



**HAL**  
open science

# A comparative study of generalized Polynomial Chaos based Approximations: integration vs. regression vs. collocation vs. kriging

Gaël Poëtte

► **To cite this version:**

Gaël Poëtte. A comparative study of generalized Polynomial Chaos based Approximations: integration vs. regression vs. collocation vs. kriging. 2018. hal-01831191

**HAL Id: hal-01831191**

**<https://hal.science/hal-01831191>**

Preprint submitted on 5 Jul 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A comparative study of generalized Polynomial Chaos based Approximations: integration vs. regression vs. collocation vs. kriging

Gaël Poëtte<sup>1</sup>

<sup>1</sup> CEA CESTA DAM, F-33114 Le Barp, France  
gael.poette@cea.fr

July 5, 2018

## Abstract

In this paper, we are interested in the numerical approximations of a transformation  $u$  of a set of independent random variables  $X = (X_1, \dots, X_Q)^t \in \mathbb{R}^Q$  into an output one  $u(X) \in \mathbb{R}$  thanks to generalized Polynomial Chaos (gPC) based methods. In particular, we aim at recalling the main results on the methods, their constructions, analysis and comparisons. Amongst the many derivations of gPC, one can count integration-gPC, regression-gPC, collocation-gPC and even recently kriging-gPC. Although the previous different methods have been compared numerically on several benchmarks in many papers, it is still hard identifying which method performs better than the other and in which conditions. The aim of this paper is to identify more easily under which conditions the strategies differ, are equivalent or are more efficient.

**keywords:** uncertainty quantification, non-intrusive, generalized Polynomial Chaos, collocation, regression, kriging

## 1 Introduction

In this paper, we are interested in the numerical approximations of a transformation  $u$  of a set of independent<sup>1</sup> random variables  $X = (X_1, \dots, X_Q)^t \in \mathbb{R}^Q$  into an output one  $u(X) \in \mathbb{R}$ . Applying the transformation  $X \in \mathbb{R}^Q \rightarrow u(X) \in \mathbb{R}$  when  $u$  resumes a complex dynamical system is commonly called uncertainty propagation. We more precisely focus on non-intrusive *generalized Polynomial Chaos* based methods to approximate  $u(X)$ . In particular, we aim at recalling the main results on the methods, their constructions, analysis and comparisons. Amongst the many derivations of gPC, one can count<sup>2</sup> integration-gPC [65, 44, 18], regression-gPC [11, 12, 10, 66, 67], collocation-gPC [41, 53, 81, 26, 42, 43, 27] and even recently kriging-gPC [64, 34, 63, 63, 35]. Although the previous different methods have been compared numerically on several benchmarks in many papers [53, 81, 54, 61, 34, 63, 64, 35], it is still hard identifying which method performs better than the other and in which conditions. The aim of this paper is to characterise some of those conditions, independently of the dimension  $Q$  of the input uncertain vector  $X$ , of its probability measure  $d\mathcal{P}_X$  and of any assumptions on  $u$ .

The paper is organized as follow: in the next section, we progressively introduce the different gPC based methods encountered in the literature, from the seminal integration-gPC, to regression-gPC, collocation-gPC and kriging-gPC. The construction of the respective approximations are recalled, their numerical analysis performed. In particular, this constitutes the first original point of the document, care will be taken to explain theoretically some already observed numerical phenomenon. Orthogonal polynomials and their properties<sup>3</sup>, of course, play a central role. All along the descriptions and analysis of the different methods, we consider a *fil rouge* problem (Runge function) on which are applied every derivations of gPC in the same conditions. It allows progressively highlighting the commons and differences of the gPC based expansions. Finally, the last section is dedicated to several additional numerical comparisons of the gPC based methods. In particular, iterative-gPC (i-gPC [57, 58]) is also amongst the latters. The paper is only stewn with very simple test-cases: we consider it is enough to recover and illustrate experimentally the results of the performed numerical analysis. It is also important for the sake of reproducibility of the numerical results. Furthermore, the literature is furnished with industrial applications of these methods and we rely on references [65, 44, 18, 11, 12, 10, 66, 67, 41, 53, 81, 26, 42, 43, 27, 54, 61, 34, 63, 64, 35] to give relevant, complex and complementary examples.

## 2 A progressive analysis of gPC based methods

In this section, we propose a brief state-of-the-art of gPC based methods together with their analysis and numerical illustrations. We begin by the very first one, here denoted by *integration-gPC*. Many examples can be found in [65, 44, 18, 19, 50, 71, 86, 82, 73, 74, 49, 48, 55] covering fields of applications ranging from the acceleration of Bayesian inference for transient diffusion to failure detection for chemical reactions *via* uncertainty propagation in prey-predator models etc.

<sup>1</sup>Independence is not mandatory in a gPC context, see [40, 22, 39, 11], but is very convenient.

<sup>2</sup>They will be recalled in this document, with coherent notations easing the comparison and with complete references.

<sup>3</sup>The most relevant ones for the present study are recalled in appendix A.

## 2.1 The construction and properties of integration-gPC approximations

Let us begin by a very general description. It may, at first glance, look like a recipe but it is representative of its practical use. Let us first suppose the random variable<sup>4</sup>  $X$  has probability measure  $d\mathcal{P}_X$ . The methodology consists in several steps:

1. it begins by building  $(\phi_k^X)_{k \in \mathbb{N}}$  the gPC basis orthonormal with respect to the inner product defined by the probability measure  $d\mathcal{P}_X$  of the input random variable  $X$ , i.e. such that

$$\int \phi_k^X \phi_t^X d\mathcal{P}_X = \delta_{k,t}, \forall (k, t) \in \mathbb{N}^2.$$

More details on algorithms to perform the latter step are given in appendix A.

2. The second step corresponds to the discretisation of the random variable and its probability measure  $(X, d\mathcal{P}_X)$  by a numerical integration method with  $N$  points:

$$(X, d\mathcal{P}_X) \approx (X_i, w_i)_{i \in \{1, \dots, N\}}. \quad (1)$$

The set  $(X_i, w_i)_{i \in \{1, \dots, N\}}$  is commonly called an *experimental design* [25, 6] and is crucial, this will be recalled and illustrated, for the approximation method.

3. The next step consists in running  $N$  independent runs of a black-box code at the *a priori* chosen points  $(X_i, w_i)_{i \in \{1, \dots, N\}}$  and gathering a new collection of output points:

$$(u(X_i), w_i)_{i \in \{1, \dots, N\}}. \quad (2)$$

This step is supposed to bear the main computational effort.

4. Once the  $N$  runs obtained, the rest is only postprocessing at the observation points of interest. The estimation of the polynomial coefficients is then made by numerical integration in this section, i.e.  $\forall k \in \{0, \dots, P\}$

$$u_k^X = \int u(X) \phi_k^X(X) d\mathcal{P}_X \approx u_k^{X,N} = \sum_{i=1}^N u(X_i) \phi_k^X(X_i) w_i. \quad (3)$$

Many authors apply other numerical methods to compute the coefficients (such as regression, collocation, kriging): they are addressed in the next sections together with their analysis.

5. Finally, one can reconstruct the truncated polynomial approximation (or the collocation or kriging ones) using the approximated coefficients (3)

$$u(X) \approx u_{P,N}^X(X) = \sum_{k=0}^P u_k^{X,N} \phi_k^X(X),$$

and perform the desired post-treatments in order to approximate the statistical quantities of interest (mean, variance, histograms, etc.) related to  $u(X)$ .

At the end of the process, one has access to an approximation<sup>5</sup>  $u_{P,N}^X(X)$ . The error between  $u(X)$  and  $u_{P,N}^X(X)$  in the  $L^2$ -norm can be decomposed in two main parts using the orthonormality of the gPC basis<sup>6</sup>:

$$\begin{aligned} \|u(X) - u_{P,N}^X(X)\|_{L^2}^2 &= \left\| \sum_{k=0}^{\infty} u_k^X \phi_k^X(X) - \sum_{k=0}^P u_k^{X,N} \phi_k^X(X) \right\|_{L^2}^2, \\ &= \underbrace{\sum_{k=0}^P (u_k^X - u_k^{X,N})^2}_{\text{integration error}} + \underbrace{\sum_{k=P+1}^{\infty} (u_k^X)^2}_{\text{truncation error}}. \end{aligned} \quad (4)$$

In (4), the error of the non-intrusive approximation have (explicitly) two parameters,  $N$  for the integration error and  $P$  for the truncation error.

Now, let us assume

- (i) the  $(u_k^{X,N})_{k \in \mathbb{N}}$  are infinitely accurately computed/available, i.e. we have  $\forall k \in \mathbb{N}, u_k^{X,N} = u_k^X$ , so that the integration error in (4) is zero,
- (ii) and the distribution of the input random variable is gaussian, i.e.  $d\mathcal{P}_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$ .

In the previous conditions, the  $(\phi_k^X)_{k \in \mathbb{N}}$  are the Hermite polynomials and the convergence of (4) is ensured by (the special case of) Cameron-Martin's theorem [16], under condition  $\|u(x)\|_{L^2}$  is bounded. This particular case is commonly called *Polynomial Chaos* (PC) whereas (4), with arbitrary measure  $d\mathcal{P}_X$  and hypothesis (i), is denoted *generalized Polynomial Chaos* (gPC). Its efficiency, i.e. the gain of gPC vs. PC, has been *numerically observed* in many fields of applications [72, 75, 74, 79, 37, 20, 83, 85, 84, 52]. Some complementary

<sup>4</sup>or vector.

<sup>5</sup>The upperscript  $X$  reminds of the approximation basis,  $P$  of the truncation order and  $N$  of the number of points for the numerical approximation of the coefficients  $(u_k^{X,N})_{k \in \{0, \dots, P\}}$ .

<sup>6</sup>assuming there are no errors in the gPC basis, see section A.

theorems ensure the convergence for some particular measures<sup>7</sup>  $d\mathcal{P}_X$  but the complete convergence theorem for gPC has been demonstrated in [24]. Note that the general convergence theorem (for gPC) demands an additional condition on the input distribution (it must have a uniquely solvable moment problem<sup>8</sup>). Paper [24] also recalls the convergence is ensured for quantiles<sup>9</sup>, relative moments<sup>10</sup>, in probability<sup>11</sup>: in other words, for every classical mathematical tools to perform a (converging) uncertainty quantification analysis. In the following, we consider random input variables  $X$  for which the gPC convergence holds [24].

Let us go back to (4), from a more practical point of view. The above description is quite simple but to be complete, two points remain to be tackled:

- We did not explain under which considerations the experimental design is chosen in practice for the discretisation (1) of the couple  $(X, d\mathcal{P}_X)$ . The notation (1) for the punctual discretisation of  $(X, d\mathcal{P}_X)$  is very general and has been chosen in order to show that the material of the rest of the chapter concerning non-intrusive gPC can be applied independently of this choice. Suppose the points  $(X_i)_{i \in \{1, \dots, N\}}$  are chosen sampled from the probability law of  $X$  and  $(w_i = \frac{1}{N})_{i \in \{1, \dots, N\}}$ , then it corresponds to the Monte-Carlo integration method for the estimation of the coefficients. With the same writing, we can conveniently consider Gauss quadrature points, Latin Hypercube Samples, Sparse Grids etc. The latter sets of points in dimension  $Q$  differ only by their asymptotic error analysis with respect to integration. We insist we here implicitly deal with *converging* discretisation of  $(X, d\mathcal{P}_X)$ : without this assumption, most of the following results do not hold<sup>12</sup>. In the following, examples are given with different experimental designs to integrate  $\{u\phi_0^X, \dots, u\phi_P^X\}$ : we aim at applying the *same* experimental design to every coefficients  $(u_k^X)_{k \in \{0, \dots, P\}}$ .
- Some authors in the literature do not exactly use numerical integration methods in order to estimate the polynomial coefficients (3). Amongst the other possibilities one can cite *regression-gPC*, *collocation-gPC* or *kriging-gPC*. Their subtleties are briefly investigated, analysed and illustrated in the next subsections 2.2–2.3–2.4.

The choice of the discretisation of the random variable  $X$  together with its probability measure  $d\mathcal{P}_X$  obviously directly impacts the quality of the gPC approximation as testifies (4). Depending on the transformation  $u$ , the integration error may be preponderant with respect to the truncation one (see for example [47]). Considerations to help choose it together with pedagogical examples can be found in [25, 6] but we would first like to focus on Gauss points as they will be central, especially to deal with collocation-gPC.

The  $N$  Gauss points of any arbitrary measure  $d\mathcal{P}_X$  are the roots of the  $(N + 1)^{th}$  degree polynomial orthonormal with respect to the inner product defined by  $d\mathcal{P}_X$ , see section A.

$$\forall i \in \{1, \dots, N\}, w_i = \int L_i(x) d\mathcal{P}_X(x). \quad (5)$$

Introducing the weights as above implicitly presents them as the coefficients ensuring exact integration of polynomials up to order  $N$ . Such  $N$ -point quadrature rule is said to have *degree of exactness*  $N$  and is denoted as *interpolary*, see [28]. Obviously, from the definition of the weights (5), given any  $N$  points, any quadrature rule can be made interpolary. Such definition is convenient but not optimal: the optimal  $N$ -point quadrature rule has degree of exactness  $2N$  and is called a Gauss quadrature rule [28]. The weights can also be defined by normalizing every eigenvectors  $(\Phi_N^X(\gamma_i))_{i \in \{1, \dots, N\}}$  and taking their first squared component: as  $\phi_0^X(x) = 1$  in our case, dealing with probability measures, we have

$$\forall i \in \{1, \dots, N\}, w_i = \frac{1}{\sum_{k=0}^P (\phi_k^X(\gamma_i))^2}.$$

The last definition of the weight emphasizes their positiveness, important in practice for robustness. The main drawback of Gauss quadratures remains linked to the curse of dimensionality. Building a multidimensional Gauss quadrature rule implies tensorizing the points in each directions. The number of Gauss points increases exponentially fast with the dimension  $Q$ . Regarding asymptotic error analysis, for a Gauss quadrature rule, we have the following very general property (see [28]):

$$\left| \int g(x) d\mathcal{P}_X - \sum_{i=1}^N g(\gamma_i) w_i \right| = \int H_{2N}(X, g) \mathcal{P}_X, \quad (6)$$

where  $H_{2N}(x, g)$  is the Hermite<sup>13</sup> interpolation polynomial of order  $2N$  relative to function  $g$ . If furthermore,  $g$  is  $2N$  times differentiable then the same asymptotic error can be expressed in term of  $\xi$ , existing in the

<sup>7</sup>For example if the support of  $d\mathcal{P}_X$  is bounded (uniform, arcsinus, beta laws etc.) the Stone-Weierstrass theorem ensures the convergence of the gPC expansion. For the Poisson distribution (unbounded discrete distribution), the convergence is ensured by the very same theorem ensuring the convergence of the Gram-Charlier expansion etc.

<sup>8</sup>It is not restrictive for the distribution of the Askey scheme [5, 84, 82, 72] nor discrete and mixed distributions but may be problematic for the lognormal one [24].

<sup>9</sup>Relevant when one is interested in approximating a probability of failure.

<sup>10</sup>Relevant when one is interested in central quantities.

<sup>11</sup>Relevant when one is interested in approximating the probability density function of the output variable by a histogram.

<sup>12</sup>For example, in such case, (4) may diverge.

<sup>13</sup>We recall the Hermite interpolation polynomials relative to the points  $(X_i)_{i \in \{1, \dots, N\}}$  satisfy  $H_{2N}(X_i, g) = g(X_i)$ , and  $H'_{2N}(X_i, g) = g'(X_i), \forall i \in \{1, \dots, N\}$ .

support of the probability measure  $d\mathcal{P}_X$ , such that

$$\left| \int g(x) d\mathcal{P}_X - \sum_{i=1}^N g(\gamma_i) w_i \right| = \frac{\langle \phi_N^{X,m}, \phi_N^{X,m} \rangle}{2N!} g^{(2N)}(\xi), \quad (7)$$

in which  $g^{(n)}$  denotes the  $n^{\text{th}}$  derivative of  $g$ . The constant in the error analysis strongly depends on the smoothness (of order  $2N$ ) of the integrand. The coefficients  $\left( \langle \phi_k^{X,m}, \phi_k^{X,m} \rangle \right)_{k \in \mathbb{N}}$  correspond to the norm of the monic orthogonal polynomial  $(\phi_k^{X,m})_{k \in \{0, \dots, P\}}$  associated to the probability measure  $d\mathcal{P}_X$ , see section A. This normalization coefficient may also depend strongly on  $N$ , cf. (42) and (46). Obviously, for smooth solutions, the convergence rate is fast. The Gauss quadrature rules for an arbitrary measure  $d\mathcal{P}_X$  are

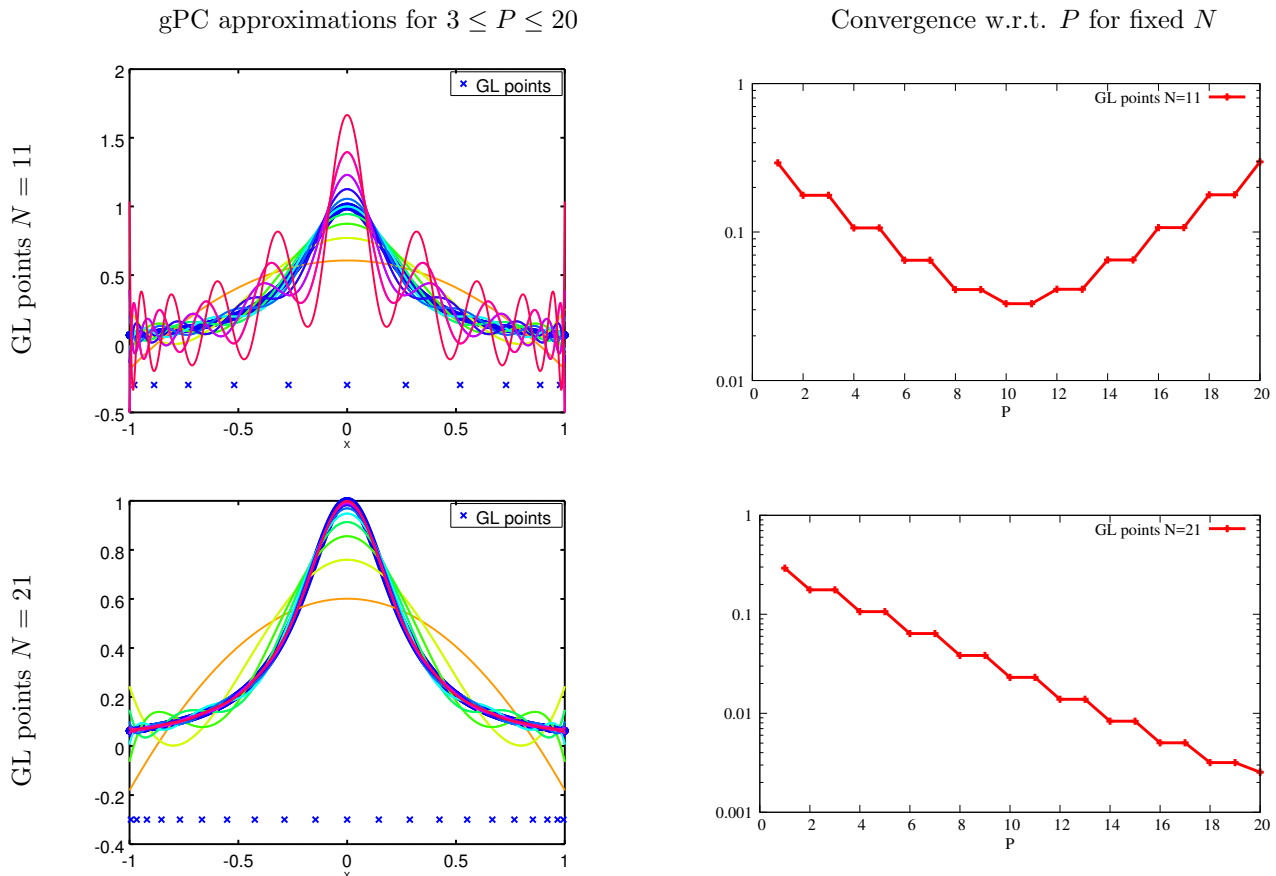


Figure 1: Application of Gauss-Legendre quadrature rule for *integration-gPC* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ .

designed/defined [4] to ensure a good accuracy up to order  $2N$  for the statistical moments of the input  $X$  i.e. up to order  $2N$  for polynomials (hence the non optimality of the definition (5)). With the above remark in mind, assume  $g$  is polynomial of order  $2N - 1$ , then  $g^{(n)} = 0, \forall n \geq 2N$  and the quadrature rule is exact, see (6).

Let us come back to our final goal, i.e. applying the quadrature rule to function  $g$  having the particular form  $g = u\phi_l^X$  with  $l \in \{0, \dots, P\}$  to build a gPC approximation of order  $P$ . Formally, we can expand  $u \in L^2$  on the gPC basis, i.e.  $u = \sum_{k=0}^{\infty} u_k^X \phi_k^X$ , and consider  $g_l = u\phi_l^X = \sum_{k=0}^{\infty} u_k^X \phi_k^X \phi_l^X, \forall l \in \{0, \dots, P\}$ . In such condition, error analysis (7) becomes

$$\begin{aligned} \left| \int g_l(x) d\mathcal{P}_X - \sum_{i=1}^N g_l(\gamma_i) w_i \right| &= \frac{\langle \phi_N^{X,m}, \phi_N^{X,m} \rangle}{2N!} \sum_{k=0}^{\infty} u_k^X (\phi_k^X \phi_l^X)^{(2N)}(\xi_l), \\ \left| u_l^X - u_l^{X,N} \right| &= \frac{\langle \phi_N^{X,m}, \phi_N^{X,m} \rangle}{2N!} \sum_{k=0}^{\infty} u_k^X \mathbf{1}_{k+l \geq 2N} (\phi_k^X \phi_l^X)^{(2N)}(\xi_l), \end{aligned} \quad (8)$$

in which the dependence with respect to  $l$  of  $\xi_l$  has been made explicit. To study (8) more in detail, recall  $\forall k \in \mathbb{N}$  we can rewrite  $\phi_k^X(X) = \Gamma_k^X \phi_k^{X,m}(X) = \Gamma_k^X \prod_{i=1}^k (X - \gamma_i^k)$  with  $(\phi_k^{X,m})_{k \in \mathbb{N}}$  the monic orthogonal polynomial relative to  $(\phi_k^X)_{k \in \mathbb{N}}$  and with  $(\gamma_i^k)_{i \in \{1, \dots, k\}}$  its roots<sup>14</sup>. Due to the fact we decomposed  $(\phi_k^X)_{k \in \mathbb{N}}$  as a product of monomials, the previous notations allow rewriting

$$(\phi_k^X \phi_l^X)^{(2N)}(\xi_l) = 2N! \Gamma_k^X \Gamma_l^X P_{k+l-2N}(\xi_l),$$

<sup>14</sup>The roots of the orthogonal polynomials are within the support of the probability measure  $d\mathcal{P}_X$ , distinct and real, see [28]. Nevertheless, the decomposition could be done in the complex plane.

where  $P_{k+l-2N}$  is a monic polynomial of order  $k+l-2N$ . With the above equality, (8) becomes

$$\begin{aligned} \left| u_l^X - u_l^{X,N} \right| &= \sum_{k=0}^{\infty} u_k^X \mathbf{1}_{k+l \geq 2N} \frac{\Gamma_k^X \Gamma_l^X}{(\Gamma_N^X)^2} P_{k+l-2N}(\xi_l). \\ \left| u_l^X - u_l^{X,N} \right|^2 &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} u_k^X u_j^X \mathbf{1}_{j+l \geq 2N} \mathbf{1}_{k+l \geq 2N} \frac{\Gamma_k^X \Gamma_j^X (\Gamma_l^X)^2}{(\Gamma_N^X)^4} P_{k+l-2N}(\xi_l) P_{j+l-2N}(\xi_l). \\ \sum_{l=0}^P \left| u_l^X - u_l^{X,N} \right|^2 &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} u_k^X u_j^X \frac{\Gamma_k^X \Gamma_j^X}{(\Gamma_N^X)^2} \underbrace{\sum_{l=0}^P \mathbf{1}_{j+l \geq 2N} \mathbf{1}_{k+l \geq 2N} \left( \frac{\Gamma_l^X}{\Gamma_N^X} \right)^2}_{\text{term (*) we can control by choosing } P} Q_{k,j,l,N}. \end{aligned} \quad (9)$$

The last term in the left hand side of (9) is exactly the  $L^2$ -norm of the integration error in (4). The term (\*) in (9) corresponds to the term we can control by choosing  $P$  with respect to  $N$ , independently of any smoothness assumptions for  $u$ . Now, for a given  $k$ , we have (cf. section A)

$$\Gamma_k^X = \sqrt{\frac{H_{2(k-1)}^X}{H_{2k}^X}} \geq 2^{2k-1}.$$

It implies that  $\forall l, N \in \mathbb{N}$ , we have

$$- \left( \frac{\Gamma_l^X}{\Gamma_N^X} \right)^2 = - \frac{H_{2(l-1)}^X}{H_{2l}^X} \frac{H_{2N}^X}{H_{2(N-1)}^X} \geq -2^{4(l-N)}. \quad (10)$$

According to the above expression, term (\*) in (9) can be controlled by making sure  $l \leq N, \forall l \in \mathbb{N}$ , i.e. choosing  $P \leq N$ , ensuring the minimization of the residue in (8) independently of the smoothness of the solution  $u$ . Obviously, the previous analysis is not optimal if  $P$  is kept  $P \leq N$ . But it is relevant in the opposite case and a gPC reconstruction with increasing  $P$  for fixed  $N$  (exponentially) accumulates errors. This explosion of the  $L^2$  error with  $P$  as soon as  $P \geq N$  has been observed on many numerical experiments, see [47, 18] for example. The above numerical analysis shows it is *independent of  $d\mathcal{P}_X$  and of  $u$*  for the associated Gauss quadrature rule.

We suggest illustrating this behaviour on a simple uncertainty propagation problem: consider the transformation of a uniform random variable  $X$  via the Runge function

$$X \sim \mathcal{U}_{[-1,1]} \longrightarrow \frac{1}{1 + 15X^2}, \quad (11)$$

and apply gPC with coefficients integrated thanks to a  $N$  points Gauss-Legendre (GL) quadrature rule. Figure 1 presents the results of the study with  $N = 11$  and  $N = 21$  points. The top right picture presents a convergence study with respect to  $P$  for  $N = 11$  GL points. The  $L^2$ -norm of the error first decreases exponentially fast (logarithmic scale for the ordinate) before *increasing as quickly* after  $P = 11 = N$ . The explosion of the error after  $P = N$  is in agreement with the previous numerical analysis (cf. (9) and (10)). The quality of the obtained gPC approximations can be observed on figure 1 top-left in the same conditions. For the bottom pictures of figure 1, care has been taken to keep  $P \leq N = 21$  GL points. The exponential convergence of the gPC approximation is ensured up to  $P = 20$  as testifies the bottom right picture: the increasing quality with  $P$  is observable on figure 1 (bottom-left) with a less and less oscillating approximation. Note that the readability of the left column of figure 1 can be discussed: this column is mainly qualitative and its sense will be revealed mainly when tackling sections 2.3–2.2–2.4 and comparisons with the approximations obtained with regression-gPC, collocation-gPC, kriging-gPC in the same conditions. Note that in every publications dealing with gPC,  $P$  is usually kept lower than  $N$ , i.e.  $P \leq N$ , in each stochastic directions. We here wanted to emphasize this strategy is more than a rule of thumb and is relevant independently of the input distributions  $d\mathcal{P}_X$  and of the regularity of  $u$  (for Gauss points).

## 2.2 The construction and properties of regression-gPC approximations

Regression has been historically at the basis of many works in statistics for modeling [25, 6, 21, 28]. It is widely used and presents the advantage of being applicable in presence of noisy outputs, i.e. *experimental* noise as well as *numerical* noise:

- Experimental noise refers to variability in the output at the different points of the experimental design *due to a finite accuracy of the measure instruments*: two identical experiments may give slightly different results. The differences could be made smaller with more accurate/less sensitive to external perturbations measurement devices.
- Numerical noise refers to variability *due to the use of a stochastic resolution scheme* [60, 1, 46, 45, 15, 76]. In this context, running twice the same simulation in identical configurations<sup>15</sup> leads to fluctuations of the observables. Those fluctuations can be made smaller with finer discretisation parameters (number of Monte-Carlo particles for example, see [36]).

In this section, we briefly detail the principles of regression together with its properties. For this, let us consider a sequence of linearly independent functions  $F_P(x) = (f_0(x), \dots, f_P(x))^t$ . Classically in regression

<sup>15</sup>but with different initial seeds for the random number generators.

approximation, the sequence is chosen as  $F_P(x) = (1, x, \dots, x^P)^t$ , see [25, 6]. Let  $X \rightarrow u(X)$  be our random variable of interest, then the regression model  $u_P^F(x)$  of  $u$  is defined as the vector product

$$u_P^F(X) = U_P^t F_P(X),$$

where  $U_P = (u_0, \dots, u_P)^t$  is defined by the vector of  $\mathbb{R}^P$  minimizing the least square error, i.e. such that

$$U_P = \underset{V \in \mathbb{R}^P}{\text{Argmin}} [J(V_P)] = \underset{V \in \mathbb{R}^P}{\text{Argmin}} \|u(X) - V_P^t F_P(X)\|_{L^2}^2. \quad (12)$$

Differentiating  $J(V_P) = V_P^t F_P(X)$  with respect to  $V_P = (v_0, \dots, v_P)^t$ , we get

$$\nabla_{V_P} J(V_P) = 2\|(u(X) - V_P^t F_P(X))F_P(X)\|_{L^2},$$

and  $U_P$  is consequently the (unique due to the convexity of  $J$ ) solution of

$$\nabla_{V_P} J(V_P) = 0 \iff \int u(X)F_P(X) d\mathcal{P}_X = \int V_P^t F_P(X)F_P(X) d\mathcal{P}_X. \quad (13)$$

From the above expression (13), in the particular case  $F_P(x) = \Phi_P^X(x) = (\phi_0^X(x), \dots, \phi_P^X(x))^t$ , the minimum of  $J$  is attained at the vector of gPC coefficients  $(u_1^X, \dots, u_P^X)^t$ .

Now, equations (12) and (13) were stated assuming *perfect* integration accuracy. The above expressions (12)–(13) are in practice discretised *via* the introduction of an experimental design  $(X_i, w_i)_{i \in \{1, \dots, N\}}$  applied to the process of interest resulting in  $(u(X_i), w_i)_{i \in \{0, \dots, P\}}$ . Equation (12) is consequently replaced in practice by

$$U_P^N = \underset{V_P \in \mathbb{R}^P}{\text{Argmin}} [J_N(V_P)] = \underset{V_P \in \mathbb{R}^P}{\text{Argmin}} \sum_{i=1}^N w_i (u(X_i) - V_P^t F_P(X_i))^2, \quad (14)$$

commonly called weighted least-squared minimization. Then  $J_N(V_P)$  is minimum for  $V_P \in \mathbb{R}^P$  satisfying

$$[(F_P^N)^t W_N F_P^N] V_P = [W_N F_P^N] \begin{pmatrix} u(X_1) \\ \dots \\ u(X_N) \end{pmatrix}, \quad (15)$$

where  $W_N = \text{diag}(w_1, \dots, w_N)$ , and

$$F_P^N = \begin{pmatrix} f_0(X_1) & \dots & f_0(X_N) \\ \dots & f_k(X_j) & \dots \\ f_P(X_1) & \dots & f_P(X_N) \end{pmatrix}.$$

This leads to

$$[W_N F_P^N] = \begin{pmatrix} w_1 f_0(X_1) & \dots & w_N f_0(X_N) \\ \dots & w_j f_k(X_j) & \dots \\ w_1 f_P(X_1) & \dots & w_N f_P(X_N) \end{pmatrix},$$

$$[(F_P^N)^t W_N F_P^N] = \begin{pmatrix} \sum_{i=1}^N w_i f_0^2(X_i) & \dots & \sum_{i=1}^N w_i f_0(X_i) f_P(X_i) \\ \dots & \sum_{i=1}^N w_i f_k(X_i) f_i(X_i) & \dots \\ \sum_{i=1}^N w_i f_0(X_i) f_P(X_i) & \dots & \sum_{i=1}^N w_i f_P^2(X_i) \end{pmatrix}. \quad (16)$$

Of course, whatever the choices of  $(N, P) \in \mathbb{N}^2$ , the matrix (16) is invertible. The solution  $U_P^N$  satisfies the well-known (unbiased estimator see [25])

$$U_P^N = [(F_P^N)^t W_N F_P^N]^{-1} [W_N F_P^N] \begin{pmatrix} u(X_1) \\ \dots \\ u(X_N) \end{pmatrix}. \quad (17)$$

The conditioning of matrix (16) depends on both the choice of the basis  $F_P(x)$  and of the experimental design  $(X_i, w_i)_{i \in \{0, \dots, N\}}$ . Suppose  $F_P(x) = (1, x, \dots, x^P)^t$ , then matrix (16) has expression

$$\begin{pmatrix} \sum_{i=1}^N w_i & \dots & \sum_{i=1}^N w_i X_i^P \\ \dots & \sum_{i=1}^N w_i X_i^{k+l} & \dots \\ \sum_{i=1}^N w_i X_i^P & \dots & \sum_{i=1}^N w_i X_i^{2P} \end{pmatrix} \xrightarrow{N \rightarrow \infty} \begin{pmatrix} s_0^X & \dots & s_P^X \\ \dots & s_{k+l}^X & \dots \\ s_P^X & \dots & s_{2P}^X \end{pmatrix}, \quad (18)$$

and tends to the Hankel matrix (38) defined in section A whose determinant is known to tend to zero quickly as  $P$  grows. It is consequently harder and harder to numerically inverse with  $P$  increasing. Figure 2 presents

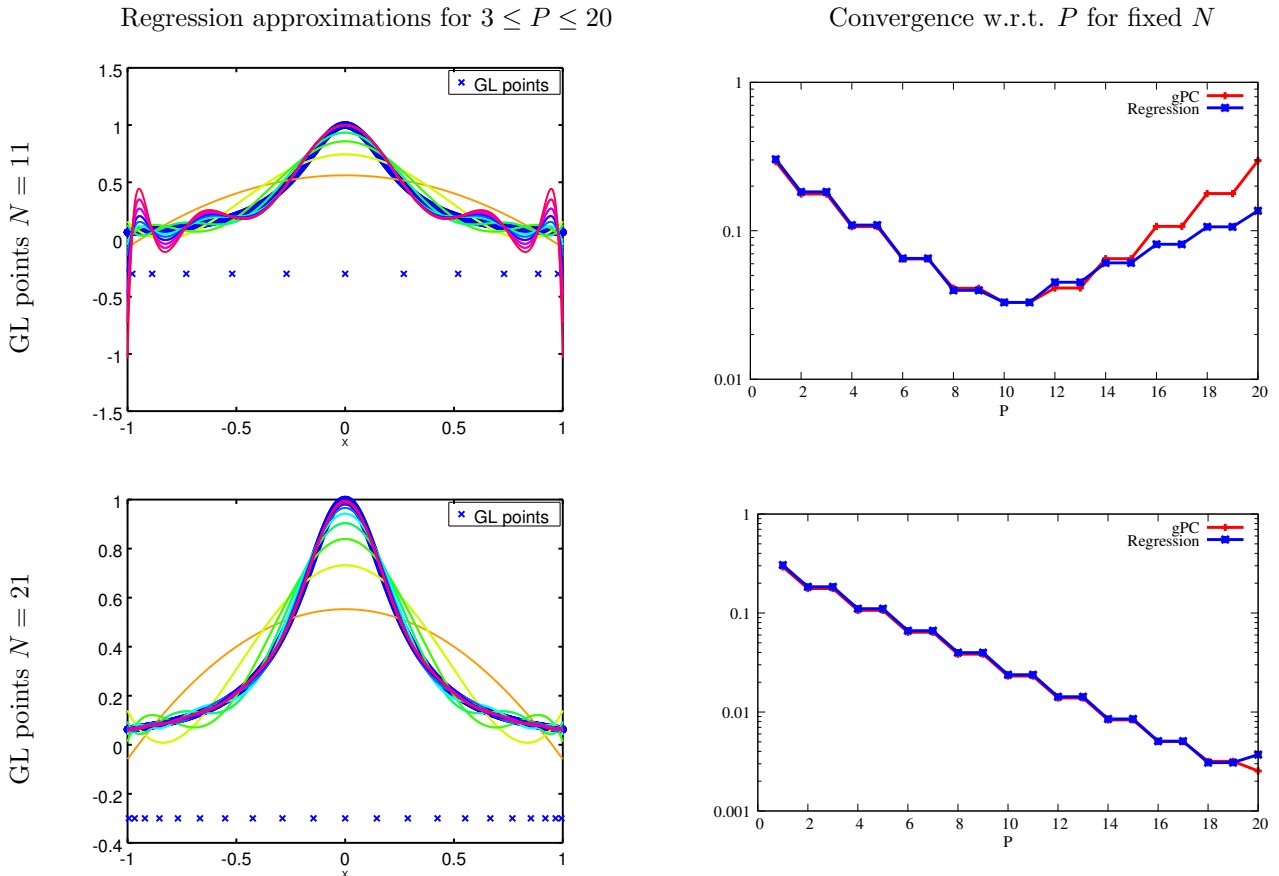


Figure 2: Application of Gauss-Legendre quadrature rule for *regression* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ .

the results obtained with regression<sup>16</sup> on the Runge function in exactly the same conditions as in section 2.1 together with the results obtained by integration-gPC. If we focus on the convergence studies (right column) with respect to  $P$  for fixed numbers of GL points  $N = 11$  and  $N = 21$ , integration-gPC and regression have the same behaviour: for a small number of points of the experimental design, the polynomial order must be kept low as the error decreases then increases for  $N = 11$  as soon as  $P$  goes beyond  $P = N = 11$ . Nonetheless, for polynomial orders higher than  $P = N = 11$ , the error is a little bit more controlled with the regression approximation. The shapes of the regression approximations for the different order for  $N = 11$  (top left of figure 2) are quite different than the one obtained with integration (top left of figure 1). When  $P$  is kept  $P \leq N$ , the results obtained with integration-gPC or regression are equivalent (up to the accuracy/cost of a matrix inversion) as testifies the bottom pictures of figure 2: by definition, the  $L^2$ -minimization is invariant with a change of basis.

Now suppose a particular form for  $F_P(x) = \Phi_P^X(x) = (\phi_0^X(x), \dots, \phi_P^X(x))^t$  with  $(\phi_k^X)_{k \in \{0, \dots, P\}}$  the components of a chosen gPC basis. The obtained approximations are denoted *regression-gPC* ones in this document and in the literature [11, 13, 67, 14, 10, 12]. Let us introduce  $U_P^{\text{int}, N} = (u_0^{X, N}, \dots, u_P^{X, N})^t$  the vector of coefficients of the gPC approximation obtained by integration. From (15), it is easy noticing that the regression solution  $U_P^N$  is related to the integration coefficients  $U_P^{\text{int}, N}$  by the relation:

$$\begin{pmatrix} \sum_{i=1}^N w_i (\phi_0^X(X_i))^2 & \dots & \sum_{i=1}^N w_i \phi_0^X(X_i) \phi_P^X(X_i) \\ \dots & \sum_{i=1}^N w_i \phi_k^X(X_i) \phi_l^X(X_i) & \dots \\ \sum_{i=1}^N w_i \phi_0^X(X_i) \phi_P^X(X_i) & \dots & \sum_{i=1}^N w_i (\phi_P^X(X_i))^2 \end{pmatrix} U_P^N = U_P^{\text{int}, N}. \quad (19)$$

According to (6), if we also assume  $P \leq N$ , the Gauss quadrature rule ensures the *exact* orthonormality of the gPC basis even with  $N < \infty$  (i.e. even in a finite integration accuracy context). Consequently, with such choice, we have

$$U_P^N = U_P^{\text{int}, N},$$

explaining their equivalent performances for  $P \leq N$ . Otherwise (i.e. if  $N > P$ ), the regression-gPC coefficients and the integration ones differ from the fact the integration coefficients may not minimize the weighted least squared error, this is emphasized in figure 3. For  $P > N$ , the regression-gPC approximations have a better control (better conditioning of (16)) of the  $L^2$  error than both classical regression and integration-gPC.

<sup>16</sup>i.e. with  $F_P(x) = (1, \dots, x^P)^t$ .



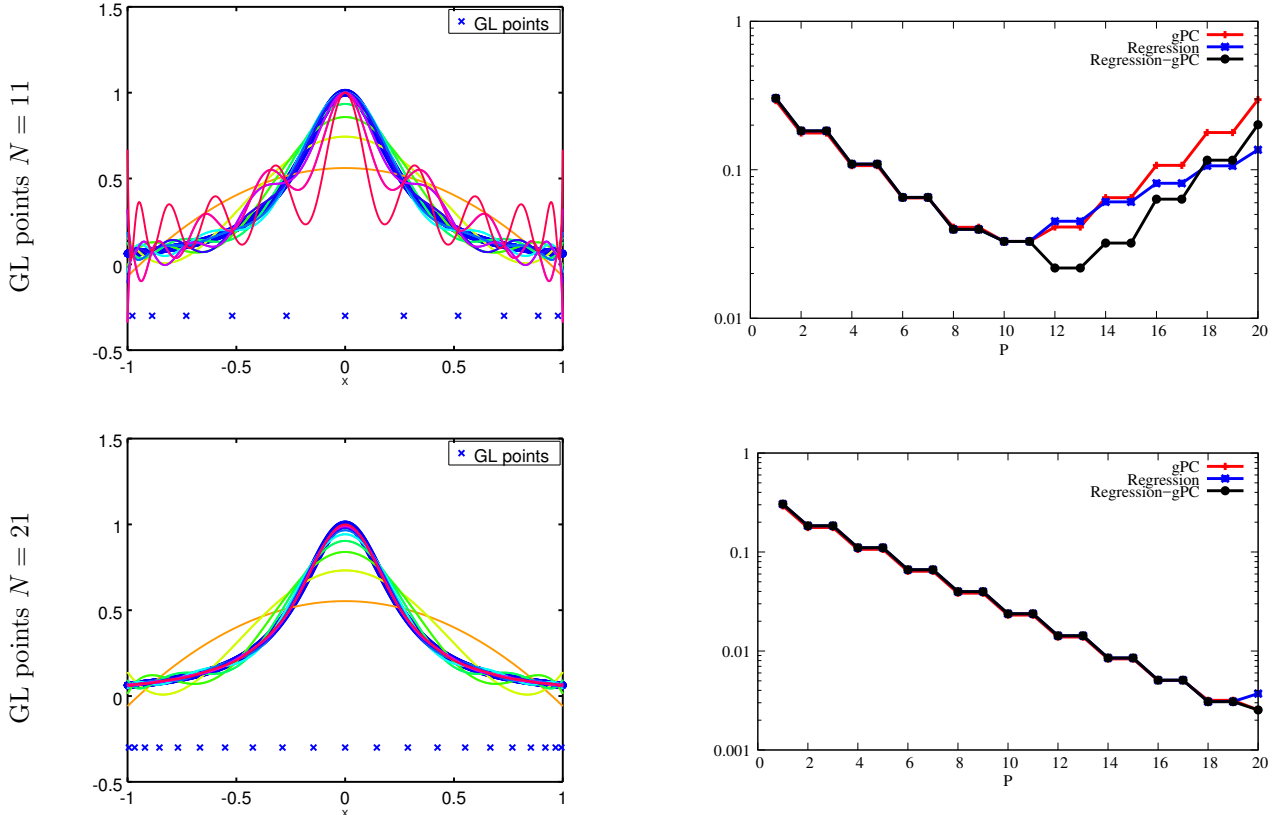
Regression-gPC approximations for  $3 \leq P \leq 20$ Convergence w.r.t.  $P$  for fixed  $N$ 

Figure 3: Application of Gauss-Legendre quadrature rule for *regression-gPC* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ .

The main interest of regression-gPC is to be able to deal with any experimental design, independently of its integration accuracy, still ensuring a relatively good conditioning of matrix (16). It is particularly conve-

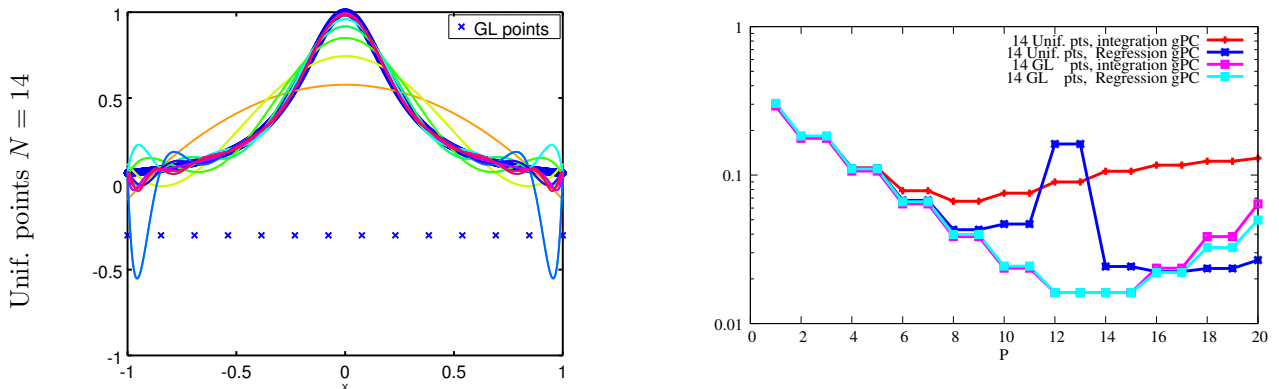
Regression-gPC for  $3 \leq P \leq 20$ Convergence w.r.t.  $P$  for fixed  $N$ 

Figure 4: Application of Gauss-Legendre quadrature rule for *regression-gPC* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ .

nient for experimental settings (in opposition to numerical experiments) and has been originally designed [25, 6, 21] for its ability to take into account experimental noise.

The above property is emphasized in figure 4 in which we briefly investigate the sensitivity to the choice of the experimental design for integration-gPC and regression-gPC. Figure 4 (left) shows the results obtained with regression-gPC with  $N = 14$  equispaced<sup>17</sup> points (uniform experimental design). Figure 4 (right) allows comparing the results obtained with integration-gPC and regression-gPC for  $N = 14$  GL points and  $N = 14$  equispaced ones. For the GL points, the behaviour is similar to what was presented in figure 3 which here corresponds to our reference. With the equispaced experimental design, the integration accuracy is lower than with the GL points. Integration-gPC consequently gives less satisfactory results with such design: the two approaches have equivalent  $L^2$ -performances only up to  $P = 5$ . For higher polynomial orders, the integration error becomes preponderant with respect to the truncation one. With regression-gPC on another hand, the accuracy of the GL and equispaced experimental designs are comparable up to order  $P = 9$ .

<sup>17</sup>Equispaced points in  $[-1, 1]$  with weights  $\frac{1}{N}$  is still a discretisation of the uniform distribution on  $[-1, 1]$ .

### 2.3 The construction and properties of collocation-gPC approximation

Regressions are convenient especially when one has to deal with experimental/numerical noise. When considering numerical experiments, some resolution schemes may be *reproducible* in the sense two runs of the same configuration gives exactly the same results. Dealing with reproducible simulation codes, one may demand the stochastic approximation method to be able to strictly recover the numerical results at the experimental design points, see [41, 53, 81, 26, 42, 43, 27]. This can be obtained using interpolation methods such as Lagrange interpolation or high-order splines for example.

Lagrange interpolation can be obtained applying formulae (17) in the particular case  $N = P$ . In practice, in order to avoid the inversion of a possibly badly conditioned matrix, the Lagrange formulae  $\forall i \in \{1, \dots, N\}$  is applied

$$L_i(x) = \prod_{\substack{j=1 \\ i \neq j}}^N \frac{x - X_j}{X_i - X_j}. \quad (20)$$

The resulting collocation approximation is then given by

$$u_N^L(X) = \sum_{j=1}^N u(X_j) L_j(X). \quad (21)$$

In term of asymptotical error analysis, we have the following well-known property: suppose  $u \in \mathcal{C}^0([a, b])$ , and an experimental design  $(X_i)_{i \in \{1, \dots, N\}}$  with  $N$  distinct nodes, then there exists  $\xi \in [a, b]$  such that

$$u(X) - u_N^L(X) = \frac{u^{(N)}(\xi)}{N+1!} \prod_{i=1}^N (X - X_i). \quad (22)$$

Note that the collocation approximation does not necessarily converge. Its converging behaviour is strongly correlated to the choice of the experimental design relative to the evolution of  $u^{(N)}(\xi)$  with  $N$ . Figure 5 (top) presents the collocation approximations of Runge function<sup>18</sup> obtained for  $N$  going from 1 to 20 of a uniform experimental design (equispaced points in  $[-1, 1]$ ): the collocation approximations seem to diverge<sup>19</sup> as  $N$  increases, as testifies the convergence study of figure 5 (top-left).

The question arising now is: is it possible to choose an experimental design ensuring the convergence of the collocation approach. It is commonly known, see [56], that  $\max_{x \in [a, b]} \left| \prod_{i=1}^N (x - X_i) \right|$  from (22) is  $\mathcal{O}(\frac{1}{2^N})$  at the roots of Chebyshev's polynomials (CC points) and that they are enough to compensate the growth of  $u^{(N)}(\xi)$  with  $N$  for Runge function. The roots of Chebyshev polynomials are Gauss points and some authors *numerically observed* the latter property of Chebyshev roots holds with Gauss ones for arbitrary distributions. This leads to *stochastic collocation, or collocation-gPC* in the literature, see [41, 53, 81, 26, 42, 43, 27, 32, 80, 78, 69, 62]. It refers to the use of Lagrange polynomials at the Gauss quadrature points associated to the probability measure of the input random variable  $X$  (or the roots of the gPC polynomials associated to the probability measure of  $X$ ). With such a choice, the  $L^2$ -norm of (22) becomes

$$\|u(X) - u_N^L(X)\|_{L^2}^2 = \left( \frac{u^{(N)}(\xi)}{N+1!} \right)^2 \int \left( \prod_{i=1}^N (X - \gamma_i) \right)^2 d\mathcal{P}_X. \quad (23)$$

The product of monomials in the above expressions with  $(\gamma_i)_{i \in \{1, \dots, N\}}$  being Gauss points is nothing more than the monic polynomial  $\phi_N^{X, m}$  associated to  $d\mathcal{P}_X$ , see section A. This leads to

$$\begin{aligned} \|u(X) - u_N^L(X)\|_{L^2}^2 &= \left( \frac{u^{(N)}(\xi)}{N+1!} \right)^2 \int \left( \phi_N^{X, m}(X) \right)^2 d\mathcal{P}_X, \\ &\stackrel{(42)}{=} \left( \frac{u^{(N)}(\xi)}{N+1!} \right)^2 \underbrace{\left( \frac{1}{\Gamma_N^X} \right)^2 \int \left( \phi_N^X(X) \right)^2 d\mathcal{P}_X}_{\delta_{N, N=1}}, \\ &\stackrel{(42)}{=} \left( \frac{u^{(N)}(\xi)}{N+1!} \right)^2 \frac{H_{2N}^X}{H_{2(N-1)}^X}, \\ &\stackrel{(46)}{\leq} \left( u^{(N)}(\xi) \right)^2 \frac{1}{2^{(4N+2)}(N+1!)^2}. \end{aligned} \quad (24)$$

The fast converging term  $\frac{1}{2^{(4N+2)}(N+1!)^2}$  proper to the use of Gauss points in the above expression allows competing with eventual fast increasing derivatives  $u^{(N)}(\xi)$ .

Figure 5 presents the collocation-gPC approximations of Runge function (i.e. Lagrange polynomials at GL points) in the same conditions as in the previous sections. As testifies the convergence study of figure 5 (right), collocation-gPC exhibits an exponential convergence behaviour, but the global accuracy at each order/number of points  $N = P$  remains slightly higher than for integration-gPC and regression-gPC. This is mainly due to systematic oscillating behaviour between the points, see figure 5 bottom-left. Of course, for  $P = 20$  in figure 5 (right), regression-gPC degenerates toward collocation-gPC.

<sup>18</sup>This example is well-known, I claim no originality here.

<sup>19</sup>In this case, it is possible to prove it diverges, see [17, 23] or <https://math.stackexchange.com/questions/775405/px-1-1x2-lagrange-polynomials-do-not-always-converge-why/807784#807784> for example.

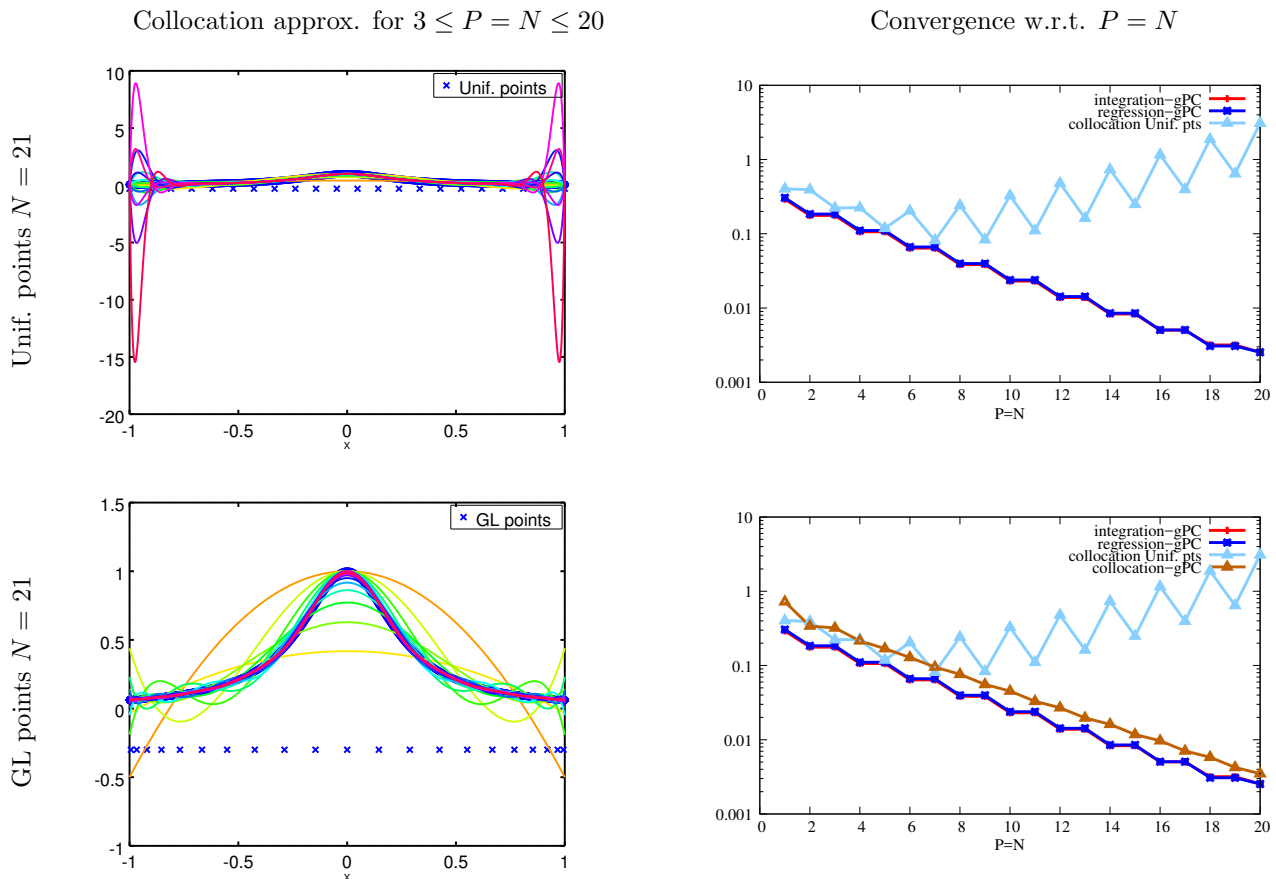


Figure 5: Application of Gauss-Legendre quadrature rule for *collocation-gPC* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ .

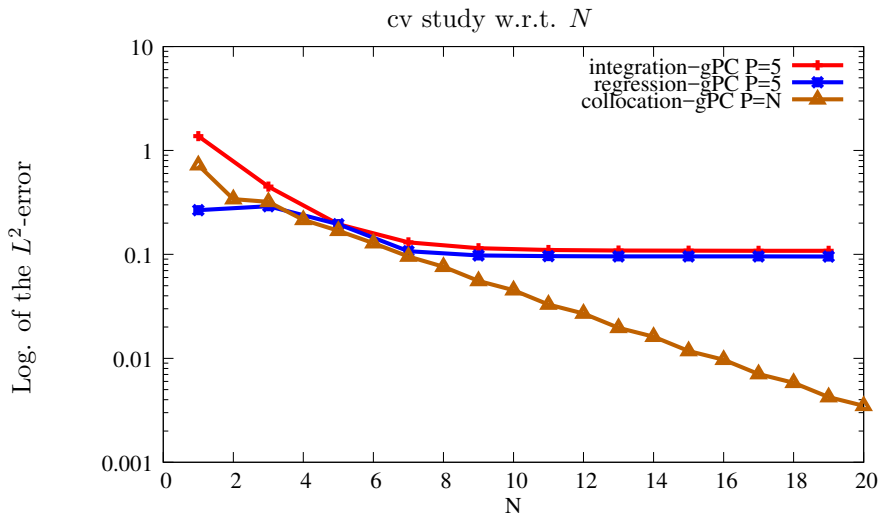


Figure 6: Convergence studies for integration-gPC, regression-gPC (both for fixed  $P = 5$ ) and collocation-gPC with respect to  $N$  for approximating Runge function (11).

Collocation-gPC has only one parameter as  $N = P$ : in figure 5, we compared collocation-gPC to integration-gPC and regression-gPC with  $N = 21$ . For the latter, the comparison may seem unfair, especially for high polynomial orders  $P$ . In figure 6, we perform some convergence studies keeping  $P = 5$  fixed for integration-gPC and regression-gPC and compare them to the collocation-gPC approximations with increasing  $N$ . The convergence studies have two regimes for the integration-gPC and regression-gPC approximations, only one for collocation-gPC:

- in the first regime, defined by  $N \leq 7$ , the three methods exhibit an exponential convergence rate (log-scale). We recall we chose  $P = 5$  for this example, implying a small numbers of points ( $N \leq P$ ).
- Beyond  $N = 7$  points, integration-gPC and regression-gPC's accuracies stagnate: this stagnation floor corresponds to the fact the truncation error (second term in (4)) becomes preponderant with respect to the integration one as  $N$  increases. This is emphasized by the fact this floor is the same for the two approximation methods as they share the same experimental design.

For collocation-gPC, the error keeps on decreasing exponentially with  $N = P$ . This is due to the fact that for collocation-gPC, the asymptotic error only depends on parameter  $N$ , see (24).

From the two previous sections, one may wonder whether it is possible to take advantage of both methods. Regression-gPC allows taking into account noisy outputs and exhibits a fast convergence rate with respect to  $P$ . Collocation-gPC ensures recovering exactly the outputs at the experimental design points and exhibits a convergence rate mainly depending on the spacing between the points once the integration accuracy reached. A compromise between regression-gPC and collocation-gPC would be a penalized regression with the introduction of Lagrange multipliers to ensure, as constraints, the approximations are interpolatory with  $u$  at the design points [70]. This idea is at the basis of *kriging*, briefly presented in the next section.

## 2.4 The construction and properties of kriging-gPC approximations

The reader interested in kriging may find very different denominations such as *simple kriging*, *Ordinary kriging*, *Universal kriging* etc. Our starting point is kriging-gPC, which is here a less elaborated version<sup>20</sup> of what can be found in [64, 34, 63, 35]. We focus on this variant of kriging as it is the most general we know and, of course, is related to a gPC basis which is the aim of this paper. Basically, kriging-gPC consists in choosing  $F_P(X) = \Phi_P(X)$  as an approximation basis in a (universal<sup>21</sup>) kriging approximation: it confers to the resolution similar advantages as the ones emphasized in the section comparing regression and regression-gPC (good conditioning). We insist the section is non-exhaustive regarding kriging technics and we rely on very pedagogical publications dedicated to them [70, 9, 34, 63] for a complete state-of-the-art. This section mainly aims at comparing and understanding the differences between kriging-gPC and gPC approximations and provide an original numerical analysis of the methods.

Kriging is also known as Gaussian process modeling. It assumes the output random variable of interest  $u(X)$  is a realization of a Gaussian random process. Let us sketch the idea behind the methodology in the next lines. Recall we aim at approximating the transformation of the known random variable  $X$  into the unknown one  $u(X)$ . The gPC approximation relied on a polynomial approximation, i.e.  $u(X) \approx \sum_{k=0}^P u_k^X \phi_k^X(X)$  where  $(\phi_k^X)_{k \in \{0, \dots, P\}}$  is the gPC basis and  $(u_k^X)_{k \in \{0, \dots, P\}}$  the gPC coefficients. Kriging-gPC relies on rewriting

$$u(X) = \sum_{k=0}^P u_k^X \phi_k^X(X) + Z_P^X(X),$$

i.e. as a gPC development plus its residue  $Z_P^X(X) = \sum_{k=P+1}^{\infty} u_k^X \phi_k^X(X)$  in the  $P$ -truncated gPC basis associated to  $X$ . In a kriging context,  $\sum_{k=0}^P u_k^X \phi_k^X(X)$  is commonly called the *trend*. Kriging introduces an *additional mathematical ingredient* whose aim is to approximate the random variable  $Z_P^X$ , the residue of the gPC development. The idea is to assume  $Z_P^X(X) \sim \sigma_K^2 Z(X)$  is a zero-mean gaussian process of variance  $\sigma^2$  independent of  $X$ , with  $K$  referring to the covariance kernel of the process. The gaussian process is fully characterised by  $K$  defined by  $K(u, v) = \mathbb{E}[Z(u)Z(v)]$  and such that  $\sigma_K^2 = \mathbb{E}[Z^2(X)]$ . Suppose  $K$  is known, then several constraints must be satisfied for  $Z$  to be a relevant gaussian process to approximate  $u(X)$ . Introduce  $\mu_Z(u) = \mathbb{E}[Z(u)]$ , then we must for example have

$$\begin{cases} \mathbb{E}[u(X)] &= u_0^X + \mathbb{E}[\mu_Z(X)], \\ \mathbb{E}[u^2(X)] &= \sum_{k=0}^P (u_k^X)^2 + \sum_{k=0}^P u_k^X \mathbb{E}[\phi_k^X(X)\mu_Z(X)] + \mathbb{E}[\mu_Z^2(X)], \\ &\dots, \end{cases} \quad (25)$$

and so on. Ensuring the constraints are satisfied for moments<sup>22</sup> higher than 2 is directly linked to how  $K$  is chosen or built with respect to  $u$ . In practice, kriging models:

- first generally assume a particular parametered shape of the covariance function  $K(u, v, \theta)$  where  $\theta$  is to be calibrated.
- The second step consists in calibrating  $\theta$ , i.e. computing  $\hat{\theta}$  minimizing differences with the above constraints in a norm which remains to be defined at this stage of the discussion. This can be done by various means (Maximum Likelihood, Cross Validation estimation, etc. see [64, 34, 63, 35, 9, 7]).
- Once  $\hat{\theta}$  obtained, we have access to the random variable  $\mu(X, \hat{\theta})$  and its *predictive variance*  $\sigma^2(X, \hat{\theta})$  with explicit matrix vector formulas (briefly detailed in the following).

In practice,  $K$  is often chosen homogeneous, i.e.  $K(u, v, \theta) = K(u - v, \theta)$ , and for a given choice of  $K$  and  $\theta$  we have (see [34])

$$\begin{aligned} \mu(X, \theta) &= U_P^N(\theta)^t \Phi_P(X) + k(X, \theta)^t W_N K^{-1}(\theta) \begin{pmatrix} u(X_1) - U_P^N(\theta)^t \Phi_P(X_1) \\ \dots \\ u(X_N) - U_P^N(\theta)^t \Phi_P(X_N) \end{pmatrix}, \\ \sigma^2(X, \theta) &= \sigma_K^2(\theta) \left( 1 - [\Phi_P^t(X), k^t(X, \theta)] \begin{bmatrix} 0 & (W_N \Phi_P^N)^t \\ (W_N \Phi_P^N)^t & W_N K(\theta) \end{bmatrix} \begin{bmatrix} \Phi_P(X) \\ k(X, \theta) \end{bmatrix} \right), \end{aligned}$$

where  $\mu(X, \theta), \sigma^2(X, \theta)$  are the mean and variance of the gaussian process approximating  $u(X)$ . The notations are almost the same as the ones of section 2.2:  $(X_i)_{i \in \{1, \dots, N\}}$  are the points of the experimental design,

<sup>20</sup>less elaborated in the sense in [64, 34, 63, 35], the authors provide an algorithm in order to choose automatically the gPC order  $P$  whereas in this document we choose it *a priori*.

<sup>21</sup>If I am not mistaking.

<sup>22</sup>By construction, kriging ensures having the same estimated first and second moments (hence mean and variance) as  $u(X)$ .

$W_N = (w_1, \dots, w_N)^t$  the vector of their weights and

$$U_P^N(\theta) = \begin{pmatrix} u_0^X(\theta) \\ \dots \\ u_P^X(\theta) \end{pmatrix}, \Phi_P(X) = \begin{pmatrix} \phi_0^X(X) \\ \dots \\ \phi_P^X(X) \end{pmatrix}, \Phi_P^N = \begin{pmatrix} \phi_0^X(X_1) & \dots & \phi_0^X(X_N) \\ \dots & \Phi_k^X(X_j) & \dots \\ \phi_P^X(X_1) & \dots & \phi_P^X(X_N) \end{pmatrix}.$$

It additionally introduces  $K(\theta)$ , the matrix of general term  $K_{i,j}(\theta) = K(X_j - X_i, \theta)$  and  $k(X, \theta) = (k(X - X_1, \theta), \dots, k(X - X_N, \theta))^t$ . Besides, the estimations of the coefficients of the development together with the variance parameter are given by

$$\begin{aligned} U_P^N(\theta) &= [(\Phi_P^N)^t W_N K^{-1}(\theta) \Phi_P^N]^{-1} W_N \Phi_P^N K^{-1}(\theta) \begin{pmatrix} u(X_1) \\ \dots \\ u(X_N) \end{pmatrix}, \\ \sigma_K^2(\theta) &= \left( \begin{pmatrix} u(X_1) \\ \dots \\ u(X_N) \end{pmatrix} - \Phi_P^N U_P^N(\theta) \right)^t W_N K^{-1}(\theta) \left( \begin{pmatrix} u(X_1) \\ \dots \\ u(X_N) \end{pmatrix} - \Phi_P^N U_P^N(\theta) \right), \end{aligned} \quad (26)$$

in which  $\theta$  remains to be chosen. Equations (26) express the minimization of the  $L^2$ -norm (least-square error, similar to regression): as hinted at in [34], if  $K = I_N$ <sup>23</sup> where  $I_N$  is the identity of size  $N$ , then (26) degenerates toward (17) for the regression-gPC approximation.

The discussion about the relevant shape of the covariance function  $K$  or the way the parameter  $\theta$  is tuned is beyond the scope of this document and we refer to [9, 7, 8] for the reader interested in deepening those considerations. The covariance kernel  $K$  can be evaluated but in general, it is chosen *a priori*. The most classical choices are gaussian, exponential or Matérn kernels<sup>24</sup>. Kriging provides  $\sigma^2$  as a measure of precision. However this measure relies on the *correctness* of the covariance function, see [7, 9]. In other words, the term *predictive* variance may be strong and it, in general, reflects an assumption. If it does not hold, the error estimation might be bad and no error estimation properties are guaranteed. However, typically, still a good interpolation is achieved for the random variable  $\mu(X)$ , mean of the Gaussian process and we focus on it in the next numerical analysis and tests.

We first go through the numerical analysis of (the mean of) the kriging-gPC approximation. We aim at helping interpreting the numerical results and comparisons with the previously presented approximations (regression-gPC and collocation-gPC mainly) displayed in figure 7. Being an interpolation method, the analysis can benefit some very well-known results: first, let us rewrite the mean of the kriging-gPC process under a more friendly form. By noticing that

$$\begin{aligned} \mu(X, \theta) &= \sum_{k=0}^P u_k^X(\theta) \phi_k^X(X) \\ &+ (k(X - X_1, \theta), \dots, k(X - X_N, \theta)) K^{-1}(\theta) \begin{pmatrix} u(X_1) - \sum_{k=0}^P u_k^X(\theta) \phi_k^X(X_1) \\ \dots \\ u(X_N) - \sum_{k=0}^P u_k^X(\theta) \phi_k^X(X_N) \end{pmatrix}, \end{aligned} \quad (27)$$

the expression can be recast as

$$\mu(X, \theta) = \sum_{k=0}^P u_k^X(\theta) \phi_k^X(X) + \sum_{i=1}^N a_i^P k(X - X_i, \theta). \quad (28)$$

Expression (28) may appear downgrading in comparison to (27) as many important properties of the approximation do not anymore explicitly appear in the coefficients  $(a_i^P)_{i \in \{1, \dots, N\}}$ . Still, it is enough for the following material. Let us introduce the functional  $F$  such that

$$F(X, \theta) = u(X) - \mu(X, \theta) - g(\bar{X}, \theta) \prod_{i=1}^N (X - X_i).$$

We define  $\bar{X}$  as an arbitrary point in  $Supp(X)$  such that  $\forall i, \bar{X} \neq X_i$  and

$$g(\bar{X}, \theta) = \frac{u(\bar{X}) - \mu(\bar{X}, \theta)}{\prod_{i=1}^N (\bar{X} - X_i)}. \quad (29)$$

The functional  $F$  has consequently  $N + 1$  roots  $X_1, \dots, X_N, \bar{X}$ . Assume furthermore that  $u(X)$  and  $k(u, v, \theta)$  are  $\mathcal{C}^{N+1}$  where, we recall,  $N$  is the number of points of the experimental design. Then according to Rolle's theorem,  $\exists \xi_0$  such that  $F^{(N+1)}(\xi_0, \theta) = 0$ . Let us now consider two situations,  $P \leq N$  in the first case and  $P > N$  in the second one:

<sup>23</sup>closely related to the particular choice  $k(u, v) = \delta_u(v)$ .

<sup>24</sup>which recovers continuously with a parameter both the gaussian and the exponential kernels.

– first, suppose  $P \leq N$  so that differentiating  $N + 1$  times  $F$  resumes to

$$F^{(N+1)}(X, \theta) = u^{(N+1)}(X) - \sum_{i=1}^N a_i^P k^{(N+1)}(X - X_i, \theta) - g_{P \leq N}(\bar{X}, \theta)(N + 1!).$$

Using the fact that  $F^{(N+1)}(\xi_0, \theta) = 0$  allows identifying  $g_{P \leq N}$  as

$$g_{P \leq N}(\bar{X}, \theta) = \frac{1}{N + 1!} \left( u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta) \right).$$

Using the above expression of  $g_{P \leq N}$  with respect to  $\xi_0$  in (29) leads to the following error estimator:

$$u(X) - \mu(X, \theta) = \frac{1}{N + 1!} \left( u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta) \right) \prod_{i=1}^N (X - X_i), \quad (30)$$

Equation (30) can be compared to (22) for collocation by noticing that

$$\begin{aligned} u(X) - \mu(X, \theta) &= \frac{1}{N + 1!} \left( u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta) \right) \prod_{i=1}^N (X - X_i), \\ &= \underbrace{\frac{u^{(N+1)}(\xi_0)}{N + 1!} \prod_{i=1}^N (X - X_i)}_{(*) \text{ recalls (22) for collocation}} - \underbrace{\frac{1}{N + 1!} \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta) \prod_{i=1}^N (X - X_i)}_{(**)}. \end{aligned} \quad (31)$$

The first term (\*) in (31) recovers the collocation error term. The error only depends on  $P$  via the coefficients  $(a_i^P)_{i \in \{1, \dots, N\}}$  in the second term (\*\*). Taking the  $L^2$ -norm of (31) leads to

$$\|u(X) - \mu(X, \theta)\|_{L^2}^2 = \left( u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta) \right)^2 \frac{1}{(N + 1!)^2} \int \left( \prod_{i=1}^N (X - X_i) \right)^2 d\mathcal{P}_X, \quad (32)$$

which is to be compared to (23). The latter testifies one can decrease the *constant multiplying the convergence rate*<sup>25</sup> of the approximation method if  $K$  (via a wise choice of  $k^{(N+1)}$  relative to  $u^{(N+1)}$ ) is sufficiently well suited. Note also that conversely, nothing prevents it from increasing it, with respect to collocation, if it is not. Now, for Gauss points (same calculations as (24)), the convergence in  $L^2$ -norm as  $N$  goes to infinity of the kriging-gPC is ensured provided the constant in (32) does not grow too fast with  $N$  as for collocation-gPC.

– Suppose now  $P > N$  and differentiate  $N + 1$  times  $F$  to obtain

$$F^{(N+1)}(X, \theta) = u^{(N+1)}(X) + \sum_{k=0}^P u_k^X(\theta) (\phi_k^X(X))^{(N+1)} - \sum_{i=1}^N a_i^P k^{(N+1)}(X - X_i) - g_{P > N}(\bar{X})(N + 1!).$$

To simplify the above expression, we can rewrite  $\phi_k^X(X) = \Gamma_k^X \phi_k^{X,m}(X) = \Gamma_k^X \prod_{i=1}^k (X - \gamma_i^k)$ ,  $\forall k \in \{0, \dots, P\}$  with  $\phi_k^{X,m}$  the monic orthogonal polynomial relative to  $\phi_k^X$ . In the latter expression,  $(\gamma_i^k)_{i \in \{1, \dots, k\}}$  are the roots<sup>26</sup> of  $\phi_k^X$ . For  $k > N$ , the  $(N + 1)^{th}$  derivative of  $\phi_k^X$  can be expressed as

$$(\phi_k^X(X))^{(N+1)} = \Gamma_k^X \sum_{i_1 + \dots + i_k = N+1} C_{i_1, \dots, i_k}^{N+1} \prod_{j=1}^k (X - \gamma_j^k)^{(i_j)},$$

with the multinomial coefficients given by  $C_{i_1, \dots, i_k}^{N+1} = \frac{(N+1)!}{\prod_{j=1}^k i_j!}$ . Each  $(X - \gamma_j^k)_{j \in \{1, \dots, k\}}$  being monomials,  $(i_j)_{j \in \{1, \dots, k\}} \in \{0, 1\}^k$ : otherwise, the above expression would be zero as  $k < N + 1$ . In other words the multinomial coefficients simplify to

$$(\phi_k^X(X))^{(N+1)} = (N + 1)! \Gamma_k^X \sum_{i_1 + \dots + i_k = N+1} \prod_{j=1}^k (X - \gamma_j^k)^{(i_j)}.$$

Now, once again using the fact that  $F^{(N+1)}(\xi_0, \theta) = 0$  ensures

$$\begin{aligned} g_{P > N}(\bar{X}, \theta) &= g_{P \leq N}(\bar{X}, \theta) + \sum_{k=0}^P u_k^X(\theta) \Gamma_k^X \sum_{i_1 + \dots + i_k = N+1} \prod_{j=1}^k (\xi_0 - \gamma_j^k)^{(i_j)}, \\ &\stackrel[N \text{ fixed}]{P \gg N} \equiv g_{P \leq N}(\bar{X}, \theta) + \mathcal{O}(\Gamma_P^X). \end{aligned}$$

<sup>25</sup>The constant is equal to  $(u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i, \theta))^2$ . The convergence rate is here related to  $\frac{1}{(N+1)!^2} \int (\prod_{i=1}^N (X - X_i))^2 d\mathcal{P}_X$  together with a choice of the experimental design  $(X_i)_{i \in \{1, \dots, N\}}$ .

<sup>26</sup>For a gPC basis, we know those roots are real, distinct in  $Supp(X)$  but the material holds for an arbitrary choice of  $F_P(X)$ , with complex roots.

Using the above expression in (29) leads to the following error estimator for the mean  $\mu(X, \theta)$ :

$$u(X) - \mu(X, \theta) = \frac{1}{N+1!} \left( \begin{aligned} &+u^{(N+1)}(\xi_0) - \sum_{i=1}^N a_i^P k^{(N+1)}(\xi_0 - X_i) \\ &+(N+1)! \sum_{k=0}^P u_k^X(\theta) \Gamma_k^X \sum_{i_1+\dots+i_k=N+1} \prod_{j=1}^k (\xi_0 - \gamma_j^k)^{i_j} \end{aligned} \right) \prod_{i=1}^N (X - X_i). \quad (33)$$

With (33), it is easy verifying for fixed  $N$  and  $P \gg N$ ,  $u(X) - \mu(X, \theta) \sim \Gamma_P^X \underset{P \gg N}{\sim} \infty$ . The analysis even shows that in this regime, the asymptotic behaviour is independent of the choice of  $K(u, v, \theta)$ . In other words, the same analysis allows explaining the behaviour of regression-gPC for  $P > N$  in the examples of the previous figures (obtained with the particular choice  $K(u, v, \theta) = \delta_u(v)$ ).

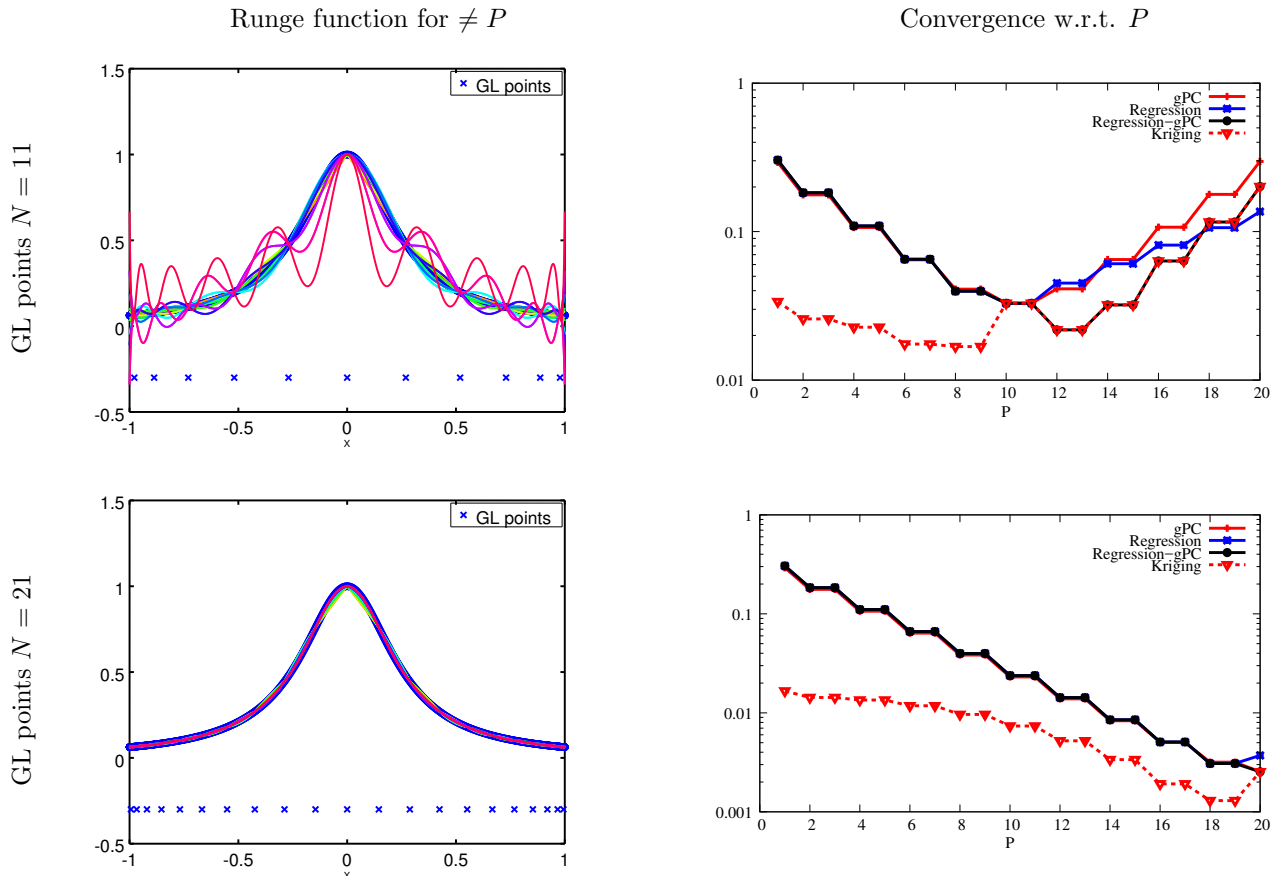


Figure 7: Application of Gauss-Legendre quadrature rule for *kriging-gPC* and the transformation of a uniform random variable through the Runge function with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the polynomial approximations for  $3 \leq P \leq 20$ . The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ . The kriging kernel is chosen exponential (34) and a dichotomy is applied to calibrate  $\theta$ .

To illustrate the above material, we suggest going through the application of kriging-gPC to the Runge function in the same conditions as in the previous paragraphs. The covariance function is here chosen as an exponential one

$$K(u, v, \theta) = \theta \exp(-|u - v|\theta), \quad (34)$$

and  $\theta$  is calibrated performing a simple dichotomy to minimize the predictive variance  $\sigma_K^2(\theta)$  as suggested in [64, 34, 63, 35]. Figure 7 presents the kriging-gPC approximations together with the ones obtained by integration-gPC and regression-gPC in the same conditions (the comparisons with collocation-gPC will be tackled later on). Let us first comment on the qualitative results of the first column of figure 7: for a low number of quadrature points ( $N = 11$ ), kriging-gPC behaves as the previous approximations and remain very oscillatory. As  $N$  increases to 21 (bottom-left picture), the oscillations are way more controlled and the kriging-gPC results are much less sensitive to the choice of the truncation order  $P$  than the other approximations. The quantitative convergence results of the right column of figure 7 show first that for  $P \leq N$ , the kriging-gPC approximations outperform the other ones. We recall kriging-gPC benefits an additional discretisation parameter  $K$  which here is very efficient in the sense it probably allows interesting compensations between term (\*) and term (\*\*) in expression (31). For higher polynomial order, i.e. in the top pictures of figure 7 with  $N = 11$  and  $P \geq N$ , analysis (33) becomes more and more relevant and we recover experimentally that the  $L^2$ -norm of the error  $\|u(X) - \mu(X, \hat{\theta})\|_{L^2} = \mathcal{O}(\Gamma_P^X)$  grows fast with  $P$ . The figure even allows recovering the fact that in such conditions ( $N$  fixed and  $P \gg N$ ) kriging-gPC and regression-gPC give equivalent results as expected by the error analysis of (33): in this regime the leading term is independent of the choice of  $K$ . If now  $P$  is kept lower than  $N = 21$  as in the bottom right picture of figure 7, error analysis (32) applies: the kriging-gPC approximations give very satisfactory results with a



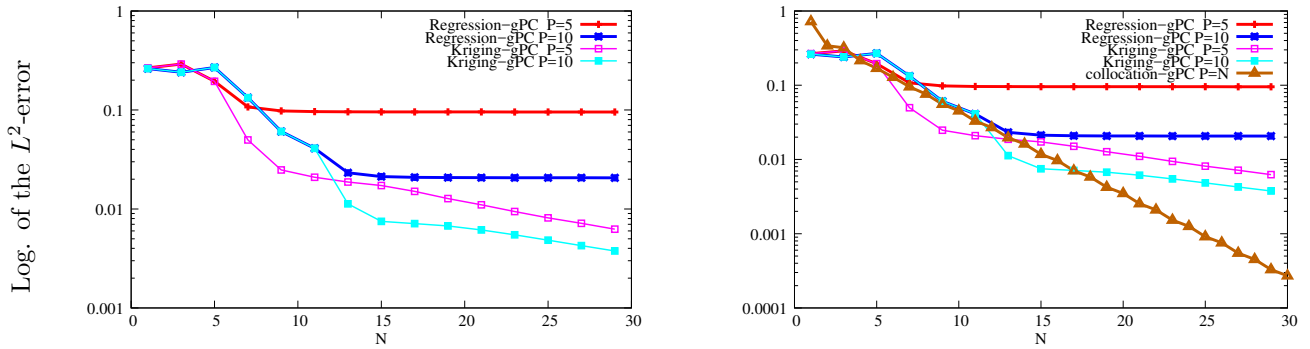


Figure 8: Convergence studies for regression-gPC, kriging-gPC (both for fixed  $P = 5$  and  $P = 10$ ) and collocation-gPC with respect to  $N$  for approximating Runge function (11).

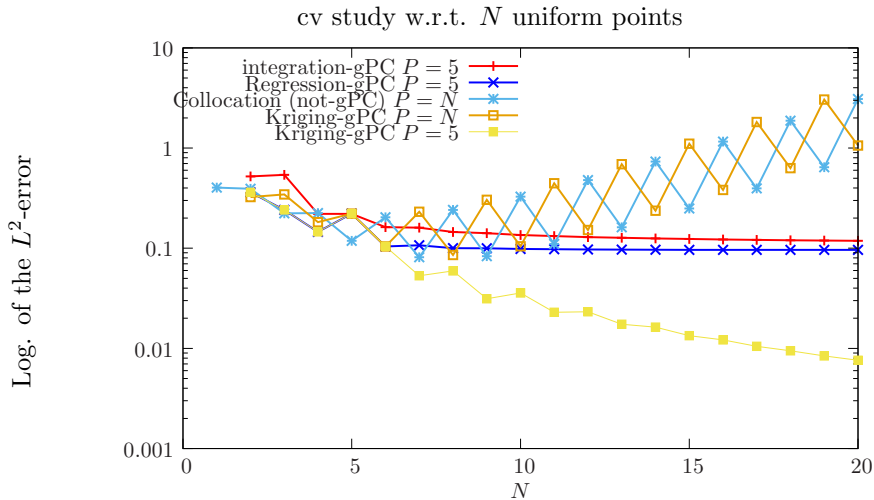


Figure 9: Convergence studies with respect to  $N$  for approximating Runge function (11) with a uniform experimental design. We compare the performances of integration-gPC, regression-gPC, kriging-gPC (all for fixed  $P = 5$ ) and collocation and kriging-gPC both with  $N = P$ .

*flatter* convergence curve testifying of a less sensitive behaviour with respect to the discretisation  $P$  than other methods.

To fully understand the influence of the covariance function, let us finally perform a convergence study with respect to  $N$  for fixed  $P = 5$  and  $P = 10$  (same conditions as in figure 6). The results are presented in figure 8: the left picture compares the convergence studies with respect to  $N$  obtained with regression-gPC as in section 2.2 and kriging-gPC for fixed  $P = 5$  and  $P = 10$ . Both methods present two regimes: in the first one, characterised by  $N \leq P$ , regression-gPC and kriging-gPC gives *exactly the same* results in  $L^2$ -norm. It is in agreement with the previous analysis (33) testifying of a relative independence of the choice of  $K$  of the approximation in such configuration. For  $N > P$ , regression-gPC approximations stagnate as the truncation error remains preponderant with respect to the integration one. On another hand, for kriging-gPC, the accuracy of the approximations continues to increase with  $N$  due to the covariance term in (32) ensuring a convergence driven by the spacing between the points of the experimental design in this regime. The right picture of figure 8 presents the same curves together with the collocation-gPC one: for  $N > P$  for which error analysis (32) applies, the kriging-gPC approximations do not systematically give better results than collocation-gPC. This is due to the fact the covariance function

- may be well-suited for some couples  $(N, P)$ , for example for  $P = 5$  and  $N \in \{5, \dots, 13\}$  or for  $P = 10$  and  $N \in \{10, \dots, 16\}$ ,
- and not so well in comparison to collocation for others: typically for large  $N > 17$  in the example of figure 8 (right).

Once again, this was predicted by the numerical analysis (32). Kriging-gPC, *via* the introduction of an additional discretisation tool (the covariance kernel  $K$ ), ensures a second convergence regime for  $N \geq P$  in comparison to integration-gPC or regression-gPC which can both lead to approximations of stagnating accuracy. The choice of the covariance kernel, in this regime, strongly affects the convergence rate of the approximation (slope of the  $L^2$ -norm of the error with respect to  $N$ ) and can lead to better approximations than integration-gPC/regression-gPC/collocation-gPC if  $K$  is well suited. Note that  $K$  needs, this is especially emphasized in (31)–(32), to depend on  $N$  to make sure kriging-gPC outperform them  $\forall N \in \mathbb{N}$ .

Finally, one may wonder what happens with kriging-gPC when applied on a uniform experimental design, in the same conditions as figure 5 on which the divergence of collocation (not gPC) has been highlighted. Expressions (24) (collocation-gPC) and (32) (kriging-gPC) are very similar: their only difference comes from the part in (32) depending on the covariance kernel  $k$ . The results obtained on Runge's function with an uniform experimental design (same conditions as in figure 5) are displayed in figure 9. It presents convergence studies with respect to  $N$  for integration-gPC, regression-gPC and kriging-gPC for  $P = 5$ . It furthermore presents the results obtained for collocation (not gPC as we do not use Gauss points) with  $P = N$  and



kriging-gPC with  $P = N$ . First, integration-gPC and regression-gPC behaves as in figure 6, the stagnation occurring once a good integration accuracy reached. Collocation (not gPC) diverges with such experimental design (same as in figure 5). Now, two different kriging-gPC convergence studies are displayed: for  $P = 5$  fixed and for  $P = N$ . For kriging-gPC with  $P = N$  (same conditions as collocation), the curve diverges: in such conditions, the introduction of the exponential covariance kernel (34) is not enough (see (31)) to compensate the fast growth of  $u^{(N)}(\xi_0)$  with  $N$  of Runge function. On another hand, for kriging-gPC and  $P = 5$  fixed (which is not feasible with collocation), the curve *first* coincides with regression-gPC but converges with increasing  $N$  after  $N = 5$ .

### 3 Few other applications and comparisons of gPC based methods

In this last section, we first compare integration-gPC, regression, regression-gPC, collocation-gPC and kriging-gPC on a discontinuous solution: the aim is to illustrate we recover the same results as predicted by the numerical analysis performed in the previous section, independently of the regularity of  $u$ . Furthermore, we also add some i-gPC approximations, see [57, 58], to the comparisons.

#### 3.1 Integration-gPC vs. regression-gPC vs. collocation-gPC vs. kriging-gPC vs. discontinuous solution

In this paragraph, we compare the previous gPC based approximations on a discontinuous solution. First because such low regularity solutions are of interest in many (of my) applications [38, 75, 26, 2, 57]. Second, to put forward the results of the numerical analysis of section 2 are independent of any regularity assumptions on  $u$ . We suggest applying integration-gPC, regression, regression-gPC, collocation-gPC and kriging-gPC to a discontinuous function  $X \rightarrow \mathbf{1}_{]-\infty, \frac{3}{10}]}(X)$  with  $X \sim \mathcal{U}([-1, 1])$ .

Figure 10 presents the results obtained with the four forementioned gPC based methods on the latter function. The presentation is slightly different than previously as the left column now displays the best

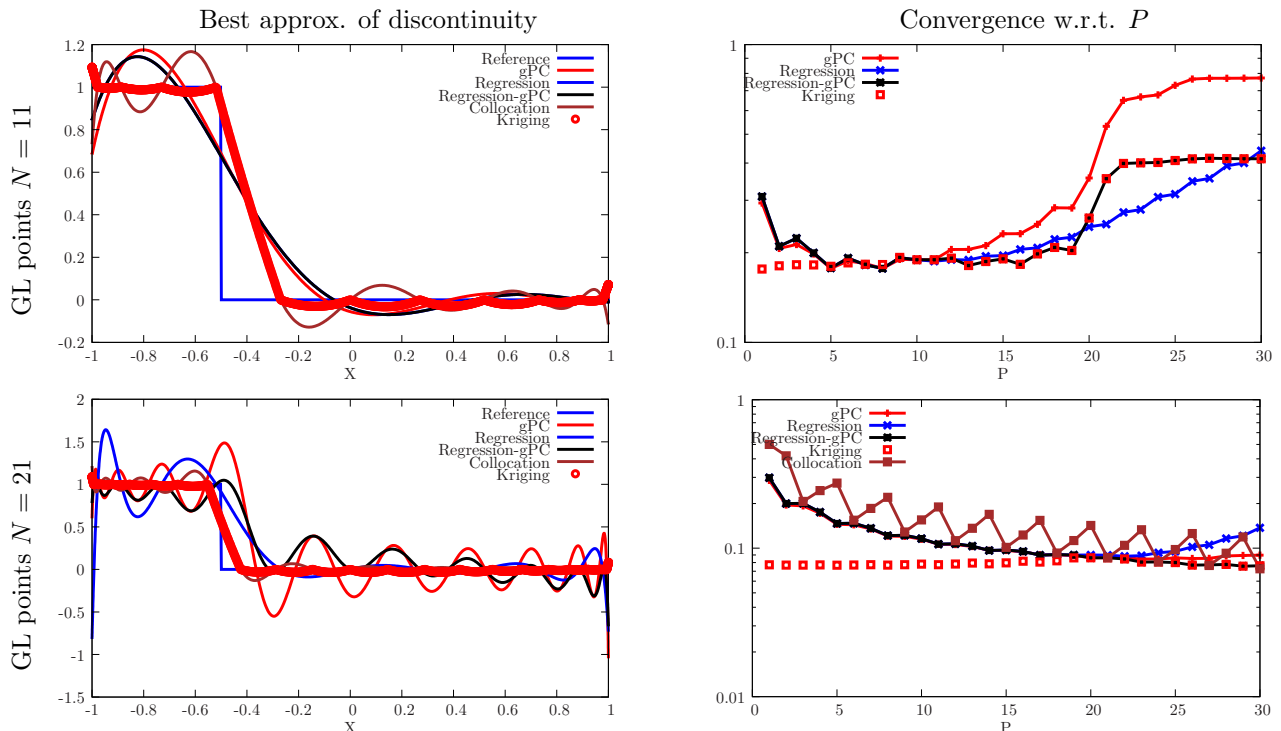


Figure 10: Application of Gauss-Legendre quadrature rule for *gPC*, *regression*, *regression-gPC*, *collocation-gPC* and *kriging-gPC* for the approximation of the transformation of a uniform random variable through a discontinuous function. The experimental designs have with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the best approximations obtained with every of the previous methods. The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ . The kriging kernel is chosen exponential (34) and a dichotomy is applied to calibrate  $\theta$ .

approximations obtained with every methods with  $N = 11$  (top) and  $N = 21$  (bottom). The right column shows convergence studies with respect to  $P$  in the same conditions with  $N = 11$  (top) and  $N = 21$  (bottom). Note that the convergence study obtained with collocation-gPC is not presented in the top right picture as it implies much more points (as  $P = N$ ) than  $N = 11$ .

We suggest beginning by commenting the convergence studies of the right column of figure 10. For  $N = 11$  (top right), integration-gPC, regression, regression-gPC have the same behaviour: the error slightly decreases before exploding. The one of kriging-gPC is more singular: the error is way lower than for the other approximations for small  $P$  but increases as fast as regression-gPC as soon as  $P > N$ . We once again recover numerically the fact that for  $P \gg N$ , the explosion rate is independent of the choice of the covariance kernel ( $K(u, v) = \delta_u(v)$  for regression-gPC,  $K(u, v)$  as in (34) for kriging-gPC), see (33). Having the same behaviour for the Runge function *and* for the discontinuous one of this section, we also recover it is independent of  $u$ . The top left picture of figure 10 presents the best approximations obtained with every methods with  $N = 11$ . For integration-gPC, regression, regression-gPC, it corresponds to  $P = 5$ . For collocation-gPC, it corresponds to  $N = P = 11$ . For kriging-gPC, it corresponds to  $P = 1$ . First, the

integration-gPC and regression approximations perfectly match and, as expected, poorly recover the function of interest. Regression-gPC presents a similar behaviour as integration-gPC and regression. Collocation-gPC and kriging-gPC coincide at the  $N = 11$  points of the experimental design. Kriging-gPC is way less oscillatory but interpolates almost linearly between the points of the experimental design on both sides of the discontinuity and not between the others.

For  $N = 21$ , the convergence studies of the resolution schemes is displayed on the bottom right picture of figure 10. First, the non-monotonous convergence of collocation-gPC is singular. Once again, integration-gPC and regression presents very similar results with an increase of the error as  $P$  becomes greater than  $N$ . The error for regression-gPC is more controlled. The kriging-gPC error, even if the lowest amongst every methods, keeps increasing from  $P = 1$  to  $P = N$  before equaling the one for regression-gPC for  $P > N$ . The fact the error is the lowest for  $P = 1$  testifies the covariance kernel ensures the accuracy for this function (at least up to  $P = 20$ ). This is interesting as the introduction of the kernel is usually justified by some smoothness hypothesis for  $u$ . Here, this assumption does not hold but the approximation remains relevant. Figure 10 bottom left presents the best results obtained with every methods: it corresponds to  $P = 20$  for integration-gPC, regression and regression-gPC, to  $P = N = 21$  for collocation-gPC and to  $P = 1$  for kriging-gPC. First, even if every approximations have a quantitatively better accuracy ( $L^2$ -norm) than in the previous case, qualitatively, the coarser approximations (for  $N = 11$ ) look better. The integration-gPC, regression and regression-gPC approximations are very oscillatory, especially in the vicinities of the boundaries of  $[-1, 1]$ . Surprisingly, collocation-gPC is less oscillatory than the three previous methods. The best approximation remains the one obtained with kriging-gPC, even if almost linearly interpolating between the points on each side of the discontinuity. The most accurate kriging-gPC approximation is once again obtained for  $P = 1$ : this implies the covariance kernel is responsible for this accuracy, more than the polynomial trend, even if relying on smoothness hypothesis of the solution.

In the next section, we suggest revisiting the two test-cases tackled in this paper adding a last gPC based approximation, i-gPC [57, 58]. The latter has been built in order to deal with discontinuous solutions.

### 3.2 Integration-gPC vs. regression-gPC vs. collocation-gPC vs. kriging-gPC vs. i-gPC

In [57, 58], we introduced a new gPC based method, iterative-gPC. The idea behind this new component of gPC is to

- exploit the fact that different gPC basis performs differently on the same output function  $u$ ,
- and design a gPC based approximation more adapted to discontinuous solutions.

The i-gPC method has been compared to integration-gPC in [57, 58] on several test-cases. The natural question arising now is how does i-gPC perform in comparison with the different approximations of section 2? We here apply integration-gPC, regression, regression-gPC, collocation-gPC, kriging-gPC and i-gPC to the same functions studied in sections 2 and 3.1 (discontinuity and Runge function).

Let us begin with function  $X \rightarrow \mathbf{1}_{]-\infty, \frac{3}{10}]}(X)$  with  $X \sim \mathcal{U}([-1, 1])$ , intensively studied in section 3.1. Figure 11 is in the same vein as figure 10: we only added the curves obtained with i-gPC. Let us begin by commenting on the convergence study with respect to  $P$  of figure 11 (top-right): the general behaviour of i-gPC is comparable to the ones of the other approximation methods. The error first decreases than explodes as soon as  $P > N$ . With such low number of points of the experimental design, it is complex controlling the integration error term in (4) along the iterations (see [58]) of i-gPC. For some polynomial orders, the i-gPC error is slightly more important than the integration-gPC one, which should not happen with an accurate numerical integration, see [57, 58]. This is the case for  $P = 4, 5, 6$  for example. For those polynomial orders, the stopping criterion should have stopped the algorithm one iteration earlier. Still, the i-gPC approximations remain controlled and of comparable qualities as the ones obtained with the other methods. Figure 11 (top-left) presents the best approximations obtained with every methods: the curves are the same as figure 10 described in section 3.1 except we added the i-gPC one. The best i-gPC approximation is obtained with  $P = 3$ . It is much less oscillatory than the other methods and the discontinuous behaviour of the solution is already captured.

The bottom pictures of figure 11 (bottom) present the same studies with  $N = 21$  GL points. With such an accurate numerical integration, i-gPC gives the best results as soon as  $P > 2$ . The convergence curve for i-gPC is always below the other ones. The best i-gPC approximation is displayed figure 11 (bottom-left) together with the best ones of the other methods: it is the less oscillatory of every approximations and it captures the discontinuous behaviour of the solution.

To complete the comparison of i-gPC with the other methods of the litterature, we perform the same study on Runge function. The results are displayed in figure 12. First, once again, for a small number of points of the experimental design, the behaviour of i-gPC is similar to the ones of the other methods. A decrease of the error before an explosion: the smoothness of Runge function does not improve this point. Furthermore, as before, the small number of points makes the control of the integration error in (4) along the iterations (see [58]) very difficult and i-gPC does not always perform better than gPC: if  $P$  is kept lower than  $N$  in another hand, see  $P = 2, 3$  in figure 12 (top-right), the results obtained with i-gPC are intermediary to the ones obtained with gPC, regression, regression-gPC (upper bound) and kriging-gPC (lower bound). Qualitatively, see figure 12 (top-left) the i-gPC approximation gives equivalent results as the other methods. Now, if  $N = 21$ , the integration is very accurate and i-gPC gives better results than gPC, regression, regression-gPC, up to  $P = N = 20$ . Kriging-gPC remains more efficient on such smooth output function even if qualitatively, every best approximations are indistinguishable (see figure 12 bottom-left).

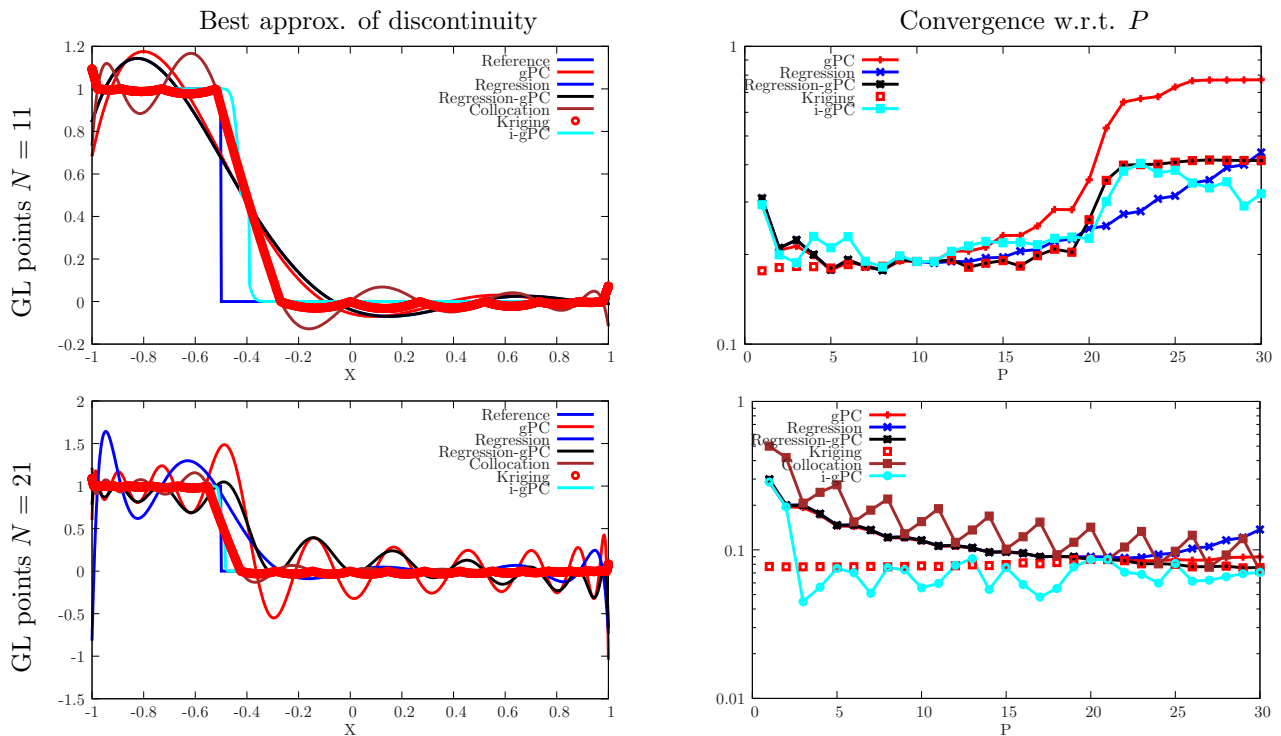


Figure 11: This figure is the same as figure 10 but we added the curves obtained with i-gPC. Application of Gauss-Legendre quadrature rule for *gPC*, *regression*, *regression-gPC*, *collocation-gPC*, *kriging-gPC* and *i-gPC* for the approximation of the transformation of a uniform random variable through a discontinuous function. The experimental designs have with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the best approximations obtained with every of the previous methods. The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ . The kriging kernel is chosen exponential (34) and a dichotomy is applied to calibrate  $\theta$ .

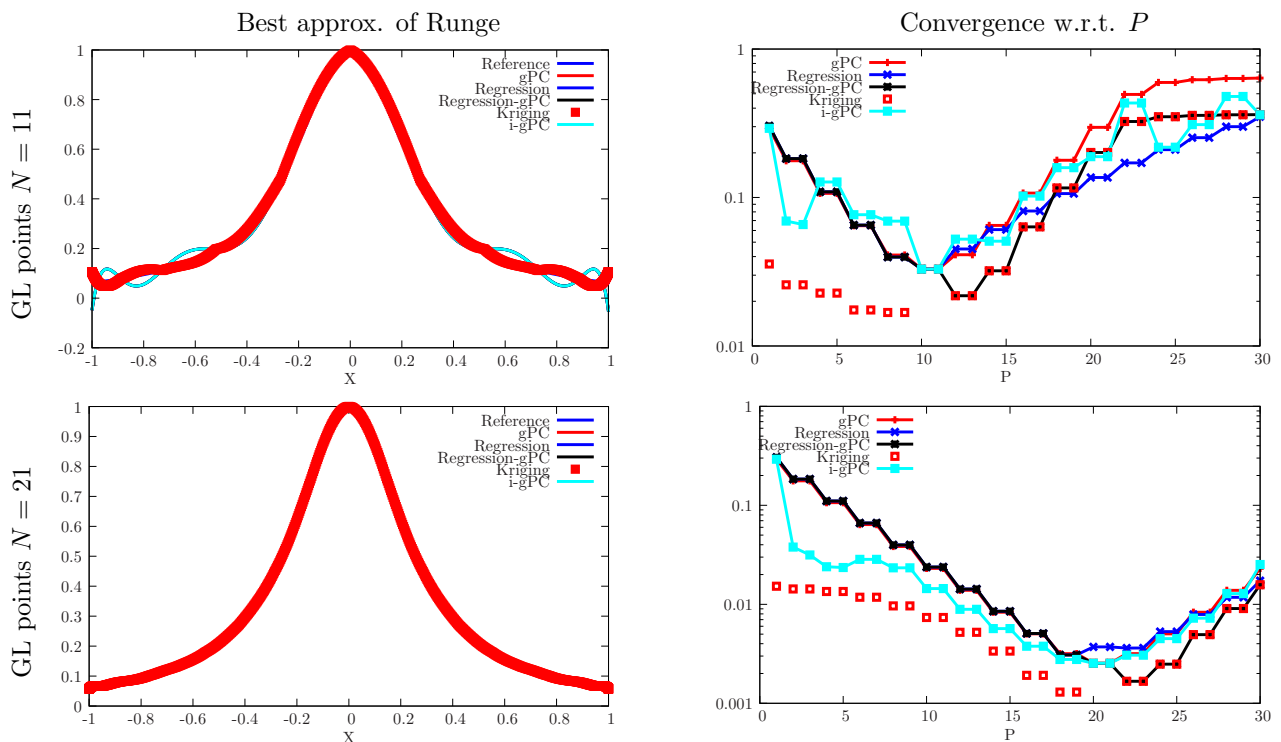


Figure 12: This figure is in the same vein as figure 11 except we consider Runge function instead of a discontinuous one. Application of Gauss-Legendre quadrature rule for *gPC*, *regression*, *regression-gPC*, *collocation-gPC*, *kriging-gPC* and *i-gPC* for the approximation of the transformation of a uniform random variable through Runge function. The experimental designs have with  $N = 11$  (top) and  $N = 21$  (bottom). The left column present the best approximations obtained with every of the previous methods. The right column present the  $L^2$ -norm of the error with respect to  $P$  for fixed  $N$ . The kriging kernel is chosen exponential (34) and a dichotomy is applied to calibrate  $\theta$ .

## 4 Conclusion

In this paper, we performed the numerical analysis of integration-gPC, regression-gPC, collocation-gPC and kriging-gPC in equivalent conditions. The paper is complementary with the furnished literature in which every gPC derivations are in general only pairwise experimentally compared<sup>27</sup>. The comparisons have mainly

<sup>27</sup>see for example [54, 61, 34, 63, 64, 35] for many interesting experimental comparison of kriging and gPC, on many different statistical observables and norms.

been made in the  $L^2$ -norm. With this numerical analysis, we aimed at identifying more easily under which conditions the strategies differ, are equivalent or are efficient. For example, we showed that taking  $P \leq N$  in gPC based approximations<sup>28</sup> with Gauss points is almost optimal, independent of the regularity of  $u$  and of the distributions of the input parameters. In such conditions  $P \leq N$ , integration-gPC is equivalent ( $L^2$ -norm) to regression-gPC and even to collocation-gPC if  $N = P$ . It ensures fast postprocessings for the gPC-reconstruction of random variable  $u(X)$  especially interesting when many outputs of interest must be approximated (see [59]). Integration-gPC may be less flexible than regression-gPC with which it is easy *a posteriori* taking into account additional points of the experimental design. Collocation-gPC ensures an interpolary approximation. Kriging-gPC presents the advantages of both regression-gPC and collocation-gPC and may avoid stagnating approximations as  $P \leq N$  with increasing  $N$  but is not ensured to converge for every experimental designs. i

## References

- [1] Brunner T. A. and P. S. Brantley. An efficient, robust, domain-decomposition algorithm for particle Monte Carlo. *Journal of Computational Physics*, 228.10:3882–3890, 2009.
- [2] R. Abgrall. A Simple, Flexible and Generic Deterministic Approach to Uncertainty Quantifications in Non Linear Problems: Application to Fluid Flow Problems. *Rapport de Recherche INRIA*, 2007.
- [3] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.
- [4] N. I. Akhiezer. *The Classical Moment Problem*. Oliver and Boyd, 1965.
- [5] R. Askey and J. Wilson. Some Basic Hypergeometric Polynomials that Generalize Jacobi Polynomials. *Memoirs Amer. Math. Soc., AMS, Providence RI*, 319, 1985.
- [6] A. Atkinson, A. Donev, and R. Tobias. *Optimum Experimental Designs, With SAS*. Oxford Statistical Science Series. OUP Oxford, 2007.
- [7] F. Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *ArXiv e-prints*, January 2013.
- [8] F. Bachoc. Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *ArXiv e-prints*, December 2014.
- [9] François Bachoc. *Estimation paramétrique de la fonction de covariance dans le modèle de Krigeage par processus Gaussiens : application à la quantification des incertitudes en simulation numérique*. PhD thesis, 2013. Thèse de doctorat dirigée par Garnier, Josselin Mathématiques appliquées Paris 7 2013.
- [10] M. Berveiller, B. Sudret, and M. Lemaire. Stochastic Finite Element: a Non Intrusive Approach by Regression. *Rev. Eur. Méc. Num.*, 15(1-2-3):81–92, 2006.
- [11] G. Blatman. *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. Thèse de doctorat, Université Blaise Pascal - Clermont II, 2009.
- [12] G. Blatman and B. Sudret. Sparse Polynomial Chaos Expansions and Adaptive Stochastic Finite Elements using a Regression Approach. *C. R. Méc.*, 336:518–523, 2008.
- [13] G. Blatman and B. Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Rel. Eng. Syst. Saf.*, 95:1216–1229, 2010.
- [14] G. Blatman, B. Sudret, and M. Berveiller. Quasi-Random Numbers in Stochastic Finite Element Analysis. *Mech. Ind. proofs*, 8:289–297, 2007.
- [15] Emeric Brun, Stéphane Chauveau, and Fausto Malvagi. Patmos: A prototype monte carlo transport code to test high performance architectures. In Proceedings of International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering, Jeju, Korea, 2017.
- [16] R.H. Cameron and W.T. Martin. The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals. *Annals of Math.*, 48:385–392, 1947.
- [17] E.W. Cheney and W.A. Light. *A Course in Approximation Theory*. Graduate studies in mathematics. American Mathematical Soc.
- [18] T. Crestaux. Polynômes de Chaos pour la Propagation et la Quantification d’Incertitudes. Technical report, CEA, 2006.
- [19] G.E. Karniadakis D. Lucor, C.-H Su. Generalized Polynomial Chaos and Random Oscillators. *Int.J.Numer.Meth.Engng.*, 60:571–596, 2004.
- [20] Bert J. Debusshere, Habib N. Najm, Philippe P. Pébay, Omar M. Knio, Roger G. Ghanem, and Olivier P. Le Maître. Numerical Challenges in the Use of Polynomial Chaos Representations for Stochastic Processes. *J. Sci. Comp.*, 26:698–719, 2004.

---

<sup>28</sup>which is a classical rule of thumb in the community.

- [21] H. Dette and W.J. Studden. *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*. A Wiley-Interscience publication. Wiley, 1997.
- [22] A Dutfoy and R Lebrun. Practical approach to dependence modelling using copulas. *Journal of Risk and Reliability*, 223(4):347–361, 2009.
- [23] James F. Epperson. On the runge example. *The American Mathematical Monthly*, 94(4):329–341, 1987.
- [24] Ernst, Oliver G., Mugler, Antje, Starkloff, Hans-Jörg, and Ullmann, Elisabeth. On the convergence of generalized polynomial chaos expansions. *ESAIM: M2AN*, 46(2):317–339, 2012.
- [25] V.V. Fedorov. *Theory Of Optimal Experiments*. Probability and Mathematical Statistics. Elsevier Science, 1972.
- [26] J. Foo and G.E. Karniadakis. Multi-element probabilistic collocation method in high dimensions. *J. Comput. Phys.*, 229:1536–1557, 2010.
- [27] B. Ganapathysubramanian and N. Zabaras. Sparse Grid Collocation Schemes for Stochastic Natural Convection Problems. *J. Comp. Phys.*, 225:652–685, 2007.
- [28] Walter Gautschi. *Orthogonal polynomials: applications and computation*, volume 5. Oxford University Press, 1996.
- [29] P.N. Gavriiadis and G.A. Athanassoulis. Moment data can be analytically completed. *Probabilistic Engineering Mechanics*, 18:329–338, 2003. doi:10.1016/j.cam.2012.05.015.
- [30] P.N. Gavriiadis and G.A. Athanassoulis. The truncated hausdorff moment problem solved by using kernel density functions. *Journal of Computational and Applied Mathematics*, 2012. doi:10.1016/j.cam.2012.05.015.
- [31] Gene H. Golub and Gerard Meurant. *Matrices, Moments and Quadrature with Applications*. Princeton University Press, Princeton, NJ, USA, 2009.
- [32] S. Hosder, R. W. Walters, and R. Perez. A Non Intrusive Polynomial Chaos Method for Uncertainty Propagation in CFD Simulations. *44th AIAA Aerospace Sciences Meeting and Exhibit*, AIAA 2006-891, 2006.
- [33] Samuel Karlin and Lloyd S. Shapley. *Geometry of moment spaces*. American Mathematical Society, 1953.
- [34] Pierrick Kersaudy, Bruno Sudret, Nadège VARSIER, Odile Picon, and Joe Wiart. A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry. *Journal of Computational Physics*, 286:130–117, January 2015.
- [35] Pierrick Kersaudy, Bruno Sudret, Nadège Varsier, Odile Picon, and Joe Wiart. A new surrogate modeling technique combining kriging and polynomial chaos expansions - application to uncertainty analysis in computational dosimetry. *J. Comput. Phys.*, 286(C):103–117, April 2015.
- [36] B. Lapeyre, E. Pardoux, and R. Sentis. *Méthodes de Monte Carlo pour les équations de transport et de diffusion*. Number 29 in Mathématiques & Applications. Springer-Verlag, 1998.
- [37] O. Le Maitre, M. Reagan, H. Najm, R. Ghanem, and O. Knio. A Stochastic Projection Method for Fluid Flow. II. Random Process. *J. Comp. Phys.*, 181:9–44, 2002.
- [38] O. Le Maitre, M. Reagan, H. Najm, R. Ghanem, and O. Knio. Multi-Resolution Analysis of Wiener-Type Uncertainty Propagation Schemes. *J. Comp. Phys.*, 197:502–531, 2004.
- [39] R. Lebrun and A. Dutfoy. A Generalization of the Nataf Transformation to Distributions with Elliptical Copula. *Prob. Eng. Mech.*, 24,2:172–178, 2009.
- [40] R. Lebrun and A. Dutfoy. An Innovating Analysis of the Nataf Transformation from the Copula viewpoint. *Prob. Eng. Mech.*, 24,3:312–320, 2009.
- [41] A. Loeven, J. Witteveen, and H. Bijl. Efficient Uncertainty Quantification using a Two-Step Approach with Chaos Collocation. *ECCOMAS CFD*, 2006.
- [42] G.J.A. Loeven and H. Bijl. Airfoil Analysis with Uncertain Geometry Using the Probabilistic Collocation Method. *49th AIAA Aerospace Sciences Meeting and Exhibit*, AIAA 2008-2070, 2008.
- [43] G.J.A Loeven, J. A. S. Witteveen, and H. Bijl. Probabilistic Collocation: an Efficient Non Intrusive Approach for Arbitrarily Distributed Parametric Uncertainties. *45th AIAA Aerospace Sciences Meeting and Exhibit*, AIAA 2007-317, 2007.
- [44] D. Lucor, J. Meyers, and P. Sagaut. Sensitivity Analysis of LES to Subgrid-Scale-Model Parametric Uncertainty using Polynomial Chaos. *J. Fluid Mech.*, 585:255–279, 2007.
- [45] Amitava Majumdar. Parallel performance study of monte carlo photon transport code on shared-, distributed-, and distributed-shared-memory architectures. 2000.
- [46] et al Martin, William R. Monte carlo photon transport on shared memory and distributed memory parallel processors. *International Journal of High Performance Computing Applications*, 1.3:57–74, 1987.

- [47] J.-M. Martinez, J. Cahen, A. Millard, D. Lucor, F. Huvelin, J. Ko, and N. Poussineau. Modélisation des Incertitudes par Polynômes de Chaos – Étude d’un Écoulement en Milieux Poreux. Technical Report Rapport DM2S/DIR/RT/06-006/A, CEA-CEMRACS, 2006.
- [48] 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):1862–1902, 2009.
- [49] Youssef M. Marzouk and Habib N. Najm. Dimensionality Reduction and Polynomial Chaos Acceleration of Bayesian Inference in Inverse Problems. *J. Comp. Phys.*, 228(6):1862–1902, 2009.
- [50] J. Le Meitour, D. Lucor, and J-C Chassaing. Prediction of Stochastic Limit Cycle Oscillations using an Adaptive Polynomial Chaos Method. *Journal of Aeroelasticity and Structural Dynamics*, 2(1):3–22, 2010.
- [51] Frederik Riis Mikkelsen. Probabilistic aspects of moment sequences: The method of moments.
- [52] H.N. Najm. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annu. Rev. Fluid Mech.*, 41:35–52, 2009.
- [53] F. Nobile, R. Tempone, and C. Webster. A Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.
- [54] N. E. Owen, P. Challenor, P. P. Menon, and S. Bennani. Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):403–435, 2017.
- [55] M. Paffrath and U. Wever. Adapted Polynomial Chaos Expansion for Failure Detection. *J. Comp. Phys.*, 226(1):263–281, 2007.
- [56] R. Peyret. *Spectral Methods for Incompressible Viscous Flow*. Applied Mathematical Sciences. Springer New York, 2002.
- [57] G. Poëtte and D. Lucor. Non Intrusive Iterative Stochastic Spectral Representation with Application to Compressible Gas Dynamics. *J. of Comput. Phys.*, 2011. DOI information: 10.1016/j.jcp.2011.12.038.
- [58] Gaël Poëtte, Alexandre Birolleau, and Didier Lucor. Iterative polynomial approximation adapting to arbitrary probability distribution. *SIAM J. Numerical Analysis*, 53(3):1559–1584, 2015.
- [59] Gaël Poëtte, Didier Lucor, and Hervé Jourden. A stochastic surrogate model approach applied to calibration of unstable fluid flow experiments. *Comptes Rendus Mathématique*, 350(5):319 – 324, 2012.
- [60] R. Procassini, M. O’Brien, and J. Taylor. Load balancing of parallel Monte Carlo transport calculations. *Mathematics and Computation, Supercomputing, Reactor Physics and Nuclear and Biological Applications*, 2005.
- [61] P. T. Roy, N. El Moçayd, S. Ricci, J.-C. Jouhaud, N. Goutal, M. De Lozzo, and M. Rochoux. Comparison of polynomial chaos and gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. *Stochastic Environmental Research and Risk Assessment*, 2017. submitted.
- [62] C. P. Rupert and C. T. Miller. An Analysis of Polynomial Chaos Approximations for Modeling Single-Fluid-Phase Flow in Porous Medium Systems. *J. Comp. Phys.*, 226:2175–2205, 2007.
- [63] R. Schöbi, B. Sudret, and S. Marelli. Rare Event Estimation Using Polynomial-Chaos Kriging. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(2):D4016002, 2017.
- [64] Roland Shöbi, Bruno Sudret, and Joe Wiart. Polynomial-chaos-based kriging. 2015.
- [65] F. Simon, P. Guillen, P. Sagaut, and D. Lucor. A gPC based approach to uncertain transonic aerodynamics. *CMAME*, 199:1091–1099, 2010.
- [66] B. Sudret. Global Sensitivity Analysis using Polynomial Chaos Expansion. *Rel. Engrg. Syst. Saf.*, 93:964–979, 2007.
- [67] B. Sudret. *Uncertainty Propagation and Sensitivity Analysis in Mechanical Models, Contribution to Structural Reliability and Stochastic Spectral Methods*. Habilitation à Diriger des Recherches, Université Blaise Pascal - Clermont II, 2007.
- [68] G. Szego. *Orthogonal Polynomials*, volume 23. American Mathematical Society, colloquim publications, 1939.
- [69] E. Novak V. Barthelmann and K. Ritter. High Dimensional Polynomial Interpolation on Sparse Grids. *Math. Subj. Class.*, 1991.
- [70] Emmanuel Vazquez. *Modélisation comportementale de systèmes non-linéaires multivariés par méthodes à noyaux et applications*. PhD thesis, Paris 11, 2005. Thèse de doctorat dirigée par Walter, Éric Sciences appliquées.
- [71] X. Wan and G. E. Karniadakis. Stochastic Heat Transfer Enhancement in a Grooved Channel. *J. Fluid Mech.*, 565:255–278, 2006.

- [72] X. Wan and G.E. Karniadakis. Beyond Wiener-Askey Expansions: Handling Arbitrary PDFs. *SIAM J. Sci. Comp.*, 27(1-3), 2006.
- [73] X. Wan and G.E. Karniadakis. Long-Term Behavior of Polynomial Chaos in Stochastic Flow Simulations. *Comp. Meth. Appl. Mech. Engrg.*, 195:5582–5596, 2006.
- [74] X. Wan and G.E. Karniadakis. Long-Term Behaviour of Polynomial Chaos in Stochastic Flow Simulations. *Comput. Meth. Appl. Mech. Engrg.*, 216(5582-5596), 2006.
- [75] X. Wan and G.E. Karniadakis. Multi-Element generalized Polynomial Chaos for Arbitrary Probability Measures. *SIAM J. Sci. Comp.*, 28(3):901–928, 2006.
- [76] Yunsong Wang. *Optimization of Monte Carlo Neutron Transport Simulations with Emerging Architectures*. PhD thesis, Université Paris-Saclay, 2017. Thse de doctorat dirigé par Calvin, Christophe Informatique Paris Saclay 2017.
- [77] N. Wiener. The Homogeneous Chaos. *Amer. J. Math.*, 60:897–936, 1938.
- [78] J. Witteveen and H. Bijl. Using Polynomial Chaos for Uncertainty Quantification in Problems with Nonlinearities. *47th AIAA*, 2006.
- [79] J. A. S. Witteveen and H. Bijl. Using Polynomial Chaos for Uncertainty Quantification in Problems with Non Linearities. *47th AIAA Aerospace Sciences Meeting and Exhibit*, AIAA 2006-2066, 2006.
- [80] J. A. S. Witteveen and H. Bijl. An Unsteady Adaptive Stochastic Finite Elements Formulation for Rigid-Body Fluid-Structure Interaction. *Comp. and Struct.*, 2008.
- [81] D. Xiu and J.S. Hesthaven. High-Order Collocation Methods for Differential Equations with Random Inputs. *J. Sci. Comput.*, 27(3):1118–1139, 2005.
- [82] D. Xiu and G. E. Karniadakis. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comp.*, 24(2):619–644, 2002.
- [83] D. Xiu and G.E. Karniadakis. Modeling Uncertainty in Steady State Diffusion Problems via generalized Polynomial Chaos. *Comp. Meth. Appl. Mech. Engrg.*, 191:4927–4948, 2002.
- [84] D. Xiu and G.E. Karniadakis. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comp.*, 24:619–644, 2002.
- [85] D. Xiu and G.E. Karniadakis. Modeling Uncertainty in Flow Simulations via generalized Polynomial Chaos. *Comp. Meth. Appl. Mech. Engrg.*, 187:137–167, 2003.
- [86] D. Xiu, D. Lucor, C.-H. Su, and G. E. Karniadakis. Stochastic Modeling of Flow-Structure Interactions using Generalized Polynomial Chaos. *J. Fluid Eng.*, 124:51–59, 2002.

## A The construction of the gPC basis

A gPC basis is nothing more than an orthonormal polynomial basis. The gPC procedure only helps the uncertainty analyst *a priori* choosing an efficient one, see section 2 and [72]. Orthonormal polynomials have been intensively studied in the literature, see amongst others [68, 3, 28, 31]. In this document, we do not aim at being exhaustive on the subject. We only recall their main properties, those useful for the construction of an arbitrary gPC basis associated to the inner product defined by an arbitrary probability measure  $d\mathcal{P}_X$ . We also prepare some notions and notations for the question of numerical integration with Gauss quadrature rules which is central for collocation-gPC (section 2.3).

### A.1 Inner product defined by an arbitrary probability measure

Let  $d\mathcal{P}_X$  be an arbitrary probability measure related to an arbitrary random variable  $X$ . Care will be taken in this section to be able to consider *any* random variable, continuous (gaussian, uniform,...) and even discrete/categorical (binomial, multinomial,...). We introduce the inner product defined by the probability measure  $d\mathcal{P}_X$  as

$$\langle f, g \rangle_X = \int f(x)g(x) d\mathcal{P}_X(x). \quad (35)$$

The above inner product, for a probability measure of a discrete random variable having  $D + 1 < \infty$  states  $(x_j)_{j \in \{0, \dots, D\}}$  with probabilities  $(p_j)_{j \in \{0, \dots, D\}} > 0$ , resumes to

$$\langle f, g \rangle_X = \int f(x)g(x) d\mathcal{P}_X(x) = \int f(x)g(x) \sum_{j=0}^D p_j \delta_{x_j}(x) = \sum_{j=0}^D p_j f(x_j)g(x_j). \quad (36)$$

In this particular case, the output random variable is in a finite vector space (of size  $D + 1 < \infty$ ). With expression (36), we insist on the fact that the notation (35) is compatible with continuous and discrete input random variables. The notations are inspired from the ones of [28].

A sequence of  $P$  orthogonal polynomials  $(\phi_k^X)_{k \in \{0, \dots, P\}}$  associated to the probability measure  $d\mathcal{P}_X$  has the following properties:

- $\phi_k^X$  is of degree  $k$ ,
- $\langle \phi_k^X, \phi_l^X \rangle_X = 0, \forall (k, l) \in \{0, \dots, P\}^2$  such that  $k \neq l$ .

The sequence is said

- *orthonormal* if  $\langle \phi_k^X, \phi_k^X \rangle_X = 1 \forall k \in \{0, \dots, P\}$ ,
- *monic* if the coefficient of the highest degree (i.e. of  $x^k$  for  $\phi_k^X$ ) for every polynomials of the sequence  $(\phi_k^X)_{k \in \{0, \dots, P\}}$  is 1. In this document, we use the additional upperscript  $m$ , i.e.  $(\phi_k^{X,m})_{k \in \mathbb{N}}$ , to denote monic (orthogonal) polynomials.

Obviously, the families  $(\phi_k^{X,m})_{k \in \mathbb{N}}$  of monic orthogonal polynomials and  $(\phi_k^X)_{k \in \mathbb{N}}$  of orthonormal polynomials are related. Their relative expression will be given in the next section once an additional notion introduced. For conciseness in the following, the sequence of polynomials orthonormal with respect to the inner product defined by a given probability measure  $d\mathcal{P}_X$  is referred as *the polynomials associated to  $X$  or  $d\mathcal{P}_X$* . Such sequence is closely related to the *statistical moments* of its corresponding random variable  $X$  or probability measure  $d\mathcal{P}_X$ . In the next section, we recall this important link.

## A.2 Moments of a probability measure and Hankel determinants

Consider an arbitrary random variable  $X$  having probability measure  $d\mathcal{P}_X$  and suppose  $\forall k \in \mathbb{N}$

$$s_k^X = \int x^k d\mathcal{P}_X(x) < \infty. \quad (37)$$

Then the sequence of numbers  $(s_k^X)_{k \in \mathbb{N}}$  is called the sequence of moments of the random variable  $X$  or of the probability measure  $d\mathcal{P}_X$ . Note that if the support of  $X/d\mathcal{P}_X$  is bounded, the existence of the sequence (37) of moments such that (37)  $\forall k \in \mathbb{N}$  is straightforward. We define the Hankel matrices/determinants of the random variable  $X$  or of the probability measure  $d\mathcal{P}_X$  by  $\forall k \in \mathbb{N}$

$$\underline{H}_{2k}^X = \begin{vmatrix} s_0^X & s_1^X & \dots & s_k^X \\ \dots & \dots & \dots & \dots \\ s_n^X & s_{n+1}^X & \dots & s_{n+k}^X \\ \dots & \dots & \dots & \dots \\ s_k^X & \dots & \dots & s_{2k}^X \end{vmatrix}, \quad \underline{H}_{2k+1}^X = \begin{vmatrix} s_1^X & s_2^X & \dots & s_{k+1}^X \\ \dots & \dots & \dots & \dots \\ s_n^X & s_{n+1}^X & \dots & s_{n+k}^X \\ \dots & \dots & \dots & \dots \\ s_{k+1}^X & \dots & \dots & s_{2k+1}^X \end{vmatrix}. \quad (38)$$

Determining whether a sequence of finite real numbers  $(s_k)_{k \in \mathbb{N}}$  are moments of a unique or not random variable/probability measure is what is commonly called the *classical moment problem*. If such random variable/probability measure *exists* and is *unique*, the problem is said *determinate*. If it *exists* and is *not unique*, the problem is said *indeterminate*. Depending on the support of the random variable/probability measure, the moment problem is called [51, 4]:

- The *Hausdorff moment problem* when  $X/d\mathcal{P}_X$  has a bounded support, i.e.  $X \in \prod_{i=1}^Q [a_i, b_i]$ . In this particular case, if a sequence  $(s_k^X)_{k \in \mathbb{N}}$  is a sequence of moments of a random variable/probability measure, then the problem is determinate [51, 33, 4].
- The *Stieltjes moment problem* when  $X/d\mathcal{P}_X$  has a half-line support, i.e.  $X \in \prod_{i=1}^Q [a_i, \infty[$ . In such conditions, if a sequence  $(s_k^X)_{k \in \mathbb{N}}$  is a sequence of moments of a random variable/probability measure, the problem may be indeterminate. A sufficient condition for uniqueness can be express as

$$\sum_{k=0}^{\infty} (s_k^X)^{-\frac{1}{2k}} = \infty. \quad (39)$$

It is called *Carleman's condition* see [4].

- The *Hamburger moment problem* when  $X/d\mathcal{P}_X$  has an unbounded support, i.e.  $X \in \mathbb{R}^Q$ . In such conditions, a sequence of moments may also be indeterminate and the *Carleman's condition* see [4] in this case is given by

$$\sum_{k=0}^{\infty} (s_{2k}^X)^{-\frac{1}{2k}} = \infty. \quad (40)$$

Consequently, depending on the support of the random variable/probability measure of interest, if existence holds, uniqueness is not always straightforward. In the indeterminate cases, the solutions of the moment problem form a convex set. Most of all for our applications, in the *determinate* moment problem case, the set of polynomials are *dense* in the associated Hilbert space and provides converging properties to the gPC basis [24].

So far, we mainly dealt with uniqueness and assumed existence. Let  $(s_k^X)_{k \in \mathbb{N}}$  be a sequence of numbers satisfying  $\forall k \in \mathbb{N} s_k^X < \infty$ . Suppose the Hankel determinants defined by (38) satisfies:

- either  $\forall k \in \mathbb{N} \underline{H}_{2k}^X > 0$  and  $\underline{H}_{2k+1}^X > 0$ ,
- or  $\forall (2k, 2k+1) \in \{0, \dots, D\}^2 \underline{H}_{2k}^X > 0$  and  $\underline{H}_{2k+1}^X > 0$  and  $\underline{H}_{2k}^X = \underline{H}_{2k+1}^X = 0$  for larger  $k$ ,



then the sequence is a sequence of moments of a random variable/probability measure for the classical moment problem (i.e. independently of being a Hausdorff, Stieltjes or Hamburger moment problem). Note that in the second case, the Hilbert space associated to the existing measure is finite-dimensional and of size  $D + 1$ .

As briefly tackled before, the existence of a set of dense polynomials associated to a given random variable/probability measure  $X/d\mathcal{P}_X$  is closely related to the existence and the determinacy of its moments. In the following section, we present Christoffel's formulae which explicit the relation between moments of a random variable and orthonormal polynomials associated this same one.

### A.3 Christoffel's formulae, Jacobi's matrix and construction procedures

Christoffel's formulae [4, 33] explicit the relation between the sequence of moments  $(s_k^X)_{k \in \mathbb{N}}$  of a random variable  $X$  and the set of orthonormal polynomials  $(\phi_k^X)_{k \in \mathbb{N}}$  associated to  $X$ . It is given by

$$\forall n \in \{0, \dots, P\}, \phi_n^X(x) = \frac{1}{\sqrt{\underline{H}_{2(n-1)}^X \underline{H}_{2n}^X}} \begin{vmatrix} s_0^X & s_1^X & \dots & s_n^X \\ \dots & \dots & \dots & \dots \\ s_k^X & s_{k+1}^X & \dots & s_{n+k}^X \\ \dots & \dots & \dots & \dots \\ 1 & x^1 & \dots & x^n \end{vmatrix}. \quad (41)$$

In (41) appears the previous Hankel determinants (38) and the fact that the polynomial sequence may only exist up to a certain order  $D$  whether the Hankel determinants are all strictly positive or if there exists an order after which they are all zero. In the following, we keep considering, for convenience, polynomials orders  $k \in \mathbb{N}$  even if for some probability measure they exist only up to a certain order  $P \in \mathbb{N}$ . With (41), it is easy verifying, see [4], the monic orthogonal polynomials  $(\phi_k^{X,m})_{k \in \mathbb{N}}$  can be expressed with respect to both the orthonormal ones  $(\phi_k^X)_{k \in \mathbb{N}}$  and the Hankel determinants as we have  $\forall k \in \mathbb{N}$ :

$$\phi_k^X(x) = \Gamma_k^X \phi_k^{X,m}(x) = \sqrt{\frac{\underline{H}_{2(k-1)}^X}{\underline{H}_{2k}^X}} \phi_k^{X,m}(x). \quad (42)$$

Christoffel's formulae (41) obviously represents a way to build the gPC basis associated to the probability measure  $d\mathcal{P}_X$ . It definitely has a theoretical interest but it is scarcely used in practice due to the difficulty to accurately numerically compute the Hankel determinants. The problem is more and more ill-conditioned as the polynomial order increases. To illustrate this, suppose the sequence of moments  $(s_k^X)_{k \in \{0, \dots, 2P\}}$  of an *existing* random variable  $X$  are not accurately known<sup>29</sup> and assume a perturbation of these moments such that

$$(s_k^\varepsilon)_{k \in \{0, \dots, 2P\}} = (s_k^X + \varepsilon_k)_{k \in \{0, \dots, 2P\}} \approx (s_k^X)_{k \in \{0, \dots, 2P\}}.$$

For the sake of simplicity of the following developments, we suppose a particular form for the perturbation  $\varepsilon = (\varepsilon_0, \dots, \varepsilon_{2P}) = (0, \dots, 0, \delta)$  so that we suppose every moments of order  $n \in \{0, \dots, 2P - 1\}$  are accurately computed whereas the last one  $s_{2P}^X$  is perturbed by  $\delta$ . The polynomials orthonormal with respect to the perturbed moments coincide with the one of  $X$  up to order  $P - 1$  and we have

$$\phi_P^\varepsilon(x) = \frac{1}{\sqrt{\underline{H}_{2(P-1)}^X \underline{H}_{2P}^\varepsilon}} \begin{vmatrix} s_0^X & s_1^X & \dots & s_n^X \\ \dots & \dots & \dots & \dots \\ s_k^X & s_{k+1}^X & \dots & s_{n+k}^X \\ \dots & \dots & \dots & \dots \\ 1 & x^1 & \dots & x^P \end{vmatrix}, \quad (43)$$

so that  $\varepsilon$  only affects the last Hankel determinant  $\underline{H}_{2P}^\varepsilon$ . Now, from the definition of  $\underline{H}_{2P}^\varepsilon$  and by a development of the last line of the determinant, we have

$$\underline{H}_{2P}^\varepsilon = \underline{H}_{2P}^X + \delta \underline{H}_{2(P-1)}^X \quad \text{so that} \quad \phi_P^\varepsilon(x) = \frac{1}{\sqrt{1 + \delta \frac{\underline{H}_{2(P-1)}^X}{\underline{H}_{2P}^X}}} \phi_P^X(x). \quad (44)$$

Consequently, the sequence of polynomials  $(\phi_k^X)_{k \in \mathbb{N}}$  associated to  $X$  and the sequence of polynomials  $(\phi_k^\varepsilon)_{k \in \mathbb{N}}$  associated to the perturbed moments are such that  $\forall n \in \{0, \dots, P - 1\}, \phi_n^\varepsilon = \phi_n^X$  and for the last component, we have

$$\phi_P^\varepsilon(x) = \phi_P^X(x) - \frac{1}{2} \frac{\underline{H}_{2(P-1)}^X}{\underline{H}_{2P}^X} \phi_P^X(x) \delta + \mathcal{O}(\delta^2). \quad (45)$$

It is known in the literature [30, 29, 4, 33] that in the previous conditions  $\forall n \in \{0, \dots, P\}$

$$\frac{\underline{H}_{2n}^X}{\underline{H}_{2(n-1)}^X} \leq 2^{-(4n+2)} \quad \text{leading to} \quad \left| \frac{d\phi_P^X}{d\delta} \right| \geq 2^{4P+1}, \quad (46)$$

which testifies for a higher and higher sensivity of the orthonormal basis components with respect to a small inaccuracy in the statistical moments/Hankel determinants ( $\delta$ ) as the polynomial order  $P$  increases. The inaccuracy  $\delta$  may come from an approximation of the moments but also from roundoff errors due to the determinant computation algorithm. In order to avoid such aliasing errors, more stable algorithm are

<sup>29</sup>they may be only computed, estimated etc.

available such as the Chebyshev one or the modified Chebyshev one (see [28]). Those algorithms intensively use another important property of orthonormal polynomials. For any set of orthonormal polynomials up to order  $P$ , there exists two sequences of coefficients  $(\alpha_k)_{k \in \{0, \dots, P\}}$  and  $(\beta_k)_{k \in \{0, \dots, P\}}$  such that  $\forall k \in \{0, \dots, P\}$

$$\sqrt{\beta_{k+1}}\phi_{k+1}^X(x) = (x - \alpha_{k+1})\phi_k^X(x) - \sqrt{\beta_k}\phi_{k-1}^X(x), \quad (47)$$

where  $\forall k \in \{0, \dots, P\}$ , we have

$$\alpha_k = \frac{\int x \phi_k^X(x) \phi_k^X(x) d\mathcal{P}_X(x)}{\langle \phi_k, \phi_k \rangle_X},$$

and  $(\beta_k)_{k \in \{0, \dots, P\}}$  are such that  $\forall k \in \{0, \dots, P\}$

$$\beta_k = \frac{\langle \phi_k^X, \phi_k^X \rangle_X}{\langle \phi_{k-1}^X, \phi_{k-1}^X \rangle_X}.$$

Equation (47) is referred to as the *three term recurrence formulae* in the literature. The Chebyshev algorithms, even if known to be more stable, may also suffer inaccuracy<sup>30</sup> in the estimations of  $(\alpha_k, \beta_k)_{k \in \mathbb{N}}$ . The relation between the moments and the coefficients of the three-term recurrence formulae can be explicitated, we refer to [68, 28] for the interested reader. It must be kept in mind that the orthonormal basis associated to any random variable with a too important order may bear unworkable inaccuracies.

The three-term recurrence formulae can be written in a matrix form by introducing the *Jacobi matrix* of order  $P$  defined by

$$J_P^X = \begin{pmatrix} \alpha_1 & \sqrt{\beta_1} & 0 & 0 & \dots & 0 \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & 0 & \dots & 0 \\ 0 & \sqrt{\beta_2} & \alpha_3 & \sqrt{\beta_3} & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \sqrt{\beta_P} \end{pmatrix}. \quad (48)$$

Denote by  $\Phi_P^X = (\phi_0^X, \dots, \phi_P^X)^t$  the vector of  $P + 1$  components of the sequence of orthonormal polynomials associated to  $X$ , the three-term recurrence formulae ensures that

$$x\Phi_P^X(x) = J_P^X \Phi_P^X(x) + \sqrt{\beta_P}\phi_{P+1}^X(x)e_P, \quad (49)$$

where  $e_P = (0, \dots, 0, 1)^t$  of size  $P + 1$ . It is then interesting noting that the  $P + 1$  roots  $(\gamma_i)_{i \in \{0, \dots, P\}}$  of polynomial  $\phi_{P+1}^X$  are the eigenvalues of the Jacobi matrix of order  $P$ . We have  $\forall k \in \{0, \dots, P\}$

$$\gamma_k \Phi_P^X(\gamma_k) = J_P^X \Phi_P^X(\gamma_k),$$

the vectors  $(\Phi_P^X(\gamma_k))_{k \in \{0, \dots, P\}}$  consequently being the corresponding eigenvectors. This property is intensively used in order to build Gauss quadrature rules, see section 2.

#### A.4 Taking into account discrete/categorical input variables with gPC

In the previous sections, care has been taken to detail the case of discrete random variables with  $D + 1$  states even if the notations and the results did not particularly need it. We here want to insist on the fact that discrete input random variables are included in the gPC framework (already in [77] or in the Askey scheme [5]). The construction procedure for the gPC basis of any discrete law is exactly the same as for continuous random variables or at least the same as for random variables having an *infinite* sequence of orthonormal polynomials. The only difference being one should not try to build the gPC basis after a certain order depending on the number of states of the categorical random variable. By convention in the following and without loss of generalities, we denote by  $(\phi_k^X(X))_{k \in \mathbb{N}}$  the gPC basis associated to any random variable  $X$ : if  $X$  is discrete then  $\exists k_0 \in \mathbb{N}$  such that  $\forall k > k_0$ ,  $\phi_k^X$  has no sense but the notation still holds.

<sup>30</sup>As the sequence  $(\beta_k)_{k \in \mathbb{N}}$  is obviously related to the sequence of moments, see (41).