathbf{H}}) = \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log(p(\boldsymbol{\eta})/p_{\mathbf{H}}(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$, and is such that $D(p, p_{\mathbf{H}}) \geq 0$ (that can be proven by applying the Jensen inequality [343, 18]) and $D(p, p_{\mathbf{H}}) = 0$ if and only if $p = p_{\mathbf{H}}$. Note that $(p, p_{\mathbf{H}}) \mapsto D(p, p_{\mathbf{H}})$ is not a distance because the symmetry property and the triangle inequality are not verified. It can easily be seen that the cross entropy $S(p, p_{\mathbf{H}}) = -\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log(p_{\mathbf{H}}(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$ and the entropy $S(p) = -\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log(p(\boldsymbol{\eta}))\, d\boldsymbol{\eta}$ are related to $D(p, p_{\mathbf{H}})$ by $S(p, p_{\mathbf{H}}) = S(p) + D(p, p_{\mathbf{H}})$. Finally, it can be proven (see for instance [50]) that $(p, p_{\mathbf{H}}) \mapsto D(p, p_{\mathbf{H}})$ is a convex function in the pair $(p, p_{\mathbf{H}})$. Using the Kullback-Leibler divergence minimum principle [46, 49, 50], the probability density function $p_{\mathbf{H}^c}$ on $\mathbb{R}^\nu$, which satisfies the constraint defined by Eq. (7.37) and which is closest to $p_{\mathbf{H}}$, is the solution of the optimization problem (see for instance [240, 241]),

$$p_{\mathbf{H}^c} = \arg \min_{p \in \mathscr{C}_{\mathrm{ad},p}} \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log\left(\frac{p(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})}\right) d\boldsymbol{\eta}\,, \tag{7.38}$$

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

$$\mathscr{C}_{\mathrm{ad},p} = \left\{ \boldsymbol{\eta} \mapsto p(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+\,, \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = 1\,, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \mathbf{b}^c \right\}. \tag{7.39}$$

7.3.2. Methodology and algorithm used for solving the optimization problem and MCMC generator

In the optimization problem defined by Eq. (7.38), the constraints are taken into account introducing Lagrange multipliers.

(i) Lagrange multipliers associated with the constraints. The constraints within the admissible set $\mathscr{C}_{\mathrm{ad},p}$ are taken into account by introducing the Lagrange multipliers $\lambda_0 - 1$ with $\lambda_0 \in \mathbb{R}^+$ (associated with the normalization condition) and $\boldsymbol{\lambda} \in \mathscr{C}_{\mathrm{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{n_c}$ (associated with the imposed moments). Under Hypothesis 1, the admissible set $\mathscr{C}_{\mathrm{ad},\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is the open subset of $\mathbb{R}^{n_c}$, assumed to be not reduced to the empty set, such that

$$\mathscr{C}_{\mathrm{ad},\boldsymbol{\lambda}} = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^{n_c} \mid 0 < E\{\exp\{-\langle \boldsymbol{\lambda}\,, \mathbf{h}^c(\mathbf{H})\rangle\} < +\infty \right\}, \tag{7.40}$$

in which the pdf of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}$ is defined by Eq. (7.36).

(ii) Construction of the optimal solution. Let us assumed that the optimization problem defined by Eq. (7.38) has almost one solution $p_{\mathbf{H}^c}$ and that $p = p_{\mathbf{H}^c}$ is a regular point of the continuously differentiable functional $p \mapsto \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - \mathbf{b}^c$. For $\lambda_0 \in \mathbb{R}^+$ and $\boldsymbol{\lambda} \in \mathscr{C}_{\mathrm{ad},\boldsymbol{\lambda}}$, we define the Lagrangian,

$$Lag(p, \lambda_0, \boldsymbol{\lambda}) = \int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, \log\left(\frac{p(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})}\right) d\boldsymbol{\eta} + (\lambda_0 - 1)\,(\int_{\mathbb{R}^\nu} p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - 1) + \langle \boldsymbol{\lambda}\,, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, d\boldsymbol{\eta} - \mathbf{b}^c \rangle\,, \tag{7.41}$$

in which $p_{\mathbf{H}}$ is defined by Eq. (7.36). Using the calculus of variations, an extremum of functional $p \mapsto Lag(p, \lambda_0, \boldsymbol{\lambda})$ is found as the function $\boldsymbol{\eta} \mapsto p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ on $\mathbb{R}^\nu$, indexed by $\boldsymbol{\lambda}$, writing as

$$p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = c_0(\boldsymbol{\lambda})\,\zeta(\boldsymbol{\eta})\,\exp\{-\langle\boldsymbol{\lambda}\,,\mathbf{h}^c(\boldsymbol{\eta})\rangle\} \quad , \quad \forall\,\boldsymbol{\eta} \in \mathbb{R}^\nu\,, \tag{7.42}$$

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 $c_0(\boldsymbol{\lambda}) = c_\nu\,\exp\{-\lambda_0\}$). Under Hypothesis 1-(H1), there exists (see [344]) $\boldsymbol{\lambda}^{\rm sol}$ in $\mathscr{C}_{ad,\boldsymbol{\lambda}}$ such that the functional $(p, \lambda_0, \boldsymbol{\lambda}) \mapsto Lag(p, \lambda_0, \boldsymbol{\lambda})$ is stationary at point $p = p_{\mathbf{H}^c}$ for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\rm sol}$ and $\lambda_0 = -\log(c_0(\boldsymbol{\lambda}^{\rm sol})/c_\nu)$. Therefore, $p_{\mathbf{H}^c} = p_{\mathbf{H}_{\boldsymbol{\lambda}^{\rm sol}}}(.\,;\boldsymbol{\lambda}^{\rm sol})$ and Eq. (7.42) yields

$$p_{\mathbf{H}^c}(\boldsymbol{\eta}) = c_0(\boldsymbol{\lambda}^{\rm sol})\,\zeta(\boldsymbol{\eta})\,\exp\{-\langle\boldsymbol{\lambda}^{\rm sol},\mathbf{h}^c(\boldsymbol{\eta})\rangle\} \quad , \quad \forall\,\boldsymbol{\eta} \in \mathbb{R}^\nu\,. \tag{7.43}$$

Under Hypothesis 1, $p_{\mathbf{H}^c}$ is the unique solution of the optimization problem defined by Eqs. (7.38) and (7.39), in which $\boldsymbol{\lambda}^{\rm sol}$ is the unique solution of a convex optimization problem defined by Theorem 3.

Theorem 3 (Construction of the probability distribution of $\mathbf{H}_{\boldsymbol{\lambda}}$ (from [242])). *Under Hypothesis 1, the admissible set $\mathscr{C}_{\rm ad,\boldsymbol{\lambda}}$ defined by Eq.* (7.41) *is a convex open subset of $\mathbb{R}^{n_c}$. For all $\boldsymbol{\lambda}$ in $\mathscr{C}_{\rm ad,\boldsymbol{\lambda}}$, let $p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ be the pdf of $\mathbf{H}_{\boldsymbol{\lambda}}$, defined by Eq.* (7.43).
(a) The $\mathbb{R}^{n_c}$-valued random variable $\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}})$ is of second-order,

$$E\{\|\,\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}})\,\|^2\} < +\infty\,. \tag{7.44}$$

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

$$\Gamma(\boldsymbol{\lambda}) = \langle\boldsymbol{\lambda}\,,\mathbf{b}^c\rangle - \log c_0(\boldsymbol{\lambda})\,, \tag{7.45}$$

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

$$\nabla_\lambda\Gamma(\boldsymbol{\lambda}) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}})\} \in \mathbb{R}^{n_c}\,, \tag{7.46}$$

$$[\Gamma\,''(\boldsymbol{\lambda})] = [cov\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}})\}] \in \mathbb{M}^+_{n_c}\,, \tag{7.47}$$

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

$$\boldsymbol{\lambda}^{\rm sol} = \arg\min_{\boldsymbol{\lambda}\in\mathscr{C}_{\rm ad,\boldsymbol{\lambda}}} \Gamma(\boldsymbol{\lambda})\,. \tag{7.48}$$

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

$$\nabla_\lambda\Gamma(\boldsymbol{\lambda}) = \mathbf{0}_{n_c}\,, \tag{7.49}$$

has a solution $\tilde{\boldsymbol{\lambda}}$ that belongs to $\mathscr{C}_{\rm ad,\boldsymbol{\lambda}}$, then this solution is unique and we have $\boldsymbol{\lambda}^{\rm sol} = \tilde{\boldsymbol{\lambda}}$.

Although the optimization problem defined by Eq. (7.48) be a convex optimization problem, since the constant of normalization $c_0(\boldsymbol{\lambda})$ cannot numerically be estimated, a classical descent algorithm cannot be used and is replaced by searching the solution of Eq. (7.49). The Newton iterative algorithm is used to solve this equation with an under-relaxation coefficient similarly to the algorithm detailed in Appendix B.

(iii) MCMC generator of $\mathbf{H}^c$. Using Eq. (7.43), the posterior pdf $p_{\mathbf{H}^c}$ is constructed as the limit of a sequence $\{p_{\mathbf{H}_\lambda}\}_\lambda$ of probability density functions of a $\mathbb{R}^\nu$-valued random variable $\mathbf{H}_{\boldsymbol{\lambda}} = (H_{\lambda,1},\ldots,H_{\lambda,\nu})$ that depends on $\boldsymbol{\lambda}$. Based on Theorem 3, an algorithm similar to the one presented in Appendix B is used to find $\boldsymbol{\lambda}^{\rm sol}$. Therefore, the construction of $\{p_{\mathbf{H}_\lambda}\}_\lambda$ requires to generate a constrained learned dataset $\mathscr{D}_{n_{\rm learn}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}}) = \{\boldsymbol{\eta}^1_{\boldsymbol{\lambda}},\ldots\boldsymbol{\eta}^{n_{\rm learn}}_{\boldsymbol{\lambda}}\}$ constituted of $n_{\rm learn} \gg n_d$ independent realizations $\{\boldsymbol{\eta}^\ell_{\boldsymbol{\lambda}}, \ell = 1,\ldots,n_{\rm learn}\}$ of $\mathbf{H}_{\boldsymbol{\lambda}}$. When the convergence is reached with respect to $\boldsymbol{\lambda}$, the constrained learned set $\mathscr{D}_{\mathbf{H}^c} = \{\boldsymbol{\eta}^1_c,\ldots,\boldsymbol{\eta}^{n_{\rm learn}}_c\}$ is generated. This set is made up of $n_{\rm learn}$ independent realizations $\{\boldsymbol{\eta}^\ell_c, \ell = 1,\ldots,n_{\rm learn}\}$ of $\mathbf{H}^c$ whose probability distribution $p_{\mathbf{H}^c}(\boldsymbol{\eta})\,d\boldsymbol{\eta}$ satisfies the constraint $E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c$. The construction of $\mathscr{D}_{n_{\rm learn}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}})$ is based on the following Theorem 4 and is carried our using the PLoM algorithm presented in Section 7.2.

Theorem 4 (MCMC generator of $\mathbf{H}_{\lambda}$ (from [242])). *Let $\mathbf{h}^c$ be the function satisfying Hypothesis 1. Let $\boldsymbol{\lambda}$ be fixed in $\mathscr{C}_{\mathrm{ad},\lambda}$. Let $\{(\mathbf{U}_{\lambda}(r), \mathbf{V}_{\lambda}(r)), r \in \mathbb{R}^+\}$ be the $\mathbb{R}^{\nu} \times \mathbb{R}^{\nu}$-valued stochastic process, satisfying, for $r > 0$ and with initial conditions,*

$$d\mathbf{U}_{\lambda}(r) = \mathbf{V}_{\lambda}(r)\, dr\,, \tag{7.50}$$

$$d\mathbf{V}_{\lambda}(r) = \mathbf{L}_{\lambda}(\mathbf{U}_{\lambda}(r))\, dr - \frac{1}{2} f_0\, \mathbf{V}_{\lambda}(r)\, dr + \sqrt{f_0}\, d\mathbf{W}^{\mathrm{wien}}(r)\,, \tag{7.51}$$

$$\mathbf{U}_{\lambda}(0) = \mathbf{u}_0\ ,\ \mathbf{V}_{\lambda}(0) = \mathbf{v}_0\ a.s. \tag{7.52}$$

(a) As previously, f_0 allows the transient part to be controlled and is such that $0 < f_0 < 4/\hat{s}$ and $\mathbf{W}^{\mathrm{wien}}$ is the $\mathbb{R}^{\nu}$-valued normalized Wiener stochastic process independent of $\mathbf{H}$.
(b) The initial condition $\mathbf{u}_0 \in \mathbb{R}^{\nu}$ is chosen from the points of the training set $\mathscr{D}_{n_d}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}_d^1, \ldots, \boldsymbol{\eta}_d^{n_d}\}$ while the initial condition $\mathbf{v}_0$ is any realization of a normalized Gaussian $\mathbb{R}^{\nu}$-valued random variable $\mathbf{V}_G$, independent of $\mathbf{W}^{\mathrm{wien}}$.
(c) For all $\mathbf{u} = (u_1, \ldots, u_{\nu})$ in $\mathbb{R}^{\nu}$, the vector $\mathbf{L}_{\lambda}(\mathbf{u})$ in $\mathbb{R}^{\nu}$ is written as

$$\mathbf{L}_{\lambda}(\mathbf{u}) = \frac{1}{\zeta(\mathbf{u})}\, \nabla_{\mathbf{u}} \zeta(\mathbf{u}) - [\nabla_{\mathbf{u}} \mathbf{h}^c(\mathbf{u})]\, \boldsymbol{\lambda}\,, \tag{7.53}$$

in which function ζ is defined by Eq. (7.36)*.*
(d) The stochastic solution $\{(\mathbf{U}_{\lambda}(r), \mathbf{V}_{\lambda}(r)), r \geq 0\}$ of the ISDE defined by Eqs. (7.50) *to* (7.52) *is unique, has almost-surely continuous trajectories, and is a second-order diffusion stochastic process. For $r \to +\infty$, this diffusion process converges to a stationary second-order diffusion stochastic process $\{(\mathbf{U}_{\lambda}^{\mathrm{st}}(\tau), \mathbf{V}_{\lambda}^{\mathrm{st}}(\tau)), \tau \geq 0\}$ associated with the unique invariant probability measure on $\mathbb{R}^{\nu} \times \mathbb{R}^{\nu}$,*

$$p_{\mathbf{H}_{\lambda}, \mathbf{V}_G}(\boldsymbol{\eta}, \mathbf{v}\,; \boldsymbol{\lambda})\, d\boldsymbol{\eta} \otimes d\mathbf{v} = (p_{\mathbf{H}_{\lambda}}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})\, d\boldsymbol{\eta}) \otimes (p_{\mathbf{V}_G}(\mathbf{v})\, d\mathbf{v})\,, \tag{7.54}$$

in which $p_{\mathbf{H}_{\lambda}}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})$ is the pdf defined by Eq. (7.42)*.*
(e) For r_s sufficiently large, we can choose $\mathbf{H}_{\lambda}$ as $\mathbf{U}_{\lambda}(r_s)$. The generation of the constrained learned dataset $\mathscr{D}_{n_{\mathrm{learn}}}(\boldsymbol{\eta}_{\lambda})$ is constructed by solving Eqs. (7.50) *to* (7.52) *for $r \in [0\,, r_s]$ and then using the realizations of $\mathbf{U}_{\lambda}(r_s)$.*

It should be noted that the PLoM algorithm for $\mathbf{H}_{\lambda}$ can easily be derived from Theorem 4, by introducing the random matrix $[\mathbf{H}_{\lambda}]$ whose columns are n_d independent copies of $\mathbf{H}_{\lambda}$ and then by projecting Eqs. (7.50) to (7.52) on the truncated diffusion-maps basis represented by matrix $[g_m]^T$ yielding the reduced-order ISDE that is similar to the one given in Theorem 1.

It can be seen that $\mathbf{L}_{\lambda}(\mathbf{u})$ defined by Eq. (7.53) involves the matrix $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})] \in \mathbb{M}_{\nu, n_c}$ (the transpose of the Jacobian matrix of $\mathbf{h}^c$). In general, function $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ from $\mathbb{R}^{\nu}$ into $\mathbb{R}^{n_c}$ is not explicitly defined by an algebraic expression, and only $\mathbf{a}^{\ell} = \mathbf{h}^c(\boldsymbol{\eta}^{\ell}) \in \mathbb{R}^{n_c}$ can be computed for any point $\boldsymbol{\eta}^{\ell}$ given in $\mathbb{R}^{\nu}$ (for instance and as previously underlined, a component of $\mathbf{h}^c$ 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 $[\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})]$ for a large number of values of $\boldsymbol{\eta}$. The construction of a surrogate model of implicit function $\mathbf{h}^c$ by using a deterministic approach, such as the meshless methods [345, 346, 347, 348, 349, 350], is not adapted taking into account a possible high dimension of the space on which $\mathbf{h}^c$ is defined. To circumvent this difficulty, the approach proposed in [241] has been generalized in [242], which consists in constructing a statistical surrogate model $\hat{\mathbf{h}}^{n_{\mathrm{learn}}}$ of $\mathbf{h}^c$, depending on the number n_{learn} of points generated in the constrained learned dataset, for which there is a simple explicit algebraic representation of the gradient function $\boldsymbol{\eta} \mapsto [\nabla_{\boldsymbol{\eta}} \hat{\mathbf{h}}^{n_{\mathrm{learn}}} \boldsymbol{\eta})]$ from $\mathbb{R}^{\nu}$ into $\mathbb{M}_{\nu, n_c}$. Such statistical surrogate model is an approximation whose convergence with respect to n_{learn} has to be studied.

(iv) Statistical surrogate model of the implicit function $\mathbf{h}^c$ and convergence analysis. In this section, we define the surrogate model $\hat{\mathbf{h}}^{n_{\mathrm{learn}}}$ of $\mathbf{h}^c$ and we study the convergence of the sequences $\{\hat{\mathbf{h}}^{n_{\mathrm{learn}}}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})\}_{n_{\mathrm{learn}}}$ and $\{[\nabla_{\boldsymbol{\eta}} \hat{\mathbf{h}}^{n_{\mathrm{learn}}}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})]\}_{n_{\mathrm{learn}}}$ as $n_{\mathrm{learn}} \to +\infty$. Then, we present the convergence of the sequence of MCMC generator using the statistical surrogate model. Let $\boldsymbol{\lambda}$ be fixed in $\mathscr{C}_{\mathrm{ad},\lambda}$ and let $\mathscr{D}_{n_{\mathrm{learn}}}(\boldsymbol{\eta}_{\lambda}) = \{\boldsymbol{\eta}_{\lambda}^1, \ldots, \boldsymbol{\eta}_{\lambda}^{n_{\mathrm{learn}}}\}$ be the constrained learned dataset whose points are $n_{\mathrm{learn}} \gg n_d$ independent realizations of the $\mathbb{R}^{\nu}$-valued random variable $\mathbf{H}_{\lambda}$ for which the pdf $\boldsymbol{\eta} \mapsto p_{\mathbf{H}_{\lambda}}(\boldsymbol{\eta}\,; \boldsymbol{\lambda})$ is defined by Eq. (7.42). Let $\mathbf{A}_{\lambda} = \mathbf{h}^c(\mathbf{H}_{\lambda})$ be the $\mathbb{R}^{n_c}$-valued random variable whose n_{learn} independent realizations $\mathbf{a}_{\lambda}^1, \ldots, \mathbf{a}_{\lambda}^{n_{\mathrm{learn}}}$

are such that $\mathbf{a}_{\boldsymbol{\lambda}}^{\ell} = \mathbf{h}^c(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^{\ell}) \in \mathbb{R}^{n_c}$ for $\ell = 1, \ldots, n_{\text{learn}}$. The surrogate model $\boldsymbol{\eta} \mapsto \hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) : \mathbb{R}^{\nu} \to \mathbb{R}^{n_c}$ of $\mathbf{h}^c$ is defined [242], for all $\boldsymbol{\eta}$ in $\mathbb{R}^{\nu}$, by

$$\hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = \sum_{\ell=1}^{n_{\text{learn}}} \mathbf{a}_{\boldsymbol{\lambda}}^{\ell} \, \frac{\beta_{\boldsymbol{\eta}}^{n_{\text{learn}}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^{\ell})}{\sum_{\ell'=1}^{n_{\text{learn}}} \beta_{\boldsymbol{\eta}}^{n_{\text{learn}}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}}^{\ell'})} \,, \tag{7.55}$$

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

$$\beta_{\boldsymbol{\eta}}^{n_{\text{learn}}}(\tilde{\boldsymbol{\eta}}) = \exp\{-\frac{1}{2s_{\text{SB}}^2} \| \, \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \, \|_H^2\} \quad , \quad \| \, \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \, \|_H^2 = \langle [\sigma_{\mathbf{H}_{\boldsymbol{\lambda}}}]^{-2} (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \, , \tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} \rangle \, , \tag{7.56}$$

in which $[\sigma_{\mathbf{H}_{\boldsymbol{\lambda}}}]$ is the diagonal positive-definite matrix in $\mathbb{M}_{\nu}^{+}$ such that $[\sigma_{\mathbf{H}_{\boldsymbol{\lambda}}}]_{\alpha\alpha}$ is the standard deviation of the real-valued random variable $H_{\boldsymbol{\lambda},\alpha}$, estimated using $\mathscr{D}_{n_{\text{learn}}}(\boldsymbol{\eta}_{\boldsymbol{\lambda}})$, and where s_{SB} is the Silverman bandwidth that depends on n_{learn} and written as $s_{\text{SB}} = (4/\{n_{\text{learn}}(2 + n_c + \nu)\})^{1/(n_c+\nu+4)}$.

Theorem 5 (Convergence of the sequence of MCMC generator using the statistical surrogate model (from [242])). *Let $\boldsymbol{\lambda}$ be fixed in $\mathscr{C}_{\text{ad},\boldsymbol{\lambda}}$ and let $\boldsymbol{\eta}$ be fixed in $\mathbb{R}^{\nu}$. Under Hypothesis 1-(H2), $\forall \varepsilon > 0$, there exists a finite integer $n_{\varepsilon}(\boldsymbol{\eta}, \boldsymbol{\lambda})$ depending on ε, $\boldsymbol{\eta}$, and $\boldsymbol{\lambda}$, such that $\forall n_{\text{learn}} \geq n_{\varepsilon}(\boldsymbol{\eta}, \boldsymbol{\lambda})$,*

$$\| \, \hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) - \mathbf{h}^c(\boldsymbol{\eta}) \, \| \;\leq\; \varepsilon \quad , \quad \| \, [\nabla_{\boldsymbol{\eta}} \hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})] - [\nabla_{\boldsymbol{\eta}} \mathbf{h}^c(\boldsymbol{\eta})] \, \|_F \;\leq\; \varepsilon \, . \tag{7.57}$$

Let $\hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda})$ be the approximation of $\mathbf{h}^c(\boldsymbol{\eta})$ defined by Eq. (7.55) *and let $\mathbf{u} \mapsto \hat{\mathbf{L}}_{\boldsymbol{\lambda}}(\mathbf{u})$ be the twice continuously differentiable function on $\mathbb{R}^{\nu}$ with values in $\mathbb{R}^{\nu}$ such that, for all $\mathbf{u}$ in $\mathbb{R}^{\nu}$,*

$$\hat{\mathbf{L}}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(\mathbf{u}) = \frac{1}{\zeta(\mathbf{u})} \, \nabla_{\mathbf{u}} \zeta(\mathbf{u}) - [\nabla_{\mathbf{u}} \hat{\mathbf{h}}^{n_{\text{learn}}}(\mathbf{u}\,;\boldsymbol{\lambda})] \, \boldsymbol{\lambda} \, , \tag{7.58}$$

in which $\boldsymbol{\zeta}$ is defined by Eq. (7.36) *and where $[\nabla_{\mathbf{u}} \hat{\mathbf{h}}^{n_{\text{learn}}}(\mathbf{u}\,;\boldsymbol{\lambda})]$ is explicitly given by differentiating function $\mathbf{u} \mapsto \hat{\mathbf{h}}^{n_{\text{learn}}}(\mathbf{u}\,;\boldsymbol{\lambda})$ defined by Eq.* (7.55)*. Let $\{(\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r), \mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r)), r \in \mathbb{R}^{+}\}$ be the stochastic process solution of the ISDE defined by Eqs.* (7.50) *to* (7.52) *in which $\mathbf{L}_{\boldsymbol{\lambda}}$ is replaced by $\hat{\mathbf{L}}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}$,*

$$d\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r) = \mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r) \, dr \, , \tag{7.59}$$

$$d\mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r) = \hat{\mathbf{L}}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r)) \, dr - \frac{1}{2} f_0 \, \mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r) \, dr + \sqrt{f_0} \, d\mathbf{W}^{\text{wien}}(r) \, , \tag{7.60}$$

$$\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(0) = \mathbf{u}_0 \;, \; \mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(0) = \mathbf{v}_0 \; a.s. \, , \tag{7.61}$$

and where $\mathbf{u}_0$, $\mathbf{v}_0$, f_0, and $\mathbf{W}^{\text{wien}}$ are the quantities defined in Theorem 4. Then the stochastic solution $\{(\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r), \mathbf{V}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r)), r \in \mathbb{R}^{+}\}$ of Eqs. (7.59) *to* (7.61) *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 $r \to +\infty$, associated with the unique invariant probability measure on $\mathbb{R}^{\nu} \times \mathbb{R}^{\nu}$, $\hat{p}_{\mathbf{H}_{\boldsymbol{\lambda}},\mathbf{V}_G}^{n_{\text{learn}}}(\boldsymbol{\eta}, \mathbf{v}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta} \otimes d\mathbf{v} = (\hat{p}_{\mathbf{H}_{\boldsymbol{\lambda}}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) \, d\boldsymbol{\eta}) \otimes (p_{\mathbf{V}_G}(\mathbf{v}) \, d\mathbf{v})$, in which $\hat{p}_{\mathbf{H}_{\boldsymbol{\lambda}}}^{n_{\text{learn}}}(\boldsymbol{\eta}\,;\boldsymbol{\lambda}) = c_0^{n_{\text{learn}}}(\boldsymbol{\lambda}) \, \zeta(\boldsymbol{\eta}) \, \exp\{-\langle \boldsymbol{\lambda}, \hat{\mathbf{h}}^{n_{\text{learn}}}(\boldsymbol{\eta})\rangle\}$. Then for all $r \in [0\,, r_s]$ with $r_s < +\infty$, the sequence $\{\mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r)\}_{n_{\text{learn}}}$ of second-order $\mathbb{R}^{\nu}$-valued random variables converges in mean-square to the second-order $\mathbb{R}^{\nu}$-valued random variable $\mathbf{U}_{\boldsymbol{\lambda}}(r)$ of Theorem 4,*

$$\lim_{n_{\text{learn}} \to +\infty} E\{\| \, \mathbf{U}_{\boldsymbol{\lambda}}^{n_{\text{learn}}}(r) - \mathbf{U}_{\boldsymbol{\lambda}}(r) \, \|^2\} = 0 \quad , \quad \forall r \in [0\,, r_s] \, . \tag{7.62}$$

8. Illustration of probabilistic learning inference for 3D stochastic homogenization of heterogeneous material with random spectrum and no scale separation

This section is an illustration of the use of the probabilistic learning under constraints presented in Section 7.3 for the 3D stochastic homogenization of an elastic heterogeneous microstructure without scale separation. The details of this application can be found in [242]. The nonseparation of the mesoscale with the macroscale means that the macroscale is another mesoscale at larger scale with random effective/apparent elastic properties. The posterior

probability distribution is constructed using the probabilistic learning inference (PLoM under constraints) in which the constraints are defined by:

- a target set made up of given "experimental" statistical moments of the random effective/apparent elasticity tensor $\mathbb{C}^{\text{eff}}$.
- the second-order moment of the random normalized residue of the random equations of the stochastic computational model.

These constraints guarantee that the probabilistic learning algorithm seeks to bring the statistical moments closer to their targets while preserving a small residue of the random equations of the stochastic computational model.

8.1. Formulation

(i) Stochastic elliptic boundary value problem. We consider the stochastic homogenization of a heterogeneous 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 $\partial\Omega$. For all m and r in $\{1,2,3\}$ the $\mathbb{R}^3$-valued displacement random field $\{\mathbf{Y}(\boldsymbol{\xi}) = (Y_1(\boldsymbol{\xi}), Y_2(\boldsymbol{\xi}), Y_3(\boldsymbol{\xi})), \boldsymbol{\xi} \in \Omega\}$, defined on $(\Theta, \mathscr{T}, \mathscr{P})$, indexed by Ω, satisfies the stochastic BVP associated with the stochastic homogenization of a random elastic medium without scale separation, which is defined by Eqs. (6.1) to (6.3). For i, j, m, and r in $\{1,2,3\}$, the component $\mathbb{C}^{\text{eff}}_{ijmr}$ of the random effective/apparent elasticity tensor $\mathbb{C}^{\text{eff}}$ at macroscale is defined by Eq. (6.4).

(ii) Prior probability model of the apparent elasticity field $\mathbb{C}$ at mesoscale. At mesoscale, the prior probability model of the random apparent elasticity field $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ is the one for heterogeneous anisotropic elastic media, which is defined in Section 5.2 and for which the spectrum is uncertain (see Section 5.2-(iv) and [74, 247]). This random field is the restriction to Ω of a non-Gaussian, homogeneous, anisotropic, fourth-order tensor-valued random field indexed by $\mathbb{R}^3$, with a random spectral measure. The level of uncertainties on the spectrum is controlled by a hyperparameter $\delta_s = 0.1$. The spatial correlation lengths at mesoscale are $\underline{\mathbf{L}}_c = (\underline{L}_{c1}, \underline{L}_{c2}, \underline{L}_{c3})$. The level of statistical fluctuations of the random medium is controlled by the dispersion coefficient $\underline{\delta}_{\mathbb{C}} = 0.3$. At mesoscale, the nominal model is isotropic with $\underline{C}_{\text{bulk}} = 1.09\times10^{11}\, N/m^2$ and $\underline{C}_{\text{shear}} = 6.85\times10^{10}\, N/m^2$. The parameterization of $\mathbb{C}$ is written as $\mathbb{C}(\boldsymbol{\xi}) = \mathbb{c}(\mathbf{G}(\boldsymbol{\xi}), \underline{\mathbf{z}})$ for $\boldsymbol{\xi} \in \Omega$ in which $\mathbb{c}$ is a tensor-valued function on $\mathbb{R}^{21}\times\mathbb{R}^3$, where $\{\mathbf{G}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \mathbb{R}^3\}$ is a non-Gaussian $\mathbb{R}^{21}$-valued random, and where $\underline{\mathbf{z}} = (\underline{C}_{\text{bulk}}, \underline{C}_{\text{shear}}, \underline{\delta}_{\mathbb{C}})$.

(iii) Spatial correlation lengths and scale separation. 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. In comparing the dimensions $1\times1\times0.1$ of domain Ω with respect to the values of the spatial correlation lengths, there will not be, *a priori*, 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.

	$\underline{L}_{c1}$	$\underline{L}_{c2}$	$\underline{L}_{c3}$	*Level of scale separation*
SC1	0.1	0.1	0.1	partial separation in ξ_1, ξ_2, not in ξ_3
SC2	0.3	0.3	0.1	not separated in ξ_1, ξ_2, and ξ_3
SC3	0.5	0.5	0.2	strongly separated in ξ_1,ξ_2, and ξ_3

(iv) Random control parameter for the probabilistic learning inference. The components of the $\mathbb{R}^3$-valued random control parameter $\mathbf{W}$ are $(\log C_{\text{bulk}}, \log C_{\text{shear}}, \log\delta_{\mathbb{C}})$ in which C_{bulk} and C_{shear} are Gamma independent random variables (see [244]) with mean values $\underline{C}_{\text{bulk}}$, $\underline{C}_{\text{shear}}$ and coefficients of variation $\delta_{\text{bulk}} = 0.5$, $\delta_{\text{shear}} = 0.25$, and where $\delta_{\mathbb{C}}$ is a uniform random variable on $[0.1\,,0.5]$.

(v) Stochastic computational model. The finite element method is used for discretizing the weak formulation of the stochastic BVP and yields the stochastic equations on $\mathbb{R}^{n_y}$,

$$\boldsymbol{\mathcal{N}}^{mr}(\boldsymbol{\mathcal{Y}}^{mr}, \boldsymbol{\mathcal{G}}, \mathbf{W}) = \mathbf{0}_{n_y} \ a.s. \quad , \quad 1 \leq m \leq r \leq 3 \, , \tag{8.1}$$

(a) $\boldsymbol{\mathcal{Y}}^{mr}$ is a random variable with values in $\mathbb{R}^{n_y}$ with $n_y = 52\,215$, which is the discretization of random field $\{\mathbf{Y}^{mr}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \overline{\Omega}\}$.

(b) $\boldsymbol{\mathcal{G}}$ is a random variable with values in $\mathbb{R}^{n_g}$ with $n_g = 3\,626\,800$, which is made up of the values of the random fourth-order tensor-valued apparent elasticity field $\{\mathbb{C}(\boldsymbol{\xi}), \boldsymbol{\xi} \in \Omega\}$ at the integration points of the finite elements.

(c) $\mathbf{W}$ is the $\mathbb{R}^{n_w}$-valued random conrol parameter with $n_w = 3$.

(vi) Definition of the statistical moments and their targets. Using the Voigt notation defined by Eq. (5.1), let $[\mathbb{C}^{\text{eff}}]$ be the second-order $\mathbb{M}_6^+$-valued random variable associated with the random tensor $\mathbb{C}^{\text{eff}}$. 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 $[\underline{\mathbb{C}}^{\text{exp}}] \in \mathbb{M}_6^+$. The corresponding constraint equation, which allows the mean value to be fitted, will then be written as

$$E\{[\mathbb{C}^{\text{eff}}]\} = [\underline{\mathbb{C}}^{\text{exp}}] \, . \tag{8.2}$$

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}, which allows the level of statistical fluctuations to be fitted. We then introduce the positive-valued random variable Δ_2^{eff},

$$\Delta_2^{\text{eff}} = \frac{1}{\| [\underline{\mathbb{C}}^{\text{eff}}] \|_F^2} \| [\mathbb{C}^{\text{eff}}] - [\underline{\mathbb{C}}^{\text{eff}}] \|_F^2 \, . \tag{8.3}$$

Let δ^{eff} be defined by $\delta^{\text{eff}} = \left(E\{ \| [\mathbb{C}^{\text{eff}}] - [\underline{\mathbb{C}}^{\text{eff}}] \|_F^2 / \| [\underline{\mathbb{C}}^{\text{eff}}] \|_F^2 \} \right)^{1/2}$. The constraint equation to control the statistical fluctuations are then written as

$$\delta^{\text{eff}} = \delta^{\text{exp}} \quad , \quad \delta^{\text{eff}} = \sqrt{E\{\Delta_2^{\text{eff}}\}} \, , \tag{8.4}$$

which can be rewritten as

$$E\{ \| [\mathbb{C}^{\text{eff}}] \|_F^2 \} = (1 + \delta^{\text{exp}})^2 \, \| [\underline{\mathbb{C}}^{\text{eff}}] \|_F^2 \, . \tag{8.5}$$

It should be noted that, if δ^{eff} goes to zero, then the statistical fluctuations (represented by $\sqrt{\Delta_2^{\text{eff}}}$) of $[\mathbb{C}^{\text{eff}}]$ around $[\underline{\mathbb{C}}^{\text{eff}}]$ 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 for a scale separation).

(vii) Training dataset computed with the prior probability model. The stochastic computational model defined by Eq. (8.1) is used for generating the training set related to the $\mathbb{R}^{n_x}$-valued random variable $\mathbf{X}$ defined by

$$\mathbf{X} = (\{\boldsymbol{\mathcal{Y}}^{mr}, 1 \leq m \leq r \leq 3\}, \boldsymbol{\mathcal{G}}, \mathbf{W}) \quad , \quad \mathbb{R}^{n_x} = \mathbb{R}^{6 \times n_y} \times \mathbb{R}^{n_g} \times \mathbb{R}^{n_w} \, , \tag{8.6}$$

with $n_x = 6\, n_y + n_g + n_w = 3\,942\,093$. The prior probability models of $\boldsymbol{\mathcal{G}}$ and $\mathbf{W}$ are used to generate their independent realizations. The Monte Carlo numerical simulation method is used with $n_d = 50$ independent realizations. We then obtain the training dataset

$$\{\mathbf{x}_d^j, j = 1, \ldots, n_d\} \quad , \quad \mathbf{x}_d^j = (\{\mathbf{y}_d^{mr,j}, 1 \leq m \leq r \leq 3\}, \mathbf{g}_d^j, \mathbf{w}_d^j) \, , \tag{8.7}$$

in which $\mathbf{y}_d^{mr,j} \in \mathbb{R}^{n_y}$ is the solution of the deterministic equation

$$\boldsymbol{\mathcal{N}}^{mr}(\mathbf{y}_d^{mr,j}, \mathbf{g}_d^j, \mathbf{w}_d^j) = \mathbf{0}_{n_y} \, . \tag{8.8}$$

The PCA of $\mathbf{X}$ is carried out by computing the SVD of the matrix $[\mathbf{x}_d^1 \ldots \mathbf{x}_d^{n_d}] \in \mathbb{M}_{n_x, n_d}$, which allows the training dataset $\mathscr{D}_{n_d}(\boldsymbol{\eta}) = \{\boldsymbol{\eta}_d^j, j = 1, \ldots, n_d\}$ of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}$ to be constructed (see Eq. (7.35)) and yields

$$\mathbf{X}^{(\nu)} = \underline{\mathbf{x}} + [\Phi] \, [\kappa]^{1/2} \, \mathbf{H} \, . \tag{8.9}$$

Since $n_d = 50$, dimension ν is chosen to its maximum value that is $\nu = n_d - 1 = 49$.

(viii) Random normalized residue induced by the use of the constrained learned set and its target. For $\boldsymbol{\lambda}$ fixed in $\mathscr{C}_{\text{ad},\boldsymbol{\lambda}}$, let $\mathbf{H}_{\boldsymbol{\lambda}}$ be the $\mathbb{R}^{\nu}$-valued random variable defined in Theorem 3, which is introduced by the learning algorithm under constraints. Let $\boldsymbol{\eta}_{\lambda}^{\ell}$ be a learned realization of $\mathbf{H}_{\boldsymbol{\lambda}}$ and let $(\mathbf{y}_{\lambda}^{mr,\ell}, \mathbf{g}_{\boldsymbol{\lambda}}^{\ell}, \mathbf{w}_{\boldsymbol{\lambda}}^{\ell})$ be the corresponding learned realization of $(\boldsymbol{\mathcal{Y}}_{\lambda}^{mr}, \boldsymbol{\mathcal{G}}_{\boldsymbol{\lambda}}, \mathbf{W}_{\boldsymbol{\lambda}})$ extracted from $\mathbf{x}_{\boldsymbol{\lambda}}^{\ell} = \underline{\mathbf{x}} + [\Phi]\,[\kappa]^{1/2}\,\boldsymbol{\eta}_{\lambda}^{\ell}$. Let $\boldsymbol{\mathcal{R}}_{\boldsymbol{\lambda}}^{mr,\ell} = \boldsymbol{\mathcal{N}}^{mr}(\mathbf{y}_{\lambda}^{mr,\ell}, \mathbf{g}_{\boldsymbol{\lambda}}^{\ell}, \mathbf{w}_{\boldsymbol{\lambda}}^{\ell})$ be the realization of the $\mathbb{R}^{n_y}$-valued random residue computed by the computational model equation and let $\hat{\rho}_{\boldsymbol{\lambda}}^{\ell}$ be defined by

$$\hat{\rho}_{\boldsymbol{\lambda}}^{\ell} = \frac{1}{\sqrt{6\,n_y}} \left(\sum_{1 \le m \le r \le 3} \| \boldsymbol{\mathcal{R}}_{\boldsymbol{\lambda}}^{mr,\ell} \|^2 \right)^{1/2} . \tag{8.10}$$

The random normalized residue is defined as the positive-valued random variable $R_{\boldsymbol{\lambda}}$ whose realization ℓ is $\rho_{\boldsymbol{\lambda}}^{\ell} = \hat{\rho}_{\boldsymbol{\lambda}}^{\ell} / \underline{\hat{\rho}}_0$ in which $\underline{\hat{\rho}}_0$ is an adapted constant for the normalization (see [242]) such that the constraint equation for the random residue is defined by

$$E\{R_{\boldsymbol{\lambda}}^2\} = b_R^c \quad , \quad b_R^c = 1\,. \tag{8.11}$$

(ix) Defining function $\mathbf{h}^c$ related to the constraints and defining the targets represented by $\mathbf{b}^c$. The constraints are defined by the function $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{n_c}$ and by the target represented by vector $\mathbf{b}^c$ given in $\mathbb{R}^{n_c}$. As previously explained, three constraints are introduced: the second-order moment of the random normalized residue (see Eq. (8.11) and two statistical moments defined by Eqs. (8.2) and (8.5). As previously, for $\boldsymbol{\lambda}$ fixed in $\mathscr{C}_{\text{ad},\boldsymbol{\lambda}}$, let $\mathbf{H}_{\boldsymbol{\lambda}}$ be the $\mathbb{R}^{\nu}$-valued random variable defined in Theorem 3.

(a) The first constraint is given by Eq. (8.11). The random normalized residue $R_{\boldsymbol{\lambda}}$ is an implicit function of $\mathbf{H}_{\boldsymbol{\lambda}}$, which is rewritten as $R_{\boldsymbol{\lambda}} = r(\mathbf{H}_{\boldsymbol{\lambda}})$. Therefore, Eq. (8.11) is rewritten as $E\{\, h_R^c(\mathbf{H}_{\boldsymbol{\lambda}})\} = b_R^c$ in which $h_R^c(\boldsymbol{\eta}) = r(\boldsymbol{\eta})^2$.

(b) The second constraint is given by Eq. (8.2). Transforming the upper triangular part of the matrices that belong to $\mathbb{M}_6^+$ as vectors in $\mathbb{R}^{21}$ yields $E\{\, \mathbf{h}_{\mathbb{C}}^c(\mathbf{H}_{\boldsymbol{\lambda}})\} = \mathbf{b}_{\mathbb{C}}^c \in \mathbb{R}^{21}$ in which $\boldsymbol{\eta} \mapsto \mathbf{h}_{\mathbb{C}}^c(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{21}$ is an implicit function of $\boldsymbol{\eta}$.

(c) The last constraint is given by Eq. (8.5), that is also rewritten as $E\{\, h_{\delta}^c(\mathbf{H}_{\boldsymbol{\lambda}})\} = b_{\delta}^c \in \mathbb{R}^+$ in which $\boldsymbol{\eta} \mapsto h_{\delta}^c(\boldsymbol{\eta}) : \mathbb{R}^{\nu} \to \mathbb{R}$ is a positive-valued implicit function.

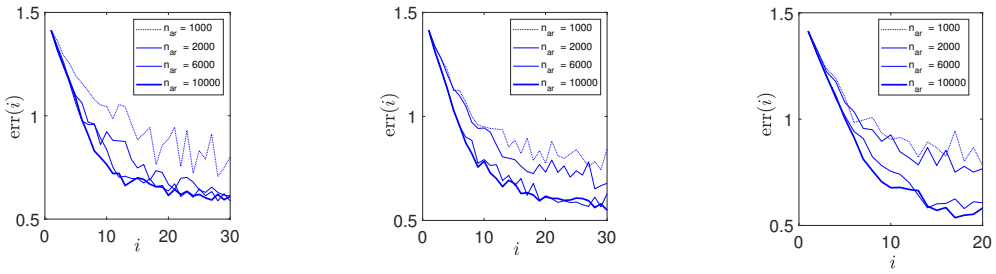
For given $\boldsymbol{\lambda}$, and in particular for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\text{sol}}$ yielding $\mathbf{H}^c = \mathbf{H}_{\boldsymbol{\lambda}^{\text{sol}}}$, the constraint $E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c$ (see Eq. (7.37)) is defined by the implicit function $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta}) = (h_R^c(\boldsymbol{\eta}), \mathbf{h}_{\mathbb{C}}^c(\boldsymbol{\eta}), h_{\delta}^c(\boldsymbol{\eta})) : \mathbb{R}^{\nu} \to \mathbb{R}^{n_c}$ and $b^c = (b_R^c, \mathbf{b}_{\mathbb{C}}^c, b_{\delta}^c) \in \mathbb{R}^{n_c}$ in which $\mathbb{R}^{n_c} = \mathbb{R} \times \mathbb{R}^{21} \times \mathbb{R}$ with $n_c = 23$.

8.2. Numerical results and validation

(i) Convergence in n_{learn} of the iterative sequence $\{\boldsymbol{\lambda}^i\}_i$ of the Lagrange multiplier $\boldsymbol{\lambda}$. We define an error function $i \mapsto \text{err}(i)$ to study the convergence of the iteration algorithm (similar to the one described in Appendix B) that allows the Lagrange multiplier $\boldsymbol{\lambda}^{\text{sol}}$ defined in Theorem 3 to be computed. This error function is defined by $\text{err}(i) = \{(\text{err}_{\mathbb{C}}(i) \,/\, \text{err}_{\mathbb{C}}(1))^2 + (\text{err}_{\delta}(i) \,/\, \text{err}_{\delta}(1))^2\}^{1/2}$ in which $\text{err}_{\mathbb{C}}(i) = \| \mathbf{b}_{\mathbb{C}}^c - E\{\, \mathbf{h}_{\mathbb{C}}^c(\mathbf{H}_{\boldsymbol{\lambda}^i})\} \| \,/\, \| \mathbf{b}_{\mathbb{C}}^c \|$ and $\text{err}_{\delta}(i) = | b_{\delta}^c - E\{\, h_{\delta}^c(\mathbf{H}_{\boldsymbol{\lambda}^i})\} | \,/\, b_{\delta}^c$. For each one of the three cases of spatial correlation lengths and for $n_{\text{learn}} = 1\,000$, $2\,000$, $6\,000$, and $10\,000$, Fig. 9 displays the graph of the error as a function of iteration number i. It can be seen that convergence is reached for $n_{\text{learn}} = 10\,000$ and that, at convergence, function $i \mapsto \text{err}(i)$ is relatively smooth. These graphs show a good illustration of the convergence of the sequence in n_{learn} of the MCMC generator using the statistical surrogate model $\hat{\mathbf{h}}^{n_{\text{learn}}}$ of $\mathbf{h}^c$ (see Theorem 5).

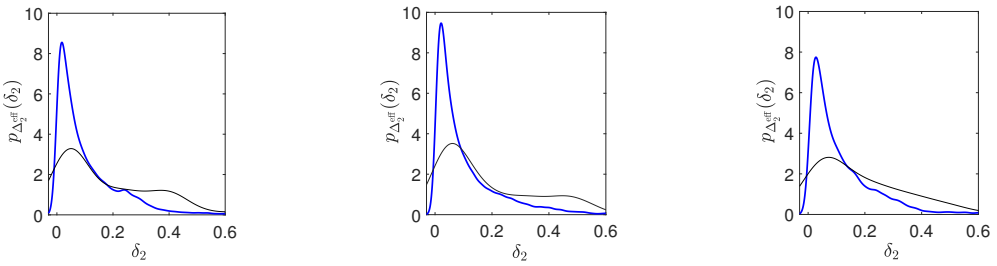
(ii) Posterior pdf of centered statistical fluctuations of the random effective/apparent elasticity matrix estimated with the constrained learned dataset. For each one of the three cases of spatial correlation lengths, Fig. 10 shows the graph of the posterior pdf of random variable Δ_2^{eff} defined by Eq. (8.3), which is estimated with the constrained learned dataset for $n_{\text{learn}} = 10\,000$ and with the training dataset (prior model).

(iii) Posterior pdf of the random residue of the stochastic equation estimated with the constrained learned set. For each one of the three cases of spatial correlation lengths, Fig. 11 displays the posterior pdf of the random normalized



(a) SC1: $\underline{\mathbf{L}}_c = (0.1, 0.1, 0.1)$ (b) SC2: $\underline{\mathbf{L}}_c = (0.3, 0.3, 0.1)$ (c) SC3: $\underline{\mathbf{L}}_c = (0.5, 0.5, 0.2)$

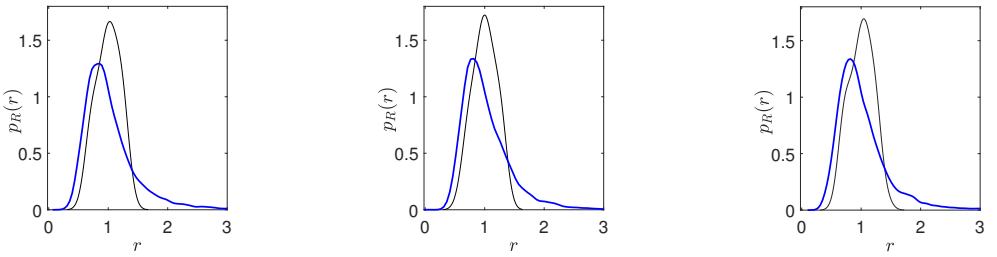
Figure 9: For each one of the three cases of spatial correlation lengths and for $n_{\text{learn}} = 1\,000$, $2\,000$, $6\,000$, and $10\,000$, graph of the error function $i \mapsto \text{err}(i)$ allowing the convergence analysis of the iteration algorithm to compute Lagrange multiplier λ^{sol}. [From [242]].



(a) SC1: $\underline{\mathbf{L}}_c = (0.1, 0.1, 0.1)$ (b) SC2: $\underline{\mathbf{L}}_c = (0.3, 0.3, 0.1)$ (c) SC3: $\underline{\mathbf{L}}_c = (0.5, 0.5, 0.2)$

Figure 10: Posterior pdf of random variable Δ_2^{eff}, estimated with the constrained learned dataset for $n_{\text{learn}} = 10\,000$ (blue thick line), and estimated with the training dataset (black thin line). [From [242]].

residue $R = R_{\lambda^{\text{sol}}}$ estimated with the constrained learned dataset for $n_{\text{learn}} = 10\,000$ and its prior counterpart $R = R_0$ estimated with the training dataset (prior model). It should be noted that the mean value of the random normalized residue for the constrained dataset remains close to the one estimated with the training dataset, which was expected.



(a) SC1: $\underline{\mathbf{L}}_c = (0.1, 0.1, 0.1)$ (b) SC2: $\underline{\mathbf{L}}_c = (0.3, 0.3, 0.1)$ (c) SC3: $\underline{\mathbf{L}}_c = (0.5, 0.5, 0.2)$

Figure 11: Posterior pdf of random normalized residue R, estimated with the constrained learned dataset for $n_{\text{learn}} = 10\,000$ (blue thick line), and its prior counterpart estimated with the training dataset (black thin line). [From [242]].

(iv) Second-order statistics of the random residue and of the effective/apparent elasticity matrix estimated with the constrained learned dataset. For cases SC1, SC2, and SC3, Table 2 gives the posterior statistics computed with the constrained learned set for $n_{\text{learn}} = 10\,000$ (subscript "c"), the prior statistics computed with the training set (subscript "d"), and the targets (superscript "exp"). It can be read,

(a) the second-order moment of the random normalized residue $E\{R_c^2\}$ and the corresponding target b_R^c.
(b) the Frobenius norm $\| [\underline{\mathbb{C}}^{\text{eff}}] \|_F$ of the mean value of the random effective/apparent elasticity matrix.
(c) the coefficient of dispersion $\delta_{\text{ML}}^{\text{eff}} = \{\max_{\delta_2} p_{\Delta_2^{\text{eff}}}(\delta_2)\}^{1/2}$ in which $p_{\Delta_2^{\text{eff}}}$ is the pdf shown in Fig. 10 of random variable Δ_2^{eff} defined by Eq. (8.3).

Table 2: For cases SC1, SC2, and SC3: posterior statistics computed with the constrained learned set for $n_{\text{learn}} = 10\,000$ (subscript "c"), prior statistics computed with the training dataset (subscript "d"), and targets (supercript "exp"). [From [242]].

	SC1	SC2	SC3
$E\{R_c^2\}$	1.2938	1.2687	1.2413
b_R^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

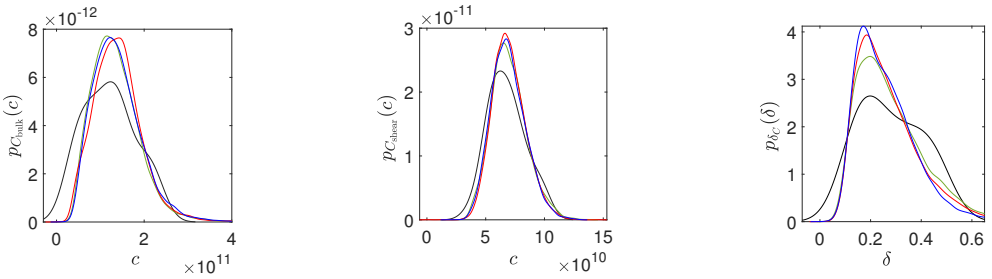
(v) Posterior mean value of the random effective/apparent elasticity matrix estimated with the constrained learned dataset. For the same three cases, Table 3 gives the values of the entries of the mean matrices $[\underline{\mathbb{C}}_d^{\text{eff}}]$ computed with the training set, $[\underline{\mathbb{C}}_c^{\text{eff}}]$ computed with the constrained learned dataset for $n_{\text{learn}} = 10\,000$, 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.

Table 3: For cases SC1, SC2, and SC3: values of $[\underline{\mathbb{C}}_d^{\text{eff}}]_{ij}$ computed with the training dataset, $[\underline{\mathbb{C}}_c^{\text{eff}}]_{ij}$ computed with the constrained learned dataset for $n_{\text{learn}} = 10\,000$, and $[\underline{\mathbb{C}}^{\text{exp}}]_{ij}$ for the targets. [From [242]].

	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

(vi) 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 $\mathbf{G}$ (discretized in a vector of 3 626 800 components) and the $\mathbb{R}^3$-valued random variable $\mathbf{W}$ with components $(C_{\text{bulk}}, C_{\text{shear}}, \delta_{\mathbb{C}})$. The random bulk modulus C_{bulk} and the random shear modulus C_{shear} control the elasticity tensor of the nominal (mean) isotropic model at mesoscale. The dispersion coefficient $\delta_{\mathbb{C}}$ controls the level of anisotropic statistical fluctuations of the random apparent elasticity field at mesoscale. For cases SC1, SC2, and SC3, Fig. 12 displays the posterior pdf $c \mapsto p_{C_{\text{bulk}}}(c)$ of C_{bulk} (figure-a), $c \mapsto p_{C_{\text{shear}}}(c)$ of C_{shear} (figure-b), and $c \mapsto p_{\delta_{\mathbb{C}}}(c)$ of $\delta_{\mathbb{C}}$ (figure-c), estimated with the constrained learned dataset for $n_{\text{learn}} = 10\,000$, and their prior counterparts estimated with the training dataset constructed using the prior probability model. It should be noted that, each random variable C_{bulk}, C_{shear}, or $\delta_{\mathbb{C}}$, has the same prior probability model for the three cases and consequently, does not depend on the case, contrary to its posterior probability model that depends on it.

(vii) Discussion about the presented results. The results obtained with the posterior model (see Tables 2 and 3) show that the constrained learned dataset significantly improves the prior probability model used for generating the training



(a) Posterior pdf $p_{C_{bulk}}(c)$ (b) Posterior pdf $p_{C_{shear}}(c)$ (c) Posterior pdf $p_{\delta_{\mathbb{C}}}(\delta)$

Figure 12: Posterior pdf estimated with the constrained learned set for $n_{\text{learn}} = 10\,000$, for cases, SC1 (green line), SC2 (red line), and SC3 (blue line), and prior pdf estimated with the training set (black line). [From [242]].

dataset. The comparison of the posterior statistics with the targets are good. The convergence of the sequence of MCMC generators with respect to the number of points generated in the constrained learned dataset is good as shown in Figs. 9 in accordance to Theorem 5. As expected, Fig. 11 and Table 2 show that the residue is controlled and remains close to the reference (the training) for the optimal solution. The target is "well" reached for the mean value of the random effective/apparent elasticity matrix (see Table 3). A last comment concerns the effects of no scale separation. As expected, for the three cases SC1, SC2, and SC3, Table 2 shows that the coefficient of dispersion is significant and increases with the spatial correlation lengths of 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.

Appendix A. Analytical examples of classical probability distributions deduced from MaxEnt and a few associated properties

Appendix A.1. Case of a real-valued random variable

Let X be a real-valued random variable, with pdf p_W on $\mathbb{R}$, whose support is $\mathscr{C}_w \subseteq \mathbb{R}$.

(i) Uniform distribution. If the available information is only the support $\mathscr{C}_w = [a, b]$ with $-\infty < a < b < +\infty$, then MaxEnt yields a uniform distribution on $[a, b]$ whose pdf is written as

$$p_W(w) = (b-a)^{-1} 1_{[a,b]}(w) \quad , \quad \forall w \in \mathbb{R}\,. \tag{A.1}$$

(ii) Gamma distribution. If the available information is made up of the support $\mathscr{C}_w =]0\,, +\infty[$, of the mean value $m_W = E\{W\}$ given in $\mathscr{C}_w$, and of the constraint $E\{\log(W)\} = c$ with $|c| < +\infty$, then MaxEnt yields a Gamma distribution whose pdf is written as

$$p_W(w) = 1_{]0,+\infty[}(w)\, \frac{(\delta^{-2})^{\frac{1}{\delta^2}}}{\Gamma(\delta^{-2})\, m_W} \left(\frac{w}{m_W}\right)^{\frac{1}{\delta^2}-1} \exp\left\{-\frac{w}{\delta^2 m_W}\right\} \quad , \quad \forall w \in \mathbb{R}\,, \tag{A.2}$$

in which $\delta = \sigma_W / m_W \in [0\,, 1/\sqrt{2}\,[$ is the coefficient of variation of W, $\sigma_W = (E\{(W - m_W)^2\})^{1/2}$ its standard deviation, and where $\Gamma(\alpha)$ is the Gamma function. Constant c has been eliminated for the benefit of δ in order to perform a reparameterization in δ instead of keeping c. It can be seen that W^{-1}, which exists almost surely (a.s.), is a second-order random variable $E\{W^{-2}\} < +\infty$ thanks to the constraint $|E\{\log(W)\}| < +\infty$.

(iii) Gaussian distribution. If the available information is made up of the support $\mathscr{C}_w =]-\infty\,, +\infty[= \mathbb{R}$, of the mean value $m_W = E\{W\}$ given in $\mathscr{C}_w$, and of the standard deviation $\sigma_W > 0$, then MaxEnt yields a Gaussian distribution

whose pdf is written as

$$p_W(w) = \frac{1}{\sqrt{2\pi}\,\sigma_W} \exp\left\{-\frac{1}{2\sigma_W^2}(w - m_W)^2\right\} \quad , \quad \forall w \in \mathbb{R}\,. \tag{A.3}$$

It can be seen that W^{-1} is not a second-order random variable because $E\{W^{-2}\} = +\infty$. Since the support of p_W is $\mathbb{R}$, a Gaussian random variable cannot be used for a positive-valued random variable.

Appendix A.2. Case of a vector-valued random variable

Let $\mathbf{W}$ be a $\mathbb{R}^{n_w}$-valued random variable with $n_w > 1$ and pdf $p_{\mathbf{W}}$ on $\mathbb{R}^{n_w}$, whose support is $\mathscr{C}_w \subseteq \mathbb{R}^{n_w}$.

(i) Multivariate exponential distribution. If the available information is made up of the support $\mathscr{C}_w =]0\,, +\infty[^{n_w}$ and of the mean value $\mathbf{m}_{\mathbf{W}} = (m_1, \ldots, m_{n_w}) = E\{\mathbf{W}\}$ given in $\mathscr{C}_w$, then MaxEnt yields the pdf,

$$p_W(w) = p_{W_1}(w_1) \times \ldots \times p_{W_{n_w}}(w_{n_w}) \quad , \quad \forall \mathbf{w} = (w_1, \ldots, w_{n_w}) \in \mathbb{R}^{n_w}\,, \tag{A.4}$$

in which $w_j \mapsto p_{W_j}(w_j)$ is the pdf of the positive-valued random variable W_j, which is written as

$$p_{W_j}(w_j) = 1_{]0,+\infty[}(w_j)\,\frac{1}{m_j} \exp\left\{-\frac{w_j}{m_j}\right\} \quad , \quad \forall w_j \in \mathbb{R}\,. \tag{A.5}$$

Equation (A.4) shows that the real-valued random variables $W_1, \ldots, W_{n_w}$ are mutually independent and, for each j, the positive-valued random variable W_j has an exponential pdf. It should be noted that, although the random variable W_j^{-1} exists almost surely, this random variable is not a second-order random variable because $E\{W_j^{-2}\} = +\infty$. It should also be noted that MaxEnt yields independent random variables $W_1, \ldots, W_{n_w}$ because there are no available information concerning some statistical moments coupling the components.

(ii) Multivariate Gaussian distribution. Let us assume that $\mathbf{W}$ is a second-order random variable, $E\{\|\mathbf{W}\|^2\} < +\infty$. Let $\mathbf{m}_{\mathbf{W}} = E\{\mathbf{W}\} \in \mathbb{R}^{n_w}$ be its mean vector and let $[C_{\mathbf{W}}] = E\{(\mathbf{W} - \mathbf{m}_{\mathbf{W}})\,(\mathbf{W} - \mathbf{m}_{\mathbf{W}})^T\} \in \mathbb{M}_{n_w}^+$ be its covariance matrix. Its correlation matrix is then $[R_{\mathbf{W}}] = E\{\mathbf{W}\,\mathbf{W}^T\} = [C_{\mathbf{W}}] + \mathbf{m}_{\mathbf{W}}\mathbf{m}_{\mathbf{W}}^T$. If the available information is the support $\mathscr{C}_w = \mathbb{R}^{n_w}$, the mean value $\mathbf{m}_{\mathbf{W}}$ given in $\mathbb{R}^{n_w}$, and the correlation matrix $[R_{\mathbf{W}}]$ given in $\mathbb{M}_{n_w}^+$, then MaxEnt yields a multivariate Gaussian distribution whose pdf is written as

$$p_{\mathbf{W}}(\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^{n_w}\det[C_{\mathbf{W}}]}} \exp\left\{-\frac{1}{2}\langle [C_{\mathbf{W}}]^{-1}(\mathbf{w} - \mathbf{m}_{\mathbf{W}}), (\mathbf{w} - \mathbf{m}_{\mathbf{W}})\rangle\right\} \quad , \quad \forall \mathbf{w} \in \mathbb{R}^{n_w}\,. \tag{A.6}$$

If the covariance matrix is not the identity matrix, then the real-valued random variables $W_1, \ldots, W_{n_w}$ are mutually dependent, but if $[C_{\mathbf{W}}] = [I_{n_w}]$ (uncorrelated random variables), then these real-values random variables are Gaussian and (note that a random vector with uncorrelated components and which is not a Gaussian vector, has generally statistically dependent components). Finally, it should be noted that $\|\mathbf{W}\|^{-1}$ is not a second-order random variable because $E\{\|\mathbf{W}\|^{-2}\} = +\infty$.

Appendix B. Algorithm for computing the Lagrange multipliers for MaxEnt

Under the hypothesis, $\mathscr{C}_w = \mathbb{R}^{n_w}$, an algorithm [1, 351] derived from [352] is presented below for computing the Lagrange multipliers that are the solution of Eq. (3.8). It is proven that $\boldsymbol{\lambda}^{\text{sol}}$ is the unique solution in $\mathscr{C}_\lambda$ of the optimization problem,

$$\boldsymbol{\lambda}^{\text{sol}} = \arg \min_{\boldsymbol{\lambda} \in \mathscr{C}_\lambda \subset \mathbb{R}^{n_c}} \Gamma(\boldsymbol{\lambda})\,, \tag{B.1}$$

in which $\Gamma(\boldsymbol{\lambda}) = \langle \boldsymbol{\lambda}\,, \mathbf{b}^c \rangle - \log(c_0(\boldsymbol{\lambda}))$ and where the constant of normalization $c_0(\boldsymbol{\lambda}) = (\int_{\mathscr{C}_w} \exp(-\langle \boldsymbol{\lambda}, \mathbf{g}^c(\mathbf{w})\rangle)\,d\mathbf{w})^{-1}$ will not be computed. For $\boldsymbol{\lambda}$ fixed in $\mathscr{C}_\lambda$, let $\mathbf{W}_{\boldsymbol{\lambda}}$ be the $\mathbb{R}^{n_w}$-valued random variable with pdf

$$p_{\mathbf{W}_{\boldsymbol{\lambda}}} = c_0(\boldsymbol{\lambda})\, \exp(-\langle \boldsymbol{\lambda},\, \mathbf{g}^c(\mathbf{w})\rangle) \quad , \quad \forall \mathbf{w} \in \mathbb{R}^{n_w}\,. \tag{B.2}$$

Comparing Eq. (B.2) with Eq. (3.7) shows that pdf $p_{\mathbf{W}}$ of $\mathbf{W}$ is equal to $p_{\mathbf{W}_{\boldsymbol{\lambda}^{\mathrm{sol}}}}$. It is then proven that $\boldsymbol{\lambda}^{\mathrm{sol}}$ is the unique solution in $\boldsymbol{\lambda}$ of

$$\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}) = \mathbf{0} \quad , \quad \boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}) = \mathbf{b}^c - E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})\}\,, \tag{B.3}$$

and the Hessian matrix $[\Gamma''(\boldsymbol{\lambda})]$ of $\Gamma(\boldsymbol{\lambda})$ is positive definite and written as

$$[\Gamma''(\boldsymbol{\lambda})] = E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})\,\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})^T\} - E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})\}\,(E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})\})^T \in \mathbb{M}^+_{n_c}\,, \tag{B.4}$$

that is to say, is the covariance matrix of the random variable $\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}})$. Consequently, function Γ is strictly convex. Since the constant of normalization cannot numerically be estimated, a classical descent algorithm cannot be used and is replaced by searching the solution of Eq. (B.3). The Newton iterative algorithm is used to solve this equation with an under-relaxation coefficient $\alpha_i \in]0\,,1[$, which can depends on iteration number i,

$$\boldsymbol{\lambda}^{i+1} = \boldsymbol{\lambda}^i - \alpha_i\,[\Gamma''(\boldsymbol{\lambda}^i)]^{-1}\,\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^i)\,. \tag{B.5}$$

For $\boldsymbol{\lambda}^i$ given in $\mathscr{C}_{\lambda} \subset \mathbb{R}^{n_c}$, $E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}^i}\}$ and $E\{\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}^i})\,\mathbf{g}^c(\mathbf{W}_{\boldsymbol{\lambda}^i})^T\}$ are estimated using realizations of $\mathbf{W}_{\boldsymbol{\lambda}^i}$ generated with a *MCMC algorithm* for which $p_{\mathbf{W}_{\boldsymbol{\lambda}^i}}(\mathbf{w}) = c_0(\boldsymbol{\lambda}^i)\,\exp(-\langle\boldsymbol{\lambda}^i,\,\mathbf{g}^c(\mathbf{w})\rangle)$ is the density of the invariant measure. It should be noted that $c_0(\boldsymbol{\lambda}^i)$ is not used in the MCMC algorithm and consequently, is not calculated. At convergence, the generated realizations of $\mathbf{W}_{\boldsymbol{\lambda}^{\mathrm{sol}}}$ are those of $\mathbf{W}$.

Appendix C. Störmer-Verlet algorithm for solving Eqs. (7.22) to (7.24)

The ISDE defined by Eqs. (7.22) to (7.24) is solved for $r \in [0, r_0]$ with $r_0 = M_0 \times \Delta r$ in which Δr is the sampling step and where M_0 is chosen in order that the solution is stationary for $r \geq r_0$. Therefore, the solution at $r = r_0$ is a random vector that follows the invariant measure. For $\ell = 0, 1, \ldots, M_0$, we consider the sampling points $r_\ell = \ell\,\Delta r$ and the following notations: $[\boldsymbol{\mathcal{Z}}_\ell] = [\boldsymbol{\mathcal{Z}}(r_\ell)]$, $[\boldsymbol{\mathcal{Y}}_\ell] = [\boldsymbol{\mathcal{Y}}(r_\ell)]$, and $[\mathbf{W}^{\mathrm{wien}}_\ell] = [\mathbf{W}^{\mathrm{wien}}(r_\ell)]$. The Störmer-Verlet scheme is written, for $\ell = 0, 1, \ldots, M_0$, as

$$[\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}] = [\boldsymbol{\mathcal{Z}}_\ell] + \frac{\Delta r}{2}\,[\boldsymbol{\mathcal{Y}}_\ell]\,,$$

$$[\boldsymbol{\mathcal{Y}}_{\ell+1}] = \frac{1-\beta}{1+\beta}\,[\boldsymbol{\mathcal{Y}}_\ell] + \frac{\Delta r}{1+\beta}\,[\boldsymbol{\mathcal{L}}_{\ell+\frac{1}{2}}] + \frac{\sqrt{f_0}}{1+\beta}\,[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}]\,[a_m]\,,$$

$$[\boldsymbol{\mathcal{Z}}_{\ell+1}] = [\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}] + \frac{\Delta r}{2}\,[\boldsymbol{\mathcal{Y}}_{\ell+1}]\,,$$

with the initial condition defined by Eq. (7.24), where $\beta = f_0\,\Delta r\,/4$, and where $[\boldsymbol{\mathcal{L}}_{\ell+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu,m}$-valued random variable such that $[\boldsymbol{\mathcal{L}}_{\ell+\frac{1}{2}}] = [\mathscr{L}([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}])] = [L([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}]\,[g_m]^T)]\,[a_m]$. In the above equation, $[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}]$ is a random variable with values in $\mathbb{M}_{\nu,n_d}$, in which the increment $[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}] = [\mathbf{W}^{\mathrm{wien}}(r_{\ell+1})] - [\mathbf{W}^{\mathrm{wien}}(r_\ell)]$. The increments are statistically independent. For all $\alpha = 1, \ldots, \nu$ and for all $j = 1, \ldots, n_d$, the real-valued random variables $\left\{[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}]_{\alpha j}\right\}_{\alpha j}$ are independent, Gaussian, second-order, and centered random variables such that $E\left\{[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}]_{\alpha j}[\Delta\mathbf{W}^{\mathrm{wien}}_{\ell+1}]_{\alpha' j'}\right\} = \Delta r\,\delta_{\alpha\alpha'}\,\delta_{jj'}$. Efficient values for the algorithm parameters have been found as $f_0 = 4$, $M_0 = 30$, and $\Delta r = 2\pi\hat{s}/20$ in which $\hat{s}$ is defined in Section 7.2.1-(iii).

References

[1] 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.

[2] R. Ghanem, P. Spanos, Stochastic Finite Elements: A spectral Approach, (revised edition) Dover Publications, New York, 2003.

[3] O. Le Maître, O. M. Knio, Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics, Springer Science & Business Media, 2010.

[4] A. D. Richardson, M. Aubinet, A. G. Barr, D. Y. Hollinger, A. Ibrom, G. Lasslop, M. Reichstein, Uncertainty quantification, Eddy covariance: A practical guide to measurement and data analysis (2012) 173–209.

[5] R. C. Smith, Uncertainty quantification: theory, implementation, and applications, Vol. 12, SIAM, 2013.

[6] T. J. Sullivan, Introduction to uncertainty quantification, Vol. 63, Springer, 2015.

[7] 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.
[8] M. Loève, Functions aleatoires du second ordre, in: P. Lévy (Ed.), Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris, 1948, pp. 366–420.
[9] J. L. Doob, Stochastic processes, John Wiley & Sons, New York, 1953.
[10] A. V. Skorokhod, M. I. Yadrenko, On absolute continuity of measures corresponding to homogeneous Gaussian fields, Theory of Probability & Its Applications 18 (1) (1973) 27–40. doi:10.1137/1118002.
[11] I. I. Guikhman, A. Skorokhod, Introduction à la Théorie des Processus Aléatoires, Edition Mir, 1980.
[12] P. Krée, C. Soize, Mathematics of Random Phenomena, Reidel Pub. Co, 1986, (first published by Bordas in 1983 and also published by Springer Science & Business Media in 2012).
[13] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differentials Equations, Springer-Verlag, Heidelberg, 1992.
[14] 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.
[15] A. Friedman, Stochastic Differential Equations and Applications, Dover Publications, Inc., Mineola, New York, 2006.
[16] R. Khasminskii, Stochastic Stability of Differential Equations, Vol. 66, Springer-Verlag, Berlin, Heidelberg, 2012, originally published in Russian, by Nauka, Moskow, 1969. First English edition published in 1980 under R.Z. Has'minski in the series Mechanics: Analysis by Sijthoff & Noordhoff. doi:10.1007/978-3-642-23280-0.
[17] M. Rosenblatt, Stationary Sequences and Random Fields, Springer Science & Business Media, 2012.
[18] R. Durrett, Probability, Theory and Examples (5th ed.), Cambridge University Press, Cambridge, 2019.
[19] R. J. Serfling, Approximation theorems of mathematical statistics, Vol. 162, John Wiley & Sons, 1980.
[20] J. E. Gentle, Computational Statistics, Springer, New York, 2019. doi:10.1007/978-0-387-98144-4.
[21] G. Givens, J. Hoeting, Computational Statistics, 2nd Edition, John Wiley and Sons, Hoboken, New Jersey, 2013.
[22] T. W. Anderson, Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, 1958.
[23] C. Fougeaud, A. Fuchs, Statistique, Dunod, Paris, 1967.
[24] E. Parzen, On estimation of a probability density function and mode, Annals of Mathematical Statistics 33 (3) (1962) 1065–1076. doi:10.1214/aoms/1177704472.
[25] G. R. Terrell, D. W. Scott, Variable kernel density estimation, The Annals of Statistics 20 (3) (1992) 1236–1265.
[26] 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.
[27] T. Duong, A. Cowling, I. Koch, M. Wand, Feature significance for multivariate kernel density estimation, Computational Statistics & Data Analysis 52 (9) (2008) 4225–4242. doi:10.1016/j.csda.2008.02.035.
[28] M. Filippone, G. Sanguinetti, Approximate inference of the bandwidth in multivariate kernel density estimation, Computational Statistics & Data Analysis 55 (12) (2011) 3104–3122. doi:10.1016/j.csda.2011.05.023.
[29] N. Zougab, S. Adjabi, C. C. Kokonendji, Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation, Computational Statistics & Data Analysis 75 (2014) 28–38. doi:10.1016/j.csda.2014.02.002.
[30] D. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, 2nd Edition, John Wiley and Sons, New York, 2015.
[31] 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.
[32] J. M. Bernardo, A. F. M. Smith, Bayesian Theory, John Wiley & Sons, Chichester, 2000.
[33] B. P. Carlin, T. A. Louis, Bayesian Methods for Data Analysis, Chapman and Hall/CRC, 2008.
[34] N. Metropolis, S. Ulam, The monte carlo method, Journal of American Statistical Association 49 (1949) 335–341.
[35] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equations of state calculations by fast computing machines, The Journal of Chemical Physics 21 (6) (1953) 1087–1092.
[36] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, Biometrika 109 (1970) 57–97.
[37] S. Geman, D. Geman, Stochastic relaxation, gibbs distribution and the Bayesian distribution of images, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol PAM I-6 (6) (1984) 721–741.
[38] G. Casella, E. George, Explaining the gibbs sampler, The American Statistician 46 (3) (1992) 167–174.
[39] M. Girolami, B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods, Journal of the Royal Statistics Society 73 (2) (2011) 123–214. doi:10.1111/j.1467-9868.2010.00765.x.
[40] R. Neal, MCMC using hamiltonian dynamics, in: S. Brooks, A. Gelman, G. Jones, X.-L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman and Hall-CRC Press, Boca Raton, 2011, Ch. 5, pp. 1–51. doi:10.1201/b10905-6.
[41] 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.
[42] 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.
[43] W. Betz, I. Papaioannou, D. Straub, Transitional markov chain monte carlo: observations and improvements, Journal of Engineering Mechanics 142 (5) (2016) 04016016.
[44] M. D. Parno, Y. M. Marzouk, Transport map accelerated markov chain Monte Carlo, SIAM/ASA Journal on Uncertainty Quantification 6 (2) (2018) 645–682. doi:10.1137/17M1134640.
[45] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (3) (1948) 379–423 & 623–659. doi:10.1002/j.1538-7305.1948.tb01338.x.
[46] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86. doi:10.1214/aoms/1177729694.
[47] E. T. Jaynes, Information theory and statistical mechanics, Physical Review 106 (4) (1957) 620–630. doi:10.1103/PhysRev.106.620.

[48] E. T. Jaynes, Information theory and statistical mechanics. ii, Physical Review 108 (2) (1957) 171–190. doi:10.1103/PhysRev.108.171.
[49] J. N. Kapur, H. K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, San Diego, 1992.
[50] T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, Hoboken, 2006.
[51] R. M. Gray, Entropy and Information Theory, 2nd Edition, Springer, New York, 2011. doi:10.1007/978-1-4419-7970-4.
[52] 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.
[53] M. L. Mehta, Random Matrices, Revised and Enlarged Second Edition, Academic Press, New York, 1991.
[54] C. Soize, A nonparametric model of random uncertainties for reduced matrix models in structural dynamics, Probabilist Engineering Mechanics 15(3) (2000) 277–294. doi:10.1016/S0266-8920(99)00028-4.
[55] C. Soize, Maximum entropy approach for modeling random uncertainties in transient elastodynamics, Journal of the Acoustical Society of America 109 (5) (2001) 1979–1996. doi:10.1121/1.1360716.
[56] C. Soize, Random matrix theory and non-parametric model of random uncertainties, Journal of Sound and Vibration 263 (4) (2003) 893–916. doi:10.1016/S0022-460X(02)01170-7.
[57] C. Soize, Random matrix theory for modeling uncertainties in computational mechanics, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1333–1366. doi:10.1016/j.cma.2004.06.038.
[58] M. Mignolet, C. Soize, Nonparametric stochastic modeling of linear systems with prescribed variance of several natural frequencies, Probabilistic Engineering Mechanics 23(2-3) (2008) 267–278. doi:10.1016/j.probengmech.2007.12.027.
[59] C. Soize, Random matrix models and nonparametric method for uncertainty quantification, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, Vol. 1, Springer, Cham, Switzerland, 2017, pp. 219–287. doi:10.1007/978-3-319-11259-6_5-1.
[60] M. Shinozuka, Simulation of multivariate and multidimensional random processes, The Journal of the Acoustical Society of America 49 (1B) (1971) 357–368. doi:10.1121/1.1912338.
[61] M. I. Yadrenko, Spectral Theory of Random Fields, Optimization Software, 1983.
[62] F. Poirion, C. Soize, Numerical methods and mathematical aspects for simulation of homogeneous and non homogeneous Gaussian vector fields, in: P. Krée, W. Wedig (Eds.), Probabilistic Methods in Applied Physics, Springer-Verlag, Berlin, 1995, pp. 17–53. doi:10.1007/3-540-60214-3-50.
[63] Y. Rozanov, Random Fields and Stochastic Partial Differential Equations, Kluwer Academic Publishers, 1998.
[64] B. Puig, F. Poirion, C. Soize, Non-Gaussian simulation using Hermite polynomial expansion: convergences and algorithms, Probabilistic Engineering Mechanics 17 (3) (2002) 253–264. doi:10.1016/S0266-8920(02)00010-3.
[65] R. Adler, The Geometry of Random Fields, SIAM, 2010.
[66] E. Vanmarcke, Random Fields: Analysis and Synthesis, World Scientific, Singapore, 2010.
[67] N. Leonenko, A. Olenko, Tauberian and abelian theorems for long-range dependent random fields, Methodology and Computing in Applied Probability 15 (4) (2013) 715–742. doi:10.1007/s11009-012-9276-9.
[68] C. Soize, Random vectors and random fields in high dimension: parametric model-based representation, identification from data, and inverse problems, in: R. Ghanem, D. Higdon, H. Owhadi (Eds.), Handbook of Uncertainty Quantification, Vol. 2, Springer, Cham, Switzerland, 2017, Ch. 26, pp. 883–936. doi:10.1007/978-3-319-11259-6_30-1.
[69] A. Malyarenko, M. Ostoja-Starzewski, Tensor-Valued Random Fields for Continuum Physics, Cambridge University Press, 2018.
[70] A. Malyarenko, M. Ostoja-Starzewski, Tensor-and spinor-valued random fields with applications to continuum physics and cosmology, Probability Surveys 20 (2023) 1–86. doi:10.1214/22-PS12.
[71] C. Soize, Non Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators, Computer Methods in Applied Mechanics and Engineering 195 (1-3) (2006) 26–64. doi:10.1016/j.cma.2004.12.014.
[72] 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.
[73] 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.
[74] 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.
[75] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (1) (1938) 897–936.
[76] R. H. Cameron, W. T. Martin, The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals, Annals of Mathematics 48 (2) (1947) 385–392. doi:10.2307/1969178.
[77] R. Ghanem, P. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics - Transactions of the ASME 57 (1) (1990) 197–202.
[78] R. Ghanem, P. D. Spanos, Stochastic Finite Elements: a Spectral Approach, Springer-Verlag, New York, 1991.
[79] 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.
[80] 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.
[81] 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.
[82] B. J. Debusschere, H. N. Najm, P. P. Pébay, O. M. Knio, R. Ghanem, O. P. Le Maître, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM journal on scientific computing 26 (2) (2004) 698–719. doi:10.1137/S1064827503427741.
[83] X. Wan, G. E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, Journal of Computational Physics 209 (2) (2005) 617–642. doi:10.1016/j.jcp.2005.03.023.
[84] 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.
[85] S. Das, R. Ghanem, S. Finette, Polynomial chaos representation of spatio-temporal random fields from experimental measurements, Journal

of Computational Physics 228 (23) (2009) 8726–8751. doi:10.1016/j.jcp.2009.08.025.
[86] C. Soize, R. Ghanem, Reduced chaos decomposition with random coefficients of vector-valued random variables and random fields, Computer Methods in Applied Mechanics and Engineering 198 (21-26) (2009) 1926–1934. doi:10.1016/j.cma.2008.12.035.
[87] O. G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, On the convergence of generalized polynomial chaos expansions, ESAIM: Mathematical Modelling and Numerical Analysis 46 (2) (2012) 317–339. doi:10.1051/m2an/2011045.
[88] 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.
[89] E. Capiez-Lernout, C. Soize, Nonparametric modeling of random uncertainties for dynamic response of mistuned bladed-disks., ASME Journal of Engineering for Gas Turbines and Power 126 (3) (2004) 610–618. doi:10.1115/1.1760527.
[90] C. Soize, A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics, Journal of Sound and Vibration 288 (3) (2005) 623–652. doi:10.1016/j.jsv.2005.07.009.
[91] C. Chen, D. Duhamel, C. Soize, Probabilistic approach for model and data uncertainties and its experimental identification in structural dynamics: Case of composite sandwich panels, Journal of Sound and Vibration 294(1-2) (2006) 64–81. doi:10.1016/j.jsv.2005.10.013.
[92] J.-F. Durand, C. Soize, L. Gagliardini, Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation, Journal of the Acoustical Society of America 124 (3) (2008) 1513–1525. doi:10.1121/1.2953316.
[93] M. Pellissetti, E. Capiez-Lernout, H. Pradlwarter, C. Soize, G. I. Schueller, Reliability analysis of a satellite structure with a parametric and a non-parametric probabilistic model, Computer Methods in Applied Mechanics and Engineering 198 (2) (2008) 344–357.
[94] C. Soize, Stochastic Models of Uncertainties in Computational Mechanics, Lecture Notes in Engineering Mechanics 2, American Society of Civil Engineers (ASCE), 2012. doi:10.1061/9780784412237.
[95] M. Mignolet, C. Soize, J. Avalos, Nonparametric stochastic modeling of structures with uncertain boundary conditions / coupling between substructures, AIAA Journal 51(6) (2013) 1296–1308. doi:10.2514/1.J051555.
[96] R. Capillon, C. Desceliers, C. Soize, Uncertainty quantification in computational linear structural dynamics for viscoelastic composite structures, Computer Methods in Applied Mechanics and Engineering 305 (2016) 154–172. doi:10.1016/j.cma.2016.03.012.
[97] O. Ezvan, A. Batou, C. Soize, L. Gagliardini, Multilevel model reduction for uncertainty quantification in computational structural dynamics, Computational Mechanics 59 (2) (2017) 219–246. doi:10.1007/s00466-016-1348-1.
[98] J. Reyes, C. Desceliers, C. Soize, L. Gagliardini, Multi-frequency model reduction for uncertainty quantification in computational vibroacoutics, Computational Mechanics 69 (2022) 661–682. doi:10.1007/s00466-021-02109-y.
[99] C. Soize, Random matrices in structural acoustics, in: R. Weaver, M. Wright (Eds.), New Directions in Linear Acoustics: Random Matrix Theory, Quantum Chaos and Complexity, Cambridge University Press, Cambridge, 2010, pp. 206–230.
[100] M. Mbaye, C. Soize, J.-P. Ousty, E. Capiez-Lernout, Robust analysis of design in vibration of turbomachines, ASME Journal of Turbomachinery 135 (2) (2013) 021008. doi:10.1115/1.4007442.
[101] R. Ohayon, C. Soize, Computational vibroacoustics in low-and medium-frequency bands: damping, rom, and uq modeling, Applied Sciences 7 (6) (2017) 586. doi:10.3390/app7060586.
[102] Q. Akkaoui, E. Capiez-Lernout, C. Soize, R. Ohayon, Uncertainty quantification for dynamics of geometrically nonlinear structures coupled with internal acoustic fluids in presence of sloshing and capillarity, Journal of Fluids and Structures 94 (2020) 102966. doi:10.1016/j.jfluidstructs.2020.102966.
[103] C. Desceliers, C. Soize, S. Cambier, Non-parametric - parametric model for random uncertainties in nonlinear structural dynamics - application to earthquake engineering, Earthquake Engineering and Structural Dynamics 33 (3) (2004) 315–327. doi:10.1002/eqe.352.
[104] M. Mignolet, C. Soize, Stochastic reduced order models for uncertain geometrically nonlinear dynamical systems, Computer Methods in Applied Mechanics and Engineering 197(45-48) (2008) 3951–3963. doi:10.1016/j.cma.2008.03.032.
[105] E. Capiez-Lernout, C. Soize, M. Mbaye, Mistuning analysis and uncertainty quantification of an industrial bladed disk with geometrical nonlinearity, Journal of Sound and Vibration 356 (10) (2015) 124–143. doi:10.1016/j.jsv.2015.07.006.
[106] E. Capiez-Lernout, C. Soize, An improvement of the uncertainty quantification in computational structural dynamics with nonlinear geometrical effects, International Journal for Uncertainty Quantification 7 (1) (2017) 83–98. doi:10.1615/Int.J.UncertaintyQuantification.2016019141.
[107] C. Soize, C. Farhat, A nonparametric probabilistic approach for quantifying uncertainties in low-dimensional and high-dimensional nonlinear models, International Journal for Numerical Methods in Engineering 109 (6) (2017) 837–888. doi:10.1002/nme .5312.
[108] C. Farhat, A. Bos, P. Avery, C. Soize, Modeling and quantification of model-form uncertainties in eigenvalue computations using a stochastic reduced model, AIAA Journal 56 (3) (2018) 1198–1210. doi:10.2514/1.J056314.
[109] H. Wang, J. Guilleminot, C. Soize, Modeling uncertainties in molecular dynamics simulations using a stochastic reduced-order basis, Computer Methods in Applied Mechanics and Engineering 354 (2019) 37–55. doi:10.1016/j.cma.2019.05.020.
[110] X. Wang, M. P. Mignolet, C. Soize, Structural uncertainty modeling for nonlinear geometric response using nonintrusive reduced order models, Probabilistic Engineering Mechanics 60 (2020) 103033. doi:10.1016/j.probengmech.2020.103033.
[111] M.-J. Azzi, C. Ghnatios, P. Avery, C. Farhat, Acceleration of a physics-based machine learning approach for modeling and quantifying model-form uncertainties and performing model updating, Journal of Computing and Information Science in Engineering 23 (1) (2022) 011009. doi:10.1115/1.4055546.
[112] E. Capiez-Lernout, M. Pellissetti, H. Pradlwarter, G. I. Schueller, C. Soize, Data and model uncertainties in complex aerospace engineering systems, Journal of Sound and Vibration 295 (3-5) (2006) 923–938. doi:10.1016/j.jsv.2006.01.056.
[113] E. Capiez-Lernout, C. Soize, Robust updating of uncertain damping models in structural dynamics for low- and medium-frequency ranges, Mechanical Systems and Signal Processing 22 (8) (2008) 1774–1792.
[114] C. Farhat, R. Tezaur, T. Chapman, P. Avery, C. Soize, Feasible probabilistic learning method for model-form uncertainty quantification in vibration analysis, AIAA Journal 57 (11) (2019) 4978–4991. doi:10.2514/1.J057797.
[115] C. Soize, C. Farhat, Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics, International Journal for Numerical Methods in Engineering 117 (2019) 819–843. doi:10.1002/nme.5980.
[116] G. Fishman, Monte Carlo: Concepts, algorithms, and applications, Springer-Verlag, New York, 1996.

[117] M. Papadrakakis, V. Papadopoulos, Robust and efficient methods for stochastic finite element analysis using monte carlo simulation, Computer Methods in Applied Mechanics and Engineering 134 (134) (1996) 325–340.
[118] S. Au, J. Beck, Important sampling in high dimensions, Structural Safety 25 (2) (2003) 139–163.
[119] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer Science & Business Media, 2005. doi:10.1007/978-1-4757-4145-2.
[120] R. Y. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo Method, Second Edition, John Wiley & Sons, New York, 2008.
[121] G. Schueller, Efficient monte carlo simulation procedures in structural uncertainty and reliability analysis - recent advances, Structural Engineering and Mechanics 32 (1) (2009) 1–20.
[122] H. Pradlwarter, G. Schueller, Local domain Monte Carlo simulation, Structural Safety 32 (5) (2010) 275–280.
[123] I. Babuska, R. Tempone, G. E. Zouraris, Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1251–1294.
[124] H. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 194 (12-16) (2005) 1295–1331.
[125] A. Nouy, A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 196 (45-48) (2007) 4521–4537.
[126] A. Nouy, Generalized spectral decomposition method for solving stochastic finite element equations: Invariant subspace problem and dedicated algorithms, Computer Methods in Applied Mechanics and Engineering 197 (51-52) (2008) 4718–4736.
[127] A. Nouy, O. P. L. Maitre, Generalized spectral decomposition for stochastic nonlinear problems, Journal of Computational Physics 228 (1) (2009) 202–235.
[128] A. Nouy, Proper generalized decomposition and separated representations for the numerical solution of high dimensional stochastic problems, Archives of Computational Methods in Engineering 17 (4) (2010) 403–434.
[129] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, Annual review of fluid mechanics 41 (2009) 35–52. doi:10.1146/annurev.fluid.010908.165248.
[130] R. Ghanem, R. M. Kruger, Numerical solution of spectral stochastic finite element systems, Computer Methods in Applied Mechanics and Engineering 129 (1996) 289–303.
[131] R. Ghanem, J. Red-Horse, Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach, Physica D 133 (1-4) (1999) 137–144.
[132] R. Ghanem, Ingredients for a general purpose stochastic finite elements formulation, Computer Methods in Applied Mechanics and Engineering 168 (1-4) (1999) 19–34. doi:10.1016/S0045-7825(98)00106-6.
[133] M. F. Pellissetti, R. G. Ghanem, Iterative solution of systems of linear equations arising in the context of stochastic finite elements, Advances in engineering software 31 (8-9) (2000) 607–616. doi:10.1016/S0965-9978(00)00034-X.
[134] M. Deb, I. Babuska, J. Oden, Solution of stochastic partial differential equations using galerkin finite element techniques, Computer Methods in Applied Mechanics and Engineering 190 (2001) 6359–6372.
[135] P. Frauenfelder, C. Schwab, R. Todor, Finite elements for elliptic problems with stochastic coefficients, Computer Methods in Applied Mechanics and Engineering 194 (2-5) (2005) 205–228.
[136] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique 15 (1-3) (2006) 81–92. doi:10.3166/remn.15.81-92.
[137] X. Xu, A multiscale stochastic finite element method on elliptic problems involving uncertainties, Computer Methods in Applied Mechanics and Engineering 196 (25-28) (2007) 2723–2736.
[138] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus Mécanique 336 (6) (2008) 518–523. doi:10.1016/j.crme.2008.02.013.
[139] H. G. Matthies, Stochastic finite elements: Computational approaches to stochastic partial differential equations, Zamm-Zeitschrift Fur Angewandte Mathematik Und Mechanik 88 (11) (2008) 849–873.
[140] 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.
[141] 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.
[142] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Vol. 160, Springer Science & Business Media, 2005. doi:10.1007/b138659.
[143] 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.
[144] A. Tarantola, Inverse Problem Theory And Methods For Model Parameter Estimation, Vol. 89, SIAM, Philadelphia, 2005.
[145] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559. doi:10.1017/S0962492910000061.
[146] H. Owhadi, C. Scovel, T. Sullivan, On the brittleness of Bayesian inference, SIAM Review 57 (4) (2015) 566–582. doi:10.1137/130938633.
[147] 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.
[148] 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.
[149] 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.
[150] 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.
[151] 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.
[152] N. Zabaras, B. Ganapathysubramanian, A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach, Journal of Computational Physics 227 (9) (2008) 4697–4735.

[153] X. Ma, N. Zabaras, An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, Inverse Problems 25 (3) (2009) Article Number: 035013.
[154] H. P. Flath, L. C. Wilcox, V. Akçelik, J. Hill, B. van Bloemen Waanders, O. Ghattas, Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations, SIAM Journal on Scientific Computing 33 (1) (2011) 407–432. doi:10.1137/090780717.
[155] 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.
[156] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Karhunen–loève expansion revisited for vector-valued random fields: Scaling, errors and optimal basis., Journal of Computational Physics 242 (2013) 607–622. doi:10.1016/j.jcp.2013.02.036.
[157] H. N. Najm, K. Chowdhary, Inference given summary statistics, in: R. Ghanem, D. Higdon, O. Houman (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 3, pp. 33–67.
[158] 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.
[159] Q. Zhou, W. Liu, J. Li, Y. Marzouk, An approximate empirical Bayesian method for large-scale linear-Gaussian inverse problems, Inverse Problems (2018).
[160] 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.
[161] 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.
[162] S. Das, R. Ghanem, J. C. Spall, Asymptotic sampling distribution for polynomial chaos representation from data: a maximum entropy and fisher information approach, SIAM Journal on Scientific Computing 30 (5) (2008) 2207–2234. doi:10.1137/060652105.
[163] M. Arnst, R. Ghanem, C. Soize, Identification of Bayesian posteriors for coefficients of chaos expansions, Journal of Computational Physics 229 (9) (2010) 3134–3154. doi:10.1016/j.jcp.2009.12.033.
[164] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data, Computer methods in applied mechanics and engineering 199 (33-36) (2010) 2150–2164. doi:10.1016/j.cma.2010.03.013.
[165] 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.
[166] 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.
[167] 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.
[168] R. Madankan, P. Singla, T. Singh, P. D. Scott, Polynomial-chaos-based Bayesian approach for state and parameter estimations, Journal of Guidance, Control, and Dynamics 36 (4) (2013) 1058–1074. doi:10.2514/1.58377.
[169] B. Chen-Charpentier, D. Stanescu, Parameter estimation using polynomial chaos and maximum likelihood, International Journal of Computer Mathematics 91 (2) (2014) 336–346. doi:10.1080/00207160.2013.809069.
[170] A. H. Elsheikh, I. Hoteit, M. F. Wheeler, Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, Computer Methods in Applied Mechanics and Engineering 269 (2014) 515–537. doi:10.1016/j.cma.2013.11.001.
[171] J. B. Nagel, B. Sudret, Spectral likelihood expansions for Bayesian inference, Journal of Computational Physics 309 (2016) 267–294.
[172] I. Sraj, O. P. Le Maître, O. M. Knio, I. Hoteit, Coordinate transformation and polynomial chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function, Computer Methods in Applied Mechanics and Engineering 298 (2016) 205–228. doi:10.1016/j.cma.2015.10.002.
[173] Q. Shao, A. Younes, M. Fahs, T. A. Mara, Bayesian sparse polynomial chaos expansion for global sensitivity analysis, Computer Methods in Applied Mechanics and Engineering 318 (2017) 474–496. doi:10.1016/j.cma.2017.01.033.
[174] C. Soize, E. Capiez-Lernout, J.-F. Durand, C. Fernandez, L. Gagliardini, Probabilistic model identification of uncertainties in computational models for dynamical systems and experimental validation, Computer Methods in Applied Mechanics and Engineering 198 (1) (2008) 150–163. doi:10.1016/j.cma.2008.04.007.
[175] C. Desceliers, C. Soize, S. Naili, G. Haïat, Probabilistic model of the human cortical bone with mechanical alterations in ultrasonic range, Mechanical Systems and Signal Processing 32 (-) (2012) 170–177. doi:10.1016/j.ymssp.2012.03.008.
[176] M. Arnst, C. Soize, Identification and sampling of Bayesian posteriors of high-dimensional symmetric positive-definite matrices for data-driven updating of computational models, Computer Methods in Applied Mechanics and Engineering 352 (2019) 300–323. doi:10.1016/j.cma.2019.04.025.
[177] J. L. Beck, L. S. Katafygiotis, Updating models and their uncertainties. i: Bayesian statistical framework, Journal of Engineering Mechanics 124 (4) (1998) 455–461.
[178] J. Beck, S. Au, Bayesian updating of structural models and reliability using markov chain monte carlo simulation, Journal of Engineering Mechanics - ASCE 128 (4) (2002) 380–391.
[179] J. Ching, J. Beck, K. Porter, Bayesian state and parameter estimation of uncertain dynamical systems, Probabilistic Engineering Mechanics 21 (1) (2006) 81–96.
[180] S. Cheung, J. Beck, Calculation of posterior probabilities for bayesian model class assessment and averaging from posterior samples based on dynamic system data, Computer-Aided Civil and Infrastructure Engineering 25 (5) (2010) 304–321.
[181] L. Parussini, D. Venturi, P. Perdikaris, G. E. Karniadakis, Multi-fidelity Gaussian process regression for prediction of random fields, Journal of Computational Physics 336 (2017) 36–50.
[182] W. Jiang, M. Bogdan, J. Josse, S. Majewski, B. Miasojedow, V. Ročková, T. Group, Adaptive bayesian SLOPE: model selection with incomplete data, Journal of Computational and Graphical Statistics 31 (1) (2022) 113–137. doi:10.1080/10618600.2021.1963263.
[183] J. P. Kleijnen, Kriging metamodeling in simulation: A review, European Journal of Operational Research 192 (3) (2009) 707–716.

doi:10.1016/j.ejor.2007.10.013.
[184] P. G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: applications to kriging surfaces, SIAM Journal on Scientific Computing 36 (4) (2014) A1500–A1524. doi:10.1137/130916138.
[185] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, J. Wiart, A new surrogate modeling technique combining kriging and polynomial chaos expansions–application to uncertainty analysis in computational dosimetry, Journal of Computational Physics 286 (2015) 103–117.
[186] J. P. Kleijnen, Regression and kriging metamodels with their experimental designs in simulation: a review, European Journal of Operational Research 256 (1) (2017) 1–16. doi:10.1016/j.ejor.2016.06.041.
[187] 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.
[188] Z. Liu, D. Lesselier, B. Sudret, J. Wiart, Surrogate modeling based on resampled polynomial chaos expansions, Reliability Engineering & System Safety 202 (2020) 107008. doi:10.1016/j.ress.2020.107008.
[189] Y. Zhou, Z. Lu, J. Hu, Y. Hu, Surrogate modeling of high-dimensional problems via data-driven polynomial chaos expansions and sparse partial least square, Computer Methods in Applied Mechanics and Engineering 364 (2020) 112906. doi:10.1016/j.cma.2020.112906.
[190] J. C. Spall, Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Vol. 65, John Wiley & Sons, 2005.
[191] E. Capiez-Lernout, C. Soize, Robust design optimization in computational mechanics, Journal of Applied Mechanics 75 (2) (2008) 021001. doi:10.1115/1.2775493.
[192] G. I. Schueller, H. A. Jensen, Computational methods in optimization considering uncertainties - an overview, Computer Methods in Applied Mechanics and Engineering 198 (1) (2008) 2–13.
[193] M. S. Pishvaee, M. Rabbani, S. A. Torabi, A robust optimization approach to closed-loop supply chain network design under uncertainty, Applied Mathematical Modelling 35 (2) (2011) 637–649. doi:10.1016/j.apm.2010.07.013.
[194] X. Gu, J. E. Renaud, S. M. Batill, R. M. Brach, A. S. Budhiraja, Worst case propagated uncertainty of multidisciplinary systems in robust design optimization, Structural and Multidisciplinary Optimization 20 (3) (2000) 190–213. doi:10.1007/s001580050148.
[195] T. W. Simpson, T. M. Mauery, J. J. Korte, F. Mistree, Kriging models for global approximation in simulation-based multidisciplinary design optimization, AIAA Journal 39 (12) (2001) 2233–2241. doi:10.2514/2.1234.
[196] H. Agarwal, J. E. Renaud, E. L. Preston, D. Padmanabhan, Uncertainty quantification using evidence theory in multidisciplinary design optimization, Reliability Engineering & System Safety 85 (1-3) (2004) 281–294. doi:10.1016/j.ress.2004.03.017.
[197] 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.
[198] S. Rao, Multiobjective optimization in structural design with uncertain parameters and stochastic processes, AIAA journal 22 (11) (1984) 1670–1678. doi:10.2514/3.8834.
[199] J. Beck, E. Chan, A.Irfanoglu, et al, Multi-criteria optimal structural design under uncertainty, Earthquake Engineering and Structural Dynamics 28 (7) (1999) 741–761.
[200] M. Li, S. Azarm, Multiobjective collaborative robust optimization with interval uncertainty and interdisciplinary uncertainty propagation, Journal of Mechanical Design 130 (8) (2008) 081402. doi:10.1115/1.2936898.
[201] W. Chen, X. Yin, S. Lee, W. K. Liu, A multiscale design methodology for hierarchical systems with random field uncertainty, Journal of Mechanical Design 132 (4) (2010) 041006. doi:10.1115/1.4001210.
[202] N. Hu, Advances in Multiscale Methods with Applications in Optimization, Uncertainty Quantification and Biomechanics, Columbia University, ProQuest Dissertations Publishing, 10151391, 2016.
[203] 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.
[204] 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.
[205] A. Bhosekar, M. Ierapetritou, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, Computers & Chemical Engineering 108 (2018) 250–267. doi:10.1016/j.compchemeng.2017.09.017.
[206] J. Qian, J. Yi, Y. Cheng, J. Liu, Q. Zhou, A sequential constraints updating approach for kriging surrogate model-assisted engineering optimization design problem, Engineering with Computers 36 (3) (2020) 993–1009.
[207] 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.
[208] R. Calandra, A. Seyfarth, J. Peters, M. P. Deisenroth, Bayesian optimization for learning gaits under uncertainty, Annals of Mathematics and Artificial Intelligence 76 (1-2) (2016) 5–23. doi:10.1007/s10472-015-9463-9.
[209] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. de Freitas, Bayesian optimization in a billion dimensions via random embeddings, Journal of Artificial Intelligence Research 55 (2016) 361–387. doi:10.1613/jair.4806.
[210] J. Xie, P. Frazier, S. Chick, Bayesian optimization via simulation with pairwise sampling and correlated pair beliefs, Operations Research 64 (2) (2016) 542–559. doi:10.1287/opre.2016.1480.
[211] R. Ghanem, C. Soize, Probabilistic nonconvex constrained optimization with fixed number of function evaluations, International Journal for Numerical Methods in Engineering 113 (4) (2018) 719–741. doi:10.1002/nme.5632.
[212] R. Ghanem, C. Soize, C. Thimmisetty, Optimal well-placement using probabilistic learning, Data-Enabled Discovery and Applications 2 (1) (2018) 4,1–16. doi:10.1007/s41688-017-0014-x.
[213] C. Soize, Design optimization under uncertainties of a mesoscale implant in biological tissues using a probabilistic learning algorithm, Computational Mechanics 62 (3) (2018) 477–497. doi:10.1007/s00466-017-1509-x.
[214] C. Soize, R. Ghanem, C. Safta, X. Huan, Z. P. Vane, J. C. Oefelein, G. Lacaze, H. N. Najm, Enhancing model predictability for a scramjet using probabilistic learning on manifolds, AIAA Journal 57 (1) (2019) 365–378. doi:10.2514/1.J057069.
[215] R. Ghanem, C. Soize, C. Safta, X. Huan, G. Lacaze, J. C. Oefelein, H. N. Najm, Design optimization of a scramjet under uncertainty using probabilistic learning on manifolds, Journal of Computational Physics 399 (2019) 108930. doi:10.1016/j.jcp.2019.108930.
[216] M. G. Marmarelis, R. G. Ghanem, Data-driven stochastic optimization on manifolds for additive manufacturing, Computational Materials

Science 181 (2020) 109750. doi:10.1016/j.commatsci.2020.109750.
[217] F. Pled, C. Desceliers, T. Zhang, A robust solution of a statistical inverse problem in multiscale computational mechanics using an artificial neural network, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113540. doi:10.1016/j.cma.2020.113540.
[218] E. Capiez-Lernout, C. Soize, Nonlinear stochastic dynamics of detuned bladed disks with uncertain mistuning and detuning optimization using a probabilistic machine learning tool, International Journal of Non-Linear Mechanics 143 (2022) 104023. doi:10.1016/j.ijnonlinmec.2022.104023.
[219] R. Ghanem, C. Soize, L. Mehrez, V. Aitharaju, Probabilistic learning and updating of a digital twin for composite material systems, International Journal for Numerical Methods in Engineering 123 (13) (2022) 3004–3020. doi:10.1002/nme.6430.
[220] O. Ezvan, C. Soize, C. Desceliers, R. Ghanem, Updating an uncertain and expensive computational model in structural dynamics based on one single target frf using a probabilistic learning tool, Computational Mechanics (2023).
[221] K. B. Korb, A. E. Nicholson, Bayesian artificial intelligence, CRC press, Boca Raton, 2010.
[222] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT press, 2012.
[223] Z. Ghahramani, Probabilistic machine learning and artificial intelligence, Nature 521 (7553) (2015) 452–459. doi:10.1038/nature14541.
[224] S. Russel, P. Norvig, Artifical Intelligence, A Modern Approach, Third Edition, Pearson, Harlow, 2016.
[225] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 2000. doi:10.1007/978-1-4757-3264-1.
[226] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Vol. 112, Springer, 2013.
[227] 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.
[228] 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.
[229] A. C. Öztireli, M. Alexa, M. Gross, Spectral sampling of manifolds, ACM Transactions on Graphics (TOG) 29 (6) (2010) 1–8. doi:10.1145/1882261.1866190.
[230] 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.
[231] G. Perrin, C. Soize, S. Marque-Pucheu, J. Garnier, Nested polynomial trends for the improvement of Gaussian process-based predictors, Journal of Computational Physics 346 (2017) 389–402. doi:10.1016/j.jcp.2017.05.051.
[232] C. Soize, R. Ghanem, Polynomial chaos representation of databases on manifolds, Journal of Computational Physics 335 (2017) 201–221. doi:10.1016/j.jcp.2017.01.031.
[233] G. Perrin, C. Soize, N. Ouhbi, Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints, Computational Statistics & Data Analysis 119 (2018) 139–154. doi:10.1016/j.csda.2017.10.005.
[234] C. Soize, R. Ghanem, C. Safta, X. Huan, Z. P. Vane, J. C. Oefelein, G. Lacaze, H. N. Najm, Q. Tang, X. Chen, Entropy-based closure for probabilistic learning on manifolds, Journal of Computational Physics 388 (2019) 528–533. doi:10.1016/j.jcp.2018.12.029.
[235] Y. Kevrekidis, Manifold learning for parameter reduction, Bulletin of the American Physical Society 65 (2020). doi:10.1016/j.jcp.2019.04.015.
[236] C. Soize, R. Ghanem, Probabilistic learning on manifolds, Foundations of Data Science 2 (3) (2020) 279–307. doi:10.3934/fods.2020013.
[237] K. Kontolati, D. Loukrezis, K. R. dos Santos, D. G. Giovanis, M. D. Shields, Manifold learning-based polynomial chaos expansions for high-dimensional surrogate models, International Journal for Uncertainty Quantification 12 (4) (2022). doi:10.1615/Int.J.UncertaintyQuantification.2022039936.
[238] C. Soize, R. Ghanem, Probabilistic learning on manifolds (PLoM) with partition, International Journal for Numerical Methods in Engineering 123 (1) (2022) 268–290. doi:10.1002/nme.6856.
[239] S. Pan, K. Duraisamy, Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability, SIAM Journal on Applied Dynamical Systems 19 (1) (2020) 480–509. doi:10.1137/19M1267246.
[240] 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.
[241] 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.
[242] C. Soize, Probabilistic learning inference of boundary value problem with uncertainties based on Kullback-Leibler divergence under implicit constraints, Computer Methods in Applied Mechanics and Engineering 395 (2022) 115078. doi:10.1016/j.cma.2022.115078.
[243] C. Soize, Probabilistic learning constrained by realizations using a weak formulation of fourier transform of probability measures, Computational Statistics (2022) 1–30, published online 23 December 2022doi:10.1007/s00180-022-01300-w.
[244] 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.
[245] C. Soize, Nonparametric probabilistic approach of uncertainties for elliptic boundary value problem, International Journal for Numerical Methods in Engineering 80 (6-7) (2009) 673–688. doi:10.1002/nme.2563.
[246] J. Guilleminot, A. Noshadravanb, C. Soize, R. Ghanem, A probabilistic model for bounded elasticity tensor random fields with application to polycrystalline microstructures, Computer Methods in Applied Mechanics and Engineering 200 (17-20) (2011) 1637–1648. doi:10.1016/j.cma.2011.01.016.
[247] 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.
[248] 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.
[249] J. Guilleminot, C. Soize, Non-Gaussian random fields in multiscale mechanics of heterogeneous materials, In: Altenbach H., Ochsner A. (eds) Encyclopedia of Continuum Mechanics (2020) 1826–1834doi:10.1007/978-3-662-55771-6_68.
[250] J. Guilleminot, C. Soize, Non-Gaussian positive-definite matrix-valued random fields with constrained eigenvalues: Application to random

elasticity tensors with uncertain material symmetries, International Journal for Numerical Methods in Engineering 88 (11) (2011) 1128–1151. doi:doi:10.1002/nme.3212.
[251] J. Guilleminot, C. Soize, A stochastic model for elasticity tensors with uncertain material symmetries, International Journal of Solids and Structures 47 (22-23) (2010) 3121–3130. doi:10.1016/j.ijsolstr.2010.07.013.
[252] L. Walpole, Elastic behavior of composite materials: theoretical foundations, Advances in Applied Mechanics 21 (1981) 169–242. doi:10.1016/S0065-2156(08)70332-6.
[253] L. Walpole, Fourth-rank tensors of the thirty-two crystal classes: multiplication tables, Proceedings of the Royal Society A 391 (1984) 149–179. doi:10.1098/rspa.1984.0008.
[254] T.-T. Le, J. Guilleminot, C. Soize, Stochastic continuum modeling of random interphases from atomistic simulations. Application to a polymer nanocomposite, Computer Methods in Applied Mechanics and Engineering 303 (2016) 430–449. doi:10.1016/j.cma.2015.10.006.
[255] E. Kröner, Statistical Continuum Mechanics, Springer-Verlag, 1971.
[256] E. Kröner, Bounds for effective elastic moduli of disordered materials, Journal of the Mechanics and Physics of Solids 25 (2) (1977) 137–155. doi:10.1016/0022-5096(77)90009-6.
[257] S. Nemat-Nasser, M. Hori, Micromechanics: overall properties of heterogeneous materials. Second revised edition, Elsevier, Amsterdam, 1999.
[258] G. W. Milton, The theory of composites, The Theory of Composites, by Graeme W. Milton, pp. 748. ISBN 0521781256. Cambridge, UK: Cambridge University Press, May 2002. (2002) 748.
[259] S. Torquato, Random Heterogeneous Materials, Microstructure and Macroscopic Properties, Springer-Verlag, New York, 2000.
[260] E. Sánchez-Palencia, Non-Homogeneous Media and Vibration Theory, Vol. 127 of Lecture notes in physics, Springer-Verlag, Berlin, 1980.
[261] S. Forest, Homogenization methods and mechanics of generalized continua-part 2, Theoretical and applied mechanics 28-29 (2002) 113–143.
[262] A. Zaoui, Continuum micromechanics: survey, Journal of Engineering Mechanics 128 (8) (2002) 808–816. doi:10.1061/(ASCE)0733-9399(2002)128:8(808).
[263] A. Bourgeat, A. Piatnitski, Approximations of effective coefficients in stochastic homogenization, Annales de l'Institut Henri Poincare (B), Probability and Statistics 40 (2) (2004) 153–165. doi:10.1016/j.anihpb.2003.07.003.
[264] J. Yvonnet, Q.-C. He, C. Toulemonde, Numerical modelling of the effective conductivities of composites with arbitrarily shaped inclusions and highly conducting interface, Composites Science and Technology 68 (13) (2008) 2818–2825. doi:10.1016/j.compscitech.2008.06.008.
[265] P. S. (Ed.), Continuum Micromechanics, Springer-Verlag, Wien, 2014.
[266] G. Papanicolaou, S. Varadhan, Boundary Value Problems with Rapidly Oscillating Random Coefficients, North Holland, 1981.
[267] 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.
[268] 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.
[269] G. Allaire, Homogenization and two-scale convergence, SIAM Journal on Mathematical Analysis 23 (6) (1992) 1482–1518. doi:10.1137/0523084.
[270] K. Sab, On the homogenization and the simulation of random materials, European Journal of Mechanics, A/Solids 11 (5) (1992) 585–607.
[271] K. T. Andrews, S. Wright, Stochastic homogenization of elliptic boundary-value problems with Lp-data, Asymptotic Analysis 17 (3) (1998) 165–184.
[272] 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.
[273] D. Jeulin, M. Ostoja-Starzewski, Mechanics of random and multiscale microstructures, Springer, 2001.
[274] 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.
[275] W. K. Liu, C. McVeigh, Predictive multiscale theory for design of heterogeneous materials, Computational Mechanics 42 (2) (2008) 147–170. doi:10.1007/s00466-007-0176-8.
[276] W. K. Liu, L. Siad, R. Tian, S. Lee, D. Lee, X. Yin, W. Chen, S. Chan, G. B. Olson, L.-E. Lindgen, et al., Complexity science of multiscale materials via stochastic computations, International Journal for Numerical Methods in Engineering 80 (6-7) (2009) 932–978. doi:10.1002/nme.2578.
[277] W. K. Liu, D. Qian, S. Gonella, S. Li, W. Chen, S. Chirputkar, Multiscale methods for mechanical science of complex materials: Bridging from quantum to stochastic multiresolution continuum, International journal for numerical methods in engineering 83 (8-9) (2010) 1039–1080. doi:10.1002/nme.2915.
[278] M. S. Greene, Y. Liu, W. Chen, W. K. Liu, Computational uncertainty analysis in multiresolution materials via stochastic constitutive theory, Computer Methods in Applied Mechanics and Engineering 200 (1-4) (2011) 309–325. doi:10.1016/j.cma.2010.08.013.
[279] A. Clément, C. Soize, J. Yvonnet, Uncertainty quantification in computational stochastic multiscale analysis of nonlinear elastic materials, Computer Methods in Applied Mechanics and Engineering 254 (2013) 61–82. doi:10.1016/j.cma.2012.10.016.
[280] B. Staber, J. Guilleminot, C. Soize, J. Michopoulos, A. Iliopoulos, Stochastic modeling and identification of a hyperelastic constitutive model for laminated composites, Computer Methods in Applied Mechanics and Engineering 347 (2019) 425–444. doi:10.1016/j.cma.2018.12.036.
[281] 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.
[282] Z. Liu, J. A. Moore, S. M. Aldousari, H. S. Hedia, S. A. Asiri, W. K. Liu, A statistical descriptor based volume-integral micromechanics model of heterogeneous material with arbitrary inclusion shape, Computational Mechanics 55 (5) (2015) 963–981. doi:10.1007/s00466-015-1145-2.
[283] M. A. Bessa, R. Bostanabad, Z. Liu, A. Hu, D. W. Apley, C. Brinson, W. Chen, W. K. Liu, A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality, Computer Methods in Applied Mechanics and Engineering 320 (2017) 633–667. doi:10.1016/j.cma.2017.03.037.

[284] H. Yang, X. Guo, S. Tang, W. K. Liu, Derivation of heterogeneous material laws via data-driven principal component expansions, Computational Mechanics 64 (2) (2019) 365–379. doi:10.1007/s00466-019-01728-w.
[285] 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.
[286] 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.
[287] 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.
[288] 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.
[289] 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.
[290] 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.
[291] Y. Yang, A. Misra, Micromechanics based second gradient continuum theory for shear band modeling in cohesive granular materials following damage elasticity, International Journal of Solids and Structures 49 (18) (2012) 2500–2514. doi:10.1016/j.ijsolstr.2012.05.024.
[292] A. Misra, L. Placidi, F. dell'Isola, E. Barchiesi, Identification of a geometrically nonlinear micromorphic continuum via granular micromechanics, Zeitschrift für angewandte Mathematik und Physik 72 (157) (2021) 1–21. doi:10.1007/s00033-021-01587-7.
[293] L. Placidi, D. Timofeev, V. Maksimov, E. Barchiesi, A. Ciallella, A. Misra, F. Dell'isola, Micro-mechano-morphology-informed continuum damage modeling with intrinsic 2nd gradient (pantographic) grain–grain interactions, International Journal of Solids and Structures 254–255 (2022). doi:10.1016/j.ijsolstr.2022.111880.
[294] J.-J. Alibert, P. Seppecher, F. dell'Isola, Truss modular beams with deformation energy depending on higher displacement gradients, Mathematics and Mechanics of Solids 8 (1) (2003) 51–73. doi:10.1177/10812865030080016.
[295] F. dell'Isola, I. Giorgio, M. Pawlikowski, N. Rizzi, Large deformations of planar extensible beams and pantographic lattices: heuristic homogenization, experimental and numerical examples of equilibrium, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 472 (2185) (2016) 20150790–1–23. doi:10.1098/rspa.2015.0790.
[296] A. Ciallella, D. Pasquali, F. D'Annibale, I. Giorgio, Shear rupture mechanism and dissipation phenomena in bias-extension test of pantographic sheets: Numerical modeling and experiments, Mathematics and Mechanics of Solids 27 (10) (2022) 2170–2188. doi:10.1177/10812865221103573.
[297] M. Stilz, F. dell'Isola, I. Giorgio, V. A. Eremeyev, G. Ganzenmüller, S. Hiermaier, Continuum models for pantographic blocks with second gradient energies which are incomplete, Mechanics Research Communications 125 (2022) 103988. doi:10.1016/j.mechrescom.2022.103988.
[298] M. Valmalle, A. Vintache, B. Smaniotto, F. Gutmann, M. Spagnuolo, A. Ciallella, F. Hild, Local-global DVC analyses confirm theoretical predictions for deformation and damage onset in torsion of pantographic metamaterial, Mechanics of Materials 172 (2022) 104379. doi:10.1016/j.mechmat.2022.104379.
[299] E. Barchiesi, F. dell'Isola, P. Seppecher, E. Turco, A beam model for duoskelion structures derived by asymptotic homogenization and its application to axial loading problems, European Journal of Mechanics - A/Solids 98 (2023) 104848. doi:10.1016/j.euromechsol.2022.104848.
[300] A. Ciallella, D. J. Steigmann, Unusual deformation patterns in a second-gradient cylindrical lattice shell: Numerical experiments, Mathematics and Mechanics of Solids 28 (1) (2023) 141–153. doi:10.1177/108128652211018.
[301] P. Trovalusci, M. Ostoja-Starzewski, M. L. De Bellis, A. Murrali, Scale-dependent homogenization of random composites as micropolar continua, European Journal of Mechanics - A/Solids 49 (2015) 396–407. doi:10.1016/j.euromechsol.2014.08.010.
[302] E. Turco, N. L. Rizzi, Pantographic structures presenting statistically distributed defects: Numerical investigations of the effects on deformation fields, Mechanics Research Communications 77 (2016) 65–69. doi:10.1016/j.mechrescom.2016.09.006.
[303] K. Berkache, S. Deogekar, I. Goda, R. Picu, J.-F. Ganghoffer, Construction of second gradient continuum models for random fibrous networks and analysis of size effects, Composite Structures 181 (2017) 347–357. doi:10.1016/j.compstruct.2017.08.078.
[304] G. La Valle, A. Ciallella, G. Falsone, The effect of local random defects on the response of pantographic sheets, Mathematics and Mechanics of Solids 27 (10) (2022) 2147–2169. doi:10.1177/10812865221103482.
[305] G. La Valle, B. E. Abali, G. Falsone, C. Soize, Sensitivity of a granular homogeneous and isotropic second-gradient continuum model with respect to uncertainties, Zeitschrift für Angewandte Mathematik und Mechanik submitted (2023).
[306] J. Guilleminot, C. Soize, D. Kondo, Mesoscale probabilistic models for the elasticity tensor of fiber reinforced composites: experimental identification and numerical aspects, Mechanics of Materials 41 (12) (2009) 1309–1322. doi:10.1016/j.mechmat.2009.08.004.
[307] C. Desceliers, C. Soize, Q. Grimal, M. Talmant, S. Naili, Determination of the random anisotropic elasticity layer using transient wave propagation in a fluid-solid multilayer: Model and experiments, Journal of the Acoustical Society of America 125 (4) (2009) 2027–2034. doi:10.1121/1.3087428.
[308] M.-T. Nguyen, C. Desceliers, C. Soize, J.-M. Allain, H. Gharbi, Multiscale identification of the random elasticity field at mesoscale of a heterogeneous microstructure using multiscale experimental observations, International Journal for Multiscale Computational Engineering 13 (4) (2015). doi:10.1615/IntJMultCompEng.2015011435.
[309] M.-T. Nguyen, J.-M. Allain, H. Gharbi, C. Desceliers, C. Soize, Experimental multiscale measurements for the mechanical identification of a cortical bone by digital image correlation, Journal of the Mechanical Behavior of Biomedical Materials 63 (2016) 125–133. doi:10.1016/j.jmbbm.2016.06.011.
[310] 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.
[311] C. Robert, G. Casella, Monte Carlo statistical methods, Springer Science & Business Media, 2013.
[312] J. Berriot, F. Martin, H. Montes, L. Monnerie, P. Sotta, Reinforcement of model filled elastomers: characterization of the cross-linking

density at the filler–elastomer interface by 1H NMR measurements, Polymer 44 (5) (2003) 1437–1447. doi:10.1016/S0032-3861(02)00882-0.
[313] S. E. Harton, S. K. Kumar, H. Yang, T. Koga, K. Hicks, H. Lee, J. Mijovic, M. Liu, R. S. Vallery, D. W. Gidley, Immobilized polymer layers on spherical nanoparticles, Macromolecules 43 (7) (2010) 3415–3421. doi:10.1021/ma902484d.
[314] D. Fragiadakis, L. Bokobza, P. Pissis, Dynamics near the filler surface in natural rubber-silica nanocomposites, Polymer 52 (14) (2011) 3175–3182. doi:10.1016/j.polymer.2011.04.045.
[315] A. Papon, K. Saalwachter, K. Schaler, L. Guy, F. Lequeux, H. Montes, Low-field NMR investigations of nanocomposites: polymer dynamics and network effects, Macromolecules 44 (4) (2011) 913–922. doi:10.1021/ma102486x.
[316] A. P. Holt, P. J. Griffin, V. Bocharova, A. L. Agapov, A. E. Imel, M. D. Dadmun, J. R. Sangoro, A. P. Sokolov, Dynamics at the polymer/nanoparticle interface in poly (2-vinylpyridine)/silica nanocomposites, Macromolecules 47 (5) (2014) 1837–1843. doi:10.1021/ma5000317.
[317] F. W. Starr, T. B. Schrøder, S. C. Glotzer, Molecular dynamics simulation of a polymer melt with a nanoscopic particle, Macromolecules 35 (11) (2002) 4481–4492. doi:10.1021/ma010626p.
[318] D. Brown, P. Mele, S. Marceau, N. Albérola, A molecular dynamics study of a model nanoparticle embedded in a polymer matrix, Macromolecules 36 (4) (2003) 1395–1406. doi:10.1021/ma020951s.
[319] G. G. Vogiatzis, E. Voyiatzis, D. N. Theodorou, Monte carlo simulations of a coarse grained model for an athermal all-polystyrene nanocomposite system, European polymer journal 47 (4) (2011) 699–712. doi:10.1016/j.eurpolymj.2010.09.017.
[320] A. Ghanbari, T. V. Ndoro, F. Leroy, M. Rahimi, M. C. Boehm, F. Mueller-Plathe, Interphase structure in silica–polystyrene nanocomposites: a coarse-grained molecular dynamics study, Macromolecules 45 (1) (2012) 572–584. doi:10.1021/ma202044e.
[321] T. V. Ndoro, M. C. Boehm, F. Mueller-Plathe, Interface and interphase dynamics of polystyrene chains near grafted and ungrafted silica nanoparticles, Macromolecules 45 (1) (2012) 171–179. doi:10.1021/ma2020613.
[322] T. Chen, H.-J. Qian, Y.-L. Zhu, Z.-Y. Lu, Structure and dynamics properties at interphase region in the composite of polystyrene and cross-linked polystyrene soft nanoparticle, Macromolecules 48 (8) (2015) 2751–2760. doi:10.1021/ma502383n.
[323] D. Brown, V. Marcadon, P. Mele, N. Albérola, Effect of filler particle size on the properties of model nanocomposites, Macromolecules 41 (4) (2008) 1499–1511. doi:10.1021/ma701940j.
[324] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, Journal of computational physics 117 (1) (1995) 1–19. doi:10.1006/jcph.1995.1039.
[325] A. Talwalkar, S. Kumar, H. Rowley, Large-scale manifold learning, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8. doi:10.1109/CVPR.2008.4587670.
[326] Y. Marzouk, T. Moselhy, M. Parno, A. Spantini, Sampling via measure transport: An introduction, Handbook of uncertainty quantification (2016) 1–41doi:10.1007/978-3-319-11259-6_23-1.
[327] 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.
[328] 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.
[329] J. O. Almeida, F. A. Rochinha, A probabilistic learning approach applied to the optimization of wake steering in wind farms, Journal of Computing and Information Science in Engineering 23 (1) (2022) 011003. doi:10.1115/1.4054501.
[330] J. O. Almeida, F. A. Rochinha, Uncertainty quantification of water-flooding in oil reservoirs computational simulations using a probabilistic learning approach, International Journal for Uncertainty Quantification online (2023).
[331] J. Guilleminot, J. E. Dolbow, Data-driven enhancement of fracture paths in random composites, Mechanics Research Communications 103 (2020) 103443. doi:10.1016/j.mechrescom.2019.103443.
[332] M. Arnst, C. Soize, K. Bulthies, Computation of sobol indices in global sensitivity analysis from small data sets by probabilistic learning on manifolds, International Journal for Uncertainty Quantification 11 (2) (2021) 1–23. doi:10.1615/Int.J.UncertaintyQuantification.2020032674.
[333] C. Soize, A. Orcesi, Machine learning for detecting structural changes from dynamic monitoring using the probabilistic learning on manifolds, Structure and Infrastructure Engineering Journal 17 (10) (2021) 1418–1430. doi:10.1080/15732479.2020.1811991.
[334] K. Zhong, J. G. Navarro, S. Govindjee, G. G. Deierlein, Surrogate modeling of structural seismic response using Probabilistic Learning on Manifolds, Earthquake Engineering and Structural Dynamics Online (2023) 1–22.
[335] R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis 21 (1) (2006) 5–30. doi:10.1016/j.acha.2006.04.006.
[336] S. Lafon, A. B. Lee, Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization, IEEE transactions on pattern analysis and machine intelligence 28 (9) (2006) 1393–1403. doi:10.1109/TPAMI.2006.184.
[337] 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.
[338] 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.
[339] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Vol. 31, Springer Science & Business Media, 2006. doi:10.1007/3-540-30666-8.
[340] C. Soize, Optimal partition in terms of independent random vectors of any non-Gaussian vector defined by a set of realizations, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 176–211. doi:10.1137/16M1062223.
[341] 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.
[342] K. Karhunen, Ueber lineare methoden in der wahrscheinlichkeitsrechnung, Amer. Acad. Sci., Fennicade, Ser. A, I 37 (1947) 3–79.
[343] 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.

[344] D. G. Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, New York, 2009.
[345] 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.
[346] 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.
[347] 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.
[348] 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.
[349] 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.
[350] 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.
[351] 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.
[352] 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.