Uncertainty quantification for nonlinear difference equations with dependent random inputs via a stochastic Galerkin projection technique

J. Calatayud[a], J.-C. Cortés[a], M. Jornet[a,*]

[a]*Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain*

Abstract

Discrete stochastic systems model discrete response data on some phenomenon with inherent uncertainty. The main goal of uncertainty quantification is to derive the probabilistic features of the stochastic system. This paper deals with theoretical and computational aspects of uncertainty quantification for nonlinear difference equations with dependent random inputs. When the random inputs are independent random variables, a generalized Polynomial Chaos (gPC) approach has been usually used to computationally quantify the uncertainty of stochastic systems. In the gPC technique, the stochastic Galerkin projections are done onto linear spans of orthogonal polynomials from the Askey-Wiener scheme or from Gram-Schmidt orthonormalization procedures. In this regard, recent results have established the algebraic or exponential convergence of these Galerkin projections to the solution process. In this paper, as the random inputs of the difference equation may be dependent, we perform Galerkin projections directly onto linear spans of canonical polynomials. The main contribution of this paper is to study the spectral convergence of these Galerkin projections for the solution process of general random difference equations. Spectral convergence is important to derive the main statistics of the response process at a cheap computational expense. In this regard, the numerical experiments bring to light the theoretical discussion of the paper.

Keywords: Nonlinear random difference equation; Stochastic Galerkin projection technique; Uncertainty quantification; Dependent random inputs.
2010 MSC: 65Q10; 65C20; 60H35; 39A50; 41A65.

1. Introduction

Stochastic systems allow getting a better understanding of the processes involved in complex phenomena with inherent uncertainty. These phenomena may belong to the applied fields of Physics, Epidemiology, Biology, Engineering, etc. Essentially, stochastic systems are dynamical

*Corresponding author
Email addresses: jucagre@doctor.upv.es (J. Calatayud), jccortes@imm.upv.es (J.-C. Cortés), marjorsa@doctor.upv.es (M. Jornet)

systems (continuous or discrete) in which the involved input parameters are random rather than constants. From a modeling standpoint, it is important to obtain the main statistics corresponding to the response process, and this is called uncertainty quantification [1].

Among the most important statistics of the response process are the expectation and the variance at each time instant. These two statistics provide measures of the average and the dispersion, respectively, and they permit having a good understanding of the uncertainty associated to the system output. Random variables with well-defined expectation and variance are encompassed in the Lebesgue space L^2, which possesses the good structure of a Hilbert space. The convergence endowed by the metric of L^2, so-called mean square convergence , preserves the convergence of the expectation and variance. Thus, representing the response process as a limit of stochastic processes in L^2 at each time instant (for instance, via infinite series expansions), allows approximating its expectation and variance.

Galerkin methods have been extensively used in computational uncertainty quantification for stochastic systems, especially as an application of the so-called generalized Polynomial Chaos (gPC) [2, 3] for continuous stochastic systems (random ordinary and partial differential equations [4, 5]). With many applications in practice, see [6, 7, 8, 9, 10, 11], the algebraic or exponential mean square convergence of gPC-based Galerkin projections for random differential equations has been established in [12, 13].

However, Galerkin methods have not been applied to random difference equations with such an emphasis. Random recursive equations are essential to model discrete response data on phenomena with uncertainty. When time is large, the explicit expression of the response to the random difference equation may be a huge complex formula involving the random input parameters, so direct uncertainty quantification becomes an impracticable task, thus the necessity of appropriate stochastic Galerkin methods. The authors of this paper have proved recently in [14] that algebraic mean square convergence of adaptive gPC-based Galerkin projections [15, 16] for random difference equations can be expected under general conditions. However, the reasoning used in [14] only works for independent random input parameters. It is both of theoretical and of practical interest to extend these results to dependent random inputs. In [17], the authors proposed a computational approach to deal with continuous stochastic systems with dependent random coefficients. The Galerkin projections, instead of being done onto orthogonal polynomials, are calculated onto multivariate polynomials from the canonical basis evaluated at the random inputs. The novelty of our paper is to apply this technique to nonlinear difference equations with dependent random inputs, and to study both from a theoretical and computational standpoint the spectral mean square convergence to the time-discrete solution stochastic process.

The structure of this paper is the following. Section 2 establishes conditions under which a random vector depending on dependent random inputs is a mean square limit of multivariate polynomials evaluated at those inputs, with spectral convergence rate. This provides a stochastic Galerkin projection technique to quantify computationally the uncertainty for difference equations with dependent random inputs. Section 3 shows a theoretical discussion on the spectral mean square convergence of these Galerkin projections. In Section 4, we illustrate our findings via numerical experiments. In Section 5 conclusions are drawn.

2. Method and application to nonlinear random difference equations

In this section, we will show how to expand random vectors as a mean square limit of multivariate polynomials. We will apply this theory to nonlinear random difference equations, via a stochastic Galerkin projection technique. Thus, a computational approach to quantify the uncertainty for discrete random data will have been developed.

2.1. Method

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, where Ω is the sample space formed by outcomes $\omega \in \Omega$ and equipped with a σ-algebra of events $\mathcal{F}$ and a probability measure $\mathbb{P}$. We will work in the Hilbert space $(\mathrm{L}^2(\Omega), \langle, \rangle)$ of random vectors $u : \Omega \to \mathbb{R}^q$ with finite second order moments: $\|u\|_{\mathrm{L}^2(\Omega)} = \sqrt{\mathbb{E}[|u|^2]} < \infty$, where $\mathbb{E}$ stands for the expectation operator.

Suppose that u is written as a deterministic function of random input parameters: $u = g(\zeta_1, \ldots, \zeta_s)$, where $\zeta_1, \ldots, \zeta_s : \Omega \to \mathbb{R}$ are random variables and $g : \mathbb{R}^s \to \mathbb{R}^q$ is a Borel measurable function. We describe a procedure to approximate u in the mean square sense via multivariate polynomials in $\zeta = (\zeta_1, \ldots, \zeta_s)$.

Assume that ζ is an absolutely continuous random vector with finite moments of all orders. We allow any probability distributions for $\zeta_1, \ldots, \zeta_s$, not necessarily the standard ones. Moreover, we do not impose any independence condition on them. Let $C_i^p = \{1, \zeta_i, \ldots, \zeta_i^p\}$, $1 \leq i \leq s$, be the canonical basis of polynomials evaluated at ζ_i up to degree p. By means of a tensor product, see [17], we obtain the canonical basis of multivariate polynomials evaluated at ζ up to degree p: $\Xi^P = \{\phi_1(\zeta), \ldots, \phi_P(\zeta)\}$, where $P = (p+s)!/(p!s!)$, $\phi_j(\zeta) = \zeta_1^{i_1} \cdots \zeta_s^{i_s}$, being $i_1, \ldots, i_s \geq 0$, $i_1 + \ldots + i_s \leq p$, and the multi-index $(i_1, \ldots, i_s)$ is associated in a bijective manner with $j \in \{1, \ldots, P\}$ in such a way that $(0, \ldots, 0)$ corresponds to $j = 1$ (that is, $\phi_1 = 1$). Letting p and P grow up to infinity, we obtain the sequence $\{\phi_i(\zeta)\}_{i=1}^{\infty}$. Formally, we may expand u as

$$u = \sum_{i=1}^{\infty} \tilde{u}_i \phi_i(\zeta) \tag{1}$$

in $\mathrm{L}^2(\Omega)$, where $\tilde{u}_i$ are coefficients to be determined.

Notice that the difference with the classical gPC method [2, 3] is that Ξ^P is not formed by orthogonal polynomials from the Askey-Wiener scheme. In the gPC approach, when the distribution of a random input, ζ_i, does not coincide with a standard distribution from the Askey-Wiener scheme, one has to deal with inverses of cumulative distribution functions, and the convergence weakens to being in probability [2, Th. 5.7]. On the other hand, the main difference with the adaptive gPC approach suggested in [15, 16] is that we do not perform an orthonormalization procedure for each C_i^p, because the random inputs may not be independent. The orthonormalization procedure, usually done via a Gram-Schmidt method, may entail numerical errors due to loss of orthogonality [14, 18]. In the approach presented in our paper, these drawbacks do not arise.

From classical results on several complex variables [19], we can analyze conditions under which (1) holds at spectral rate.

Proposition 2.1. *Suppose that $|\zeta_i| \leq A_i$, for certain constants A_i, $i = 1, \ldots, s$. Assume that g is real analytic on $\mathbb{R}^s$. Then (1) holds in $\mathrm{L}^2(\Omega)$ at exponential rate.*

Proof. Since g is real analytic on $\mathbb{R}^s$, we may write $u = g(\zeta) = \sum_{i=1}^{\infty} \tilde{u}_i \phi_i(\zeta)$ pointwise on $\zeta(\omega) \in \mathbb{R}^s$. In multi-index notation, $u = g(\zeta) = \sum_{|\alpha| \geq 0} \tilde{u}_\alpha \zeta^\alpha$ pointwise on $\zeta(\omega) \in \mathbb{R}^s$, where $\zeta^\alpha = \zeta_1^{\alpha_1} \cdots \zeta_s^{\alpha_s}$, $\alpha = (\alpha_1, \ldots, \alpha_s)$. We will prove that, given any $b > 1$, $b^N \| \sum_{|\alpha| \geq N} \tilde{u}_\alpha \zeta^\alpha \|_{\mathrm{L}^2(\Omega)} \to 0$

as $N \to \infty$ (in fact, this means that the convergence to 0 is much faster than exponentially). By the triangular inequality and the boundedness condition on ζ_i, we have the first estimates

$$\begin{aligned} b^N \left\| \sum_{|\alpha| \geq N} \tilde{u}_\alpha \zeta^\alpha \right\|_{\mathrm{L}^2(\Omega)} &\leq b^N \sum_{|\alpha| \geq N} |\tilde{u}_\alpha| \|\zeta^\alpha\|_{\mathrm{L}^2(\Omega)} = b^N \sum_{|\alpha| \geq N} |\tilde{u}_\alpha| \|\zeta_1^{\alpha_1} \cdots \zeta_s^{\alpha_s}\|_{\mathrm{L}^2(\Omega)} \\ &\leq b^N \sum_{|\alpha| \geq N} |\tilde{u}_\alpha| A_1^{\alpha_1} \cdots A_s^{\alpha_s} = b^N \sum_{|\alpha| \geq N} |\tilde{u}_\alpha| A^\alpha, \end{aligned} \tag{2}$$

where $A = (A_1, \dots, A_s)$ and $A^\alpha = A_1^{\alpha_1} \cdots A_s^{\alpha_s}$. By Cauchy-Hadamard Theorem in several variables, see [19, Th. 4, p. 32],

$$\limsup_{|\alpha| \to \infty} \sqrt[|\alpha|]{|\tilde{u}_\alpha| \rho^\alpha} = 1,$$

for all $\rho = (\rho_1, \dots, \rho_s) \in (0, \infty)^s$. Thus, for large $|\alpha|$, $\sqrt[|\alpha|]{|\tilde{u}_\alpha| \rho^\alpha} \leq 2$, that is, $|\tilde{u}_\alpha| \leq 2^{|\alpha|}/\rho^\alpha$. Taking $\rho_1 = \ldots = \rho_s = r$ that will be determined later, $|\tilde{u}_\alpha| \leq (2/r)^{|\alpha|} = \delta^{|\alpha|}$. Take r such that $\delta A_i \leq 1/(6b)$, for $i = 1, \dots, s$. Then, from (2),

$$\begin{aligned} b^N \left\| \sum_{|\alpha| \geq N} \tilde{u}_\alpha \zeta^\alpha \right\|_{\mathrm{L}^2(\Omega)} &\leq b^N \sum_{|\alpha| \geq N} |\tilde{u}_\alpha| A^\alpha \leq b^N \sum_{|\alpha| \geq N} \delta^{|\alpha|} A^\alpha = b^N \sum_{|\alpha| \geq N} (\delta A_1)^{\alpha_1} \cdots (\delta A_s)^{\alpha_s} \\ &\leq b^N \sum_{|\alpha| \geq N} \left(\frac{1}{6b}\right)^{\alpha_1} \cdots \left(\frac{1}{6b}\right)^{\alpha_s} = b^N \sum_{|\alpha| \geq N} \left(\frac{1}{6b}\right)^{|\alpha|} \leq \sum_{|\alpha| \geq N} \frac{1}{6^{|\alpha|}} = \sum_{j=N}^{\infty} \sum_{|\alpha| = j} \frac{1}{6^{|\alpha|}} \\ &= \sum_{j=N}^{\infty} \frac{1}{6^j} \cdot \operatorname{card}\{\alpha = (\alpha_1, \dots, \alpha_s) : |\alpha| = j\} = \sum_{j=N}^{\infty} \frac{1}{6^j} \binom{j+s-1}{j}, \end{aligned} \tag{3}$$

where the identity $\operatorname{card}\{\alpha = (\alpha_1, \dots, \alpha_s) : |\alpha| = j\} = \binom{j+s-1}{j}$ comes from [2, (5.26), p. 65]. If $j \geq N$ and N is sufficiently large, the inequality $j + s - 1 \leq 2j$ holds. Then the binomial coefficient may be bound as follows:

$$\binom{j+s-1}{j} \leq \left(\frac{\mathrm{e}(j+s-1)}{j}\right)^j \leq (2\mathrm{e})^j \tag{4}$$

(the first inequality is a standard upper bound for binomial coefficients: $(\frac{n\mathrm{e}}{k})^k = \frac{n^k}{k^k} \sum_{j=0}^{\infty} \frac{k^j}{j!} \geq \frac{n^k}{k!} \geq \binom{n}{k}$). Thus, from (3) and (4), the statement of the proposition follows:

$$b^N \left\| \sum_{|\alpha| \geq N} \tilde{u}_\alpha \zeta^\alpha \right\|_{\mathrm{L}^2(\Omega)} \leq \sum_{j=N}^{\infty} \left(\frac{2\mathrm{e}}{6}\right)^j = \sum_{j=N}^{\infty} \left(\frac{\mathrm{e}}{3}\right)^j = \frac{(\mathrm{e}/3)^N}{1 - \mathrm{e}/3} \xrightarrow{N \to \infty} 0.$$

□

Compare the convergence established in Proposition 2.1 with the characterization proved in [12] for the convergence of gPC expansions. An implication of [12, Th. 3.4 and Th. 3.6] is that gPC expansions by means of orthogonal polynomials with bounded random inputs converge in $\mathrm{L}^2(\Omega)$. The result derived from Proposition 2.1 establishes convergence for expansions via non-orthogonal multivariate polynomials.

2.2. Application to nonlinear random difference equations

Consider a system of difference equations

$$u(m+1) = R(\zeta, u(m)), \quad \zeta = (\zeta_1, \ldots, \zeta_s), \tag{5}$$

where $u(m) : \Omega \to \mathbb{R}^q$ is a random vector for each step m (i.e., a time-discrete stochastic process) and $R : \mathbb{R}^s \times \mathbb{R}^q \to \mathbb{R}^q$ is a nonlinear Borel measurable function. The terms $\zeta_1, \ldots, \zeta_s$ are random variables that represent the inputs of the stochastic system (5). The initial condition $u(0) = u_0$ is supposed to be a constant vector in $\mathbb{R}^q$.

As we have previously seen, we may formally write the solution stochastic process of the system of random difference equations (5) as

$$u(m) = \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta). \tag{6}$$

Truncating the series from (6) up to order P gives the motivation to look for an approximate solution to (5) of the form

$$\hat{u}^P(m) = \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\zeta). \tag{7}$$

This is called the stochastic Galerkin projection onto the span of $\{\phi_i(\zeta)\}_{i=1}^P$. This method has been already developed in [17] for continuous stochastic systems (random ordinary differential equations), but only from a computational point of view.

Substituting (7) into (5) gives

$$\sum_{i=1}^{P} \hat{u}_i^P(m+1)\phi_i(\zeta) = R\left(\zeta, \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\zeta)\right),$$

with initial condition

$$\sum_{i=1}^{P} \hat{u}_i^P(0)\phi_i(\zeta) = u_0. \tag{8}$$

We perform the inner product with each $\phi_k(\zeta)$, $k = 1, \ldots, P$ (stochastic Galerkin projection technique):

$$\sum_{i=1}^{P} \hat{u}_i^P(m+1)\langle \phi_i(\zeta), \phi_k(\zeta)\rangle = \langle R\left(\zeta, \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\zeta)\right), \phi_k(\zeta)\rangle, \tag{9}$$

$$\sum_{i=1}^{P} \hat{u}_i^P(0)\langle \phi_i(\zeta), \phi_k(\zeta)\rangle = u_0\mathbb{E}[\phi_k(\zeta)]. \tag{10}$$

This gives a deterministic system of difference equations for $\{\hat{u}_i^P(m)\}_{i=1}^P$, which can be solved numerically by repeated iteration in (9) from the initial condition (10) up to the time m desired, whatever the degree of nonlinearity of the map R. Thus, the approximation $\hat{u}^P(m)$ is numerically computable.

Notice that, since u_0 is constant and $\phi_1 = 1$, equations (8) and (10) are trivial, because $\hat{u}_1^P(0) = u_0$ and $\hat{u}_i^P(0) = 0$ for $2 \leq i \leq P$. However, for the sake of completeness, we have shown the full development of the stochastic Galerkin projection technique.

Based on intuition and numerical experiments, one expects $\hat{u}^P(m) \to u(m)$ as $P \to \infty$ in $\mathrm{L}^2(\Omega)$, for each $m \geq 1$. In Section 3, we discuss theoretically the convergence of the Galerkin projections, while Section 4 illustrates the theoretical findings via numerical experiments. The mean square convergence allows approximating the main statistics of the response stochastic process $u(m)$, say the expectation $\mathbb{E}[u(m)]$ and the covariance matrix, via

$$\mathbb{E}[\hat{u}^P(m)] = \sum_{i=1}^{P} \hat{u}_i^P(m)\mathbb{E}[\phi_i(\zeta)] \tag{11}$$

and

$$\mathbb{C}\mathrm{ov}[\hat{u}^P(m), \hat{u}^P(m')] = \sum_{i,j=1}^{P} \hat{u}_i^P(m)\hat{u}_j^P(m')\mathbb{C}\mathrm{ov}[\phi_i(\zeta), \phi_j(\zeta)], \tag{12}$$

respectively.

In the system from (9), an important matrix G which one has to deal with is $G_{ik} = \langle \phi_i(\zeta), \phi_k(\zeta)\rangle$. This matrix G is symmetric and positive definite, hence invertible, and it corresponds to the Gram matrix of $\{\phi_i(\zeta)\}_{i=1}^P$ [20, Cor. 7.2.9, Th. 7.2.10]. Furthermore, G^{-1} is positive definite too [20, p. 397]. Numerically, it is checked that this matrix is ill-conditioned, which may entail numerical errors for large P. This is a limitation of our computational approach. In this paper, we will not deal with specific numerical procedures to deal with ill-conditioned linear systems [21].

3. Convergence for nonlinear random difference equations

In this section, we analyze the theoretical convergence of the Galerkin projection $\hat{u}^P(m)$ given by (7) to the solution process $u(m)$ of the system of nonlinear random difference equations given by (5).

The main result to be proved is the following Theorem 3.1. For the proof of this result we will use some ideas from [13, 14]. Reference [13] studies, in the context of random differential equations, the convergence of stochastic Galerkin projections based on gPC expansions for independent random inputs. In [14], we have analyzed the convergence of stochastic Galerkin projections based on adaptive gPC [15, 16], in the context of difference equations with independent random inputs.

Theorem 3.1. *Consider the system of random difference equations $u(m+1) = R(\zeta, u(m))$, where $\zeta = (\zeta_1, \ldots, \zeta_s) : \Omega \to \mathbb{R}^s$ is an absolutely continuous random vector, $u(m) : \Omega \to \mathbb{R}^q$ is a random vector, $R : \mathbb{R}^s \times \mathbb{R}^q \to \mathbb{R}^q$ is a Borel measurable function, and the initial condition $u(0) = u_0$ is a deterministic constant in $\mathbb{R}^q$. Let $\zeta_1, \ldots, \zeta_s$ be bounded random variables, not necessarily independent. Let $J_1, \ldots, J_s \subseteq \mathbb{R}$ be compact intervals that contain the support of $\zeta_1, \ldots, \zeta_s$, respectively. Let $I = \prod_{i=1}^s J_i$ be the multidimensional compact rectangle that contains the support of ζ. Let g_m be the solution of the system of difference equations for $u(m)$, i.e., $u(m) = g_m(\zeta)$. Assume that g_m is real analytic on $\mathbb{R}^s$. Suppose that R is Lipschitz on $I \times \mathbb{R}^q$: there exists a constant $K > 0$ such that $\|R(\zeta, v_1) - R(\zeta, v_2)\|_{\mathrm{L}^2(\Omega)} \leq K\|v_1 - v_2\|_{\mathrm{L}^2(\Omega)}$, for every pair of random vectors $v_1, v_2 \in \mathrm{L}^2(\Omega)$. Let G be the $P \times P$ matrix defined as $G_{ik} = \mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)]$.*

Then the Galerkin projection $\hat{u}^P(m)$ defined by (7) and the partial sum $\tilde{u}^P(m) = \sum_{i=1}^P \tilde{u}_i(m)\phi_i(\zeta)$

from (6) satisfy the following inequality:

$$\|\hat{u}^P(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)} \leq \|\tilde{u}^P(0) - u_0\|_{\mathrm{L}^2(\Omega)}(K\Delta_P)^m$$
$$+ \sum_{j=1}^{m} \left(K\Delta_P\|u(j-1) - \tilde{u}^P(j-1)\|_{\mathrm{L}^2(\Omega)} + \|u(j) - \tilde{u}^P(j)\|_{\mathrm{L}^2(\Omega)}\right)(K\Delta_P)^{m-j}, \tag{13}$$

where

$$\Delta_P = \sqrt{\sum_{l,k=1}^{P} \|\phi_l\|_{\mathrm{L}^2(\Omega)}\|\phi_k\|_{\mathrm{L}^2(\Omega)}|(G^{-1})_{lk}|}. \tag{14}$$

Proof. Since g_m is real analytic on $\mathbb{R}^s$, we derive that $\tilde{u}^P(m) \to u(m)$ as $P \to \infty$ in $\mathrm{L}^2(\Omega)$, for each $m \geq 0$, as a consequence of Proposition 2.1. Thereby, we may expand $u(m)$ as in (6): $u(m) = \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta)$. Substituting this expression for $u(m)$ into (5), $\sum_{i=1}^{\infty} \tilde{u}_i(m+1)\phi_i(\zeta) = R(\zeta, \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta))$. Multiplying by $\phi_k(\zeta)$, $k = 1, \ldots, P$, and applying the expectation operator, we obtain

$$\sum_{i=1}^{\infty} \tilde{u}_i(m+1)\mathbb{E}[\phi_i(\zeta)\phi_k(\zeta)] = \langle R(\zeta, \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta)), \phi_k(\zeta)\rangle. \tag{15}$$

On the other hand, recall the deterministic recursive equations (9) satisfied by the Galerkin projection $\hat{u}^P(m)$. Combining both (9) and (15) implies

$$\sum_{i=1}^{P}(\tilde{u}_i(m+1) - \hat{u}_i^P(m+1))G_{ik}$$
$$= \underbrace{\langle R\left(\zeta, \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta)\right) - R\left(\zeta, \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\zeta)\right), \phi_k(\zeta)\rangle}_{:=b_k^P(m)} - \underbrace{\sum_{i=P+1}^{\infty} \tilde{u}_i(m+1)G_{ik}}_{:=g_k^P(m+1)}.$$

Let us put this expression in matrix form. Let

$$\tilde{u}_\vee^P(m) = (\tilde{u}_i(m))_{i=1}^P, \quad \hat{u}_\vee^P(m) = (\hat{u}_i^P(m))_{i=1}^P,$$
$$b^P(m) = (b_k^P(m))_{k=1}^P, \quad g^P(m) = (g_k^P(m))_{k=1}^P.$$

If $G = (G_{ik})_{1\leq i,k\leq P}$, then

$$\tilde{u}_\vee^P(m+1) - \hat{u}_\vee^P(m+1) = G^{-1}(b^P(m) - g^P(m+1)). \tag{16}$$

Since G is symmetric and positive definite, we can consider the norms $\|x\|_G = \sqrt{x^\top G x}$ and $\|x\|_{G^{-1}} = \sqrt{x^\top G^{-1} x}$, where $\top$ stands for the transpose operator. From (16),

$$\|\tilde{u}^P(m+1) - \hat{u}^P(m+1)\|_{\mathrm{L}^2(\Omega)} = \|\tilde{u}_\vee^P(m+1) - \hat{u}_\vee^P(m+1)\|_G$$
$$= \|G^{-1}(b^P(m) - g^P(m+1))\|_G = \|b^P(m) - g^P(m+1)\|_{G^{-1}}$$
$$\leq \|b^P(m)\|_{G^{-1}} + \|g^P(m+1)\|_{G^{-1}}. \tag{17}$$

Let us estimate both $\|b^P(m)\|_{G^{-1}}$ and $\|g^P(m+1)\|_{G^{-1}}$. By Cauchy-Schwarz inequality, the Lipschitz condition for R and the triangular inequality, we have

$$|b_k^P(m)| \leq \left\|R\left(\zeta, \sum_{i=1}^{\infty} \tilde{u}_i(m)\phi_i(\zeta)\right) - R\left(\zeta, \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\zeta)\right)\right\|_{\mathrm{L}^2(\Omega)} \|\phi_k(\zeta)\|_{\mathrm{L}^2(\Omega)}$$
$$\leq \left(K\|\tilde{u}^P(m) - \hat{u}^P(m)\|_{\mathrm{L}^2(\Omega)} + K\|u(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)}\right)\|\phi_k(\zeta)\|_{\mathrm{L}^2(\Omega)}.$$

As a consequence, we get an upper bound for $\|b^P(m)\|_{G^{-1}}$:

$$\|b^P(m)\|_{G^{-1}} = \sqrt{\sum_{l,k=1}^{P} b_l^P(m) b_k^P(m)(G^{-1})_{lk}}$$
$$\leq \left(K\|\tilde{u}^P(m) - \hat{u}^P(m)\|_{\mathrm{L}^2(\Omega)} + K\|u(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)}\right)\Delta_P, \quad (18)$$

where Δ_P is already defined in (14). On the other hand, the estimate for the second term $\|g^P(m+1)\|_{G^{-1}}$ is derived as follows:

$$\|g^P(m+1)\|_{G^{-1}}^2 = \sum_{l,k=1}^{P} g_l^P(m+1) g_k^P(m+1)(G^{-1})_{lk}$$
$$= \sum_{l,k=1}^{P} \sum_{i=P+1}^{\infty} \sum_{j=P+1}^{\infty} \tilde{u}_i(m+1) G_{il} \tilde{u}_j(m+1) G_{jk} (G^{-1})_{lk}$$
$$= \sum_{i=P+1}^{\infty} \sum_{j=P+1}^{\infty} \left(\sum_{l,k=1}^{P} G_{il}(G^{-1})_{lk} G_{kj} \right) \tilde{u}_i(m+1)\tilde{u}_j(m+1)$$
$$= \sum_{i=P+1}^{\infty} \sum_{j=P+1}^{\infty} G_{ij} \tilde{u}_i(m+1)\tilde{u}_j(m+1)$$
$$= \|u(m+1) - \tilde{u}^P(m+1)\|_{\mathrm{L}^2(\Omega)}^2. \quad (19)$$

Taking into account the estimates obtained in (18) and (19), a new inequality arises from (17):

$$\|\tilde{u}^P(m+1) - \hat{u}^P(m+1)\|_{\mathrm{L}^2(\Omega)} \leq K\Delta_P \|\tilde{u}^P(m) - \hat{u}^P(m)\|_{\mathrm{L}^2(\Omega)}$$
$$+ K\Delta_P \|u(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)} + \|u(m+1) - \tilde{u}^P(m+1)\|_{\mathrm{L}^2(\Omega)}.$$

This inequality may be seen as a non-autonomous linear recursive equation (with inequality) for $\|\tilde{u}^P(m) - \hat{u}^P(m)\|_{\mathrm{L}^2(\Omega)}$. Using the general solution for non-autonomous first-order linear difference equations [22, Ch. 1], we conclude that

$$\|\hat{u}^P(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)} \leq \|\tilde{u}^P(0) - u_0\|_{\mathrm{L}^2(\Omega)}(K\Delta_P)^m$$
$$+ \sum_{j=1}^{m} \left(K\Delta_P \|u(j-1) - \tilde{u}^P(j-1)\|_{\mathrm{L}^2(\Omega)} + \|u(j) - \tilde{u}^P(j)\|_{\mathrm{L}^2(\Omega)} \right) (K\Delta_P)^{m-j},$$

which is exactly the inequality (13) that we wanted to prove.

□

An important consequence of this theorem is the following. Suppose that g_m is real analytic on $\mathbb{R}^s$, for all $m \geq 1$. This is not an unrealistic assumption, since going backwards in (5) consists of compositions of functions, and the composition of analytic functions is analytic again. As it was previously discussed in Section 2, we have that $\|u(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)} \to 0$ as $P \to \infty$ at exponential rate, for each $m \geq 0$. If Δ_P had an exponential growth, i.e., $\Delta_P \leq C\mathrm{e}^{rP}$ for certain $C, r > 0$, then we would conclude that $\|\hat{u}^P(m) - \tilde{u}^P(m)\|_{\mathrm{L}^2(\Omega)} \to 0$ as $P \to \infty$ at exponential rate, for each $m \geq 0$. For example, if our sequence of polynomials $\{\phi_i(\zeta)\}_{i=1}^P$ were orthonormal, then $G = I_P$ and $\Delta_P = P$, and we recover the main result established in [14]. Thus, if g_m is real analytic on $\mathbb{R}^s$, the convergence of the stochastic Galerkin projection depends upon the growth rate of Δ_P. It is an open question for us whether Δ_P always grows exponentially. This would solve the problem of the spectral convergence of the Galerkin projection. Nonetheless, in practice, it is possible to analyze empirically the growth of Δ_P. By fitting a regression line for the set of points $(P, \log(P))$, one gets an idea on whether the growth of Δ_P is exponential. If this is the case, we ensure that the Galerkin projections converge to the solution process $u(m)$ at exponential, $m \geq 0$. See Section 4 for examples of this methodology. We will see numerically that, in general, Δ_P increases at most exponentially in P. In fact, there are examples, for instance when G is the Hilbert matrix of size $P \times P$ [23, 24, 25, 26, 27] (one random input parameter in (5) with Uniform(0, 1) distribution), in which Δ_P grows exactly exponentially.

Our development emphasizes the importance of a detailed analysis for the matrix G. If G^{-1} has large entries as P increases, two problems arise: from a theoretical point of view, the Galerkin projection may not converge; from a computational standpoint, the condition number of G increases, which makes the computations in the computer less accurate. Section 4 illustrates how disastrous error may appear in practice when P grows.

Concerning the hypothesis of boundedness for ζ in Theorem 3.1, which was necessary to approximate g_m by polynomials on I, see Proposition 2.1, we would like to remark that, in practice, this is not a restrictive assumption. If one works with an unbounded random input parameter ζ_i, then one may truncate this random variable, preserving nearly all the probabilistic features of it. This assertion is supported by Chebyshev's inequality [28]. In addition, in some evolution equations, for instance epidemic models, parameters usually refer to proportions (proportion of infectives, proportion of vaccinated individuals, etc.), so that their domain of definition is [0, 1].

Remark 3.2. The Lipschitz condition $\|R(\zeta, v_1) - R(\zeta, v_2)\|_{\mathrm{L}^2(\Omega)} \leq K\|v_1 - v_2\|_{\mathrm{L}^2(\Omega)}$, for every pair of random vectors $v_1, v_2 \in \mathrm{L}^2(\Omega)$, may be difficult to check in practice. We show a stronger condition but which might be helpful in numerical examples. Let $\mathcal{D}_m$ be a subset of $\mathbb{R}^q$, independent of ζ, that contains the support of the random vector $u(m)$, and let $\mathcal{D} = \cup_{m=0}^{\infty} \mathcal{D}_m$. Such a set $\mathcal{D}_m$ exists because $|u(m)| \leq \|g_m\|_{\mathrm{L}^\infty(I)}$, being $\|g_m\|_{\mathrm{L}^\infty(I)} < \infty$ by compactness of I and continuity of g_m. If

$$|R(\xi, w_1) - R(\xi, w_2)| \leq K|w_1 - w_2|, \quad w_1, w_2 \in \mathcal{D},\ \xi \in I, \tag{20}$$

then the Lipschitz condition from Theorem 3.1 holds. Indeed, for each $\xi \in I$, we apply Tietze Extension Theorem [29, Th. 1] to $R(\xi, \cdot) : \mathcal{D} \to \mathbb{R}^q$. This gives an extension of R to $I \times \mathbb{R}^q$ such that $|R(\xi, w_1) - R(\xi, w_2)| \leq K|w_1 - w_2|$ for all $w_1, w_2 \in \mathbb{R}^q$ and $\xi \in I$. Applying the $\mathrm{L}^2(\Omega)$ norm, we derive $\|R(\zeta, v_1) - R(\zeta, v_2)\|_{\mathrm{L}^2(\Omega)} \leq K\|v_1 - v_2\|_{\mathrm{L}^2(\Omega)}$, for every pair of random vectors $v_1, v_2 \in \mathrm{L}^2(\Omega)$. In Section 4, we show the generality of the Lipschitz condition (20).

Remark 3.3. There are certainly examples in which the Galerkin projections do not converge. Consider the simple difference equation

$$\begin{cases} u(m+1) = u(m) + \frac{1}{\eta}, \\ u(0) = 0, \end{cases}$$

where $\eta \sim \mathrm{logN}(0, 1)$. Its solution is given by $u(m) = m/\eta$. By [12, Prop. 4.2], the gPC expansion of the random variable $\zeta = 1/\eta$ with respect to the orthonormal polynomials $\{\psi_i(\eta)\}_{i=1}^{\infty}$ in η (these polynomials can be constructed in terms of Stieltjes-Wigert polynomials, see [12, Appendix A]) does not converge in mean square to ζ. That is, if $\bar{\zeta}_i = \mathbb{E}[\zeta\psi_i(\eta)]$ is the i-th Fourier coefficient of ζ (see [12, Appendix B] for an explicit expression of $\bar{\zeta}_i$), then $\zeta \neq \sum_{i=1}^{\infty} \bar{\zeta}_i\psi_i(\eta)$. By [2, Th. 3.3],

$$\left\| u(m) - \sum_{i=1}^{P} (m\bar{\zeta}_i)\psi_i(\eta) \right\|_{\mathrm{L}^2(\Omega)} \leq \left\| u(m) - \sum_{i=1}^{P} \hat{u}_i^P(m)\phi_i(\eta) \right\|_{\mathrm{L}^2(\Omega)},$$

therefore the Galerkin projections $\hat{u}^P(m)$ cannot converge to $u(m)$, $m \geq 1$.

4. Numerical experiments

In this section we perform numerical experiments for time-discrete models with randomness. The goal is to assess the theoretical convergence of the Galerkin projections in order to carry out uncertainty quantification. The computations are performed in the software Mathematica®, version 11.2 [30], installed on an Intel® Core™ i7 CPU 3.1 GHz.

Example 4.1. Let us consider the nonlinear random difference equation

$$u(m+1) = \sin(r)u(m)(T - u(m)). \tag{21}$$

The terms r and T are considered random variables, so that $u(m)$ is a random variable for each $m \geq 1$. The initial condition $u(0) = u_0$ is assumed to be constant.

The equation (21) corresponds to a logistic model. We can interpret $u(m)$ as the number of infected individuals at time m in a population of size T. The product $u(m)(T - u(m))$ is the number of contacts between infected and susceptible individuals. The parameter $\sin(r)$ represents the proportion of those contacts that gives rise to a new infective (in terms of modeling this is somewhat artificial, but we want to test our methodology with nonlinear terms).

In order that model (21) makes sense, we need $u(m) \in [0, T]$ for all $m \geq 0$. If $0 \leq r \leq 4/T$ and $u_0 \in [0, T]$, then it is easily proved by induction on m that $u(m) \in [0, T]$. Thus, we will set probability distributions for r and T such that $0 \leq r \leq 4/T$.

Let $r \sim \text{Normal}(0.02, \sigma = 0.005)|_{[0.01,0.04]}$ and $T \sim \text{Triangular}(80, 100)$. These random variables are assumed to be independent. Notice that $0.01 \leq r \leq 0.04 = 4/100 \leq 4/T$, as required. As initial condition, we assume that $u(0) = u_0 = 3$ individuals were infected at the beginning.

Let us check that the conditions of Theorem 3.1 are satisfied. We have $q = 1$, $s = 2$ and $\zeta = (r, T)$. The probability distributions are absolutely continuous, with finite moments and bounded. We have $J_1 = [0.01, 0.04]$, $J_2 = [80, 100]$ and $I = J_1 \times J_2$. The function R is given by $R(r, T, u) = \sin(r)u(T - u)$. In the notation of Remark 3.2, take $\mathcal{D} = \mathcal{D}_m = [0, 100]$. For $\xi = (\xi_1, \xi_2) \in I$ and $w_1, w_2 \in \mathcal{D}$, $|R(\xi, w_1) - R(\xi, w_2)| = |\sin(\xi_1)w_1(\xi_2 - w_1) - \sin(\xi_1)w_2(\xi_2 - w_2)| \leq |\sin(\xi_1)||\xi_2||w_1 - w_2| + |\sin(\xi_1)||w_1 + w_2||w_1 - w_2| \leq 100|w_1 - w_2| + 200|w_1 - w_2| = 300|w_1 - w_2|$, therefore the Lipschitz condition (20) is satisfied. By Remark 3.2, the Lipschitz condition of Theorem 3.1 holds. Finally, the function g_m is real analytic on $\mathbb{R}^s = \mathbb{R}^2$, because the sine function is analytic.

By Theorem 3.1, the inequality (13) holds. Depending on the rate of growth of Δ_P, the Galerkin projection $\hat{u}^P(m)$ will converge or not. In Table 1, we analyze the increase of Δ_P (some inaccuracies in the computations of Table 1 might have occurred due to G being badly conditioned, especially for large p). Figure 1 shows a regression line for the set of points $(P, \log(\Delta_P))$. This gives an exponential model for (P, Δ_P), which is depicted in Figure 2. Since Δ_P grows slower than the exponential model, we conclude empirically that Δ_P has at most exponential growth. As g_m is real analytic on $\mathbb{R}^s = \mathbb{R}^2$, $\tilde{u}^P(m)$ converges to $u(m)$ at exponential rate. Thus, as a consequence of Theorem 3.1, the Galerkin projection $\hat{u}^P(m)$ will indeed converge to $u(m)$ at exponential rate.

p	1	2	3	4	5	6	7
P	3	6	10	15	21	28	36
Δ_P	45	1691	$67,036$	$557,137$	$2.00412 \cdot 10^7$	$1.7873 \cdot 10^8$	$8.5111 \cdot 10^8$

Table 1: Values of Δ_P. Example 4.1.

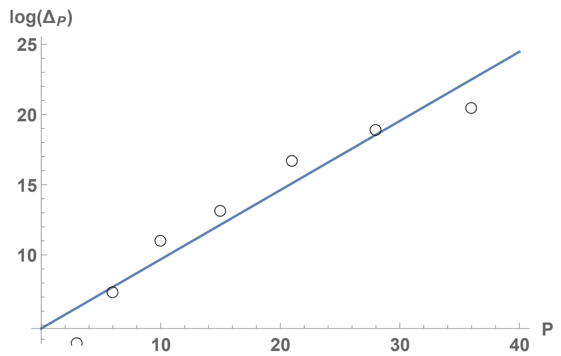

Figure 1: Regression line for the set of points $(P, \log(\Delta_P))$. Example 4.1.

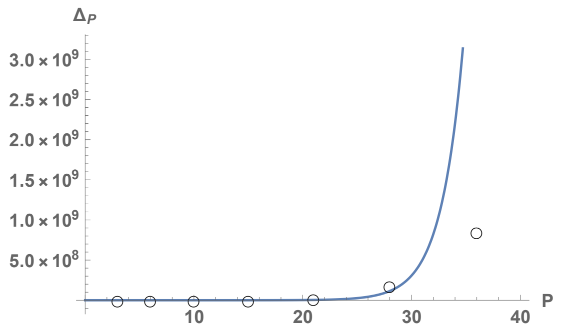

Figure 2: Exponential model for (P, Δ_P). Since Δ_P grows slower than the exponential model (solid line), we conclude empirically that Δ_P has at most exponential growth. Example 4.1.

In Figure 3, we perform numerical experiments for $p = 1, 2, 3, 4, 5, 6$. We show the average number of infected individuals together with a confidence interval with the rule [mean ± standard deviation], for $0 \leq m \leq 30$. The mean and variance are computed with (11)–(12). Since r and T are independent, the method based on adaptive gPC from [14] can also be applied. We observe in Figure 3 that both approaches yield the same results. We notice disastrous errors from $p = 5$ in our approach and from $p = 6$ in [14] method, due to inaccuracies in the computations: for [14], the main computational drawback is the loss of orthogonality in the Gram-Schmidt procedure; whereas in our approach, the ill-conditioned matrix G entails computational errors for large p. The best choice for uncertainty quantification in this example is $p = 4$.

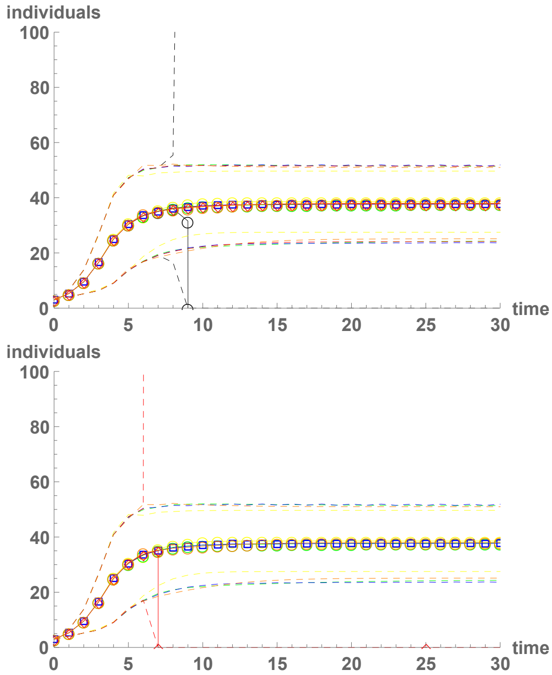

Figure 3: Average number of infected individuals (continuous lines) together with a confidence interval with the rule [mean ± standard deviation] (dashed lines), for $0 \leq m \leq 30$. Up: technique from [14]. Down: our approach. Observe that both techniques coincide for $p = 1$ (yellow), $p = 2$ (orange), $p = 3$ (green) and $p = 4$ (blue). For $p = 5$ (red) and/or $p = 6$ (black), disastrous errors appear due to the accumulation of errors. Example 4.1.

From a dynamics standpoint, we observe that the expected number of infected individuals, as well as the confidence intervals, tend to stabilize as $m \to \infty$. More concretely, the average number of infective tends to the expectation of the (random) fixed point $T - 1/\sin(r)$, $\mathbb{E}[T - 1/\sin(r)] = 37.7452$, while the typical deviation approaches $\sqrt{\mathbb{V}[T - 1/\sin(r)]} = 13.853$.

Example 4.2. We consider the same example as before, but we change the distribution of $\zeta = (r, T)$. We pick

$$\zeta \sim \text{Normal}(\mu, \Sigma)|_{[0.01,0.04]\times[80,100]}, \quad \mu = \begin{pmatrix} 0.02 \\ 90 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.000025 & 0.01 \\ 0.01 & 9 \end{pmatrix}.$$

The initial condition is taken $u(0) = u_0 = 3$. This example cannot be addressed with the techniques exposed in [14], since both random input parameters are not independent.

As in Example 4.1, the hypotheses of Theorem 3.1 hold. As a consequence, if the rate of growth of Δ_P is exponential, then the Galerkin projection $\hat{u}^P(m)$ will converge to $u(m)$ at exponential rate. In Table 2, we present how Δ_P increases as P augments (significant errors might have occurred in the computations of Table 2 due to G being badly conditioned, mainly for large p). Figure 4 presents a regression line for the set of points $(P, \log(\Delta_P))$. From this regression line, Figure 5 shows an exponential model for (P, Δ_P). It seems that there is at most exponential growth of Δ_P (because Δ_P grows slower than the exponential model), therefore Theorem 3.1 ensures the convergence of the Galerkin projections as $P \to \infty$.

p	1	2	3	4	5	6
P	3	6	10	15	21	28
Δ_P	80	4660	56, 764	114, 327	180, 509	396, 342

Table 2: Values of Δ_P. Example 4.2.

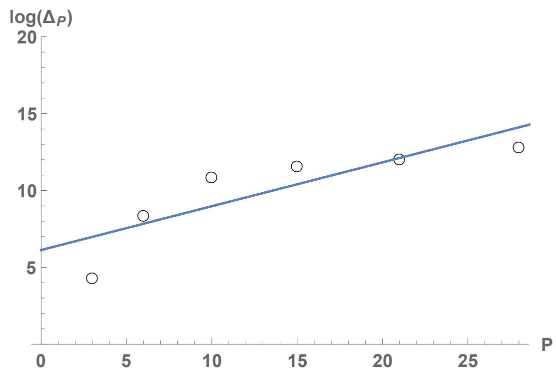

Figure 4: Regression line for the set of points $(P, \log(\Delta_P))$. Example 4.2.

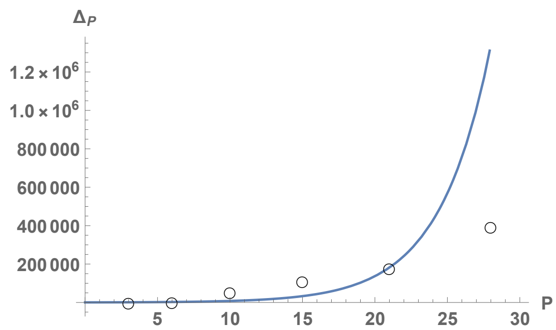

Figure 5: Exponential model for (P, Δ_P). Since Δ_P grows slower than the exponential model (solid line), we conclude empirically that Δ_P has at most exponential growth. Example 4.2.

In Figure 6, we show the numerical experiments for $p = 1, 2, 3$ corresponding to the average of infected individuals and the confidence interval [mean ± standard deviation]. The means and variances have been calculated with (11)–(12). For $p = 1$ and $p = 2$, similar results are obtained. However, from $p = 3$ catastrophic numerical errors appear. This implies that the best choice for uncertainty quantification in this example is $p = 2$.

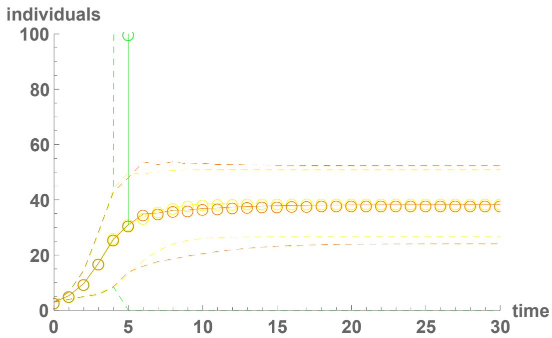

Figure 6: Average number of infected individuals (continuous lines) together with a confidence interval with the rule [mean ± standard deviation] (dashed lines), for $0 \leq m \leq 30$. For $p = 1$ (yellow) and $p = 2$ (orange) similar results are observed. For $p = 3$ (green), disastrous errors appear due to the accumulation of errors. Example 4.2.

Concerning the asymptotic behavior of the random discrete dynamical system, we observe that the expected number of infected individuals and the confidence intervals tend to stabilize as $m \to \infty$. The average number of infective tends to the expectation of the (random) fixed point $T - 1/\sin(r)$, $\mathbb{E}[T - 1/\sin(r)] = 37.8474$, whereas the standard deviation approaches $\sqrt{\mathbb{V}[T - 1/\sin(r)]} = 15.192$.

Example 4.3. In this example, we deal with the recursive equation

$$u(m+1) = 5\cos(\zeta_1 u(m) + \zeta_2).$$

Here, $\zeta = (\zeta_1, \zeta_2)$ are the random input parameters, with joint distribution

$$\zeta \sim \text{Dirichlet}(80, 4, 316).$$

The initial condition u_0 is taken as $u_0 = -10$. The degree of nonlinearity of this difference equation is higher than in the previous two examples. Due to the non-independence of the random input coefficients, this stochastic equation cannot be tackled with the techniques from [14].

Let us see that the assumptions of Theorem 3.1 hold. We have $q = 1$ and $s = 2$. The support of ζ_1 and ζ_2 is contained in $J_1 = J_2 = [0, 1]$. Let $I = J_1 \times J_2 = [0, 1]^2$. The function g_m is real analytic on ζ, because it consists of a composition of cosine functions. In the notation of Remark 3.2, $\mathcal{D}_m = \mathcal{D} = [-5, 5]$. Let us check the Lipschitz condition (20). For $w_1, w_2 \in \mathcal{D}$ and $\xi_1, \xi_2 \in [0, 1]$, by the Mean Value Theorem we have $|R(\xi_1, \xi_2, w_1) - R(\xi_1, \xi_2, w_2)| = 5|\cos(\xi_1 w_1 + \xi_2) - \cos(\xi_1 w_2 + \xi_2)| \leq 5|w_1 - w_2|$, as wanted.

As a consequence of Theorem 3.1, the inequality (13) holds. Depending on the growth of Δ_P, there will be exponential convergence of the Galerkin projections $\hat{u}^P(m)$ in $L^2(\Omega)$. Table 3 shows the values of Δ_P for $p = 1, 2, 3, 4, 5$ (significant errors might have occurred in the calculations of Table 3 due to G being ill-conditioned, mainly for large p). Figure 7 draws a regression line for the set of points $(P, \log(\Delta_P))$, from which Figure 8 depicts an exponential model for (P, Δ_P), in order to check empirically whether Δ_P grows exponentially. We deduce that Δ_P grows at most exponentially, because Δ_P lies below the exponential model. Hence, the Galerkin projections converge to the solution process at exponential rate, as wanted.

p	1	2	3	4	5
P	3	6	10	15	21
Δ_P	21	309	3758	20,296	30,072

Table 3: Values of Δ_P. Example 4.3.

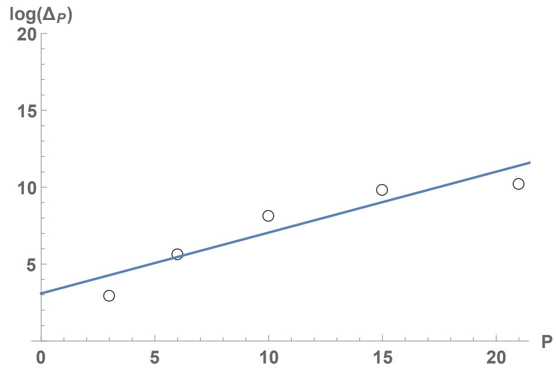

Figure 7: Regression line for the set of points $(P, \log(\Delta_P))$. Example 4.3.

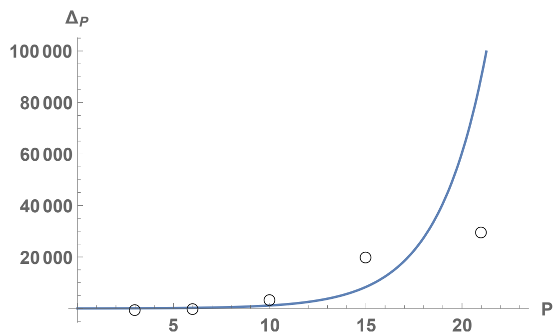

Figure 8: Exponential model for (P, Δ_P). Since Δ_P grows slower than the exponential model (solid line), we conclude empirically that Δ_P has at most exponential growth. Example 4.3.

In Figure 9, we perform numerical experiments for $p = 2, 3, 4$ corresponding to the average and standard deviation. The means and variances have been determined with the formulas (11)–(12). Similar results are obtained for $p = 2, 3, 4$, which agrees with the exponential convergence stated by Theorem 3.1.

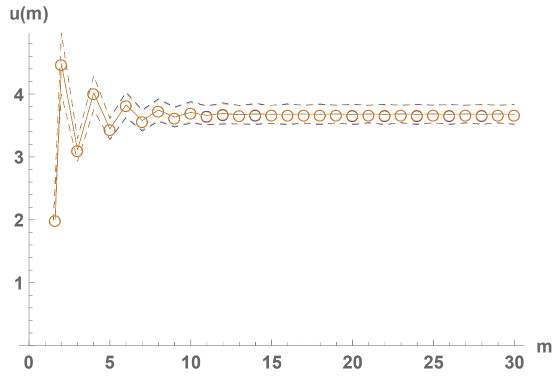

Figure 9: Average statistic together with a confidence interval with the rule [mean ± standard deviation] (dashed lines), for $0 \leq m \leq 30$. For $p = 2$ (orange), $p = 3$ (green) and $p = 4$ (blue) similar results are obtained. Example 4.3.

Concerning the asymptotic dynamics of the discrete stochastic system, we observe stabilization of both the mean and variance statistics in Figure 9. The mean goes towards the value 3.676, while the variance to 0.025, as $m \to \infty$.

5. Conclusions

In this paper, we have analyzed the application of a stochastic Galerkin method for uncertainty quantification for nonlinear difference equations with dependent random input parameters. The Galerkin projections have been done onto canonical multivariate polynomials evaluated at the random inputs. A theoretical discussion has been developed to analyze the mean square spectral convergence of these Galerkin projections to the time-discrete solution process. In the numerical experiments, we have performed uncertainty quantification for some specific random difference equations with different degrees of nonlinearity. We have observed rapid convergence of the Galerkin projections, although large truncation orders may entail significant numerical errors because the Gram matrix associated to the canonical polynomial basis is ill-conditioned. Thus, for proper uncertainty quantification in practice, a correct and cautious choice of the basis length is proved to be important.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía y Competitividad grant MTM2017–89664–P. The co-author Marc Jornet acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

Conflict of Interest Statement

The authors declare that there is no conflict of interests regarding the publication of this article.

[1] Smith RC. *Uncertainty Quantification. Theory, Implementation, and Application*. SIAM Computational Science & Engineering; 2014.
[2] Xiu D. *Numerical Methods for Stochastic Computations. A Spectral Method Approach*. Princeton University Press; 2010.
[3] Xiu D, Karniadakis GE. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computation*. 2002; 24 (2): 619–644.
[4] Soong TT. *Random Differential Equations in Science and Engineering*. Elsevier; 1973.
[5] Strand JL. Random ordinary differential equations. *Journal of Differential Equations*. 1970; 7 (3): 538–553.
[6] Chen-Charpentier BM, Stanescu D. Epidemic models with random coefficients. *Mathematical and Computer Modelling*. 2010; 52 (7-8): 1004–1010.
[7] Stanescu D, Chen-Charpentier BM. Random coefficient differential equation models for bacterial growth. *Mathematical and Computer Modelling*. 2009; 50 (5-6): 885–895.
[8] Santonja F, Chen-Charpentier BM. Uncertainty quantification in simulations of epidemics using polynomial chaos. *Computational and Mathematical Methods in Medicine*. 2012; 2012.
[9] Lucor D, Su CH, Karniadakis GE. Generalized polynomial chaos and random oscillators. *International Journal for Numerical Methods in Engineering*. 2004; 60 (3): 571–596.
[10] Xiu D, Karniadakis GE. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*. 2002; 191 (43): 4927–4948.
[11] Villegas M, Augustin F, Gilg A, Hmaidi A, Wever U. Application of the polynomial chaos expansion to the simulation of chemical reactors with uncertainties. *Mathematics and Computers in Simulation*. 2012; 82 (12): 805–817.
[12] Ernst OG, Mugler A, Starkloff HJ, Ullmann E. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*. 2012; 46 (2): 317–339.
[13] Shi W, Zhang C. Error analysis of generalized polynomial chaos for nonlinear random ordinary differential equations. *Applied Numerical Mathematics*. 2012; 62 (12): 1954–1964.
[14] Calatayud J, Cortés JC, Jornet M. On the convergence of adaptive gPC for non-linear random difference equations: Theoretical analysis and some practical recommendations. *Journal of Nonlinear Sciences and Applications*. 2018; 11(9): 1077–1084.
[15] Chen-Charpentier BM, Cortés JC, Licea JA, Romero JV, Roselló MD, Santonja FJ, Villanueva RJ. Constructing adaptive generalized polynomial chaos method to measure the uncertainty in continuous models: A computational approach. *Mathematics and Computers in Simulation*. 2015; 109: 113–129.
[16] Cortés JC, Romero JV, Roselló MD, Villanueva RJ. Improving adaptive generalized polynomial chaos method to solve nonlinear random differential equations by the random variable transformation technique. *Communications in Nonlinear Science and Numerical Simulations*. 2017; 50: 1–15.
[17] Cortés JC, Romero JV, Roselló MD, Santonja FJ, Villanueva RJ. (2013). Solving continuous models with dependent uncertainty: a computational approach. *Abstract and Applied Analysis*. 2013; 2013.
[18] Giraud L, Langou J, Rozloznik M. The loss of orthogonality in the Gram-Schmidt orthogonalization process. *Computers & Mathematics with Applications*. 2005; 50 (7): 1069–1075.
[19] Shabat BV. *Introduction to Complex Analysis. II: Functions of Several Variables*. Translations of Mathematical Monographs, 110, American Mathematical Society, Providence, RI; 1992.
[20] Horn RA, Johnson CR. *Matrix Analysis*. Cambridge University Press; 1985.
[21] Neumaier A. Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM review*. 1998; 40 (3): 636–666.
[22] Elaydi SN. *An introduction to difference equations*. 3rd edition, Springer; 2005.
[23] Hilbert D. Ein Beitrag zur Theorie des Legendre'schen Polynoms. *Acta mathematica*. 1894; 18 (1): 155–159.
[24] Choi MD. Tricks or treats with the Hilbert matrix. *The American Mathematical Monthly*. 1983; 90 (5): 301–312.
[25] Todd J. The condition of the finite segments of the Hilbert matrix. *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*. 1954; 39: 109-116.
[26] Taussky O. A remark concerning the characteristic roots of the finite segments of the Hilbert matrix. *The Quarterly Journal of Mathematics*. 1949; (1): 80–83.
[27] Otte P. Upper bounds for the spectral radius of the $n \times n$ Hilbert matrix. *Pacific Journal of Mathematics*. 2005; 219 (2): 323–331.
[28] Casella G, Berger RL. *Statistical Inference*. Brooks/Cole; 2008.
[29] McShane EJ. Extension of range of functions. *Bulletin of the American Mathematical Society*. 1934; 40 (12): 837–842.
[30] Wolfram Research, Inc. Mathematica. Version 11.2, Champaign, Illinois (2017).