Physics-constrained non-Gaussian probabilistic learning on manifolds

Christian Soize, Roger Ghanem

Physics-constrained non-Gaussian probabilistic learning on manifolds

C. Soize[a,*], R. Ghanem[b]

[a] *Université Paris-Est Marne-la-Vallée, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-La-Vallée, France*
[b] *University of Southern California, 210 KAP Hall, Los Angeles, CA 90089, United States*

Abstract

An extension of the probabilistic learning on manifolds (PLoM), recently introduced by the authors, is presented: in addition to the initial dataset given for performing the probabilistic learning, constraints are given, which correspond to statistics of experiments or of physical models. We consider a non-Gaussian random vector whose unknown probability distribution has to satisfy constraints. The method consists in constructing a generator using the PLoM and the classical Kullback-Leibler minimum cross-entropy principle. The resulting optimization problem is reformulated using Lagrange multipliers associated with the constraints. The optimal solution of the Lagrange multipliers is computed using an efficient iterative algorithm. At each iteration, the MCMC algorithm developed for the PLoM is used, consisting in solving an Itô stochastic differential equation that is projected on a diffusion-maps basis. The method and the algorithm are efficient and allow the construction of probabilistic models for high-dimensional problems from small initial datasets and for which an arbitrary number of constraints are specified. The first application is sufficiently simple in order to be easily reproduced. The second one is relative to a stochastic elliptic boundary value problem in high dimension.

Keywords: probabilistic learning, data driven, statistical constraints, Kullback-Leibler, uncertainty quantification, machine learning

1. Introduction

The consideration of constraints in learning algorithms remains a very important and active research topic. Bayesian updating (see for instance [1, 2, 3, 4, 5]) provides a rational framework for integrating data into predictive models and has been successfully adapted to situations where likelihoods are not readily available either because of expense or because of the nature of available information [6, 7]. In many instances, however, relevant information is available in the form of sample statistics rather than raw data or samplewise constraints. In these settings, an alternative to the Bayesian framework has evolved over the past decades in the form of constrained maximum entropy [8] and constrained minimum cross-entropy [9, 10, 11]. This latter, also known as the Kullback-Liebler divergence forms the basis of the method developed in the present paper.

*Corresponding author: C. Soize, christian.soize@u-pem.fr
Email addresses: christian.soize@u-pem.fr (C. Soize), ghanem@usc.edu (R. Ghanem)

Preprint submitted to International Journal for Numerical Methods in Engineering, 2019, doi:10.1002/nme.6202, September 11, 2019

The approach using the Kullback-Leibler divergence has been used extensively over the last three decades for imposing constraints in the framework of learning with statistical models. Many developments and applications have been published such as, for instance, the following interesting papers devoted to the training of neural network classifiers [12], learning topological [13], ensemble learning related to multi-layer networks [14], constraints on parameters during learning [15], recognition of human activities [16], text categorization [17], exploration of data analysis [18] classification problems with constrained data [19], statistical learning algorithms with linguistic constraints [20], reinforcement learning in finite Markov Decision Processes [21], visual recognition [22, 23, 24], optimal sequential allocation [25], identifiability of parameter learning machines [26], speech enhancement [27], the training of generative neural samplers [28], pessimistic uncertainty quantification problem [29], inference given summary statistics [30, 31], optimization framework concerning distance metric learning [32] unsupervised machine learning techniques to learn latent parameters[33], graph-based semi-supervised learning [34], speech enhancement [35], and finally, to the broad learning systems [36].

In spite of these developments, little progress has been made for non-Gaussian probabilistic learning under statistical constraints, for the case of small datasets. The present paper is devoted to this framework: learning from a high-dimensional small dataset, without invoking the Gaussian assumption. The first ingredient of the proposed approach is the probabilistic learning on manifolds (PLoM) presented in [37], for which complementary developments can be found in [38, 39, 40], and for which applications with validation are presented in [41, 42, 43]. This PLoM allows for constructing a *generated dataset* D_M^g made up of M additional independent realizations $\{\mathbf{x}_{\rm ar}^\ell, \ell = 1, \ldots, M\}$ of a non-Gaussian random vector $\mathbf{X}$, defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{R}^n$, for which the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ is unknown, using only an *initial dataset* D_N (training set) which is a small dataset, made up of N independent realizations $\{\mathbf{x}^j, j = 1, \ldots, N\}$ of $\mathbf{X}$ with $\mathbf{x}^j = \mathbf{X}(\theta_j)$ for $\theta_j \in \Theta$ and such that $M \gg N$.

In this paper, we present an extension of this PLoM for which, not only the initial dataset D_N is given, but in addition, constraints are specified, in the form of statistics synthesized from experimental data, from theoretical considerations, or from numerical simulations. The "physics-constrained" terminology used in the title will be explained in Section 4. We thus construct a non-Gaussian random vector $\mathbf{X}^c$ whose probability distribution $P_{\mathbf{X}^c}(d\mathbf{x})$ must satisfy the specified constraints. The method consists of constructing a generator of $P_{\mathbf{X}^c}(d\mathbf{x})$ using the PLoM, for which $P_{\mathbf{X}^c}(d\mathbf{x})$ is closest to $P_{\mathbf{X}}(d\mathbf{x})$, while satisfying the constraints. For that, the classical Kullback-Leibler minimum cross-entropy principle is used, which is the second ingredient of the proposed method. Consequently, an optimization problem as is formulated using Lagrange multipliers associated with the constraints. The Lagrange multipliers are evaluated as the solution to a convex optimization problem. An efficient iterative algorithm based on the Newton method is proposed. At each Newton iteration, a MCMC algorithm has to be used for estimating the gradient and the Hessian of the objective function. The PLoM, which is detailed in [37], is used for the generation of realizations. It consists in solving an Itô stochastic differential equation (ISDE) that is projected on a diffusion-maps basis. This ISDE corresponds to a nonlinear stochastic dissipative Hamiltonian dynamical system. The method and the algorithm presented are efficient and allow high-dimensional problems to be solved from small initial datasets and for which an arbitrary number of constraints are specified.

The paper is organized as follows. In Section 2, the problem that has to be solved is defined. Section 3 deals with the methodology proposed to take into account the constraints that correspond to statistics coming from experiments or defined by a physical/computational model. Examples of constraints are presented in Section 4. Section 5 is devoted to the computational

statistical method for generating realizations of $\mathbf{X}^c$ using the PLoM and involving the iterative Newton algorithm that requires the use of the MCMC generator based on the ISDE projected on a diffusion-maps basis. Finally, two applications are presented. The first one, presented in Section 6, is completely specified and can thus be easily reproduced by the reader. The second one, presented in Section 7, corresponds to a physical system, which is represented by a stochastic computational model that corresponds to the finite element discretization of a stochastic elliptic boundary value problem.

Notations

Lower-case letters such as q or η are deterministic real variables.
Boldface lower-case letters such as $\mathbf{q}$ or $\boldsymbol{\eta}$ are deterministic vectors.
Lower-case letters $\mathfrak{q}$, $\mathfrak{w}$, and $\mathfrak{x}$ are deterministic vectors.
Upper-case letters such as X or H are real-valued random variables.
Boldface upper-case letters such as $\mathbf{X}$ or $\mathbf{H}$ are vector-valued random variables.
Upper-case letters $\mathbb{Q}$, $\mathbb{U}$, $\mathbb{W}$, and $\mathbb{X}$ are vector-valued random variables.
Lower-case letters between brackets such as $[x]$ or $[\eta]$ are deterministic matrices.
Boldface upper-case letters between brackets such as $[\mathbf{X}]$ or $[\mathbf{H}]$ are matrix-valued random variables.

m_c: number of constraints.
n: dimension ($n = n_q + n_w$) of vector $\mathbf{x}$, $\mathbf{X}$, and $\mathbf{X}^c$.
n_q: dimension of vectors $\mathbf{q}$, $\mathbf{Q}$, and $\mathbf{Q}^c$.
n_w: dimension of vectors $\mathbf{w}$, $\mathbf{W}$, and $\mathbf{W}^c$.
ν: dimension of $\mathbf{H}$ and $\mathbf{H}^c$.
$\mathbf{q}^j$, $\mathbf{q}^{c,\ell}$: realization of $\mathbf{Q}$, $\mathbf{Q}^c$.
$\mathbf{w}^j$, $\mathbf{w}^{c,\ell}$: realization of $\mathbf{W}$, $\mathbf{W}^c$.
$\mathbf{x}^j$, $\mathbf{x}^{c,\ell}$: realization of $\mathbf{X}$, $\mathbf{X}^c$.
M : number of points in the generated dataset D_M^g.
N : number of points in initial dataset D_N.
D_N : initial dataset for $\mathbf{X}$.
D_M^g : generated dataset for $\mathbf{X}^c$.
$\mathbb{D}_N$: initial dataset for $\mathbb{X}$.
$\mathbf{Q}$, $\mathbf{Q}^c$: random QoI (output).
$\mathbf{W}$, $\mathbf{W}^c$: random system parameter (input).
$\mathbf{X}$, $\mathbf{X}^c$: random vector ($\mathbf{Q}$, $\mathbf{W}$), ($\mathbf{Q}^c$, $\mathbf{W}^c$).

$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\mathbb{M}_{n,N}$: set of all the $(n \times N)$ real matrices.
$\mathbb{M}_n$: set of all the square $(n \times n)$ real matrices.
$\mathbb{M}_n^{+0}$: set of all the positive symmetric $(n \times n)$ real matrices.
$\mathbb{M}_n^{+}$: set of all the positive-definite symmetric $(n \times n)$ real matrices.
$\mathbb{R}$: set of all the real numbers.
$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$[x]_{kj}$: entry of matrix $[x]$.
$[x]^T$: transpose of matrix $[x]$.
$\delta_{kk'}$: Kronecker's symbol such that $\delta_{kk'} = 0$ if $k \neq k'$ and $= 1$ if $k = k'$.

$\|\mathbf{x}\|$: usual Euclidean norm in $\mathbb{R}^n$.
$<\mathbf{x},\mathbf{y}>$: usual Euclidean inner product in $\mathbb{R}^n$.
$\delta_{kk'}$: Kronecker's symbol.

In this paper, any Euclidean space $\mathcal{E}$ (such as $\mathbb{R}^n$) is equipped with its Borel field $\mathcal{B}_\mathcal{E}$, which means that $(\mathcal{E}, \mathcal{B}_\mathcal{E})$ is a measurable space on which a probability measure can be defined. The mathematical expectation is denoted by E.

2. Setting the problem to be solved

A typical problem for the use of the approach proposed in this paper is the following. Let $(\mathbf{w}, \mathbb{u}) \mapsto \mathbf{f}(\mathbf{w}, \mathbb{u})$ be any measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^{n_u}$ with values in $\mathbb{R}^{n_q}$ representing a mathematical/computational model. Let $\mathbf{W}$ and $\mathbb{U}$ be two independent (non-Gaussian) random variables defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_w}$ and $\mathbb{R}^{n_u}$, for which the probability distributions $P_\mathbf{W}(d\mathbf{w}) = p_\mathbf{W}(\mathbf{w})\, d\mathbf{w}$ and $P_\mathbb{U}(d\mathbb{u}) = p_\mathbb{U}(\mathbb{u})\, d\mathbb{u}$ are defined by the probability density functions $p_\mathbf{W}$ and $p_\mathbb{U}$ with respect to the Lebesgue measures $d\mathbf{w}$ and $d\mathbb{u}$ on $\mathbb{R}^{n_w}$ and $\mathbb{R}^{n_u}$. The role played by $\mathbf{W}$ and $\mathbb{U}$ are the following. The boundary value problem used for constructing the computational model depends on uncertain parameters (epistemic and/or aleatory uncertainties). These random parameters are, for instance, those used for modeling geometry uncertainties, uncertain boundary conditions, constitutive equations of random media, uncertain physical properties, and can also be a finite spatial discretization of random fields (such as the one used for modeling the elasticity field of a heterogeneous material). Random vector $\mathbf{W}$ is made up of a part of these random parameters, which are used for controlling the system, while random vector $\mathbb{U}$ is made up of the other part of these random parameters, which are not used for controlling the system. Let $\mathbf{Q}$ be the vector of the quantities of interest (QoI) that is a random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^{n_q}$ such that

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbb{U})\,. \tag{1}$$

Random QoI, $\mathbf{Q}$, represents the vector of all the observations performed in the system and is expressed as a function of the solution of the computational model. Let us assume that N calculations have been performed with the mathematical/computational model whose solution is represented by Eq. (1), allowing N independent realizations $\{\mathbf{q}^j, j = 1, \ldots, N\}$ of $\mathbf{Q}$ to be computed such that

$$\mathbf{q}^j = \mathbf{f}(\mathbf{w}^j, \mathbb{u}^j)\,, \tag{2}$$

in which $\{\mathbf{w}^j, j = 1, \ldots, N\}$ and $\{\mathbb{u}^j, j = 1, \ldots, N\}$ are N independent realizations of $(\mathbf{W}, \mathbb{U})$, which have been generated using an adapted generator for $p_\mathbf{W}$ and $p_\mathbb{U}$. We then consider the random variable $\mathbf{X}$ with values in $\mathbb{R}^n$, such that

$$\mathbf{X} = (\mathbf{Q}, \mathbf{W}) \quad , \quad n = n_q + n_w\,. \tag{3}$$

The probabilistic learning will be performed for this random vector $\mathbf{X}$ that includes the control parameter $\mathbf{W}$ and the QoI $\mathbf{Q}$, but that does not include random parameter $\mathbb{U}$. The initial dataset related to random vector $\mathbf{X}$ is then made up of the N independent realizations

$$\{\mathbf{x}^j, j = 1, \ldots, N\} \quad , \quad \mathbf{x}^j = (\mathbf{q}^j, \mathbf{w}^j) \in \mathbb{R}^n. \tag{4}$$

It is assumed that the measurable mapping $\mathbf{f}$ is such that the conditional probability distribution $P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}|\mathbf{w})$ admits a conditional probability density function $p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})$ with respect to the Lebesgue measure $d\mathbf{q}$ on $\mathbb{R}^{n_q}$ for $P_{\mathbf{W}}(d\mathbf{w})$-almost all $\mathbf{w}$ in $\mathbb{R}^{n_w}$ (note that the existence of a pdf for $\mathbb{U}$, $P_{\mathbb{U}}(d\mathbb{u}) = p_{\mathbb{U}}(\mathbb{u})\, d\mathbb{u}$, is necessary but is not sufficient). Under this hypothesis, it can be deduced that the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ of $\mathbf{X}$ admits a density $\mathbf{x} \mapsto p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on $\mathbb{R}^n$. The proof is the following. Let $P_{\mathbf{Q}|\mathbf{W}}(d\mathbf{q}|\mathbf{w})$ be the conditional probability distribution of the conditional random variable $\mathbf{Q}|\mathbf{W}$ given $\mathbf{W} = \mathbf{w}$ for $\mathbf{w} \in \mathbb{R}^{n_w}$, which is such that $\mathbf{Q}|\mathbf{w} = \mathbf{f}(\mathbf{w}, \mathbb{U})$. Since $P_{\mathbf{W}}(d\mathbf{w}) = p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w}$, under the above hypothesis (existence of conditional pdf $\mathbf{q} \mapsto p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})$), the probability distribution of $(\mathbf{Q}, \mathbf{W})$ can be written as $P_{\mathbf{Q},\mathbf{W}}(d\mathbf{q}, d\mathbf{w}) = p_{\mathbf{Q}|\mathbf{W}}(\mathbf{q}|\mathbf{w})\; p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{q}\, d\mathbf{w}$, which shows that the probability distribution of $(\mathbf{Q}, \mathbf{W})$ admits a pdf with respect to $d\mathbf{q}\, d\mathbf{w}$ and consequently, shows that the probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ of $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ admits a density $p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on $\mathbb{R}^n$.

The PLoM allows for generating additional realizations $\{(\mathbf{q}_{\text{ar}}^{\ell}, \mathbf{w}_{\text{ar}}^{\ell}), \ell = 1, \ldots, M\}$ for $M \gg N$ without using the computational model, but using only the training dataset which we denote by D_N. These additional realizations allow, for instance, a cost function $J(\mathbf{w}) = E\{\mathcal{J}(\mathbf{Q}, \mathbf{W})|\mathbf{W} = \mathbf{w}\}$ to be evaluated, in which $(\mathbf{q}, \mathbf{w}) \mapsto \mathcal{J}(\mathbf{q}, \mathbf{w})$ is a given measurable real-valued mapping on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ as well as constraints related to a nonconvex optimization problem [39, 41, 42] and this, without calling the mathematical/computational model.

Sometimes, additional information in the form of statistics may become available, synthesized from partial measurements, published data, or numerical simulations. The goal is then to generate additional realizations by taking into account the constraints defined by these statistics. For instance, statistics can correspond to a specified mean value $\mathbb{m}_k^{\text{exp}}$ and standard deviation $\mathbb{s}_k^{\text{exp}}$ of several components Q_k of $\mathbf{Q}$ (see Section 4). Statistics can also correspond to the mean value of a given stochastic equation involving $\mathbf{Q}$ and $\mathbf{W}$, which models a physical system or a part of it (see Appendix A).

Remark concerning the scaling of the initial dataset. In practice, the initial dataset can be made up of heterogeneous numerical values and must be scaled for performing computational statistics. In this remark, it is then assumed that the initial dataset $\mathbb{D}_N$ is made up of N independent realizations $\{\mathbb{x}^j, j = 1, \ldots, N\}$ of the $\mathbb{R}^n$-valued random variable $\mathbb{X}$. Let $\mathbb{x}^{\max} = \max_j\{\mathbb{x}^j\}$ and $\mathbb{x}^{\min} = \min_j\{\mathbb{x}^j\}$ be vectors in $\mathbb{R}^n$. Let $[\alpha_x]$ be the invertible diagonal $(n \times n)$ real matrix such that $[\alpha_x]_{kk} = (\mathbb{x}_k^{\max} - \mathbb{x}_k^{\min})$ if $\mathbb{x}_k^{\max} - \mathbb{x}_k^{\min} \neq 0$, and $[\alpha_x]_{kk} = 1$ if $\mathbb{x}_k^{\max} - \mathbb{x}_k^{\min} = 0$. The scaling of random vector $\mathbb{X}$ is the $\mathbb{R}^n$-valued random variable $\mathbf{X}$ (that has previously been introduced) such that

$$\mathbb{X} = [\alpha_x]\, \mathbf{X} + \mathbb{x}^{\min} \quad , \quad \mathbf{X} = [\alpha_x]^{-1}(\mathbb{X} - \mathbb{x}^{\min})\,, \tag{5}$$

which shows that the N independent realizations $\{\mathbf{x}^j, j = 1, \ldots, N\}_j$ of $\mathbf{X}$ are such that

$$\mathbf{x}^j = [\alpha_x]^{-1}(\mathbb{x}^j - \mathbb{x}^{\min})\,. \tag{6}$$

The scaled initial dataset D_N is then defined by

$$D_N = \{\mathbf{x}^j, j = 1, \ldots, N\} \quad , \quad (\mathbf{q}^j, \mathbf{w}^j) = \mathbf{x}^j\,. \tag{7}$$

3. Methodology proposed to take into account constraints defined by experiments or by physical models

To take into account statistical constraints in the algorithm of the PLoM, the proposed methodology consists of minimizing the probabilistic distance (or the divergence) between the probability distribution $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$ of the $\mathbb{R}^n$-valued random variable $\mathbf{X}$ estimated in the framework of the PLoM and the probability distribution $P_{\mathbf{X}^c}(d\mathbf{x}) = p_{\mathbf{X}^c}(\mathbf{x})\, d\mathbf{x}$ of the $\mathbb{R}^n$-valued random variable $\mathbf{X}^c$ that, in addition, satisfies the constraints. We thus have to find among all the probability density functions (pdf) $p_{\mathbf{X}^c}$ that satisfy the constraints, the one that is closest to the pdf $p_{\mathbf{X}}$. For that, we propose to use the classical Kullback-Leibler minimum cross-entropy principle [9, 10, 11], which is written as

$$p_{\mathbf{X}^c} = \arg \min_{\widehat{p} \in C_{\mathrm{ad},\mathbf{X}^c}} \int_{\mathbb{R}^n} \widehat{p}(\mathbf{x}) \, \log \frac{\widehat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} \, d\mathbf{x} \,, \tag{8}$$

in which the admissible set $C_{\mathrm{ad},\mathbf{X}^c}$ is defined by

$$C_{\mathrm{ad},\mathbf{X}^c} = \{\mathbf{x} \mapsto \widehat{p}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^+, \int_{\mathbb{R}^n} \widehat{p}(\mathbf{x})\, d\mathbf{x} = 1, \int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x})\, \widehat{p}(\mathbf{x})\, d\mathbf{x} = \boldsymbol{\beta}^c\}. \tag{9}$$

In Eq. (9), $\boldsymbol{\beta}^c$ is a given vector in $\mathbb{R}^{m_c}$ with $m_c \geq 1$ and $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x})$ is a measurable mapping from $\mathbb{R}^n$ into $\mathbb{R}^{m_c}$. This means that, in addition to the normalization condition, pdf $p_{\mathbf{X}^c}$ must be such that the following constraint is satisfied,

$$E\{\mathbf{g}^c(\mathbf{X}^c)\} = \int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x})\, p_{\mathbf{X}^c}(\mathbf{x})\, d\mathbf{x} = \boldsymbol{\beta}^c \,. \tag{10}$$

Remarks

(1) It should be noted that, in the framework of the PLoM, the estimation $p_{\mathbf{X}}$ of the pdf of $\mathbf{X}$, which is performed using the Gaussian kernel-density estimation method and initial dataset D_N, is such that $p_{\mathbf{X}}(\mathbf{x}) > 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

(2) The functional that has to be minimized on $C_{\mathrm{ad},\mathbf{X}^c}$ is the Kullback-Leibler divergence (relative entropy),

$$D(\widehat{p}\,; p_{\mathbf{X}}) = \int_{\mathbb{R}^n} \widehat{p}(\mathbf{x}) \, \log \frac{\widehat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} \, d\mathbf{x} \,, \tag{11}$$

that is nonnegative, additive, and non-symmetric ($D(\widehat{p}\,; p_{\mathbf{X}}) \neq D(p_{\mathbf{X}}; \widehat{p})$). The symmetric divergence

$$D^s(\widehat{p}\,; p_{\mathbf{X}}) = \int_{\mathbb{R}^n} (\widehat{p}(\mathbf{x}) - p_{\mathbf{X}}(\mathbf{x})) \, \log \frac{\widehat{p}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} \, d\mathbf{x} \,,$$

could also be introduced. As explained in [10], the symmetric divergence has exactly the same mathematical properties as the relative entropy, such as nonnegativeness and additivity properties. The introduction of the symmetry property could have only an interest if pdf $p_{\mathbf{X}}$ was constructed as an *a priori* pdf, which is not the case, because $p_{\mathbf{X}}$ is constructed using the nonparametric statistics. In addition, the choice of the relative divergence instead of the symmetric divergence has allowed for developing simpler and more efficient algorithms for solving the optimization problem defined by Eqs. (8) and (9). Consequently, we will pursue the formulation

defined by Eqs. (8) and (9).

(3) The objective of this paper is therefore to use the PLoM for the random variable $\mathbf{X}^c$ with values in $\mathbb{R}^n$ for which its probability distribution $P_{\mathbf{X}^c}(d\mathbf{x}) = p_{\mathbf{X}^c}(\mathbf{x})\, d\mathbf{x}$ is defined by the pdf $p_{\mathbf{X}^c}$ that is unknown but that can possibly be concentrated in the neighborhood of a manifold in $\mathbb{R}^n$.

(i) Data of the problem are only those defined by

(i-1) initial dataset $D_N = \{\mathbf{x}^j, j = 1, \ldots, N\}$ of independent realizations of the $\mathbb{R}^n$-valued random variable $\mathbf{X}$ for which its probability distribution $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$ (defined by a density) is unknown. We consider the small dataset case, which means that N is small.

(i-2) a given vector-valued constraint equation on $\mathbb{R}^{m_c}$ (see Eq. (10)),

$$\int_{\mathbb{R}^n} \mathbf{g}^c(\mathbf{x})\, \widehat{p}_{\mathbf{X}^c}(\mathbf{x})\, d\mathbf{x} = \boldsymbol{\beta}^c \,.$$

(ii) The problem, which has to be solved, consists in constructing M realizations $\{\mathbf{x}^{c,\ell}, \ell = 1, \ldots, M\}$ of random vector $\mathbf{X}^c$ with $M \gg N$, for which pdf $p_{\mathbf{X}^c}$ satisfies the constraint while being the closest pdf to $p_{\mathbf{X}}$.

4. Example of constraints

Using the previous notations, random vector $\mathbf{X}^c$ is written as

$$\mathbf{X}^c = (\mathbf{Q}^c, \mathbf{W}^c) \quad \text{with values in} \quad \mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w} \,. \tag{12}$$

In this section, we give two examples of constraints. The first example corresponds to given statistical constraints defined by experimental second-order moments of $\mathbf{Q}^c$. The second one corresponds to physical constraints defined by the mean value of a stochastic equation that relates $\mathbf{Q}^c$ and $\mathbf{W}^c$. For illustrations, two other examples are given in Appendix A.

4.1. Statistical constraints defined by experimental second-order moments of $\boldsymbol{Q}^c$

Let $\mathcal{J}_{\text{mom}} = \{k_1, \ldots, k_{\kappa_{\text{mom}}}\}$ be $\kappa_{\text{mom}} \leq n_q$ integers such that $\mathcal{J}_{\text{mom}} \subseteq \{1, \ldots, n_q\}$. For $\kappa \in \{1, \ldots, \kappa_{\text{mom}}\}$, let us assume that the experimental mean value, $\mathbb{m}_\kappa^{\text{exp}}$, and the experimental standard deviation, $\mathbb{s}_\kappa^{\text{exp}}$, of the component $Q^c_{k_\kappa}$ of random vector $\mathbf{Q}^c$ are specified. For instance, these statistics could come from measurements. Note that the measured realizations that have been used for estimating these experimental quantities are not available. On the other hand, $\{(\mathbb{m}_\kappa^{\text{exp}}, \mathbb{s}_\kappa^{\text{exp}}), \kappa = 1, \ldots, \kappa_{\text{mom}}\}$ do not correspond to the initial dataset D_N. For instance, D_N could come from numerical simulations which include modeling errors. The experimental second-order moment $\mathbb{r}_\kappa^{\text{exp}} = E\{(X^c_{k_\kappa})^2\}$ of random variable $X^c_{k_\kappa}$, can be deduced and is written as $\mathbb{r}_\kappa^{\text{exp}} = (\mathbb{s}_\kappa^{\text{exp}})^2 + (\mathbb{m}_\kappa^{\text{exp}})^2$. In this case, $m_c = 2\,\kappa_{\text{mom}} \leq 2\, n_q$ and the m_c functions $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g^c_1(\mathbf{x}), \ldots, g^c_{m_c}(\mathbf{x}))$ from $\mathbb{R}^n$ into $\mathbb{R}^{m_c}$ and vector $\boldsymbol{\beta}^c = (\beta^c_1, \ldots, \beta^c_{m_c})$, which define the constraints (see Eq. (10)), are such that, for $\kappa = 1, \ldots, \kappa_{\text{mom}}$,

$$g^c_\kappa(\mathbf{X}^c) = Q^c_{k_\kappa} \quad , \quad \beta^c_\kappa = \mathbb{m}_\kappa^{\text{exp}} \,, \tag{13}$$

$$g^c_{\kappa+\kappa_{\text{mom}}}(\mathbf{X}^c) = (Q^c_{k_\kappa})^2 \quad , \quad \beta^c_{\kappa+\kappa_{\text{mom}}} = \mathbb{r}_\kappa^{\text{exp}} \,. \tag{14}$$

4.2. Physical constraints defined by the mean value of a stochastic equation relating $\boldsymbol{Q}^c$ and $\boldsymbol{W}^c$

Let us assume that the initial dataset $D_N = \{\mathbf{x}^j, j = 1, \ldots, N\}$ corresponds to N experimental realizations (measurements or numerical simulations obtained by the training a stochastic computational model) of random vector $\mathbf{X} = (\mathbf{Q}, \mathbf{W})$ whose probability distribution on $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ admits a density with respect to the Lebesgue measure. In this appendix, we consider the case for which the PLoM is used for generating M additional realizations (using only initial dataset D_N) under the constraints defined by statistical moments (for instance, the mean value) of a stochastic equation corresponding to a stochastic computational model of a physical system. As previously, let $p_{\mathbf{X}^c}$ be the pdf of the random vector $\mathbf{X}^c = (\mathbf{Q}^c, \mathbf{W}^c)$ with values in $\mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$, which has to verify, in statistical mean, the equation $\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c) = \boldsymbol{\beta}^{\text{const}}$, that is to say $E\{\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c)\} = \boldsymbol{\beta}^{\text{const}}$ in which $(\mathbf{q}^c, \mathbf{w}^c) \mapsto \mathbf{f}^{\text{const}}(\mathbf{q}^c, \mathbf{w}^c)$ is a given (deterministic) measurable mapping defined on $\mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$ with values in $\mathbb{R}^{m_c}$ and where $\boldsymbol{\beta}^{\text{const}}$ is a given vector in $\mathbb{R}^{m_c}$. This constraint can then be rewritten as Eq. (10) with $\mathbf{g}^c(\mathbf{X}^c) = \mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c)$ and $\boldsymbol{\beta}^c = \boldsymbol{\beta}^{\text{const}}$. For instance, function $\mathbf{f}^{\text{const}}$ could be written, for $m_c = n_q$, as

$$\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c) = \mathbf{Q}^c - [\underline{B}]\, \mathbf{W}^c \,, \tag{15}$$

in which $[\underline{B}]$ is a given matrix in $\mathbb{M}_{n_q, n_w}$.

It should be noted that this formulation can be extended to a stochastic equation that would be written, for instance, as $E\{\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c, \mathbb{U})\} = \boldsymbol{\beta}^{\text{const}}$ in which $\mathbb{U}$ is a random vector independent of $\mathbf{W}^c$ and where $\mathbf{f}^{\text{const}}(\mathbf{Q}^c, \mathbf{W}^c, \mathbb{U}) = \mathbf{Q}^c - [B(\mathbb{U})]\, \mathbf{W}^c$. Clearly, any more complex stochastic equation can be used.

5. Computational statistical method for generating realizations of $\mathbf{X}^c$ using probabilistic learning on manifolds

In this section, we present a computational method for generating M independent realizations $\mathbf{x}^{c,1}, \ldots, \mathbf{x}^{c,M}$ of random variable $\mathbf{X}^c$ whose pdf $p_{\mathbf{X}^c}$ is defined by Eqs. (8) and (9). Since it is assumed that N is small and since the available information is only made up of $\{\mathbf{x}^j, j = 1, \ldots, N\}$ and of the constraint equation $E\{\mathbf{g}^c(\mathbf{X}^c)\} = \boldsymbol{\beta}^c$, we propose to use the PLoM that will be developed in Section 5.8. The main steps of the computational statistical method are as follows:

1. principal component analysis (PCA) of $\mathbf{X}$ at order $\nu \leq n$ using the set $\{\mathbf{x}^j, j = 1, \ldots, N\}$, which will introduce the normalized non-Gaussian random vector $\mathbf{H}$ with values in $\mathbb{R}^\nu$, centered, with covariance matrix $[I_\nu]$, for which the N independent realizations are $\{\boldsymbol{\eta}^j, j = 1, \ldots, N\}$.
2. Gaussian kernel-density estimation of the pdf $p_{\mathbf{H}}$ of random vector $\mathbf{H}$.
3. reformulating the optimization problem defined by Eqs. (8) and (9) for the construction of pdf $p_{\mathbf{H}^c}$ of $\mathbf{H}^c$ in terms of pdf $p_{\mathbf{H}}$ of $\mathbf{H}$ and of the constraint equation rewritten in terms of $p_{\mathbf{H}^c}$.
4. representation of the solution of the optimization problem with constraints using Lagrange multipliers $(\lambda_0, \boldsymbol{\lambda})$ where λ_0 is associated with the normalization constraint.
5. reformulation introducing a random vector $\mathbf{H}_{\boldsymbol{\lambda}}$ and eliminating λ_0.
6. introducing an iterative algorithm for computing the optimal value $\boldsymbol{\lambda}^{\text{sol}}$ of the Lagrange multiplier $\boldsymbol{\lambda}$.
7. constructing a nonlinear Itô Stochastic Differential Equation (ISDE) as a generator of random variable $\mathbf{H}_{\boldsymbol{\lambda}}$.

8. computing the additional realizations of $\mathbf{H}^c$ using the PLoM and then deducing additional realizations of $\mathbf{X}^c$.
9. estimating statistics of $\mathbf{X}^c$ using the additional realizations.

In the following, each of these steps is expanded into a subsection.

5.1. PCA of random vector $\boldsymbol{X}$

The purpose of the PCA transformation is to improve the statistical condition of $\mathbf{X}$ through de-correlation and normalization. Dimensional reduction is only accepted if it reproduces the original data with a given tolerance. Let $\widehat{\underline{\mathbf{x}}} \in \mathbb{R}^n$ and $[\widehat{C}_{\mathbf{X}}] \in \mathbb{M}_n^+$ be the classical empirical estimates of the mean vector and the covariance matrix of $\mathbf{X}$,

$$\widehat{\underline{\mathbf{x}}} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{x}^j \quad , \quad [\widehat{C}_{\mathbf{X}}] = \frac{1}{N-1}\sum_{j=1}^{N} (\mathbf{x}^j - \widehat{\underline{\mathbf{x}}})\,(\mathbf{x}^j - \widehat{\underline{\mathbf{x}}})^T \,. \tag{16}$$

Let $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n \geq 0$ be the eigenvalues and let $\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^n$ be the orthonormal eigenvectors of $[\widehat{C}_{\mathbf{X}}]$ such that $[\widehat{C}_{\mathbf{X}}]\,\boldsymbol{\varphi}^\alpha = \mu_\alpha\,\boldsymbol{\varphi}^\alpha$. For $\varepsilon > 0$ fixed, let ν be the integer such that $1 \leq \nu \leq n$, verifying,

$$\mathrm{err}_{\mathrm{PCA}}(\nu) = 1 - \frac{\sum_{\alpha=1}^{\nu} \mu_\alpha}{\mathrm{tr}\,[\widehat{C}_{\mathbf{X}}]} \leq \varepsilon \,. \tag{17}$$

The PCA of $\mathbf{X}$ allows for representing $\mathbf{X}$ by $\mathbf{X}^\nu$ such that

$$\mathbf{X}^\nu = \widehat{\underline{\mathbf{x}}} + [\Phi]\,[\mu]^{1/2}\,\mathbf{H} \quad , \quad E\{\|\mathbf{X} - \mathbf{X}^\nu\|^2\} \leq \varepsilon\, E\{\|\mathbf{X}\|^2\}\,, \tag{18}$$

in which $[\Phi] = [\varphi^1 \ldots \varphi^\nu] \in \mathbb{M}_{n,\nu}$ such that $[\Phi]^T\,[\Phi] = [I_\nu]$ and $[\mu]$ is the diagonal $(\nu \times \nu)$ matrix such that $[\mu]_{\alpha\beta} = \mu_\alpha \delta_{\alpha\beta}$. The $\mathbb{R}^\nu$-valued random variable $\mathbf{H}$ is obtained by projection,

$$\mathbf{H} = [\mu]^{-1/2}\,[\Phi]^T\,(\mathbf{X} - \widehat{\underline{\mathbf{x}}})\,, \tag{19}$$

and its N independent realizations $\{\boldsymbol{\eta}^j, j = 1, \ldots, N\}$ are such that

$$\boldsymbol{\eta}^j = [\mu]^{-1/2}\,[\Phi]^T\,(\mathbf{x}^j - \widehat{\underline{\mathbf{x}}}) \in \mathbb{R}^\nu \,. \tag{20}$$

The empirical estimates of the mean vector and the covariance matrix of $\mathbf{H}$ verify the following conditions,

$$\widehat{\underline{\boldsymbol{\eta}}} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\eta}^j = \mathbf{0} \quad , \quad [\widehat{C}_{\mathbf{H}}] = \frac{1}{N-1}\sum_{j=1}^{N} (\boldsymbol{\eta}^j - \widehat{\underline{\boldsymbol{\eta}}})\,(\boldsymbol{\eta}^j - \widehat{\underline{\boldsymbol{\eta}}})^T = [I_\nu]\,. \tag{21}$$

From a numerical point of view, if $N < n$, the estimate $[\widehat{C}_{\mathbf{X}}]$ of covariance matrix of $\mathbf{X}$ is not computed and $\nu, [\mu]$, and $[\Phi]$ are directly computed using a thin SVD of the matrix whose N columns are $(\mathbf{x}^j - \widehat{\underline{\mathbf{x}}})$ for $j = 1, \ldots N$.

Hypothesis. Let $L^2(\Theta, \mathbb{R}^n)$ be the space of second-order random variables $\mathbf{X}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^n$ such that $E\{\|\mathbf{X}\|^2\} < +\infty$. Similarly, let $L^2(\Theta, \mathbb{R}^\nu)$ be the space of all the second-order random variables $\mathbf{H}$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^\nu$ such that $E\{\|\mathbf{H}\|^2\} < +\infty$. It should be noted that we do not consider the subset $\mathcal{H}_\nu$ of $L^2(\Theta, \mathbb{R}^\nu)$ made up of all the second-order $\mathbb{R}^\nu$-valued random variables $\mathbf{H}$ such that $E\{\mathbf{H}\} = \mathbf{0}$ and $[C_{\mathbf{H}}] = [I_\nu]$, since the constraints

will incur a shift and scaling that is *a priori* unknown. For fixed $\nu \leq n$ that verifies Eq. (17), random vector $\mathbf{X}^\nu$ defined by Eq. (18) spans a subspace $\mathcal{X}_\nu$ of $L^2(\Theta, \mathbb{R}^n)$ when $\mathbf{H}$ runs through $L^2(\Theta, \mathbb{R}^\nu)$. It is then assumed that random vector $\mathbf{X}^c$, for which pdf $p_{\mathbf{X}^c}$ is closest to pdf $p_{\mathbf{X}}$ under the constraint $E\{\mathbf{g}^c(\mathbf{X}^c)\} = \boldsymbol{\beta}^c$ (see Section 3), belongs to $\mathcal{X}_\nu$. It should be noted that there always exists ν such that $\nu \leq n$ with $\mathrm{err}_{\mathrm{PCA}}(\nu) \leq \varepsilon$ in order that $\mathbf{X}^c \in \mathcal{X}_\nu$, because for $\nu = n$ (that is to say, for ε sufficiently small), we have $\mathcal{X}_n = L^2(\Theta, \mathbb{R}^n)$.

5.2. Gaussian kernel-density estimation of the pdf of random vector $\boldsymbol{H}$

The pdf $\boldsymbol{\eta} \mapsto p_{\mathbf{H}}(\boldsymbol{\eta})$ on $\mathbb{R}^\nu$ of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}$ is constructed by the Gaussian kernel-density estimation method using the N independent realizations $\{\boldsymbol{\eta}^j, j = 1, \ldots, N\}$ defined by Eq. (20). We use the modification proposed in [44] of the classical formulation [45], which can be written, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, as

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = c_\nu \, \rho(\boldsymbol{\eta}) \quad , \quad c_\nu = \frac{1}{(\sqrt{2\pi}\, \widehat{s}_\nu)^\nu} \, , \tag{22}$$

$$\rho(\boldsymbol{\eta}) = \frac{1}{N} \sum_{j=1}^{N} exp\{-\frac{1}{2\widehat{s}_\nu^2} \, \| \frac{\widehat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \boldsymbol{\eta} \|^2 \} \, , \tag{23}$$

in which s_ν is the usual multidimensional optimal Silverman bandwidth (taking into account that the covariance matrix of $\mathbf{H}$ is $[I_\nu]$), and where $\widehat{s}_\nu$ has been introduced in order that the second equation in Eq. (21) be satisfied,

$$s_\nu = \left\{ \frac{4}{N(2+\nu)} \right\}^{1/(\nu+4)} \quad , \quad \widehat{s}_\nu = \frac{s_\nu}{\sqrt{s_\nu^2 + (N-1)/N}} \, . \tag{24}$$

With this formulation, Eqs. (21) to (24) yield

$$E\{\mathbf{H}\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta} \, p_{\mathbf{H}}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \frac{\widehat{s}_\nu}{s_\nu} \, \underline{\widehat{\boldsymbol{\eta}}} = \mathbf{0} \, , \tag{25}$$

$$E\{\mathbf{H}\mathbf{H}^T\} = \int_{\mathbb{R}^\nu} \boldsymbol{\eta}\boldsymbol{\eta}^T \, p_{\mathbf{H}}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \widehat{s}_\nu^2 \, [I_\nu] + \frac{\widehat{s}_\nu^2}{s_\nu^2} \, \frac{(N-1)}{N} \, [\widehat{C}_{\mathbf{H}}] = [I_\nu] \, . \tag{26}$$

The pdf of $\mathbf{H}$, defined by Eqs. (22) to (24), is rewritten, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, as

$$p_{\mathbf{H}}(\boldsymbol{\eta}) = c_\nu \, e^{-\psi(\boldsymbol{\eta})} \quad , \quad \psi(\boldsymbol{\eta}) = -\log \rho(\boldsymbol{\eta}) \, . \tag{27}$$

5.3. Reformulating the optimization problem as a function of random vector $\boldsymbol{H}$

The objective is to reformulate the optimization problem defined by Eqs. (8) and (9) to evaluate the pdf $p_{\mathbf{H}}$ defined by Eq. (20) and (22) to (24). Taking into account Eq. (18) and the hypothesis introduced in Section 5.1, the $\mathbb{R}^\nu$-valued random variable $\mathbf{X}^{c,\nu}$ can be written as

$$\mathbf{X}^{c,\nu} = \underline{\widehat{\mathbf{x}}} + [\Phi] \, [\mu]^{1/2} \, \mathbf{H}^c \, , \tag{28}$$

in which the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}^c$ is defined by

$$\mathbf{H}^c = [\mu]^{-1/2} \, [\Phi]^T \, (\mathbf{X}^c - \underline{\widehat{\mathbf{x}}}) \, . \tag{29}$$

Note that, in general, $\mathbf{H}^c$ is not centered and its covariance matrix is not $[I_\nu]$. In addition, for ε sufficiently small verifying Eq. (17), the mean-square error $E\{\|\mathbf{X}^c - \mathbf{X}^{c,\nu}\|^2\}$ is sufficiently small (note that such condition always exists since the limiting case corresponds to $\nu = n$). Due to the introduction of the reduced-order representation $\mathbf{X}^{c,\nu}$ of $\mathbf{X}^c$ defined by Eq. (28), the stochastic dimension of $\mathbf{X}^{c,\nu}$ is less than or equal to ν, because $\mathbf{H}^c$ is a $\mathbb{R}^\nu$-valued random variable.

Let $\boldsymbol{\eta} \mapsto \widetilde{\mathbf{h}}^c(\boldsymbol{\eta})$ be the measurable mapping from $\mathbb{R}^\nu$ into $\mathbb{R}^{m_c}$ such that, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$,

$$\widetilde{\mathbf{h}}^c(\boldsymbol{\eta}) = \mathbf{g}^c(\underline{\widehat{\mathbf{x}}} + [\Phi]\,[\mu]^{1/2}\,\boldsymbol{\eta})\,. \tag{30}$$

If it is assumed that the value of ν is such that $\boldsymbol{\beta}^c$ (defined by Eq. (10)) belongs to the range space $\widetilde{\mathbf{h}}^c(\mathbb{R}^\nu)$ of mapping $\widetilde{\mathbf{h}}^c$, then the constraint defined by Eq. (10) is transported on $\mathbf{H}^c$ and writes,

$$E\{\widetilde{\mathbf{h}}^c(\mathbf{H}^c)\} = \boldsymbol{\beta}^c \quad , \quad \int_{\mathbb{R}^\nu} \widetilde{\mathbf{h}}^c(\boldsymbol{\eta})\, p_{\mathbf{H}^c}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = \boldsymbol{\beta}^c\,, \tag{31}$$

in which the pdf $p_{\mathbf{H}^c}$ of $\mathbf{H}^c$ is closest to pdf $p_{\mathbf{H}}$ while satisfying the constraint defined by Eq. (31).

We propose to perform the projection of the $\mathbb{R}^{m_c}$-valued random variable $\widetilde{\mathbf{h}}^c(\mathbf{H})$ in a $\mathbb{R}^{\nu_c}$-valued random variable $\mathbf{h}^c(\mathbf{H}) = [\Psi]^T\,\widetilde{\mathbf{h}}^c(\mathbf{H})$ with $\nu_c \leq \nu$ in which the matrix $[\Psi] \in \mathbb{M}_{m_c,\nu_c}$ verifies $[\Psi]^T\,[\Psi] = [I_{\nu_c}]$. This projection is carried out in order that the projected constraints be algebraically independent (see Section 5.4-(iv) and that the components of the random vector $\mathbf{h}^c(\mathbf{H})$ be de-correlated and not degenerated. Note that, in addition, with such a projection, the computation is less costly and more robust for large value of m_c. For the second criterion, matrix $[\Psi]$ and ν_c are constructed as follows. Let $[\mathrm{COV}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]$ be the covariance matrix in $\mathbb{M}^{+0}_{m_c}$ of the $\mathbb{R}^{m_c}$-valued random variable $\widetilde{\mathbf{h}}^c(\mathbf{H})$ in which $\mathbf{H}$ is the $\mathbb{R}^\nu$-valued random variable defined by Eq. (19) for which the N independent realizations are defined by Eq. (20). Let

$$[\mathrm{COV}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]\,\boldsymbol{\psi}^\alpha = \lambda^c_\alpha\,\boldsymbol{\psi}^\alpha$$

be the eigenvalue problem such that

$$\lambda^c_1 \geq \ldots \geq \lambda^c_{\nu_c} > 0 = \lambda^c_{\nu_c+1} = \ldots = \lambda^c_{m_c}\,.$$

Consequently, the rank of $[\mathrm{COV}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]$ is

$$\mathrm{rank}\{[\mathrm{COV}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]\} = \nu_c \leq m_c\,. \tag{32}$$

Let $[\Psi] = [\boldsymbol{\psi}^1 \ldots \boldsymbol{\psi}^{\nu_c}] \in \mathbb{M}_{m_c,\nu_c}$ be the matrix whose columns are the orthonormal eigenvectors $\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^{\nu_c}$ associated with $\lambda^c_1 \geq \ldots \geq \lambda^c_{\nu_c} > 0$. Therefore, we have $[\Psi]^T\,[\Psi] = [I_{\nu_c}]$. Consequently, the covariance matrix $[\mathrm{COV}\{\mathbf{h}^c(\mathbf{H})\}] \in \mathbb{M}^+_{\nu_c}$ is diagonal and invertible (the components of random vector $\mathbf{h}^c(\mathbf{H})$ are de-correlated and not degenerated).

From a numerical point of view, if $N < m_c$, the estimate $[\widehat{\mathrm{COV}}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]$ of covariance matrix of $\widetilde{\mathbf{h}}^c(\mathbf{H})$ is not computed, and ν_c, $\lambda^c_1, \ldots, \lambda^c_{\nu_c}$, $[\Psi]$ are directly computed using a thin SVD of the matrix whose N columns are $(\widetilde{\mathbf{h}}^c(\boldsymbol{\eta}^j) - \underline{\widetilde{\mathbf{h}}^c})$ for $j = 1, \ldots N$, where $\underline{\widetilde{\mathbf{h}}^c}$ is the estimate of the mean value of random vector $\widetilde{\mathbf{h}}^c(\mathbf{H})$, which is performed using the realizations $\{\widetilde{\mathbf{h}}^c(\boldsymbol{\eta}^j),\, j = 1, \ldots, N\}$. The projection of Eq. (31) is performed using the basis $[\Psi]$ that spans a subspace of $\mathbb{R}^{m_c}$ of dimension ν_c and is written as,

$$E\{\mathbf{h}^c(\mathbf{H}^c)\} = \mathbf{b}^c\,, \tag{33}$$

in which $\boldsymbol{\eta} \mapsto \mathbf{h}^c(\boldsymbol{\eta})$ is the measurable mapping from $\mathbb{R}^\nu$ into $\mathbb{R}^{\nu_c}$, with

$$\nu_c \leq \nu \,, \tag{34}$$

such that, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$,

$$\mathbf{h}^c(\boldsymbol{\eta}) = [\Psi]^T \, \widetilde{\mathbf{h}}^c(\boldsymbol{\eta}) = [\Psi]^T \, \mathbf{g}^c(\widehat{\underline{\mathbf{x}}} + [\Phi]\,[\mu]^{1/2}\,\boldsymbol{\eta}) \in \mathbb{R}^{\nu_c} \,, \tag{35}$$

and where $\mathbf{b}^c$ is written as

$$\mathbf{b}^c = [\Psi]^T \, \boldsymbol{\beta}^c \in \mathbb{R}^{\nu_c} \,. \tag{36}$$

Using again the Kullback-Leibler minimum cross-entropy principle (as done in Section 3) for random variable $\mathbf{H}^c$, the optimization problem defined by Eqs. (8) and (9) is reformulated as

$$p_{\mathbf{H}^c} = \arg \min_{\widehat{p} \in C_{\mathrm{ad},\mathbf{H}^c}} \int_{\mathbb{R}^\nu} \widehat{p}(\boldsymbol{\eta}) \, \log \frac{\widehat{p}(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})} \, d\boldsymbol{\eta} \,, \tag{37}$$

in which the admissible set $C_{\mathrm{ad},\mathbf{H}^c}$ is defined by

$$C_{\mathrm{ad},\mathbf{H}^c} = \{\boldsymbol{\eta} \mapsto \widehat{p}(\boldsymbol{\eta}) : \mathbb{R}^\nu \to \mathbb{R}^+, \int_{\mathbb{R}^\nu} \widehat{p}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = 1, \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta}) \, \widehat{p}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \mathbf{b}^c. \tag{38}$$

Remark concerning the calculation of the rank ν_c. The rank ν_c defined by Eq. (32) is a function of the tolerance $\tau_c > 0$ that is used for numerically testing the condition $\{\lambda^c_{\nu_c} > 0\} \cap \{\lambda^c_{\nu_c+1} = 0\}$. This test is rewritten as follows,

$$\text{If } \{\lambda^c_{\nu_c} > \tau_c \, \lambda^c_1\} \cap \{\lambda^c_{\nu_c+1} \leq \tau_c \, \lambda^c_1\} \text{ then the rank is } \nu_c \,. \tag{39}$$

In the linear algebra libraries devoted to the computation of the rank of a matrix, τ_c is generally chosen as 10^{-12}. It should be noted that the smaller τ_c the larger ν_c. Therefore, the value of ν_c that is less than or equal to m_c will plays an important role on the existence of a unique solution of the optimization problem defined by Eq. (37) with Eq. (38). This aspect will be discussed in Section 5.11.

5.4. Existence of a solution and its representation using Lagrange multipliers $(\lambda_0, \boldsymbol{\lambda})$

The optimization problem defined by Eq. (37) with Eq. (38) is solved by introducing Lagrange multipliers as done for the Maximum Entropy Principle (see for instance [46, 11, 8, 47, 48, 49]).

(i) Lagrange multipliers associated with the constraints. A Lagrange multipliers is introduced for each constraint:

$\lambda_0 \in \mathbb{R}^+$ for the constraint: $\int_{\mathbb{R}^\nu} \widehat{p}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = 1$,

$\boldsymbol{\lambda} \in C_{\mathrm{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{\nu_c}$ for the constraint: $\int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta}) \, \widehat{p}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \mathbf{b}^c$,

in which the admissible set $C_{\mathrm{ad},\boldsymbol{\lambda}}$ is defined as the open subset of $\mathbb{R}^{\nu_c}$ such that,

$$C_{\mathrm{ad},\boldsymbol{\lambda}} = \{\boldsymbol{\lambda} \in \mathbb{R}^{\nu_c} \,, \int_{\mathbb{R}^\nu} e^{-\psi(\boldsymbol{\eta}) - <\boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta})>} \, d\boldsymbol{\eta} < +\infty\} \,, \tag{40}$$

where ψ is the potential of $\mathbf{H}$ defined by Eq. (27). The form of the admissible set $C_{\mathrm{ad},\boldsymbol{\lambda}}$ follows from the representation using Lagrange multipliers.

(ii) Expression of the Lagrangian. For $\lambda_0 \in \mathbb{R}^+$ and $\boldsymbol{\lambda} \in C_{\text{ad},\lambda}$, the Lagrangian associated with Eqs. (37) and (38) is written as

$$\mathcal{L}ag(\widehat{p}\,;\lambda_0,\boldsymbol{\lambda}) = \int_{\mathbb{R}^\nu} \widehat{p}(\boldsymbol{\eta})\,\log\frac{\widehat{p}(\boldsymbol{\eta})}{p_{\mathbf{H}}(\boldsymbol{\eta})}\,d\boldsymbol{\eta} - (\lambda_0-1)\Big(\int_{\mathbb{R}^\nu}\widehat{p}(\boldsymbol{\eta})\,d\boldsymbol{\eta}-1\Big) - <\boldsymbol{\lambda}\,,\int_{\mathbb{R}^\nu}\mathbf{h}^c(\boldsymbol{\eta})\,\widehat{p}(\boldsymbol{\eta})\,d\boldsymbol{\eta}-\mathbf{b}^c> \ . \tag{41}$$

(iii) Reformulation of the optimization problem and construction of the solution. If it is assumed that there exists a unique solution $p_{\mathbf{H}^c}$ to the optimization problem defined by Eq. (37) with Eq. (38), then, using simple arguments from the calculus of variations, it can be seen that there exists $(\lambda_0^{\text{sol}},\boldsymbol{\lambda}^{\text{sol}}) \in \mathbb{R}^+\times C_{\text{ad},\lambda}$, such that the functional $(\widehat{p},\lambda_0,\boldsymbol{\lambda}) \mapsto \mathcal{L}ag(\widehat{p}\,;\lambda_0,\boldsymbol{\lambda})$ is stationary at $p_{\mathbf{H}^c}$ for $\lambda_0=\lambda_0^{\text{sol}}$ and $\boldsymbol{\lambda}=\boldsymbol{\lambda}^{\text{sol}}$, and $p_{\mathbf{H}^c}$ can be written, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$, as

$$p_{\mathbf{H}^c}(\boldsymbol{\eta}) = c_\nu\,\exp\{-\psi(\boldsymbol{\eta})-\lambda_0^{\text{sol}}- <\boldsymbol{\lambda}^{\text{sol}},\mathbf{h}^c(\boldsymbol{\eta})>\}\,, \tag{42}$$

in which c_ν and $\psi(\boldsymbol{\eta})$ are defined by Eqs. (22) and (27). In addition, always under the hypothesis that there exists a unique solution to the optimization problem defined by Eq. (37) with Eq. (38), and under mild additional hypotheses, it is proven in Appendix B pp. 126-128 of [50] that the unique solution is given by Eq. (42) in which $(\lambda_0^{\text{sol}},\boldsymbol{\lambda}^{\text{sol}})$ is the unique solution in $\mathbb{R}^+\times C_{\text{ad},\lambda}$ of the following nonlinear algebraic equations,

$$\int_{\mathbb{R}^\nu} c_\nu\,\exp\{-\psi(\boldsymbol{\eta})-\lambda_0- <\boldsymbol{\lambda},\mathbf{h}^c(\boldsymbol{\eta})>\}\,d\boldsymbol{\eta} = 1\,, \tag{43}$$

$$\int_{\mathbb{R}^\nu} c_\nu\,\mathbf{h}^c(\boldsymbol{\eta})\,\exp\{-\psi(\boldsymbol{\eta})-\lambda_0- <\boldsymbol{\lambda},\mathbf{h}^c(\boldsymbol{\eta})>\}\,d\boldsymbol{\eta} = \mathbf{b}^c\,. \tag{44}$$

(iv) Comments about the existence of a unique solution. The hypothesis of the existence of a unique solution for the optimization problem defined by Eqs. (37) with (38) does not necessarily hold for arbitrary combinations of initial dataset D_N and constraint equations. As explained in Section 5.3, a first necessary condition to guarantee the existence of a unique solution of the optimization problem is that the constraints be algebraically independent, which can be checked using the following criterion: there exists a bounded subset B of $\mathbb{R}^{\nu_c}$ with $\int_B d\boldsymbol{\eta} > 0$ such that, for any nonzero vector $\mathbf{v}$ in $\mathbb{R}^{1+\nu_c}$,

$$\int_B <\mathbf{v},\widehat{\mathbf{h}}^c(\boldsymbol{\eta})>^2\,d\boldsymbol{\eta} \ > 0 \quad \text{with} \quad \widehat{\mathbf{h}}^c(\boldsymbol{\eta}) = (1\,,\mathbf{h}^c(\boldsymbol{\eta})) \in \mathbb{R}^{1+\nu_c}\,. \tag{45}$$

It should ne noted that, in general, this criterion requires a numerical evaluation. If the given constraints are not consistent with D_N, the admissible set $C_{\text{ad},\lambda}$ may be the empty set. This necessary condition is therefore not sufficient and there is no mathematical criterion that can easily be implemented, which gives a necessary and sufficient condition to guarantee the existence of a unique solution as a function of D_N, $\mathbf{h}^c$, and $\mathbf{b}^c$. This problem relative to existence and uniqueness will be revisited from a numerical point of view in Section 5.11 .

5.5. Reformulation introducing a random vector $\boldsymbol{H}_{\boldsymbol{\lambda}}$ and eliminating λ_0

In general, only a numerical approximation to the solution of Eqs. (43) and (44) can be achieved. However, a major difficulty is introduced by the normalization constant $c_\nu\,e^{-\lambda_0}$ that

induces considerable difficulties when ν is large. In addition, integrals in high dimension have to be computed for calculating the Lagrange multipliers. In this section, a procedure is introduced to address this difficulty. This is done by first eliminating λ_0, followed by introducing a convex objective function that lends itself to efficient iterative schemes, and finally, by using the PLoM as a MCMC generator of realizations allowing the integrals to be computed.

5.5.1. Introduction of the $\mathbb{R}^\nu$-valued random variable $\boldsymbol{H}_\lambda$ and eliminating λ_0

For $\boldsymbol{\lambda}$ given in $C_{\text{ad},\lambda}$, let $\mathbf{H}_\lambda$ be the $\mathbb{R}^\nu$-valued random variable whose pdf with respect to $d\boldsymbol{\eta}$ on $\mathbb{R}^\nu$ is written, for all $\boldsymbol{\eta} \in \mathbb{R}^\nu$, as

$$p_{\mathbf{H}_\lambda}(\boldsymbol{\eta}) = c_0(\boldsymbol{\lambda})\, \exp\{-\psi(\boldsymbol{\eta}) - < \boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta}) >\}\,, \tag{46}$$

in which the function $\boldsymbol{\lambda} \mapsto c_0(\boldsymbol{\lambda})$ from $C_{\text{ad},\lambda} \subset \mathbb{R}^{\nu_c}$ into $]0, +\infty[$ is defined by

$$c_0(\boldsymbol{\lambda}) = \Big(\int_{\mathbb{R}^\nu} \exp\{-\psi(\boldsymbol{\eta}) - < \boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta}) >\}\, d\boldsymbol{\eta} \Big)^{-1}\,. \tag{47}$$

It can be easily verified that $c_0(\boldsymbol{\lambda}) = c_\nu\, \exp\{-\lambda_0\}$ and consequently, for all $\boldsymbol{\eta}$ in $\mathbb{R}^\nu$,

$$p_{\mathbf{H}^c}(\boldsymbol{\eta}) = \{p_{\mathbf{H}_\lambda}(\boldsymbol{\eta})\}_{\lambda=\lambda^{\text{sol}}} \quad , \quad p_{\mathbf{H}}(\boldsymbol{\eta}) = \{p_{\mathbf{H}_\lambda}(\boldsymbol{\eta})\}_{\lambda=\mathbf{0}}\,. \tag{48}$$

Because $c_0(\boldsymbol{\lambda})$ is defined as the constant of normalization of pdf $p_{\mathbf{H}_\lambda}$, the constraint defined by Eq. (43), which is rewritten as $\int_{\mathbb{R}^\nu} p_{\mathbf{H}_\lambda}(\boldsymbol{\eta})\, d\boldsymbol{\eta} = 1$ is automatically satisfied for all $\boldsymbol{\lambda}$ in $C_{\text{ad},\lambda}$. Using Eqs. (46) and (47), it can be seen that the constraint defined by Eq. (44) can be rewritten, for all $\boldsymbol{\lambda}$ in $C_{\text{ad},\lambda}$, as

$$E\{\mathbf{h}^c(\mathbf{H}_\lambda)\} = \mathbf{b}^c\,, \tag{49}$$

because $E\{\mathbf{h}^c(\mathbf{H}_\lambda)\} = \int_{\mathbb{R}^\nu} \mathbf{h}^c(\boldsymbol{\eta})\, p_{\mathbf{H}_\lambda}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$. From Section 5.4-(iv), it can then be deduced that $\boldsymbol{\lambda}^{\text{sol}} \in C_{\text{ad},\lambda}$ is the unique solution in $\boldsymbol{\lambda}$ of Eq. (49).

5.5.2. Construction of an optimization problem with a convex objective function for calculating $\boldsymbol{\lambda}^{\text{sol}}$

Similarly to the discrete-case approach proposed in [51], let $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ be the objective function from the open subset $C_{\text{ad},\lambda}$ of $\mathbb{R}^{\nu_c}$ into $\mathbb{R}$ such that

$$\Gamma(\boldsymbol{\lambda}) = < \boldsymbol{\lambda}, \mathbf{b}^c > - \log c_0(\boldsymbol{\lambda})\,. \tag{50}$$

It can easily be deduced that the gradient of Γ with respect to $\boldsymbol{\lambda}$ is written as

$$\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_\lambda)\}\,, \tag{51}$$

and that the Hessian matrix, $[\Gamma''(\boldsymbol{\lambda})]$, of Γ at $\boldsymbol{\lambda}$ is written as

$$[\Gamma''(\boldsymbol{\lambda})] = [\text{COV}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}]\,, \tag{52}$$

in which the covariance matrix $[\text{COV}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}]$ of random vector $\mathbf{h}^c(\mathbf{H}_\lambda)$ is such that

$$[\text{COV}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}] = E\{\mathbf{h}^c(\mathbf{H}_\lambda)\, \mathbf{h}^c(\mathbf{H}_\lambda)^T\} - E\{\mathbf{h}^c(\mathbf{H}_\lambda)\}\, E\{\mathbf{h}^c(\mathbf{H}_\lambda)\}^T\,. \tag{53}$$

By construction (see Section 5.3), we have $[\text{COV}\{\mathbf{h}^c(\mathbf{H})\}] = [\Psi]^T\, [\text{COV}\{\widetilde{\mathbf{h}}^c(\mathbf{H})\}]\, [\Psi]$. We will assume that, for all $\boldsymbol{\lambda} \in C_{\text{ad},\lambda}$, the covariance matrix $[\text{COV}\{\mathbf{h}^c(\mathbf{H}_\lambda)\}]$ is positive definite (this is

a reasonable hypothesis due to the construction of $[\Psi]$ and due to the hypothesis of algebraical independence of the constraints (see Eq. (45)). Consequently, $\boldsymbol{\lambda} \mapsto \Gamma(\boldsymbol{\lambda})$ is a strictly convex function in $C_{\text{ad},\boldsymbol{\lambda}}$. Since $\boldsymbol{\lambda}^{\text{sol}} \in C_{\text{ad},\boldsymbol{\lambda}}$ is the unique solution of Eq. (49), we have $E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\text{sol}}})\} = \mathbf{b}^c$. Therefore, Eq. (51) yields $\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\text{sol}}) = \mathbf{0}$. Consequently, $\boldsymbol{\lambda}^{\text{sol}}$ is the unique solution of the following optimization problem,

$$\boldsymbol{\lambda}^{\text{sol}} = \arg \min_{\boldsymbol{\lambda} \in C_{\text{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{n_{u_c}}} \Gamma(\boldsymbol{\lambda}) \,. \tag{54}$$

Since Γ is convex, this optimization problem would be convex if $C_{\text{ad},\boldsymbol{\lambda}}$ was convex and then $\boldsymbol{\lambda}^{\text{sol}}$ would be the unique global minimum that could be computed with the efficient algorithms devoted to convex optimization problems. However, it cannot be proven that $C_{\text{ad},\boldsymbol{\lambda}}$ is convex for arbitrary constraints $\mathbf{h}^c$.

5.6. Iterative algorithm for computing the Lagrange multipliers $\boldsymbol{\lambda}^{\text{sol}}$

Taking into account the convexity of Γ, we propose to solve the minimization problem defined by Eq. (54) with a Newton iterative method applied to function $\boldsymbol{\lambda} \mapsto \boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda})$ [52, 49] starting from an initial point $\boldsymbol{\lambda}_1$ in $C_{\text{ad},\boldsymbol{\lambda}}$. This algorithm allows for finding $\boldsymbol{\lambda}^{\text{sol}}$ such that $\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\text{sol}}) = \mathbf{0}$ and is written, for $\iota \geq 1$, as

$$\boldsymbol{\lambda}^{\iota+1} = \boldsymbol{\lambda}^{\iota} - [\Gamma''(\boldsymbol{\lambda}^{\iota})]^{-1} \, \boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\iota}) \,, \tag{55}$$

in which $\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\iota})$ and $[\Gamma''(\boldsymbol{\lambda}^{\iota})]$ are given by Eqs. (51) and (52), that is to say,

$$\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\iota}) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}})\} \quad , \quad [\Gamma''(\boldsymbol{\lambda}^{\iota})] = [\text{cov}\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}})\}] \,. \tag{56}$$

The initial point, which corresponds to $\iota = 1$, is chosen such that

$$\boldsymbol{\lambda}^{1} = -[\Gamma''(\boldsymbol{\lambda}^{0})]^{-1} \, \boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{0}) \,, \tag{57}$$

with $\boldsymbol{\lambda}^0 = \mathbf{0}$, and since $\mathbf{H}_{\mathbf{0}} = \mathbf{H}$ (see Eq. (48)), we have

$$\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{0}) = \mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H})\} \quad , \quad [\Gamma''(\boldsymbol{\lambda}^{0})] = [\text{cov}\{\mathbf{h}^c(\mathbf{H})\}] \,. \tag{58}$$

Consequently, the necessary condition concerning the initial point in the Newton method is satisfied: $\boldsymbol{\lambda}^1$ is not close to origin $\mathbf{0}$ because $\mathbf{b}^c \neq E\{\mathbf{h}^c(\mathbf{H})\}$ and we have $\boldsymbol{\lambda}^1 \in C_{\text{ad},\boldsymbol{\lambda}}$. The right-hand sides of the two equations in Eq. (58) can be estimated using the empirical estimators of the mean vector and covariance matrix of $\mathbf{h}^c(\mathbf{H})$ using the the initial dataset $\{\boldsymbol{\eta}^j, j = 1, \ldots, N\}$ defined by Eq. (20). At each iteration ι, the convergence of the iteration algorithm is controlled by calculating the error

$$\text{err}(\iota) = \frac{\|\mathbf{b}^c - E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}})\}\|}{\|\mathbf{b}^c\|} = \frac{\|\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^{\iota}\|}{\|\mathbf{b}^c\|} \,. \tag{59}$$

During the iteration, for each $\boldsymbol{\lambda}^{\iota}$ given in $C_{\text{ad},\boldsymbol{\lambda}} \subset \mathbb{R}^{\nu_c}$, the following mathematical expectations,

$$E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}})\} \quad , \quad E\{\mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}}) \, \mathbf{h}^c(\mathbf{H}_{\boldsymbol{\lambda}^{\iota}})^T\} \,, \tag{60}$$

have to be computed using an MCMC algorithm (see Section 5.7). It should be noted that admissible set $C_{\text{ad},\boldsymbol{\lambda}}$ cannot, in the general case, explicitly be described. An additional discussion concerning the convergence of the iteration algorithm in the framework of the existence of a solution is given in Section 5.11.

5.7. Nonlinear Itô stochastic differential equation (ISDE) for the generator of random variable $\boldsymbol{H}_{\boldsymbol{\lambda}}$

For $\boldsymbol{\lambda}$ fixed in $C_{\mathrm{ad},\lambda}$, we have to construct a generator for $\mathbf{H}_{\boldsymbol{\lambda}}$ in order to compute an estimation of the quantities listed in Eq. (60). We propose to use the approach presented in [53, 49], which is based on a nonlinear ISDE associated with a nonlinear stochastic dissipative Hamiltonian dynamical system. This choice of MCMC generator will allow for implementing, in Section 5.8, the probabilistic learning on manifolds proposed in [37] for generating a large number of realizations of random variable $\mathbf{H}_{\lambda'}$ and then of $\mathbf{H}^c$ (at convergence with respect to ι).

5.7.1. Introduction of the potential function associated with pdf $p_{\boldsymbol{H}_{\lambda}}$

For $\boldsymbol{\lambda}$ fixed in $C_{\mathrm{ad},\lambda}$, Eq. (46) is rewritten, for all $\boldsymbol{\eta}$ in $\mathbb{R}^{\nu}$, as

$$p_{\mathbf{H}_{\lambda}}(\boldsymbol{\eta}) = c_0(\boldsymbol{\lambda})\, \rho_{\lambda}(\boldsymbol{\eta})\,, \tag{61}$$

in which function $\boldsymbol{\eta} \mapsto \rho_{\lambda}(\boldsymbol{\eta})$ from $\mathbb{R}^{\nu}$ into $\mathbb{R}^{+}$ is defined by

$$\rho_{\lambda}(\boldsymbol{\eta}) = \exp\{-\mathcal{V}_{\lambda}(\boldsymbol{\eta})\} \quad , \quad \mathcal{V}_{\lambda}(\boldsymbol{\eta}) = \psi(\boldsymbol{\eta}) + < \boldsymbol{\lambda}, \mathbf{h}^c(\boldsymbol{\eta}) > \ . \tag{62}$$

5.7.2. Nonlinear ISDE for random vector $\boldsymbol{H}_{\boldsymbol{\lambda}}$

Following [37], we have to consider the nonlinear stochastic dissipative Hamiltonian dynamical system whose potential is $\mathcal{V}_{\lambda}$ defined by Eq. (62), which is represented by the nonlinear ISDE expressed, for $t > 0$, as

$$d[\mathbf{U}_{\lambda}(t)] = [\mathbf{V}_{\lambda}(t)]\, dt\,, \tag{63}$$

$$d[\mathbf{V}_{\lambda}(t)] = [L_{\lambda}([\mathbf{U}_{\lambda}(t)])]\, dt - \frac{1}{2}\, f_0 [\mathbf{V}_{\lambda}(t)]\, dt + \sqrt{f_0}\, d[\mathbf{W}^{\mathrm{wien}}_{\lambda}(t)]\,, \tag{64}$$

with the initial condition at $t = 0$,

$$[\mathbf{U}_{\lambda}(0)] = [\eta^{\mathrm{init}}] \quad , \quad [\mathbf{V}_{\lambda}(0)] = [v^{\mathrm{init}}] \quad a.s.\,, \tag{65}$$

in which $\{([\mathbf{U}_{\lambda}(t)], [\mathbf{V}_{\lambda}(t)]), t \in \mathbb{R}^{+}\}$ is a stochastic process with values in $\mathbb{M}_{\nu,N} \times \mathbb{M}_{\nu,N}$, which is made up of the independent stochastic processes $\{(\mathbf{U}^{\ell}_{\lambda}(t), \mathbf{V}^{\ell}_{\lambda}(t)),\, t \in \mathbb{R}^{+}\}$ such that $[\mathbf{U}_{\lambda}(t)] = [\mathbf{U}^1_{\lambda}(t) \ldots \mathbf{U}^N_{\lambda}(t)]$ and $[\mathbf{V}_{\lambda}(t)] = [\mathbf{V}^1_{\lambda}(t) \ldots \mathbf{V}^N_{\lambda}(t)]$, and where:

(i) $f_0 > 0$ is a free parameter allowing the dissipation to be controlled in the stochastic dynamical system. This parameter is chosen such that $f_0 < 4$. The value, 4, corresponds to the critical damping rate of the linearized ISDE associated with Eqs. (63) and (64).

(ii) $\{[\mathbf{W}^{\mathrm{wien}}_{\lambda}(t)], t \in \mathbb{R}^{+}\}$ is the stochastic process, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by $\mathbb{R}^{+}$, with values in $\mathbb{M}_{\nu,N}$, for which the columns of $[\mathbf{W}^{\mathrm{wien}}_{\lambda}(t)]$ are N independent copies of the $\mathbb{R}^{\nu}$-valued normalized Wiener process whose matrix-valued autocorrelation function is $\min(t, t')\, [I_{\nu}]$. The dependence on $\boldsymbol{\lambda}$ is such that, if $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are two distinct values in $C_{\mathrm{ad},\lambda}$, then the two stochastic processes $\{[\mathbf{W}^{\mathrm{wien}}_{\lambda}(t)], t \in \mathbb{R}^{+}\}$ and $\{[\mathbf{W}^{\mathrm{wien}}_{\lambda'}(t)], t \in \mathbb{R}^{+}\}$ are independent.

(iii) $[u] \mapsto [L_{\lambda}([u])]$ is a nonlinear mapping from $\mathbb{M}_{\nu,N}$ into $\mathbb{M}_{\nu,N}$, expressed in the following form as negative of the gradient of the potential $\mathcal{V}_{\lambda}$

$$[L_{\lambda}([u])]_{\alpha\ell} = -\frac{\partial}{\partial u^{\ell}_{\alpha}} \mathcal{V}_{\lambda}(\mathbf{u}^{\ell}) \quad , \quad \alpha = 1, \ldots, \nu \quad , \quad \ell = 1, \ldots, N\,, \tag{66}$$

in which $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N] \in \mathbb{M}_{\nu,N}$ with $\mathbf{u}^\ell = (u_1^\ell, \dots u_\nu^\ell) \in \mathbb{R}^\nu$. For ℓ fixed in $\{1, \dots, N\}$, the Hamiltonian of the associated conservative homogeneous dynamical system related to stochastic process $\{(\mathbf{U}^\ell(t), \mathbf{V}^\ell(t)), t \in \mathbb{R}^+\}$ is thus written as $\mathbb{H}_{\boldsymbol{\lambda}}(\mathbf{u}^\ell, \mathbf{v}^\ell) = \frac{1}{2}\|\mathbf{v}^\ell\|^2 + \mathcal{V}_{\boldsymbol{\lambda}}(\mathbf{u}^\ell)$.

(iv) $[\eta^{\text{init}}]$ and $[v^{\text{init}}]$ are given matrices in $\mathbb{M}_{\nu,N}$ that will be defined in Section 5.7.5 as a function of iteration ι of the algorithm given in Section 5.6.

5.7.3. Existence and uniqueness of a stationary solution associated with the invariant measure

Using Theorems 4 to 7 in Pages 211 to 216 of [54], it can be proven, that for $\boldsymbol{\lambda} \in C_{\text{ad},\boldsymbol{\lambda}}$, Eqs. (63) to (65) admits the unique invariant measure,

$$\otimes_{\ell=1}^N \{p_{\mathbf{H}_{\boldsymbol{\lambda}}}(\mathbf{u}^\ell)\, p_{\mathbf{G}}(\mathbf{v}^\ell)\, d\mathbf{u}^\ell\, d\mathbf{v}^\ell\}\,, \tag{67}$$

in which $p_{\mathbf{H}_{\boldsymbol{\lambda}}}$ is defined by Eqs. (61) and (62), and where the pdf $p_{\mathbf{G}}$ is written as $p_{\mathbf{G}}(\mathbf{v}^\ell) = (2\pi)^{-\nu/2} \exp\{-1/2\|\mathbf{v}^\ell\|^2\}$. In addition, these Theorems allow for proving that Eqs. (63) to (65) has a unique solution $\{([\mathbf{U}_{\boldsymbol{\lambda}}(t)], [\mathbf{V}_{\boldsymbol{\lambda}}(t)]), t \in \mathbb{R}^+\}$, which is a second-order diffusion stochastic process that is asymptotic for $t \to +\infty$ to the stationary stochastic process $\{([\mathbf{U}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})], [\mathbf{V}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})]), t_{\text{st}} \in \mathbb{R}^+\}$ for the right-shift semi-group on $\mathbb{R}^+$. For all fixed t_{st}, the joint probability distribution of the random matrices $[\mathbf{U}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})]$ and $[\mathbf{V}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})]$ is the invariant measure defined by Eq. (67). The probability distribution of random matrix $[\mathbf{U}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})]$ is

$$\otimes_{\ell=1}^N p_{\mathbf{H}_{\boldsymbol{\lambda}}}(\mathbf{u}^\ell)\, d\mathbf{u}^\ell\,, \tag{68}$$

that is to say, is the probability distribution $p_{[\mathbf{H}_{\boldsymbol{\lambda}}]}([u])\, d[u]$ with $d[u] = \otimes_{\ell=1}^N d\mathbf{u}^\ell$ of the random matrix $[\mathbf{H}_{\boldsymbol{\lambda}}]$ with values in $\mathbb{M}_{\nu,N}$. The columns $\mathbf{H}_{\boldsymbol{\lambda}}^1, \dots, \mathbf{H}_{\boldsymbol{\lambda}}^N$ of $[\mathbf{H}_{\boldsymbol{\lambda}}]$ are N independent copies of random vector $\mathbf{H}_{\boldsymbol{\lambda}}$ with values in $\mathbb{R}^\nu$ whose pdf is $p_{\mathbf{H}_{\boldsymbol{\lambda}}}$ defined by Eqs. (61) and (62). It can then be deduced that for any fixed t_{st},

$$[\mathbf{H}_{\boldsymbol{\lambda}}] = [\mathbf{U}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})] = \lim_{t \to +\infty} [\mathbf{U}_{\boldsymbol{\lambda}}(t)]\,. \tag{69}$$

Equation (69) means that, for $f_0 > 0$, there exists $t_{\boldsymbol{\lambda}}^{\min} > 0$, such that $\{[\mathbf{U}_{\boldsymbol{\lambda}}(t)], t > t_{\boldsymbol{\lambda}}^{\min}\}$ is a stationary stochastic process that is a copy of stochastic process $\{[\mathbf{U}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})],\, t_{\text{st}} \in \mathbb{R}^+\}$. As previously explained, f_0 allows for controlling the transient response generated by the initial conditions for quickly reaching the stationary regime. Finally, we introduce the Gaussian random matrices $[\mathbf{V}_{\boldsymbol{\lambda}}]$ with values in $\mathbb{M}_{\nu,N}$ such that

$$[\mathbf{V}_{\boldsymbol{\lambda}}] = [\mathbf{V}_{\boldsymbol{\lambda}}^{\text{st}}(t_{\text{st}})] = \lim_{t \to +\infty} [\mathbf{V}_{\boldsymbol{\lambda}}(t)]\,. \tag{70}$$

5.7.4. Expression of the mapping $[L_{\boldsymbol{\lambda}}]$

In this subsection a general expression for the mapping $[L_{\boldsymbol{\lambda}}]$ is developed. In addition, specializations to particular common examples are derived.

Using Eqs. (62) and (66), it can be proven that, for $\alpha = 1, \dots, \nu$, for $\ell = 1, \dots, N$, and for $[u] = [\mathbf{u}^1 \dots \mathbf{u}^N]$ in $\mathbb{M}_{\nu,N}$, we have

$$[L_{\boldsymbol{\lambda}}([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)} \boldsymbol{\nabla}_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) - [\mu]^{1/2} [\Phi]^T [D_{\mathbf{x}} \mathbf{g}^c(\widehat{\underline{\mathbf{x}}} + [\Phi][\mu]^{1/2} \mathbf{u}^\ell)] [\Psi] \boldsymbol{\lambda} \right\}_\alpha\,, \tag{71}$$

in which ρ is defined by Eq. (23), where the gradient of ρ is such that

$$\boldsymbol{\nabla}_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) = \frac{1}{\widehat{s}_\nu^2} \frac{1}{N} \sum_{j=1}^N (\frac{\widehat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \mathbf{u}^\ell) \exp\{-\frac{1}{2\widehat{s}_\nu^2} \|\frac{\widehat{s}_\nu}{s_\nu} \boldsymbol{\eta}^j - \mathbf{u}^\ell\|^2\}\,, \tag{72}$$

and where $[D_{\mathbf{x}}\mathbf{g}^c(\mathbf{x})] \in \mathbb{M}_{n,m_c}$ is the derivative at point $\mathbf{x} \in \mathbb{R}^n$ of the mapping $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_{m_c}(\mathbf{x}))$ from $\mathbb{R}^n$ into $\mathbb{R}^{m_c}$ such that

$$[D_{\mathbf{x}}\mathbf{g}^c(\mathbf{x})]_{ki} = \frac{\partial g_i^c(\mathbf{x})}{\partial x_k} \quad , \quad k = 1, \ldots, n \quad , \quad i = 1, \ldots m_c \, .$$

(1) Case of the example presented in Section 4.1. Let us consider the example presented in Section 4.1. We have $\mathcal{J}_{\rm mom} = \{k_1, \ldots, k_{\kappa_{\rm mom}}\}$, $\kappa_{\rm mom} \leq n_q$, and $m_c = 2\kappa_{\rm mom} \leq 2n_q$. Therefore,

$$\boldsymbol{\beta}^c = (\mathbb{m}_{k_1}^{\rm exp}, \ldots, \mathbb{m}_{k_{\kappa_{\rm mom}}}^{\rm exp}, \mathbb{r}_{k_1}^{\rm exp}, \ldots, \mathbb{r}_{k_{\kappa_{\rm mom}}}^{\rm exp}) \in \mathbb{R}^{m_c} \, , \tag{73}$$

and, for $k = 1, \ldots, n$ and $\mathbf{v} = [\Psi]\,\boldsymbol{\lambda}$ in $\mathbb{R}^{m_c}$, we have

$$\{[D_{\mathbf{x}}\mathbf{g}^c(\mathbf{x})]\,\mathbf{v}\}_k = \sum_{\kappa=1}^{\kappa_{\rm mom}} \delta_{kk_\kappa}(v_\kappa + 2\, x_{k_\kappa} v_{\kappa+\kappa_{\rm mom}}) \, .$$

Let us consider the block expression of $[\Phi] \in \mathbb{M}_{n,\nu}$ such that $[\Phi]^T = [[\Phi_q]^T[\Phi_w]^T]$ in which $[\Phi_q] \in \mathbb{M}_{n_q,\nu}$. For $\kappa = 1, \ldots, \kappa_{\rm mom}$, we define $q_{k_\kappa}(\mathbf{u}^\ell)$ such that

$$q_{k_\kappa}(\mathbf{u}^\ell) = \widehat{\underline{q}}_{k_\kappa} + \{[\Phi_q][\mu]^{1/2}\mathbf{u}^\ell\}_{k_\kappa} \, . \tag{74}$$

Consequently, for $\alpha = 1, \ldots, \nu$, for $\ell = 1, \ldots, N$, and for $\mathbf{v} = [\Psi]\,\boldsymbol{\lambda}$, Eq. (71) yields

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)} \boldsymbol{\nabla}_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^{c,1}([u])]_{\alpha\ell} \, . \tag{75}$$

$$[L_\lambda^{c,1}([u])]_{\alpha\ell} = -\sum_{\kappa=1}^{\kappa_{\rm mom}} \mu_\alpha^{1/2} [\Phi_q]_{k_\kappa\alpha}(v_\kappa + 2\, q_{k_\kappa}(\mathbf{u}^\ell) v_{\kappa+\kappa_{\rm mom}}) \, . \tag{76}$$

For facilitating the numerical implementation and enhancing computational efficiency, the right-hand side of Eq. (76) can be rewritten in a matrix form as follows. Let $\underline{\widetilde{\mathbf{q}}} = (\underline{\widetilde{q}}_1, \ldots, \underline{\widetilde{q}}_{\kappa_{\rm mom}})$ be the vector in $\mathbb{R}^{\kappa_{\rm mom}}$ and let $[\widetilde{\Phi}_q]$ be the matrix in $\mathbb{M}_{\kappa_{\rm mom},\nu}$ such that, for $\kappa = 1, \ldots, \kappa_{\rm mom}$ and for all $\alpha = 1, \ldots, \nu$, we have

$$\underline{\widetilde{q}}_\kappa = \widehat{\underline{q}}_{k_\kappa} \quad , \quad [\widetilde{\Phi}_q]_{\kappa\alpha} = [\Phi_q]_{k_\kappa\,\alpha} \, . \tag{77}$$

Let $\widetilde{\mathbf{q}}(\mathbf{u}^\ell) = (\widetilde{q}_1(\mathbf{u}^\ell), \ldots, \widetilde{q}_{\kappa_{\rm mom}}(\mathbf{u}^\ell))$ be the vector in $\mathbb{R}^{\kappa_{\rm mom}}$ such that $\widetilde{q}_\kappa(\mathbf{u}^\ell) = q_{k_\kappa}(\mathbf{u}^\ell)$, which can be written as

$$\widetilde{\mathbf{q}}(\mathbf{u}^\ell) = \underline{\widetilde{\mathbf{q}}} + [\widetilde{\Phi}_q]\,[\mu]^{1/2}\mathbf{u}^\ell \, . \tag{78}$$

Let $[\widetilde{q}(\mathbf{u}^\ell)]$ be the diagonal matrix in $\mathbb{M}_{\kappa_{\rm mom}}$ such that $[\widetilde{q}(\mathbf{u}^\ell)]_{\kappa\kappa'} = \delta_{\kappa\kappa'}\,\widetilde{q}_\kappa(\mathbf{u}^\ell)$ in which $\widetilde{q}_\kappa(\mathbf{u}^\ell)$ is component κ of vector $\widetilde{\mathbf{q}}(\mathbf{u}^\ell)$ defined by Eq. (78). Let us consider the block writing of $[\Psi] \in \mathbb{M}_{m_c,\nu_c}$ such that $[\Psi]^T = [[\Psi_{\mathbb{m}}]^T[\Psi_{\mathbb{r}}]^T]$ in which $[\Psi_{\mathbb{m}}]$ and $[\Psi_{\mathbb{r}}]$ are matrices that belong to $\mathbb{M}_{\kappa_{\rm mom},\nu_c}$. Using the above notations, Eq. (76) can be rewritten as

$$[L_\lambda^{c,1}([u])]_{\alpha\ell} = \left\{ -[\mu]^{1/2}[\widetilde{\Phi}_q]^T \left([\Psi_{\mathbb{m}}] + 2\,[\widetilde{q}(\mathbf{u}^\ell)]\,[\Psi_{\mathbb{r}}] \right) \boldsymbol{\lambda} \right\}_\alpha \, . \tag{79}$$

(2) Case of the example presented in Section 4.2. Let us now consider Eq. (15) of the example presented in Section 4.2 with $m_c = n_q$. We then have $\mathbf{g}^c(\mathbf{x}) = \mathbf{q} - [\underline{B}]\,\mathbf{w}$ with $\mathbf{x} = (\mathbf{q}, \mathbf{w}) \in \mathbb{R}^n = \mathbb{R}^{n_q} \times \mathbb{R}^{n_w}$, which can be rewritten as $\mathbf{g}^c(\mathbf{x}) = [G^c]\,\mathbf{x}$ with $[G^c] = [\,[I_{n_q}] \,|\, -[\underline{B}]\,] \in \mathbb{M}_{n_q,n}$ with $n = n_q + n_w$. Consequently, we have $[D_{\mathbf{x}}\mathbf{g}^c(\mathbf{x})] = [G^c]^T \in \mathbb{M}_{n,n_q}$ and for $\alpha = 1, \ldots, \nu$, for $\ell = 1, \ldots, N$, and for $\mathbf{v} = [\Psi]\,\boldsymbol{\lambda}$, Eq. (71) yields

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)} \boldsymbol{\nabla}_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^{c,2}([u])]_{\alpha\ell}\,. \tag{80}$$

$$[L_\lambda^{c,2}([u])]_{\alpha\ell} = \left\{ -[\mu]^{1/2} [\Phi]^T\, [G^c]^T\, [\Psi]\,\boldsymbol{\lambda} \right\}_\alpha\,. \tag{81}$$

It should be noted that $[L_\lambda^{c,2}([u])]$ depends only on $\boldsymbol{\lambda}$ and is independent of $[u]$.

(3) Case of the example presented in Appendix A.1. Finally, we consider the example presented in Appendix A.1 for which $\mathcal{J}_{\text{mom}} = \{k_1, \ldots, k_{\kappa_{\text{mom}}}\}$, $\kappa_{\text{mom}} \leq n_q$, and $m_c = 2\,\kappa_{\text{mom}} + n_q \leq 3\,n_q$. Therefore,

$$\boldsymbol{\beta}^c = (\mathbb{m}_{k_1}^{\text{exp}}, \ldots, \mathbb{m}_{k_{\kappa_{\text{mom}}}}^{\text{exp}}, \mathbb{r}_{k_1}^{\text{exp}}, \ldots, \mathbb{r}_{k_{\kappa_{\text{mom}}}}^{\text{exp}}, \beta_1^{\text{const}}, \ldots, \beta_{n_q}^{\text{const}}) \in \mathbb{R}^{m_c}\,, \tag{82}$$

and using Eqs. (75), (76), (80), and (81), we have, for $\alpha = 1, \ldots, \nu$, for $\ell = 1, \ldots, N$, and for $\boldsymbol{\lambda} \in \mathbb{R}^{\nu_c}$,

$$[L_\lambda([u])]_{\alpha\ell} = \left\{ \frac{1}{\rho(\mathbf{u}^\ell)} \boldsymbol{\nabla}_{\mathbf{u}^\ell} \rho(\mathbf{u}^\ell) \right\}_\alpha + [L_\lambda^{c}([u])]_{\alpha\ell}\,, \tag{83}$$

$$[L_\lambda^{c}([u])] = [L_\lambda^{c,1}([u])] + [L_\lambda^{c,2}([u])]\,. \tag{84}$$

5.7.5. Expression of the initial conditions for the MCMC algorithm

In Eq. (65), the initial conditions $[\eta^{\text{init}}]$ and $[v^{\text{init}}]$ are constructed in $\mathbb{M}_{\nu,N}$ as a function of iteration ι of the algorithm presented in Section 5.6 in order to calculate the sequence $\{\boldsymbol{\lambda}^\iota\}_{\iota \geq 2}$ from the initial value $\boldsymbol{\lambda}^1$ as follows:

- $\boldsymbol{\lambda}^1$ is computed (see Eq. (55)) by $\boldsymbol{\lambda}^1 = -[\Gamma''(\boldsymbol{\lambda}^0)]^{-1}\,\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^0)$ with $\boldsymbol{\lambda}^0 = \mathbf{0}$, using Eq. (58) as explained in Section 5.6.

- $\boldsymbol{\lambda}^2$ is calculated using Eq. (55), in which $[\Gamma''(\boldsymbol{\lambda}^1)]$ and $\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^1)$ are computed with Eqs. (51) and (52) in which $E\{\mathbf{h}^c(\mathbf{H}_{\lambda^1})\}$ and $[\text{COV}\{\mathbf{h}^c(\mathbf{H}_{\lambda^1})]$ are estimated with the MCMC algorithm for which the initial conditions at $t = 0$ are chosen as $[\eta^{\text{init}}] = [\eta^{\text{init},1}]$ and $[v^{\text{init}}] = [v^{\text{init},1}]$ with

  $$[\eta^{\text{init},1}] = [\boldsymbol{\eta}^1 \ldots \boldsymbol{\eta}^N] \in \mathbb{M}_{\nu,N}\,, \tag{85}$$

  and where the N columns of $[v^{\text{init},1}]$ are N independent realizations of a Gaussian, centered, $\mathbb{R}^\nu$-valued random variable for which its covariance matrix is $[I_\nu]$. The MCMC generator (see Section 5.8) allows for generating n_{MC} realizations $[\eta_{\lambda^1}^1], \ldots, [\eta_{\lambda^1}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{H}_{\lambda^1}]$ and $[v_{\lambda^1}^1], \ldots, [v_{\lambda^1}^{n_{\text{MC}}}]$ of random matrix $[\mathbf{V}_{\lambda^1}]$ (see Eqs. (69) and (70)).

- Knowing $\boldsymbol{\lambda}^\iota$ at iteration $\iota \geq 1$, the vector $\boldsymbol{\lambda}^{\iota+1}$ is calculated using Eq. (55) in which $[\Gamma''(\boldsymbol{\lambda}^\iota)]$ and $\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^\iota)$ are computed with Eqs. (51) and (52)) using the MCMC algorithm for which the initial conditions at $t = 0$ are chosen as $[\eta^{\text{init}}] = [\eta^{\text{init},\iota}]$ and $[v^{\text{init}}] = [v^{\text{init},\iota}]$ with

  $$[\eta^{\text{init},\iota}] = [\eta_{\lambda^{\iota-1}}^{n_{\text{MC}}}] \quad , \quad [v^{\text{init},\iota}] = [v_{\lambda^{\iota-1}}^{n_{\text{MC}}}]\,. \tag{86}$$

5.8. Computing the additional realizations of $\boldsymbol{H}_{\lambda^i}$ using the probabilistic learning on manifolds

For given $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\iota$ in $\mathbb{R}^{\nu_c}$ in which $\boldsymbol{\lambda}^\iota$ comes from the iteration algorithm detailed in Section 5.6, and for given initial conditions $[\eta^{\text{init},\iota}]$ and $[v^{\text{init},\iota}]$ defined by Eq. (86), n_{MC} realizations $[\eta^1_{\lambda^\iota}], \ldots, [\eta^{n_{\text{MC}}}_{\lambda^\iota}]$ of random matrix $[\mathbf{H}_{\lambda^\iota}]$ and $[v^1_{\lambda^\iota}], \ldots, [v^{n_{\text{MC}}}_{\lambda^\iota}]$ of random matrix $[\mathbf{V}_{\lambda^\iota}]$ have to be computed using the MCMC generator defined by Eqs. (63) to (65) in which $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\iota$. However, as it has previously been explained, N is assumed to be small and consequently, the probabilistic learning on manifolds presented in [37] is used. This algorithm consists in projecting the nonlinear ISDE defined by Eqs. (63) to (65) on the diffusion-maps basis. This approach allows for preserving the concentration of the invariant measure $p_{\mathbf{H}_{\lambda^\iota}}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$ and to avoid the scattering of the realizations computed with the MCMC generator. In order to simplify the reading of this paper, we briefly summarized the algorithm that is proposed in [37] in adapting the notations to the problem under consideration.

5.8.1. Construction of the diffusion-maps basis

The diffusion-maps basis is independent of $\boldsymbol{\lambda}^\iota$, only depends on the initial dataset $\{\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^N\}$, and is represented by the matrix

$$[g] = [\mathbf{g}^1 \ldots \mathbf{g}^m] \in \mathbb{M}_{N,m} \quad \text{with} \quad 1 < m \leq N\,. \tag{87}$$

For $m = N$, it is an algebraic basis of vector space $\mathbb{R}^N$. This basis is constructed using the diffusion maps proposed in [55]. Let $[\mathbb{K}]$ be the symmetric $(N \times N)$ real matrix such that $[\mathbb{K}]_{jj'} = \exp(-\frac{1}{4\,\varepsilon_{\text{diff}}}\|\boldsymbol{\eta}^j - \boldsymbol{\eta}^{j'}\|^2)$ that depends on a real smoothing parameter $\varepsilon_{\text{diff}} > 0$. Let $[\mathbb{P}]$ be the transition matrix in $\mathbb{M}_N$ of a Markov chain such that $[\mathbb{P}] = [\mathbb{b}]^{-1}\,[\mathbb{K}]$ in which $[\mathbb{b}]$ is the positive-definite diagonal real matrix such that $[\mathbb{b}]_{ij} = \delta_{ij} \sum_{j'=1}^N [\mathbb{K}]_{jj'}$. For m fixed in $\{1, \ldots, N\}$, let $\mathbf{g}^1, \ldots, \mathbf{g}^m$ be the right eigenvectors in $\mathbb{R}^N$ of matrix $[\mathbb{P}]$ such that $[\mathbb{P}]\,\mathbf{g}^\alpha = \Lambda_\alpha\,\mathbf{g}^\alpha$ in which the eigenvalues are sorted such that $1 = \Lambda_1 > \Lambda_2 > \ldots > \Lambda_m$. It can easily be proven that the eigenvalues are positive and the largest is Λ_1 for which all the components of the corresponding eigenvector $\mathbf{g}^1$ are equal. Considering the normalization $[g]^T\,[\mathbb{b}]\,[g] = [I_m]$, the right-eigenvalue problem of the nonsymmetric matrix $[\mathbb{P}]$ can then be done solving the eigenvalue problem $[\mathbb{b}]^{-1/2}\,[\mathbb{K}]\,[\mathbb{b}]^{-1/2}\,\boldsymbol{\xi}^\alpha = \Lambda_\alpha\,\boldsymbol{\xi}^\alpha$ related to a positive-definite symmetric real matrix with the orthonormalization $< \boldsymbol{\xi}^\alpha, \boldsymbol{\xi}^\beta >= \delta_{\alpha\beta}$. Therefore, $\mathbf{g}^\alpha$ can be deduced from $\boldsymbol{\xi}^\alpha$ by $\mathbf{g}^\alpha = [\mathbb{b}]^{-1/2}\,\boldsymbol{\xi}^\alpha$.

The construction introduces two hyperparameters: the dimension $m \leq N$ and the smoothing parameter $\varepsilon_{\text{diff}} > 0$. An algorithm is proposed in [40] for estimating their values. Most of the time, m and $\varepsilon_{\text{diff}}$ can be chosen as follows. Let $\varepsilon_{\text{diff}} \mapsto \widehat{m}(\varepsilon_{\text{diff}})$ be the function from $]0\,, +\infty[$ into the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of all the integers such that

$$\widehat{m}(\varepsilon_{\text{diff}}) = \arg \min_{\alpha\,|\,\alpha \geq 3} \left\{ \frac{\Lambda_\alpha(\varepsilon_{\text{diff}})}{\Lambda_2(\varepsilon_{\text{diff}})} < 0.1 \right\}\,. \tag{88}$$

If function $\widehat{m}$ is a decreasing function of $\varepsilon_{\text{diff}}$ in the broad sense (if not, see [40]), then the optimal value $\varepsilon^{\text{opt}}_{\text{diff}}$ of $\varepsilon_{\text{diff}}$ can be chosen as the smallest value of the integer $\widehat{m}(\varepsilon^{\text{opt}}_{\text{diff}})$ such that

$$\{\widehat{m}(\varepsilon^{\text{opt}}_{\text{diff}}) < \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]0, \varepsilon^{\text{opt}}_{\text{diff}}[\,\} \cap \{\widehat{m}(\varepsilon^{\text{opt}}_{\text{diff}}) = \widehat{m}(\varepsilon_{\text{diff}})\,, \forall \varepsilon_{\text{diff}} \in]\varepsilon^{\text{opt}}_{\text{diff}}, 1.5\,\varepsilon^{\text{opt}}_{\text{diff}}[\,\}\,. \tag{89}$$

The corresponding optimal value m^{opt} of m is then given by $m^{\text{opt}} = \widehat{m}(\varepsilon^{\text{opt}}_{\text{diff}})$. Random vector $\mathbf{H}$ is centered, but due to the constraints, $\mathbf{X}^c$ can be not centered. Consequently, vector $\mathbf{g}^1$ associated with Λ_1 has to be kept in the basis.

5.8.2. Reduced-order representation of random matrix $[\boldsymbol{H}_{\lambda^t}]$

The diffusion-maps basis represented by matrix $[g] \in \mathbb{M}_{N,m}$ spans a subspace of $\mathbb{R}^N$ that characterizes the local geometry structure of dataset $\{\boldsymbol{\eta}^j, j = 1, \ldots, N\}$. The reduced-order representation is obtained by projecting each column of the $\mathbb{M}_{N,\nu}$-valued random matrix $[\mathbf{H}_{\lambda^t}]^T$ on the subspace of $\mathbb{R}^N$, spanned by $\{\mathbf{g}^1, \ldots, \mathbf{g}^m\}$. Let $[\mathbf{Z}_{\lambda^t}]$ be the random matrix with values in $\mathbb{M}_{\nu,m}$ such that

$$[\mathbf{H}_{\lambda^t}] = [\mathbf{Z}_{\lambda^t}]\,[g]^T\,. \tag{90}$$

Since the matrix $[g]^T\,[g] \in \mathbb{M}_m^+$ is invertible, the least-square approximation of $[\mathbf{Z}_{\lambda^t}]$ can be written as

$$[\mathbf{Z}_{\lambda^t}] = [\mathbf{H}_{\lambda^t}]\,[a]\,, \tag{91}$$

in which the matrix $[a]$ is defined by

$$[a] = [g]\,([g]^T\,[g])^{-1} \in \mathbb{M}_{N,m}\,. \tag{92}$$

It should be noted that the least-square approximation defined by Eq. (91) is only used for calculating the initial conditions for the reduced-order nonlinear ISDE.

5.8.3. Generation of realizations of random matrix $[\boldsymbol{H}_{\lambda^t}]$

The objective is to compute n_{MC} independent realizations $[\eta_{\lambda^t}^1], \ldots, [\eta_{\lambda^t}^{n_{\mathrm{MC}}}]$ of random matrix $[\mathbf{H}_{\lambda^t}]$ in which n_{MC} is assumed to be independent of $\boldsymbol{\lambda}^t$. The reduced-order nonlinear ISDE is constructed as the projection of the nonlinear ISDE defined by Eqs. (63) to (65) on the diffusion-maps basis represented by matrix $[g] \in \mathbb{M}_{N,m}$. We therefore define the $\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m}$- valued stochastic process $\{([\mathcal{Z}_{\lambda^t}(t)], [\boldsymbol{\mathcal{Y}}_{\lambda^t}(t)]), t \geq 0\}$ such that, for all $t \geq 0$,

$$[\mathbf{U}_{\lambda^t}(t)] = [\mathcal{Z}_{\lambda^t}(t)]\,[g]^T \quad , \quad [\mathbf{V}_{\lambda^t}(t)] = [\boldsymbol{\mathcal{Y}}_{\lambda^t}(t)]\,[g]^T\,. \tag{93}$$

The projection of Eqs. (63) to (65) yields

$$d[\mathcal{Z}_{\lambda^t}(t)] = [\boldsymbol{\mathcal{Y}}_{\lambda^t}(t)]\,dt\,, \tag{94}$$

$$d[\boldsymbol{\mathcal{Y}}_{\lambda^t}(t)] = [\mathcal{L}_{\lambda^t}([\mathcal{Z}_{\lambda^t}(t)])]\,dt - \frac{1}{2}\,f_0[\boldsymbol{\mathcal{Y}}_{\lambda^t}(t)]\,dt + \sqrt{f_0}\,d[\boldsymbol{\mathcal{W}}_{\lambda^t}^{\mathrm{wien}}(t)]\,, \tag{95}$$

with the projected initial conditions (see Eqs. (86) and (91)) at $t = 0$,

$$[\mathcal{Z}_{\lambda^t}(0)] = [\eta_{\lambda^{t-1}}^{n_{\mathrm{MC}}}]\,[a] \quad , \quad [\boldsymbol{\mathcal{Y}}_{\lambda^t}(0)] = [v_{\lambda^{t-1}}^{n_{\mathrm{MC}}}]\,[a] \quad , a.s.\,, \tag{96}$$

in which the random matrix $[\mathcal{L}_{\lambda^t}([\mathcal{Z}_{\lambda^t}(t)])]$ with values in $\mathbb{M}_{\nu,m}$ is written as

$$[\mathcal{L}_{\lambda^t}([\mathcal{Z}_{\lambda^t}(t)])] = [L_{\lambda^t}([\mathcal{Z}_{\lambda^t}(t)]\,[g]^T)]\,[a]\,, \tag{97}$$

$$[\boldsymbol{\mathcal{W}}_{\lambda^t}^{\mathrm{wien}}(t)] = [\mathbf{W}_{\lambda^t}^{\mathrm{wien}}(t)]\,[a]\,. \tag{98}$$

Using the results presented in Section 5.7.3, it can be deduced that random matrix $[\mathbf{Z}_{\lambda^t}] = \lim_{t\to+\infty}[\mathcal{Z}_{\lambda^t}(t)]$ in probability distribution. We can then compute n_{MC} independent realizations $[z_{\lambda^t}^1], \ldots, [z_{\lambda^t}^{n_{\mathrm{MC}}}]$ of random matrix $[\mathbf{Z}_{\lambda^t}]$, using the stationary regime of stochastic process $\{[\mathcal{Z}_{\lambda^t}(t)], t > 0\}$ that is the solution of Eqs. (94) to (98). The realizations of $[\mathbf{H}_{\lambda^t}]$ are then obtained using Eq. (90),

$$[\eta_{\lambda^t}^\ell] = [z_{\lambda^t}^\ell]\,[g]^T \in \mathbb{M}_{\nu,N} \quad , \quad \ell = 1, \ldots, n_{\mathrm{MC}}\,. \tag{99}$$

The Störmer-Verlet algorithm is used for solving the reduced-order nonlinear ISDE defined by Eqs. (94) to (98) as proposed in [44, 37]. This algorithm is summarized in Appendix B.

5.8.4. Deducing the realization of random vector $\boldsymbol{H}_{\boldsymbol{\lambda}^\iota}$

Let be $M = n_{\mathrm{MC}} \times N$. Once the realizations $\{[\eta^\ell_{\boldsymbol{\lambda}^\iota}], \ell = 1, \ldots, n_{\mathrm{MC}}\}$ of the $\mathbb{M}_{\nu,N}$-valued random variable $[\mathbf{H}_{\boldsymbol{\lambda}^\iota}]$ have been computed using Eq. (99), the $M \gg N$ realizations $\{\boldsymbol{\eta}^{\ell'}_{\boldsymbol{\lambda}^\iota}, \ell' = 1, \ldots, M\}$ of the $\mathbb{R}^\nu$-valued random variable $\mathbf{H}_{\boldsymbol{\lambda}^\iota}$ are deduced by reshaping the matrices $\{[\eta^\ell_{\boldsymbol{\lambda}^\iota}], \ell = 1, \ldots, n_{\mathrm{MC}}\}$.

5.9. Computing the additional realizations of random vector $\boldsymbol{X}^c$

As soon as the error function defined by Eq. (59) is less that a given small tolerance (that is to say, considering the iteration number ι such that $\mathrm{err}(\iota) \leq \mathrm{tol}$), the iteration algorithm detailed in Section 5.6 is converged and the solution $\boldsymbol{\lambda}^{\mathrm{sol}}$ defined by Eq. (54) is such that $\boldsymbol{\lambda}^{\mathrm{sol}} \simeq \boldsymbol{\lambda}^\iota$. The realizations $\{\boldsymbol{\eta}^{\ell'}_{\boldsymbol{\lambda}^\iota}, \ell' = 1, \ldots, M\}$ of the random vector $\mathbf{H}^c$ constructed in Section 5.8.4, which correspond to the last iteration $\boldsymbol{\lambda}^\iota$, are rewritten as $\{\boldsymbol{\eta}^{c,\ell'}, \ell' = 1, \ldots, M\}$ (that is to say, $\boldsymbol{\eta}^{c,\ell'} = \boldsymbol{\eta}^{\ell'}_{\boldsymbol{\lambda}^\iota}$). Using Eq. (28), the M realizations $\{\mathbf{x}^{c,\ell'}, \ell' = 1, \ldots, M\}$ of random vector $\mathbf{X}^c$ are such that

$$\mathbf{x}^{c,\ell'} = \widehat{\underline{\mathbf{x}}} + [\Phi]\,[\mu]^{1/2}\,\boldsymbol{\eta}^{c,\ell'} \quad , \quad \ell' = 1, \ldots, M\,. \tag{100}$$

5.10. Convergence analysis of the probabilistic learning with respect dimension N of initial dataset

Let us assumed that ε, which has been introduced in Eq. (18), is fixed independently of the value of N. Except for the case for which ε is chosen very small, in general, $\nu = \nu(N)$ depends on N for N belonging to the ordered subset $\mathcal{J}_N = \{N_{\min}, ..., N_{\max}\} \subset \mathbb{N}$ of integers, with $0 < N_{\min} < N_{\max}$. The value of $N_{\max}$ corresponds to the maximum value that is available for constructing the initial dataset $D_{N_{\max}}$. Consequently, the random variable $\mathbf{X}^{c,\nu(N)}$ depends on $N \in \mathcal{J}_N$ and will simply be denoted as $\mathbf{X}^{c,N}$. A question related to the proposed probabilistic learning on manifolds is the convergence of the probability distribution $P_{\mathbf{X}^{c,N}}$ when N goes through the ordered set $\mathcal{J}_N$. This means that, for a given tolerance, when N goes from $N_{\min}$ to $N_{\max}$, the probability distribution $P_{\mathbf{X}^{c,N}}$ must converge. If it is not the case, this means that the learning is not convergent and then $N_{\max}$ has to be increased. Therefore, additional calculations have to be carried out with the computational models for constituting a new bigger training set.

(i) Since the mean-square convergence implies the convergence in probability distribution, we can study the convergence of the learning in analyzing the function

$$N \mapsto \mathrm{conv}_{L^2,\mathbf{X}^c}(N) = E\{\|\mathbf{X}^{c,N}\|^2\}\,. \tag{101}$$

In fact, the limit is unknown and a Cauchy sequence could be studied (Hilbert space $L^2(\Theta, \mathbb{R}^n)$ is a complete vector space). In practice, since set $\mathcal{J}_N$ is not sufficiently big for using such a Cauchy criterion, we can check that $N \mapsto \mathrm{conv}_{L^2,\mathbf{X}^c}(N)$ is flat on $\{N_f, \ldots, N_{\max}\}$ for $N_{\max} < N_f < N_{\max}$.

(ii) Because the contribution of $\mathbf{Q}^{c,N}$ and $\mathbf{W}^{c,N}$, such that $\mathbf{X}^{c,N} = (\mathbf{Q}^{c,N}, \mathbf{W}^{c,N})$, can be very different in the value of $\mathrm{conv}_{L^2,\mathbf{X}^c}(N)$, the definition of $\mathrm{conv}_{L^2,\mathbf{X}^c}(N)$ given by Eq. (101) can be replaced by the following one,

$$N \mapsto \mathrm{conv}_{L^2,\mathbf{X}^c}(N) = \frac{1}{\sqrt{2}} \left\{ \frac{E\{\|\mathbf{Q}^{c,N}\|^2\}}{E\{\|\mathbf{Q}^{N_{\max}}\|^2\}} + \frac{E\{\|\mathbf{W}^{c,N}\|^2\}}{E\{\|\mathbf{W}^{N_{\max}}\|^2\}} \right\}. \tag{102}$$

(iii) Sometimes, for a given $\mathcal{J}_N$, the criteria defined by Eqs. (101) or (102) is not sufficiently sensitive when N goes through $\mathcal{J}_N$. In such a case, and if the constraints are only applied to the

quantities of interest, indicators of the convergence of the learning can be analyzed by studying, for a given subset of components k, the graphs of the functions

$$N \mapsto \mathbb{m}_{\mathbf{Q}_k^c}(N) \quad , \quad N \mapsto \mathbb{s}_{\mathbf{Q}_k^c}(N) \quad , \quad N \mapsto \{q \mapsto p_{\mathbf{Q}_k^c}(q;N)\}\,, \tag{103}$$

of the mean value, the standard deviation, and the pdf of $\mathbf{Q}_k^c$, which are estimated by using the M additional realizations generated with the PLoM.

5.11. Remarks concerning the convergence of the iteration algorithm in the framework of the existence of a solution

If there exists a unique solution (see the comments formulated in Section 5.4-(iv)), then, for all $\iota \geq 1$, $[\Gamma''(\boldsymbol{\lambda}^\iota)]$ is a positive-definite matrix and reciprocally. Let us assume that this hypothesis holds. Since the initial point is correctly chosen as explained in Section 5.6, then the iteration algorithm defined by Eq. (55) is convergent and it is not necessary to introduce an under-relaxation of the type $\boldsymbol{\lambda}^{\iota+1} = \boldsymbol{\lambda}^\iota - \beta_{\text{relax}}\,[\Gamma''(\boldsymbol{\lambda}^\iota)]^{-1}\,\boldsymbol{\nabla}\Gamma(\boldsymbol{\lambda}^\iota)$ with $\beta_{\text{relax}} \in]0,1[$. If for a given iteration number ι, the rank of positive matrix $[\Gamma''(\boldsymbol{\lambda}^\iota)]$ becomes less than ν_c, then solution $\boldsymbol{\lambda}^{\text{sol}}$ cannot be constructed in $C_{\text{ad},\boldsymbol{\lambda}}$, which means that there is no solution. This difficulty is related to the value of ν_c and cannot be compensated using an under-relaxation. The introduction of an under-relaxation does not modify the fact that there will exist $\iota' > \iota$ (before reaching the convergence) such that the rank of $[\Gamma''(\boldsymbol{\lambda}^{\iota'})]$ is less than ν_c (all the numerical experiments that have been done have confirmed this analysis). In fact, the existence of a solution is related to the number ν_c of projected constraint equations defined by Eq. (33), which is the projection by $[\Psi] \in \mathbb{M}_{m_c,\nu_c}$ of the constraint equations defined by the Eq. (10). If τ_c is too small, then ν_c is too big compared to the stochastic dimension ν_s that have been introduced in Section 5.4-(iv). This means that the number ν_c of projected constraint equations is too large with respect to the number ν_s of "stochastic degrees of freedom" and consequently, the projected constraint equations defined by Eq. (33) cannot be exactly satisfied. An effective method to solve such a non-convergence problem is to slightly increase the value τ_c, which makes it possible to obtain a slightly smaller rank ν_c and thus to reduce the number of projected constraint equations (all the numerical experiments that have been conducted validate this analysis).

Assumptions for the applicability of the proposed method. The proposed method is general. The given constraints for performing the probabilistic learning on manifolds are not limited to second-order moments but can be expressed as the mathematical expectation of any measurable mapping of the control parameters and the quantities of interest. For example, marginal probability densities can be imposed as shown in Appendix A. There is obviously a domain of validity, which is not due to the method proposed itself, but which is due to the coherence of the given constraints with respect to the initial dataset (that is used for the probabilistic learning). If the introduced constraints are not consistent with the initial dataset, then the optimization problem will have no solution. The existence of this solution is therefore conditioned by hypotheses that are discussed above and in Section 5.4-(iv).

6. Application (AP1)

In this section, an application is presented and is used for performing the validation of the methodology and the algorithms presented. All the random variables are defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$. This application that will be referenced as (AP1) is sufficiently simple in order

to be easily reproducible by the reader and will also allow for presenting a detailed analysis for several aspects of the proposed method.

6.1. Stochastic model

The stochastic model used for generating the initial dataset $\mathbb{D}_N = \{\mathbb{x}^j = (\mathbb{q}^j, \mathbb{w}^j),\, j = 1, \ldots, N\}$ (see the Remark at the end of Section 2), related to random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$ in which $\mathbb{Q} = (\mathbb{Q}_1, \ldots, \mathbb{Q}_{n_q})$ and $\mathbb{W} = (\mathbb{W}_1, \ldots, \mathbb{W}_{n_w})$, is written as

$$\mathbb{Q} = [\mathbb{B}(\mathbb{U})]\,(\mathbb{W} + V\,\mathbb{b}_{\text{AP1}})\,, \tag{104}$$

in which $\mathbb{U}$, V, and $\mathbb{W}$ are independent random variables. The maximum value of N is 300 and $n_w = 20$. We have $n_q = 200$. The deterministic vector $\mathbb{b}_{\text{AP1}}$ in $\mathbb{R}^{n_w}$ is written as $\mathbb{b}_{\text{AP1}} = 0.2\,\mathbf{u} + 0.9$ in which all the components of $\mathbf{u}$ belong to $]0, 1[$ (generated with the Matlab script: *rng*('default'); $\mathbf{u} = rand(n_w, 1)$). The real-valued random variable $V = 0.2\,\mathcal{U} + 0.9$ in which $\mathcal{U}$ is a uniform random variable on $[0, 1]$. The random vector $\mathbb{U} = (\mathbb{U}_1, \ldots, \mathbb{U}_{n_u})$ with $n_u = 6$ is written, for $\alpha = 1, \ldots, n_u$, as $\mathbb{U}_\alpha = 2\,u_\alpha\,\mathcal{U}_\alpha + 1 - u_\alpha$ in which $\mathcal{U}_1, \ldots, \mathcal{U}_{n_u}$ are n_u independent uniform random variables on $[0, 1]$ and where, $u_\alpha = 0.2(\alpha - 1)/(n_u - 1)$. The entries $[\mathbb{B}(\mathbb{U})]_{kj}$ of the $(n_q \times n_w)$ random matrix are defined by $[\mathbb{B}(\mathbb{U})]_{kj} = \sum_{\alpha=1}^{n_u} \omega_\alpha(\mathbb{U}_\alpha)\,\phi_k^\alpha(\mathbb{U}_\alpha)\,\phi_{j+n_q/2}^\alpha(\mathbb{U}_\alpha)$ in which $\phi_k^\alpha(\mathbb{U}_\alpha) = \sin\{\alpha\, k\pi/(n_q + 1)\}$ is independent of $\mathbb{U}_\alpha$ (deterministic) and $\omega_\alpha(\mathbb{U}_\alpha) = 1/(\alpha\mathbb{U}_\alpha)^2$. The random vector $\mathbb{W}$ is written as $\mathbb{W} = \sum_{\beta=1}^{3} \sqrt{\varpi_\beta}\,\boldsymbol{\phi}_{\mathbb{w}}^\beta\,\mathbb{A}_\beta$, in which $\varpi_\beta = 1/\beta^2$ and $\boldsymbol{\phi}_{\mathbb{w}}^\beta = (\phi_{\mathbb{w},1}^\beta, \ldots, \phi_{\mathbb{w},n_w}^\beta)$ with $\phi_{\mathbb{w},j}^\beta = \sin\{\beta\pi\, j/(1 + n_w)\}$. The non-Gaussian centered random vector $\mathbb{A} = (\mathbb{A}_1, \mathbb{A}_2, \mathbb{A}_3)$ is written as $\mathbb{A} = \sum_{\gamma=1}^{27} \mathbb{a}^\gamma\,\psi_{\alpha_1^{(\gamma)}}(\Xi_1)\,\psi_{\alpha_2^{(\gamma)}}(\Xi_2)$ in which Ξ_1 and Ξ_2 are independent normalized Gaussian random variables. The indices $\alpha_1^{(\gamma)}$ and $\alpha_2^{(\gamma)}$ are such that $0 < \alpha_1^{(\gamma)} + \alpha_2^{(\gamma)} \leq 6$, and $\psi_{\alpha_1^{(\gamma)}}(\Xi_1)$ and $\psi_{\alpha_2^{(\gamma)}}(\Xi_2)$ are the polynomial chaos in Gaussian variables. The matrix $[\mathbb{a}] = [\mathbb{a}^1 \ldots \mathbb{a}^{27}]$ is such that $[\mathbb{a}]\,[\mathbb{a}]^T = [I_3]$ and is generated using the Matlab script: *rng*('default'); $M_1 = randn(27,27)$; $[M_2,\sim] = eig(M_1{*}(M_1)')$; $M_2(:, 4{:}27) = []$; $[\mathbb{a}] = (M_2)'$.

6.2. Simulated experiments

Simulated experiments are not required for the proposed methodology. Only some statistics corresponding to the constraints must be specified. Nevertheless, in order to generate coherent statistics for the constraints (but also for validating the methodology), we introduce an experimental dataset that is generated with an "experimental model". This experimental dataset is used for estimating the statistics that correspond to the constraints. We insist on the fact that for applying the method proposed, the experimental dataset is not used and consequently, is not required.

The experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 200$ independent experimental realizations $\{\mathbb{q}^{\text{exp},r}, r = 1, \ldots n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \ldots, \mathbb{Q}_{n_q}^{\text{exp}})$, which are used for estimating the right-hand side member of Eq. (10). The simulated experiments correspond to model uncertainties induced by modeling errors and are constructed using significant perturbations of the stochastic model. The experimental stochastic model is then written as

$$\mathbb{Q}^{\text{exp}} = [\mathbb{B}(\mathbb{U}^{\text{exp}})]\,(\mathbb{W}^{\text{exp}} + V^{\text{exp}}\,\mathbb{b}_{\text{AP1}})\,, \tag{105}$$

in which $\mathbb{U}^{\text{exp}}$, V^{exp}, and $\mathbb{W}^{\text{exp}}$ are independent random variables that are also independent of $\mathbb{U}$, V, and $\mathbb{W}$. The real-valued random variable $V^{\text{exp}} = 0.2\,\mathcal{U}^{\text{exp}} + 0.9$ in which $\mathcal{U}^{\text{exp}}$ is a uniform random variable on $[0, 1]$ independent of $\mathcal{U}$. The random vector $\mathbb{U}^{\text{exp}} = (\mathbb{U}_1^{\text{exp}}, \ldots, \mathbb{U}_{n_u}^{\text{exp}})$ is

written, for $\alpha = 1, \ldots, n_u$, as $\mathbb{U}_\alpha^{\rm exp} = 2\, u_\alpha^{\rm exp}\, \mathcal{U}_\alpha^{\rm exp} + 1 - u_\alpha^{\rm exp}$ in which $\mathcal{U}_1^{\rm exp}, \ldots, \mathcal{U}_{n_u}^{\rm exp}$ are n_u independent uniform random variables on $[0, 1]$ and where, $u_\alpha^{\rm exp} = 0.3(\alpha - 1)/(n_u - 1)$. Note that the coefficient is 0.3 and not 0.2 as in the stochastic model. The mapping $\mathbb{u} \mapsto [\mathbb{B}(\mathbb{u})]$ is the same as the one of the stochastic model. The random vector $\mathbb{W}^{\rm exp}$ is written as

$$\mathbb{W}^{\rm exp} = \zeta_{\rm MEAN} \times \mathbf{1}_{n_w} + \zeta_{\rm STD} \times \widetilde{\mathbb{W}}^{\rm exp}\,, \tag{106}$$

in which $\mathbf{1}_{n_w} \in \mathbb{R}^{n_w}$ is the vector whose components are equal to 1 and where $\widetilde{\mathbb{W}}^{\rm exp}$ is an independent copy of the stochastic model of $\mathbb{W}$. The parameter $\zeta_{\rm MEAN}$ controls the mean value of $\mathbb{W}^{\rm exp}$ while $\zeta_{\rm STD}$ controls the amplitude of its covariance matrix. Many values of $(\zeta_{\rm MEAN}, \zeta_{\rm STD})$ have been considered in the subset $[-0.2, +0.7] \times [0, 1.4]$ for analyzing the robustness of the algorithm. In order to limit the number of figures, we will limit the presentation of the results to the value $\zeta_{\rm MEAN} = 0.4$ and $\zeta_{\rm STD} = 1.4$.

6.3. Scaling

Using the block writing $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, the scaling of the $\mathbb{R}^{n_q}$-valued random variable $\mathbb{Q}$ and the scaling of the $\mathbb{R}^{n_w}$-valued random variable $\mathbb{W}$ are written, using Eq. (5), as

$$\mathbb{Q} = [\alpha_q]\, \mathbf{Q} + \mathbb{q}^{\rm min} \quad , \quad \mathbf{Q} = [\alpha_q]^{-1} (\mathbb{Q} - \mathbb{q}^{\rm min})\,, \tag{107}$$

$$\mathbb{W} = [\alpha_w]\, \mathbf{W} + \mathbb{w}^{\rm min} \quad , \quad \mathbf{W} = [\alpha_w]^{-1} (\mathbb{W} - \mathbb{w}^{\rm min})\,, \tag{108}$$

in which $\mathbb{x}^{\rm min} = (\mathbb{q}^{\rm min}, \mathbb{w}^{\rm min})$ and where $[\alpha_q]$ and $[\alpha_w]$ correspond to the block writing of $[\alpha_x]$. The scaling of Eq. (104) is then written as

$$\mathbf{Q} = [B(\mathbb{U})]\, (\mathbf{W} + V\, \mathbf{b}_{\rm AP1} + \mathbf{c}_{\rm AP1}) + \mathbf{d}_{\rm AP1}\,, \tag{109}$$

in which $[B(\mathbb{U})] = [\alpha_q]^{-1}\, [\mathbb{B}(\mathbb{U})]\, [\alpha_w]$, $\mathbf{b}_{\rm AP1} = [\alpha_w]^{-1}\, \mathbb{b}_{\rm AP1}$, $\mathbf{c}_{\rm AP1} = [\alpha_w]^{-1}\, \mathbb{w}^{\rm min}$, and $\mathbf{d}_{\rm AP1} = -[\alpha_q]^{-1}\, \mathbb{q}^{\rm min}$. The scaling of $\mathbb{Q}^{\rm exp}$ is such that $\mathbf{Q}^{\rm exp} = [\alpha_q]^{-1} (\mathbb{Q}^{\rm exp} - \mathbb{q}^{\rm min})$, which shows that the n_r independent experimental realizations $\{\mathbf{q}^{{\rm exp},r}, r = 1, \ldots n_r\}$ of $\mathbf{Q}^{\rm exp}$ are such that

$$\mathbf{q}^{{\rm exp},r} = [\alpha_q]^{-1} (\mathbb{q}^{{\rm exp},r} - \mathbb{q}^{\rm min})\,. \tag{110}$$

6.4. Constraints

The constraints are those that are defined in Section 4, that is to say by Eqs. (13) and (14) in which $\kappa_{\rm mom} = n_q$ and where the experimental moments $\{(\mathbb{m}_\kappa^{\rm exp}, \mathbb{s}_\kappa^{\rm exp}), \kappa = 1, \ldots, \kappa_{\rm mom}\}$ are estimated using the n_r independent experimental realizations $\{\mathbf{q}^{{\rm exp},r}, r = 1, \ldots n_r\}$ defined by Eq. (110). Consequently, the number of constraints is $m_c = 2\, n_q = 400$.

6.5. Values of the parameters of the algorithms

In this paragraph, all the values of the parameters are related to $N = 300$. The PCA of $\mathbf{X}$ (see Section 5.1) is performed with $\varepsilon = 10^{-6}$ in Eq. (17) and yields $\nu = 9$. The values of the two parameters of the Gaussian kernel-density estimation defined by Eq. (24) are $s_\nu = 0.5966$ and $\widehat{s}_\nu = 0.5129$. Concerning the construction of the diffusion-maps basis, the use of Eqs. (88) and (89) yields $\varepsilon_{\rm diff}^{\rm opt} = 48$ and the optimal value of m is $m^{\rm opt} = 12$. The graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1, which corresponds to these optimal values, is shown in Fig. 1-(a). For solving the reduced-order nonlinear ISDE defined in Section 5.8.3, the parameters related to the algorithm are those introduced in Appendix B, have

been estimated using the procedure that is detailed in Section 4.7.2 in [37], and are the following: $f_0 = 1.5$, $n_{\text{MC}} = 50$, $\ell_0 = 10$, $M_0 = 10$, $\Delta t = 0.16115$, and then, $M = 15\,000$. The projection of the constraints introduced in Section 5.3 has been performed with $\tau_c = 0.001$ that yields $\nu_c = 3$ (see Eq. (39)). Figure 1-(b) displays the graph of the error function $\iota \mapsto \text{err}(\iota)$ that is defined by Eq. (59), which shows the convergence of the iteration algorithm for computing the Lagrange multiplier $\boldsymbol{\lambda}^{\text{sol}}$. It can be seen that the convergence is fast, the error decreases exponentially as a function of the iteration number ι.

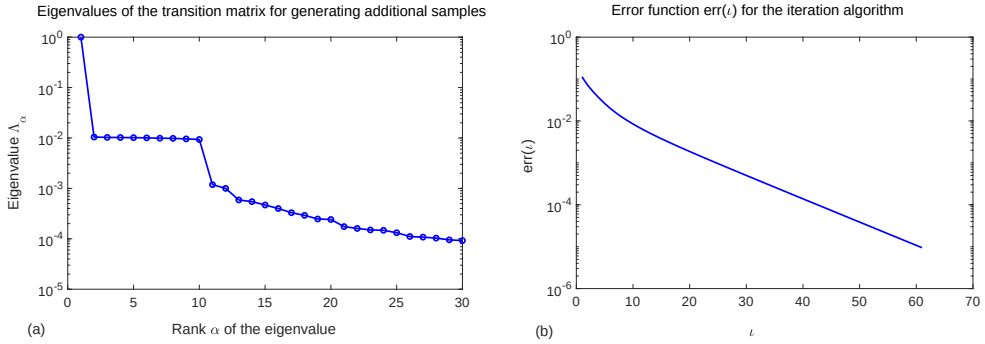

Figure 1: Application AP1. Figure (a): graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1. Figure (b): graph of the error function $\iota \mapsto \text{err}(\iota)$ defined by Eq. (59) showing the convergence of the iteration algorithm for computing the Lagrange multiplier $\boldsymbol{\lambda}^{\text{sol}}$. The vertical axis is in log.

6.6. Convergence of the learning with respect to the value of N

For this application, the use of the criterion defined by Eq. (101) or by Eq. (102) for analyzing the convergence of the learning with respect to dimension N of the initial dataset is not sufficiently informative. For $N \in [125, 300]$, $\text{conv}_{L^2,\mathbf{X}^c}(N)$ fluctuates between 9.86 and 9.92. Consequently, for this application, the convergence of the learning is analyzed by using Section 5.10-(iii), that is to say by studying the sequence in N of graphs of the functions $\{k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}(N)\}_N$ and $\{q \mapsto p_{\mathbb{Q}_{110}^c}(q;N)\}_N$ for $N \in \{125, 150, 250, 300\}$. These quantities are related to the unscaled vector-valued random variable $\mathbb{Q}^c$ with values in $\mathbb{R}^{n_q}$ for which $n_q = 200$. For fixed N, $\mathbb{s}_{\mathbb{Q}_k^c}(N)$ is the standard deviation of component $\mathbb{Q}_k^c$ and $q \mapsto p_{\mathbb{Q}_{110}^c}(q;N)$ is the pdf of $\mathbb{Q}_{110}^c$ (only component 110 that corresponds to the maximum of response has been selected for limiting the number of figures). Figure 2-(a) shows the graphs of standard-deviation functions $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}(N)$ for which a zoom in k has been carried out in order to better separate the curves for viewing the convergence in N. Figure 2-(b) shows the graphs of pdf-functions $q \mapsto p_{\mathbb{Q}_{110}^c}(q;N)$ for the same values of N. The convergence of the learning is particularly clear for the standard-deviation functions. For $N = 250$ and $N = 300$ the two curves are practically overlapping.

6.7. Results obtained and validation of the method for N = 300 and M = 15 000

For $k = 1, \ldots, 200$, let $\mathbb{Q}_k$ be the component k of the random unscaled QoI $\mathbb{Q}$ (the scaled one is $\mathbf{Q}$) associated with the initial dataset $\mathbb{D}_N$, let $\mathbb{Q}_k^{\text{exp}}$ be the component k of the random unscaled experimental QoI $\mathbb{Q}^{\text{exp}}$ (the scaled one is $\mathbf{Q}^{\text{exp}}$), and let $\mathbb{Q}_k^c$ be the component k of the random unscaled QoI $\mathbb{Q}^c$ that verifies the constraints (the scaled one is $\mathbf{Q}^c$). The quality assessment

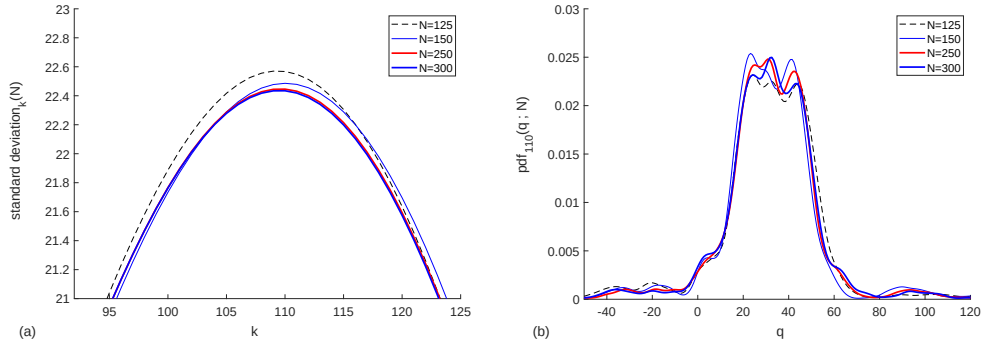

Figure 2: Convergence of the probabilistic learning for application AP1 with $M = 15\,000$ and for N in $\{125, 150, 250, 300\}$. Figure (a): graphs of the standard-deviation function $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}(N)$ for $k = 92, \ldots, 125$. Figure (b): graphs of probability density function $q \mapsto p_{\mathbb{Q}_{110}^c}(q; N)$.

consists in comparing $\mathbb{Q}^c$ with the random vector $\mathbb{Q}^{c,\text{model}}$ that is transformed from $\mathbb{W}^c$ using the stochastic computational model, that is to say, $\mathbb{Q}^{c,\text{model}} = [\mathbb{B}(\mathbb{U})]\,(\mathbb{W}^c + V\,\mathbb{b}_{\text{AP1}})$. For that the pdf of component $\mathbb{Q}_k^c$ can be compared to the pdf of component $\mathbb{Q}_k^{c,\text{model}}$.

(i) For $k = 1, \ldots, 200$, Fig. 3-(a) displays the graphs of the mean functions $k \mapsto \mathbb{m}_{\mathbb{Q}_k}$, $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{\text{exp}}}$, $k \mapsto \mathbb{m}_{\mathbb{Q}_k^c}$, and $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{c,\text{model}}}$. Figure 3-(b) shows the graphs of the standard-deviation functions $k \mapsto \mathbb{s}_{\mathbb{Q}_k}$, $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{\text{exp}}}$, $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}$, and $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{c,\text{model}}}$. In each one of these two figures, it can be seen that the two curves $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{\text{exp}}}$ and $k \mapsto \mathbb{m}_{\mathbb{Q}_k^c}$ are indistinguishable and that the two curves $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{\text{exp}}}$ and $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}$, what validates the method proposed: the constraints on the mean and the standard deviation are well taken into account. Concerning the quality assessment, it can also be seen that the two curves $k \mapsto \mathbb{m}_{\mathbb{Q}_k^c}$ and $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{c,\text{model}}}$ are overlapping and also that the two curves $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}$ and $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{c,\text{model}}}$ are also overlapping, demonstrating the capability of the method at least for the second-order moments of the QoI.

(ii) To complete the analysis and the validation, we present the results obtained for the probability density functions. In order to limit the number of figures, we have selected, as previously, component 110 for $\mathbb{Q}$. Figure 4 displays the graphs of the probability density functions $q \mapsto p_{\mathbb{Q}_{110}}(q)$, $q \mapsto p_{\mathbb{Q}_{110}^{\text{exp}}}(q)$, $q \mapsto p_{\mathbb{Q}_{110}^c}(q)$, and $q \mapsto p_{\mathbb{Q}_{110}^{c,\text{model}}}(q)$. The constraints that are imposed concern only the second-order moments of $\mathbb{Q}$. Consequently, there is no reason for that the pdf of $\mathbb{Q}_{110}^c$ coincides with the pdf of $\mathbb{Q}_{110}^{\text{exp}}$. Nevertheless, it can be seen that the probabilistic learning on manifolds under these constraints yields a pdf that is not so bad. Note also that, in practice and in the framework of this application, the pdf of $\mathbb{Q}_{110}^{\text{exp}}$ cannot be estimated because the n_r realizations of $\mathbb{Q}^{\text{exp}}$ are not available. Nevertheless, it can be seen that the quality assessment is good enough because pdf $q \mapsto p_{\mathbb{Q}_{110}^{c,\text{model}}}(q)$ is very close to pdf $q \mapsto p_{\mathbb{Q}_{110}^c}(q)$.

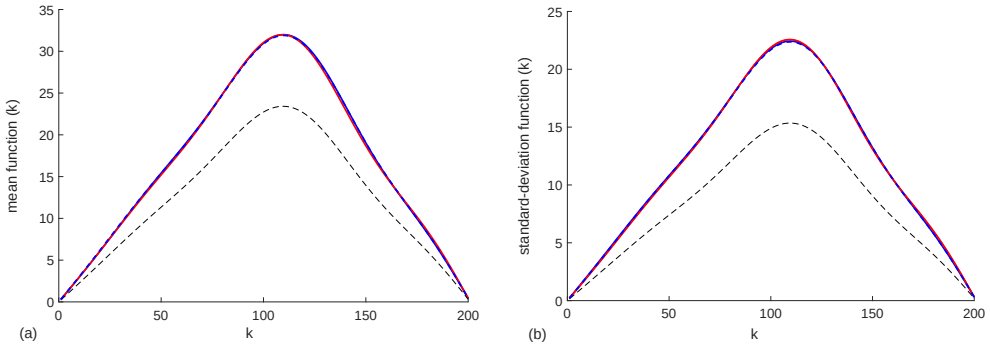

Figure 3: Application AP1 with $M = 15\,000$, for $N = 300$, for $n_r = 200$, and for $k = 1, \ldots, 200$. Figure (a): graphs of the mean functions $k \mapsto \mathbb{m}_{\mathbb{Q}_k}$ (black dashed line), $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{\exp}}$ (red line), $k \mapsto \mathbb{m}_{\mathbb{Q}_k^c}$ (blue line), and $k \mapsto \mathbb{m}_{\mathbb{Q}_k^{c,\mathrm{model}}}$ (blue dashed line). Figure (b): graphs of the standard-deviation functions $k \mapsto \mathbb{s}_{\mathbb{Q}_k}$ (black dashed line), $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{\exp}}$ (red line), $k \mapsto \mathbb{s}_{\mathbb{Q}_k^c}$ (blue line), and $k \mapsto \mathbb{s}_{\mathbb{Q}_k^{c,\mathrm{model}}}$ (blue dashed line). In each figure, the red line, the blue line, and the blue dashed line are overlapping.

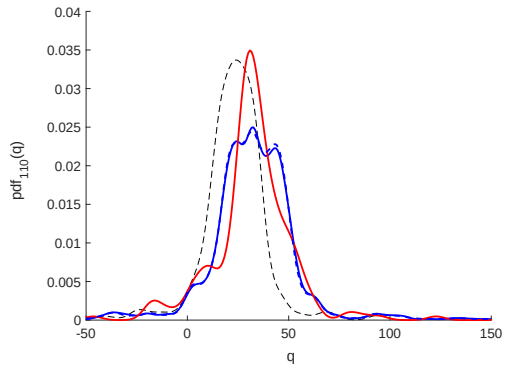

Figure 4: Application AP1: For $M = 15\,000$, for $N = 300$, for $n_r = 200$, graphs of the probability density functions $q, \to p_{\mathbb{Q}_{110}}(q)$ (black dashed line), $q \mapsto p_{\mathbb{Q}_{110}^{\exp}}(q)$ (red line), $q \mapsto p_{\mathbb{Q}_{110}^c}(q)$ (blue line), and $q \mapsto p_{\mathbb{Q}_{110}^{c,\mathrm{model}}}(q)$ (blue dashed line, almost overlapping with the blue line).

7. Application (AP2): Probabilistic learning of the random elasticity field for a heterogeneous anisotropic medium

In this section, we present an application in high dimension related to the learning of the random elasticity field for a heterogeneous anisotropic medium. This application is described in details in order that it can be reproduced. The $\mathbb{R}^{n_w}$-valued random system parameter, $\mathbb{W}$, corresponds to the spatial discretization of the tensor-valued random elasticity field for which $n_w = 420\,000$. The $\mathbb{R}^{n_q}$-valued random quantity of interest, $\mathbb{Q}$, has dimension $n_q = 1\,683$ corresponding to observed components of the displacement field at 561 points located on the boundary of the domain occupied by the heterogeneous material, and $m_c = 3\,366$ constraints defined in Section 4 are introduced. All the random variables are defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$.

7.1. Construction of the initial dataset

(i) Stochastic model defined by a stochastic boundary value problem. The stochastic model is used for generating the initial dataset and is based on the stochastic boundary value problem defined hereinafter. The medium occupies the domain $\Omega =]0, 1.0[\times]0, 0.2[\times]0, 0.1[\, m^3$ with generic point $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3)$. The boundary $\partial\Omega$ of Ω is written as $\partial\Omega = \Gamma_0 \cup \Gamma \cup \Gamma_{\text{obs}} \cup \Gamma_c$ in which

$$\Gamma_0 = \{\zeta_1 = 1.0\,,\ 0 \leq \zeta_2 \leq 0.2\,,\ 0 \leq \zeta_3 \leq 0.1\}\,,$$

$$\Gamma = \{\zeta_1 = 0.0\,,\ 0 \leq \zeta_2 \leq 0.2\,,\ 0 \leq \zeta_3 \leq 0.1\}\,,$$

$$\Gamma_{\text{obs}} = \{0 \leq \zeta_1 \leq 1.0\,,\ 0 \leq \zeta_2 \leq 0.2\,,\ \zeta_3 = 0.1\}\}\,,$$

and where $\Gamma_c = \partial\Omega^{\text{meso}} \backslash \{\Gamma_0 \cup \Gamma \cup \Gamma_{\text{obs}}\}$. The outward unit normal to $\partial\Omega$ is denoted by $\mathbf{n}(\boldsymbol{\zeta})$. The heterogeneous complex material is modeled by a heterogeneous and anisotropic elastic random medium for which the elastic properties are defined by the fourth-order tensor-valued non-Gaussian random field $\{\mathbb{C} = \{\mathbb{C}_{ijkh}(\boldsymbol{\zeta})\}_{ijkh}, \boldsymbol{\zeta} \in \Omega\}$, which is rewritten, for i, j, k, and h in $\{1, 2, 3\}$, as

$$\mathbb{C}_{ijkh}(\boldsymbol{\zeta}) = [\mathbf{K}(\boldsymbol{\zeta})]_{IJ} \quad \text{with} \quad I = (i, j) \quad \text{and} \quad J = (k, h)\,, \tag{111}$$

in which indices I and J belong to $\{1, \ldots, 6\}$ and where the matrix-valued non-Gaussian random elasticity field $\{[\mathbf{K}(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \Omega\}$ will be defined after. The $\mathbb{R}^3$-valued displacement field $\boldsymbol{\Xi} = (\Xi_1, \Xi_2, \Xi_3)$ is defined in Ω. A Dirichlet condition $\boldsymbol{\Xi} = \mathbf{0}$ is given on Γ_0 while a Neumann condition is given on $\Gamma \cup \Gamma_{\text{obs}} \cup \Gamma_c$. The Neumann condition is zero on $\Gamma_{\text{obs}} \cup \Gamma_c$ while on Γ, a given random surface force field $\boldsymbol{\mathcal{G}}^{\Gamma}$ is applied (note that there is no surface force field applied to Γ_{obs}, which is the part of the boundary for which field $\boldsymbol{\Xi}$ is observed). The stochastic boundary value problem is written as

$$-\text{div}\ \boldsymbol{\Sigma} = \mathbf{0} \quad \text{in} \quad \Omega\,, \tag{112}$$

$$\boldsymbol{\Xi} = \mathbf{0} \quad \text{on} \quad \Gamma_0\,, \tag{113}$$

$$\boldsymbol{\Sigma}\,\mathbf{n} = \boldsymbol{\mathcal{G}}^{\Gamma} \quad \text{on} \quad \Gamma\,, \tag{114}$$

$$\boldsymbol{\Sigma}\,\mathbf{n} = \mathbf{0} \quad \text{on} \quad \Gamma_{\text{obs}} \cup \Gamma_c\,. \tag{115}$$

The stress tensor $\boldsymbol{\Sigma} = \{\Sigma_{ij}\}_{ij}$ is related to the strain tensor $\mathbf{E}(\boldsymbol{\Xi}) = \{E_{kh}(\boldsymbol{\Xi})\}_{kh}$ by the constitutive equation,

$$\Sigma_{ij}(\boldsymbol{\zeta}) = \mathbb{C}_{ijkh}(\boldsymbol{\zeta})\, E_{kh}(\boldsymbol{\Xi}(\boldsymbol{\zeta})) \tag{116}$$

in which the strain tensor is such that $E_{kh}(\boldsymbol{\Xi}) = (\partial\Xi_k/\partial\zeta_h + \partial\Xi_h/\partial\zeta_k)/2$.

(ii) Stochastic model of the applied forces. The surface force field applied on Γ induces the superposition (1) of a mean tension in the ζ_1-axis induced by a surface force field applied in the $-\zeta_1$-axis direction with a constant amplitude g^{ζ_1}, (2) of a mean bending around ζ_3-axis induced by a surface force field applied in the $-\zeta_2$-axis direction with a constant amplitude g^{ζ_2}, and (3) of a mean bending around ζ_2-axis induced by a surface force field applied in the $-\zeta_3$-axis direction with a constant amplitude g^{ζ_3}. Consequently, the $\mathbb{R}^3$-valued random surface force field $\{\boldsymbol{\mathcal{G}}^{\Gamma}(\boldsymbol{\zeta}) = (\mathcal{G}_1^{\Gamma}(\boldsymbol{\zeta}), \mathcal{G}_2^{\Gamma}(\boldsymbol{\zeta}), \mathcal{G}_3^{\Gamma}(\boldsymbol{\zeta})), \boldsymbol{\zeta} \in \Gamma\}$ applied on surface Γ is defined as follows. For all $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3)$ in Γ,

$$\mathcal{G}_1^{\Gamma}(\boldsymbol{\zeta}) = -g^{\zeta_1} \cos(\mathbb{U}_1)\,, \tag{117}$$

$$\mathcal{G}_2^{\Gamma}(\boldsymbol{\zeta}) = -g^{\zeta_1} \sin(\mathbb{U}_1) - g^{\zeta_2} \cos(\mathbb{U}_2) - g^{\zeta_3} \sin \mathbb{U}_3\,, \tag{118}$$

$$\mathcal{G}_3^{\Gamma}(\boldsymbol{\zeta}) = -g^{\zeta_2} \sin(\mathbb{U}_2) - g^{\zeta_3} \cos \mathbb{U}_3 \,, \tag{119}$$

where $\mathbb{U} = (\mathbb{U}_1, \mathbb{U}_2, \mathbb{U}_3)$ is a $\mathbb{R}^3$-valued random variable such that

$$\mathbb{U}_k = \alpha_0 \, (2\, \mathcal{U}_k - 1) \quad , \quad \alpha_0 = 2\pi \alpha_0^d / 360 \quad , \quad \alpha_0^d = 1^o \,, \tag{120}$$

in which $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$ are three independent uniform random variables on $[0, 1]$, which are also independent of random field $\mathbb{C}$. It can then be deduced that $\mathbb{U}$ is independent of random field $\mathbb{C}$. The amplitudes are such that $g^{\zeta_1} = 15\,000\, N/m^2$, $g^{\zeta_2} = 130\, N/m^2$, and $g^{\zeta_3} = 50\, N/m^2$.

(iii) Stochastic model of the random elasticity field. The non-Gaussian fourth-order tensor-valued random field $\{\mathbb{C}(\boldsymbol{\zeta}), \boldsymbol{\zeta} \in \Omega\}$ is assumed to be of second order and statistically homogeneous. Its mean function is thus independent of $\boldsymbol{\zeta}$ and is an elasticity tensor denoted by $\underline{\mathbb{C}}$, which is defined using its representation $[\underline{K}] \in \mathbb{M}_6^+$ (see Eq. (111)) such that $\underline{\mathbb{C}}_{ijkh} = [\underline{K}]_{IJ}$. Elasticity tensor $\underline{\mathbb{C}}$ is associated with a homogeneous isotropic elastic material whose Young modulus is $10^{10}\, N/m^2$ and Poisson coefficient 0.15 (note that the fluctuations are those of a heterogeneous anisotropic elastic material). The non-Gaussian $\mathbb{M}_6^+$-valued random field $\{[\mathbf{K}(\boldsymbol{\zeta})]\,, \boldsymbol{\zeta} \in \Omega\}$ is constructed using the stochastic model [56, 49] of random elasticity fields for heterogeneous anisotropic elastic media that are isotropic in statistical mean and exhibit anisotropic statistical fluctuations, for which the parameterization consists of spatial-correlation lengths and of a positive-definite lower bound. The random field $\{[\mathbf{K}(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \Omega\}$ is written as,

$$[\mathbf{K}(\boldsymbol{\zeta})] = [C_\ell] + [\,\underline{C}\,]^{1/2}\, [\mathbf{G}_0(\boldsymbol{\zeta})]\, [\,\underline{C}\,]^{1/2} \quad , \quad \forall \boldsymbol{\zeta} \in \Omega \,. \tag{121}$$

The lower-bound matrix is defined by $[C_\ell] = \varepsilon_\ell (1 + \varepsilon_\ell)^{-1}\, [\underline{K}] \in \mathbb{M}_6^+$ in which ε_ℓ is chosen equal to 10^{-6}. Consequently, $[\,\underline{C}\,] = [\,\underline{K}\,] - [C_\ell] = (1 + \varepsilon_\ell)^{-1}\, [\underline{K}]$ belongs to $\mathbb{M}_6^+$. Note that $[\,\underline{C}\,]^{1/2}$ is the square root of matrix $[\,\underline{C}\,]$ in $\mathbb{M}_n^+$. The non-Gaussian random field $\{[\mathbf{G}_0(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$, which is indexed by $\mathbb{R}^3$ with values in $\mathbb{M}_6^+$, is homogeneous in $\mathbb{R}^3$ and is a second-order random field such that, for all $\boldsymbol{\zeta}$ in $\mathbb{R}^3$,

$$E\{[\mathbf{G}_0(\boldsymbol{\zeta})]\} = [I_6] \quad , \quad [\mathbf{G}_0(\boldsymbol{\zeta})] > 0 \quad \text{a.s.} \tag{122}$$

It can then be deduced that, for all $\boldsymbol{\zeta}$ in Ω,

$$E\{[\mathbf{K}(\boldsymbol{\zeta})]\} = [\,\underline{K}\,] \in \mathbb{M}_6^+ \quad , \quad [\mathbf{K}(\boldsymbol{\zeta})] - [C_\ell] > 0 \quad \text{a.s.} \tag{123}$$

Random field $\{[\mathbf{G}_0(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$ depends on three spatial correlation lengths, L_1, L_2, and L_3, and on a dispersion parameter, δ_{G_0}. Its construction and its generator are summarized in Appendix C.

With such a stochastic model of the elasticity field, it is proven in [56] that the stochastic boundary value problem defined by Eqs. (112) to (116) admits a unique second-order solution.

(iv) Finite element approximation of the stochastic boundary value problem and construction of random vectors $\mathbb{Q}$ *and* $\mathbb{W}$. Domain $\Omega =]0, 1.0[\times]0, 0.2[\times]0, 0.1[$ is meshed with $50 \times 10 \times 5 = 2\,500$ finite elements using 8-nodes finite elements. There are 3 366 nodes and 10 098 dofs (degrees of freedom). The displacements are locked at all the 66 nodes belonging to surface Γ_0 and therefore, there are 198 zero Dirichlet conditions. The $\mathbb{R}^{n_q}$-valued random variable $\mathbb{Q}$ of the QoIs are constituted of the three dofs of the 561 observed nodes on surface Γ_{obs}; we then have $n_q = 1\,683 = 3 \times 561$. There are 8 integration points in each finite element. Consequently, there

are $N_i = 20\,000$ integration points $\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^{N_i}$. The $\mathbb{R}^{n_w}$-valued random variable $\mathbb{W}$ is generated as follows. Let us consider the set $\{[\mathbf{L}(\boldsymbol{\zeta}^1)], \ldots, [\mathbf{L}(\boldsymbol{\zeta}^{N_i})]\}$ of the random values at the N_i integration points of the random field $\{[\mathbf{L}(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$ with values in the set of all the upper triangular (6×6) matrices with positive-valued diagonal entries. The random matrix $[\mathbf{L}(\boldsymbol{\zeta})]$ is the factor in the Cholesky factorization $[\mathbf{G}_0(\boldsymbol{\zeta})] = [\mathbf{L}(\boldsymbol{\zeta})]^T\,[\mathbf{L}(\boldsymbol{\zeta})]$ given by Eq. (C.1) (from Appendix C) of random matrix $[\mathbf{G}_0(\boldsymbol{\zeta})]$ used in Eq. (121). The components of random vector $\mathbb{W}$ are constituted, for $i = 1, \ldots N_i$ and for $1 \leq j \leq k \leq 6$, of the family of random variables $\log\{[\mathbf{L}(\boldsymbol{\zeta}^i)]_{jj}\}$ and $[\mathbf{L}(\boldsymbol{\zeta}^i)]_{jk}$ for $j < k$. Since there are 21 entries in the upper triangular random matrix $[\mathbf{L}(\boldsymbol{\zeta}^i)]$, we have $n_w = 420\,000 = 20\,000{\times}21$.

(v) Generation of the initial dataset. The finite element discretization of the stochastic boundary value problem allows for constructing $\mathbb{Q}$ as a function of $\mathbb{W}$ and $\mathbb{U}$ that is formally written as

$$\mathbb{Q} = \mathbb{f}(\mathbb{W}, \mathbb{U})\,, \tag{124}$$

in which $(\mathbb{w}, \mathbb{u}) \mapsto \mathbb{f}(\mathbb{w}, \mathbb{u})$ is a measurable mapping on $\mathbb{R}^{n_w} \times \mathbb{R}^3$ with values in $\mathbb{R}^{n_q}$. Note that the deterministic mapping $\mathbb{f}$ cannot explicitly be described because this mapping is associated with the solution of the finite element approximation of the stochastic boundary value problem. The initial dataset $\mathbb{D}_N = \{\mathbb{x}^j = (\mathbb{q}^j, \mathbb{w}^j),\, j = 1, \ldots, N\}$, relative to the random variable $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, is then constructed by using the Monte Carlo simulation method for solving the discretized stochastic equation, in which $\mathbb{q}^j = \mathbb{f}(\mathbb{w}^j, \mathbb{u}^j)$ where $\{\mathbb{w}^j,\, j = 1, \ldots, N\}$ and $\{\mathbb{u}^j,\, j = 1, \ldots, N\}$ are N independent realizations of random vectors $\mathbb{W}$ and $\mathbb{U}$ (note that $\mathbb{u}^j$ is also an independent realization of $\mathbb{w}^j$). The generation is carried out with $N = 200$, $L_1 = L_2 = L_3 = 0.2$, and $\delta_{G_0} = 0.25$. For illustration, Fig. 5 displays the 2D plot in the plane $\zeta_3 = 0.0958\,m$ of one realization of the random fields $(\zeta_1, \zeta_2) \mapsto \log\{[\mathbf{L}(\zeta_1, \zeta_2, 0.0958)]_{jj}\}$ for $j = 1, 2, 4$, and $(\zeta_1, \zeta_2) \mapsto [\mathbf{L}((\zeta_1, \zeta_2, 0.0958))]_{12}$.

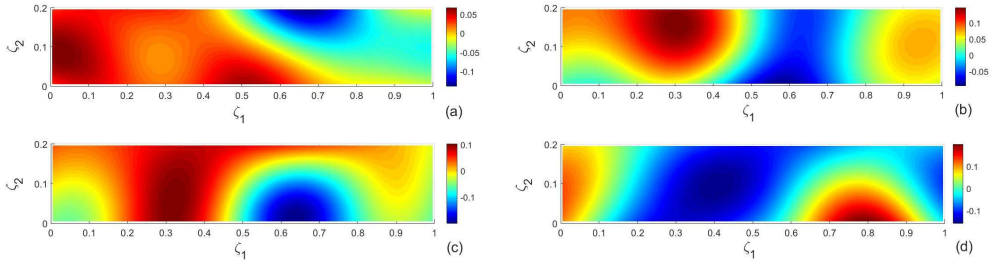

Figure 5: Application AP2: 2D plot in the plane $\zeta_3 = 0.0958\,m$ of one realization of the random field $(\zeta_1, \zeta_2) \mapsto \log\{[\mathbf{L}(\zeta_1, \zeta_2, 0.0958)]_{jj}\}$ for $j = 1$ (a), $j = 2$ (b), $j = 4$ (c), and of the random field $(\zeta_1, \zeta_2) \mapsto [\mathbf{L}((\zeta_1, \zeta_2, 0.0958))]_{12}$ (d).

7.2. Generation of the experimental dataset

As explained in Section 6 devoted to Application (AP1), simulated experiments are not required for the proposed methodology. Only some statistics that correspond to the constraints must be specified. Nevertheless, similarly to Application (AP1), in order to generate coherent statistics for the constraints, we introduce an experimental dataset generated as a perturbation of the stochastic boundary value problem (as previously, we insist on the fact that for applying the method proposed, the experimental dataset is not used and consequently, is not required). The

experimental dataset $\mathbb{D}_{n_r}^{\text{exp}}$ is generated with $n_r = 1\,000$ independent experimental realizations $\{\mathbb{q}^{\text{exp},r}, r = 1, \dots n_r\}$ of $\mathbb{Q}^{\text{exp}} = (\mathbb{Q}_1^{\text{exp}}, \dots, \mathbb{Q}_{n_q}^{\text{exp}})$ such that

$$\mathbb{Q}^{\text{exp}} = \mathbb{f}(\mathbb{W}^{\text{exp}}, \mathbb{U}^{\text{exp}}), \tag{125}$$

in which the deterministic mapping $\mathbb{f}$ is the same as the one used in Eq. (124). The $\mathbb{R}^3$-valued random variable $\mathbb{U}^{\text{exp}}$ is an independent copy of the $\mathbb{R}^3$-valued random variable $\mathbb{U}$ introduced in Section 7.1-(ii). The random vector $\mathbb{W}^{\text{exp}}$ is constructed as an independent copy of random vector $\mathbb{W}$ defines in in Section 7.1-(iv) for which the spatial correlation lengths are $L_1^{\text{exp}} = L_2^{\text{exp}} = L_3^{\text{exp}} = 0.16$ and the dispersion parameter is $\delta_{G_0}^{\text{exp}} = 0.45$.

7.3. Scaling

Using the block notation $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, the scaling of the $\mathbb{R}^{n_q}$-valued random variable $\mathbb{Q}$ and the scaling of the $\mathbb{R}^{n_w}$-valued random variable $\mathbb{W}$ is carried out as explained in Section 6.3 and is obtained by using Eqs. (107) and (108) whose values of the parameterized transformation are those of the present application (AP2). The scaling of Eq. (124) is then written as

$$\mathbf{Q} = \mathbf{f}(\mathbf{W}, \mathbb{U}), \tag{126}$$

in which $\mathbf{f}(\mathbf{W}, \mathbb{U}) = [\alpha_q]^{-1}\,(\mathbb{f}([\alpha_w]\mathbf{W} + \mathbb{w}^{\min}, \mathbb{U}) - \mathbb{q}^{\min})$. The scaling of $\mathbb{Q}^{\text{exp}}$ is such that $\mathbf{Q}^{\text{exp}} = [\alpha_q]^{-1}(\mathbb{Q}^{\text{exp}} - \mathbb{q}^{\min})$, which shows that the n_r independent experimental realizations $\{\mathbf{q}^{\text{exp},r}, r = 1, \dots n_r\}$ of $\mathbf{Q}^{\text{exp}}$ are such that

$$\mathbf{q}^{\text{exp},r} = [\alpha_q]^{-1}(\mathbb{q}^{\text{exp},r} - \mathbb{q}^{\min}). \tag{127}$$

7.4. Constraints

The constraints are those that are defined by Eqs. (13) and (14) in which $\kappa_{\text{mom}} = n_q$ and where the experimental moments $\{(\mathbb{m}_\kappa^{\text{exp}}, \mathbb{s}_\kappa^{\text{exp}}), \kappa = 1, \dots, \kappa_{\text{mom}}\}$ are estimated using the n_r independent experimental realizations $\{\mathbf{q}^{\text{exp},r}, r = 1, \dots n_r\}$ defined by Eq. (127). Consequently, the number of constraints is $m_c = 3\,366$.

7.5. Values of the parameters of the algorithms

All the values of the parameters given in this paragraph are related to $N = 200$.

(i) The PCA of $\mathbf{X}$ (see Section 5.1) is performed with $\varepsilon = 10^{-3}$ in Eq. (17) and yields $\nu = 193$. The graph of the error function $\nu \mapsto \text{err}_{\text{PCA}}(\nu)$ for $\nu \in [1, 193]$, defined by Eq. (17), is shown in Fig. 6 (a).

(ii) The parameters of the Gaussian kernel-density estimation defined by Eq. (24) are $s_\nu = 0.9544$ and $\widehat{s}_\nu = 0.6913$.

(iii) For the diffusion-maps basis, the use of Eqs. (88) and (89) yields $\varepsilon_{\text{diff}}^{\text{opt}} = 800$ and the optimal value of m is $m^{\text{opt}} = 195$. The graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1, which corresponds to these optimal values, is shown in Fig. 6 (b).

(iv) For solving the reduced-order nonlinear ISDE defined in Section 5.8.3, the parameters related to the algorithm are those introduced in Appendix B, have been estimated using the procedure that is detailed in Section 4.7.2 in [37], and are the following: $f_0 = 1.5$, $n_{\text{MC}} = 250$, $\ell_0 = 10$, $M_0 = 10$, $\Delta t = 0.217192$, and then, $M = 50\,000$.

(v) The projection of the constraints introduced in Section 5.3 has been performed with $\tau_c = 0.004$ that yields $\nu_c = 14$ (see Eq. (39)). Figure 7 (a) displays the graph of the error function

$\iota \mapsto \mathrm{err}(\iota)$ that is defined by Eq. (59), which shows the convergence of the iteration algorithm for computing the Lagrange multiplier $\boldsymbol{\lambda}^{\mathrm{sol}}$. It can be seen that the convergence is fast, the error decreases exponentially as a function of the iteration number ι. Figure 7 (b) shows the graph of the components $k \mapsto \lambda_k^{\mathrm{sol}}$, for $k \in [1, \nu_c]$ with $\nu_c = 14$, of the optimal solution $\boldsymbol{\lambda}^{\mathrm{sol}} = (\lambda_1^{\mathrm{sol}}, \ldots, \lambda_{\nu_c}^{\mathrm{sol}})$ of the Lagrange multipliers $\boldsymbol{\lambda}$ computed by Eq. (54).

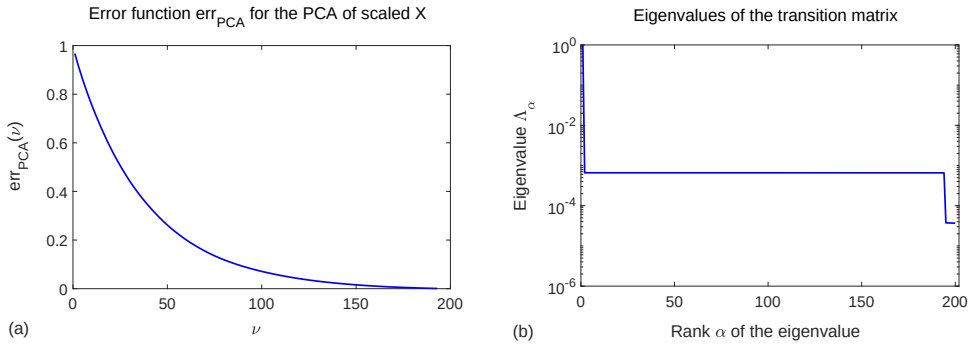

Figure 6: Application AP2. Figure (a): graph of PCA-error function $\nu \mapsto \mathrm{err}_{\mathrm{PCA}}(\nu)$ for $\nu \in [1, 193]$, defined by Eq. (17). Figure (b): graph $\alpha \mapsto \Lambda_\alpha$ of the eigenvalues of the transition matrix defined in Section 5.8.1 (the vertical axis is in log).

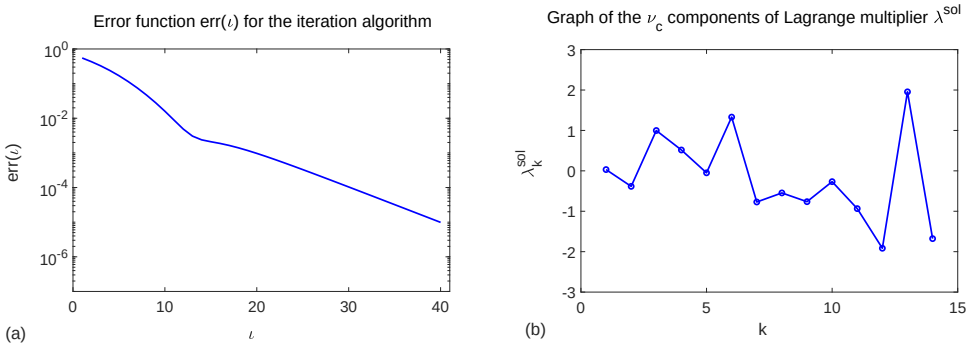

Figure 7: Application AP2. Figure (a): graph of the error function $\iota \mapsto \mathrm{err}(\iota)$ defined by Eq. (59) showing the convergence of the iteration algorithm for computing the Lagrange multiplier $\boldsymbol{\lambda}^{\mathrm{sol}}$ (the vertical axis is in log). Figure (b): graph of the components $k \mapsto \lambda_k^{\mathrm{sol}}$, for $k \in [1, \nu_c]$ with $\nu_c = 14$, of the optimal solution $\boldsymbol{\lambda}^{\mathrm{sol}} = (\lambda_1^{\mathrm{sol}}, \ldots, \lambda_{\nu_c}^{\mathrm{sol}})$ of the Lagrange multipliers $\boldsymbol{\lambda}$ computed by Eq. (54).

7.6. Convergence of the learning with respect to the value of N

Similarly to Application (AP1), the use of the criterion defined by Eq. (101) or by Eq. (102) for analyzing the convergence with respect to N (convergence of the learning with respect to dimension of the initial dataset) is not sufficiently informative for this application. In addition, taking into account the high dimensions $n_w = 420\,000$ and $n_q = 1\,683$, it is neither easy nor efficient to perform the analysis as carried out for Application (AP1). In this condition, we

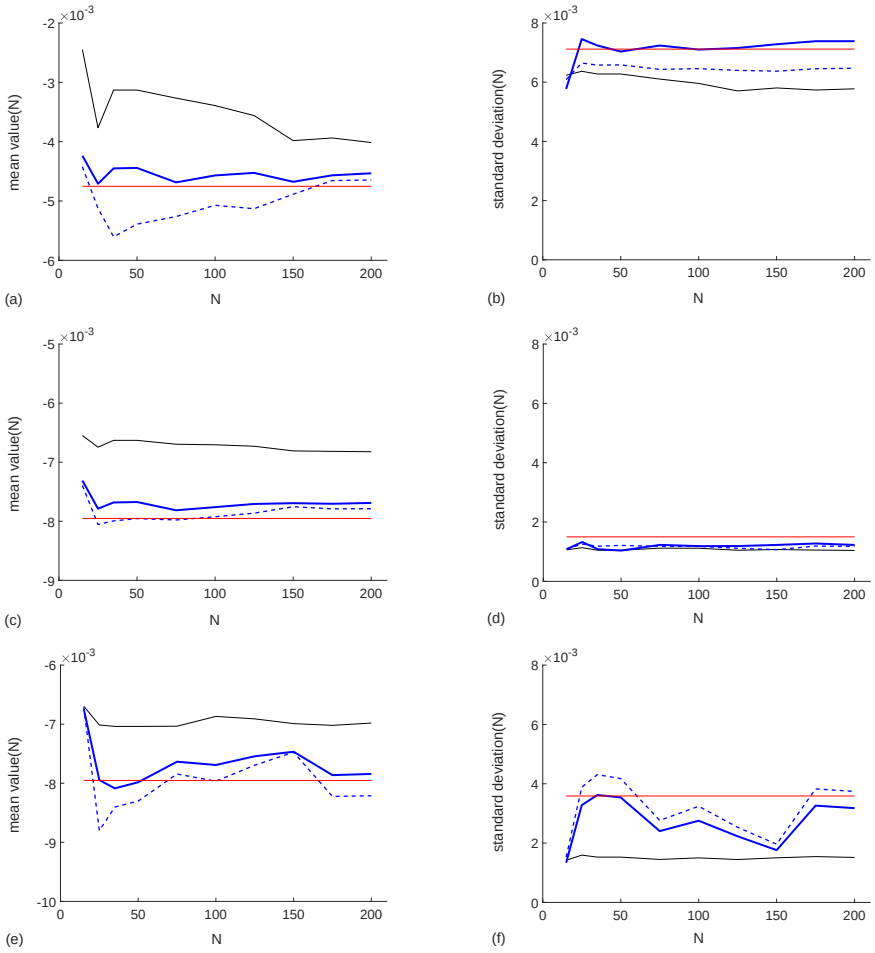

Figure 8: Application AP2: For component $k = 2$ (Figs. (a) and (b)), $k = 31$ (Figs. (c) and (d), and $k = 33$ (Figs. (e) and (f)), the left figures plot the graphs of the mean functions $N \mapsto \mathbb{m}_{\mathbb{Q}_k}(N)$ (black thin line), $N \mapsto \mathbb{m}_{\mathbb{Q}_k}^{\exp}$ (red horizontal line), $N \mapsto \mathbb{m}_{\mathbb{Q}_k^c}(N)$ (blue thick line), and $N \mapsto \mathbb{m}_{\mathbb{Q}_k^{c,\mathrm{model}}}(N)$ (blue dashed line); the right figures plot the graphs of the standard deviation of functions $N \mapsto \mathbb{s}_{\mathbb{Q}_k}(N)$ (black dashed line), $N \mapsto \mathbb{s}_{\mathbb{Q}_k}^{\exp}$ (red line), $N \mapsto \mathbb{s}_{\mathbb{Q}_k^c}(N)$ (blue line), and $N \mapsto \mathbb{s}_{\mathbb{Q}_k^{c,\mathrm{model}}}(N)$ (blue dashed line).

therefore propose to analyze the convergence of the mean value and the standard deviation of unscaled displacements $\mathbb{Q}_k$ of three selected nodes, for which the displacements are maximum for the mean responses: $\mathbb{Q}_2$ corresponding to the ζ_2-displacement of the node of coordinates $(0, 0, 0.1)$, $\mathbb{Q}_{31}$ corresponding to the ζ_1-displacement of the node of coordinates $(0, 0.2, 0.1)$, and $\mathbb{Q}_{33}$ corresponding to the ζ_3-displacement of node of coordinates $(0, 0.2, 0.1)$. The computation has been carried out with $M = 50\,000$, $n_r = 1\,000$, and for ten values of integer N such that $15 \leq N \leq 200$.

For $k \in \{2, 31, 33\}$, Figure 8 allows for analyzing the convergence with respect to N of the mean value $\mathbb{m}(N)$ and of the standard deviation $\mathbb{s}(N)$ of the following random variables: $\mathbb{Q}_k$

Table 1: Applications (AP2): For $N = 200$ and $M = 50\,000$, and for $k = 2, 31, 33$, values of the mean $\mathbb{m}_{\mathbb{Q}_k}$, $\mathbb{m}_{\mathbb{Q}_k^{\exp}}$, $\mathbb{m}_{\mathbb{Q}_k^{c}}$, $\mathbb{m}_{\mathbb{Q}_k^{c,\mathrm{model}}}$, and values of the standard deviations $\mathbb{s}_{\mathbb{Q}_k}$, $\mathbb{s}_{\mathbb{Q}_k^{\exp}}$, $\mathbb{s}_{\mathbb{Q}_k^{c}}$, $\mathbb{s}_{\mathbb{Q}_k^{c,\mathrm{model}}}$.

Millimeter ($10^{-3}\,m$)	**Component** 2	**Component** 31	**Component** 33
$\mathbb{m}_{\mathbb{Q}_k}$	−4.01	−6.82	−6.98
$\mathbb{m}_{\mathbb{Q}_k^{c}}$	−4.53	−7.69	−7.84
$\mathbb{m}_{\mathbb{Q}_k^{c,\mathrm{model}}}$	−4.65	−7.79	−8.21
$\mathbb{m}_{\mathbb{Q}_k^{\exp}}$	−4.75	−7.95	−7.95
$\mathbb{s}_{\mathbb{Q}_k}$	5.78	1.04	1.52
$\mathbb{s}_{\mathbb{Q}_k^{c}}$	7.38	1.23	3.18
$\mathbb{s}_{\mathbb{Q}_k^{c,\mathrm{model}}}$	6.47	1.17	3.74
$\mathbb{s}_{\mathbb{Q}_k^{\exp}}$	7.12	1.50	3.59

(estimated with the N points of the initial dataset), $\mathbb{Q}_k^c$ (estimated with the M additional realizations of the generated dataset computed with the probabilistic learning on manifolds under the $m_c = 2\,366$ constraints), $\mathbb{Q}_k^{\exp}$ (corresponding to the experimental dataset), and finally, for the quality assessment, $\mathbb{Q}_k^{c,\mathrm{model}}$ that results from the transformation of $\mathbb{W}^c$ using the stochastic computational model, that is to say, $\mathbb{Q}^{c,\mathrm{model}} = \mathbb{f}(\mathbb{W}^c, \mathbb{U})$. It can be seen that the effects of the constraints are significant: the thin lines (without constraints) are different from the thick lines (with constraints). The convergence with respect to N is good enough even for the quality assessment: the thick lines (with constraints) are close to the dashed lines (QoI computed with the stochastic model using $\mathbb{W}^c$ learned with the constraints).

7.7. Results obtained for $N = 200$ and $M = 50\,000$

Similarly to Section 7.6, we consider, for $k = 2, 31, 33$, $\mathbb{Q}_k$ defined by the initial dataset, $\mathbb{Q}_k^{\exp}$ associated with the experiments, $\mathbb{Q}_k^c$ constructed using the constraints, and $\mathbb{Q}_k^{c,\mathrm{model}}$ transformed from $\mathbb{W}^c$ using the stochastic computational model.

(i) For $N = 200$ and $M = 50\,000$, for components $k = 2, 31, 33$, Table 1 gives the mean values $\mathbb{m}_{\mathbb{Q}_k}$, $\mathbb{m}_{\mathbb{Q}_k^{\exp}}$, $\mathbb{m}_{\mathbb{Q}_k^{c}}$, $\mathbb{m}_{\mathbb{Q}_k^{c,\mathrm{model}}}$ and the standard deviations $\mathbb{s}_{\mathbb{Q}_k}$, $\mathbb{s}_{\mathbb{Q}_k^{\exp}}$, $\mathbb{s}_{\mathbb{Q}_k^{c}}$, $\mathbb{s}_{\mathbb{Q}_k^{c,\mathrm{model}}}$. This table allows comparisons to be carried out. As already pointed out in Section 7.6, it can be seen that the constraints play an important role because the values relative to $\mathbb{Q}_k$ (without the constraints) are very different from the corresponding values relative to $\mathbb{Q}_k^c$ (with the constraints). The predictions obtained under the constraints are good enough when comparing the values of $\mathbb{Q}_k^c$ (with the constraints) with the corresponding target values for $\mathbb{Q}_k^{\exp}$. Finally, the quality assessment is good taking into account the high dimension of the problem and the constraints that are applied, which are only relative to a subset of all the dofs and related to second-order moments. This quality can be analyzed by comparing the values relative to $\mathbb{Q}_k^c$ (with the constraints) with the values relative to $\mathbb{Q}_k^{c,\mathrm{model}}$.

(ii) To complete the analysis and validation, we present the results obtained for the probability density functions for components $k = 2$ (Fig. 9-a), $k = 31$ (Fig. 9-b), and $k = 33$ (Fig. 9-c). These figures display the graphs of the probability density functions of $\mathbb{Q}_k$, $\mathbb{Q}_k^{\exp}$, $\mathbb{Q}_k^c$, and $\mathbb{Q}_k^{c,\mathrm{model}}$. The

results are very good and even a little unexpected since only the mean and the standard deviation are imposed for a subset of dofs of the stochastic computational model. The pdf of $\mathbb{Q}_k^c$ is close to the pdf of $\mathbb{Q}_k^{c,\text{model}}$, which is itself close to the pdf of $\mathbb{Q}_k^{\text{exp}}$ (this experimental pdf is not used in the methodology, is not required, and is not available in the framework of the method proposed).

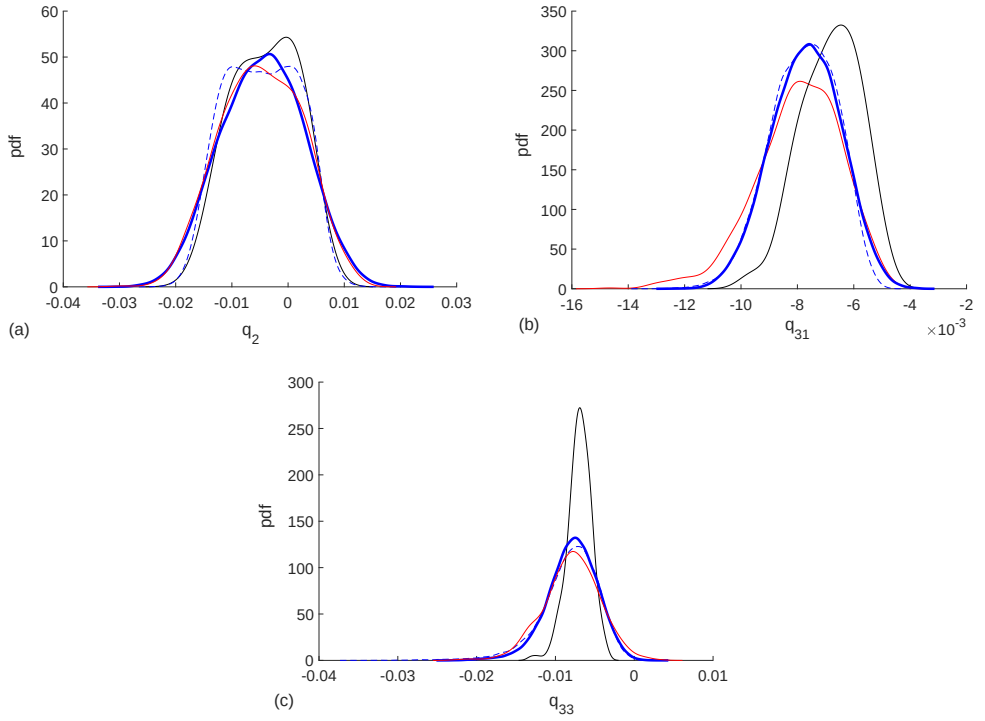

Figure 9: Application AP2: For $M = 50\,000$ and $N = 200$, the three figures show, for components $k = 2$ (a), $k = 31$ (b), and $k = 33$ (c), the pdf of $\mathbb{Q}_k$ (black thin line), $\mathbb{Q}_k^{\text{exp}}$ (red line), $\mathbb{Q}_k^c$ (blue thick line), and $\mathbb{Q}_k^{c,\text{model}}$ (blue dashed line).

(iii) In order to illustrate how the constraints qualitatively impact the concentration of the probability measure of random vector $\mathbb{X} = (\mathbb{Q}, \mathbb{W})$, Fig. 10-left shows the 50 000 additional realizations computed with the probabilistic learning on manifolds for random vector $(\mathbb{W}_{21\,000}, \mathbb{W}_{121\,000}, \mathbb{Q}_{33})$ (without the constraints) and $(\mathbb{W}_{21\,000}, \mathbb{W}_{121\,000}, \mathbb{Q}_{33}^c)$ (with the constraints), while Fig. 10-right shows the results for $(\mathbb{W}_{121\,000}, \mathbb{W}_{321\,000}, \mathbb{Q}_{33})$ (without the constraints) and $(\mathbb{W}_{121\,000}, \mathbb{W}_{321\,000}, \mathbb{Q}_{33}^c)$ (with the constraints). It should be noted that the two submanifolds that are considered are not deterministic due to the presence of the random vector $\mathbb{U}$ and to the other components of $\mathbb{W}$, which are not used in the chosen representation. The relatively large dispersion that seems to be obtained is only an appearance that is related to the choice of the scale that has been taken for $\mathbb{Q}_{33}$. On the other hand, according to Table 1, $\mathbb{Q}_{33}^c$ is more scattered than $\mathbb{Q}_{33}$ due to the imposed constraints, so the concentration is less for $\mathbb{Q}_{33}^c$ than for $\mathbb{Q}_{33}$, which is normal.

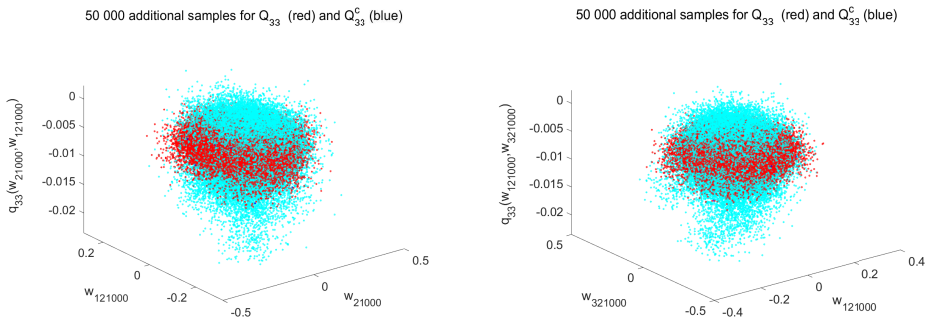

Figure 10: Application AP2. Left figure: 50 000 realizations of $(\mathbb{W}_{21\,000}, \mathbb{W}_{121\,000}, \mathbb{Q}_{33})$ (red symbols) and of $(\mathbb{W}_{21\,000}, \mathbb{W}_{121\,000}, \mathbb{Q}^c_{33})$ (blue symbols). Right figure: 50 000 realizations of $(\mathbb{W}_{121\,000}, \mathbb{W}_{321\,000}, \mathbb{Q}_{33})$ (red symbols) and of $(\mathbb{W}_{121\,000}, \mathbb{W}_{321\,000}, \mathbb{Q}^c_{33})$ (blue symbols).

8. Conclusion

In this paper, we have presented a methodology that extends the probabilistic leaning on manifolds from a small dataset to the case for which constraints are imposed, during the learning process, to a subset of quantities of interest. These constraints are, for instance, experimental second-order statistical moments (mean and standard deviation) that are given. The method has the capability to consider more general constraints than the second-order statistical moments. The proposed method allows for analyzing non-Gaussian cases in high dimension related to functional inputs and outputs. The iteration algorithm presented is very robust and seems to be exponentially convergent with respect to the number of iterations. Two applications have been presented, which allow for analyzing the behavior of the method and for validating it.

ACKNOWLEDGEMENTS

Part of this research was supported by the PIRATE project funded under DARPA's AIRA program.

Appendix A. Examples of constraints

In this Appendix, two examples of constraints are given in addition to the those given in Section 4.

Appendix A.1. Learning from a training dataset under constraints defined by experimental second-order moments given for components of $\boldsymbol{Q}^c$ and by the mean value of a physical constaint

This example corresponds to the concatenation of the two examples presented in Section 4.1 and in 4.2. Consequently, using the notations introduced in these two examples, $m_c = 2\,\kappa_{\text{mom}} + n_q$, function $\mathbf{x} \mapsto \mathbf{g}^c(\mathbf{x}) = (g^c_1(\mathbf{x}), \ldots, g^c_{m_c}(\mathbf{x}))$ from $\mathbb{R}^n$ into $\mathbb{R}^{m_c}$, and vector $\boldsymbol{\beta}^c = (\beta^c_1, \ldots, \beta^c_{m_c}) \in \mathbb{R}^{m_c}$, which define the constraints (see Eq. (10)), are such that, for $\kappa = 1, \ldots, \kappa_{\text{mom}}$ and for $k = 1, \ldots, n_q$,

$$g^c_\kappa(\mathbf{X}^c) = Q^c_{k_\kappa} \quad , \quad \beta^c_\kappa = \mathbb{m}^{\text{exp}}_\kappa \, , \tag{A.1}$$

$$g^c_{\kappa+\kappa_{\rm mom}}(\mathbf{X}^c) = (Q^c_{k_\kappa})^2 \quad , \quad \beta^c_{\kappa+\kappa_{\rm mom}} = \mathbb{r}^{\rm exp}_\kappa \,, \tag{A.2}$$

$$g^c_{k+2\,\kappa_{\rm mom}}(\mathbf{X}^c) = Q^c_k - ([\underline{B}]\,\mathbf{W}^c)_k \quad , \quad \beta^c_{k+2\,\kappa_{\rm mom}} = \beta^{\rm const}_k \,. \tag{A.3}$$

Appendix A.2. Probability density functions given for components of $\boldsymbol{Q}^c$

Let $\mathcal{J}_{\rm pdf} = \{k_1, \ldots, k_{\kappa_{\rm pdf}}\}$ be $\kappa_{\rm pdf} \leq n_q$ integers such that $\mathcal{J}_{\rm pdf} \subseteq \{1, \ldots, n_q\}$. For $\kappa \in \{1, \ldots, \kappa_{\rm pdf}\}$, let $q \mapsto p^{\rm exp}_\kappa(q)$ be the given experimental pdf assigned to component $Q^c_{k_\kappa}$ of random vector $\mathbf{Q}^c$, which is assumed to be defined by its values $\{p^{\rm exp}_\kappa(q_{\kappa i}), i = 1, \ldots, m_\kappa\}$ that are given in m_κ points $q_{\kappa 1}, \ldots, q_{\kappa m_\kappa}$ in $\mathbb{R}$. We then have to take into account the $m_c = \sum_{\kappa=1}^{\kappa_{\rm pdf}} m_\kappa$ constraints that could be written, for $\kappa = 1, \ldots, \kappa_{\rm pdf}$, as

$$p_{X^c_{k_\kappa}}(q_{\kappa i}) = p^{\rm exp}_\kappa(q_{\kappa i}) \quad , \quad i = 1, \ldots, m_\kappa \,. \tag{A.4}$$

Since $p_{X^c_{k_\kappa}}(q_{\kappa i}) = E\{\delta_0(Q^c_{k_\kappa} - q_{\kappa i}\}$, the Dirac δ_0 can be regularized and Eq. (A.4) is replaced by the following approximation,

$$E\{\widetilde{g}^c_{\kappa i}(\mathbf{X}^c)\} = \widetilde{\beta}^c_{\kappa i} \quad , \quad i = 1, \ldots, m_\kappa \,, \tag{A.5}$$

with

$$\widetilde{g}^c_{\kappa i}(\mathbf{X}^c) = \frac{1}{\sqrt{2\pi}\,\varepsilon_\kappa} \exp\{-\frac{1}{2\varepsilon^2_\kappa}(Q^c_{k_\kappa} - q_{\kappa i})^2\} \quad , \quad \widetilde{\beta}^c_{\kappa i} = p^{\rm exp}_\kappa(q_{\kappa i}) \,, \tag{A.6}$$

in which an estimation of the hyperparameter ε_κ is detailed below. Consequently, function $\mathbf{g}^c$ and vector $\boldsymbol{\beta}^c$ is obtained by gathering $\widetilde{g}^c_{\kappa i}$ and $\widetilde{\beta}^c_{\kappa i}$ for all the κ in $\{1, \ldots, \kappa_{\rm pdf}\}$ and for all i in $\{1, \ldots, m_\kappa\}$,

$$\mathbf{g}^c(\mathbf{X}^c) = \{\widetilde{g}^c_{\kappa i}(\mathbf{X}^c)\}_{\kappa i} \quad , \quad \boldsymbol{\beta}^c = \{\widetilde{\beta}^c_{\kappa i}\}_{\kappa i} \quad , \quad m_c = \sum_{\kappa=1}^{\kappa_{\rm pdf}} m_\kappa \,. \tag{A.7}$$

In Eq. (A.6), the hyperparameter ε_κ can be estimated as follows. Since $p_{Q^c_{k_\kappa}}(q_{\kappa i}) = p^{\rm exp}_\kappa(q_{\kappa i})$ is the value of the closest pdf to $p_{Q_{k_\kappa}}$, ε_κ can be estimated using the Gaussian kernel-density estimation method of $p_{Q_{k_\kappa}}(q_{\kappa i})$ constructed with the initial dataset $D_N = \{(\mathbf{q}^j, \mathbf{w}^j), j = 1, \ldots, N\}$ in which the component k_κ of $\mathbf{q}^j$ is $q^j_{k_\kappa}$. We then have,

$$p_{Q_{k_\kappa}}(q_{\kappa i}) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}\,s_s\widehat{\sigma}_\kappa} \exp\{-\frac{1}{2s^2_s\widehat{\sigma}^2_\kappa}(q^j_{k_\kappa} - q_{\kappa i})^2\} \,, \tag{A.8}$$

in which $s_s = (4/(3N))^{1/5}$ is the Sylverman bandwidth and where $\widehat{\sigma}_\kappa$ is the estimate of the standard deviation of Q_{k_κ} performed with D_N. For N sufficiently large, the right-hand side of Eq. (A.8) is

$$E\{\frac{1}{\sqrt{2\pi}\,s_s\widehat{\sigma}_\kappa} \exp\{-\frac{1}{2s^2_s\widehat{\sigma}^2_\kappa}(Q^c_{k_\kappa} - q_{\kappa i})^2\} \,,$$

that compares to Eqs. (A.5) and (A.6) yields $\varepsilon_\kappa = s_s\widehat{\sigma}_\kappa$.

Appendix B. Störmer-Verlet algorithm for solving the reduced-order nonlinear ISDE defined by Eqs. (94) to (98)

The number M of realizations is reparameterized as

$$M = n_{\mathrm{MC}} \times N \,, \tag{B.1}$$

in which n_{MC} is the integer that has been introduced in Section 5.7.5. The reduced-order ISDE defined by Eqs. (94) to (98) is solved for $t \in [0, t_{\max}]$ with $t_{\max} = (\ell_0 + n_{\mathrm{MC}} \times M_0)\, \Delta t$ in which Δt is the sampling step, where ℓ_0 is chosen in order that the solution of Eqs. (94) to (96) has reached the stationary regime, and where M_0 allows for controlling the number of sampling steps between two consecutive realizations for obtaining a quasi-independence of the computed realizations For $\ell = 0, 1, \ldots, n_{\mathrm{MC}} \times M_0$, we consider the sampling points $t_\ell = \ell\, \Delta t$ and the following notations: $[\boldsymbol{\mathcal{Z}}_\ell] = [\boldsymbol{\mathcal{Z}}_{\lambda'}(t_\ell)]$, $[\boldsymbol{\mathcal{Y}}_\ell] = [\boldsymbol{\mathcal{Y}}_{\lambda'}(t_\ell)]$, and $[\boldsymbol{\mathcal{W}}_\ell^{\mathrm{wien}}] = [\boldsymbol{\mathcal{W}}_{\lambda'}^{\mathrm{wien}}(t_\ell)]$. The Störmer-Verlet scheme is used for solving the reduced-order ISDE, which is written, for $\ell = 0, 1, \ldots, n_{\mathrm{MC}} \times M_0$, as

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

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

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

with the initial condition defined by Eq. (96), where $\beta = f_0\, \Delta t\, /4$, and where $[\boldsymbol{\mathcal{L}}_{\ell+\frac{1}{2}}]$ is the $\mathbb{M}_{\nu,m}$-valued random variable such that

$$[\boldsymbol{\mathcal{L}}_{\ell+\frac{1}{2}}] = [\mathcal{L}_{\lambda'}([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}])] = [L_{\lambda'}([\boldsymbol{\mathcal{Z}}_{\ell+\frac{1}{2}}]\, [g]^T)]\, [a] \,.$$

In the above equation, $[\Delta \boldsymbol{\mathcal{W}}_{\ell+1}^{\mathrm{wien}}] = [\Delta \mathbf{W}_{\ell+1}^{\mathrm{wien}}]\, [a]$ is a random variable with values in $\mathbb{M}_{\nu,m}$, in which the increment $[\Delta \mathbf{W}_{\ell+1}^{\mathrm{wien}}] = [\mathbf{W}_{\lambda'}^{\mathrm{wien}}(t_{\ell+1})] - [\mathbf{W}_{\lambda'}^{\mathrm{wien}}(t_\ell)]$. The increments are statistically independent. For all $\alpha = 1, \ldots, \nu$ and for all $j = 1, \ldots, N$, the real-valued random variables $\{[\Delta \mathbf{W}_{\ell+1}^{\mathrm{wien}}]_{\alpha j}\}_{\alpha j}$ are independent, Gaussian, second-order, and centered random variables such that

$$E\{[\Delta \mathbf{W}_{\ell+1}^{\mathrm{wien}}]_{\alpha j} [\Delta \mathbf{W}_{\ell+1}^{\mathrm{wien}}]_{\alpha' j'}\} = \Delta t\, \delta_{\alpha\alpha'}\, \delta_{jj'} \,.$$

Appendix C. Construction of the non-Gaussian random field $[\mathbf{G}_0]$ and its generator of realizations

In this Appendix, we summarize the construction of the non-Gaussian random field $[\mathbf{G}_0]$ and its generator of realizations (whose details can be found in [56, 49]). This construction starts with the introduction of a Gaussian random field.

Appendix C.1. Gaussian random field U allowing the stochastic germ of random field $[\boldsymbol{G}_0]$ to be constructed

Let $\{U(\boldsymbol{\zeta}), \boldsymbol{\zeta} \in \mathbb{R}^3\}$ be the real-valued second-order, centered, homogeneous, Gaussian, and normalized random field, defined on probability space $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by $\mathbb{R}^3$. For all $\boldsymbol{\zeta}$ in $\mathbb{R}^3$, we then have,

$$E\{U(\boldsymbol{\zeta})\} = 0 \quad , \quad E\{U(\boldsymbol{\zeta})^2\} = 1 \,.$$

The random field U is then completely and uniquely defined by the its autocorrelation function

$$\boldsymbol{\chi} = (\chi_1, \chi_2, \chi_3) \mapsto R_U(\boldsymbol{\chi}) = E\{U(\boldsymbol{\zeta} + \boldsymbol{\chi})\, U(\boldsymbol{\zeta})\}\,,$$

from $\mathbb{R}^3$ into $\mathbb{R}$, such that $R_U(0) = 1$. The spatial-correlation lengths of U, denoted as L_1, L_2, L_3, are such that

$$L_k = \int_0^{+\infty} |R_U(\boldsymbol{\chi}^k)|\, d\chi_k \quad , \quad k = 1, 2, 3\,,$$

in which $\boldsymbol{\chi}^1 = (\chi_1, 0, 0)$, $\boldsymbol{\chi}^2 = (0, \chi_2, 0)$, and $\boldsymbol{\chi}^3 = (0, 0, \chi_3)$. The autocorrelation function is written as

$$R_U(\boldsymbol{\chi}) = \rho_1(\chi_1) \times \rho_2(\chi_2) \times \rho_3(\chi_3)\,,$$
$$\rho_k(\chi_k) = \{4L_k^2/(\pi^2\chi_k^2)\}\ \sin^2(\pi\chi_k/(2L_k)) \quad , \quad k = 1, 2, 3\,.$$

Random field U is mean-square continuous on $\mathbb{R}^3$ and its power spectral density function defined on $\mathbb{R}^3$ has a compact support that is written as,

$$[-\pi/L_1\,, \pi/L_1] \times [-\pi/L_2\,, \pi/L_2] \times [-\pi/L_3\,, \pi/L_3]\,.$$

Such a model allows for sampling random field U as a function of the spatial-correlation lengths by using the Shannon theorem and has only 3 real parameters. The details concerning the generation of realizations of random field U can be found in [56]. One possible method is based on the usual numerical simulation of homogeneous Gaussian vector-valued random field constructed with the stochastic integral representation of homogeneous stochastic fields (see [57]).

Appendix C.2. Defining the random field $[\boldsymbol{G}_0]$ *and its generator of realizations*

Let δ_{G_0} be such that $0 < \delta_{G_0} < \sqrt{7/11}$. For $1 \le j \le k \le 6$, let $\{U_{jk}(\boldsymbol{\zeta}), \boldsymbol{\zeta} \in \mathbb{R}^3\}_{jk}$ be 21 independent copies of Gaussian random field $\{U(\boldsymbol{\zeta}), \boldsymbol{\zeta} \in \mathbb{R}^3\}$ that is defined in Appendix C.1. The non-Gaussian random field $\{[\mathbf{G}_0(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$ is then constructed as follows:

- For all $\boldsymbol{\zeta}$ fixed in $\mathbb{R}^3$, the random matrix $[\mathbf{G}_0(\boldsymbol{\zeta})]$ is written as

$$[\mathbf{G}_0(\boldsymbol{\zeta})] = [\mathbf{L}(\boldsymbol{\zeta})]^T\, [\mathbf{L}(\boldsymbol{\zeta})]\,, \tag{C.1}$$

  in which $[\mathbf{L}(\boldsymbol{\zeta})]$ is an upper (6×6) real triangular random matrix.

- For $1 \le j \le k \le 6$, the random fields $\{[\mathbf{L}(\boldsymbol{\zeta})]_{jk}, \boldsymbol{\zeta} \in \mathbb{R}^3\}$ are independent.

- For $j < k$, the real-valued random field $\{[\mathbf{L}(\boldsymbol{\zeta})]_{jk}, \boldsymbol{\zeta} \in \mathbb{R}^3\}$ is defined by $[\mathbf{L}(\boldsymbol{\zeta})]_{jk} = \sigma_6\, U_{jk}(\boldsymbol{\zeta})$ in which σ_6 is such that $\sigma_6 = \delta_{G_0}/\sqrt{6+1}$.

- For $j = k$, the positive-valued random field $\{[\mathbf{L}(\boldsymbol{\zeta})]_{jj}, \boldsymbol{\zeta} \in \mathbb{R}^3\}$ is defined by $[\mathbf{L}(\boldsymbol{\zeta})]_{jj} = \sigma_6\, \sqrt{2\, h(U_{jj}(\boldsymbol{\zeta}), a_j)}$ in which $a_j = (6+1)/(2\delta_{G_0}^2) + (1-j)/2$. Function $u \mapsto h(\alpha, u)$ is such that $\Gamma_\alpha = h(\alpha, \mathcal{U})$ is a gamma random variable with parameter α ($\mathcal{U}$ being a normalized Gaussian random variable).

References

[1] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer-Verlag, New York, 2005.

[2] 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.

[3] B. P. Carlin, T. A. Louis, Bayesian Methods for Data Analysis, Third Edition, Chapman & Hall / CRC Press, Boca Raton, 2009.

[4] A. Stuart, Inverse problems: A Bayesian perspective, Acta Numerica 19 (2010) 451–559. doi:10.1017/S0962492910000061.

[5] 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.

[6] J. Marin, P. Pudlo, C. Robert, R. Ryder, Approximate bayesian computational methods, Statistics and Computing 22 (6) (2012) 1167–1180. doi:10.1007/s11222-011-9288-2.

[7] P. Marjoram, J. Molitor, V. Plagnol, S. Tavare, Markov chain monte carlo without likelihoods, Proceedings of the National Academy of Sciences of the United States of America 100 (26) (2003) 15324–15328. doi:10.1073/pnas.0306899100.

[8] E. T. Jaynes, Information theory and statistical mechanics, Physical Review 106 (4) (1957) 620–630.

[9] A. Bhattacharyya, On the measures of divergence between two statistical populations defined by their probability distributions, Bulletin of the Calcultta Mathematical Society 35 (1943) 99–109.

[10] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[11] J. N. Kapur, H. K. Kesavan, Entropy Optimization Principles with Applications, Academic Press, San Diego, 1992.

[12] H. Gish, A probabilistic approach to the understanding and training of neural network classifiers, Proceedings of the IEEE International Conference on Acoustics, speech, and signal processing (1990) 1361–1364.

[13] H. Shatkay, L. P. Kaelbling, Learning topological maps with weak local odometric information, Proceedings of IJCAI-97, INCIJCAI (2) (1997) 920–929.

[14] D. Barber, C. M. Bishop, Ensemble learning for multi-layer networks, Proceedings of Advances in neural information processing systems (1998) 395–401.

[15] J. Hollmén, V. Tresp, O. Simula, A learning vector quantization algorithm for probabilistic models, Proceedings of the 10th European Signal Processing Conference (2000) 1–4.

[16] A. Galata, N. Johnson, D. Hogg, Learning variable-length markov models of behavior, Computer Vision and Image Understanding 81 (3) (2001) 398–413. doi:10.1006/cviu.2000.0894.

[17] B. Bigi, Using kullback-leibler distance for text categorization, Proceedings of the European Conference on information retrieval (2003) 305–319.

[18] S. Kaski, J. Sinkkonen, Principle of learning metrics for exploratory data analysis, Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology 37 (2-3) (2004) 177–188.

[19] T. Lange, M. H. Law, A. K. Jain, J. M. Buhmann, Learning with constrained and unlabelled data, Proceedings of the Computer Society Conference on Computer vision and pattern recognition (CVPR'05) 1 (2005) 731–738.

[20] S. Peperkamp, R. Le Calvez, J.-P. Nadal, E. Dupoux, The acquisition of allophonic rules: Statistical learning with linguistic constraints, Cognition 101 (3) (2006) B31–B41. doi:10.1016/j.cognition.2005.10.006.

[21] S. Filippi, O. Cappé, A. Garivier, Optimism in reinforcement learning and kullback-leibler divergence, Proceedings of the 48th Annual Allerton IEEE Conference on Communication, control, and computing (2010) 115–122.

[22] N. Vasconcelos, P. Ho, P. Moreno, The kullback-leibler kernel as a framework for discriminant and localized representations for visual recognition, Proceedings of the European Conference on Computer Vision (2004) 430–441.

[23] W. Zhang, S. Shan, X. Chen, W. Gao, Local gabor binary patterns based on kullback–leibler divergence for partially occluded face recognition, IEEE Signal Processing Letters 14 (11) (2007) 875–878. doi:10.1109/LSP.2007.903260.

[24] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, Proceedings of the IEEE Conference on Computer vision and pattern recognition (2012) 2288–2295.

[25] 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.

[26] Z.-Y. Ran, B.-G. Hu, Determining structural identifiability of parameter learning machines, Neurocomputing 127 (2014) 88–97. doi:10.1016/j.neucom.2013.08.039.

[27] M. Sun, Y. Li, J. F. Gemmeke, X. Zhang, Speech enhancement under low snr conditions via noise estimation using sparse and low-rank nmf with kullback–leibler divergence, IEEE Transactions on Audio, Speech, and Language Processing 23 (7) (2015) 1233–1242. doi:10.1109/TASLP.2015.2427520.

[28] S. Nowozin, B. Cseke, R. Tomioka, f-gan: Training generative neural samplers using variational divergence minimization, Proceedings of the Conference on Advances in neural information processing systems (2016) 271–279.

[29] G. A. Hanasusanto, V. Roitch, D. Kuhn, W. Wiesemann, Ambiguous joint chance constraints under mean and dispersion information, Operations Research 65 (3) (2017) 751–767. doi:doi.org/10.1287/opre.2016.1583.
[30] H. N. Najm, R. D. Berry, C. Safta, K. Sargsyan, B. J. Debusschere, Data-free inference of uncertain parameters in chemical models, International Journal for Uncertainty Quantification 4 (2) (2014) 111–132. doi:10.1615/Int.J.UncertaintyQuantification.2013005679.
[31] H. N. Najm, K. Chowdhary, Inference given summary statistics, in: R. Ghanem, D. Higdon, O. H. (Eds.), Handbook of Uncertainty Quantification, Springer, Cham, Switzerland, 2017, Ch. 3, pp. 33–67.
[32] B. Nguyen, C. Morell, B. De Baets, Supervised distance metric learning through maximization of the jeffrey divergence, Pattern Recognition 64 (2017) 215–225. doi:10.1016/j.patcog.2016.11.010.
[33] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, Physical Review E 96 (2) (2017) 022140. doi:10.1103/PhysRevE.96.022140.
[34] M. Fan, X. Zhang, L. Du, L. Chen, D. Tao, Semi-supervised learning through label propagation on geodesics, IEEE Transactions on Cybernetics 48 (5) (2018) 1486–1499. doi:10.1109/TCYB.2017.2703610.
[35] N. Saleem, G. Ijaz, Low rank sparse decomposition model based speech enhancement using gammatone filterbank and kullback–leibler divergence, International Journal of Speech Technology 21 (2) (2018) 217–231. doi:10.1007/s10772-018-9500-2.
[36] M. Xu, M. Han, C. P. Chen, T. Qiu, Recurrent broad learning systems for time series prediction, IEEE Transactions on Cybernetics (99) (2018) 1–13. doi:10.1109/TCYB.2018.2863020.
[37] 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.
[38] 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.
[39] 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.
[40] C. Soize, R. G. 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.
[41] 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.
[42] R. G. 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 submitted december 2018 (2019) xx–xx.
[43] 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.
[44] 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.
[45] A. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis, Oxford University Press, Oxford, UK, 1997.
[46] C. E. Shannon, A mathematical theory of communication, Bell System Technology Journal 27 (14) (1948) 379–423 & 623–659.
[47] E. T. Jaynes, Information theory and statistical mechanics, Physical Review 108 (2) (1957) 171–190.
[48] T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, John Wiley & Sons, Hoboken, 2006.
[49] 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.
[50] 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.
[51] 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.
[52] 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.
[53] 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.
[54] C. Soize, The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions, World Scientific Publishing Co Pte Ltd, Singapore, 1994.
[55] R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, PNAS 102 (21) (2005) 7426–7431.
[56] C. Soize, Non gaussian positive-definite matrix-valued random fields for elliptic stochastic partial dif-

ferential operators, Computer Methods in Applied Mechanics and Engineering 195 (1-3) (2006) 26–64. doi:10.1016/j.cma.2004.12.014.

[57] 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.