Optimal partition in terms of independent random vectors of any non-Gaussian vector defined by a set of realizations

Christian Soize

Optimal partition in terms of independent random vectors of any non-Gaussian vector defined by a set of realizations

C. Soize[†]

Abstract. We propose a fast algorithm for constructing an optimal partition, in terms of mutually independent random vectors, of the components of a non-Gaussian random vector that is only defined by a given set of realizations. The method proposed and its objective are different from the independent component analysis (ICA) that was introduced to extract independent source signals from a linear mixture of independent stochastic processes. The algorithm that is proposed is based on the use of the mutual entropy from information theory and on the use of graph theory for constructing an optimal partition. The method has especially been developed for random vectors in high dimension and for which the number of realizations that constitute the data set can be small. The proposed algorithm allows for improving the identification of any stochastic model of a non-Gaussian random vector in high dimension for which a data set is given. Instead of directly constructing a unique stochastic model for which its stochastic dimension, which is identified by solving a statistical inverse problem, can be large, the proposed preprocessing of the data set allows for constructing several mutually independent stochastic models with smaller stochastic dimensions. Consequently, such a method allows for decreasing the cost of the identification and/or to make possible an identification for a case that is a priori in high dimension and that could not be identified through a direct and global approach. The algorithm is completely defined in the paper and can easily be implemented.

Notations. The following notations are used:

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 (except for E, K, S).

A boldface upper case letter, $\mathbf{X}$, $\mathbf{H}$, or $\mathbf{U}$, is a real random vector.

A letter between brackets such as $[x]$, $[\eta]$, $[u]$ or $[C]$, 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.

n: dimension of the random vector $\mathbf{H}^\nu$.

m: number of independent groups in the partition of $\mathbf{H}^\nu$.

E: mathematical expectation.

N: dimension of the random vector $\mathbf{X}$ for which the set of realizations is given.

ν: number of independent realizations of $\mathbf{X}$.

$\mathbb{N}$: set of all the null and positive integers.

$\mathbb{N}^*$: set of the positive integers.

$\mathbb{R}$: real line.

$\mathbb{R}^N$: Euclidean vector space of dimension N.

$\mathbb{M}_{n,N}$: set of all the $(n \times N)$ real matrices.

$\mathbb{M}_n$: $\mathbb{M}_{n,n}$.

$\mathbb{M}_n^S$: set of all the symmetric $(n \times n)$ real matrices.

[†]Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallee, France (christian.soize@univ-paris-est.fr).

$\mathbb{M}_n^+$: set of all the positive-definite $(n \times n)$ real matrices.
$(\Theta, \mathcal{T}, \mathcal{P})$: probability space.
$L^2(\Theta, \mathbb{R}^n)$: space of all the $\mathbb{R}^n$-valued second-order random variables on $(\Theta, \mathcal{T}, \mathcal{P})$.
$\mathbf{x}^T$: transpose of the column matrix representing a vector $\mathbf{x}$ in $\mathbb{R}^N$.
$[x]_{kj}$: entry of matrix $[x]$.
$[x]^T$: transpose of matrix $[x]$.
$\mathrm{tr}\{[x]\}$: trace of a square matrix $[x]$.
$[I_n]$: identity matrix in $\mathbb{M}_n$.
$\|\mathbf{x}\|$: Euclidean norm of vector $\mathbf{x}$.
δ_{ij}: Kronecker's symbol such that $\delta_{ij} = 0$ if $i \neq j$ and $= 1$ if $i = j$.
pdf: probability density function.

1. Introduction. In this introduction, we explain the objective of the paper, we give some explanations concerning the utility of the statistical tool that is proposed, we present a very brief state of the art, and finally, we give the organization of the paper and we describe the methodology proposed.

(i) *Objective*. The objective of this paper is to propose an algorithm for constructing an optimal partition of the components of a non-Gaussian second-order random vector $\mathbf{H}^\nu = (\mathrm{H}_1^\nu, \ldots \mathrm{H}_n^\nu)$ in terms of m mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ with values in $\mathbb{R}^{\mu_1}, \ldots, \mathbb{R}^{\mu_m}$, such that $\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$, where m must be identified as the largest possible integer and where $\mu_1, \ldots, \mu_m$ must be identified as the smallest integers greater than or equal to one, such that $\mu_1 + \ldots + \mu_m = n$. For all $j = 1, \ldots, m$, the components $Y_1^{\nu,j}, \ldots, Y_{\mu_j}^{\nu,j}$ of the $\mathbb{R}^{\mu_j}$-random vector $\mathbf{Y}^{\nu,j}$ are mutually dependent. For such a construction, it is assumed that the solely available information is made up of a given data set made up of ν "experimental" realizations $\boldsymbol{\eta}^{\mathrm{exp},1}, \ldots, \boldsymbol{\eta}^{\mathrm{exp},\nu}$ of the random vector $\mathbf{H}^\nu$. In the method proposed, a probability density function $\boldsymbol{\eta} \mapsto p_{\mathbf{H}^\nu}(\boldsymbol{\eta})$ on $\mathbb{R}^n$ will be constructed by the Gaussian kernel estimation method from the set of the ν realizations. Then, the random vector $\mathbf{H}^\nu$ will be defined as the random vector for which its probability distribution will be $p_{\mathbf{H}^\nu}(\boldsymbol{\eta})\, d\boldsymbol{\eta}$. This is the reason why the random vector is denoted as $\mathbf{H}^\nu$ (and not $\mathbf{H}$) for indicating its dependence on ν). We are interested in the challenging case for which n can be big (high dimension) and for which the number ν of realizations can be small (the statistical estimators can then not be well converged). Note that the "experimental" realizations can come from experimental measurements or come from numerical simulations. Such an objective requires *a priori* (i) to construct an adapted criterion for testing the mutual independence of random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ in a given approximation (the exact independence will not be reached due to the fact that the number ν of "experimental" realizations is finite, and can be small enough), and (ii) to develop an algorithm that is faster than the one that would be based on the exploration all the possible partitions of the n components of $\mathbf{H}^\nu$ (without repetition of the indices). The number of possible partitions that should be constructed and tested with the adapted criterion would then be $\sum_{j=1}^{n-1} C_n^j$ with $C_n^j = ((n - j + 1) \times \ldots \times n)/j!$, which is really too big for a high value of n. In this paper, we propose a statistical tool with a fast algorithm for solving this problem in high dimension and for which the value of ν can be small enough.

(ii) *What would be the utility of the statistical tool proposed*? For many years, probabilistic and statistical methodologies and tools have been developed for uncertainty quantification in computational

sciences and engineering. In particular advanced methodologies and algorithms are necessary for modeling, solving, and analyzing problems in high dimension. A recent state of the art concerning all these methodologies can be found in the Handbook of Uncertainty Quantification [26]. The statistical tool presented in this paper, which consists in constructing an optimal partition of a non-Gaussian random vector in several mutually independent random vectors, each one having a smaller dimension than the original random vector, can be used for decreasing the numerical cost of many existing methodologies devoted to the construction of stochastic models and their identification by solving statistical inverse problems, both for the uncertain parameters of the computational models and the stochastic representation of their random responses.

Let us consider, for instance, the case of the polynomial chaos expansion (PCE) of random vectors, stochastic processes, and random fields, which has been pioneered by Ghanem and Spanos [23] in the community of computational sciences and engineering, on the basis of the mathematical works performed by Wiener [57] and Cameron and Martin [9]. This initial work has then been extended by Ghanem and his co-authors and by many researchers in the community of uncertainty quantification, concerning the development of many methodologies and tools, and for which numerous applications have been performed in many fields of science and engineering with great success. In the area of the methodologies, we refer, for instance, to [2, 14, 15, 16, 17, 24, 25, 40, 44, 45, 53, 55] for the spectral approaches of linear and nonlinear stochastic boundary value problems and some associated statistical inverse problems; to [18, 41, 51, 56, 58] for the generalized chaos expansions; to [3, 4, 5, 21, 36, 42] for works devoted to the dimension reduction and to the acceleration of stochastic convergence of the PCE. If we were interested in the identification of the PCE for the random vector $\mathbf{H}^\nu$ with values in $\mathbb{R}^n$, the use of the statistical tool that is proposed in this paper would allow for obtaining a significant gain if the optimal partition was such that $1 \ll m \le n$. We briefly explain hereinafter the reason why. For $N_d \ge 1$ (N_d will be the maximum degree of the polynomials) and for $1 \le N_g \le n$ (N_g will be the length of the random germ for the polynomial chaos), let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{N_g}) \in \mathbb{N}^{N_g}$ be the multi-index such that $0 \le \alpha_1 + \ldots + \alpha_{N_g} \le N_d$. The first $1 + K$ multi-indices are denoted by $\boldsymbol{\alpha}^{(0)}, \ldots, \boldsymbol{\alpha}^{(K)}$ in which $\boldsymbol{\alpha}^{(0)}$ is the multi-index $(0, \ldots, 0)$, and where $K = (N_g + N_d)!/N_g!\, N_d! - 1$. For N_g fixed, let $\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}), \, \boldsymbol{\alpha} \in (\alpha_1, \ldots, \alpha_{N_g}) \in \mathbb{N}^{N_g}\}$ be the family of multivariate orthonormal polynomials with respect to a given arbitrary probability measure $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})\, d\boldsymbol{\xi}$ on $\mathbb{R}^{N_g}$ such that, for all $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in $\mathbb{N}^{N_g}$, $\int_{\mathbb{R}^{N_g}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\, \Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi})\, p_{\boldsymbol{\Xi}}(\boldsymbol{\xi})\, d\boldsymbol{\xi} = E\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\Xi})\, \Psi_{\boldsymbol{\beta}}(\boldsymbol{\Xi})\} = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$. For $\boldsymbol{\alpha} = \mathbf{0}$, $\Psi_{\mathbf{0}}(\boldsymbol{\eta}) = 1$ is the constant normalized polynomial. The second-order random vector $\mathbf{H}^\nu$ can then be expanded in polynomial chaos $\Psi_{\boldsymbol{\alpha}}$ as $\mathbf{H}^\nu = \lim_{N_g \to n, N_d \to +\infty} \mathbf{H}^{\nu, N_d, N_g}$ with convergence in the vector space of all the $\mathbb{R}^n$-valued second-order random variables, in which $\mathbf{H}^{\nu, N_d, N_g} = \sum_{k=0}^{K} \mathbf{h}^{(k)}\, \Psi_{\boldsymbol{\alpha}^{(k)}}(\boldsymbol{\Xi})$ where $\mathbf{h}^{(0)}, \ldots, \mathbf{h}^{(K)}$ are the coefficients in $\mathbb{R}^n$. The statistical inverse problem consists in estimating integers $N_g \le n$ and $N_d \ge 1$, and the coefficients $\mathbf{h}^{(0)}, \ldots, \mathbf{h}^{(K)}$ in $\mathbb{R}^n$, by using the ν "experimental" realizations $\boldsymbol{\eta}^{\exp,1}, \ldots, \boldsymbol{\eta}^{\exp,\nu}$ in order that $E\{\|\mathbf{H}^\nu - \mathbf{H}^{\nu, N_d, N_g}\|^2\} \le \varepsilon\, E\{\|\mathbf{H}^\nu\|^2\}$. An adapted method such as those proposed in [2, 5, 14, 16, 42, 45, 47, 53] can be used for solving this statistical inverse problem for which the numerical effort is directly related to the value of $N_g \le n$. If an optimal partition $\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ is constructed with $m \le n$, then the PCE can be identified for each $\mathbb{R}^{\mu_j}$-valued random vector $\mathbf{Y}^{\nu,j}$ such that $\mathbf{Y}^{\nu,j,N_d^{(j)}} = \sum_{k=0}^{K_j} \mathbf{h}^{(j,k)}\, \Psi_{\boldsymbol{\alpha}^{(j,k)}}(\boldsymbol{\Xi}^j)$ in which $\mathbf{h}^{(j,0)}, \ldots, \mathbf{h}^{(j,K_j)}$ are the coefficients in $\mathbb{R}^{\mu_j}$, where Ξ^j is a random variable with values in $\mathbb{R}^{N_g^{(j)}}$ with $N_g^{(j)} \le \mu_j < n$ for which its arbitrary probability measure $p_{\boldsymbol{\Xi}^j}(\boldsymbol{\xi}^j)\, d\boldsymbol{\xi}^j$ on $\mathbb{R}^{N_g^{(j)}}$ is given, and where

$K_j = (N_g^{(j)} + N_d^{(j)})!/N_g^{(j)}! \, N_d^{(j)}! - 1$. As the random vectors $\mathbf{\Xi}^1, \ldots, \mathbf{\Xi}^m$ are mutually independent, if m (with $m \leq n$) is big, then the gain obtained is very big, because $\sum_{j=1}^m K_j \ll K$.

(iii) *Brief state of the art concerning the methodologies for testing independence.*
(1) A popular method for testing the statistical independence of the N components of a random vector from a given set of ν realizations is the use of the frequency distribution [19] coupled with the use of the Pearson chi-squared (χ^2) test [46, 29]. For the high dimension (N big) and for a relatively small value of ν, such a an approach does not give sufficiently good results. In addition, if this type of method allows for testing the independence, we always need an additional fast algorithm for constructing the optimal partition that we are looking for.
(2) The independent component analysis (ICA) [31, 35, 10, 11, 33, 12, 34, 39, 6], which is also referred to as the blind source separation, is an efficient method that consists in extracting independent source signals from a linear mixture of mutually statistically independent signals, which is used for source-separation problems, and which is massively used in signal analysis for analyzing, for instance, financial time series or damage in materials, for image processing, in particular for analyzing neuroimage data such as electroencephalogram (EEG) images, neuroimaging of the brain in computational neuroscience, data compression of spectroscopic data sets, etc. The fundamental hypothesis that is introduced in such an ICA methodology is that the observed vector-valued signal is a linear transformation of statistically independent real-valued signals (that is to say, is a linear transformation of a vector-valued signal whose components are mutually independent) and the objective of the ICA algorithms is to identify the best linear operator. The ICA method can be formulated as follows. Let $T = \{t_1, \ldots, t_{N_T}\}$ be a finite set (corresponding to a time index or any other index of interest). For all t in T, let $\mathbf{z}^{\exp,1}(t), \ldots, \mathbf{z}^{\exp,\nu}(t)$ be ν observations (the realizations) of a time series $\{\mathbf{Z}(t) = (Z_1(t), \ldots, Z_N(t)), t \in T\}$ with values in $\mathbb{R}^N$. In the ICA methodology, it is assumed that, for all t in T, the random vector $\mathbf{Z}(t)$ can be modeled as a linear combination of hidden mutually independent random variables $Y_1(t), \ldots, Y_m(t)$, with some unknown coefficients $[c]_{ij}$, that is to say, $Z_i(t) = \sum_{j=1}^m [c]_{ij} Y_j(t)$, or $\mathbf{Z}(t) = [c]\,\mathbf{Y}(t)$, for which matrix $[c]$ and the time series $\{Y_1(t), t \in T\}, \ldots, \{Y_m(t), t \in T\}$, which are assumed to be mutually independent, must be estimated using only $\{\mathbf{z}^{\exp,1}(t), t \in T\}, \ldots, \{\mathbf{z}^{\exp,\nu}(t), t \in T\}$.
(3) In this paper, the method proposed and its objective are different from the ICA. The common building block with the existing methods developed for ICA is the use of the mutual information for measuring the statistical independence. All the other parts of the methodology presented in this paper are different and have been constructed in order to obtain a robust construction of an optimal partition of a non-Gaussian random vector in high dimension, which is defined by a relatively small number of realizations.

(iv) *Organization of the paper and the methodology proposed.* The problem related to the construction of an optimal partition of a non-Gaussian random vector in terms of several mutually independent random vectors is a difficult problem for a random vector in high dimension.
The given data set is made up of ν realizations of a non-Gaussian second-order random vector $\mathbf{X}$ with values in $\mathbb{R}^N$, for which its probability distribution is unknown.
■ In Section 2, a principal component analysis of $\mathbf{X}$ is performed and the ν realizations of the random vector $\mathbf{H}^\nu$ with values in $\mathbb{R}^n$ with $n \leq N$ is carried out. Then, the random vector $\mathbf{H}^\nu$ is defined by its

probability distribution $p_{\mathbf{H}^\nu}(\boldsymbol{\eta})\,d\boldsymbol{\eta}$ in which the pdf $p_{\mathbf{H}^\nu}$ is constructed by the kernel density estimation method using the ν realizations of $\mathbf{H}^\nu$.

■ Section 3 is devoted to the proposed algorithm for constructing an optimal partition of random vector $\mathbf{H}^\nu$ in terms of mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$. For testing the mutual independence of the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ of a given partition of random vector $\mathbf{H}^\nu$, a numerical criterion is defined in coherence with the statistical estimates made with the ν realizations. We have chosen to construct this numerical criterion on the base of the mutual information that is expressed as a function of the Shannon entropy (instead of using another approach such as the one based on the use of the conditional pdf). This choice is motivated by the two following reasons:

(i) The entropy of a pdf is expressed as the mathematical expectation of the logarithm of this pdf. Consequently, for a pdf defined on a high-dimension set (n big), the presence of the logarithm allows for solving the numerical problems induced by the constant of normalization of the pdf.

(ii) As the entropy is a mathematical expectation, this mathematical expectation can be estimated by the Monte Carlo method for which the convergence rate is independent of the dimension thanks to the law of large numbers, a property that is absolutely needed for the high dimensions.

(iii) The criterion that will be constructed by using the Monte Carlo estimation of the entropy will be coherent with the maximum likelihood method, which is often use for solving statistical inverse problems (such as the identification of the coefficients of the PCE of the random vector from a set of realizations).

(iv) We then formulate the problem related to the identification of the optimal partition as an optimization problem. However, the exploration of all the possible partitions that would be necessary for solving this optimization problem is tricky in high dimension. This optimization problem is then replaced by an equivalent one that can be solved with a fast algorithm from the graph theory (detailed in Section 3.6).

■ Section 4 deals with four numerical experiments, which allow for obtaining numerical validations of the approach proposed.

2. Construction of a non-Gaussian reduced-order statistical model from a set of realizations.

2.1. Data description and estimates of the second-order moments.
Let $\mathbf{X} = (X_1, \ldots, X_N)$ be a $\mathbb{R}^N$-valued non-Gaussian second-order random vector defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, for which its probability distribution is represented by an unknown pdf $\mathbf{x} \mapsto p_{\mathbf{X}}(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on $\mathbb{R}^N$. It is assumed that ν (with $\nu \gg 1$) independent realizations $\mathbf{x}^{\exp,1}, \ldots, \mathbf{x}^{\exp,\nu}$ of $\mathbf{X}$ are given (coming from experimental data or from numerical simulations). Let $\widehat{\mathbf{m}}_{\mathbf{X}}^\nu$ and $[\widehat{C}_{\mathbf{X}}^\nu]$ be the empirical estimates of the mean vector $\mathbf{m}_{\mathbf{X}} = E\{\mathbf{X}\}$ and of the covariance matrix $[C_{\mathbf{X}}] = E\{(\mathbf{X} - \mathbf{m}_{\mathbf{X}})(\mathbf{X} - \mathbf{m}_{\mathbf{X}})^T\}$, such that

$$\widehat{\mathbf{m}}_{\mathbf{X}}^\nu = \frac{1}{\nu}\sum_{\ell=1}^{\nu} \mathbf{x}^{\exp,\ell} \quad , \quad [\widehat{C}_{\mathbf{X}}^\nu] = \frac{1}{\nu-1}\sum_{\ell=1}^{\nu}(\mathbf{x}^{\exp,\ell} - \widehat{\mathbf{m}}_{\mathbf{X}}^\nu)(\mathbf{x}^{\exp,\ell} - \widehat{\mathbf{m}}_{\mathbf{X}}^\nu)^T . \tag{2.1}$$

We introduce the non-Gaussian second-order random vector $\mathbf{X}^\nu$ with values in $\mathbb{R}^N$, defined on $(\Theta, \mathcal{T}, \mathcal{P})$, for which the ν independent realizations are $\mathbf{x}^{\exp,1}, \ldots, \mathbf{x}^{\exp,\nu}$,

$$\mathbf{X}^\nu(\theta_\ell) = \mathbf{x}^{\exp,\ell} \in \mathbb{R}^N \quad , \quad \theta_\ell \in \Theta \quad , \quad \ell = 1, \ldots, \nu, \tag{2.2}$$

and for which the mean vector and the covariance matrix are exactly $\widehat{\mathbf{m}}_{\mathbf{X}}^{\nu}$ and $[\widehat{C}_{\mathbf{X}}^{\nu}]$,

$$E\{\mathbf{X}^{\nu}\} = \widehat{\mathbf{m}}_{\mathbf{X}}^{\nu} \quad , \quad E\{(\mathbf{X}^{\nu} - E\{\mathbf{X}^{\nu}\})\,(\mathbf{X}^{\nu} - E\{\mathbf{X}^{\nu}\})^{T}\} = [\widehat{C}_{\mathbf{X}}^{\nu}]\,. \tag{2.3}$$

2.2. Reduced-order statistical model $\mathbf{X}^{(n,\nu)}$ of $\mathbf{X}^{\nu}$. Let n be the reduced-order dimension such that $n \leq N$, which is defined (see hereinafter) by analyzing the convergence of the principal component analysis of random vector $\mathbf{X}^{\nu}$. The eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$ and the associated orthonormal eigenvectors $\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^N$, such that $(\boldsymbol{\varphi}^i)^T \boldsymbol{\varphi}^j = \delta_{ij}$, are such that $[\widehat{C}_{\mathbf{X}}^{\nu}]\,\boldsymbol{\varphi}^i = \lambda_i\,\boldsymbol{\varphi}^i$. The principal component analysis allows for constructing a reduced-order statistical model $\mathbf{X}^{(n,\nu)}$ of $\mathbf{X}^{\nu}$ such that

$$\mathbf{X}^{(n,\nu)} = \widehat{\mathbf{m}}_{\mathbf{X}}^{\nu} + \sum_{i=1}^{n} \sqrt{\lambda_i}\,\mathrm{H}_i^{\nu}\,\boldsymbol{\varphi}^i\,, \tag{2.4}$$

in which

$$\mathrm{H}_i^{\nu} = (\mathbf{X}^{\nu} - \widehat{\mathbf{m}}_{\mathbf{X}}^{\nu})^T\,\boldsymbol{\varphi}^i/\sqrt{\lambda_i} \quad , \quad E\{\mathrm{H}_i^{\nu}\} = 0 \quad , \quad E\{\mathrm{H}_i^{\nu}\mathrm{H}_j^{\nu}\} = \delta_{ij}\,. \tag{2.5}$$

It should be noted that the second-order random variables $H_1^{\nu}, \ldots, H_n^{\nu}$ are non-Gaussian, centered, and uncorrelated, but are statistically dependent. Let $(n, \nu) \mapsto \mathrm{err}(n, \nu)$ be the error function defined on $\{1, \ldots, N\} \times \mathbb{N}^*$ such that

$$\mathrm{err}(n, \nu) = 1 - \frac{\sum_{i=1}^{n} \lambda_i}{\mathrm{tr}[\widehat{C}_{\mathbf{X}}^{\nu}]}\,, \tag{2.6}$$

in which λ_i depends on ν. Note that $n \mapsto \mathrm{err}(n, \nu)$ depends on ν, but that $\mathrm{err}(N, \nu) = 0$ for all ν. For given $\varepsilon > 0$ (sufficiently small with respect to 1), it is assumed that

$$\exists\,\nu_p \in \mathbb{N}^* \;,\; \exists\,n \in \{1, \ldots, N\} \;:\; \forall\,\nu \geq \nu_p \;,\; \mathrm{err}(n, \nu) \leq \varepsilon\,, \tag{2.7}$$

in which n and ε are independent of ν (for $\nu \geq \nu_p$). In the following, it is assumed that n is fixed such that (2.7) is verified. Consequently, since $\mathbf{X}^{(N,\nu)} = \mathbf{X}^{\nu}$, it can be deduced that

$$\forall\,\nu \geq \nu_p \quad , \quad E\{\|\mathbf{X}^{\nu} - \mathbf{X}^{(n,\nu)}\|^2\} \leq \varepsilon\,\mathrm{tr}[\widehat{C}_{\mathbf{X}}^{\nu}]\,. \tag{2.8}$$

The left-hand side of (2.8) represents the square of the norm of random vector $\mathbf{X}^{\nu} - \mathbf{X}^{(n,\nu)}$ in $L^2(\Theta, \mathbb{R}^N)$ and consequently, allows for measuring the distance between $\mathbf{X}^{\nu}$ and $\mathbf{X}^{(n,\nu)}$ in $L^2(\Theta, \mathbb{R}^N)$.
Note that the dimension reduction constructed by using a principal component analysis is not systematically mandatory but is either required for constructing a reduced model such as for a random field or is recommended in order to avoid some possible numerical difficulties that could occur during the computation of the mutual information for which the Gaussian kernel density estimation method is used and therefore, introduces the computation of exponentials.

2.3. Probabilistic construction of the random vector $\mathbf{H}^{\nu}$. In this section, we give the probabilistic construction of the non-Gaussian random vector $\mathbf{H}^{\nu}$ whose components are the coordinates $\mathrm{H}_1^{\nu}, \ldots, \mathrm{H}_n^{\nu}$ introduced in the reduced-order statistical model $\mathbf{X}^{(n,\nu)}$ of $\mathbf{X}^{\nu}$, defined by (2.4).

2.3.1. Independent realizations and second-order moments of the random vector $\mathbf{H}^\nu$.

Let $\mathbf{H}^\nu = (\mathrm{H}_1^\nu, \ldots, \mathrm{H}_n^\nu)$ be the $\mathbb{R}^n$-valued non-Gaussian second-order random variable defined on $(\Theta, \mathcal{T}, \mathcal{P})$. The ν independent realizations $\{\boldsymbol{\eta}^{\mathrm{exp},\ell} = (\eta_1^{\mathrm{exp},\ell}, \ldots, \eta_n^{\mathrm{exp},\ell}) \in \mathbb{R}^n, \ell = 1, \ldots, \nu\}$ of the $\mathbb{R}^n$-valued random variable $\mathbf{H}^\nu$ are computed by

$$\eta_i^{\mathrm{exp},\ell} = \frac{1}{\sqrt{\lambda_i}} (\mathbf{x}^{\mathrm{exp},\ell} - \widehat{\mathbf{m}}_{\mathbf{X}}^\nu)^T \boldsymbol{\varphi}^i \quad , \quad \ell = 1, \ldots, \nu \quad , \quad i = 1, \ldots, n, \tag{2.9}$$

and depend on ν, and for all fixed ν,

$$\widehat{\mathbf{m}}_{\mathbf{H}}^\nu = \frac{1}{\nu} \sum_{\ell=1}^\nu \boldsymbol{\eta}^{\mathrm{exp},\ell} = \mathbf{0} \quad , \quad [\widehat{R}_{\mathbf{H}}^\nu] = \frac{1}{\nu-1} \sum_{\ell=1}^\nu \boldsymbol{\eta}^{\mathrm{exp},\ell} (\boldsymbol{\eta}^{\mathrm{exp},\ell})^T = [I_n]. \tag{2.10}$$

Equations (2.5) and (2.10) show that

$$E\{\mathbf{H}^\nu\} = \widehat{\mathbf{m}}_{\mathbf{H}}^\nu = \mathbf{0} \quad , \quad E\{\mathbf{H}^\nu(\mathbf{H}^\nu)^T\} = [\widehat{R}_{\mathbf{H}}^\nu] = [I_n]. \tag{2.11}$$

2.3.2. Probability distribution of the non-Gaussian random vector $\mathbf{H}^\nu$.

At this step of the construction of random vector $\mathbf{H}^\nu$, its ν realizations $\{\boldsymbol{\eta}^{\mathrm{exp},\ell}, \ell = 1, \ldots, \nu\}$ (see (2.9)) and its second-order moments (see (2.11)) are defined. For completing the probabilistic construction of $\mathbf{H}^\nu$, we define its probability distribution by the pdf $\boldsymbol{\eta} \mapsto p_{\mathbf{H}^\nu}(\boldsymbol{\eta})$ on $\mathbb{R}^n$ that is constructed by the Gaussian kernel density estimation method on the basis of the knowledge of the ν independent realizations $\boldsymbol{\eta}^{\mathrm{exp},1}, \ldots, \boldsymbol{\eta}^{\mathrm{exp},\nu}$. The modification proposed in [54] of the classical Gaussian kernel density estimation method [8] is used so that, for all ν, (2.11) is preserved. The positive valued function $p_{\mathbf{H}^\nu}$ on $\mathbb{R}^n$ is then written as

$$p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) = c_n \, q^\nu(\boldsymbol{\eta}) \quad , \quad \forall \boldsymbol{\eta} \in \mathbb{R}^n, \tag{2.12}$$

in which the positive constant c_n and the positive-valued function $\boldsymbol{\eta} \mapsto q^\nu(\boldsymbol{\eta})$ on $\mathbb{R}^n$ are such that

$$c_n = \frac{1}{(\sqrt{2\pi}\,\widehat{s}_n)^n} \quad , \quad q^\nu(\boldsymbol{\eta}) = \frac{1}{\nu} \sum_{\ell=1}^\nu \exp\{-\frac{1}{2\widehat{s}_n^2}\|\frac{\widehat{s}_n}{s_n}\boldsymbol{\eta}^{\mathrm{exp},\ell} - \boldsymbol{\eta}\|^2\}, \tag{2.13}$$

and where the positive parameters s_n and $\widehat{s}_n$ are defined by

$$s_n = \left\{\frac{4}{\nu(2+n)}\right\}^{1/(n+4)} \quad , \quad \widehat{s}_n = \frac{s_n}{\sqrt{s_n^2 + \frac{\nu-1}{\nu}}}. \tag{2.14}$$

Parameter s_n is the usual multidimensional optimal Silverman bandwidth (taking into account that the standard deviation of each component of $\mathbf{H}^\nu$ is unity) and parameter $\widehat{s}_n$ has been introduced in order that the second equation of (2.11) be verified. It should be noted that, for n fixed, parameters s_n and $\widehat{s}_n$ go to 0^+, and $\widehat{s}_n/s_n$ goes to 1^-, when ν goes to $+\infty$. Using (2.12) to (2.14), it can easily be verified that

$$E\{\mathbf{H}^\nu\} = \int_{\mathbb{R}^n} \boldsymbol{\eta} \, p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \frac{\widehat{s}_n}{s_n} \widehat{\mathbf{m}}_{\mathbf{H}}^\nu = \mathbf{0}, \tag{2.15}$$

$$E\{\mathbf{H}^\nu(\mathbf{H}^\nu)^T\} = \int_{\mathbb{R}^n} \boldsymbol{\eta} \, \boldsymbol{\eta}^T \, p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} = \widehat{s}_n^2 [I_n] + \left(\frac{\widehat{s}_n}{s_n}\right)^2 \frac{(\nu-1)}{\nu}[\widehat{R}_{\mathbf{H}}^\nu] = [I_n]. \tag{2.16}$$

Remark. By construction of $p_{\mathbf{H}^\nu}$, the sequence $\{p_{\mathbf{H}^\nu}\}_\nu$ has a limit, denoted by $p_{\mathbf{H}} = \lim_{\nu \to +\infty} p_{\mathbf{H}^\nu}$, in the space of all the integrable positive-valued functions on $\mathbb{R}^n$. Consequently, the sequence $\{\mathbf{H}^\nu\}_\nu$ of $\mathbb{R}^n$-valued random variables has a limit, denoted by $\mathbf{H} = \lim_{\nu \to +\infty} \mathbf{H}^\nu$, in probability distribution. The probability distribution of the $\mathbb{R}^n$-valued random variable $\mathbf{H}$ is defined by the pdf $p_{\mathbf{H}}$.

2.3.3. Marginal distribution related to a subset of the components of random vector $\mathbf{H}^\nu$.
Let $j_1 < j_2 < \ldots < j_{\mu_j}$ with $\mu_j < n$ be a subset of $\{1, 2, \ldots, n\}$. Let $\mathbf{Y}^{\nu,j} = (H_{j_1}, H_{j_2}, \ldots, H_{j_{\mu_j}})$ be the random vector with values in $\mathbb{R}^{\mu_j}$ and let be $\boldsymbol{\eta}^j = (\eta_1^j, \ldots, \eta_{\mu_j}^j) \in \mathbb{R}^{\mu_j}$. From (2.12) and (2.13), it can easily be deduced that the pdf $p_{\mathbf{Y}^{\nu,j}}$ of $\mathbf{Y}^{\nu,j}$ with respect to $d\boldsymbol{\eta}^j$, which is such that $p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^j) = p_{H_{j_1}, H_{j_2}, \ldots, H_{j_{\mu_j}}}(\eta_1^j, \ldots, \eta_{\mu_j}^j)$, can be written as

$$p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^j) = \widetilde{c}_{\mu_j} \, q_j^\nu(\boldsymbol{\eta}^j) \quad , \quad \forall \boldsymbol{\eta}^j \in \mathbb{R}^{\mu_j} , \tag{2.17}$$

in which the positive constant $\widetilde{c}_{\mu_j}$ and the positive-valued function $\boldsymbol{\eta}^j \mapsto q_j^\nu(\boldsymbol{\eta}^j)$ on $\mathbb{R}^{\mu_j}$ are such that

$$\widetilde{c}_{\mu_j} = \frac{1}{(\sqrt{2\pi}\,\widehat{s}_n)^{\mu_j}} \quad , \quad q_j^\nu(\boldsymbol{\eta}^j) = \frac{1}{\nu} \sum_{\ell=1}^{\nu} \exp\{-\frac{1}{2\widehat{s}_n^2} \sum_{k=1}^{\mu_j} (\frac{\widehat{s}_n}{s_n} \eta_{j_k}^{\exp,\ell} - \eta_{j_k})^2\} . \tag{2.18}$$

Note that, in (2.18), s_n and $\widehat{s}_n$ must be used and not s_{μ_j} and $\widehat{s}_{\mu_j}$.

2.4. Remark concerning convergence aspects.
It can be seen that, for $\nu \to +\infty$, the sequence of random vectors $\{\mathbf{X}^\nu\}_\nu$ converges in probability distribution to random vector $\mathbf{X}$ for which the probability distribution $P_{\mathbf{X}}(d\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$ on $\mathbb{R}^N$ is the limit (in the space of the probability measures on $\mathbb{R}^N$) of the sequence of the probability distributions $\{P_{\mathbf{X}^\nu}(d\mathbf{x}) = p_{\mathbf{X}^\nu}(\mathbf{x})\, d\mathbf{x}\}_\nu$ in which $p_{\mathbf{X}^\nu}$ is the pdf of the random vector $\mathbf{X}^\nu = \mathbf{X}^{(N,\nu)}$ given by (2.4) with $n = N$, and where the pdf $p_{\mathbf{H}^\nu}$ of random vector $\mathbf{H}^\nu$ is defined by (2.12). It should be noted that this result holds because it is assumed that $\mathbf{X}$ is a second-order random vector and that its probability distribution $P_{\mathbf{X}}(d\mathbf{x})$ admits a density $p_{\mathbf{X}}$ with respect to $d\mathbf{x}$.

3. An algorithm for constructing an optimal partition of random vector $\mathbf{H}^\nu$ in terms of mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$.
For n fixed such that $n \leq N$ and for any ν fixed such that $\nu \geq \nu_p$, as $\mathbf{H}^\nu = (H_1^\nu, \ldots, H_n^\nu)$ is a normalized random vector (centered and with a covariance matrix equal to $[I_n]$, see (2.11)), if $\mathbf{H}^\nu$ was Gaussian, then the components $H_1^\nu, \ldots, H_n^\nu$ would be independent. As $\mathbf{H}^\nu$ is assumed to be a non-Gaussian random vector, although $\mathbf{H}^\nu$ is normalized, its components $H_1^\nu, \ldots, H_n^\nu$ are, *a priori*, mutually dependent (it could also be non-Gaussian and mutually independent). In this section that is central for this paper, the question solved is related to the development of an algorithm for constructing an optimal partition of $\mathbf{H}^\nu$, which consists in finding the largest value $m_{\max} \geq 1$ of the number m of random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ made up of the components $H_1^\nu, \ldots, H_n^\nu$, such that $\mathbf{H}^\nu$ can be written as $\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ in which the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent, but for which the components of each random vector $\mathbf{Y}^{\nu,j}$ are, *a priori*, mutually dependent. If the largest value $m_{\max}$ of m is such that
- $m_{\max} = 1$, then there is only one random vector, $\mathbf{Y}^{\nu,1} = \mathbf{H}^\nu$, and all the components of $\mathbf{H}^\nu$ are mutually dependent and cannot be separated into several random vectors;
- $m_{\max} = n$, then each random vector of the partition is made up of one component of $\mathbf{H}^\nu$, which means that, $Y^{\nu,j} = H_j^\nu$ and all the components $H_1^\nu, \ldots, H_n^\nu$ of $\mathbf{H}^\nu$ are mutually independent.

As we have explained in Section 1, for a random vector $\mathbf{H}^\nu$ in high dimension (big value of n) defined by a set of realizations, this problem is difficult. An efficient algorithm based on the use of information theory and of graph theory is proposed hereinafter.

3.1. Defining a partition of $\mathbf{H}^\nu$ in terms of mutually independent random vectors.
Let m be an integer such that
$$1 \leq m \leq n\,.$$

A partition of the components $H_1^\nu, \ldots, H_n^\nu$ of $\mathbf{H}^\nu$, denoted by $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$, consists in writing $\mathbf{H}^\nu$ in terms of m random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ such that

$$\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})\,, \tag{3.1}$$

in which, for all j in $\{1, \ldots, m\}$, μ_j is an integer such that $1 \leq \mu_j \leq n$ and $\mathbf{Y}^{\nu,j} = (Y_1^{\nu,j}, \ldots, Y_{\mu_j}^{\nu,j})$ is a $\mathbb{R}^{\mu_j}$-valued random vector, which is written as

$$\mathbf{Y}^{\nu,j} = (H_{r_1^j}^\nu, \ldots, H_{r_{\mu_j}^j}^\nu) \quad, \quad 1 \leq r_1^j < r_2^j < \ldots < r_{\mu_j}^j \leq n\,, \tag{3.2}$$

in which the equality $r_{\mu_j}^j = n$ can exist only if $m = 1$ (and thus $\mu_1 = n$ that yields $\mathbf{Y}^{\nu,1} = \mathbf{H}^\nu$). For fixed ν, the integers $\mu_1 \geq 1, \ldots, \mu_m \geq 1$ are such that

$$\cup_{j=1}^m \{r_1^j, \ldots, r_{\mu_j}^j\} = \{1, \ldots, n\} \quad, \quad \cap_{j=1}^m \{r_1^j, \ldots, r_{\mu_j}^j\} = \{\emptyset\} \quad, \quad \mu_1 + \ldots + \mu_m = n\,. \tag{3.3}$$

It should be noted that (3.2) implies that there is no possible repetition of indices $r_1^j, \ldots, r_{\mu_j}^j$ and that no permutation is considered (any permutation would yield the same probability distribution for $\mathbf{Y}^{\nu,j}$). If the partition $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$ is made up of m mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$, then the joint pdf $(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m) \mapsto p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m)$ defined on $\mathbb{R}^{\mu_1} \times \ldots \times \mathbb{R}^{\mu_m}$ is such that, for all $\boldsymbol{\eta} = (\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m)$ in $\mathbb{R}^n = \mathbb{R}^{\mu_1} \times \ldots \times \mathbb{R}^{\mu_m}$,

$$p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) = p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m) = p_{\mathbf{Y}^{\nu,1}}(\boldsymbol{\eta}^1) \times \ldots \times p_{\mathbf{Y}^{\nu,m}}(\boldsymbol{\eta}^m)\,, \tag{3.4}$$

where, for all j in $\{1, \ldots, m\}$, $\boldsymbol{\eta}^j \mapsto p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^j)$ is the pdf on $\mathbb{R}^{\mu_j}$ of random vector $\mathbf{Y}^{\nu,j}$. From (2.12) and (3.1), it can be deduced that the joint pdf $p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}$ on $\mathbb{R}^{\mu_1} \times \ldots \times \mathbb{R}^{\mu_m}$ of the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ can be written, for all $\boldsymbol{\eta} = (\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m)$ in $\mathbb{R}^n = \mathbb{R}^{\mu_1} \times \ldots \times \mathbb{R}^{\mu_m}$, as

$$p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}(\boldsymbol{\eta}^1, \ldots, \boldsymbol{\eta}^m) = p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) = c_n\, q^\nu(\boldsymbol{\eta})\,, \tag{3.5}$$

in which c_n and q^ν are defined by (2.13).

3.2. Setting the problem for finding an optimal partition $\mathbb{P}_{\mathrm{opt}}^\nu$ in terms of mutually independent random vectors.
For n and ν fixed such that (2.7) is satisfied, the problem related to the construction of an optimal partition $\mathbb{P}_{\mathrm{opt}}^\nu = \mathbb{P}^\nu(m_{\max}; \mu_1^{\mathrm{opt}}, \ldots, \mu_{m_{\max}}^{\mathrm{opt}})$ in terms of mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m_{\max}}$ consists in finding the largest number $m_{\max}$ of integer m on the set of all the possible partitions $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$. As we will explain later, a numerical criterion must be constructed for testing the mutual independence of the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ in a context for which, in general, the number ν of the given realizations of $\mathbf{H}^\nu$ is not sufficiently large so that the statistical estimators are well converged (negligible statistical fluctuations of the estimates).

3.3. Construction of a numerical criterion for testing a partition in terms of mutually independent random vectors.
In order to characterize the mutual independence of the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ of a partition $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$ of random vector $\mathbf{H}^\nu$, we need to introduce a numerical criterion that we propose to construct by using the mutual information [37, 13]), which is defined as the relative entropy (introduced by Kullback and Leibler [38]) between the joint pdf of all the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ (that is to say, the pdf of $\mathbf{H}^\nu$) and the product of the pdf of each random vector $\mathbf{Y}^{\nu,j}$. This mutual information and the relative entropy can be expressed with the entropy (also called the differential entropy or the Shannon entropy, introduced by Shannon [50]) of the pdf of the random vectors. This choice of constructing a numerical criterion based on the use of the entropy (instead of using another approach such as the one based on the use of the conditional pdf) is motivated by the following reasons:

- As it has previously been explained, we recall that the entropy of a pdf is expressed as the mathematical expectation of the logarithm of this pdf. Consequently, for a pdf whose support is a set having a high dimension, the presence of the logarithm allows for avoiding the numerical problems that are induced by the presence of the constant of normalization of the pdf.
- As the entropy is a mathematical expectation, this mathematical expectation can be estimated by the Monte Carlo method for which the convergence rate is independent of the dimension thanks to the law of large numbers (central limit theorem) [27, 48, 49], property that is absolutely needed for the high dimensions.
- The numerical criterion that will be constructed by using the Monte Carlo estimation of the entropy will be coherent with the maximum likelihood method, which is, for instance (see Section 1), used for the identification of the coefficients of the PCE of random vector $\mathbf{H}^\nu$ by using the set of realizations $\boldsymbol{\eta}^{\mathrm{exp},1}, \ldots, \boldsymbol{\eta}^{\mathrm{exp},\nu}$ [47, 53].

Nevertheless, a numerical criterion that is based on the use of the mutual information cannot be constructed as a direct application of the mutual information, and some additional specific ingredients must be introduced for obtaining a numerical criterion that is robust with respect to the statistical fluctuations induced by a nonperfect convergence of the statistical estimators used, because, for many applications, the number ν of the available independent realizations of the random vector $\mathbf{H}^\nu$ is not sufficiently big.

3.3.1. Entropy of the pdfs related to a partition.
For all j in $\{1, \ldots, m\}$, the entropy of the pdf $p_{\mathbf{Y}^{\nu,j}}$ of the $\mathbb{R}^{\mu_j}$-valued random variable $\mathbf{Y}^{\nu,j}$ is $S(p_{\mathbf{Y}^{\nu,j}}) \in \mathbb{R}$ that is defined (using the convention $0 \log(0) = 0$) by

$$S(p_{\mathbf{Y}^{\nu,j}}) = -E\{\log(p_{\mathbf{Y}^{\nu,j}}(\mathbf{Y}^{\nu,j}))\} = -\int_{\mathbb{R}^{\mu_j}} p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^j) \, \log(p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^j)) \, d\boldsymbol{\eta}^j \,, \qquad (3.6)$$

in which pdf $p_{\mathbf{Y}^{\nu,j}}$ is defined by (2.17). Similarly, the entropy $S(p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}) \in \mathbb{R}$ of the pdf $p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}$ is defined by

$$S(p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}) = -E\{\log(p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}))\} \,, \qquad (3.7)$$

which can be rewritten (using (3.5)) as

$$S(p_{\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}}) = S(p_{\mathbf{H}^\nu}) = -E\{\log(p_{\mathbf{H}^\nu}(\mathbf{H}^\nu))\} = -\int_{\mathbb{R}^n} p_{\mathbf{H}^\nu}(\boldsymbol{\eta}) \, \log(p_{\mathbf{H}^\nu}(\boldsymbol{\eta})) \, d\boldsymbol{\eta} \,. \qquad (3.8)$$

3.3.2. Mutual information of the random vectors related to a partition. The mutual information $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ between the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ is defined by

$$i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) = -E\left\{\log\left(\frac{p_{\mathbf{Y}^{\nu,1}}(\mathbf{Y}^{\nu,1}) \times \ldots \times p_{\mathbf{Y}^{\nu,m}}(\mathbf{Y}^{\nu,m})}{p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})}\right)\right\}, \tag{3.9}$$

in which the conventions $0 \log(0/a) = 0$ for $a \geq 0$ and $b \log(b/0) = +\infty$ for $b > 0$ are used. The mutual information is related to the relative entropy (Kullback-Leibler divergence) of the pdf $p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}$ with respect to the pdf $p_{\mathbf{Y}^{\nu,1}} \otimes \ldots \otimes p_{\mathbf{Y}^{\nu,m}}$,

$$d(p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}} : p_{\mathbf{Y}^{\nu,1}} \otimes \ldots \otimes p_{\mathbf{Y}^{\nu,m}}) = i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}). \tag{3.10}$$

In the following, we will use the mutual information that is equivalent to the Kullback-Leibler divergence. From (3.6) and (3.7), it can be deduced that the mutual information defined by (3.9) can be written as

$$i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) = S(p_{\mathbf{Y}^{\nu,1}}) + \ldots + S(p_{\mathbf{Y}^{\nu,m}}) - S(p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}). \tag{3.11}$$

3.3.3. Inequalities verified by the mutual information and by the entropy. We have the classical important following inequalities that allow for testing the mutual independence of the random vectors defined by a partition $\mathbb{P}^{\nu}(m; \mu_1, \ldots, \mu_m)$:
(i) Using the Jensen inequality, it can be proved [13] that the mutual information is non negative and can be equal to $+\infty$,

$$0 \leq i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \leq +\infty. \tag{3.12}$$

(ii) From (3.11) and (3.12), it can be deduced that

$$S(p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}) \leq S(p_{\mathbf{Y}^{\nu,1}}) + \ldots + S(p_{\mathbf{Y}^{\nu,m}}). \tag{3.13}$$

(iii) If random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent, then from (3.4) and (3.9), it can be deduced that

$$i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) = 0. \tag{3.14}$$

Therefore, (3.11) and (3.14) yield

$$S(p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}) = S(p_{\mathbf{Y}^{\nu,1}}) + \ldots + S(p_{\mathbf{Y}^{\nu,m}}). \tag{3.15}$$

3.3.4. Defining a theoretical criterion for testing the mutual independence of the random vectors of a partition. Let us assume that n and ν are fixed such that (2.7) is satisfied. Taking into account the properties of the mutual information, for testing the mutual independence of random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ of a partition $\mathbb{P}^{\nu}(m; \mu_1, \ldots, \mu_m)$, we choose the mutual information $0 \leq i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \leq +\infty$ as the theoretical criterion that is such that $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) = 0$ if $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent. Equations (2.13) and (2.18) yield $c_n = \widetilde{c}_{\mu_1} \times \ldots \times \widetilde{c}_{\mu_m}$. Consequently, (2.12) and (2.17) show that the mutual information $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$, which is defined by (3.11) and deduced from (3.9), is independent of the constants of normalization $c_n, \widetilde{c}_{\mu_1}, \ldots, \widetilde{c}_{\mu_m}$. However, such a theoretical criterion cannot directly be used (and will be replaced by a numerical criterion derived from the theoretical criterion) for the following reasons:

- The number of realizations ν is fixed and can be relatively small (this means that the asymptotic value $\nu \to +\infty$ is not reached at all for obtaining a good convergence of the statistical estimators). Consequently, there are statistical fluctuations of the theoretical criterion as a function of ν, which exclude the possibility for obtaining exactly the value 0 (by superior values).

- Although the pdf $p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}} = p_{\mathbf{H}^\nu}$ is explicitly defined by (2.12) and the pdf $p_{\mathbf{Y}^{\nu,j}}$ by (2.17), the entropy $S(p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}) = S(p_{\mathbf{H}^\nu})$ and the entropies $S(p_{\mathbf{Y}^{\nu,1}}), \ldots, S(p_{\mathbf{Y}^{\nu,m}})$ cannot exactly be calculated. As we have explained before, for the high dimensions, the Monte Carlo method must be used and, consequently, only a numerical approximation can be obtained. As $p_{\mathbf{H}^\nu}$ is known, a generator of realizations could be used for generating a large number of independent realizations in order to obtain a very good approximation of these entropies. Such a generator could be constructed by using a Markov Chain Monte Carlo (MCMC) algorithm such as the popular Metropolis-Hastings algorithm [27, 30, 43], the Gibbs sampling [20], or other algorithms. Nevertheless, in a following section, for estimating these entropies, we will propose to use only the given set of realizations $\boldsymbol{\eta}^{\exp,1}, \ldots, \boldsymbol{\eta}^{\exp,\nu}$ without adding some additional realizations computed with an adapted generator. The reason why is we want to use exactly the same data for testing the independence and for constructing the estimators related to the identification of the representation of $\mathbf{H}^\nu$ under consideration in order to preserve the same level of statistical fluctuations in all the estimators used.

Remarks concerning the choice of the theoretical criterion. The theoretical criterion consists in testing the mutual information $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ with respect to zero (by superior values) because $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \geq 0$, and the mutual independence is obtained for 0. As this type of test (with respect to 0) has no numerical meaning, this theoretical criterion must be replaced by a numerical criterion that will be derived, and which will be presented in Section 3.3.5. However, it could be interesting to examine what could be other choices for defining a theoretical criterion for mutual independence. Hereinafter, we examine two possibilities among others:

(i) Considering (3.11), (3.13), and (3.14), a first alternative choice would consist in introducing the quantity $S(p_{\mathbf{H}^\nu})/\sum_{j=1}^m S(p_{\mathbf{Y}^{\nu,j}}) \leq 1$ that is equal to 1 when $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent, but which requires us to introduce the hypothesis: $\sum_{j=1}^m S(p_{\mathbf{Y}^{\nu,j}}) > 0$. Unfortunately, such a hypothesis is not satisfied in the general case.

(ii) A second choice could be introduced on the base of the following result. Let $\mathbf{G}^\nu$ be the Gaussian second-order centered $\mathbb{R}^n$-valued random vector for which its covariance matrix is $[I_n]$. Consequently, its components are mutually independent and its entropy is $S(p_{\mathbf{G}^\nu}) = n\,(1 + \log(2\pi))/2 > 1$. On the other hand, it is known (see [13]) that for any non-Gaussian centered random vector $\mathbf{H}^\nu$ with covariance matrix equal to $[I_n]$, we have $S(p_{\mathbf{H}^\nu}) \leq S(p_{\mathbf{G}^\nu})$. It should be noted that the equality $S(p_{\mathbf{H}^\nu}) = S(p_{\mathbf{G}^\nu})$ is verified only if $\mathbf{H}^\nu$ is Gaussian. The components of $\mathbf{H}^\nu$ can be independent and not Gaussian, and in such a case, we have $S(p_{\mathbf{H}^\nu}) < S(p_{\mathbf{G}^\nu})$. Thus, for all $j = 1, \ldots, m$, we have $S(p_{\mathbf{Y}^{\nu,j}}) \leq S(p_{\mathbf{G}^{\nu,j}})$ and it can then be deduced that $0 \leq \sum_{j=1}^m S(p_{\mathbf{Y}^{\nu,j}}) - S(p_{\mathbf{H}^\nu}) \leq \sum_{j=1}^m S(p_{\mathbf{G}^{\nu,j}}) - S(p_{\mathbf{H}^\nu}) = S(p_{\mathbf{G}^\nu}) - S(p_{\mathbf{H}^\nu})$. For the non-Gaussian case, $S(p_{\mathbf{G}^\nu}) - S(p_{\mathbf{H}^\nu}) > 0$ and, consequently, the theoretical criterion $0 \leq \{\sum_{j=1}^m S(p_{\mathbf{Y}^{\nu,j}}) - S(p_{\mathbf{H}^\nu})\}/\{S(p_{\mathbf{G}^\nu}) - S(p_{\mathbf{H}^\nu})\} \leq 1$ could be constructed. Unfortunately, such a criterion would be ambiguous because 0 (by superior values) corresponds to the independence of non-Gaussian vectors, but for a small non-Gaussian perturbation of a Gaussian vector, that is to say if $\mathbf{H}^\nu \to \mathbf{G}^\nu$ in probability distribution, then $S(p_{\mathbf{G}^\nu}) - S(p_{\mathbf{H}^\nu}) \to 0$

(by superior values) and, consequently, an indeterminate form of the criterion is, *a priori*, obtained.

3.3.5. Numerical criterion for testing the mutual independence of the random vectors of a partition.
As we have explained in Section 3.3.4, the theoretical criterion cannot directly be used and a numerical criterion must be derived. Let us introduce the positive-valued random variable Z^ν such that

$$Z^\nu = \frac{p_{\mathbf{Y}^{\nu,1}}(\mathbf{Y}^{\nu,1}) \times \ldots \times p_{\mathbf{Y}^{\nu,m}}(\mathbf{Y}^{\nu,m})}{p_{\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}}(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})} . \tag{3.16}$$

The theoretical criterion (mutual information) defined by (3.9) can be rewritten as

$$i(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = -E\{\log Z^\nu\} . \tag{3.17}$$

(i) *Defining an approximate criterion for testing the mutual independence of the random vectors of a partition.* An approximation of the theoretical criterion defined by (3.9), and rewritten as (3.17), is constructed by introducing the usual estimator of the mathematical expectation, but, as we have previously explained, by using only the ν independent realizations $\boldsymbol{\eta}^{\exp,1},\ldots,\boldsymbol{\eta}^{\exp,\nu}$ of $\mathbf{H}^\nu$. As $\boldsymbol{\eta} = (\boldsymbol{\eta}^1,\ldots,\boldsymbol{\eta}^m) \in \mathbb{R}^n = \mathbb{R}^{\mu_1} \times \ldots \times \mathbb{R}^{\mu_m}$ is the deterministic vector associated with random vector $\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$, for $\ell = 1,\ldots,\nu$, the realization $\boldsymbol{\eta}^{\exp,\ell} \in \mathbb{R}^n$ is rewritten as $\boldsymbol{\eta}^{\exp,\ell} = (\boldsymbol{\eta}^{1,\exp,\ell},\ldots,\boldsymbol{\eta}^{m,\exp,\ell})$ with $\boldsymbol{\eta}^{j,\exp,\ell} = (\eta_1^{j,\exp,\ell},\ldots,\eta_{\mu_j}^{j,\exp,\ell}) = (\eta_{r_1^j}^{\exp,\ell},\ldots,\eta_{r_{\mu_j}^j}^{\exp,\ell}) \in \mathbb{R}^{\mu_j}$. Let $Z^{\exp,1},\ldots,Z^{\exp,\nu}$ be ν independent copies of random variable Z^ν. The approximate criterion for testing the mutual independence of random vectors $\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}$ is defined as the real-valued random variable

$$I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = -\frac{1}{\nu}\sum_{\ell=1}^{\nu}\log Z^{\exp,\ell} . \tag{3.18}$$

If the random vectors $\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}$ are mutually independent, then $I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = 0$ almost surely.

(ii) *Computing a realization $i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$ of the random variable $I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$.* The realization of $I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$ associated with the ν independent realizations $\boldsymbol{\eta}^{\exp,1},\ldots,\boldsymbol{\eta}^{\exp,\nu}$ of $\mathbf{H}^\nu$ is denoted by $i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$, which can be written (see (3.6), (3.8), (3.11), and (3.16) to (3.18)) as

$$i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = s_1^\nu + \ldots + s_m^\nu - s^\nu , \tag{3.19}$$

in which the real numbers $\{s_j^\nu, j = 1,\ldots,m\}$ and s^ν can be computed with the formulas,

$$s_j^\nu = -\frac{1}{\nu}\sum_{\ell=1}^{\nu}\log(p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^{j,\exp,\ell})) \quad , \quad j = 1,\ldots,m , \tag{3.20}$$

$$s^\nu = -\frac{1}{\nu}\sum_{\ell=1}^{\nu}\log(p_{\mathbf{H}^\nu}(\boldsymbol{\eta}^{\exp,\ell})) . \tag{3.21}$$

If the random vectors $\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}$ are mutually independent, then $i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = 0$.

(iii) *Connection with the likelihood function.* From (3.20) and (3.21), s_j^ν and s^ν can be rewritten with the log likelihood functions $L_j^\nu = \sum_{\ell=1}^\nu \log(p_{\mathbf{Y}^{\nu,j}}(\boldsymbol{\eta}^{j,\exp,\ell}))$ and $L^\nu = \sum_{\ell=1}^\nu \log(p_{\mathbf{H}^\nu}(\boldsymbol{\eta}^{\exp,\ell}))$ as $s_j^\nu = -L_j^\nu/\nu$ and $s^\nu = -L^\nu/\nu$. The realization of the approximate criterion can then be rewritten as $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) = -(L_1^\nu + \ldots + L_m^\nu - L^\nu)/\nu$.

(iv) *Random lower bound of the approximate criterion* $I^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$. As the independent random variables $Z^{\exp,1}, \ldots, Z^{\exp,\nu}$ are positive almost surely, and as the arithmetic mean $(\Sigma_{\ell=1}^\nu Z^{\exp,\ell})/\nu$ is greater than or equal to the geometric mean $(\Pi_{\ell=1}^\nu Z^{\exp,\ell})^{1/\nu}$, it can be deduced that

$$\log\{(\Sigma_{\ell=1}^\nu Z^{\exp,\ell})/\nu\} \geq \log\{(\Pi_{\ell=1}^\nu Z^{\exp,\ell})^{1/\nu}\} = \frac{1}{\nu}\sum_{\ell=1}^\nu \log(Z^{\exp,\ell}) \quad a.s. \tag{3.22}$$

Consequently, (3.18) and (3.22) yield the following inequality of real-valued random variables,

$$B^\nu \leq I^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \quad a.s, \tag{3.23}$$

in which the real-valued random variable B^ν is written as

$$B^\nu = -\log\left(\frac{1}{\nu}\sum_{\ell=1}^\nu Z^{\exp,\ell}\right). \tag{3.24}$$

If the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent, then $B^\nu = 0$ almost surely. Considering the ν independent realizations $\boldsymbol{\eta}^{\exp,1}, \ldots, \boldsymbol{\eta}^{\exp,\nu}$ of $\mathbf{H}^\nu$, (3.16) and (3.24) yield the deterministic inequality

$$0 \leq b^\nu \leq i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}), \tag{3.25}$$

in which the real number b^ν is given by

$$b^\nu = -\log\left(\frac{1}{\nu}\sum_{\ell=1}^\nu \frac{p_{\mathbf{Y}^{\nu,1}}(\boldsymbol{\eta}^{1,\exp,\ell}) \times \ldots \times p_{\mathbf{Y}^{\nu,m}}(\boldsymbol{\eta}^{m,\exp,\ell})}{p_{\mathbf{H}^\nu}(\boldsymbol{\eta}^{\exp,\ell})}\right). \tag{3.26}$$

If the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ are mutually independent, then $b^\nu = 0$. It should be noted that ν_p is assumed to be such that, for all $\nu \geq \nu_p$, we have effectively $0 \leq b^\nu$, which implies that for all $\nu \geq \nu_p$, $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \geq 0$. Unfortunately, as ν is generally not sufficiently large for getting good convergence of the estimator of the pdfs in the right-hand side of (3.26), the lower bound b^ν cannot directly be used in the algorithm. An adapted numerical criterion is introduced in paragraph (vi) hereafter.

(v) *Convergence analysis.* Taking into account (2.7), n is fixed and is independent of ν. From the remark given in Section 2.3.2, the sequence $\{\mathbf{H}^\nu\}_\nu$ of random vectors tends in probability distribution to the random vector $\mathbf{H}$ when ν goes to $+\infty$. The probability distribution of $\mathbf{H}$ is defined by the pdf $p_{\mathbf{H}}$. Similarly to (3.1), the limit $\mathbf{H}$ is written as $\mathbf{H} = (\mathbf{Y}^1, \ldots, \mathbf{Y}^m)$ in which for all j in $\{1, \ldots, m\}$, $\mathbf{Y}^j = (Y_1^j, \ldots, Y_{\mu_j}^j)$ is an $\mathbb{R}^{\mu_j}$-valued random vector, where the integers m and $\mu_1, \ldots, \mu_m$ are fixed, independent of ν, and verify (3.3) (see Section 3.1). Consequently, for $\nu \to +\infty$, $i(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \to i(\mathbf{Y}^1, \ldots, \mathbf{Y}^m)$. It is assumed that pdf $p_{\mathbf{H}}$ (that is unknown) is such that there exists ν_0 such that, for

all $\nu > \nu_0$, $E\{(\log Z^\nu)^2\} \le c_Z < +\infty$ in which the positive constant c_Z is independent of ν. For all ν fixed, it can easily be proved that $m_{I^\nu} = E\{I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})\} = i(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$ and $\sigma_{I^\nu}^2 = E\{(I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) - m_{I^\nu})^2\} = \frac{1}{\nu}\sigma_{\log Z^\nu}^2$ with $\sigma_{\log Z^\nu}^2 = E\{(\log Z^\nu)^2\} - i(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})^2$. Consequently, $\sigma_{I^\nu}^2 \to 0$ as $\nu \to 0$. For all $\varepsilon > 0$, the use of the Chebyshev inequality yields $\mathcal{P}\{|I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) - i(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})| \ge \varepsilon\} \le \sigma_{I^\nu}^2/\varepsilon^2$ that allows for obtaining the following convergence property in probability,

$$\lim_{\nu\to+\infty} I^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}) = i(\mathbf{Y}^1,\ldots,\mathbf{Y}^m). \tag{3.27}$$

Equation (3.16) shows that $E\{Z^\nu\} = 1$ and (3.24) shows that B^ν tends to $-\log(1)$. Consequently,

$$\lim_{\nu\to+\infty} B^\nu = 0 \quad \text{in probability}. \tag{3.28}$$

It should be noted that the mutual information of the random vectors $\mathbf{Y}^1,\ldots,\mathbf{Y}^m$ is such that

$$0 \le i(\mathbf{Y}^1,\ldots,\mathbf{Y}^m), \tag{3.29}$$

and, if the random vectors $\mathbf{Y}^1,\ldots,\mathbf{Y}^m$ are independent, then $i(\mathbf{Y}^1,\ldots,\mathbf{Y}^m) = 0$.

(vi) *Defining a numerical criterion for testing the mutual independence of the random vectors of a partition* $\mathbb{P}^\nu(m;\mu_1,\ldots,\mu_m)$. For fixed ν, we have proved that $b^\nu \le i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$ and, if the random vectors $\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}$ are mutually independent, then $0 = b^\nu = i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$. Unfortunately, often, ν is not sufficiently large for getting good convergence of the estimator of the pdfs, and consequently, the mutual information $i^\nu(\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m})$ cannot be compared to the value 0 for testing the mutual independence. We thus propose to construct a numerical criterion derived from the approximate criterion in order to test the mutual independence, as explained hereinafter.

Let $\mathbf{G} = (G_1,\ldots,G_n)$ be the Gaussian second-order centered $\mathbb{R}^n$-valued random vector for which its covariance matrix is $[I_n]$. It is assumed that $\mathbf{G}$ is statistically independent of random vector $\mathbf{H}^\nu$. Therefore the components $G_1,\ldots,G_n$ are mutually independent real-valued random variables. Let $\{\mathbf{g}^{\exp,\ell}, \ell = 1,\ldots,\nu\}$ be ν independent realizations of $\mathbf{G}$ such that $\mathbf{g}^{\exp,\ell} = (g_1^{\exp,\ell},\ldots,g_n^{\exp,\ell}) \in \mathbb{R}^n$.

- By applying the partition $\mathbb{P}^\nu(m;\mu_1,\ldots,\mu_m)$ defined by (3.1) to (3.3) to random vector $\mathbf{G}$, this vector is rewritten as $\mathbf{G} = (\mathbf{G}^1,\ldots,\mathbf{G}^m)$ in which the Gaussian random vector $\mathbf{G}^j = (G_{r_1^j}^j,\ldots,G_{r_{\mu_j}^j}^j)$ has values in $\mathbb{R}^{\mu_j}$, where the indices $1 \le r_1^j < r_2^j < \ldots < r_{\mu_j}^j \le n$ are those defined in (3.2) and (3.3), and where the ν independent realizations of $\mathbf{G}^j$ are $\{\mathbf{g}^{j,\exp,\ell}, \ell = 1,\ldots,\nu\}$ such that $\mathbf{g}^{j,\exp,\ell} = (g_1^{j,\exp,\ell},\ldots,g_{\mu_j}^{j,\exp,\ell}) = (g_{r_1^j}^{\exp,\ell},\ldots,g_{r_{\mu_j}^j}^{\exp,\ell}) \in \mathbb{R}^{\mu_j}$.

- The mutual information $i^\nu(\mathbf{G}^1,\ldots,\mathbf{G}^m)$ of random vectors $\mathbf{G}^1,\ldots,\mathbf{G}^m$ is calculated by using (3.19) to (3.21) after replacing random vectors $\mathbf{Y}^{\nu,1},\ldots,\mathbf{Y}^{\nu,m}$ by the mutually independent Gaussian random vectors $\mathbf{G}^1,\ldots,\mathbf{G}^m$. However, although the values of the pdf, $p_{\mathbf{G}}(\mathbf{g}^{\exp,\ell})$ and $p_{\mathbf{G}^j}(\mathbf{g}^{j,\exp,\ell}))$, could be calculated exactly, these quantities are approximated by using the Gaussian kernel density estimation method as explained in Sections 2.3.2 and 2.3.3 (in order to define a numerical criterion that is coherent from a numerical point of view).

- The Gaussian random vectors $\mathbf{G}^1,\ldots,\mathbf{G}^m$ are mutually independent. For $\nu \to +\infty$, $b_{\text{Gauss}}^\nu = i^\nu(\mathbf{G}^1,\ldots,\mathbf{G}^m)$ goes to 0. Nevertheless, as ν is finite (and sometimes small enough), the numerical value $i^\nu(\mathbf{G}^1,\ldots,\mathbf{G}^m)$ will not be zero, but will be such that $0 < b_{\text{Gauss}}^\nu \le i^\nu(\mathbf{G}^1,\ldots,\mathbf{G}^m)$.

Consequently, for ν fixed, the numerical value $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m)$ is chosen as the lower bound of the mutual information for testing the mutual independence.

- The random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ of the partition $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$ will be considered as *mutually independent* if

$$0 \leq i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \leq i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m), \tag{3.30}$$

while they will be considered as *mutually dependent* if

$$0 \leq i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m) < i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}). \tag{3.31}$$

- Taking into account (3.25), (3.30), and (3.31), for testing the independence of the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ of partition $\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m)$, we introduce $\widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)$ as the *numerical criterion*, which is defined by

$$\widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m) = 1 - \frac{i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})}{i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m)}. \tag{3.32}$$

This numerical criterion is such that random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ will be considered as

$$\text{mutually } \textit{independent} \text{ if } \quad 0 \leq \widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m) \leq 1,$$
$$\text{mutually } \textit{dependent} \text{ if } \quad \widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m) < 0.$$

By convention, $\widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)$ will be taken equal to 1 if $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m) \ll 1$ with $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}) \leq i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m)$.

The numerical criterion defined by (3.32) has been constructed for analyzing the high-dimension case (n big) with a number of realizations, which can be relatively small (ν small). Consequently, the estimate $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ of the mutual information performed by using the nonparametric statistics is not sufficiently accurate and there exist statistical fluctuations. In order to increase the robustness of the prevision of the mutual independence, the theoretical lower bound "zero" is replaced by the lower bound b^ν_{Gauss} (that would be equal to zero if ν was equal to infinity), and which is estimated with the same number ν of realizations as for $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$, in order to construct a robust numerical criterion that takes into account the statistical fluctuations of the lower bound and that allows for testing the mutual independence of the random vectors of a given partition. Finally, it should be noted that the numerical criterion defined by (3.32) corresponds to a lower bound b^ν_{Gauss} that depends on the partition analyzed and, consequently, the numerical criterion changes for every partition, but, this is the effect that is searched in order to adapt the level of the statistical fluctuations of the numerical criterion to every partition analyzed.

(vii) *Comment about the introduction of random variable B^ν.* It should be noted that the random variable B^ν and its estimate b^ν have been introduced for justifying the introduction of the lower bound for the Gaussian case, which allows for constructing the numerical criterion defined by (3.32). Consequently, B^ν and b^ν are introduced for a theoretical understanding of the proposed construction, but do not play any role in the algorithm that will be introduced in Sections 3.5 and 3.6.

3.4. Optimization problem for finding the optimal partition of the random vector $\mathbf{H}^\nu$ in terms of mutually independent random vectors.

Using the numerical criterion defined by (3.32), we propose to construct the solution of the problem formulated in Section 3.2 by solving the following optimization problem:

$$(m_{\max}; \mu_1^{\mathrm{opt}}, \ldots, \mu_{m_{\max}}^{\mathrm{opt}}) = \inf_m \{\arg \max_{1 \leq m \leq n} \{\max_{(\mu_1, \ldots, \mu_m) \in \mathcal{C}_m} \widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)\}\}, \qquad (3.33)$$

in which the admissible set $\mathcal{C}_m \subset \{1, \ldots, n\}^m$ is such that $\mathcal{C}_m = \{\mu_1 \geq 1, \ldots, \mu_m \geq 1, \mu_1 + \ldots + \mu_m = n\}$ and where the notation used in (3.33) is detailed in the fourth paragraph of the following remarks.

Remarks about the construction proposed.
- The optimal partition that is constructed depends on ν. Such a formulation is reasonable because the available information is only made up of the set of the ν realizations of random vector $\mathbf{H}^\nu$.
- The principle that corresponds to the formulation defined by (3.33) consists in finding the partition in terms of mutually independent vectors for which (3.30) is obtained with the largest positive value of the difference $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m) - i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$, that is to say, with the largest positive value of $\widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)$. Such a criterion yields the largest robustness with respect to the statistical fluctuations induced by the fact that ν can be small enough.
- The mathematical analysis of the optimization problem defined by (3.33) seems difficult enough and we have not been able to obtain interesting properties in this way, as against many numerical experiments have confirmed that this choice was efficient with respect to the problem that we had to solve.
- For $m_{\max} \geq 2$, let $T = \arg \max_{1 \leq m \leq n} \{\max_{(\mu_1, \ldots, \mu_m) \in \mathcal{C}_m} \widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)\}$ be the set of all the integers m with $1 \leq m \leq n$ such that $m \mapsto \max_{(\mu_1, \ldots, \mu_m) \in \mathcal{C}_m} \widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)$ is maximum with respect to m. *A priori*, set T is not always reduced to a single point. If $m_{\max}^{\mathrm{inf}}$ is the number of independent random vectors of the optimal partition corresponding to $\inf_m\{T\}$ and $m_{\max}^{\mathrm{sup}}$ the one corresponding to $\sup_m\{T\}$, then it can be deduced that $m_{\max}^{\mathrm{inf}} \leq m_{\max}^{\mathrm{sup}}$. Although we are interested in identifying the largest value of m, the most robust solution leads us to choose the largest number of dependent random vectors for the identified optimal partition. Therefore, $m_{\max} = m_{\max}^{\mathrm{inf}} = \inf_m\{T\}$ is selected as the maximum number of independent random vectors for this optimal partition. This is the reason why the inferior value of set T has been introduced in (3.33).

3.5. Reformulation of the optimization problem for constructing the optimal partition of the random vector $\mathbf{H}^\nu$ in terms of mutually independent random vectors.

The reason why a reformulation of the optimization problem defined by (3.33) is necessary is explained hereinafter, and a new formulation that will be adapted to the development of an algorithm based on graph theory (presented in Section 3.6) is proposed.

3.5.1. Why a reformulation of the optimization problem is necessary.

The optimization problem defined by (3.33) consists in finding the optimal partition $\mathbb{P}_{\mathrm{opt}}^\nu = \mathbb{P}(m_{\max}; \mu_1^{\mathrm{opt}}, \ldots, \mu_{m_{\max}}^{\mathrm{opt}})$ on the set of all the possible partitions $\{\mathbb{P}^\nu(m; \mu_1, \ldots, \mu_m), 1 \leq m \leq n, \mu_1 + \ldots + \mu_m = n\}$, for which $\mu_1^{\mathrm{opt}} + \ldots + \mu_{m_{\max}}^{\mathrm{opt}} = n$ and where $m_{\max}$ is the largest value of the number m of independent

random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ such that, for $j = 1, \ldots, m$,

$$\mathbf{Y}^{\nu,j} = (H^{\nu}_{r^j_1}, \ldots, H^{\nu}_{r^j_{\mu_j}}) \quad , \quad 1 \leq r^j_1 < r^j_2 < \ldots < r^j_{\mu_j} \leq n \quad \text{verifying (3.3)} . \tag{3.34}$$

For solving (3.33), a natural algorithm would consist in computing $\widetilde{\tau}_{\nu}(m; \mu_1, \ldots, \mu_m)$ for each partition $\mathbb{P}^{\nu}(m; \mu_1, \ldots, \mu_m)$ among all the possible partitions. However, as the number of all the possible partitions is $\sum_{j=1}^{n-1} C^j_n$ with $C^j_n = ((n-j+1) \times \ldots \times n)/j!$, it can be seen that, for a high value of n, such a computation would be very tricky. We then propose another algorithm, which is faster than the natural one. This algorithm will be based on the use of the graph theory. The edges of the graph that will be constructed are obtained by analyzing the statistical dependence of the components $(H^{\nu}_{r_1}, H^{\nu}_{r_2})$, two by two, for which the indices r_1 and r_2 (with $r_1 \neq r_2$) are taken in the set of the indices of components $H^{\nu}_1, \ldots, H^{\nu}_n$ of $\mathbf{H}^{\nu}$. Consequently, first, we need to adapt the previous notations relative to the mutual information for two scalar random components and, second, we need to reformulate the optimization problem defined by (3.33).

3.5.2. Adapting the notations for the computation of the mutual information of the real-valued random variables $H^{\nu}_{r_1}$ and $H^{\nu}_{r_2}$. Let r_1 and r_2 be any two integers belonging to $\{1, \ldots, n\}$ with $r_1 \neq r_2$. Hereinafter, the results related to the mutual information and to the numerical criterion presented in Section 3.3 are adapted for analyzing the independence of the real-valued random variables $H^{\nu}_{r_1}$ and $H^{\nu}_{r_2}$.

Let $i^{\nu}(H^{\nu}_{r_1}, H^{\nu}_{r_2})$ be the realization of the random mutual information $I^{\nu}(H^{\nu}_{r_1}, H^{\nu}_{r_2})$, computed by using the ν independent realizations $\boldsymbol{\eta}^{\mathrm{exp},1}, \ldots, \boldsymbol{\eta}^{\mathrm{exp},\nu}$ of $\mathbf{H}^{\nu}$, in which $\boldsymbol{\eta}^{\mathrm{exp},\ell} = (\eta^{\mathrm{exp},\ell}_1, \ldots, \eta^{\mathrm{exp},\ell}_n) \in \mathbb{R}^n$. From (3.19) to (3.21), it can be deduced that

$$i^{\nu}(H^{\nu}_{r_1}, H^{\nu}_{r_2}) = s^{\nu,2}_{r_1} + s^{\nu,2}_{r_2} - s^{\nu,2}_{r_1,r_2} , \tag{3.35}$$

in which the real numbers $s^{\nu,2}_{r_1}$, $s^{\nu,2}_{r_2}$, and $s^{\nu,2}_{r_1,r_2}$ are computed with the formulas,

$$s^{\nu,2}_r = -\frac{1}{\nu} \sum_{\ell=1}^{\nu} \log(p_{H^{\nu}_r}(\eta^{\mathrm{exp},\ell}_r)) \quad , \quad r \in \{r_1, r_2\} , \tag{3.36}$$

$$s^{\nu,2}_{r_1,r_2} = -\frac{1}{\nu} \sum_{\ell=1}^{\nu} \log(p_{H^{\nu}_{r_1}, H^{\nu}_{r_2}}(\eta^{\mathrm{exp},\ell}_{r_1}, \eta^{\mathrm{exp},\ell}_{r_2})) . \tag{3.37}$$

From Sections 2.3.2 and 2.3.3, it can be seen that the pdf of H^{ν}_r is written as

$$p_{H^{\nu}_r}(\eta_r) = \widetilde{c}_1 \, q^{\nu}_r(\eta_r) \quad , \quad \forall \eta_r \in \mathbb{R} , \tag{3.38}$$

in which the positive constant $\widetilde{c}_1$ and the positive-valued function $\eta_r \mapsto q^{\nu}_r(\eta_r)$ on $\mathbb{R}$ are such that

$$\widetilde{c}_1 = \frac{1}{\sqrt{2\pi}\,\widehat{s}_n} \quad , \quad q^{\nu}_r(\eta_r) = \frac{1}{\nu} \sum_{\ell=1}^{\nu} \exp\{-\frac{1}{2\widehat{s}^2_n}(\frac{\widehat{s}_n}{s_n}\eta^{\mathrm{exp},\ell}_r - \eta_r)^2\} . \tag{3.39}$$

The joint pdf of $(H^{\nu}_{r_1}, H^{\nu}_{r_2})$ is written as

$$p_{H^{\nu}_{r_1}, H^{\nu}_{r_2}}(\eta_{r_1}, \eta_{r_2}) = \widetilde{c}_2 \, q^{\nu}_{r_1,r_2}(\eta_{r_1}, \eta_{r_2}) \quad , \quad \forall (\eta_{r_1}, \eta_{r_2}) \in \mathbb{R}^2 , \tag{3.40}$$

in which the positive constant $\widetilde{c}_2$ and the positive-valued function $(\eta_{r_1}, \eta_{r_2}) \mapsto q^\nu_{r_1,r_2}(\eta_{r_1}, \eta_{r_2})$ on $\mathbb{R}^2$ are such that

$$\widetilde{c}_2 = \frac{1}{(\sqrt{2\pi}\,\widehat{s}_n\,)^2} \quad , \quad q^\nu_{r_1,r_2}(\eta_{r_1}, \eta_{r_2}) = \frac{1}{\nu} \sum_{\ell=1}^\nu \exp\{-\frac{1}{2\widehat{s}_n^2} \sum_{k=1}^2 (\frac{\widehat{s}_n}{s_n} \eta_{r_k}^{\exp,\ell} - \eta_{r_k})^2\}. \qquad (3.41)$$

In (3.39) and (3.41), the constants s_n and $\widehat{s}_n$ are given by (2.14).

3.5.3. Reformulation of the optimization problem.

Let $i_{\text{ref}} \geq 0$ be any fixed real value of the mutual information for two real-valued random variables. Let r_1 and r_2 be any two integers such that $1 \leq r_1 < r_2 \leq n$. The random variables $H^\nu_{r_1}$ and $H^\nu_{r_2}$ will be considered as

$$\textit{dependent} \quad \text{if} \quad 0 \leq i_{\text{ref}} < i^\nu(H^\nu_{r_1}, H^\nu_{r_2}),$$

$$(3.42)$$

$$\textit{independent} \quad \text{if} \quad 0 \leq i^\nu(H^\nu_{r_1}, H^\nu_{r_2}) \leq i_{\text{ref}}.$$

We then introduce the following rule given by the probability theory: If r_1, r_2, and r_3 are three indices such that $1 \leq r_1 < r_2 < r_3 \leq n$, then

- if $H^\nu_{r_1}$ is dependent on $H^\nu_{r_2}$ and if $H^\nu_{r_1}$ is dependent on $H^\nu_{r_3}$, then $H^\nu_{r_1}$, $H^\nu_{r_2}$, and $H^\nu_{r_3}$ are mutually dependent;
- if $H^\nu_{r_1}$ is dependent on $H^\nu_{r_2}$ and if $H^\nu_{r_2}$ is dependent on $H^\nu_{r_3}$, then $H^\nu_{r_1}$, $H^\nu_{r_2}$, and $H^\nu_{r_3}$ are mutually dependent.

From (3.42) and this rule, it can be deduced that,

- when $i_{\text{ref}} = 0$, if $0 < i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$ for all $1 \leq r_1 < r_2 < r_3 \leq n$, then all the components $H^\nu_1, \ldots, H^\nu_n$ of $\mathbf{H}^\nu$ are considered as mutually dependent and, consequently, $m = 1$ (there is only one subset in the partition, which coincides with $\mathbf{H}^\nu$);
- when $i_{\text{ref}} > \max_{1 \leq r_1 < r_2 \leq n} i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$, if $i^\nu(H^\nu_{r_1}, H^\nu_{r_2}) \leq i_{\text{ref}}$ for all $1 \leq r_1 < r_2 < r_3 \leq n$, then all the components $H^\nu_1, \ldots, H^\nu_n$ of $\mathbf{H}^\nu$ are considered as independent and consequently, $m = n$ (there are n independent subsets in the partition);
- when i_{ref} is such that $0 < i_{\text{ref}} < \max_{1 \leq r_1 < r_2 \leq n} i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$, the construction of the partition $\mathbb{P}^\nu_{i_{\text{ref}}}(m; \mu_1, \ldots, \mu_m)$ of $\mathbf{H}^\nu$ in terms of random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ considered as independent for the given level i_{ref}, is deduced from the previous rule and from graph theory [7]. The set of nodes of the graph is $\mathcal{I} = \{1, \ldots, n\}$. Let $[\mathcal{N}]$ be the symmetric adjacency $(n \times n)$ matrix made up of 0 and 1, for which all its diagonal entries are 0 and such that, for all integers r_1 and r_2 verifying $1 \leq r_1 < r_2 \leq n$,

$$[\mathcal{N}]_{r_1 r_2} = [\mathcal{N}]_{r_2 r_1} = 0 \quad \text{if} \quad i^\nu(H^\nu_{r_1}, H^\nu_{r_2}) \leq i_{\text{ref}},$$

$$(3.43)$$

$$[\mathcal{N}]_{r_1 r_2} = [\mathcal{N}]_{r_2 r_1} = 1 \quad \text{if} \quad i_{\text{ref}} < i^\nu(H^\nu_{r_1}, H^\nu_{r_2}).$$

For such a given level i_{ref} and from the knowledge of adjacency matrix $[\mathcal{N}]$, the algorithm presented in Section 3.6 allows for constructing the unique partition $\mathbb{P}^\nu_{i_{\text{ref}}}(m; \mu_1, \ldots, \mu_m)$ associated with i_{ref} and then for computing the numerical criterion $\widetilde{\tau}_\nu(m; \mu_1, \ldots, \mu_m)$, defined by (3.32), which is rewritten as $\tau_\nu(i_{\text{ref}})$,

$$\tau_\nu(i_{\text{ref}}) = 1 - \frac{i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})}{i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m)}. \qquad (3.44)$$

Consequently, the optimization problem defined by (3.33) can be rewritten as the following one,

$$\mathbb{P}^\nu(m_{\max}; \mu_1^{\text{opt}}, \dots, \mu_{m_{\max}}^{\text{opt}}) = \mathbb{P}^\nu_{i_{\text{ref}}^{\text{opt}}}(m; \mu_1, \dots, \mu_m), \qquad (3.45)$$

in which the optimal level $i_{\text{ref}}^{\text{opt}}$ of i_{ref} is given by

$$i_{\text{ref}}^{\text{opt}} = \inf_{i_{\text{ref}}}\{\arg \max_{i_{\text{ref}} \geq 0} \ \tau_\nu(i_{\text{ref}})\}. \qquad (3.46)$$

Remarks.
- In (3.46), the *inf* operator is introduced for the same reason as the one given for justifying (3.33). Let $T = \arg \max \tau_\nu(i_{\text{ref}})$ be the set of all the $i_{\text{ref}} \geq 0$ such that $\tau_\nu(i_{\text{ref}})$ is maximum. The numerical experiments have shown that set T is not always reduced to a single point. The solution is thus chosen as $i_{\text{ref}}^{\text{opt}} = \inf\{T\}$.
- If for all $i_{\text{ref}} > 0$, we have $\tau_\nu(i_{\text{ref}}) \leq \tau(i_{\text{ref}}^{\text{opt}}) < 0$ with $|\tau(i_{\text{ref}}^{\text{opt}})| \leq \epsilon \ll 1$ (for instance, $\epsilon = 10^{-3}$), then all the components $H_1^\nu, \dots, H_n^\nu$ of $\mathbf{H}^\nu$ are Gaussian and independent.
- The algorithm used for solving the optimization problem defined by (3.46) is the trial method.
- Equation (3.45) means that, once the optimal value $i_{\text{ref}}^{\text{opt}}$ of the level i_{ref} has been estimated by (3.46), then the partition $\mathbb{P}^\nu_{i_{\text{ref}}^{\text{opt}}}(m; \mu_1, \dots, \mu_m)$ of $\mathbf{H}^\nu$ in terms of random vectors $\mathbf{Y}^{\nu,1}, \dots, \mathbf{Y}^{\nu,m}$ that are assumed to be mutually independent for this optimal value $i_{\text{ref}}^{\text{opt}}$ of i_{ref}, is calculated and, consequently, the optimal values $m_{\max}$ and $\mu_1^{\text{opt}}, \dots, \mu_{m_{\max}}^{\text{opt}}$ such that $\mu_1^{\text{opt}} + \dots + \mu_{m_{\max}}^{\text{opt}} = n$ are deduced.

3.6. Algorithm for constructing the partition $\mathbb{P}^\nu_{i_{\text{ref}}}(m; \mu_1, \dots, \mu_m)$ of H^ν in terms of random vectors considered as independent for a given level i_{ref}. For i_{ref} such that $0 < i_{\text{ref}} < \max_{1 \leq r_1 < r_2 \leq n} i^\nu(H_{r_1}^\nu, H_{r_2}^\nu)$, an algorithm using the graph theory is proposed for constructing the partition $\mathbb{P}^\nu_{i_{\text{ref}}}(m; \mu_1, \dots, \mu_m)$ of $\mathbf{H}^\nu$ in terms of random vectors $\mathbf{Y}^{\nu,1}, \dots, \mathbf{Y}^{\nu,m}$ considered as independent for the given level i_{ref}. This algorithm, written using the Matlab language, is detailed in Algorithm 1 given in Appendix A, for which it is assumed that the symmetric adjacency $(n \times n)$ matrix $[\mathcal{N}]$ has previously been calculated using (3.43). The proposed algorithm uses notions of graph theory, such as the "nodes" and the "symmetric adjacency" for which the definition can be found in [7]. The algorithm that is detailed in Appendix A has especially been developed for constructing the partition by following the development presented in Section 3.5.3.

The CPU time generated by the proposed algorithm is totally due to the construction of the symmetric adjacency matrix $[\mathcal{N}]$, which requires the calculation of (3.35) to (3.41) (use of the classical Gaussian kernel density estimation method for computing a realization of the random mutual information). Consequently, the CPU time is proportional to $n \times \nu \times n(n+1)/2$ (dimension of the random vector) $\times$ (number of realizations) $\times$ (number of entries in the symmetric adjacency matrix) and to the number of constructions of matrix $[\mathcal{N}]$. The CPU time generated by the graph algorithm detailed in Appendix A is completely negligible. It should be noted that the sequences that generate the CPU time can easily be parallelized due to the algebraic structure of (3.39) and (3.41), and due to the fact that the computation of the entries of matrix $[\mathcal{N}]$ can also easily be parallelized.

4. Numerical experiments and numerical validation. Four numerical experiments are presented. For the first three, the construction of each numerical experiment is performed in order that

the optimal partition in terms of mutually independent random vectors is known *a priori*. Therefore, these numerical experiments yield a validation of the method proposed. For the last numerical experiment (the fourth), the optimal partition is unknown. Consequently, this numerical experiment does not directly contribute to the validation, but this example that is relative to a non-Gaussian $\mathbb{M}_6^+$-valued random field indexed by $\mathbb{R}^3$, allows for testing the capability of the method proposed to identify an optimal partition for a case in high dimension.

4.1. Numerical experiment 1. The generation of the set of realizations and the results are presented below.

(i) Generation of the set of realizations. This numerical experiment consists in analyzing a very simple case of "almost" independent uniform real-valued random variables.

- For $N = 10$, let $\mathbf{W} = (W_1, \ldots, W_N)$ be the random vector, defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$, whose components are written as $\mathbf{W} = \sqrt{3}\,(2\,\mathbf{U} - 1)$ in which the components $U_1, \ldots, U_N$ of random vector $\mathbf{U}$ are N mutually independent uniform random variables on $[0\,,1]$. For $\nu = 100$, let $\{\mathbf{U}(\theta_\ell), \ell = 1, \ldots, \nu\}$ be ν independent realizations of $\mathbf{U}$ with $\theta_\ell \in \Theta$. Let $\mathbf{W}(\theta_\ell) = \sqrt{3}\,(2\,\mathbf{U}(\theta_\ell) - 1)$ be the corresponding realizations of $\mathbf{W}$.
- We now defined the $\mathbb{R}^N$-valued random vector $\mathbf{X}^\nu$ from the set of its ν independent realizations $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ as explained in Section 2.1, such that $\mathbf{x}^{\exp,\ell} = \mathbf{W}(\theta_\ell)$.
- The reduced-order statistical model defined in Section 2.2 and the probabilistic construction of the random vector $\mathbf{H}^\nu$ given in Section 2.3 are applied in taking $n = N$ (thus $\mathrm{err}(n, \nu) = 0$), which allows for constructing the ν independent realizations $\{\boldsymbol{\eta}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ of the $\mathbb{R}^n$-valued random vector $\mathbf{H}^\nu$ such that (2.15) and (2.16) are verified.

It should be noted that, if ν went to $+\infty$, then the components of $\mathbf{H}^\nu$ would be mutually independent. However, since the value 100 for ν is not sufficiently large, the components of $\mathbf{H}^\nu$ are not exactly mutually independent. The test for the validation of the method that is proposed is to evaluate its capability to find the mutual independence of all the components of $\mathbf{H}^\nu$ for $\nu = 100$. As an illustration, Figure 4.1 (left) shows the graph of the pdf of the random variable H_3^ν, computed by using (3.38) with (3.39), which is compared to a Gaussian pdf with the same empirical mean value and the same standard deviation. Figure 4.1 (right) displays the graph of the joint pdf of the random variables (H_1^ν, H_3^ν) computed by using (3.40) with (3.41).

(ii) Results. The graph of function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$, defined by (3.44) is displayed in Figure 4.2 (left). The optimal level $i_{\mathrm{ref}}^{\mathrm{opt}}$ of i_{ref}, defined by (3.46), is $i_{\mathrm{ref}}^{\mathrm{opt}} = 0.02$. Figure 4.2 (right) displays the graph $p = (r_1, r_2) \mapsto i^\nu(H_{r_1}^\nu, H_{r_2}^\nu)$ defined by (3.35) and the horizontal line corresponding to the ordinate $i_{\mathrm{ref}}^{\mathrm{opt}} = 0.02$. The notation $p = (r_1, r_2)$ means that p is the integer belonging to $\{1, \ldots, n(n-1)/2\}$, which is associated with the pair (r_1, r_2) when r_1 and r_2 run through the ensemble $\{1 \leq r_1 < r_2 \leq n\}$. The multi-indices (r_1, r_2) are ordered such that for r_1 given by increasing value in $1 \leq r_1 \leq n - 1$, r_2 is ordered by increasing value such that $r_1 + 1 \leq r_2 \leq n$. This order is also used in the similar figures later in the paper. For instance, in Figure 4.2 (right), $p = 1$ corresponds to $r_1 = 1$ and $r_2 = 2$; $p = 9$, to $r_1 = 1$ and $r_2 = 10$; $p = 10$ to $r_1 = 2$ and $r_2 = 3$, etc. It can be seen that all the values of $i^\nu(H_{r_1}^\nu, H_{r_2}^\nu)$ are smaller than $i_{\mathrm{ref}}^{\mathrm{opt}}$ and therefore, taking into account (3.42), there are no dependent random variables $H_{r_1}^\nu$ with $H_{r_2}^\nu$ for all $1 \leq r_1 < r_2 \leq n$. We then obtained an optimal partition $\mathbb{P}_{\mathrm{opt}}^\nu = \mathbb{P}(m_{\max}; \mu_1^{\mathrm{opt}}, \ldots, \mu_{m_{\max}}^{\mathrm{opt}})$ with $m_{\max} = n = N = 10$ and $\mu_1^{\mathrm{opt}} = \ldots = \mu_{m_{\max}}^{\mathrm{opt}} = 1$.
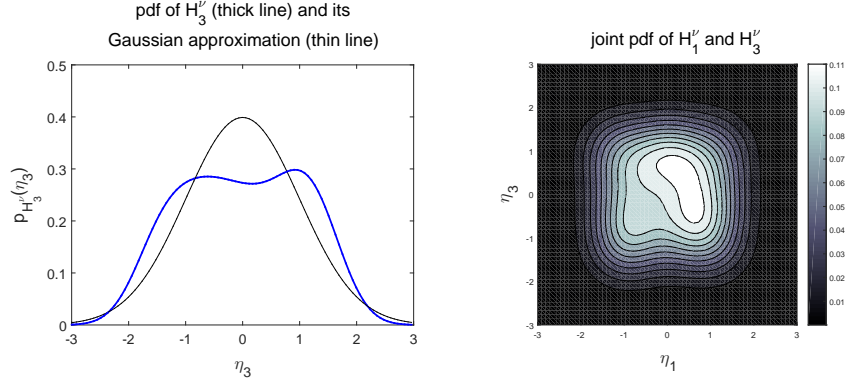
Figure 4.1. *Numerical experiment 1. Graph of $\eta_3 \mapsto p_{\mathrm{H}_3^\nu}(\eta_3)$ (thick line) compared to the graph of a Gaussian pdf with the same empirical mean value and with the same standard deviation as H_3^ν (thin line) (left figure) and graph of $(\eta_1, \eta_3) \mapsto p_{\mathrm{H}_1^\nu, \mathrm{H}_3^\nu}(\eta_1, \eta_3)$ (right figure).*
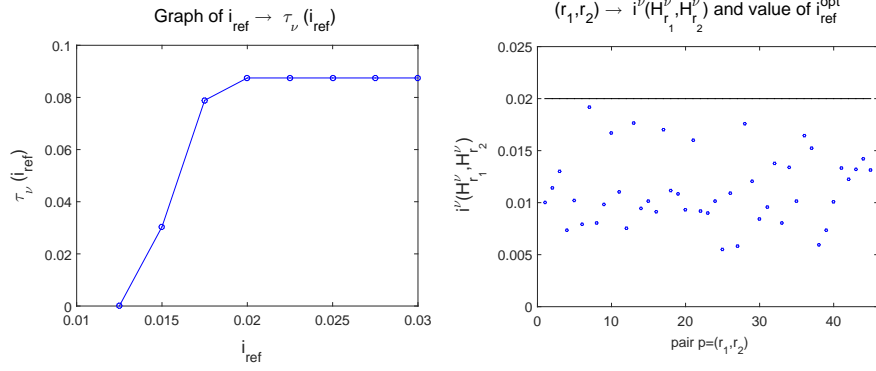

Figure 4.2. *Numerical experiment 1. Graph of function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$ (left figure) and graph of $p = (r_1, r_2) \mapsto i^\nu(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ (symbols) and value of $i_{\mathrm{ref}}^{\mathrm{opt}}$ (horizontal line) (right figure).*

Consequently, the optimal partition identified corresponds to $n = 10$ mutually independent real-valued random variables $Y^{\nu,1}, \ldots, Y^{\nu,m}$ with $Y^{\nu,j} = \mathrm{H}_j^\nu$, that is to say, all the components of random vector $\mathbf{H}^\nu$ are found as mutually independent for $\nu = 100$ (that was expected). For this optimal partition, the criterion defined by (3.30) is verified with $i^\nu(Y^{\nu,1}, \ldots, Y^{\nu,m_{\max}}) = 4.83$ and $i^\nu(G^1, \ldots, G_{\max}^m) = 5.29$. The total CPU time for obtaining the results with a laptop computer is 7 seconds.

4.2. Numerical experiment 2. For this numerical experiment, we test the capability of the method proposed to identify an optimal partition of a random vector having 78 non-Gaussian components, in terms of 12 mutually independent non-Gaussian random vectors of different lengths.

(i) Generation of the set of realizations. For this numerical example, we consider a random vector with $N = 78$ components, which is constituted of $m = 12$ independent random vectors with length $\mu_j = j$ for $j = 1, \ldots, m$. The generation is performed as follows.

- Let $\Xi_1, \ldots, \Xi_m$ be m mutually independent real-valued random variables (the random germs)

such that $\Xi_j = (2 U_j - 1)/4$ in which $U_1, \ldots, U_m$ are m mutually independent uniform random variables on $[0, 1]$, defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$. The number of independent realizations in the set that we are constructing is $\nu = 100$. For $\ell = 1, \ldots, \nu$, the ℓ-th independent realization of $\{\Xi_1, \ldots, \Xi_m\}$ is denoted by $\{\xi_1^\ell, \ldots, \xi_m^\ell\}$.

- For j fixed in $\{1, \ldots, m\}$, we define the random vector $\mathbf{Z}^j = (Z_1^j, \ldots, Z_{\mu_j}^j)$ with $\mu_j = j$ such that, for all $k = 1, \ldots, \mu_j$, $Z_k^j = \sum_{i=0}^k \sqrt{i!} \, (\Xi_j)^i$. Let $\widehat{\mathbf{m}}_{\mathbf{Z}}^j \in \mathbb{R}^{\mu_j}$ and $[\widehat{C}_{\mathbf{Z}}^j] \in \mathbb{M}_{\mu_j}$ be the empirical estimates of the mean value and of the covariance matrix of $\mathbf{Z}^j$, computed by using the ν independent realizations $\xi_j^1, \ldots, \xi_j^\nu$ of Ξ_j. Let $[L_{\mathbf{Z}}^j]$ be the upper triangular matrix such that $[\widehat{C}_{\mathbf{Z}}^j] = [L_{\mathbf{Z}}^j]^T [L_{\mathbf{Z}}^j]$.

- We introduce the normalized $\mathbb{R}^{\mu_j}$-random vector $\mathbf{W}^j$ such that $\mathbf{W}^j = [L_{\mathbf{Z}}^j]^{-T} (\mathbf{Z}^j - \widehat{\mathbf{m}}_{\mathbf{Z}}^j)$ for which we compute the ν independent realizations $\{\mathbf{w}^{j,\ell} \in \mathbb{R}^{\mu_j}, \ell = 1, \ldots, \nu\}$. The empirical estimates of the mean value $\widehat{\mathbf{m}}_{\mathbf{W}}^j \in \mathbb{R}^{\mu_j}$ and of the covariance matrix $[\widehat{C}_{\mathbf{W}}^j] \in \mathbb{M}_{\mu_j}$ of random vector $\mathbf{W}^j$, computed by using $\{\mathbf{w}^{j,\ell}, \ell = 1, \ldots, \nu\}$ are such that $\widehat{\mathbf{m}}_{\mathbf{W}}^j = \mathbf{0}_{\mu_j}$ and $[\widehat{C}_{\mathbf{W}}^j] = [I_{\mu_j}]$.

- We now introduce the $\mathbb{R}^N$-random vector $\mathbf{W} = (\mathbf{W}^1, \ldots, \mathbf{W}^m)$ and we denote by $\{\mathbf{w}^\ell, \ell = 1, \ldots, \nu\}$ its ν independent realizations $\mathbf{w}^\ell = (\mathbf{w}^{1,\ell}, \ldots, \mathbf{w}^{m,\ell}) \in \mathbb{R}^N$.

- The $\mathbb{R}^N$-valued random vector $\mathbf{X}^\nu$ is now defined by its ν independent realizations $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ as explained in Section 2.1 such that $\mathbf{x}^{\exp,\ell} = \mathbf{w}^\ell$.

- Let $\widehat{\mathbf{m}}_{\mathbf{X}}^\nu \in \mathbb{R}^N$ and $[\widehat{C}_{\mathbf{X}}^\nu] \in \mathbb{M}_N$ be the empirical estimates of the mean value and of the covariance matrix of $\mathbf{X}^\nu$, computed by using $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ (see (2.1)). We thus have $\widehat{\mathbf{m}}_{\mathbf{X}}^\nu = \mathbf{0}_N$, but although $[\widehat{C}_{\mathbf{W}}^j] = [I_{\mu_j}]$ for all $j = 1, \ldots, m$, since the value 100 of ν is not very big, the statistical estimate $[\widehat{C}_{\mathbf{X}}^\nu]$ of the covariance matrix of $\mathbf{X}^\nu$ is not equal to the identity matrix and we have only $[\widehat{C}_{\mathbf{X}}^\nu] \simeq [I_N]$ (the diagonal blocks are effectively $[I_{\mu_1}], \ldots, [I_{\mu_m}]$, but the entries of the extra-diagonal blocks are not exactly zero).

- Despite the fact that $[\widehat{C}_{\mathbf{X}}^\nu]$ is not exactly equal to $[I_N]$, in order to preserve the mutual independence for the m random vectors that we are generating, the reduced-order statistical model defined in Section 2.2 is not applied and we define the random vector $\mathbf{H}^\nu$ such that $\mathbf{H}^\nu = \mathbf{X}^\nu$ with $n = N$, and therefore, the ν independent realizations $\{\boldsymbol{\eta}^{\exp,\ell} \in \mathbb{R}^n, \ell = 1, \ldots, \nu\}$ of $\mathbf{H}^\nu$ are such that such that (2.15) is verified but (2.16) is approximatively verified.

As an illustration, Figure 4.3 shows the graph of the pdf for two components of $\mathbf{H}^\nu$, computed by using (3.38) with (3.39), which is compared to a Gaussian pdf with the same empirical mean value and with the same standard deviation. Figure 4.4 displays the graph of the joint pdf of the random variables $(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ for two pairs of indices, computed by using (3.40) with (3.41). Note that the first one (left figure) corresponds to two dependent components belonging to subset $j = 6$ while the second pair corresponds to two independent components that belong to subsets $j = 9$ and $j = 11$ (see (4.1)). Therefore, the construction proposed allows for writing the following partition $\mathbf{H}^\nu = (\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m})$ in which $\mathbf{Y}^{\nu,j}$ is a random vector with values in $\mathbb{R}^{\mu_j}$ with $\mu_j = j$. The ν independent realizations of $\mathbf{Y}^{\nu,j}$ are $\{\boldsymbol{\eta}^{j,\exp,\ell}, \ell = 1, \ldots, \nu\}$ such that $\boldsymbol{\eta}^{\exp,\ell} = (\boldsymbol{\eta}^{1,\exp,\ell}, \ldots, \boldsymbol{\eta}^{m,\exp,\ell})$ with $\boldsymbol{\eta}^{j,\exp,\ell} = (\eta_1^{j,\exp,\ell}, \ldots, \eta_{\mu_j}^{j,\exp,\ell}) = (\eta_{r_1^j}^{\exp,\ell}, \ldots, \eta_{r_{\mu_j}^j}^{\exp,\ell}) \in \mathbb{R}^{\mu_j}$. With the construction proposed, if ν was sufficiently big, then the number of mutually independent random vectors would be surely $m = 12$ and, for $j = 1, \ldots, m$, the number of components of $\mathbf{Y}^{\nu,j} = (\mathrm{H}_{r_1^j}^\nu, \ldots, \mathrm{H}_{r_{\mu_j}^j}^\nu)$ would be surely $\mu_j = j$. Let $\{[\mathcal{S}]_{jk}\}_{jk}$ be the $(m \times m)$ matrix of integers for which only the entries $1 \leq j \leq k \leq m$ are defined and such that $[\mathcal{S}]_{jk} = r_k^j$ where $r_1^j, \ldots, r_{\mu_j}^j$ are the indices of the mutually dependent components
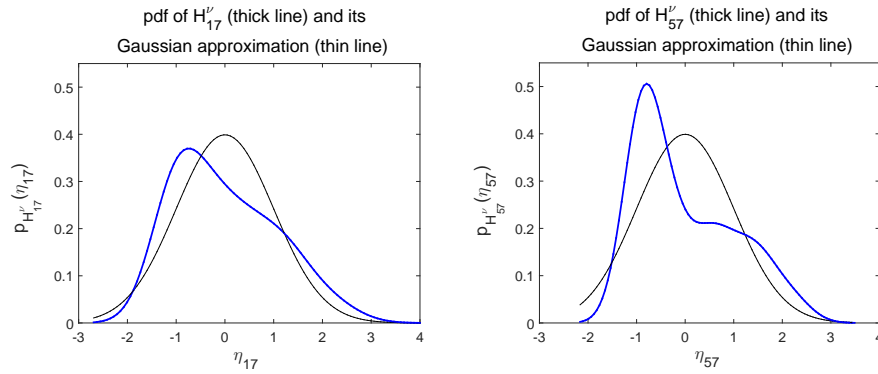
Figure 4.3. *Numerical experiment 2. Graph of $\eta_j \mapsto p_{H_j^\nu}(\eta_j)$ (thick line) compared to the graph of a Gaussian pdf with the same empirical mean value and with the same standard deviation as H_j^ν (thin line), for $j = 17$ (left figure) and $j = 57$ (right figure).*
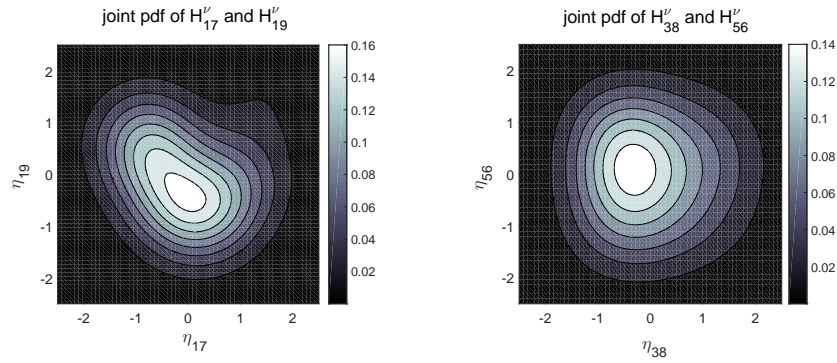

Figure 4.4. *Numerical experiment 2. Graph of the joint pdf $(\eta_{r_1}, \eta_{r_2}) \mapsto p_{H_{r_1}^\nu, H_{r_2}^\nu}(\eta_{r_1}, \eta_{r_2})$ for $r_1 = 17$ and $r_2 = 19$ (left figure), $r_1 = 38$ and $r_2 = 56$ (right figure).*

$H_{r_1^j}^\nu, \ldots, H_{r_{\mu_j}^j}^\nu$ of $\mathbf{H}^\nu$. The generation proposed yields (if ν is sufficiently large) the following mutually independent random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$. For $j = 1, \ldots, m$, the j-th row in (4.1) yields the indices

of the dependent random variables $H^\nu_{r^j_1}, \ldots, H^\nu_{r^j_{\mu_j}}$ that are the components of $\mathbf{Y}^{\nu,j}$,

$$[\mathcal{S}] = \begin{bmatrix} 1 \\ 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 & 21 \\ 22 & 23 & 24 & 25 & 26 & 27 & 28 \\ 29 & 30 & 31 & 32 & 33 & 34 & 35 & 36 \\ 37 & 38 & 39 & 40 & 41 & 42 & 43 & 44 & 45 \\ 46 & 47 & 48 & 49 & 50 & 51 & 52 & 53 & 54 & 55 \\ 56 & 57 & 58 & 59 & 60 & 61 & 62 & 63 & 64 & 65 & 66 \\ 67 & 68 & 69 & 70 & 71 & 72 & 73 & 74 & 75 & 76 & 77 & 78 \end{bmatrix}. \tag{4.1}$$

For instance, for $j = 4$, the indices of $\mathbf{H}^\nu$ that constitute the components of $\mathbf{Y}^{4,\nu}$ are 7, 8, 9, and 10, which means that $\mathbf{Y}^{4,\nu} = (H^\nu_7, H^\nu_8, H^\nu_9, H^\nu_{10})$. As previously commented, It should be noted that, if ν was sufficiently big, then we should find as the optimal partition of $\mathbf{H}^\nu$, the one defined by (4.1). Nevertheless, as the value 100 for ν is not sufficiently big, the random vectors $\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m}$ defined above are not exactly mutually independent. In the framework of the validation of the method proposed, this numerical experiment consists in evaluating the capability of the method to identify the optimal partition defined by (4.1) with a small value of ν ($\nu = 100$). Note that the $[\mathcal{S}]$ matrix obtained with the optimal partitioning using the proposed algorithm is exactly the one defined by (4.1).
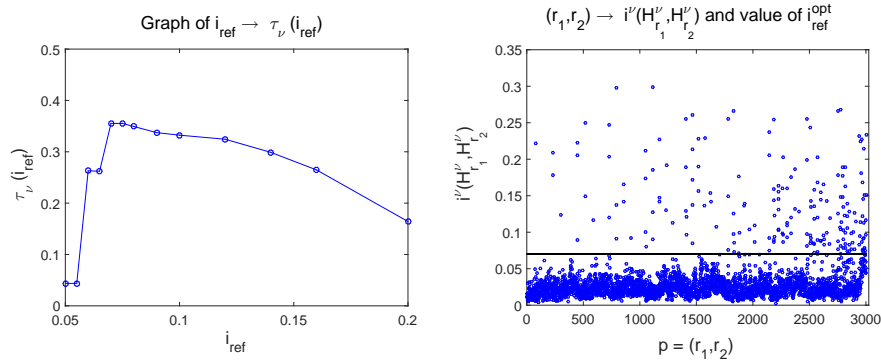

Figure 4.5. *Numerical experiment 2. Graph of function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$ (left figure) and graph of $p = (r_1, r_2) \mapsto i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$ (symbols) and value of $i^{\mathrm{opt}}_{\mathrm{ref}}$ (horizontal line) (right figure).*

(ii) Results. The graph of function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$, defined by (3.44), is displayed in Figure 4.5 (left). The optimal level $i^{\mathrm{opt}}_{\mathrm{ref}}$ of i_{ref}, defined by (3.46), is $i^{\mathrm{opt}}_{\mathrm{ref}} = 0.07$. Figure 4.5 (right) displays the graph $p = (r_1, r_2) \mapsto i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$ defined by (3.35) and the horizontal line corresponds to the ordinate $i^{\mathrm{opt}}_{\mathrm{ref}} = 0.07$. The optimal partition $\mathbb{P}^\nu_{\mathrm{opt}} = \mathbb{P}(m_{\max}; \mu^{\mathrm{opt}}_1, \ldots, \mu^{\mathrm{opt}}_{m_{\max}})$ that is identified with the method

proposed yields $m_{\max} = 12$ and $\mu_j^{\mathrm{opt}} = j$ with $j = 1, \ldots, 12$. Consequently, the partition defined by (4.1) is identified by the method proposed for $\nu = 100$ (as expected). For this optimal partition, the criterion defined by (3.30) is verified with $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m_{\max}}) = 23.87$ and $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m_{\max}) = 37.02$. The total CPU time for obtaining the results with a laptop computer is 105 seconds.

4.3. Numerical experiment 3. This numerical experiment is the same as numerical experiment 2, except that the uniform random germ is replaced by a Gaussian random germ, but, obviously, random vector $\mathbf{H}^\nu$ is not Gaussian. This experiment is carried out in order to analyze the sensitivity of the method proposed with respect to the probability distribution of the random vector defined by the set of realizations. The reader will see that the graph of the function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$, the graph of the function $p = (r_1, r_2) \mapsto i^\nu(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$, and the value $i_{\mathrm{ref}}^{\mathrm{opt}}$ are different from the results obtained with the numerical experiment 2 (that is normal because the probability distribution is not the same) but, however, the method proposed gives the solution, which is the same as the one for numerical experiment 2.
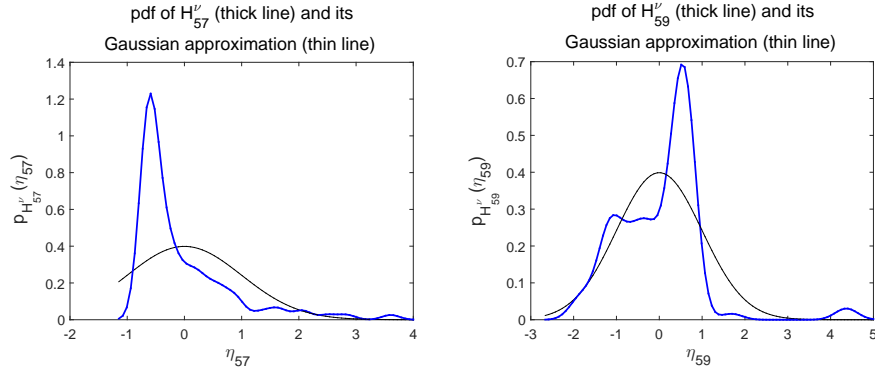

Figure 4.6. *Numerical experiment 3. Graph of $\eta_j \mapsto p_{\mathrm{H}_j^\nu}(\eta_j)$ (thick line) compared to the graph of a Gaussian pdf with the same empirical mean value and with the same standard deviation as H_j^ν (thin line), for $j = 57$ (left figure) and $j = 59$ (right figure).*

(i) Generation of the set of realizations. For this numerical example, we consider the generation presented in Section 4.2 for which $\Xi_1, \ldots, \Xi_m$ are m mutually independent real-valued normalized Gaussian random variables (centered and unit variance). All the other parameters are those of numerical experiment 2. As an illustration, Figure 4.6 displays the graph of the pdf for two components of $\mathbf{H}^\nu$, computed by using (3.38) with (3.39), which is compared to a Gaussian pdf with the same empirical mean value and with the same standard deviation. Figure 4.7 displays the graph of the joint pdf of the random variables $(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ for two pairs of indices, computed by using (3.40) with (3.41). Note that the first one (left figure) corresponds to two dependent components belonging to subset $j = 11$ while the second pair corresponds to two independent components that belong to subsets $j = 5$ and $j = 8$ (see (4.1)).

(ii) Results. The graph of function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$, defined by (3.44), is displayed in Figure 4.8 (left). The optimal level $i_{\mathrm{ref}}^{\mathrm{opt}}$ of i_{ref}, defined by (3.46), is $i_{\mathrm{ref}}^{\mathrm{opt}} = 0.13$. Figure 4.8 (right) displays the graph
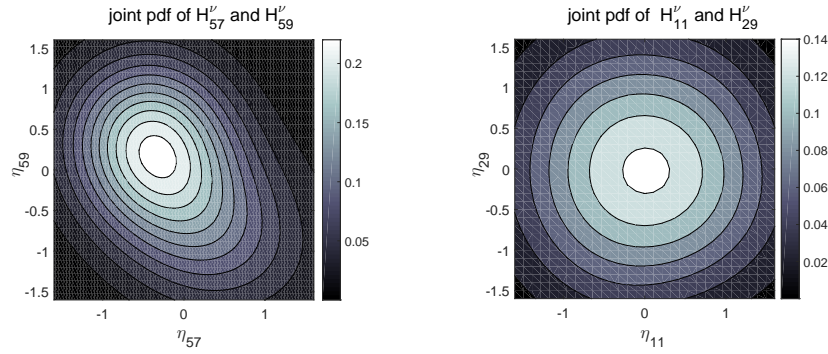
Figure 4.7. *Numerical experiment 3. Graph of the joint pdf $(\eta_{r_1}, \eta_{r_2}) \mapsto p_{H^\nu_{r_1}, H^\nu_{r_2}}(\eta_{r_1}, \eta_{r_2})$ for $r_1 = 57$ and $r_2 = 59$ (left figure), $r_1 = 11$ and $r_2 = 29$ (right figure).*
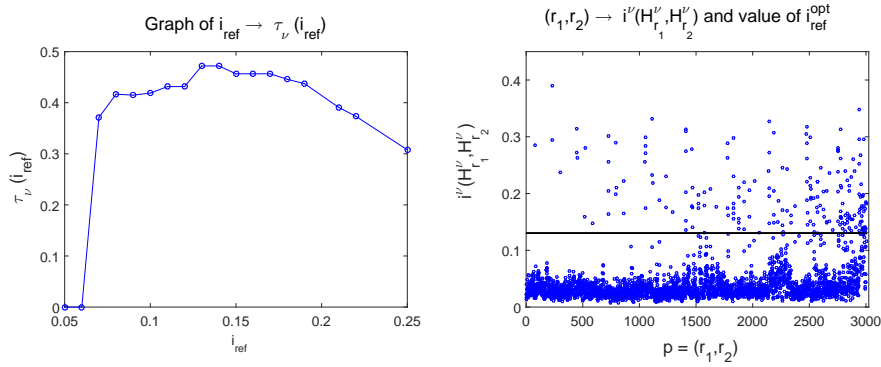

Figure 4.8. *Numerical experiment 3. Graph of function $i_{\rm ref} \mapsto \tau_\nu(i_{\rm ref})$ (left figure) and graph of $p = (r_1, r_2) \mapsto i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$ (symbols) and value of $i^{\rm opt}_{\rm ref}$ (horizontal line) (right figure).*

$p = (r_1, r_2) \mapsto i^\nu(H^\nu_{r_1}, H^\nu_{r_2})$ defined by (3.35) and the horizontal line corresponding to the ordinate $i^{\rm opt}_{\rm ref} = 0.13$. The optimal partition $\mathbb{P}^\nu_{\rm opt} = \mathbb{P}(m_{\max}; \mu^{\rm opt}_1, \ldots, \mu^{\rm opt}_{m_{\max}})$ that is identified with the method proposed yields $m_{\max} = 12$ and $\mu^{\rm opt}_j = j$ with $j = 1, \ldots, 12$. Consequently, the partition defined by (4.1) is identified by the method proposed for $\nu = 100$ (as expected). For this optimal partition, the criterion defined by (3.30) is verified with $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, {}_b f Y^{\nu, m_{\max}}) = 19.55$ and $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^m_{\max}) = 37.02$. The total CPU time for obtaining the results with a laptop computer is 140 seconds.

4.4. Numerical experiment 4. For this numerical experiment, there are no references that allow for validating the optimal partition identified by the method proposed, but the example is in high dimension. The data set, which defines the random vector $\mathbf{X}^\nu$ that has $N = 27,951$ components, is constituted of $\nu = 1,000$ independent realizations $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ with $\mathbf{x}^{\exp,\ell} \in \mathbb{R}^N$. This random vector is generated as the spatial discretization of a non-Gaussian positive-definite matrix-valued random field.

(i) Generation of the set of realizations. We begin by defining the stochastic model of the random field and then we explain how the set of realizations is constructed.

- Let $\{[\mathbb{G}(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$ be the non-Gaussian second-order random field, defined on the probability space $(\Theta, \mathcal{T}, \mathcal{P})$, indexed by $\mathbb{R}^3$, with values in $\mathbb{M}_6^+$, which is completely defined in [52] (in this reference, this field $[\mathbb{G}]$ is denoted as $[\mathbf{G}_n]$). The stochastic model of this field depends on hyperparameters that are (1) a dispersion parameter δ that allows for controlling the level of statistical fluctuations of the random field, and (2) three lengths L_1, L_2, and L_3 that allow for controlling the three spatial correlation lengths of the random field, associated with the three coordinates $\zeta_1, \zeta_2, \zeta_3$ of $\boldsymbol{\zeta}$. In the present numerical experiment, $\delta = 0.7$ and $L_1 = L_2 = L_3 = L_{\text{corr}}$ for which three values $L_{\text{corr}} = 0.2, 0.5$, and 1.0 will be considered.
- Let $\{[\mathbb{G}^{\log}(\boldsymbol{\zeta})], \boldsymbol{\zeta} \in \mathbb{R}^3\}$ be the non-Gaussian second-order random field, indexed by $\mathbb{R}^3$, with values in $\mathbb{M}_6^S$, such that, for all $\boldsymbol{\zeta}$ in $\mathbb{R}^3$, $[\mathbb{G}^{\log}(\boldsymbol{\zeta})] = \log_{\mathbb{M}}([\mathbb{G}(\boldsymbol{\zeta})])$ in which $\log_{\mathbb{M}}$, which is defined on $\mathbb{M}_6^+$ with values in $\mathbb{M}_6^S$, is the reciprocity mapping of the exponential mapping $\exp_{\mathbb{M}}$ from $\mathbb{M}_6^S$ into $\mathbb{M}_6^+$.
- Let Ω be the open bounded domain of $\mathbb{R}^3$ defined by $\Omega = (]0, 1[)^3$. We consider a regular Cartesian grid of Ω made up of $N_p = 11 \times 11 \times 11 = 1,331$ nodes with constant steps for each one of the three coordinates. Let $\boldsymbol{\zeta}^1, \ldots, \boldsymbol{\zeta}^{N_p}$ be the N_p points in $\mathbb{R}^3$ defining the nodes of this grid. For $n_p = 1, \ldots, N_p$, the 21 entries $\{[\mathbb{G}^{\log}(\boldsymbol{\zeta}^{n_p})]_{kk'}, 1 \leq k \leq k' \leq 6\}$ are reshaped in a $\mathbb{R}^{21}$-valued random vector $\mathbb{G}_{\text{resh}}^{\log}(\boldsymbol{\zeta}^{n_p})$. Let $N = N_p \times 21 = 27,951$. We then introduce the $\mathbb{R}^N$-valued random vector $\mathbf{G}_{\text{resh}}^{\log}$ such that $\mathbf{G}_{\text{resh}}^{\log} = (\mathbb{G}_{\text{resh}}^{\log}(\boldsymbol{\zeta}^1), \ldots, \mathbb{G}_{\text{resh}}^{\log}(\boldsymbol{\zeta}^{N_p}))$. The random generator presented in [52] allows for constructing the $\nu = 1,000$ independent realizations $\{\mathbf{g}^\ell, \ell = 1, \ldots, \nu\}$ with $\mathbf{g}^\ell \in \mathbb{R}^N$ of the non-Gaussian random vector $\mathbf{G}_{\text{resh}}^{\log}$.
- Let $\widehat{\mathbf{m}}_{\mathbf{G}}^\nu$ be the empirical mean value of $\mathbf{G}_{\text{resh}}^{\log}$ computed by using $\{\mathbf{g}^\ell, \ell = 1, \ldots, \nu\}$. We then introduced the ν centered realizations $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ such that $\mathbf{x}^{\exp,\ell} = \mathbf{g}^\ell - \widehat{\mathbf{m}}_{\mathbf{G}}^\nu$. We now defined the $\mathbb{R}^N$-valued random vector $\mathbf{X}^\nu$ from the set of the ν independent realizations $\{\mathbf{x}^{\exp,\ell}, \ell = 1, \ldots, \nu\}$ as explained in Section 2.1, such that its independent realizations are $\mathbf{X}^\nu(\theta_\ell) = \mathbf{x}^{\exp,\ell}$. Note that the empirical mean value of $\mathbf{X}^\nu$ is $\mathbf{0}_N$.
- For each value of L_{corr}, a reduced-order statistical model $\mathbf{X}^{(n,\nu)}$ of $\mathbf{X}^\nu$ is carried out as explained in Section 2.2. Taking into account the dimension N, the empirical covariance matrix $[\widehat{C}_{\mathbf{X}}^\nu]$ defined by (2.1) with $\widehat{\mathbf{m}}_{\mathbf{X}}^\nu = \mathbf{0}_N$ is not computed and the reduced-statistical model is constructed by using the method derived from [1, 22, 32], which is based on a "thin SVD" [28] of the rectangular $(N \times \nu)$ matrix $[\mathbf{x}^{\exp,1} \ldots \mathbf{x}^{\exp,\nu}]$. For the three values of L_{corr}, Figure 4.9 (left) displays the graph of the function $n \mapsto \text{err}(n, \nu)$ defined by (2.6). The value n of the reduced-order statistical model is chosen in order that the relative error $\text{err}(n, \nu)$ is equal to 0.99, which yields $n = 706, 245$, and 77 for $L_{\text{corr}} = 0.2, 0.5$, and 1.0, respectively.

As an illustration, for $L_{\text{corr}} = 1.0$, Figure 4.9 (central and right) displays the graph of the pdf for two components of $\mathbf{H}^\nu$, computed by using (3.38) with (3.39), which is compared to a Gaussian pdf with the same empirical mean value and the same standard deviation. Figure 4.10 displays the graph of the joint pdf of the random variables $(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ for two pairs of indices, computed by using (3.40) with (3.41). The optimal partition obtained, for which the results are presented hereinafter, shows that the first one (left figure) corresponds to two dependent components H_{17}^ν and H_{33}^ν that belong to subset $j = 14$ while the second pair corresponds to two independent components H_{48}^ν and H_{56}^ν that belong to subsets $j = 34$ and $j = 3$ (see figure 4.12-right).
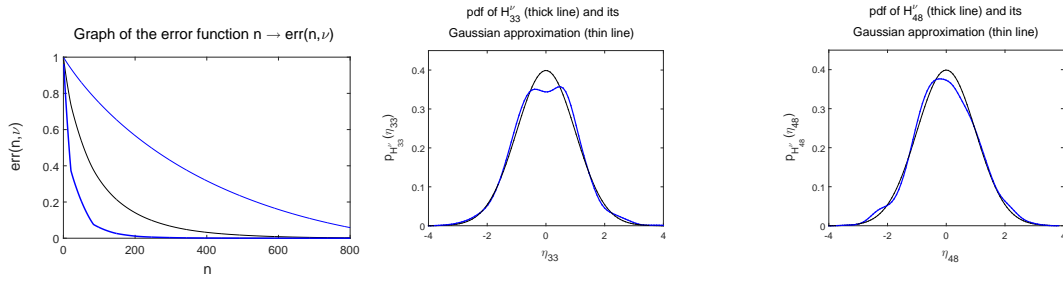
Figure 4.9. *Numerical experiment 4. Graph of the relative error function $n \mapsto err(n,\nu)$ defined by (2.6) for the three values of L_{corr} (left figure): 0.2 (blue thin line), 0.5 (black thin line), and 1.0 (thick blue line). For $L_{\mathrm{corr}} = 1.0$, graph of $\eta_j \mapsto p_{\mathrm{H}_j^\nu}(\eta_j)$ (thick line) compared to the graph of a Gaussian pdf with the same empirical mean value and with the same standard deviation as H_j^ν (thin line), for $j = 33$ (central figure) and $j = 48$ (right figure).*

(ii) Results. For $L_{\mathrm{corr}} = 0.2$, 0.5, and 1.0, figure 4.11 (left) displays the graph of the function $i_{\mathrm{ref}} \mapsto \tau_\nu(i_{\mathrm{ref}})$, defined by (3.44). As an illustration, for $L_{\mathrm{corr}} = 1.0$, Figure 4.11 (right) displays the graph $p = (r_1, r_2) \mapsto i^\nu(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ defined by (3.35) and the horizontal line corresponding to the ordinate $i_{\mathrm{ref}}^{\mathrm{opt}} = 0.0095$. For each one of the value of $L_{\mathrm{corr}} \in \{0.2, 0.5, 1.0\}$, we have
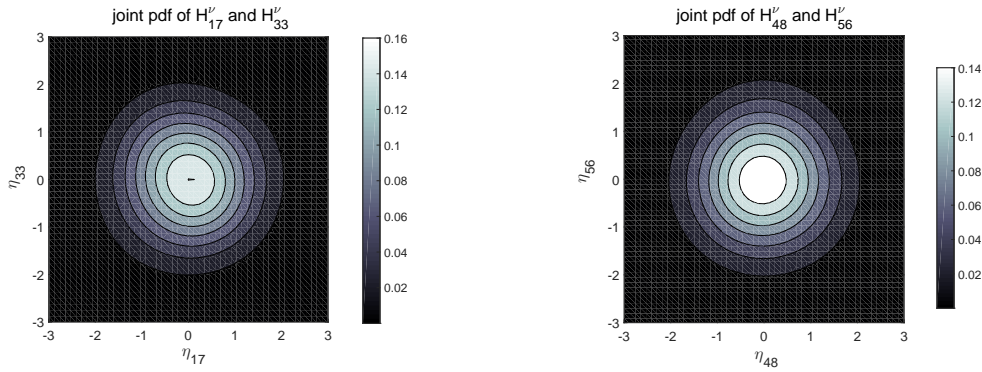

Figure 4.10. *Numerical experiment 4. For $L_{\mathrm{corr}} = 1.0$, graph of the joint pdf $(\eta_{r_1}, \eta_{r_2}) \mapsto p_{\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu}(\eta_{r_1}, \eta_{r_2})$ for $r_1 = 17$ and $r_2 = 33$ (left figure) and $r_1 = 48$ and $r_2 = 56$ (right figure).*

Table 4.1

Results obtained for numerical experiment 4

L_{corr}	n	$i_{\mathrm{ref}}^{\mathrm{opt}}$	$m_{\max}$	r	$i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m_{\max}})$	$i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^{m_{\max}})$
0.2	706	0.0084	563	0.7975	521.04	521.71
0.5	245	0.0082	187	0.7633	183.54	183.86
1.0	77	0.0095	51	0.6623	56.05	56.31

- Table 4.1 yields the dimension n of $\mathbf{H}^\nu$, the optimal level $i_{\mathrm{ref}}^{\mathrm{opt}}$ of i_{ref}, defined by (3.46), the number $m_{\max}$ of mutually independent random vectors for the optimal partition $\mathbb{P}_{\mathrm{opt}}^\nu = \mathbb{P}(m_{\max}; \mu_1^{\mathrm{opt}}, \ldots, \mu_{m_{\max}}^{\mathrm{opt}})$, the rate r of mutually independent random vectors defined as $r =$

$m_{\max}/n$, and for the optimal partition, the values of $i^\nu(\mathbf{Y}^{\nu,1}, \ldots, \mathbf{Y}^{\nu,m_{\max}})$ and $i^\nu(\mathbf{G}^1, \ldots, \mathbf{G}^{m_{\max}})$ that are used for the criterion defined by (3.30). It can be seen that rate r of independent random vectors increases when the correlation length L_{corr} decreases, which is physically coherent;

- the optimal partition defined by Figure 4.12 that displays the graph $j \mapsto \{[\mathcal{S}]_{jk} = r_k^j, 1 \le k \le \mu_j^{\text{opt}}\}$ in which $r_1^j, \ldots, r_{\mu_j^{\text{opt}}}^j$ are the indices of the mutually dependent components $\mathrm{H}_{r_1^j}^\nu, \ldots,$ $\mathrm{H}_{r_{\mu_j^{\text{opt}}}^j}^\nu$ of $\mathbf{H}^\nu$, which belong to the subset j (see (3.2)).

For $L_{\text{corr}} = 0.2$, for which $n = 706$ and $\nu = 1000$, the CPU time with one core for constructing one symmetric adjacency matrix is about 500 minutes, but as explained in Section 3.6, with J cores, the elapsed time would be $500/J$ minutes (for the presented applications, the computation has be carried out with $J = 12$).

The results presented in table 4.1 and in figure 4.12 show that a very important gain would be obtained for the construction of a stochastic representation of such a non-Gaussian tensor-valued random field for which a set of its realizations would be given. For instance, instead of directly identifying the coefficients of its PCE from the set $\{\boldsymbol{\eta}^{\text{exp},\ell} \in \mathbb{R}^n, \ell = 1, \ldots, \nu\}$ of its realizations, this preprocessing of the data set allows for identifying, independently, for each j fixed in $\{1, \ldots, m\}$, the coefficients of the polynomial chaos expansion of the random vector $\mathbf{Y}^{\nu,j}$ with values in $\mathbb{R}^{\mu_j}$ for which $\mu_j < n$, using solely its realizations $\{\boldsymbol{\eta}^{j,\text{exp},\ell} \in \mathbb{R}^{\mu_j}, \ell = 1, \ldots, \nu\}$ that are directly extracted from the set $\{\boldsymbol{\eta}^{\text{exp},\ell} \in \mathbb{R}^n, \ell = 1, \ldots, \nu\}$ as soon as the optimal partition of $\mathbf{H}^\nu$ has been identified with the algorithm presented. Consequently, the methods proposed, for instance in [2, 5, 47, 53], could be reused with a larger efficiency.
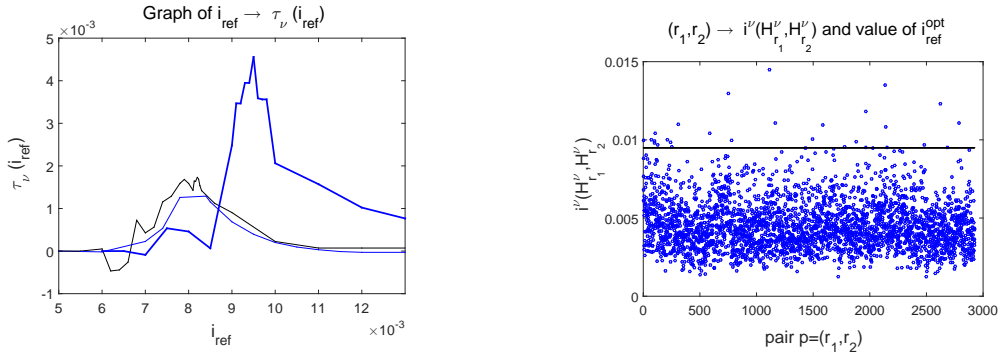

Figure 4.11. *Numerical experiment 4. Graph of function $i_{\text{ref}} \mapsto \tau_\nu(i_{\text{ref}})$ (left figure) for $L_{\text{corr}} = 0.2$ (blue thin line), $= 0.5$ (black thin line), $= 1.0$ (blue thick line) and, for $L_{\text{corr}} = 1.0$, graph of $p = (r_1, r_2) \mapsto i^\nu(\mathrm{H}_{r_1}^\nu, \mathrm{H}_{r_2}^\nu)$ (symbols) and value of $i_{\text{ref}}^{\text{opt}}$ (horizontal line) (right figure).*

Note that the partition that is constructed is relative to vector $\mathbf{H}^\nu$, for which the components are non-Gaussian, centered, and uncorrelated (but statistically dependent) random variables that are related to the representation of the $\mathbb{M}_6^S$-valued random field $[\mathbb{G}^{\log}]$ such that $[\mathbb{G}^{\log}(\boldsymbol{\zeta})] = \log_{\mathbb{M}}([\mathbb{G}(\boldsymbol{\zeta})])$, and
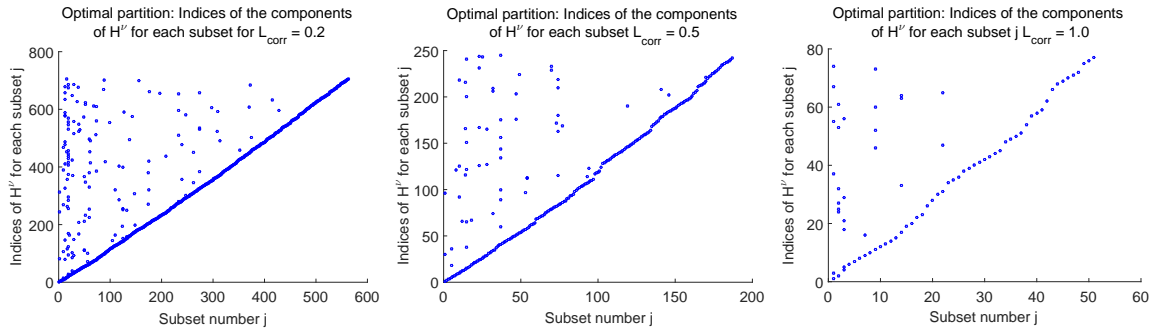
Figure 4.12. *Numerical experiment 4. Graph of $j \mapsto \{[\mathcal{S}]_{jk} = r_k^j, 1 \leq k \leq \mu_j^{\mathrm{opt}}\}$ corresponding to the optimal partition for $L_{\mathrm{corr}} = 0.2$ (left figure), $L_{\mathrm{corr}} = 0.5$ (central figure), $L_{\mathrm{corr}} = 1.0$ (right figure).*

not related to the $\mathbb{M}_6^+$-valued random field $[\mathbb{G}]$ such that $[\mathbb{G}(\zeta)] = \exp_{\mathbb{M}}([\mathbb{G}^{\log}(\zeta)])$ and which is a strongly non-Gaussian $\mathbb{M}_6^+$-valued random field (see [52]). This means that, even if all the uncorrelated random components of $\mathbf{H}^\nu$ were mutually independent (and consequently, $\mathbf{H}^\nu$ would be a Gaussian random vector), the components of the non-Gaussian random vector $\mathbf{X}^\nu$ related to the $\mathbb{M}_6^+$-valued random field $[\mathbb{G}]$ would be dependent. As shown in Figure 4.9 (central and right) and in Figure 4.10, some components of random vector $\mathbf{H}^\nu$ are relatively close to Gaussian random variables but are not Gaussian. This is a reasonable explanation for justifying the relatively large values of parameter r given in Table 4.1, but again, it should be noted the coherence of the results obtained, because r decreases when the spatial correlation length of the random field increases.

5. Conclusions. In this paper, we have proposed a fast algorithm based on the use of mutual entropy from information theory and on the use of graph theory for constructing an optimal partition, in terms of mutually independent random vectors, of the components of a non-Gaussian random vector that is only defined by a given set of realizations. The method has especially been developed for random vectors in high dimension and for which the number of realizations that constitute the data set can be small. The method proposed and its objective are different from the ICA that was introduced to extract independent source signals from a linear mixture of independent stochastic processes, which is used for source-separation problems. The statistical tool that is proposed allows for improving the identification of any stochastic model of a random vector in high dimension for which a data set of realizations is given. Instead of directly constructing a unique stochastic model for which its stochastic dimension, which is identified by solving a statistical inverse problem, can be large, the proposed preprocessing of the data set allows for constructing several mutually independent stochastic models with smaller stochastic dimensions. Consequently, such a method allows for decreasing the cost of the identification and/or to make possible an identification for a case that is a priori in high dimension and that could not be identified through a direct and global approach. For instance, instead of directly identifying the coefficients of a PCE for a random vector from a given set of its realizations, the preprocessing of the data set allows for identifying the coefficients of several PCEs for several random vectors that have smaller dimensions. The random vector defined by the given data set can come from the discretization of a stochastic process or of a random field for which the set of realizations come from experimental measurements or computational simulations. The algorithm is completely defined in the paper, which can easily be implemented. Three numerical experiments have been proposed

to validate the proposed approach. Clearly, many other cases could be analyzed. A non-Gaussian matrix-valued random field has been analyzed and the results obtained show that a very important gain can be obtained for the construction of a stochastic representation of this random field. Finally, it should be noted that the sole information that is supposed to be available for the random vector considered, is a limited set of its realizations that come, for example, from experimental measurements, and for which one seeks to build a stochastic representation by solving for example, a statistical inverse problem. It is assumed that no further information is available. So if a stochastic model is identified to represent this random vector, it will depend solely on this limited number of realizations. It is therefore consistent, in such a construction process, to use the same limited set of data to separate its components as an ensemble of several mutually independent random vectors that have smaller dimensions, before discussing the construction of a stochastic model for its representation.

Appendix A. Algorithm 1.

In this appendix, we give Algorithm 1 that allows for constructing the partition $\mathbb{P}^{\nu}_{i_{\text{ref}}}(m; \mu_1, \ldots, \mu_m)$ of $\mathbf{H}^{\nu}$ in terms of random vectors that are considered as independent for a given level i_{ref}. This basic algorithm is derived from the graph theory and uses the symmetric adjacency matrix $[\mathcal{N}]$. The algorithm is described using the Matlab language.

- The input variables are: n and, for i_{ref} given, the $(n \times n)$ symmetric adjacency matrix $[\mathcal{N}]$.

- The output variables are: m, the m-column matrix M such that $M(j) = \mu_j$ for $j = 1, \ldots, m$, and the $(m \times \max_j \mu_j)$-matrix $[\mathcal{S}]$ such that the j-th row of $[\mathcal{S}]$ is made up of the indices $[r_1^j \ldots r_{\mu_j}^j]$ of the random vector $\mathbf{H}^{\nu}$ such that $\mathbf{Y}^{\nu,j} = (H^{\nu}_{r_1^j}, \ldots, H^{\nu}_{r_{\mu_j}^j})$.

Algorithm 1

$j = 0;$	%
$M = \text{zeros}(n, 1); \mathcal{S} = \text{zeros}(n, n);$	% $M(j) = \mu_j; \mathcal{S}(j, k) = r_k^j$ (see (3.2)).
$U = \text{zeros}(n, 1);$	% $U(r) = 0$, node r not treated and $= 1$, treated.
while isempty$(\text{find}(U == 0)) == 0$	% **do**: there are nodes that have not been treated.
$\quad \mathcal{U} = \text{find}(U == 0);$	% nodes that have not been treated.
$\quad x = \mathcal{U}(1);$	% node used for starting a new subset.
$\quad P = [\,];$	% initializing the list of the nodes to be analyzed.
$\quad V = [\,];$	% initializing the list of nodes already analyzed.
$\quad R = [x];$	% R contains the nodes of the present subset.
$\quad j = j + 1;$	% increasing the present number of subsets.
$\quad P = \text{union}(P, x);$	% union of the nodes with no repetitions.
$\quad$ **while** isempty$(P) == 0$	% P not empty **do**.
$\quad\quad y = P(1); P(1) = [\,];$	% load a node and unstack P.
$\quad\quad V = \text{union}(V, y);$	% union of the nodes with no repetitions.
$\quad\quad$ **for** $z = 1 : n$	% exploring all nodes z such that $\mathcal{N}(y, z) = 1$
$\quad\quad\quad W = \text{find}(\text{union}(P, V) == z);$	% and $z \notin \{P \cup V\}$.
$\quad\quad\quad$ **if** $\mathcal{N}(y, z) == 1 \,\&\, \text{isempty}(W) == 1$	%
$\quad\quad\quad\quad P = \text{union}(P, z);$	% stack (P, z).
$\quad\quad\quad\quad R = \text{union}(R, z);$	% y belongs to the subset.
$\quad\quad\quad$ **end**	% end if.
$\quad\quad$ **end**	% end for.
$\quad$ **end**	% end while.
$\quad \mu_j = \text{size}(R, 2); M(j) = \mu_j; \mathcal{S}(j, 1 : \mu_j) = R;$	%
$\quad U(R) = 1;$	% all nodes in R have then been treated.
$\quad \mathcal{N}(R, R) = 0;$	% setting to zero the nodes in subset j.
end	% end while.
m = j;	% loading the number m of independent subsets.
if $m < n$	% adjusting the size of the matrices.
$\quad M(m{+}1{:}n) = [\,];$	% adjusting the length of M.
$\quad \mathcal{S}(m{+}1{:}n, :) = [\,];$	% adjusting the number of rows of $[\mathcal{S}]$.
$\quad \mathcal{S}(:, \max(M){+}1{:}n) = [\,];$	% adjusting the number of columns of $[\mathcal{S}]$.
end	% end if.

REFERENCES

[1] H.C. ANDREWS AND C.L. PATTERSON, *Singular value decomposition and digital image processing*, Transactions on Acoustics, Speech, and Signal Processing - IEEE, 24 (1976), pp. 26-53.

[2] M. ARNST, R. GHANEM, AND C. SOIZE, *Identification of Bayesian posteriors for coefficients of chaos expansion*, Journal of Computational Physics, 229(9) (2010), pp. 3134-3154.

[3] M. ARNST, C. SOIZE, AND R. GHANEM, *Hybrid sampling/spectral method for solving stochastic coupled problems*, SIAM/ASA Journal on Uncertainty Quantification, 1(1) (2013), pp. 218243.

[4] M. ARNST, R. GHANEM, E. PHIPPS, AND J. RED-HORSE, *Reduced chaos expansions with random coefficients in reduced-dimensional stochastic modeling of coupled problems*, International Journal for Numerical Methods in Engineering, 29(5) (2014), pp. 352-376.

[5] G. BLATMAN AND B. SUDRET, *Adaptive sparse polynomial chaos expansion based on least angle regression*, Journal of Computational Physics, 230(6) (2011), pp. 23452367.

[6] R. BOLOSCO, H. PAN, *Independent component analysis based on nonparametric density estimation*, IEEE Transactions on Neural Networks, 15(1) (2004), pp. 55-65.

[7] J.A. BONDY AND U.S.R MURTY, *Graph Theory with Applications*, Fifth printing, Elsevier Science Publishing, New York, 1982.

[8] A.W. BOWMAN AND A. AZZALINI, *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, 1997.

[9] R.H. CAMERON AND W.T. MARTIN, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, The Annals of Mathematics, Second Series, 48(2) (1947), pp. 385-392.

[10] P. COMON, C. JUTTEN, J. HERAULT, *Blind separation of sources, part II: Problems statement*, Signal Processing, 24(1) (1991), pp. 11-20.

[11] P. COMON, *Independent component analysis - a new concept?*, Signal Processing, 36(3) (1994), pp. 287-314.

[12] P. COMON AND C. JUTTEN, *Handbook of blind source separation. Independent Component Analysis and Applications*, Elsevier, Oxford, 2010.

[13] T.M. COVER AND J.A. THOMAS, *Elements of Information Theory*, 2nd edition, John Wiley & Sons, Hoboken, New Jersey, 2006.

[14] S. DAS, R. GHANEM, AND J. SPALL, *Asymptotic sampling distribution for polynomial chaos representation of data : A maximum-entropy and fisher information approach*, SIAM Journal on Scientific Computing, 30(5) (2008), pp. 2207-2234.

[15] B.J. DEBUSSCHERE, H.N. NAJM, P.P. PEBAY, O.M. KNIO, R. GHANEM, AND O.P. LE MAITRE, *Numerical challenges in the use of polynomial chaos representations for stochastic processes*, SIAM Journal on Scientific Computing, 26(2) (2004), pp. 698-719.

[16] C. DESCELIERS, R. GHANEM, AND C. SOIZE, *Maximum likelihood estimation of stochastic chaos representations from experimental data*, International Journal for Numerical Methods in Engineering, 66(6) (2006), pp. 978-1001 .

[17] A. DOOSTAN, R. GHANEM, AND J. RED-HORSE, *Stochastic model reduction for chaos representations*, Computer Methods in Applied Mechanics and Engineering, 196(37-40) (2007), pp. 3951-3966.

[18] O.G. ERNST, A. MUGLER, H.J. STARKLOFF, AND E. ULLMANN, *On the convergence of generalized polynomial chaos expansions*, ESAIM: Mathematical Modelling and Numerical Analysis, 46(2) (2012), pp. 317-339.

[19] J.L. FLEISS, B. LEVIN, M.C. PAIK, *Statistical Methods for Rates and Proportions*, Third Edition, John Wiley & Sons, Hoboken, New Jersey, 2003.

[20] S. GEMAN AND D. GEMAN, *Stochastic relaxation, Gibbs distribution and the Bayesian distribution of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol PAM I-6 (1984), pp. 721-741.

[21] D. GHOSH AND R. GHANEM, *Stochastic convergence acceleration through basis enrichment of polynomial chaos expansions*, International Journal for Numerical Methods in Engineering, 73(2) (2008), pp. 162-184.

[22] J.J. GERBRANDS, *On the relationships between SVD, KLT and PCA*, Pattern Recognition, 14(1) (1981), pp. 375-381.

[23] R. GHANEM AND P.D. SPANOS, *Stochastic Finite Elements: A spectral Approach*, Springer-verlag, New-York, 1991 (revised edition, Dover Publications, New York, 2003).

[24] R. GHANEM AND R.M. KRUGER, *Numerical solution of spectral stochastic finite element systems*, Computer Methods in Applied Mechanics and Engineering, 129(3) (1996), pp. 289-303.

[25] R. GHANEM, R. DOOSTAN, AND J. RED-HORSE, *A probability construction of model validation*, Computer Methods in Applied Mechanics and Engineering, 197(29-32) (2008), pp. 2585-2595.

[26] R. GHANEM, D. HIGDON, AND H. OWHADI (Eds.), *Handbook for Uncertainty Quantification*, Springer, 1st ed., 2016.

[27] G.H. GIVENS AND J.A. HOETING, *Computational Statistics*, 2nd edition, John Wiley & Sons, Hoboken, New Jersey, 2013.

[28] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, 4th edition, The Johns Hopkins University Press, Baltimore, 2013.

[29] P.E. GREENWOOD AND M.S. NIKULIN, A Guide to Chi-Squared Testing, John Wiley & Sons, New York, 1996.

[30] W. K. HASTINGS, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 109 (1970), pp. 57-97.

[31] J. HERAULT AND C. JUTTEN, *Space or time adaptive signal processing by neural network models*, pp. 206-211, in

J.S. Denker (ed.), *Neural Networks for Computing*, Proceedings of AIP Conference, American Institute of Physics, New York, 1986.

[32] T.S. HUANG, *Picture Processing and Digital Filtering*, Springer, Berlin, 1975.

[33] A. HYVÄRINEN, *Fast and robust fixed-point algorithms for independent component analysis*, IEEE Transactions On Neural Networks, 10(3) (1999),pp. 626- 634.

[34] A. HYVÄRINEN AND E. OJA, *Independent component analysis: algorithms and applications*, Neural Networks, 13(4-5) (2000), pp. 411-430.

[35] C. JUTTEN, J. HERAULT, *Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*, Signal Processing, 24(1) (1991), pp. 1-10.

[36] V. KESHAVARZZADEH, R. GHANEM, S.F. MASRI, AND O.J. ALDRAIHEM, *Convergence acceleration of polynomial chaos solutions via sequence transformation*, Computer Methods in Applied Mechanics and Engineering, 271 (2014), pp. 167-184.

[37] A.N. KOLMOGOROV, *On the Shannon theory of information transmission in the case of continuous signals*, IRE Transactions on Information Theory, 2(4) (1958), pp. 102-108.

[38] S. KULLBACK AND R.A. LEIBLER, *On information and sufficiency*, The Annals of Mathematical Statistics, 22(1) (1951), pp. 79-86.

[39] T.-W. LEE, *A unifying information-theoretic framework for independent component analysis*, Computers and Mathematics with Applications, 39 (2000), pp. 1-21.

[40] O.P. LE MAITRE AND O.M. KNIO, *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*, Springer, Heidelberg, 2010.

[41] D. LUCOR, C.H. SU, AND G.E. KARNIADAKIS, *Generalized polynomial chaos and random oscillators*, International Journal for Numerical Methods in Engineering, 60(3) (2004), pp. 571-596.

[42] Y.M. MARZOUK AND H.N. NAJM, *Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems*, Journal of Computational Physics, 228(6) (2009), pp. 1862-1902.

[43] N. METROPOLIS AND S. ULAM, *The Monte Carlo method*, Journal of the American Statistical Association, 44 (1949), pp. 335341.

[44] H.N. NAJM, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, Journal Review of Fluid Mechanics, 41 (2009), pp. 35-52.

[45] A. NOUY AND C. SOIZE, *Random fields representations for stochastic elliptic boundary value problems and statistical inverse problems*, European Journal of Applied Mathematics, 25(3) (2014), pp. 339-373.

[46] K. PEARSON, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philosophical Magazine Series 5, 50(302) (1900), pp. 157-175, doi:10.1080/14786440009463897.

[47] G. PERRIN, C. SOIZE, D. DUHAMEL, AND C. FUNFSCHILLING, *Identification of polynomial chaos representations in high dimension from a set of realizations*, SIAM Journal on Scientific Computing, 34(6) (2012), pp. A2917-A2945.

[48] R.Y. RUBINSTEIN AND D.P. KROESE, *Simulation and the Monte Carlo Method*, 2nd edition, John Wiley & Sons, Hoboken, New Jersey, 2008.

[49] R.J. SERFLING, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, Hoboken, New Jersey, 1980.

[50] C.E. SHANNON, A MATHEMATICAL THEORY OF COMMUNICATION, Bell System Technical Journal, 27 (1948), pp. 379-423 and 623-659.

[51] C. SOIZE AND R. GHANEM, *Physical systems with random uncertainties : Chaos representation with arbitrary probability measure*, SIAM Journal on Scientific Computing, 26(2) (2004), pp. 395-410.

[52] C. SOIZE, *Non Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators*, Computer Methods in Applied Mechanics and Engineering, 195(1-3) (2006), pp. 26-64.

[53] C. SOIZE, *Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data*, Computer Methods in Applied Mechanics and Engineering, 199(33-36) (2010), pp. 2150-2164.

[54] C. SOIZE, *Polynomial chaos expansion of a multimodal random vector*, SIAM/ASA Journal on Uncertainty Quantification, 3(1) (2015), pp. 34-60.

[55] B. SUDRET, *Global sensitivity analysis using polynomial chaos expansions*, Reliability Engineering & System Safety, 93(7) (2008), pp. 964-979.

[56] X.L. WAN AND G.E. KARNIADAKIS, *Multi-element generalized polynomial chaos for arbitrary probability measures*, SIAM Journal on Scientific Computing, 28(3) (2006), pp. 901-928.

[57] N. WIENER, *The homogeneous chaos*, The American Journal of Mathematics, 60(1) (1938), pp. 897936.

[58] D.B. XIU AND G.E. KARNIADAKIS, *Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM Journal on Scientific Computing, 24(2) (2002), pp. 619-644.