Probabilistic learning on manifold for optimization under uncertainties

Christian Soize, Roger Ghanem

▶ To cite this version:

Christian Soize, Roger Ghanem. Probabilistic learning on manifold for optimization under uncertainties. (Plenary Lecture), UNCECOMP 2017, 2nd International Conference on Uncertainty Quantification in Computational Sciences and Engineering and COMPDYN 2017, 6th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Jun 2017, Rhodes Island, Greece. pp.1-15. hal-01541216

PROBABILISTIC LEARNING ON MANIFOLD FOR OPTIMIZATION UNDER UNCERTAINTIES

C. Soize[1] and R. Ghanem[2]

[1]Université Paris-Est
Laboratoire MSME, UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France
e-mail: christian.soize@univ-paris-est.fr

[2]University of Southern California
210 KAP Hall, Los Angeles, CA 90089, United States
e-mail: ghanem@usc.edu

Keywords: Optimization under uncertainty, Probabilistic optimization, Nonconvex constrained optimization, Probability distribution on manifolds, Concentration of probability, Random sampling generator, MCMC generator, Diffusion maps, Statistics on manifolds.

Abstract. *This paper presents a challenging problem devoted to the probabilistic learning on manifold for the optimization under uncertainties and a novel idea for solving it. The methodology belongs to the class of the statistical learning methods and allows for solving the probabilistic nonconvex constrained optimization with a fixed number of expensive function evaluations. It is assumed that the expensive function evaluator generates samples (defining a given dataset) that randomly fluctuate around a "manifold". The objective is to develop an algorithm that uses a number of expensive function evaluations at a level essentially equal to that of the deterministic problem. The methodology proposed consists in using an algorithm to generate additional samples in the neighborhood of this manifold from the joint probability distribution of the design parameters and of the random quantities that defined the objective and the constraint functions. This is achieved by using the probabilistic learning on manifold from the given dataset generated by the optimizer without performing additional expensive function evaluations. A statistical smoothing technique is developed for estimating the mathematical expectations in the computation of the objective and constraint functions at any point of the admissible set by using only the additional samples. Several numerical illustrations are presented for validating the proposed approach.*

1 INTRODUCTION

The present paper addresses a novel approach for solving the probabilistic nonconvex constrained optimization by using only a fixed number of the expensive function evaluations. It is assumed that the expensive function evaluator generates samples that fluctuate around a manifold. An algorithm is then introduced to sample the neighborhood of this manifold from the joint probability distribution of the random parameters and the design variables of the stochastic computational model. The underlying manifold is learned from a diffusion process on the dataset that is generated by the optimizer. This paper extends recent work by the authors [1], where the above sampling on manifolds was first introduced, to the case where the joint probability distribution of multiple vectors, is constructed and used to evaluate the conditional expectations that define objective functions and constraints in optimization under uncertainties. The paper is organized as follows. In Section 2, the problem and novel methodology proposed for solving it is presented. Section 3 deals with the probabilistic learning on manifold. Section 4 is devoted to the probabilistic nonconvex constrained optimization to be solved with a fixed number of function evaluations. A numerical illustration is presented for validating the method proposed.

Notations

A lower case letter such as x, η, or u, is a real deterministic variable.
A boldface lower case letter such as $\mathbf{x}$, $\boldsymbol{\eta}$, or $\mathbf{u}$ is a real deterministic vector.
An upper case letter such as X, H, or U, is a real random variable.
A boldface upper case letter, $\mathbf{X}$, $\mathbf{H}$, or $\mathbf{U}$, is a real random vector.
A lower case letter between brackets such as $[x]$, $[\eta]$, or $[u]$), is a real deterministic matrix.
A boldface upper case letter between brackets such as $[\mathbf{X}]$, $[\mathbf{H}]$, or $[\mathbf{U}]$, is a real random matrix.

$\mathbb{N} = \{0, 1, 2, \ldots\}$: set of all the null and positive integers.
$\mathbb{R}$: set of all the real numbers.
$\mathbb{R}^n$: Euclidean vector space on $\mathbb{R}$ of dimension n.
$\|\mathbf{x}\|$: usual Euclidean norm in $\mathbb{R}^n$.
$\mathbb{M}_{n,N}$: set of all the $(n \times N)$ real matrices.
$\mathbb{M}_\nu$: set of all the square $(\nu \times \nu)$ real matrices.
$[x]_{kj}$: entry of matrix $[x]$.
$[x]^T$: transpose of matrix $[x]$.
$\|[x]\|_F$: Frobenius norm of matrix $[x]$ such that $\|x\|_F^2 = \mathrm{tr}\{[x]^T [x]\}$.
$[I_\nu]$: identity matrix in $\mathbb{M}_\nu$.
$\delta_{kk'}$: Kronecker's symbol such that $\delta_{kk'} = 0$ if $k \neq k'$ and $= 1$ if $k = k'$.
$\mathbb{1}_\mathcal{A}(a)$ is the indicator function of set $\mathcal{A}$: $\mathbb{1}_\mathcal{A}(a) = 1$ if $a \in \mathcal{A}$ and $= 0$ if $a \notin \mathcal{A}$.
E: Mathematical expectation.
pdf: probability density function.
ISDE: Itô Stochastic Differential Equation.
MCMC: Markov Chain Monte Carlo.

2 PRESENTATION OF THE PROBLEM AND NOVEL METHODOLOGY PROPOSED FOR SOLVING IT

Increasingly, the design of engineered systems that either involve complex interacting processes or new composite materials, must rely on computational models that resolve the under-

lying physics with great detail. The computational burden associated with such a design, in particular in the presence of model uncertainty or parametric uncertainty, quickly becomes prohibitive as it entails iterating over an already expensive function evaluation. Novel perspectives, methodologies, and algorithms must be developed to fulfill the promise of model-assisted design for such complex systems. The probabilistic learning on manifold that is proposed in this paper is one possible way for solving design optimization problems under uncertainties.

2.1 What is the problem that is considered?

In order to properly define the problem that is analyzed in the framework of this paper, we first detail what we mean by the "probabilistic learning on manifold from a dataset" and what we mean by the "optimization under uncertainties".

Meaning of "probabilistic learning on manifold from a dataset"

In the framework of this paper, "probabilistic learning on manifold from a dataset" is related to the novel methodology that is proposed in [1, 2]

- for identifying, from a database made up of N samples $\boldsymbol{\eta}^\ell = (\eta_1^\ell, \ldots, \eta_\nu^\ell)$ in $\mathbb{R}^\nu$ with $\ell = 1, \ldots, N$ of a $\mathbb{R}^\nu$-valued random variable $\mathbf{H} = (H_1, \ldots, H_\nu)$, its non-Gaussian probability distribution that is unknown and that is concentrated on an unknown subset $\mathcal{S}_\nu$ of $\mathbb{R}^\nu$.

- for generating additional samples that follow the unknown probability distribution in preserving the concentration on $\mathcal{S}_\nu$ and consequently, by avoiding the scattering of the generated samples.

(i) What is a dataset generated by a probability distribution that is concentrated on a subset $\mathcal{S}_\nu$ of $\mathbb{R}^\nu$? In Figure 1, the three figures are related to a subset $\mathcal{S}_\nu$ of $\mathbb{R}^\nu$ with $\nu = 3$ and N samples $\boldsymbol{\eta}^\ell = (\eta_1^\ell, \eta_2^\ell, \eta_3^\ell)$. In the left figure, the statistical mean line of the dataset can easily be identified as a helical. In the central figure, the statistical mean surface of the dataset is concentrated around a surface with a complex geometry as shown in the right figure. For instance, for such examples, the general method proposed in this paper will allow for estimating the statistics of the real-valued random variable $H_3(\eta_1, \eta_2)$ by using only dataset $\{\boldsymbol{\eta}^\ell\}_{\ell=1,\ldots,N}$. It will be not assumed that $N_s' \gg 1$ points $\eta_3^{\ell,\ell'}$ are available for $\ell' = 1, \ldots, N_s'$, and consequently, the classical empirical estimation

$$E\{H_3(\eta_1^\ell, \eta_2^\ell)\} \simeq \tfrac{1}{N_s'} \sum_{\ell'=1}^{N_s'} \eta_3^{\ell,\ell'},$$

cannot be used. In addition, we want to estimate the statistics of $H_3(\eta_1, \eta_2)$ at any point (η_1, η_2) and not only at the points $\{(\eta_1^\ell, \eta_2^\ell), \ell = 1, \ldots N\}$ of the dataset.

(ii) What is the scattering of the generated samples if a classical generator is used? Using the nonparametric statistics for estimating the probability distribution of random vector $\mathbf{H}$ with the dataset made up of $N = 400$ samples that are plotted in Figure 2 (left figure) and if a MCMC generator is used for generating $8,000$ additional samples plotted in Figure 2 (right figure), then it can be seen that a scattering of the generated samples is obtained and the concentration around the statistical mean helical line is lost.

Meaning of "optimization under uncertainties" in the framework of this paper

The terminology "Optimization Under Uncertainties" (OUU) refers to as optimization algorithms with underlying stochastic operators and stochastic constraints. An efficient exploration
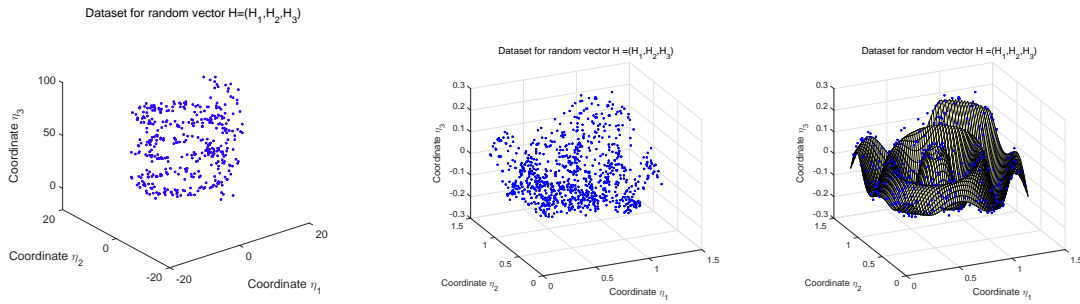
Figure 1: Data set with $N = 400$ for which the statistical mean line of this dataset can easily be identified as a helical (left figure). Data set with $N = 900$ (central figure) for which the statistical mean surface of this dataset has a complex geometry (right figure).
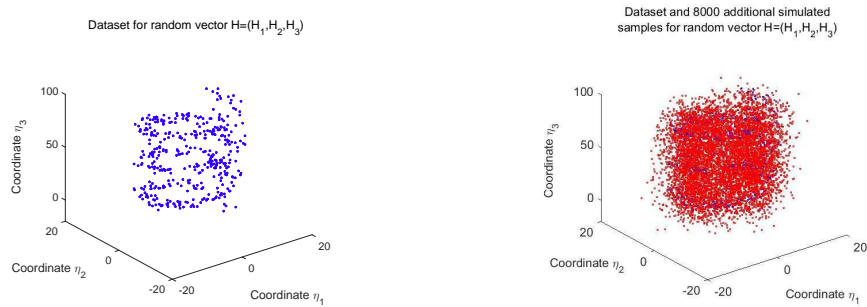

Figure 2: Data set for which the $N = 400$ samples are concentrated around a statistical mean helical line (left figure). Generation with a MCMC generator of $8,000$ additional samples that are scattered (right figure) and consequently, that are not concentrated around the statistical mean helical line.

of the admissible set of the design parameters is crucial to the optimization of a problem with expensive functions (nonconvex objective function and nonlinear constraint function). The development of mathematical and algorithmic constructs that promote learning with successive optimization steps continues to be a key challenge in that regard. In the framework of this paper, "optimization under uncertainties" is related to the novel methodology that is proposed in [3]

- for solving a probabilistic nonconvex constrained optimization (an OUU),

- by using a dataset made up of a small number of points generated by the optimizer for which only a small number of expensive function evaluations is carried out.

A few words about optimization under uncertainties

For solving optimization problems under uncertainties, the methods have progressed along many directions, including gradient-based learning, adapted to convex problems [4, 5], and global search algorithms including stochastic, genetic, and evolutionary algorithms [6, 7]. Statistical learning methods, whereby a deterministic problem is construed as the representative from a class of stochastic problems have also been developed with the benefit of enabling statistical learning [8]. The learning process is typically manifested in the form of a surrogate model from which approximations of the expensive function can be readily evaluated [9]. The resulting error and its repercussions on the attained optimal solution distinguish the various algorithms.

The global character of the surrogate is typically achieved either through a deterministic interpolation process, or a stochastic model whereby biases induced by complex dependencies between model outputs and design parameters are captured through statistical correlations over parameter space. Although Gaussian process models are most commonly used in this context [10, 11], more robust alternatives based on Bayesian optimization [8, 12] have also proven useful. Recent research in the field of uncertainty quantification [13, 14, 15, 16, 17] has underscored the need for optimization algorithms with underlying stochastic operators and constraints. In these situations, that we have previously referred to as OUU, the challenge is magnified since for each design point along the optimization path, a sufficiently large statistical sample of function outputs must be computed to evaluate the required expectations. In essence, the function output must be characterized as a stochastic process over the set of design variables in order to facilitate such evaluations. For expensive function evaluations exhibiting uncertainty, computational challenges remain currently significant enough to require simplifying assumptions in the form of surrogate models for the stochastic function itself or approximations to relevant probabilities [18, 19, 20].

2.2 What is the novel methodology proposed for solving the problem that is considered?

Class of the methodology, fundamental hypothesis, and objective

The methodology belongs to the class of the statistical learning methods. It allows for solving a probabilistic nonconvex constrained optimization with a fixed number of expensive function evaluations. It is assumed that the expensive function evaluator generates samples (the given dataset) that randomly fluctuate around a "manifold". The objective is to develop an algorithm that uses a number of expensive function evaluations at a level essentially equal to that of the deterministic problem.

Principle of the methodology proposed

The methodology proposed [3] consists

- in using an algorithm to generate additional samples in the neighborhood of this manifold from the joint probability distribution of the design parameters and of the random quantities that defined the objective and the constraint functions. This is achieved by using the probabilistic learning on manifold without performing additional expensive function evaluations.

- in developing a statistical smoothing technique for estimating the mathematical expectations in the computation of the objective and constraint functions at any point of the admissible set, by using only the given dataset and the additional samples.

3 PROBABILISTIC LEARNING ON MANIFOLD

3.1 Short summary of the methodology and algorithm for a concentrated probability distribution

In this Section, we summarize the methodology and the algorithm for generating additional samples from a given dataset whose points are the samples of a random vector that follows an unknown concentrated probability distribution, which allows for avoiding the scattering of the generated samples. The details of this approach can be found in [1, 2].

Defining the random matrix $[X]$ and the dataset $[x_d]$ as its given sample

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a $\mathbb{R}^n$-valued random variable defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$. Let $p_{\mathbf{X}}$ be the pdf of $\mathbf{X}$, which is unknown and concentrated on an unknown subset $\mathcal{S}_n$ of $\mathbb{R}^n$. The *dataset* is defined by N given points that are the vectors $\mathbf{x}^{d,1}, \ldots, \mathbf{x}^{d,N}$ in $\mathbb{R}^n$, which correspond to N independent samples of random vector $\mathbf{X}$, and which is represented by the $(n \times N)$ real matrix $[x_d]$ such that

$$[x_d] = [\mathbf{x}^{d,1} \ldots \mathbf{x}^{d,N}] \quad \in \quad \mathbb{M}_{n,N} \, .$$

We define the random matrix $[\mathbf{X}]$ on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{M}_{n,N}$ such that

$$[\mathbf{X}] = [\mathbf{X}^1 \ldots \mathbf{X}^N] \, ,$$

in which the columns columns are N independent copies $\mathbf{X}^1, \ldots, \mathbf{X}^N$ of random vector $\mathbf{X}$. Consequently, the dataset represented by matrix $[x_d]$ is a sample of $[\mathbf{X}]$.

Reduced normalized random matrix $[H]$ and its sample $[\eta_d]$ constructed by the principal component analysis

For $\nu \leq n$, the normalized random matrix $[\mathbf{H}] = [\mathbf{H}^1, \ldots, \mathbf{H}^N]$ with values in $\mathbb{M}_{\nu,N}$ for which the columns are N independent copies of a random vector $\mathbf{H} = (H_1, \ldots H_\nu)$ defined on $(\Theta, \mathcal{T}, \mathcal{P})$ with values in $\mathbb{R}^\nu$, is defined by the following equation that corresponds to a principal component analysis,

$$[\mathbf{X}] = [\underline{x}] + [\varphi] \, [\lambda]^{1/2} \, [\mathbf{H}] \, ,$$

in which $[\lambda]$ is the $(\nu \times \nu)$ diagonal matrix of the ν positive eigenvalues of the empirical estimate $[\text{cov}] \in \mathbb{M}_n$ of the covariance matrix of $\mathbf{X}$ (computed with the dataset), where $[\varphi]$ is the $(n \times \nu)$ matrix of the associated eigenvectors such $[\varphi]^T [\varphi] = [I_\nu]$, and where $[\underline{x}]$ is the matrix in $\mathbb{M}_{n,N}$ with identical columns, each equal to the empirical estimate $\underline{\mathbf{x}} \in \mathbb{R}^n$ of the mean value of random vector $\mathbf{X}$ (computed with the dataset). The sample

$$[\eta_d] = [\boldsymbol{\eta}^{d,1} \ldots \boldsymbol{\eta}^{d,N}] \quad \in \quad \mathbb{M}_{\nu,N}$$

of $[\mathbf{H}]$ (associated with the sample $[x_d]$ of $[\mathbf{X}]$) is computed by

$$[\eta_d] = [\lambda]^{-1/2}[\varphi]^T \, ([x_d] - [\underline{x}]) \, .$$

The empirical estimates of the mean value and of the covariance matrix of random vector $\mathbf{H}$ are therefore $\mathbf{0}_\nu$ and $[I_\nu]$.

Methodology of the mathematical formulation

The methodology of the proposed mathematical formulation that is detailed in [1] can be summarized by the five following steps.

Step 1. A multidimensional kernel-density estimation [21] of the pdf $[\eta] \mapsto p_{[\mathbf{H}]}([\eta])$ of random matrix $[\mathbf{H}]$ is constructed by using the normalized dataset represented by matrix $[\eta_d]$.

Step 2. A Markov chain Monte Carlo (MCMC) generator for random matrix $[\mathbf{H}]$ is constructed by using [22], which belongs to the class of Hamiltonian Monte Carlo methods [22, 23, 24]. The samples are obtained by solving an Itô stochastic differential equation (ISDE) corresponding to

a stochastic nonlinear dissipative Hamiltonian dynamical system, for which $p_{\mathbf{H}}(\boldsymbol{\eta})\,d\boldsymbol{\eta}$ is the unique invariant measure.

Step 3. A diffusion-map approach [25] is used to discover and to characterize the local geometry structure of the normalized dataset concentrated in the neighborhood of the unknown subset $\mathcal{S}_\nu$ of $\mathbb{R}^\nu$. The method consists in introducing the transition matrix $[\mathbb{P}]$ in $\mathbb{M}_N$ such that

$$[\mathbb{P}] = [\mathbb{b}]^{-1}\,[K] \quad , \quad [\mathbb{b}]_{ij} = \delta_{ij}\sum_{j'=1}^{N}[K]_{ij'} \quad , \quad [K]_{ij'} = \exp(-\frac{1}{4\varepsilon}\|\boldsymbol{\eta}^{d,i} - \boldsymbol{\eta}^{d,j'}\|^2)\,,$$

in which $\varepsilon > 0$ is a real smoothing parameter. Let $\boldsymbol{\psi}^1,\ldots,\boldsymbol{\psi}^m$ be the right eigenvectors associated with the m positive eigenvalues $1 = \Lambda_1 > \ldots \geq \Lambda_m$ of the eigenvalue problem $[\mathbb{P}]\,\boldsymbol{\psi}^\alpha = \Lambda_\alpha\,\boldsymbol{\psi}^\alpha$. The eigenvectors are normalized such that $[\psi]^T\,[\mathbb{b}]\,[\psi] = [I_m]$. A reduced order diffusion-maps basis of $\mathbb{R}^N$, of order $m < N$, is defined by

$$[g] = [\mathbf{g}^1\ldots\mathbf{g}^m] \in \mathbb{M}_{N,m}\,.$$

in which $\mathbf{g}^1,\ldots,\mathbf{g}^m$ are the vectors in $\mathbb{R}^\nu$ that are associated with the first m eigenvalues of transition matrix $[\mathbb{P}]$ relative to the local geometric structure of the given normalized dataset, and that are written as

$$\mathbf{g}^\alpha = \Lambda_\alpha^\zeta\,\boldsymbol{\psi}^\alpha \in \mathbb{R}^N \quad , \quad \alpha = 1,\ldots,m\,,$$

in which ζ is a given positive integer. For $m = N$, $\{\mathbf{g}^1,\ldots,\mathbf{g}^N\}$ is an algebraic basis of $\mathbb{R}^N$.

Step 4. The following reduced-order representation of random matrix $\mathbf{H}$,

$$[\mathbf{H}] = [\mathbf{Z}]\,[g]^T\,,$$

is constructed on the manifold in which $[\mathbf{Z}]$ is a random matrix with values in $\mathbb{M}_{\nu,m}$. The value of m is chosen as explained in [1]. As $m < N$, this equation defines a statistical reduction of random matrix $[\mathbf{H}]$ with respect to data dimension N, which allows for keeping the concentration in $\mathcal{S}_\nu \subset \mathbb{R}^\nu$ and consequently, for avoiding the scattering of the generated samples.

Step 5. A reduced-order ISDE is constructed for generating additional samples concentrated in subset $\mathcal{S}_\nu$ without scattering of the generated samples. This MCMC generator on the manifold is obtained by projecting the ISDE introduced in Step 2 onto the diffusion manifold by using the reduced-order diffusion-maps basis represented by matrix $[g]^T$. The constructed reduced-order ISDE is then used for generating n_{MC} additional samples,

$$[z_{\text{ar}}^1],\ldots,[z_{\text{ar}}^{n_{\text{MC}}}] \quad \in \quad \mathbb{M}_{\nu,m}$$

of random matrix $[\mathbf{Z}]$, and therefore, for deducing the n_{MC} additional samples

$$[\eta_{\text{ar}}^1],\ldots,[\eta_{\text{ar}}^{n_{\text{MC}}}] \quad \in \quad \mathbb{M}_{\nu,N}$$

of random matrix $[\mathbf{H}]$, such that $[\eta_{\text{ar}}^\ell] = [z_{\text{ar}}^\ell]\,[g]^T$ for $\ell = 1,\ldots,n_{\text{MC}}$. Let $\{([\boldsymbol{\mathcal{Z}}(r)], [\boldsymbol{\mathcal{Y}}(r)]),$ $r \in \mathbb{R}^+\}$ be the unique asymptotic (for $r \to +\infty$) stationary and ergodic diffusion stochastic process with values in $\mathbb{M}_{\nu,m} \times \mathbb{M}_{\nu,m}$, of the following reduced-order ISDE (stochastic nonlinear second-order dissipative Hamiltonian dynamical system [26, 22]), for $r > 0$,

$$d[\boldsymbol{\mathcal{Z}}(r)] = [\boldsymbol{\mathcal{Y}}(r)]\,dr\,, \tag{1}$$

$$d[\boldsymbol{\mathcal{Y}}(r)] = [\mathcal{L}([\boldsymbol{\mathcal{Z}}(r)])]\,dr - \frac{1}{2}f_0\,[\boldsymbol{\mathcal{Y}}(r)]\,dr + \sqrt{f_0}\,[d\boldsymbol{\mathcal{W}}(r)]\,,$$

with appropriate initial conditions for $r = 0$, and where

- $[\mathcal{L}([\boldsymbol{\mathcal{Z}}(r)])] = [L([\boldsymbol{\mathcal{Z}}(r)]\,[g]^T)]\,[a]$ in which $[\eta] \mapsto [L(\eta)]$ is a nonlinear function from $\mathbb{M}_{\nu,N}$ into $\mathbb{M}_{\nu,N}$, which is expressed as a function of pdf $p_{[\mathbf{H}]}$,

- $[d\boldsymbol{\mathcal{W}}(r)] = [d\mathbf{W}(r)]\,[a]$ where $[d\mathbf{W}(r)]$ is the normalized Wiener process with values in $\mathbb{M}_{\nu,N}$,

- the matrix $[a]$ belongs to $\mathbb{M}_{N,m}$ and is such that $[a] = [g]\,([g]^T\,[g])^{-1}$,

- the free parameter $f_0 > 0$ allows the dissipation term of the nonlinear second-order dynamical system (dissipative Hamiltonian system) to be controlled in order to kill the transient part induced by the initial conditions.

We then have

$$[\mathbf{Z}] = \lim_{r \to +\infty} [\boldsymbol{\mathcal{Z}}(r)] \quad \text{in probability distribution, ,}$$

which allows for generating the additional samples, $[z_{\mathrm{ar}}^1], \ldots, [z_{\mathrm{ar}}^{n_{\mathrm{MC}}}]$.

Remark on the methodology proposed

The stochastic germ of the reduced-order ISDE that is used for generating the samples of $[\mathbf{Z}]$, lives on the "manifold" that is identified by the diffusion maps, that is to say, lives on a subset of the set $\mathbb{M}_{\nu,m}$, which has a small dimension because $m \ll N$. The samples of $[\mathbf{Z}]$ are directly generated by the reduced-order ISDE on the "manifold", which is the subset of $\mathbb{M}_{\nu,m}$ with a small dimension.

3.2 Numerical illustration

As the numerical illustration, we present the simple one introduced in Section 2 for which $n = \nu = 3$ with $N = 400$ given points in the dataset. Another one, corresponding to a petro-physics data base of experiments for which $n = 35$, $\nu = 32$, and $N = 13,056$ given points in the dataset, can be found in [1]. Figure 3 displays the dataset that is made up of 400 given points $\{\boldsymbol{\eta}^\ell\}_\ell$ concentrated around a statistical mean helical line (left figure), and the eigenvalues $\Lambda_{\alpha\,\alpha=1,\ldots,20}$. The convergence analysis leads us a reduction order $m = 4$. Figure 4
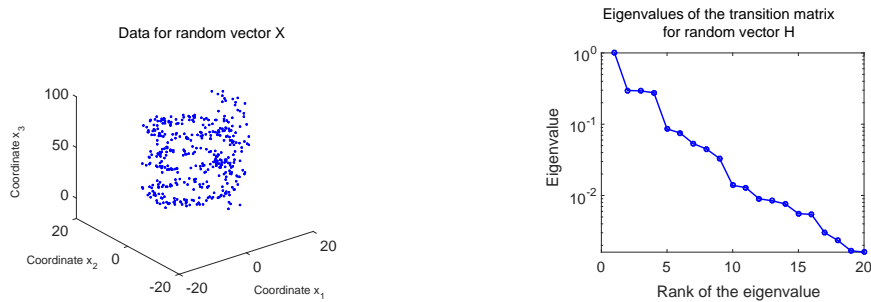

Figure 3: Dataset of 400 given points $\boldsymbol{\eta}^\ell$ concentrated around a statistical mean helical line (left figure). Eigenvalues in $\log_{10}$-scale of the transition matrix $[\mathbb{P}]$ (right figure).

(left) shows the 400 given points of the dataset and the $8,000$ additional realizations generated by using the reduced-order ISDE with $m = 4$. It can be seen that the concentration of the additional samples is kept. Figure 4 (right) displays $8,000$ additional realizations generated by the MCMC generator without using the reduced diffusion-maps basis. In such a case the samples are scattered and the concentration is lost.
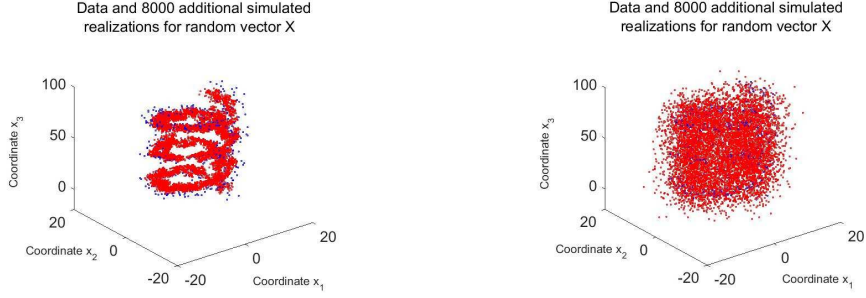
Figure 4: Left figure: 400 given points of the dataset (blue symbols) and the $8,000$ additional realizations (red symbols) generated by using the reduced-order ISDE with $m = 4$. Right figure: $8,000$ additional realizations generated by the MCMC generator without using the reduced diffusion-maps basis.

4 PROBABILISTIC NONCONVEX CONSTRAINED OPTIMIZATION TO BE SOLVED WITH FIXED NUMBER OF FUNCTION EVALUATIONS

4.1 Definition of a probabilistic nonconvex constrained optimization problem

Algebraic form of the optimization problem

Let $\mathbf{w} \in \mathcal{C}_{\mathbf{w}} \subset \mathbb{R}^{m_w}$ be the admissible set of the vector of the design parameters. The nonconvex objective function is defined by the function $\mathbf{w} \mapsto f(\mathbf{w})$ from $\mathcal{C}_{\mathbf{w}}$ into $\mathbb{R}$. The nonlinear constraints are represented by the function $\mathbf{w} \mapsto \mathbf{c}(\mathbf{w})$ from $\mathcal{C}_{\mathbf{w}}$ into $\mathbb{R}^{m_c}$. The Probabilistic Nonconvex Optimization Problem with nonlinear constraints (PNOP) is written as

$$\mathbf{w}^{\text{opt}} = \arg \min_{\mathbf{w} \in \mathcal{C}_{\mathbf{w}}, \, \mathbf{c}(\mathbf{w}) < 0} f(\mathbf{w}) \, .$$

Probabilistic aspects of the optimization problem

The objective function and the constraint function are assumed to be written as

$$f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\} \quad , \quad \mathbf{c}(\mathbf{w}) = E\{\boldsymbol{\mathcal{B}}(\mathbf{w})\} \, .$$

The stochastic process $\{\mathcal{Q}(\mathbf{w}), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ and $\{\boldsymbol{\mathcal{B}}(\mathbf{w}) = (\mathcal{B}_1(\mathbf{w}), \ldots, \mathcal{B}_{m_c}(\mathbf{w})), \mathbf{w} \in \mathcal{C}_{\mathbf{w}}\}$ are defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, are indexed by $\mathcal{C}_{\mathbf{w}}$, are with values in $\mathbb{R}$ and $\mathbb{R}^{m_c}$ respectively, are statistically dependent, are second-order stochastic processes. Consequently, for all $\mathbf{w}$ fixed in $\mathcal{C}_{\mathbf{w}}$, the real-valued random variable $\mathcal{Q}(\mathbf{w}) : \theta \mapsto \mathcal{Q}(\mathbf{w}; \theta)$ and the $\mathbb{R}^{m_c}$-valued random variable $\boldsymbol{\mathcal{B}}(\mathbf{w}) : \theta \mapsto \boldsymbol{\mathcal{B}}(\mathbf{w}; \theta)$ are such that

$$E\{\mathcal{Q}(\mathbf{w})^2\} = \int_{\Theta} \mathcal{Q}(\mathbf{w}; \theta)^2 \, d\mathcal{P}(\theta) < +\infty \, ,$$

$$E\{\|\boldsymbol{\mathcal{B}}(\mathbf{w})\|^2\} = \int_{\Theta} \|\boldsymbol{\mathcal{B}}(\mathbf{w}; \theta)\|^2 \, d\mathcal{P}(\theta) < +\infty \, .$$

4.2 Framework and objective

Framework

For $\mathbf{w}$ given in $\mathcal{C}_{\mathbf{w}}$, $f(\mathbf{w})$ and $\mathbf{c}(\mathbf{w})$ are calculated by using the Stochastic Computational Model (SCM) in which a probabilistic model of uncertainties is implemented. It is assumed that the PNOP defined before has a unique solution $\mathbf{w}^{\text{opt}}$ in $\mathcal{C}_{\mathbf{w}}$.

Objective

The objective is the development of a formulation that permits to solve the PNOP by using a small number of numerical evaluations of $f(\mathbf{w})$ and $\mathbf{c}(\mathbf{w})$ in order to limit the calls to the expensive SCM.

4.3 Methodology

Ingredients

The first ingredient is the probabilistic learning on manifold that has been presented in Section 3, which allows for generating additional samples concentrated on the manifold that has been identified by using the dataset [1, 2], without performing additional function evaluations by the use of the SCM. The second ingredient is a smoothing technique that allows for estimating the mathematical expectations in the computation of $f(\mathbf{w}^0)$ and $\mathbf{c}(\mathbf{w}^0)$ at any point $\mathbf{w}^0$ in $\mathcal{C}_{\mathbf{w}}$, by using only the given dataset and the additional samples [3].

What would be the consequences of the use of the classical procedure?

Let us assume that the PNOP requires N evaluations $f(\mathbf{w}_\ell)$ and $\mathbf{c}(\mathbf{w}_\ell)$ at points $\mathbf{w}^\ell$ for $\ell = 1, \ldots, N$. For a given $\mathbf{w}^\ell$, the use of the classical estimation would lead us to compute the samples $\mathcal{Q}(\mathbf{w}^\ell; \theta_{\ell'})$ and $\mathcal{B}(\mathbf{w}^\ell; \theta_{\ell'})$ for $\ell' = 1, \ldots, N_s'$ with the SCM. For N_s' sufficiently large, the following empirical estimations would be performed,

$$
f(\mathbf{w}_\ell) \simeq \frac{1}{N_s'} \sum_{\ell'=1}^{N_s'} \mathcal{Q}(\mathbf{w}^\ell; \theta_{\ell'}) \quad , \quad \mathbf{c}(\mathbf{w}_\ell) \simeq \frac{1}{N_s'} \sum_{\ell'=1}^{N_s'} \mathcal{B}(\mathbf{w}^\ell; \theta_{\ell'}) \, .
$$

With such a classical approach, the SCM would be called $N_s' \times N$ times, which would be prohibitive for expensive function calls.

Method proposed for avoiding $N_s' \times N$ evaluations with the SCM and based on the use of only N evaluations

Step 1: Construction of the dataset by using only a fixed number N of evaluations. For $\ell = 1, \ldots, N$ (with N fixed), let $\mathbf{w}^\ell$ be the N values of $\mathbf{w}$, which correspond either to a training procedure applied to $\mathbf{w}$ or are some values of $\mathbf{w}$ generated by an optimizer as it explores the admissible domain. Let $q^\ell = \mathcal{Q}(\mathbf{w}^\ell, \theta_\ell)$ and $\mathbf{b}^\ell = \mathcal{B}(\mathbf{w}^\ell, \theta_\ell)$ be the N corresponding samples that are computed by using the SCM (therefore, there are only N evaluations). The dataset is made up of the N data points $\mathbf{x}^1, \ldots, \mathbf{x}^N$ in $\mathbb{R}^n$ such that,

$$
\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell) \quad , \quad \ell = 1, \ldots, N \, ,
$$

with $n = m_w + 1 + m_c$.

Step 2: Construction of the diffusion-maps basis and generating additional samples with the reduced ISDE. We introduce the random variable $\mathbf{X} = (\mathbf{W}, Q, \mathbf{B})$ with values in $\mathbb{R}^n$ such that $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell)$ are N independent samples. The diffusion-maps basis is constructed by using $\{\mathbf{x}^\ell\}_{\ell=1,\ldots,N}$. We can then generate $\nu_{\text{sim}} \gg N$ additional samples such that,

$$
\mathbf{x}_{\text{ar}}^\ell = (\mathbf{w}_{\text{ar}}^\ell, q_{\text{ar}}^\ell, \mathbf{b}_{\text{ar}}^\ell) \quad , \quad \ell = 1, \ldots, \nu_{\text{sim}} \, ,
$$

without performing additional function evaluations.

Step 3: Introduction of a smoothing technique by using a nonparametric statistical estimation and solving the PNOP. A smoothing technique has been developed and is written as

$$f(\mathbf{w}^0) \simeq E\{Q \,|\, \mathbf{W} = \mathbf{w}^0\} \quad , \quad \mathbf{c}(\mathbf{w}^0) \simeq E\{\mathbf{B} \,|\, \mathbf{W} = \mathbf{w}^0\} ,$$

in which the conditional mathematical expectations are estimated by using the kernel density estimation method with the additional samples $\{\mathbf{x}_{\mathrm{ar}}^{\ell} = (\mathbf{w}_{\mathrm{ar}}^{\ell}, q_{\mathrm{ar}}^{\ell}, \mathbf{b}_{\mathrm{ar}}^{\ell})\}_{\ell=1,\dots,\nu_{\mathrm{sim}}}$. It should be noted that an explicit numerical formula exists for $f(\mathbf{w}^0)$ and for $\mathbf{c}(\mathbf{w}^0)$ for any value of dimensions m_w and m_c. The probabilistic nonconvex optimization problem with nonlinear constraints is therefore solved,

$$\mathbf{w}^{\mathrm{opt}} = \arg \min_{\mathbf{w}^0 \in \mathcal{C}_{\mathbf{w}} , \, \mathbf{c}(\mathbf{w}^0) < 0} f(\mathbf{w}^0) .$$

4.4 Numerical illustration

Description

The design parameter $\mathbf{w} = (w_1, w_2)$ is with values in $\mathbb{R}^2$ ($m_w = 2$). The nonconvex objective function $\mathbf{w} \mapsto f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\}$ with values in $\mathbb{R}$ is defined on a subset $\mathcal{C}_{\mathbf{w}} = [0 , 1.1] \times [0 , 1.1]$ of $\mathbb{R}^2$. The constraint function is an affine function $\mathbf{w} \mapsto \mathbf{c}(\mathbf{w}) = E\{\mathcal{B}(\mathbf{w})\}$ defined on $\mathcal{C}_{\mathbf{w}}$ with values in $\mathbb{R}^4$ ($m_c = 4$). Consequently, we have $n = m_w + 1 + m_c = 7$.

Solution of reference

The solution of reference is computed by using the classical procedure with a Cartesian grid of 3,600 points uniformly distributed in $\mathcal{C}_{\mathbf{w}}$ and for which $N_s' = 10,000$ samples. The optimal solution is $w_{r,1}^{\mathrm{opt}} = 0.74$, $w_{r,2}^{\mathrm{opt}} = 0.49$, $f(\mathbf{w}_r^{\mathrm{opt}}) = -0.123$ for which the 4 constraints are active. In Fig. 6, the six figures show the reference solution. Figure 5-(a) displays the graph of objective function $\mathbf{w} \mapsto f(\mathbf{w})$ while Figure 5-(b) corresponds to its contour plot in which the white lozenge marks the location of the optimal solution. Figures 5-(c) to 5-(f) correspond to the graphs of the four components of constraint function $\mathbf{w} \mapsto \mathbf{c}(\mathbf{w})$.
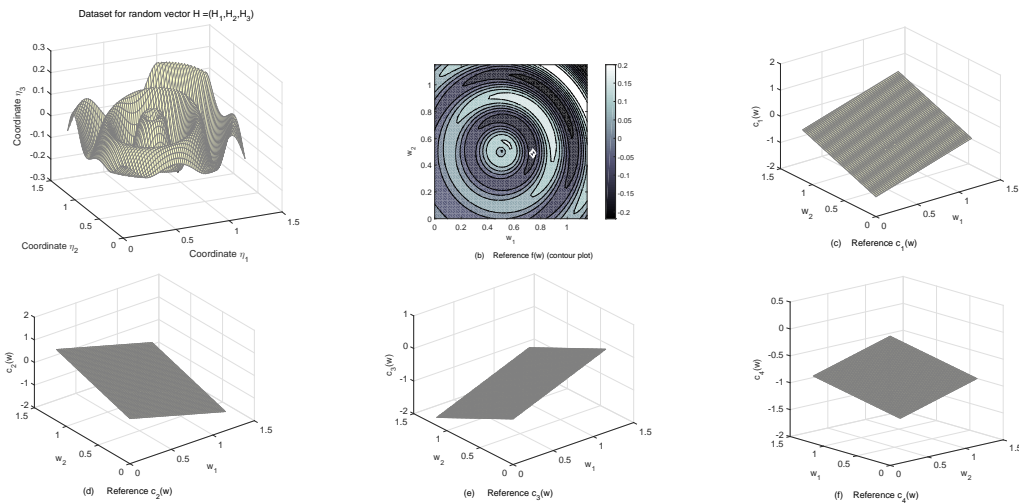

Figure 5: Reference solution: Graph of $\mathbf{w} \mapsto f(\mathbf{w})$ (up left figure), contour plot of $\mathbf{w} \mapsto f(\mathbf{w})$ (b-figure), graphs of $\mathbf{w} \mapsto c_1(\mathbf{w})$ (c-figure), $c_2(\mathbf{w})$ (d-figure), $c_3(\mathbf{w})$ (e-figure), and $c_4(\mathbf{w})$ (f-figure). In b-figure, the white lozenge marks the location of the optimal solution.

Dataset generated by the optimizer

The number of numerical evaluations of the objective function f and of the constraint function $\mathbf{c}$ by using the SCM is $N = 900$. We thus have computed $\mathbf{x}^\ell = (\mathbf{w}^\ell, q^\ell, \mathbf{b}^\ell)$ for $\ell = 1, \ldots, N$ with $q^\ell = \mathcal{Q}(\mathbf{w}^\ell; \theta_\ell)$ and $\mathbf{b}^\ell = \mathcal{B}(\mathbf{w}^\ell; \theta_\ell)$. Figure 6 (left) shows the $N = 900$ given/computed data points $q^\ell = \mathcal{Q}(\mathbf{w}^\ell; \theta_\ell)$ generated by the optimizer for estimating the objective function. Figure 6 (right) gives a view of the statistical fluctuations of these 900 given/computed data points around the surface corresponding to the objective function, $\mathbf{w} \mapsto f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\}$.
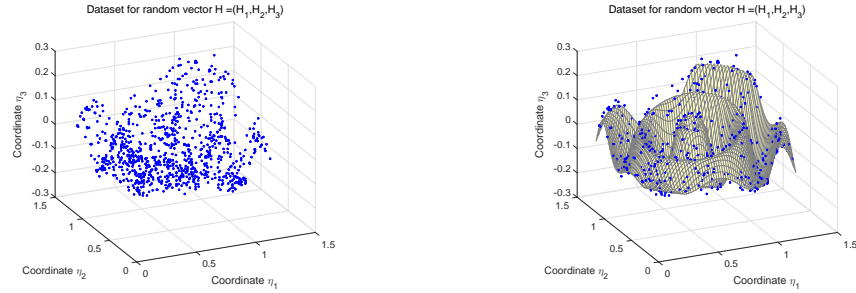

Figure 6: Left figure: $N = 900$ given/computed data points $q^\ell = \mathcal{Q}(\mathbf{w}^\ell; \theta_\ell)$ generated by the optimizer for estimating the objective function. Right figure: View of the statistical fluctuations of these 900 given/computed data points around the surface corresponding to the objective function, $\mathbf{w} \mapsto f(\mathbf{w}) = E\{\mathcal{Q}(\mathbf{w})\}$.

Solution given by the proposed probabilistic learning and data smoothing

The same grid of $3,600$ points is used for the computation. Figure 7 displays the contour plot of the graph of objective function $\mathbf{w} \mapsto f(\mathbf{w})$ constructed with the $N = 900$ given/computed data points (Figure 7-(left)), with $\nu_{\text{sim}} = 9,000$ additional samples (Figure 7-(central)), and with $\nu_{\text{sim}} = 90,000$ additional samples (Figure 7-(right)). In these figures, the white lozenge marks the location of the reference solution while the white circle marks the location of the optimal solution computed with $N = 900$ given/computed data points, $\nu_{\text{sim}} = 9,000$ additional samples, and $\nu_{\text{sim}} = 90,000$ additional samples. For $\nu_{\text{sim}} = 9,000$ additional samples $\{\mathbf{x}_{\text{ar}}^\ell = (\mathbf{w}_{\text{ar}}^\ell, q_{\text{ar}}^\ell, \mathbf{b}_{\text{ar}}^\ell)\}_{\ell=1,\ldots,\nu_{\text{sim}}}$ (Figure 7-(central)), we obtain a good approximation $w_1^{\text{opt}} = 0.70$, $w_2^{\text{opt}} = 0.49$, $f(\mathbf{w}^{\text{opt}}) = -0.112$ of the reference solution, which is confirmed for $\nu_{\text{sim}} = 90,000$. It should be noted that the image of the contour plot of the objective function is really well represented for $\nu_{\text{sim}} = 9,000$ and for $\nu_{\text{sim}} = 90,000$ when comparing these two figures to Figure 5-(b).
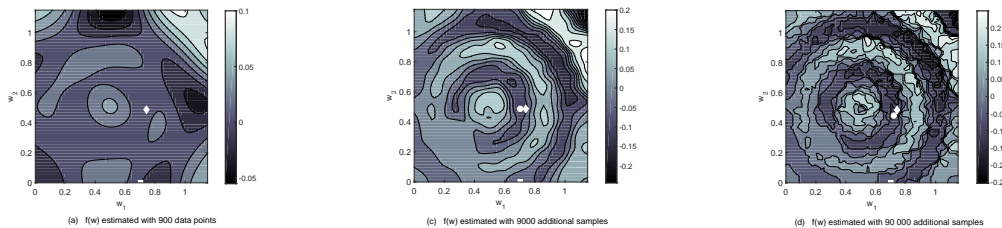

Figure 7: Contour plot of the objective function computed with $N = 900$ given/computed data points (left figure), with $\nu_{\text{sim}} = 9,000$ additional samples (central figure), and with $\nu_{\text{sim}} = 90,000$ additional samples (right figure). The white lozenge marks the location of the reference solution while the white circle marks the location of the optimal solution computed in each case.

5   CONCLUSIONS

A novel methodology has been presented for generating samples of an $\mathbb{R}^n$-valued random vector from a dataset of length N, for which the probability distribution is unknown and is concentrated on an unknown subset $\mathcal{S}_n$ of $\mathbb{R}^n$. The method is robust and efficient for high dimension n and for a big dataset of length N. A new perspective on optimization under uncertainty is given for a nonconvex objective function with a nonlinear constraints function that are viewed as the average of data concentrated around a manifold. For such a case, standard procedures typically require a very large number of function evaluations at each design point. With the method proposed, only a limited number of expensive function evaluations is used.

REFERENCES

[1] C. Soize, R. Ghanem, Data-driven probability concentration and sampling on manifold. *Journal of Computational Physics*, **321**, 9242–258, 2016.

[2] C. Soize, R. Ghanem, Polynomial chaos representation of databases on manifolds. *Journal of Computational Physics*, submitted May 21, 2016, in minor revision October 2016.

[3] R. Ghanem, C. Soize, Probabilistic nonconvex constrained optimization with fixed number of function evaluations. *SIAM Journal on Optimization*, accepted for publication 08 November 2016.

[4] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, pp. 2278-2324, in *Proceedings of the IEEE*, 1998.

[5] J.C. Spall, *Introduction to Stochastic Search and Optimization*, John Wiley and Sons, Hoboken, New Jersey, 2003.

[6] A. Konak, D. Coit, A. Smith, Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety*, **91**(9), 992–1007, 2006.

[7] C. Coello, A. Carlos, Evolutionary multi-objective optimization: A historical view of the field. *IEEE Computational Intelligence Magazine*, **1**(1), 28–36, 2006.

[8] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13** 455–492, 1998.

[9] N. Queipo, R. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, K. Tucker, Surrogate-based analysis and optimization. *Progress in Aerospace Science*, **41**, 1–28, 2005.

[10] J. Kleijnen, W. van Beers, I. van Nieuwenhuyse, Constrained optimization in expensive simulation: Novel approach. *European Journal of Operational Research*, **202**, 164–174, 2010.

[11] Y. Sui, A. Gotovos, J. Burdick, A. Krause, Safe Exploration for Optimization with Gaussian Processes. *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, JMLR: W&CP Volume 37, 2015.

[12] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. de Freitas, Bayesian Optimization in a Billion Dimensions via Random Embeddings. *Journal of Artificial Intelligence Research*, **55**, 361–387, 2016.

[13] R. Ghanem, P.D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, 1991.

[14] C. Soize, A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, **15**(3), 277–294, 2000.

[15] C. Soize, R. Ghanem, Physical systems with random uncertainties: Chaos representation with arbitrary probability measure. *SIAM Journal on Scientific Computing*, **26**(2), 395–410, 2004.

[16] C. Thimmisetty, A. Khodabakhshnejad, N. Jabbari, F. Aminzadeh, R. Ghanem, K. Rose, J. Bauer, C. Disenhof, Multiscale stochastic representation in high-dimensional data using Gaussian processes with implicit diffusion metrics. *Lecture Notes in Computer Science*, **8964**, 157–166, 2015.

[17] R. Ghanem, D. Higdon, H. Owhadi (Editors), *Handbook of Uncertainty Quantification*, Springer-Verlag, 2017.

[18] Xiaoping Du, Wei Chen, Sequential optimization and reliability assessment method for efficient probabilistic design. *ASME Journal of Mechanical Design*, **126**(2), 225–233, 2004.

[19] M. Eldred, Design under uncertainty employing stochastic expansion methods. *International Journal for Uncertainty Quantification*, **1**(2), 119–146, 2011.

[20] W. Yao, X. Chen, W. Luo, M. vanTooren, J. Guo, Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences*, **47**, 450–479, 2011.

[21] A.W. Bowman, A. Azzalini, *Applied Smoothing Techniques for Data Analysis*, Oxford UP, 1997.

[22] 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), 1583-1611 (2008).

[23] R.M. Neal, MCMC using Hamiltonian dynamics. S. Brooks, A. Gelman, G. Jones, X.L. Meng eds. *Handbook of Markov Chain Monte Carlo*, Chapman and Hall-CRC Press, Boca Raton, 2010.

[24] M. Girolami, B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistics Society*, **73**, 123–214, 2011.

[25] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S.W. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, **102**(21), 7426–7431, 2005.

[26] C. Soize, *The Fokker-Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions*, World Scientific, Singapore, 1994.