Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints

Guillaume Perrin, Christian Soize, N. Ouhbi

Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints

G. Perrin[a], C. Soize[b], N. Ouhbi[c]

[a] *CEA/DAM/DIF, F-91297, Arpajon, France*
[b] *Université Paris-Est, MSME UMR 8208 CNRS, Marne-la-Vallée, France*
[c] *Innovation and Research Department, SNCF, Paris, France*

Abstract

The multidimensional Gaussian kernel-density estimation (G-KDE) is a powerful tool to identify the distribution of random vectors when the maximal information is a set of independent realizations. For these methods, a key issue is the choice of the kernel and the optimization of the bandwidth matrix. To optimize these kernel representations, two adaptations of the classical G-KDE are presented. First, it is proposed to add constraints on the mean and the covariance matrix in the G-KDE formalism. Secondly, it is suggested to separate in different groups the components of the random vector of interest that could reasonably be considered as independent. This block by block decomposition is carried out by looking for the maximum of a cross-validation likelihood quantity that is associated with the block formation. This leads to a tensorized version of the classical G-KDE. Finally, it is shown on a series of examples how these two adaptations can improve the nonparametric representations of the densities of random vectors, especially when the number of available realizations is relatively low compared to their dimensions.

Key words:
Kernel density estimation, optimal bandwidth, nonparametric representation, data-driven sampling

Email addresses: guillaume.perrin2@cea.fr (G. Perrin)

1. Introduction

The generation of independent realizations of a second-order $\mathbb{R}^d$-valued random vector $\boldsymbol{X}$, whose distribution, $P_{\boldsymbol{X}}(d\boldsymbol{x})$, is unknown but can only be approximated from a finite set of $N \geq 1$ realizations, is a central issue in uncertainty quantification, signal processing and data analysis. One possible approach to address this problem is to suppose that the searched distribution belongs to an algebraic class of distributions, which can be mapped from a relatively small number of parameters (for instance, the multidimensional Gaussian distribution). Generating new realizations of random vector $\boldsymbol{X}$ amounts therefore at identifying the parameters that best suit the available data and then, at sampling independent realizations associated with the identified parametric distribution. However, when the dependence structure associated with the components of $\boldsymbol{X}$ is complex, such that its distribution can be concentrated on an unknown subset of $\mathbb{R}^d$, the definition of a relevant parametric class to represent $P_{\boldsymbol{X}}(d\boldsymbol{x})$ can become very difficult. In that case, nonparametric approaches are generally preferred to these parametric constructions [**?** **?**]. In particular, the multidimensional Gaussian kernel-density estimation (G-KDE) method approximates the probability density function (PDF) of $\boldsymbol{X}$, if it exists, as a sum of N multidimensional Gaussian PDFs, which are centred at each available independent realization of $\boldsymbol{X}$. Optimizing the covariance matrices associated with these N PDFs is a central issue, as they control the influence of each realization of $\boldsymbol{X}$ on the final approximation of $P_{\boldsymbol{X}}(d\boldsymbol{x})$. Even if there are many contributions on this subject (see for instance [**?** **?** **?** **?** **?**]), when the dimension d of $\boldsymbol{X}$ is high ($d \sim 10 - 100$), constant covariance matrices parametrized by a unique scaling parameter are generally considered. In particular, the Silverman rule of thumb [**?**] for choosing this scaling parameter is widely used because of its simplicity and its good asymptotic behaviour when N tends to infinity. However, for fixed values of N, this Silverman choice often overestimates the scattering of $P_{\boldsymbol{X}}(d\boldsymbol{x})$, and can have difficulties to correctly concentrate the new generated realizations of $\boldsymbol{X}$ on their regions of high probability.

To overcome this problem, a two-step procedure is introduced. First, we suggest to center and to uncorrelate the random vector $\boldsymbol{X}$ (using a Principal Component Analysis for instance). Then, based on the maximization of a global "Leave-One-Out" likelihood, the idea is to separate in different blocks the elements of $\boldsymbol{X}$, which could reasonably be considered as statistically independent. A tensorized version of the classical G-KDE that is adapted

to this dependence structure is eventually proposed. Indeed, for a finite number of realizations of $\boldsymbol{X}$, the less elements there are in each group, the more chance we have to correctly infer the multidimensional distribution of each sub-vector constituted of each group elements, and so the better should be the estimation of the PDF of $\boldsymbol{X}$. Nevertheless, the identification of this (unknown) block decomposition is a difficult combinatorial problem. This paper presents therefore two algorithms to find relevant block decompositions in a reasonable computational time.

The outline of this work is as follows. Section 2 presents the theoretical framework associated with the G-KDE and the optimization of the covariance matrices on which it is based. The block decomposition we propose is then detailed in Section 3. At last, the efficiency of the method is illustrated on a series of analytic and industrial examples in Section 4.

2. Theoretical framework

Let $\boldsymbol{X} := \{\boldsymbol{X}(\omega),\ \omega \in \Omega\}$ be a second-order random vector defined on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$, with values in $\mathbb{R}^d$. We assume that the probability density function (PDF) of $\boldsymbol{X}$ exists. By definition, this PDF, which is denoted by $p_{\boldsymbol{X}}$, is an element of $\mathcal{M}_1(\mathbb{R}^d, \mathbb{R}^+)$, the set of positive-valued functions, whose integral over $\mathbb{R}^d$ is 1. It is assumed that the maximal available information about $p_{\boldsymbol{X}}$ is a set of $N > d$ independent and distinct realizations of $\boldsymbol{X}$, which are gathered in the deterministic set $\mathcal{S}(N) := \{\boldsymbol{X}(\omega_n),\ 1 \leq n \leq N\}$. Given these realizations of $\boldsymbol{X}$, the kernel estimator of $p_{\boldsymbol{X}}$ is

$$\widehat{p}_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{H}, \mathcal{S}(N)) = \frac{\det(\boldsymbol{H})^{-1/2}}{N} \sum_{n=1}^{N} K\left(\boldsymbol{H}^{-1/2}\left(\boldsymbol{x} - \boldsymbol{X}(\omega_n)\right)\right), \qquad (1)$$

where $\det(\cdot)$ is the determinant operator, K is any function of $\mathcal{M}_1(\mathbb{R}^d, \mathbb{R}^+)$, and $\boldsymbol{H}$ is a $(d \times d)$-dimensional positive definite symmetric matrix that is generally referred as the "bandwidth matrix". In the following, we focus on the classical case when K is the Gaussian multidimensional density. Hence, the PDF $p_{\boldsymbol{X}}$ is approximated by a mixture of N Gaussian PDFs, for which the means are the available realizations of $\boldsymbol{X}$ and the covariance matrices are all equal to $\boldsymbol{H}$:

$$\widehat{p}_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{H},\mathcal{S}(N)) = \frac{1}{N}\sum_{n=1}^{N}\phi\left(\boldsymbol{x};\boldsymbol{X}(\omega_n),\boldsymbol{H}\right), \quad \boldsymbol{x}\in\mathbb{R}^d, \tag{2}$$

where for any $\mathbb{R}^d$-dimensional vector $\boldsymbol{\mu}$ and for any $(\mathbb{R}^d \times \mathbb{R}^d)$-dimensional symmetric positive definite matrix $\boldsymbol{C}$, $\phi(\cdot;\boldsymbol{\mu},\boldsymbol{C})$ is the PDF of an $\mathbb{R}^d$-dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$:

$$\phi\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{C}\right) := \frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\boldsymbol{C}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right)}{(2\pi)^{d/2}\sqrt{\det(\boldsymbol{C})}}, \quad \boldsymbol{x}\in\mathbb{R}^d. \tag{3}$$

By construction, the matrix $\boldsymbol{H}$ in Eq. (2) characterizes the local contribution of each realization of $\boldsymbol{X}$. Thus, its value has to be optimized to minimize the difference between $p_{\boldsymbol{X}}$, which is unknown, and $\widehat{p}_{\boldsymbol{X}}(\cdot;\boldsymbol{H},\mathcal{S}(N))$. The mean integrated squared error (MISE) performance criterion

$$\mathrm{MISE}(\boldsymbol{H};d,N) = \mathbb{E}\left[\int_{\mathbb{R}^d}\left(p_{\boldsymbol{X}}(\boldsymbol{x}) - \widehat{p}_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{H},\mathcal{S}(N))\right)^2 d\boldsymbol{x}\right] \tag{4}$$

is generally considered to quantify such a difference. Here $\mathbb{E}\left[\cdot\right]$ is the mathematical expectation. For this criterion, it can be noticed that the set $\mathcal{S}(N)$ is random, whereas in the rest of this paper it is deterministic. Given sufficient regularity conditions on $p_{\boldsymbol{X}}$, an asymptotic approximation of this criterion can be derived. In low dimension, the value of $\boldsymbol{H}$ that minimizes this asymptotic criterion can be explicitly calculated, but its value depends on the unknown PDF $p_{\boldsymbol{X}}$ and its derivatives (see [**?**] for more details). Studies have therefore been conducted to estimate these functions (generally iteratively) from the only available information given by $\mathcal{S}(N)$ (see for instance [**?** **?**]). However, the convergence of these methods is rather slow in high dimension, such that in practice, a widely used value for $\boldsymbol{H}$ is given by the Silverman bandwidth matrix

$$\boldsymbol{H}^{\mathrm{Silv}}(d,N) := \left(h^{\mathrm{Silv}}(d,N)\right)^2 \begin{bmatrix} \widehat{\sigma}_1^2 & 0 & \cdots & 0 \\ 0 & \widehat{\sigma}_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \widehat{\sigma}_d^2 \end{bmatrix} \tag{5}$$

where for all $1 \leq i \leq d$, $\widehat{\sigma}_i^2$ is the empirical estimation of the variance of X_i, and where

$$h^{\text{Silv}}(d,N) := \left(\frac{1}{N}\frac{4}{(d+2)}\right)^{\frac{1}{d+4}}. \tag{6}$$

This expression, which is derived from a Gaussian assumption on $p_{\boldsymbol{X}}$, is thought to be a good compromise between complexity and precision. However, it is generally observed that, for fixed values of N, when the distribution of $\boldsymbol{X}$ is concentrated on an unknown subset of $\mathbb{R}^d$, the more complex and disconnected this subset, the less relevant the value of $\boldsymbol{H}^{\text{Silv}}(d,N)$. To face this problem, the diffusion maps theory [**?**] can be used to bias the generation of independent realizations under $\widehat{p}_{\boldsymbol{X}}(\cdot;\boldsymbol{H}^{\text{Silv}}(d,N),\mathcal{S}(N))$ and make them closer to the ones we could have got if they had been generated under the true PDF $p_{\boldsymbol{X}}$. Indeed, diffusion maps are a very powerful mathematical tool to discover and characterize sets on which the distribution of $\boldsymbol{X}$ is concentrated, and their coupling to nonparametric statistical representations has shown promising results, even when dealing with very high values of d [**?**].

From another point of view, the likelihood $\mathcal{L}(\mathcal{S}(N)|\boldsymbol{H})$ associated with $\boldsymbol{H}$ can also directly be used to identify relevant values of $\boldsymbol{H}$. From Eq. (1), it follows that

$$\mathcal{L}(\mathcal{S}(N)|\boldsymbol{H}) := \prod_{n=1}^{N} \widehat{p}_{\boldsymbol{X}}(\boldsymbol{X}(\omega_n);\boldsymbol{H},\mathcal{S}(N)) = \frac{1}{N^N}\prod_{n=1}^{N}\sum_{m=1}^{N}\phi_{n,m}(\boldsymbol{H}), \tag{7}$$

$$\phi_{n,m}(\boldsymbol{H}) := \phi\left(\boldsymbol{X}(\omega_n);\boldsymbol{X}(\omega_m),\boldsymbol{H}\right), \quad 1 \leq n,m \leq N. \tag{8}$$

The function $\mathcal{L}(\mathcal{S}(N)|\boldsymbol{H})$ uses twice the same information (to compute $\widehat{p}_{\boldsymbol{X}}(\cdot;\boldsymbol{H},\mathcal{S}(N))$ and to evaluate it). Hence, it tends to infinity when $\boldsymbol{H}$ tends to zero, which can be seen as an overfitting of the available data. In order to avoid this phenomenon, it is proposed in [**?**] to consider its "Leave-One-Out" (LOO) expression

$$\mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{H}) := \prod_{n=1}^{N}\frac{1}{N-1}\sum_{m=1,m\neq n}^{N}\phi_{n,m}(\boldsymbol{H}) \tag{9}$$

instead. Given this approximate likelihood obtained from an LOO cross-validation, and an *a priori* density $p_{\boldsymbol{H}}$ for $\boldsymbol{H}$, Bayesian approaches can be used to compute the posterior density of $\boldsymbol{H}$ [**?**]:

$$p_{\boldsymbol{H}}(\boldsymbol{H}|\mathcal{S}(N)) := c\ \mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{H}) p_{\boldsymbol{H}}(\boldsymbol{H}),\ \boldsymbol{H} \in \mathbb{M}^{+}(d). \tag{10}$$

Here, c is a normalizing constant and $\mathbb{M}^{+}(d)$ is the set of all $(d \times d)$-dimensional symmetric positive definite matrices. In particular, the maximum likelihood estimate of $\boldsymbol{H}$ is denoted by

$$\boldsymbol{H}^{\mathrm{MLE}}(d, N) := \arg \max_{\boldsymbol{H} \in \mathbb{M}^{+}(d)} \mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{H}). \tag{11}$$

Additionally, considering that the best available approximations of the true mean and covariance matrix of $\boldsymbol{X}$ are given by their empirical estimations

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} := \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{X}(\omega_n),$$

$$\widehat{\boldsymbol{R}}_{\boldsymbol{X}} := \frac{1}{N-1} \sum_{n=1}^{N} (\boldsymbol{X}(\omega_n) - \widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}) \otimes (\boldsymbol{X}(\omega_n) - \widehat{\boldsymbol{\mu}}_{\boldsymbol{X}}),$$

the expression given by Eq. (1) can be slightly modified to ensure that the mean and the covariance matrix of the G-KDE approximation of $\boldsymbol{X}$ are equal to these estimations. Following [**?**], this can be done by considering the subsequent proposition. The proof is given in Appendix.

Proposition 1. *If the PDF of* $\widetilde{\boldsymbol{X}}$ *is equal to*

$$\widetilde{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}, \mathcal{S}(N)) := \frac{1}{N} \sum_{n=1}^{N} \phi\left(\cdot; \boldsymbol{A}\boldsymbol{X}(\omega_n) + \boldsymbol{\beta}, \boldsymbol{H}\right), \tag{12}$$

$$\boldsymbol{\beta} := (\boldsymbol{I}_d - \boldsymbol{A})\widehat{\boldsymbol{\mu}}, \quad \boldsymbol{H} := \widehat{\boldsymbol{R}}_{\boldsymbol{X}} - \frac{N-1}{N} \boldsymbol{A} \widehat{\boldsymbol{R}}_{\boldsymbol{X}} \boldsymbol{A}^T, \tag{13}$$

where $\boldsymbol{A}$ *is any* $(d \times d)$*-dimensional matrix such that* $\boldsymbol{H}$ *is positive definite, then the mean and the covariance matrix of* $\widetilde{\boldsymbol{X}}$ *are equal to* $\widehat{\boldsymbol{\mu}}$ *and* $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$ *respectively.*

Given $\mathcal{S}(N)$, the G-KDE of the PDF of $\boldsymbol{X}$ under constraints on its mean and its covariance matrix is denoted by $\widetilde{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}^{\mathrm{MLE}}(d, N), \mathcal{S}(N))$. Here, $\boldsymbol{H}^{\mathrm{MLE}}(d, N)$ is the argument that maximizes the LOO likelihood of $\boldsymbol{H}$ associated with $\widetilde{p}_{\boldsymbol{X}}$.

Given $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$, and $\boldsymbol{H}^{\mathrm{MLE}}(d,N)$, the generation of independent realizations of $\widetilde{\boldsymbol{X}} \sim \widetilde{p}_{\boldsymbol{X}}(\cdot;\boldsymbol{H}^{\mathrm{MLE}}(d,N),\mathcal{S}(N))$ is straightforward. Indeed, for any $M \geq 1$, the Algorithm 1 (defined below) can be used to generate a $(d \times M)$-dimensional matrix $\boldsymbol{Z}$, whose columns are independent realizations of $\widetilde{\boldsymbol{X}}$. There, $\mathcal{U}\{1,\ldots,N\}$ denotes the discrete uniform distribution over $\{1,\ldots,N\}$ and $\mathcal{N}(0,1)$ denotes the standard Gaussian distribution.

1 Let $Q(\omega_1'),\ldots,Q(\omega_M')$ be M independent realizations that are drawn from $\mathcal{U}\{1,\ldots,N\}$;
2 Let $\boldsymbol{M}$ be a $(d \times M)$-dimensional matrix whose columns are all equal to $\widehat{\boldsymbol{\mu}}$;
3 Compute $\boldsymbol{A}$ such that $\boldsymbol{H} := \widehat{\boldsymbol{R}}_{\boldsymbol{X}} - \frac{N-1}{N}\boldsymbol{A}\widehat{\boldsymbol{R}}_{\boldsymbol{X}}\boldsymbol{A}^T$;
4 Define $\bar{\boldsymbol{X}} := \left[\boldsymbol{X}(\omega_{Q(\omega_1')}) \ \cdots \ \boldsymbol{X}(\omega_{Q(\omega_M')})\right]$;
5 Let $\boldsymbol{\Xi}$ be a $(d \times M)$-dimensional matrix, whose components are dM independent realizations that are drawn from $\mathcal{N}(0,1)$;
6 Assemble $\boldsymbol{Z} = \boldsymbol{M} + \boldsymbol{A}(\bar{\boldsymbol{X}} - \boldsymbol{M}) + \boldsymbol{H}^{\mathrm{MLE}}(d,N)^{1/2}\boldsymbol{\Xi}$.

Algorithm 1: Generation of M independent realizations of $\widetilde{\boldsymbol{X}}$.

Finally, this section has presented the general framework to nonparametrically approximate the PDF of a random vector when the maximal information is a set of N independent realizations. Some adjustments of the classical formulation have been proposed to take into account constraints on the first and second statistical moments of the approximated PDF, and it has been proposed to search the kernel density bandwidth as the solution of a computationally demanding LOO likelihood maximization problem.

However, from the analysis of a series of test cases, it appears that $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$ is a rather good approximation of $\boldsymbol{H}^{\mathrm{MLE}}(d,N)$ for the nonparametric modelling of high dimensional random vectors ($d \sim 10-100$) with limited information ($N \sim 10d$ for instance). From Eqs. (12) and (13), this means that we are approximating the PDF of $\boldsymbol{X}$ as a unique Gaussian PDF, whose parameters correspond to the empirical mean and covariance matrix of $\boldsymbol{X}$:

$$\lim_{\boldsymbol{H}\to\widehat{\boldsymbol{R}}_{\boldsymbol{X}}} \widetilde{p}_{\boldsymbol{X}}(\cdot;\boldsymbol{H},\mathcal{S}(N)) = \phi(\cdot;\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{R}}_{\boldsymbol{X}}). \tag{14}$$

This could prevent us from recovering the subset of $\mathbb{R}^d$ on which $\boldsymbol{X}$ is

actually concentrated. To face this problem, we can be tempted to impose smaller values for the components of $\boldsymbol{H}$ in the nonparametric model. If all the components of $\boldsymbol{X}$ are actually dependent, there is however no reason to do so without biasing the final constructed distribution in focusing too much on the available data. Thus, instead of artificially decreasing the most likely value of $\boldsymbol{H}$ (according to the available data), the next section proposes several adaptations of this G-KDE formalism.

3. Data-driven tensor-product representation

This section presents some adaptations of the classical G-KDE to improve the nonparametric representations of $p_{\boldsymbol{X}}$ when the number N of available realizations of $\boldsymbol{X}$ is relatively small compared to its dimension d. Following [**?**] and [**?**], we first suggest to pre-process the realizations of $\boldsymbol{X}$ (from a Principal Component Analysis for instance) such that $\boldsymbol{X}$ is now supposed to be centred and uncorrelated:

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} = \boldsymbol{0}, \quad \widehat{\boldsymbol{R}}_{\boldsymbol{X}} = \boldsymbol{I}_d.$$

Here, $\boldsymbol{I}_d$ is the $(d \times d)$-dimensional identity matrix. This makes independent the components of $\boldsymbol{X}$ that were only linearly dependent. Then, the idea is to identify groups of components of $\boldsymbol{X}$ that can reasonably be considered as statistically independent, if they exist. Instead of using statistical tests, we propose to search these groups by looking for the maximum of a cross-validation likelihood quantity that is associated with each block formation. Thus, given a block by block decomposition of the components of $\boldsymbol{X}$, the PDF $p_{\boldsymbol{X}}$ is approximated as the product of the nonparametric estimations of the PDFs associated with each sub-vector of $\boldsymbol{X}$. For instance, if the d components of $\boldsymbol{X}$ are sorted in d distinct groups, the approximation of $p_{\boldsymbol{X}}$ corresponds to the product of the d nonparametric estimations of the marginal PDFs of $\boldsymbol{X}$. Indeed, if the identified block decomposition is correctly adapted to the (unknown) dependence structure of $\boldsymbol{X}$, there are good chances for the nonparametric representation of $p_{\boldsymbol{X}}$ to be improved.

More details about this block decomposition are presented in the rest of this section. First, we introduce the notations and the formalism on which this decomposition is based. Then, several algorithms are proposed for its practical identification.

3.1. Block by block decomposition

For any $\boldsymbol{b}$ in $\{1,\dots,d\}^d$ and for all $1 \leq i \leq d$, b_i can be used as a block index for the i^{th} component X_i of $\boldsymbol{X}$. This means that if $b_i = b_j$, X_i and X_j are supposed to be dependent and have to belong to the same block. On the contrary, if $b_i \neq b_j$, X_i and X_j are supposed to be independent and they can belong to two different blocks. In order to avoid any redundancy in the block by block parametrization of $\boldsymbol{X}$, the following subset of $\{1,\dots,d\}^d$ is considered:

$$\mathbb{B}(d) := \left\{ \boldsymbol{b} \in \{1,\dots,d\}^d \mid b_1 = 1,\ 1 \leq b_j \leq 1 + \max_{1 \leq i \leq j-1} b_i,\ 2 \leq j \leq d \right\}. \tag{15}$$

Additionally, for any $\boldsymbol{b}$ in $\mathbb{B}(d)$, let

- $\text{Max}(\boldsymbol{b})$ be the maximal value of $\boldsymbol{b}$,
- $\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})$ be the random vector that gathers all the components of $\boldsymbol{X}$ with a block index equal to ℓ,
- d_ℓ be the number of elements of $\boldsymbol{b}$ that are equal to ℓ,
- $\mathcal{S}^\ell(N)$ be the set that gathers the N independent realizations of $\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})$ that have been extracted from the N independent realizations of $\boldsymbol{X}$ in $\mathcal{S}(N)$.

There exists a bijection between $\mathbb{B}(d)$ and the set of all block by block decompositions of $\boldsymbol{X}$. For instance, for $d = 5$, all the elements of $\{(i,j,i,k,k),\ 1 \leq i \neq j \neq k \leq 5\}$ correspond to the same block decomposition of $\boldsymbol{X}$, but only $\boldsymbol{b} = (1,2,1,3,3)$ is in $\mathbb{B}(d)$. We can also identify

$$\boldsymbol{s}^{(1)}(\boldsymbol{X};\boldsymbol{b}) = (X_1, X_3), \quad \boldsymbol{s}^{(2)}(\boldsymbol{X};\boldsymbol{b}) = X_2, \quad \boldsymbol{s}^{(3)}(\boldsymbol{X};\boldsymbol{b}) = (X_4, X_5), \tag{16}$$

$$\text{Max}(\boldsymbol{b}) = 3, \quad d_1 = 2 \quad d_2 = 1, \quad d_3 = 2. \tag{17}$$

According to Eq. (12), for any $\boldsymbol{H}_\ell$ in $\mathbb{M}^+(d_\ell)$, the PDF of $\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})$ can be approximated by $\widetilde{p}_{\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})}(\cdot;\boldsymbol{H}_\ell,\mathcal{S}^\ell(N))$. It follows that the PDF of $\boldsymbol{X}$ can be constructed as the product of these $\text{Max}(\boldsymbol{b})$ PDFs:

$$\widetilde{p}_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{H}_1,\ldots,\boldsymbol{H}_{\mathrm{Max}(\boldsymbol{b})},\mathcal{S}(N),\boldsymbol{b}) := \prod_{\ell=1}^{\mathrm{Max}(\boldsymbol{b})} \widetilde{p}_{\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})}(\boldsymbol{s}^{(\ell)}(\boldsymbol{x};\boldsymbol{b});\boldsymbol{H}_\ell,\mathcal{S}^\ell(N)). \tag{18}$$

Such a construction for the PDF of $\boldsymbol{X}$ means that the vectors $\boldsymbol{s}^{(\ell)}(\boldsymbol{X};\boldsymbol{b})$, $1 \leq \ell \leq \mathrm{Max}(\boldsymbol{b})$, are assumed to be independent. For any $\boldsymbol{b}$ in $\mathbb{B}(d)$, let $\boldsymbol{H}_1^{\mathrm{MLE}}(\boldsymbol{b}),\ldots,\boldsymbol{H}_d^{\mathrm{MLE}}(\boldsymbol{b})$ be the arguments that maximize the LOO likelihood associated with $\widetilde{p}_{\boldsymbol{X}}$. Hence, for a given block by block decomposition of $\boldsymbol{X}$ that is characterized by a given value of $\boldsymbol{b}$, the most likely G-KDE of $p_{\boldsymbol{X}}$ is given by

$$\widetilde{p}_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{H}_1^{\mathrm{MLE}}(\boldsymbol{b}),\ldots,\boldsymbol{H}_d^{\mathrm{MLE}}(\boldsymbol{b}),\mathcal{S}(N),\boldsymbol{b}). \tag{19}$$

Using Eqs. (9), (12) and (18), for any $\boldsymbol{b}$ in $\mathbb{B}(d)$ and any $(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_{\mathrm{Max}(\boldsymbol{b})})$ in $\mathbb{M}^+(d_1)\times\cdots\times\mathbb{M}^+(d_{\mathrm{Max}(\boldsymbol{b})})$, this LOO likelihood is given by

$$\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{H}_1\ldots,\boldsymbol{H}_d,\boldsymbol{b}) \;=\; \prod_{\ell=1}^{\mathrm{Max}(\boldsymbol{b})}\prod_{n=1}^{N}\frac{1}{N-1}\sum_{m=1,m\neq n}^{N}\widetilde{\phi}_{n,m}(\boldsymbol{H}_\ell,\boldsymbol{b}), \tag{20}$$

$$\widetilde{\phi}_{n,m}(\boldsymbol{H}_\ell,\boldsymbol{b}) := \phi\left(\boldsymbol{s}^{(\ell)}(\boldsymbol{X}(\omega_n);\boldsymbol{b});\boldsymbol{A}_\ell\boldsymbol{s}^{(\ell)}(\boldsymbol{X}(\omega_m);\boldsymbol{b}),\boldsymbol{H}_{(\ell)}\right), \tag{21}$$

$$\boldsymbol{H}_\ell := \boldsymbol{I}_{d_\ell} - \frac{N-1}{N}\boldsymbol{A}_\ell\boldsymbol{A}_\ell^T. \tag{22}$$

Noticing that

$$\begin{aligned}\max_{\boldsymbol{H}_1,\ldots,\boldsymbol{H}_{\mathrm{Max}(\boldsymbol{b})},\boldsymbol{b}} &\prod_{\ell=1}^{\mathrm{Max}(\boldsymbol{b})}\prod_{n=1}^{N}\frac{1}{N-1}\sum_{m=1,m\neq n}^{N}\widetilde{\phi}_{n,m}(\boldsymbol{H}_\ell,\boldsymbol{b})\\ &= \max_{\boldsymbol{b}}\prod_{\ell=1}^{\mathrm{Max}(\boldsymbol{b})}\max_{\boldsymbol{H}_\ell}\prod_{n=1}^{N}\frac{1}{N-1}\sum_{m=1,m\neq n}^{N}\widetilde{\phi}_{n,m}(\boldsymbol{H}_\ell,\boldsymbol{b}),\end{aligned} \tag{23}$$

it follows that for a given block by block decomposition of $\boldsymbol{X}$, the most likely values of $\boldsymbol{H}_1,\ldots,\boldsymbol{H}_{\mathrm{Max}(\boldsymbol{b})}$ can be computed independently, and saved for a possible re-use for an other value of $\boldsymbol{b}$. Indeed, if $\boldsymbol{b}^{(1)} = (1,1,2,2)$, two

values $\boldsymbol{H}_1^{(1)}$ and $\boldsymbol{H}_2^{(1)}$ have to be chosen for the bandwidth matrices (one for each block). This means that two independent LOO likelihood maximization problems have to be solved. In the same manner, if $\boldsymbol{b}^{(2)} = (1, 1, 2, 3)$, three values $\boldsymbol{H}_1^{(2)}$, $\boldsymbol{H}_2^{(2)}$ and $\boldsymbol{H}_3^{(2)}$ have to be chosen. However, given the same set of realizations of $\boldsymbol{X}$, it is clear that the most likely value of $\boldsymbol{H}_1^{(1)}$ is equal to the most likely value of $\boldsymbol{H}_1^{(2)}$. Hence, the most likely value of $\boldsymbol{b}$, which is denoted by $\boldsymbol{b}^{\mathrm{MLE}}$, is eventually solution of

$$\boldsymbol{b}^{\mathrm{MLE}} := \arg \max_{\boldsymbol{b} \in \mathbb{B}(d)} \mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N) | \boldsymbol{H}_1^{\mathrm{MLE}}(\boldsymbol{b}), \ldots, \boldsymbol{H}_d^{\mathrm{MLE}}(\boldsymbol{b}), \boldsymbol{b}). \tag{24}$$

There, we remind that for any $\boldsymbol{b}$ in $\mathbb{B}(d)$ and any $1 \leq \ell \leq \mathrm{Max}(\boldsymbol{b})$,

$$\boldsymbol{H}_\ell^{\mathrm{MLE}}(\boldsymbol{b}) := \arg \max_{\boldsymbol{H}_\ell \in \mathbb{M}^+(d_\ell)} \prod_{n=1}^{N} \frac{1}{N-1} \sum_{m=1, m \neq n}^{N} \widetilde{\phi}_{n,m}(\boldsymbol{H}_\ell, \boldsymbol{b}). \tag{25}$$

Analyzing the value of $\boldsymbol{b}^{\mathrm{MLE}}$ can give information on the actual dependence structure for the components of $\boldsymbol{X}$. Indeed, if $\boldsymbol{b}^{\mathrm{MLE}} = (1, \ldots, 1)$, the most appropriate representation for the PDF of $\boldsymbol{X}$ is its classical multidimensional Gaussian kernel estimation. This would mean that all the components of $\boldsymbol{X}$ are likely to be dependent. On the contrary, if $\boldsymbol{b}^{\mathrm{MLE}} = (1, 2, \ldots, d)$, the most likely representation corresponds to the assumption that all the components of $\boldsymbol{X}$ are independent. Other values of $\boldsymbol{b}^{\mathrm{MLE}}$ can also be used to identify groups of dependent components of $\boldsymbol{X}$, which are likely to be independent the ones to the others.

3.2. Practical solving of the block by block decomposition problem

The optimization problem defined by Eq. (24) being very complex, we suggest to search the most likely block by block decomposition of $\boldsymbol{X}$ using very simple parametrizations of the bandwidth matrices. Indeed, once vector $\boldsymbol{X}$ has been centred and uncorrelated, it is reasonable to parametrize each bandwidth matrix $\boldsymbol{H}_\ell$ by a unique scalar h_ℓ, such that $\boldsymbol{H}_\ell = h_\ell^2 \boldsymbol{I}_{d_\ell}$. From Eq. (22), it follows that

$$\boldsymbol{A}_\ell = \frac{N}{N-1} \sqrt{1 - h_\ell^2} \boldsymbol{I}_{d_\ell}. \tag{26}$$

Hence, for a given precision ϵ, the complex problem of searching the most likely values of $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_{\mathrm{Max}(\boldsymbol{b})}$ can be reduced to minimizing $\mathrm{Max}(\boldsymbol{b})$ non convex but explicit functions over the closed interval $[\epsilon, 1]$. This can be done

value of d	1	2	3	4	5	6	7	8	9	10
value of $N_{\mathbb{B}(d)}$	1	2	5	15	52	203	877	4140	21147	115975
value of $N^{\max}_{\text{greedy}}(d)$	1	3	8	17	31	51	78	113	157	211

Table 1: Evolution of $N_{\mathbb{B}(d)}$ and $N^{\max}_{\text{greedy}}(d)$ with respect to d.

in parallel, and each minimization problem can be solved very efficiently using a combination of golden section search and successive parabolic interpolations (see [**?**] for further details about this method). However, solving the optimization problem defined by Eq. (24) can still be computationally demanding when d increases. Indeed, as it can be seen in Table 1, the number of admissible values of $\boldsymbol{b}$, which is denoted by $N_{\mathbb{B}(d)}$, increases exponentially with respect to d. Hence, a brute force approach, which would consist in testing all the possible values of $\boldsymbol{b}$, can not be used to identify $\boldsymbol{b}^{\text{MLE}}$.

As an alternative, we propose to consider a greedy algorithm, whose computational cost can be bounded. Starting from a configuration where all the components of $\boldsymbol{X}$ are in the same block, which corresponds to $\boldsymbol{b} = (1, \ldots, 1)$, the idea of this algorithm is to remove iteratively one element of this initial block, and to put it in a block that would be already built, or in a new block where it is the only element. The Algorithm 2 provides a more detailed description of this procedure. By construction, the number $N_{\text{greedy}}(d)$ of evaluations of $\boldsymbol{b} \mapsto \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}, \boldsymbol{h})$ verifies

$$N_{\text{greedy}}(d) \leq N^{\max}_{\text{greedy}}(d) := 1 + \sum_{i=0}^{d-2} (d-i)(i+1) \leq d^3. \tag{27}$$

For $d > 4$, such an algorithm can therefore be used to approximate $\boldsymbol{b}^{\text{MLE}}$ at a computational cost that is much more affordable than a direct identification based on $N_{\mathbb{B}(d)}$ evaluations of $\boldsymbol{b} \mapsto \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}, \boldsymbol{h})$.

When modelling high dimensional random vectors ($d \sim 50 - 100$), the value of $N^{\max}_{\text{greedy}}(d)$, which is definitely much smaller than $N_{\mathbb{B}(d)}$, can also become very high:

$$N^{\max}_{\text{greedy}}(d = 50) = 22051, \quad N^{\max}_{\text{greedy}}(d = 100) = 171601. \tag{28}$$

To identify relevant values for $\boldsymbol{b}$ at a lower computational cost in such a constrained discrete set $\mathbb{B}(d)$, the genetic algorithms (see [**?**] for further

Initialization: $\boldsymbol{b}^* = (1, \ldots, 1)$, ind.blocked $= \emptyset$;
for $k = 1 : d$ **do**
  $L^{(k)} = \emptyset$, $\boldsymbol{b}^{(k)} = \emptyset$, index$^{(k)} = \emptyset$, $\ell = 1$;
  for $i \in \{1, \ldots, d\} \setminus ind.blocked$ **do**
    for $j = 2 : \min(d, Max(\boldsymbol{b}^\star) + 1)$ **do**
      Adapt the value of the block index: $\boldsymbol{b}^{\text{temp}} := \boldsymbol{b}^*$, $b_i^{\text{temp}} = j$;
      Compute: $L^{\text{temp}} = \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}^{\text{temp}}, \boldsymbol{h})$;
      Save results: $L^{(k)}\{\ell\} = L^{\text{temp}}$, $\boldsymbol{b}^{(k)}\{\ell\} = \boldsymbol{b}^{\text{temp}}$, index$^{(k)}\{\ell\} = i$;
      Increment: $\ell \leftarrow \ell + 1$;
    end
  end
  Find the best block index at iteration k: $\ell^* = \arg\max_\ell L^{(k)}\{\ell\}$;
  Actualize: $\boldsymbol{b}^* \leftarrow \boldsymbol{b}^{(k)}\{\ell^*\}$, ind.blocked $\leftarrow$ ind.blocked $\cup$ index$^{(k)}\{\ell^*\}$;
end
Maximize over all iterations: $(\ell^{\text{greedy}}, k^{\text{greedy}}) := \arg\max_{\ell,k} L^{(k)}\{\ell\}$;
Approximate $\boldsymbol{b}^{\text{MLE}} \approx \boldsymbol{b}^{(k^{\text{greedy}})}\left\{\ell^{\text{greedy}}\right\}$.

Algorithm 2: Greedy search of $\boldsymbol{b}^{\text{MLE}}$.

details) seem to be particularly adapted. Hence, an adaptation of these algorithms to the case of the identification of the most likely block by block decomposition of $\boldsymbol{X}$ is proposed. The fusion and the mutation processes on which such algorithms are generally based, as well as a pseudo-projection in $\mathbb{B}(d)$ are therefore detailed in Appendix. In these algorithms, for any set $\mathcal{S}$ (which can be discrete or continuous), we denote by $\mathcal{U}(\mathcal{S})$ the uniform distribution over $\mathcal{S}$. Based on these three functions, the Algorithm 3 shows the genetic procedure we suggest for solving Eq. (24). The results given by this genetic algorithm are dependent on three parameters:

- the maximum number of iterations $i^{\max}$,
- the probability of mutation p^{Mut},
- the size of the population we are considering in the genetic algorithm N_{pop}.

For this algorithm, the number of evaluations of $\boldsymbol{b} \mapsto \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}, \boldsymbol{h})$ is equal to $N^{\text{tot}} = i^{\max} \times N_{\text{pop}}$. For a given value of N^{tot}, it is however hard to infer the optimal values for these three parameters, as it depends on d and on the optimal block-by-block structure of the considered random vector of interest. However, from the analysis of a series of numerical examples, it is generally interesting to choose small values for p^{Mut} to limit the number of spontaneous mutations, and favour high values for the number of iterations $i^{\max}$ rather than for the population size N_{pop}.

Once a satisfying value $\widehat{\boldsymbol{b}}^{\text{MLE}}$ of $\boldsymbol{b}$ has been identified using the scalar parametrization of the bandwidth matrices, it is possible to enrich the parametrization of the bandwidth matrices to improve the nonparametric representation of the PDF of $\boldsymbol{X}$. This amounts at solving

$$\boldsymbol{H}_{\ell}^{\text{MLE}}(\widehat{\boldsymbol{b}}^{\text{MLE}}) = \arg \max_{\boldsymbol{H}_{\ell} \in \mathbb{M}^{+}(d_{\ell})} \frac{1}{N-1} \sum_{m=1, m \neq n}^{N} \widetilde{\phi}_{n,m}(\boldsymbol{H}_{\ell}, \widehat{\boldsymbol{b}}^{\text{MLE}}) \tag{29}$$

for all $1 \leq \ell \leq \text{Max}(\widehat{\boldsymbol{b}}^{\text{MLE}})$. In practice, we observed on a series of test cases that the interest of such an enrichment of the bandwidth matrix was relatively limited.

Choose $N_{\text{pop}} \geq 2$, $0 \leq p^{\text{Mut}} \leq 1$ and $i^{\max} \geq 1$;
Initialization ;
Define $B = \emptyset$, $L = \emptyset$, inc $= 1$;
Choose at random N_{pop} elements of $\mathbb{B}(d)$, $\left\{ \boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(N_{\text{pop}})} \right\}$;
for *n = 1 : N_{pop}* **do**
    Compute: $L^{\text{temp}} = \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}^{(n)}, \boldsymbol{h})$;
    Save results: $L\{\text{inc}\} = L^{\text{temp}}$, $B\{\text{inc}\} = \boldsymbol{b}^{(n)}$, inc $=$ inc $+ 1$;
end
Iteration ;
for *i = 2 : i^{max}* **do**
    Gather in $\mathcal{S}$ the N_{pop} elements of B associated with the N_{pop} highest values of L ;
    Choose at random N_{pop} distinct pairs of elements of $\mathcal{S}$: $\left\{ \left(\boldsymbol{b}^{(n,1)}, \boldsymbol{b}^{(n,2)} \right),\ 1 \leq n \leq N_{\text{pop}} \right\}$;
    for *n = 1 : N_{pop}* **do**
        Fusion: $\boldsymbol{b}^{\text{Fus}} = \text{Fusion}(\boldsymbol{b}^{(n,1)}, \boldsymbol{b}^{(n,2)})$;
        Mutation: $\boldsymbol{b}^{\text{Mut}} = \text{Mutation}(\boldsymbol{b}^{\text{Fus}}, p^{\text{Mut}})$;
        Compute: $L^{\text{temp}} = \max_{\boldsymbol{h}} \mathcal{L}^{\text{LOO}}(\mathcal{S}(N)|\boldsymbol{b}^{\text{Mut}}, \boldsymbol{h})$;
        Save results: $L\{\text{inc}\} = L^{\text{temp}}$, $B\{\text{inc}\} = \boldsymbol{b}^{\text{Mut}}$, inc $=$ inc $+ 1$;
    end
end
Maximize over all iterations: $k^{\text{gene}} = \arg\max_{1 \leq k \leq \text{inc}-1} L\{k\}$;
Approximate $\boldsymbol{b}^{\text{MLE}} \approx B\{k^{\text{gene}}\}$.

Algorithm 3: Genetic search of $\boldsymbol{b}^{\text{MLE}}$. The functions Mutation() and Fusion() are presented in Appendix, and are detailed in Algorithms 4 and 5.

4. Simulation and application studies

The purpose of this section is to illustrate the interest of the correlation constraints and the tensorized formulation for the nonparametric representation of PDFs when the maximal information is a finite set of independent realizations. To this end, a series of examples will be presented. The first examples will be based on generated data, so that the errors can be controlled, whereas the last example presents an industrial application based on experimental data.

4.1. Monte Carlo simulation studies

4.1.1. Lemniscate function

Let U be a random value that is uniformly distributed on $[-0.85\pi, 0.85\pi]$, $\boldsymbol{\xi} = (\xi_1, \xi_2)$ be a 2-dimensional random vector whose components are two independent standard Gaussian variables, and $\boldsymbol{X}^{\mathrm{L}} = (X_1^{\mathrm{L}}, X_2^{\mathrm{L}})$ be the random vector so that

$$\boldsymbol{X}^{\mathrm{L}} = \left(\frac{\sin(U)}{1 + \cos(U)^2}, \frac{\sin(U)\cos(U)}{1 + \cos(U)^2} \right) + 0.05\boldsymbol{\xi}. \tag{30}$$

We assume that $N = 200$ independent realizations of $\boldsymbol{X}^{\mathrm{L}}$ have been gathered in $\mathcal{S}(N)$. Given this information, we would like to generate additional points that could sensibly be considered as new independent realizations of $\boldsymbol{X}^{\mathrm{L}}$. Based on the G-KDE formalism presented in Section 2, four kinds of generators are compared in Figure 1, depending on the value of the bandwidth and on the constraints on the statistical moments of $\boldsymbol{X}^{\mathrm{L}}$.

- Case 1: $p_{\boldsymbol{X}^{\mathrm{L}}}$ is approximated by $p_{\widehat{\boldsymbol{X}}^{\mathrm{L}}}(\cdot; (h^{\mathrm{Silv}}(d,N))^2 \boldsymbol{I}_d, \mathcal{S}(N))$, which is defined by Eq. (1) (no constraints).
- Case 2: $p_{\boldsymbol{X}^{\mathrm{L}}}$ is approximated by $p_{\widetilde{\boldsymbol{X}}^{\mathrm{L}}}(\cdot; (h^{\mathrm{Silv}}(d,N))^2 \boldsymbol{I}_d, \mathcal{S}(N))$, which is defined by Eq. (12) (constraints on the mean and the covariance).
- Case 3: $p_{\boldsymbol{X}^{\mathrm{L}}}$ is approximated by $p_{\widehat{\boldsymbol{X}}^{\mathrm{L}}}(\cdot; (h^{\mathrm{MLE}}(d,N))^2 \boldsymbol{I}_d, \mathcal{S}(N))$ (no constraints).
- Case 4: $p_{\boldsymbol{X}^{\mathrm{L}}}$ is approximated by $p_{\widetilde{\boldsymbol{X}}^{\mathrm{L}}}(\cdot; (h^{\mathrm{MLE}}(d,N))^2 \boldsymbol{I}_d, \mathcal{S}(N))$ (constraints on the mean and the covariance).

The relevance of the different approximations of $p_{\boldsymbol{X}^{\mathrm{L}}}$ can be analysed from a graphical point of view in Figure 1. It is instructive to compare the associated values of the LOO likelihood, which is denoted by $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{H})$, as the higher this value, the more likely the approximation. Hence, for this example, introducing constraints on the mean and the covariance of the G-KDE tends to slightly increase the values of $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{H})$. Moreover, these results are strongly improved when choosing $h^{\mathrm{MLE}}(d,N)$ instead of $h^{\mathrm{Silv}}(d,N)$. Then, for these four cases, Figure 2 compares the evolution of $h^{\mathrm{Silv}}(d,N)$ and $h^{\mathrm{MLE}}(d,N)$ with respect to N, and shows the associated values of the LOO likelihood. For this example, it can therefore be seen that $h^{\mathrm{Silv}}(d,N)$ strongly overestimates the scattering of the distribution of $\boldsymbol{X}^{\mathrm{L}}$, for any considered values of N. This is not the case when working with $h^{\mathrm{MLE}}(d,N)$. It is also interesting to notice that for values of N lower than 10^4 (which is very high for 2-dimensional cases), the difference between $h^{\mathrm{MLE}}(d,N)$ and $h^{\mathrm{Silv}}(d,N)$ is always important.

4.1.2. Four branches clover-knot function

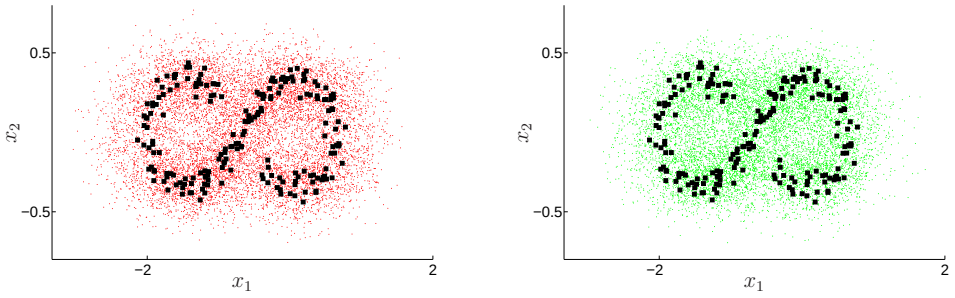
In the same manner than in the previous section, let U be a random value that is uniformly distributed on $[-\pi,\pi]$, $\boldsymbol{\xi}=(\xi_1,\xi_2,\xi_3)$ be a 3-dimensional random vector whose components are three independent standard Gaussian variables, and $\boldsymbol{X}^{\mathrm{FB}}$ be the random vector so that

$$\boldsymbol{X}^{\mathrm{FB}} = (\cos(U)+2\cos(3U), \sin(U)-2\sin(3U), 2\sin(4U)) + \boldsymbol{\xi}. \tag{31}$$

Once again, starting from a data set of $N=200$ independent realizations, we would like to be able to generate additional realizations of $\boldsymbol{X}^{\mathrm{FB}}$. For this 3-dimensional case, as in the previous section, Figures 3 and 4 allow us to underline the interest of considering G-KDE representations that are constrained in terms of mean and covariance, for which the bandwidths are optimized from the likelihood maximization point of view.

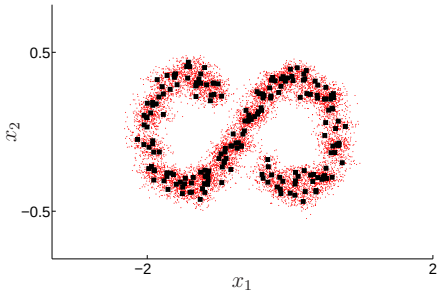
4.1.3. Interest of the block-by-block decomposition in higher dimensions

As explained in Section 3, when d is high, the G-KDE of $p_{\boldsymbol{X}}$ requires very high values of N to be able to identify the manifold on which the distribution of $\boldsymbol{X}$ is concentrated. In other words, if N is fixed, the higher d, the higher $h^{\mathrm{MLE}}(d,N)$ and the more scattered the new realizations of $\boldsymbol{X}$. As an illustration of this phenomenon, let us consider the two following random vectors, for $d \leq 1$:

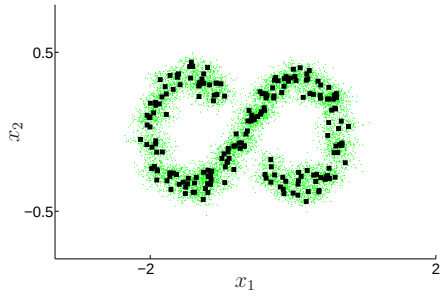


(a) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -147$

(b) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -142$

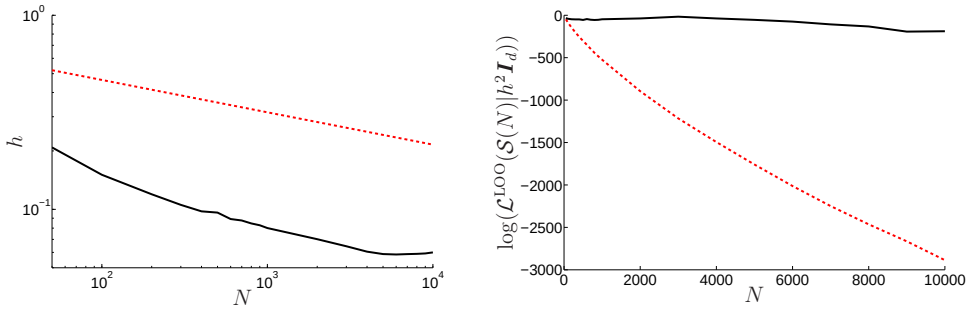


(c) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -39.0$



(d) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -38.7$

Figure 1: Lemniscate case: $N = 200$ given data points (big black squares) and 10^4 additional realizations (small red and green points) generated from a G-KDE approach for $h = h^{\mathrm{Silv}}(d, N)$ (first row) and $h = h^{\mathrm{MLE}}(d, N)$ (second row). The first column corresponds to the case where no constraints on the mean and the covariance of the generated points are introduced, whereas the second column corresponds to the case where the mean and the covariance of the generated points are equal to their empirical estimations that are computed from the available data. Under each graph is shown the value of the LOO likelihood for the associated value of h.



(a) Evolution of the bandwidth

(b) Evolution of $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)$

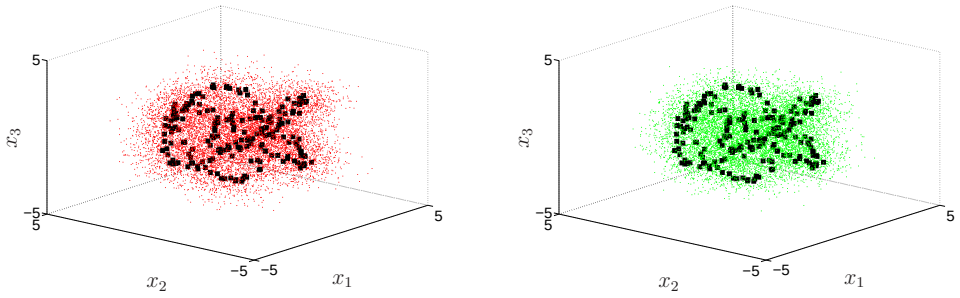
Figure 2: Evolution of the bandwidth (left) and of the LOO-likelihood (right) with respect to N for the Lemniscate function (2D). The red dotted lines correspond to the Silverman case: $h = h^{\mathrm{Silv}}(d, N)$. The black solid lines correspond to the MLE case: $h = h^{\mathrm{MLE}}(d, N)$. For this 2D example, the distinctions between the cases with correlation constraints or without were negligible compared to the difference between the Silverman and the MLE cases. Hence, only the cases where correlation constraints are imposed on the G-KDE are represented. Each curve corresponds to the mean values of h and $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d))$, which have been computed from 50 independent generated 200-dimensional sets of independent realizations of $\boldsymbol{X}^{\mathrm{L}}$.

- Case 1: $\boldsymbol{X}^{(2D)} = (\boldsymbol{X}^{\mathrm{L}}, \Xi_3, \ldots, \Xi_d)$.
- Case 2: $\boldsymbol{X}^{(3D)} = (\boldsymbol{X}^{\mathrm{FB}}, \Xi_4, \ldots, \Xi_d)$.

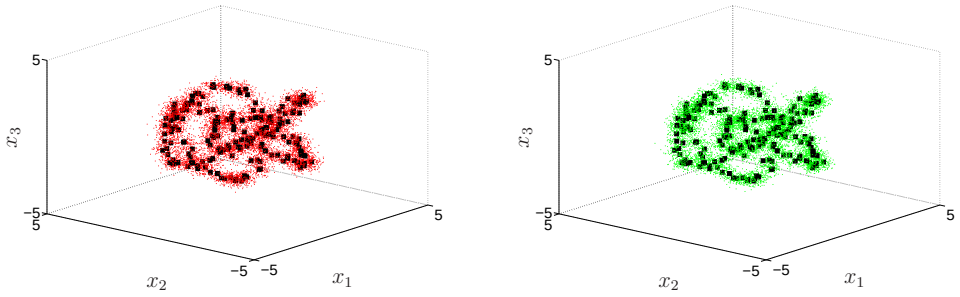
Here, $\Xi_3, \ldots, \Xi_d$ denote d independent standard Gaussian random variables, whereas the random vectors $\boldsymbol{X}^{\mathrm{L}}$ and $\boldsymbol{X}^{\mathrm{FB}}$ have been introduced in Section 4.1. For these two cases, two configurations are compared.

- On the first hand, a classical G-KDE of the PDFs of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$ is computed. In that case, no block decomposition is carried out. The block by block vectors associated with these modelling, which are respectively denoted by $\boldsymbol{b}^{(2D,1)}$ and $\boldsymbol{b}^{(3D,1)}$, are equal to $(1, \ldots, 1)$.
- On the second hand, we impose $\boldsymbol{b}^{(2D,2)} = (1, 1, 2, \ldots, d-1)$ and $\boldsymbol{b}^{(3D,2)} = (1, 1, 1, 2, \ldots, d-2)$, and we build the associated tensorized versions of the G-KDE of the PDFs of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$.

Hence, when no block decomposition is carried out, we can verify in Figure 5 that $h^{\mathrm{MLE}}(d, N)$ quickly converges to 1 when d increases, for the two

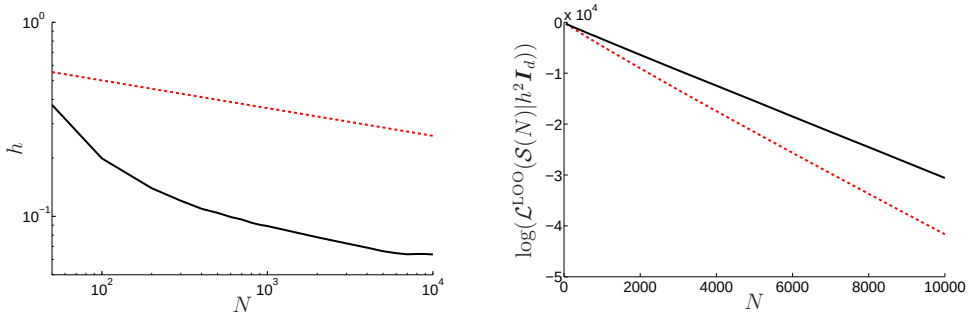


(a) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -1.102 \times 10^3$ (b) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -1.101 \times 10^3$



(c) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -8.00 \times 10^2$ (d) $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)) = -7.98 \times 10^2$

Figure 3: Four branches clover-knot case: $N = 200$ given data points (big black squares) and 10^4 additional realizations (small red and green points) generated from a G-KDE approach for $h = h^{\mathrm{Silv}}(d, N)$ (first row) and $h = h^{\mathrm{MLE}}(d, N)$ (second row). The first column corresponds to the case where no constraints on the mean and the covariance of the generated points are introduced, whereas the second column corresponds to the case where the mean and the covariance of the generated points are equal to their empirical estimations that are computed from the available data. Under each graph is shown the value of the LOO likelihood for the associated value of h.

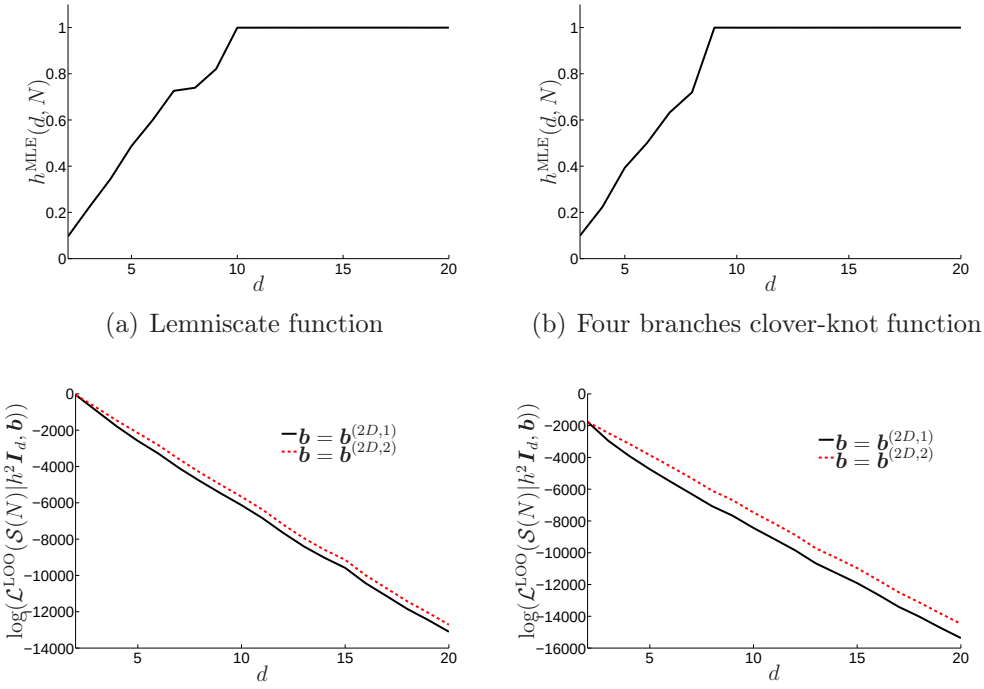


(a) Evolution of the bandwidth (b) Evolution of $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d)$

Figure 4: Evolution of the bandwidth (left) and of the LOO-likelihood (right) with respect to N for the four branches clover-knot function (3D). The red dotted lines correspond to the Silverman case: $h = h^{\mathrm{Silv}}(d, N)$. The black solid lines correspond to the MLE case: $h = h^{\mathrm{MLE}}(d, N)$. For this 3D example, the distinctions between the cases with correlation constraints or without were negligible compared to the difference between the Silverman and the MLE cases. Hence, only the cases where correlation constraints are imposed on the G-KDE are represented. Each curve corresponds to the mean values of h and $\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|h^2\boldsymbol{I}_d))$, which have been computed from 50 independent generated 200-dimensional sets of independent realizations of $\boldsymbol{X}^{\mathrm{KB}}$.

considered cases. As a consequence, the capacity of the classical G-KDE formalism to concentrate the new realizations of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$ on the correct subspaces of $\mathbb{R}^d$ decreases when d increases. To illustrate this phenomenon, for $N = 500$, Figure 6 compares the positions of the first components of the available realizations of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$, and the corresponding positions of 10^4 additional points generated from a G-KDE approach. Hence, just by working on the optimization of the value of the bandwidth, it is quickly impossible to recover the subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$ on which the true distributions of $\boldsymbol{X}^{\mathrm{L}}$ and $\boldsymbol{X}^{\mathrm{FB}}$ are concentrated. On the contrary, when the block by block decompositions given by $\boldsymbol{b}^{(2D,2)}$ and $\boldsymbol{b}^{(3D,2)}$ are considered, the approximation of the PDFs of the two first components of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$ is not affected by the presence of the additional random variables $\Xi_3, \ldots, \Xi_d$. As a consequence, for each considered values of d, the new generated points are concentrated on the correct subspaces, as it can be seen in Figure 6.

At last, the high interest of introducing the block by block decomposition for these two examples is emphasized by comparing in Figure 5 the values of the LOO likelihood in each case.



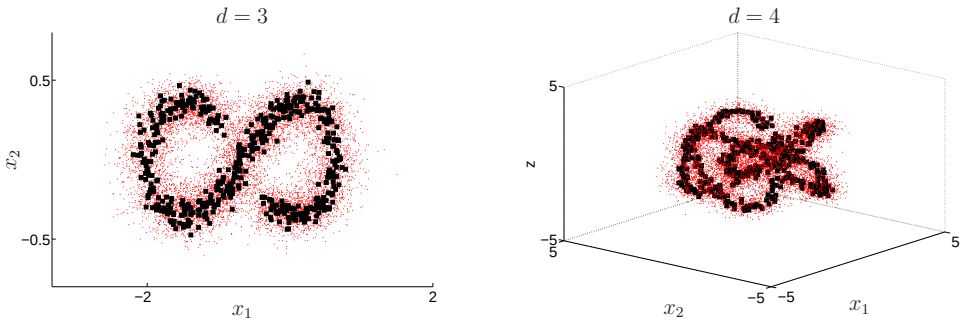
(a) Lemniscate function

(b) Four branches clover-knot function

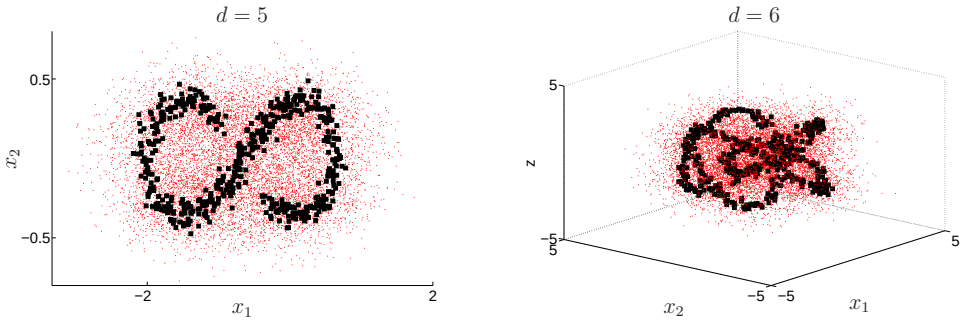
(c) Lemniscate function

(d) Four branches clover-knot function

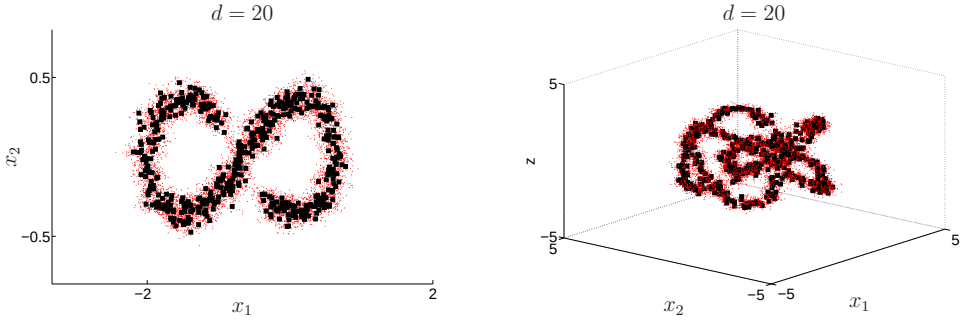
Figure 5: Evolutions of $h^{\mathrm{MLE}}(d,N)$ and $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|(h^{\mathrm{MLE}}(d,N))^2\boldsymbol{I}_d)$ with respect to d, for $N=500$. The left column correspond to the modelling of $\boldsymbol{X}^{(2D)}$, whereas the right column corresponds to the modelling of $\boldsymbol{X}^{(3D)}$.



(a) $\boldsymbol{b} = \boldsymbol{b}^{(2D,1)}$ (b) $\boldsymbol{b} = \boldsymbol{b}^{(3D,1)}$



(c) $\boldsymbol{b} = \boldsymbol{b}^{(2D,1)}$ (d) $\boldsymbol{b} = \boldsymbol{b}^{(3D,1)}$



(e) $\boldsymbol{b} = \boldsymbol{b}^{(2D,2)}$ (f) $\boldsymbol{b} = \boldsymbol{b}^{(3D,2)}$

Figure 6: Comparison between the positions of $N = 500$ given values of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$ (big black squares) and the positions of 10^4 additional values (small red points) generated from a G-KDE approach for several values of d and two configurations of the block by block decomposition for the G-KDE of the PDF of $\boldsymbol{X}^{(2D)}$ and $\boldsymbol{X}^{(3D)}$. The left column corresponds to the modelling of $\boldsymbol{X}^{(2D)}$, whereas the right column corresponds to the modelling of $\boldsymbol{X}^{(3D)}$.

value of $\boldsymbol{b}$	value of $\boldsymbol{h}^{\mathrm{MLE}}(d,N)$	$\log(\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)\|\boldsymbol{h}^{\mathrm{MLE}}(d,N),\boldsymbol{b}))$
(1,1,1,1,1)	0.395	-1.21×10^{3}
(1,2,3,4,5)	(0.115,0.163,0.0971,0.108, 0.118)	-1.15×10^{3}
(1,2,1,2,1)	(0.290,0.226)	-1.19×10^{3}
(1,1,2,2,2)	(0.113,0.140)	$\mathbf{-8.35\times10^{2}}$
(1,1,2,2,3)	(0.113,0.119,0.118)	-9.96×10^{2}

Table 2: Influence of the choice of $\boldsymbol{b}$ on the LOO log-likelihood of the G-KDE for the modeling of the PDF of $\boldsymbol{X}=(\boldsymbol{X}^{\mathrm{L}},\boldsymbol{X}^{\mathrm{FB}})$ with $N=200$ independent realizations.

In the same manner, if we define $\boldsymbol{X}$ as the concatenation of $\boldsymbol{X}^{\mathrm{L}}$ and $\boldsymbol{X}^{\mathrm{FB}}$, which are chosen independent, the interest of introducing the correct block by block decomposition of $\boldsymbol{X}$ in terms of likelihood maximization is shown in Table 2. Indeed, choosing $\boldsymbol{b}=(1,1,2,2,2)$ instead of the two classical *a priori* choices $\boldsymbol{b}=(1,1,1,1,1)$ (all the components are modelled at the same time) and $\boldsymbol{b}=(1,2,3,4,5)$ (all the components are modelled separately), allows us to strongly increase the likelihood associated with the approximation of the PDF of $\boldsymbol{X}$. Reciprocally, such an example seems to confirm the fact that maximizing $\mathcal{L}^{\mathrm{LOO}}(\mathcal{S}(N)|\boldsymbol{h}^{\mathrm{MLE}}(d,N),\boldsymbol{b})$ should help us to find the dependence structure in the components of $\boldsymbol{X}$.

4.1.4. Efficiency of the proposed algorithms for the block-by-block decomposition

This section aims at comparing the efficiency of the proposed algorithms for solving the optimization problem given by Eq. (24). To this end, using the same notations than in Section 3.1, we denote by $\boldsymbol{X}$ the random vector such that for all $1\leq\ell\leq\mathrm{Max}(\boldsymbol{b})$,

$$\boldsymbol{s}^{(\ell)}(\boldsymbol{X},\boldsymbol{b})=\boldsymbol{\xi}^{(\ell)}/\left\|\boldsymbol{\xi}^{(\ell)}\right\|+0.15\boldsymbol{\Xi}^{(\ell)}. \tag{32}$$

Here $\boldsymbol{\xi}^{(\ell)}$ and $\boldsymbol{\Xi}^{(\ell)}$ denote independent standard Gaussian random vectors, and $\|\cdot\|$ denotes the classical Euclidean norm. By construction, the random vectors $\boldsymbol{s}^{(\ell)}(\boldsymbol{X},\boldsymbol{b})$ are concentrated on d_ℓ-dimensional hyper-spheres, d_ℓ being the dimension of $\boldsymbol{s}^{(\ell)}(\boldsymbol{X},\boldsymbol{b})$. Thus, random vector $\boldsymbol{X}$ presents a known block by block structure, and its distribution is concentrated on a subset of $\mathbb{R}^d$.

Then, we assume that the maximal available information is a set of $N=$

Chosen values of $\boldsymbol{b}$	d	$N_{\mathbb{B}(d)}$	$N_{\text{greedy}}(d)$	$\widehat{N}_{\text{gene}}^{(10,0.01)}(d)$
(1,2,2,1,3,4,1)	7	877	68 (52)	32.4
(1,2,2,1,3,4,1,2,4,5)	10	115975	174 (141)	25.5
(1,2,2,1,3,4,1,2,4,5,5,6,3,4,7,6,8,1,2,7)	20	5.17×10^{17}	968 (879)	51.1

Table 3: Comparison of the efficiency of the greedy and the genetic algorithms for the identification of the block-by-block structure of $\boldsymbol{X}$.

$p^{\text{Mut}} \setminus N_{\text{pop}}$	2	5	10	20	50
0	∞	13.5	16.0	35.8	58.1
0.005	∞	15.9	19.5	39.7	77.7
0.01	13.4	15.7	25.5	36.0	62.9
0.1	22.9	44.3	41.0	64.7	78.7

Table 4: Influence of the parameters p^{Mut} and N_{pop} on the mean number of tested values of $\boldsymbol{b}$, which is denoted by $\widehat{N}_{\text{gene}}^{(N_{\text{pop}},p^{\text{Mut}})}(d)$.

500 independent realizations of $\boldsymbol{X}$. For different values of d and $\boldsymbol{b}$, the ability of the greedy and the genetic algorithms to find back the correct block by block structure of $\boldsymbol{X}$ is compared in Table 3. In this table, $N_{\text{pop}} = 10$, $p^{\text{Mut}} = 0.01$, and we denote by $\widehat{N}_{\text{gene}}^{(N_{\text{pop}},p^{\text{Mut}})}(d)$ the mean number of distinct values of $\boldsymbol{b}$ that were tested for the genetic algorithm to identify the optimal value of $\boldsymbol{b}$. These values were computed from 20 runs of the algorithm initialized in 20 different initial populations chosen at random in $\mathbb{B}(d)$. For the greedy case, the algorithm, which is deterministic, was run until it stopped, and we indicate in Table 3 two quantities: the total number of iterations $N_{\text{greedy}}(d)$, and, in parenthesis, the number of iterations that was actually needed to get the best value of $\boldsymbol{b}$. Hence, for these particular examples, the genetic algorithm was more efficient than the greedy one.

The influence of parameters N_{pop} and p^{Mut} is then analysed in Table 4, for $d = 10$ and $\boldsymbol{b} = (1, 2, 2, 1, 3, 4, 1, 2, 4, 5)$. A value of $\widehat{N}_{\text{gene}}^{(N_{\text{pop}},p^{\text{Mut}})}(d)$ equal to ∞ means that the correct value was never found after 10^5 iterations. Therefore, this example (the same thing was observed for the other examples we tried) seem to encourage the use of small (but not zero) values of p^{Mut}, as well as small values of N_{pop} such that several mutation processes can be achieved.

4.2. Application to the generation of relevant ballast grain shapes

The mechanical behaviour of the railway track strongly depends on the track superstructure and substructure components. In particular, the mechanical properties of the ballast layer are very important. Therefore, a series of studies are in progress to better analyse the influence of the ballast shape on the railway track performance. In that prospect, the shapes of $N = 975$ ballast grains have been measured very precisely. As an illustration, Figure 7 shows the scans of three ballast grains. These measurements can be considered as independent realizations of a complex random field. From this finite set of realizations, a Karhunen-Loève expansion (see [**?** **?**] for more details about this method) has been carried out to reduce the statistical dimension of this random field. Without entering too much into details, we admit in this paper that the random field associated with the varying ballast shape can finally be parametrized by a 117-dimensional random vector, which is denoted by $\boldsymbol{X}$. As a consequence of the Karhunen-Loève expansion, this random vector is centred and its covariance matrix is equal to the 117-dimensional identity matrix:

$$\mathbb{E}\left[\boldsymbol{X}\right] = \boldsymbol{0}, \quad \mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}] = \boldsymbol{I}_{117}. \tag{33}$$

From the experimental data, we have access to $N = 975$ independent realizations of $\boldsymbol{X}$, which are gathered in $\mathcal{S}(N)$. Based on this maximal available information, we would like to identify the PDF of $\boldsymbol{X}$ from a G-KDE approach. The results associated with several modellings based on the G-KDE formalism are summarized in Table 5. In this table, we notice the high interest of introducing correlation constraints. Indeed, for such a very high dimensional problem with relatively little data, if no constraints are introduced, we get very poor models associated with very low values of the LOO likelihood. In that case, assuming that all the components are independent leads to better results than assuming that they are all dependent. This can be explained by the fact that if all the component of $\boldsymbol{X}$ are chosen independent, we impose a diagonal structure for $\mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}]$, which is, in that case, very close to imposing that $\mathbb{E}[\boldsymbol{X} \otimes \boldsymbol{X}] = \boldsymbol{I}_{117}$.

On the contrary, much higher values of the LOO likelihood are obtained by adding constraints on the mean value and the covariance matrix of the G-KDE of the PDF of $\boldsymbol{X}$. In both cases, it can be noticed that it is worth working on the values of the bandwidth. Indeed, passing from $h^{\mathrm{Silv}}(d, N)$ to $h^{\mathrm{MLE}}(d, N)$ makes a big difference when looking at the LOO likelihood.

Figure 7: Three scanned ballast grains (provided by SNCF).

Value of $\boldsymbol{b}$	Value of $\boldsymbol{h}$	Correlation constraints	LOO Log-likelihood
$(1,\dots,1)$	$h^{\mathrm{Silv}}(d,N)$	no	-179379
$(1,\dots,1)$	$h^{\mathrm{MLE}}(d,N)$	no	-176886
$(1,\dots,d)$	$\boldsymbol{h}^{\mathrm{MLE}}(d,N)$	no	-162398
$(1,\dots,1)$	$h^{\mathrm{Silv}}(d,N)$	yes	-161745
$(1,\dots,1)$	$h^{\mathrm{MLE}}(d,N)$	yes	-161262
$(1,\dots,d)$	$\boldsymbol{h}^{\mathrm{MLE}}(d,N)$	yes	-161775
$\boldsymbol{b}^{MLE}$	$\boldsymbol{h}^{\mathrm{MLE}}(d,N)$	**yes**	**-160930**

Table 5: Influence of the value of the bandwidth, of the presence of constraints on the covariance, and of the choice of the block by block decomposition for the approximation of the PDF of $\boldsymbol{X}$.

At last, introducing the tensorized representation as it is done in Section 3, and working on the value of the block-by-block decomposition of $\boldsymbol{X}$ leads to another high increase of the LOO likelihood. For this application, the value of $\boldsymbol{b}^{MLE}$ has been approximated from the coupling of the greedy algorithm and the genetic algorithm presented in Section 3. The greedy algorithm was first launched, and stopped after 30000 iterations. Then, based on these results, additional 20000 iterations were performed using the genetic algorithm with $N_{\mathrm{pop}} = 500$ and $p^{\mathrm{Mut}} = 0.005$.

Finally, by working on both the correlation constraints and the block by block decomposition of $\boldsymbol{X}$, it is possible to construct, for this example, very interesting statistical models for $\boldsymbol{X}$. Such models can then be used for the analysis of the ballast statistical properties.

5. Conclusion

This work considers the challenging problem of identifying complex PDFs when the maximal available information is a set of independent realizations. In that prospect, the multidimensional G-KDE method plays a key role, as it presents a good compromise between complexity and efficiency. Two adaptations of this method have been presented. First, a modified formalism is presented to make the mean and the covariance matrix of the estimated PDF equal to their empirical estimations. Then, tensorized representations are proposed. These constructions are based on the identification of a block by block dependence structure of the random vectors of interest. The interest of these two adaptations has finally been illustrated on a series of analytical examples and on a high-dimensional industrial example.

The identification of the bandwidth matrices and of the block structure is carried out in the frequency domain. Investigating Bayesian sampling for the bandwidth matrices and the block structure selection could be interesting for future work.

Appendix

A1. Proof of Proposition 1

We can calculate:

$$\mathbb{E}\left[\widetilde{\boldsymbol{X}}\right] \;=\; \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{A}\boldsymbol{X}(\omega_n) + \boldsymbol{\beta} \;=\; \widehat{\boldsymbol{\mu}}. \tag{34}$$

$$\begin{aligned}
\mathrm{Cov}(\widetilde{\boldsymbol{X}}) &= \int_{\mathbb{R}^d} \boldsymbol{x}\otimes\boldsymbol{x}\; \widetilde{p}_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{H},\mathcal{S}(N))d\boldsymbol{x} - \widehat{\boldsymbol{\mu}}\otimes\widehat{\boldsymbol{\mu}} \\
&= \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{H} + (\boldsymbol{A}\boldsymbol{X}(\omega_n)+\boldsymbol{\beta})\otimes(\boldsymbol{A}\boldsymbol{X}(\omega_n)+\boldsymbol{\beta}) - \widehat{\boldsymbol{\mu}}\otimes\widehat{\boldsymbol{\mu}} \\
&= \boldsymbol{H} + \frac{N-1}{N}\boldsymbol{A}\widehat{\boldsymbol{R}}_{\boldsymbol{X}}\boldsymbol{A}^T \\
&= \widehat{\boldsymbol{R}}_{\boldsymbol{X}}.
\end{aligned} \tag{35}$$

A2. Description of three algorithms used in the genetic algorithm

This section presents the three algorithms that are used in the genetic algorithm defined in Section 3. Algorithm 4 presents the fusion function,

Algorithm 5 describes the mutation function, and Algorithm 6 shows the pseudo projection on $\mathbb{B}(d)$ on which they are based.

Let $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ be two elements of $\mathbb{B}(d)$;
Initialization: $\boldsymbol{b} = (0, \ldots, 0)$, index $= \{1, \ldots, d\}$, $n = 1$;
while *index is not empty* **do**
    Choose $i \sim \mathcal{U}(\{\text{index}\})$, $j \sim \mathcal{U}(\{1, 2\})$, $k \sim \mathcal{U}(\{1, 2\})$;
    Find $\boldsymbol{u}^{(1)} = \text{which}(\boldsymbol{b}^{(1)} == b_i^{(1)})$, $\boldsymbol{u}^{(2)} = \text{which}(\boldsymbol{b}^{(2)} == b_i^{(2)})$;
    if *k==1* **then**
        Define $\boldsymbol{v} = \boldsymbol{u}^{(j)} \cap \text{index}$;
    end
    else
        Define $\boldsymbol{v} = (\boldsymbol{u}^{(1)} \cup \boldsymbol{u}^{(2)}) \cap \text{index}$;
    end
    Fill $\boldsymbol{b}[\boldsymbol{v}] = n$;
    Actualize $n \leftarrow n + 1$, index $\leftarrow$ index$\backslash \boldsymbol{v}$.
end
Fusion$(\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}) := \Pi^{\mathbb{B}(d)}(\boldsymbol{b})$.

Algorithm 4: Algorithm for the fusion of two elements $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ of $\mathbb{B}(d)$.

Let $\boldsymbol{b}$ be an element of $\mathbb{B}(d)$ and $0 \leq p^{\text{Mut}} \leq 1$;
for $i = 1 : d$ **do**
    Choose $u \sim \mathcal{U}([0,1])$;
    if $u < p^{Mut}$ **then**
        $b_i \sim \mathcal{U}(\{1,\ldots,d\}) \setminus \{b_i\}$;
    end
end
Mutation$(\boldsymbol{b}, p^{\text{Mut}}) := \Pi^{\mathbb{B}(d)}(\boldsymbol{b})$.

Algorithm 5: Algorithm for the mutation of an element $\boldsymbol{b}$ of $\mathbb{B}(d)$.

Let $\boldsymbol{b}$ be an element of $\{1,\ldots,d\}^d$, index $= (1,\ldots,1)$, $n = 1$, $\boldsymbol{b}^* = (0,\ldots,0)$;
for $i = 1 : d$ **do**
    if *sum(index)==0* **then**
        break ;
    end
    else
        Find $\boldsymbol{u} = \text{which}(\boldsymbol{b} == b_i)$;
        Fill $\boldsymbol{b}^*[\boldsymbol{u}] = n$;
        Actualize $n = n + 1$, index$[\boldsymbol{u}] = 0$.
    end
end
$\Pi^{\mathbb{B}(d)}(\boldsymbol{b}) := \boldsymbol{b}^*$.

Algorithm 6: Pseudo-projection $\Pi^{\mathbb{B}(d)}(\boldsymbol{b})$ of any element $\boldsymbol{b}$ in $\{1,\ldots,d\}^d$ on $\mathbb{B}(d)$.