East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.150616.071216a

Vol. 7, No. 2, pp. 306-324
May 2017

Application of gPCRK Methods to Nonlinear Random Differential Equations with Piecewise Constant Argument

Chengjian Zhang[1] and Wenjie Shi[1,2,*]

[1] *School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China.*
[2] *School of Mathematics and Computer Science, Wuhan Textile University, Wuhan 430073, China.*

Received 15 June 2016; Accepted (in revised version) 7 December 2016.

Abstract. We propose a class of numerical methods for solving nonlinear random differential equations with piecewise constant argument, called gPCRK methods as they combine generalised polynomial chaos with Runge-Kutta methods. An error analysis is presented involving the error arising from a finite-dimensional noise assumption, the projection error, the aliasing error and the discretisation error. A numerical example is given to illustrate the effectiveness of this approach.

AMS subject classifications: 65L70, 65C30

Key words: Random differential equations, piecewise constant argument, generalised polynomial chaos, finite-dimensional noise, error analysis.

1. Introduction

In recent years, delay differential equations (DDEs)

$$\begin{cases} x'(t) = f\big(t, x(t), x(\lfloor t \rfloor)\big), \quad t \in [0, T], \\ x(0) = x_0 \end{cases} \tag{1.1}$$

with piecewise constant argument $\lfloor t \rfloor$ (where $\lfloor \cdot \rfloor$ denotes the greatest-integer function) have been considered quite widely — e.g. see Refs. [6, 7, 12–14] and references therein. In particular, the solvability, stability, oscillations, dissipativity and other related properties of the analytical and numerical solutions were investigated. However, previous research

*Corresponding author. *Email addresses:* cjzhang@mail.hust.edu.cn (C. Zhang), wenjieshihust@gmail.com (W. Shi)

 ©2017 Global-Science Press

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

has focused on deterministic equations, and in this article we consider random differential equations with piecewise constant argument (RDEPCA) as follows:

$$\begin{cases} u'(t,\omega) = f\big(t,\gamma(t,\omega),u(t,\omega),u(\lfloor t\rfloor,\omega)\big), \quad t\in[0,T],\ \omega\in\Omega, \\ u(0)=\varphi_0 . \end{cases} \tag{1.2}$$

Here $(\Omega,\mathscr{F},\mathscr{P})$ denotes a complete probability space with a sample space Ω, a σ-algebra $\mathscr{F}\subset 2^{\Omega}$ and a probability measure $\mathscr{P}$; f is a random function in the random field $\gamma(t,\omega)$ with an assigned distribution; and the initial value φ_0 is given. For convenience, $u(t,\omega)$ will be denoted by $u(t)$ hereafter. The regularity of the solution of the RDEPCA (1.2) can be discussed as in Ref. [18].

We combine generalised polynomial chaos (gPC) with Runge-Kutta (RK) methods, producing so-called gPCRK methods to solve the RDEPCA (1.2). The gPC method of Xiu *et al.* [17] has been applied successfully to some random models involving ordinary and partial differential equations [9,10,16]) — cf. also the review article [8,11,15]. A gPC method can change a random model into a deterministic one, so one can transform the RDEPCA (1.2) into a set of deterministic DDEs and then obtain its solution by applying a suitable RK method. In Section 2, we introduce a finite-dimensional noise assumption and gPC expansion, and then construct our class numerical methods for the RDEPCA (1.2) by combining the gPC with Runge-Kutta methods. In Section 3, the error analysis involves the error arising from the finite-dimensional noise assumption, the projection error, the aliasing error and the discretisation error. In Section 4, we give a numerical example to illustrate the effectiveness of the method, followed by our conclusions in Section 5.

2. Construction of the Numerical Methods

In order to describe the solution of the RDEPCA (1.2) by a finite number of random variables, we introduce a so-called finite-dimensional noise assumption — e.g. see Babuska *et al.* [1]. Thus a random parameter $\gamma(t,\omega)$ can be expressed approximately as

$$\gamma(t,\omega)\approx\gamma\big(t,\xi_1(\omega),\xi_2(\omega),\cdots,\xi_N(\omega)\big),\ \ (t,\omega)\in[0,T]\times\Omega, \tag{2.1}$$

where $N\in\mathbb{Z}^+$, and $\{\xi_n\}_{n=1}^N$ are real-valued random variables with zero mean and unit variance.

Here we assume that the random variables $\{\xi_n\}_{n=1}^N$ are independent, so the random parameter $\gamma(t,\omega)$ admits a truncated series representation [4,16]:

$$\gamma(t,\omega)=\mathbb{E}\{\gamma\}+\sum_{n=1}^{\infty}\gamma_n(t)\xi_n(\omega)\approx\mathbb{E}\{\gamma\}+\sum_{n=1}^{N}\gamma_n(t)\xi_n(\omega)\triangleq\gamma_N(t,y) \tag{2.2}$$

where $y=(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_N(\omega))$, with a joint probability density

$$\rho(y)=\prod_{i=1}^{N}\rho(\xi_i),\quad \Gamma=\prod_{i=1}^{N}\Gamma_i\to\mathbb{R}^+ \ \text{ where } \Gamma_i=\xi_i(\Omega),$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

$\mathbb{E}\{\cdot\}$ the ensemble mean operator, and $\{\gamma_n(t)\}$ a set of deterministic functions. This follows via Karhunen-Loéve and Fourier expansions [4, 16].

A general second-order random process $X(\omega)$ can be expressed as [16]

$$\begin{aligned} X(\omega) =& c_0 I_0 + \sum_{i_1=1}^{\infty} c_{i_1} I_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} c_{i_1 i_2} I_2(\xi_{i_1}, \xi_{i_2}) \\ &+ \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} c_{i_1 i_2 i_3} I_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \cdots , \end{aligned} \tag{2.3}$$

where $I_n(\cdot)$ denotes the gPC of order n with respect to the random variables $\{\xi_i\}_{i=1}^{\infty}$, and the coefficients $c_{i_1 \cdots i_r}$ are deterministic constants. For convenience, we rewrite Eq. (2.3) as

$$X(\omega) = \sum_{j=0}^{\infty} \hat{c}_j \Psi_j(\xi_1, \cdots, \xi_N, \cdots) , \tag{2.4}$$

with a one-to-one correspondence between the coefficients and basis functions in Eqs. (2.3) and (2.4). Indeed, we assume that the basis $\{\Psi_n(\cdot)\}$ admits $\mathbb{E}\{\Psi_i \Psi_j\} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. An appropriate selection of the type of gPC often leads us to an effective gPC method with an optimal convergence rate [16, 17].

Letting $\gamma_N(t, y)$ be the first N-term expansion of $\gamma(t, \omega)$ and replacing $\gamma(t, \omega)$ in Eq. (1.2) by $\gamma_N(t, y)$, we obtain

$$\begin{cases} u_N'(t) = f\big(t, \gamma_N(t, y), u_N(t), u_N(\lfloor t \rfloor)\big), \quad t \in [0, T], \ y \in \Gamma , \\ u_N(0) = \varphi_0 , \end{cases} \tag{2.5}$$

and from Ref. [2] the solution $u_N(t)$ of this problem (2.5) can be approximated as follows:

$$u_N(t) \approx \sum_{i=0}^{M-1} \breve{u}_i(t) \Psi_i(y), \tag{2.6}$$

where $M = (N + p)!/N!p!$ with N and p the respective number of random variables in y and the highest order of the gPC basis $\{\Psi_i(y)\}_{i=0}^{M-1}$, and where a total degree polynomial space is employed. Substituting Eq. (2.6) into Eq. (2.5) we obtain a new random system, which when projected onto $\Psi_k(y)$ yields the following deterministic system of DDEs:

$$\begin{cases} \breve{u}_k'(t) = \mathbb{E}\left\{ \Psi_k(y) f\left(t, \gamma_N(t, y), \sum_{i=0}^{M-1} \breve{u}_i(t) \Psi_i(y) , \sum_{i=0}^{M-1} \breve{u}_i(\lfloor t \rfloor) \Psi_i(y) \right) \right\}, \quad t \in [0, T], \\ \breve{u}_k(0) = \mathbb{E}\Big\{ \Psi_k(y) \varphi_0 \Big\}, \quad k = 0, 1, \cdots, M-1 . \end{cases} \tag{2.7}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

Setting

$$z(t)=\big(\breve{u}_0(t),\breve{u}_1(t),\cdots,\breve{u}_{M-1}(t)\big)^T,\qquad \phi=(\varphi_0,0,\cdots,0)^T,$$
$$g_k\big(t,z(t),z(\lfloor t\rfloor)\big)=\mathbb{E}\left\{\Psi_k(y)f\left(t,\gamma_N(t,y),\sum_{i=0}^{M-1}\breve{u}_i(t)\Psi_i(y),\sum_{i=0}^{M-1}\breve{u}_i(\lfloor t\rfloor)\Psi_i(y)\right)\right\},$$
$$g\big(t,z(t),z(\lfloor t\rfloor)\big)=\Big(g_0\big(t,z(t),z(\lfloor t\rfloor)\big),g_1\big(t,z(t),z(\lfloor t\rfloor)\big),\cdots,g_{M-1}\big(t,z(t),z(\lfloor t\rfloor)\big)\Big)^T,$$

we can rewrite the system (2.7) as

$$\begin{cases} z'(t)=g\big(t,z(t),z(\lfloor t\rfloor)\big), & t\in[0,T],\\ z(0)=z_0\,. \end{cases} \tag{2.8}$$

The solution of (2.8) is a function $z(t)$ subject to the conditions (cf. Wiener [14]): $z(t)$ is continuous on $[0,T]$; the derivative $z'(t)$ exists at each point $t\in[0,T]$, with the possible exception of the points $\lfloor t\rfloor\in[0,T]$ where one-sided derivatives exist; and Eq. (2.8) is satisfied on each interval $[r,r+1)\subset[0,T]$ where $r\in\mathbb{Z}^+\cup\{0\}$.

Extended RK methods [6, 7, 12, 13] are a good option to solve the delay system (2.8). Thus

$$\begin{cases} Z_i^{(n)}=z_n+h\sum_{j=1}^{s}a_{ij}g\big(t_n+c_jh,Z_j^{(n)},\tilde{Z}_{n,j}\big), & i=1,2,\cdots,s,\\ z_{n+1}=z_n+h\sum_{j=1}^{s}b_jg\big(t_n+c_jh,Z_j^{(n)},\tilde{Z}_{n,j}\big), & n\ge 0, \end{cases} \tag{2.9}$$

where $t_n=nh$, the stepsize $h=1/m$ defines the mesh points for any given positive integer m, and $z_n,Z_j^{(n)},\tilde{Z}_{n,j}$ are the respective approximations to $z(t_n),z(t_n+c_jh),z(\lfloor t_n+c_jh\rfloor)$. These extended methods are obtained by adapting underlying RK methods for ODE, for which the computational effectiveness, stability, dissipativity and related properties are well known — cf. Refs. [6, 7, 12, 13] and references therein. For convenience, we write the underlying RK method for ODE as (A,b,c), with $A=(a_{ij})\in\mathbb{R}^{s\times s}$, $b=(b_1,\cdots,b_s)^T$ and $c=(c_1,\cdots,c_s)^T\in\mathbb{R}^s$, and always assume that $\sum_{i=1}^{s}b_i=1$ and $0\le c_1\le c_2\le\cdots\le c_s<1$. By combining the gPC approximation with extended RK methods, we have a class of numerical schemes that we call *gPCRK methods* to solve the RDEPCA (1.2).

3. Error Estimation for the gPCRK Methods

We restrict the error analysis for our gPCRK methods to the one-dimensional case of the RDEPCA (1.2) for simplicity. Let

$$\tilde{u}_i(t)=\mathbb{E}\{\Psi_i(y)u_N(t)\},\qquad \tilde{u}(t)\triangleq\sum_{i=0}^{M-1}\tilde{u}_i(t)\Psi_i(y),$$
$$\breve{u}(t)\triangleq\sum_{i=0}^{M-1}\breve{u}_i(t)\Psi_i(y),\qquad \bar{u}(t)\triangleq\sum_{i=0}^{M-1}\bar{u}_i(t)\Psi_i(y),$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

where $(\bar{u}_0(t), \bar{u}_1(t), \cdots, \bar{u}_{M-1}(t))^T$ is the numerical solution of Eq. (2.8), and in addition $\|f(t,\omega)\|_{L^2} = [\int_\Omega f^2(t,\omega)\mathrm{d}\mathscr{P}(\omega)]^{1/2}$ for all $t \in [0,T]$ denotes the L^2-norm of a random function $f(t,\omega)$, where $\mathscr{P}$ is a probability measure. Consequently, the error arising from the finite-dimensional noise assumption is $\|u(t)-u_N(t)\|_{L^2}^2$, the projection error is $\|u_N(t)-\tilde{u}(t)\|_{L^2}^2$, the aliasing error is $\|\tilde{u}(t)-\check{u}(t)\|_{L^2}^2$, and the discretisation error is $\|\check{u}(t)-\bar{u}(t)\|_{L^2}^2$. The global error of a gPCRK method involving these four errors is then estimated as

$$\begin{aligned}&\|u(t)-\bar{u}(t)\|_{L^2}^2\\ &\le 4\Big(\|u(t)-u_N(t)\|_{L^2}^2 + \|u_N(t)-\tilde{u}(t)\|_{L^2}^2 + \|\tilde{u}(t)-\check{u}(t)\|_{L^2}^2 + \|\check{u}(t)-\bar{u}(t)\|_{L^2}^2\Big).\end{aligned} \tag{3.1}$$

To analyse the individual errors on the right-hand of Eq. (3.1), we assume that for all $\gamma_1, \gamma_2, w_1, w_2, v_1, v_2$ the function f in the RDEPCA (1.2) satisfies the Lipschitz condition

$$\big\|f(t,\gamma_1,w_1,w_2)-f(t,\gamma_2,v_1,v_2)\big\|_{L_2} \le k(t)\left(\|\gamma_1-\gamma_2\|_{L^2} + \sum_{i=1}^{2}\|w_i-v_i\|_{L^2}\right), \tag{3.2}$$

where $k(t) > 0$ is a given function such that $\int_0^T k(s)\,ds < \infty$. On writing

$$c_i(t) = \begin{cases} 3\displaystyle\int_0^t k(s)\,ds \int_0^t k(s)\big\|\gamma(s,\omega)-\gamma_N(s,y)\big\|_{L^2}^2\,ds\,, \quad i=1\,,\\ c_1(t) + \displaystyle\sum_{l=1}^{i-2} 3c_l(l)e(l)\int_l^{l+1} k(s)\,ds\int_0^t k(s)\,ds\\ \qquad + 3c_{i-1}(i-1)e(i-1)\displaystyle\int_{i-1}^t k(s)\,ds\int_0^t k(s)\,ds\,, \quad i>1\,, \end{cases} \tag{3.3}$$

and

$$e(t) = \exp\left\{3\int_0^T k(s)\,ds\int_0^t k(s)\,ds\right\}, \tag{3.4}$$

we obtain an estimation of the error arising from the finite-dimensional noise assumption stated in the following theorem.

Theorem 3.1. *Assuming that the Lipschitz condition* (3.2) *holds, the error arising from the finite-dimensional noise assumption can be estimated from*

$$\big\|u(t)-u_N(t)\big\|_{L^2}^2 \le c_i(t)e(t)\,, \quad \forall t \in (i-1,i]\,. \tag{3.5}$$

Proof. From (1.2) and (2.5) we have

$$u(t) = \varphi_0 + \int_0^t f\big(s,\gamma(s,\omega),u(s),u(\lfloor s\rfloor)\big)\,ds\,, \quad t\in[0,T]\,,\ \omega\in\Omega\,, \tag{3.6}$$

and

$$u_N(t) = \varphi_0 + \int_0^t f\big(s,\gamma_N(s,y),u_N(s),u_N(\lfloor s\rfloor)\big)\,ds\,, \quad t\in[0,T]\,,\ y\in\Gamma\,. \tag{3.7}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

It follows from Eqs. (3.6) and (3.7) and the Cauchy-Schwarz inequality that

$$\begin{aligned}\|u(t)-u_N(t)\|_{L^2}^2 &= \Big\| \int_0^t \Big(f\big(s,\gamma(s,\omega),u(s),u(\lfloor s\rfloor)\big)-f\big(s,\gamma_N(s,y),u_N(s),u_N(\lfloor s\rfloor)\big)\Big)\,ds\Big\|_{L^2}^2 \\ &\le \int_\Omega \int_0^t k(s)\,ds \int_0^t \frac{1}{k(s)}\Big(f\big(s,\gamma(s,\omega),u(s),u(\lfloor s\rfloor)\big) \\ &\qquad -f\big(s,\gamma_N(s,y),u_N(s),u_N(\lfloor s\rfloor)\big)\Big)^2 ds\,d\mathscr{P}(\omega) \\ &= \int_0^t k(s)\,ds \int_0^t \frac{1}{k(s)}\Big\|f\big(s,\gamma(s,\omega),u(s),u(\lfloor s\rfloor)\big) \\ &\qquad -f\big(s,\gamma_N(s,y),u_N(s),u_N(\lfloor s\rfloor)\big)\Big\|_{L^2}^2 ds\,. \end{aligned} \tag{3.8}$$

Applying the Lipschitz condition (3.2) and the generalised mean inequality

$$\frac{|f_1|+|f_2|+|f_3|}{3} \le \sqrt{\frac{|f_1|^2+|f_2|^2+|f_3|^2}{3}}, \quad \forall f_1,f_2,f_3\in\mathbb{R}$$

to the inequality (3.8) yields

$$\|u(t)-u_N(t)\|_{L^2}^2 \le c_1(t)+3\int_0^t k(s)\,ds\int_0^t k(s)\Big(\big\|u(s)-u_N(s)\big\|_{L^2}^2+\big\|u(\lfloor s\rfloor)-u_N(\lfloor s\rfloor)\big\|_{L^2}^2\Big)ds. \tag{3.9}$$

We now proceed to prove the error estimation result (3.5) by mathematical induction. For $t\in[0,1)$ we have

$$\int_0^t k(s)\big\|u(\lfloor s\rfloor)-u_N(\lfloor s\rfloor)\big\|_{L^2}^2 ds = \int_0^t k(s)\big\|u(0)-u_N(0)\big\|_{L^2}^2 ds = 0\,, \tag{3.10}$$

and substituting Eq. (3.10) into the inequality (3.9) yields

$$\big\|u(t)-u_N(t)\big\|_{L^2}^2 \le c_1(t)+3\int_0^T k(s)\,ds\int_0^t k(s)\big\|u(s)-u_N(s)\big\|_{L^2}^2 ds\,. \tag{3.11}$$

Since $k(t)>0$ and $c_1(t)$ is non-decreasing, application of the Gronwall inequality leads to

$$\big\|u(t)-u_N(t)\big\|_{L^2}^2 \le c_1(t)e(t)\,,\quad \forall t\in[0,1)\,, \tag{3.12}$$

and because $\|u(t)-u_N(t)\|_{L^2}^2$ is continuous in t we consequently have that inequality (3.5) holds for $i=1$. Assuming (3.5) holds for $i: 1\le i\le n$, when $t\in(n,n+1)$ we have that

$$\begin{aligned}&\int_0^t k(s)\big\|u(\lfloor s\rfloor)-u_N(\lfloor s\rfloor)\big\|_{L^2}^2 ds \\ &=\sum_{l=0}^{n-1}\int_l^{l+1} k(s)\big\|u(l)-u_N(l)\big\|_{L^2}^2 ds+\int_n^t k(s)\big\|u(n)-u_N(n)\big\|_{L^2}^2 ds \\ &\le \sum_{l=1}^{n-1} c_l(l)e(l)\int_l^{l+1} k(s)\,ds + c_n(n)e(n)\int_n^t k(s)ds\,. \end{aligned} \tag{3.13}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

Substituting inequality (3.13) into inequality (3.9) gives

$$\left\|u(t)-u_N(t)\right\|_{L^2}^2 \le c_{n+1}(t)+3\int_0^T k(s)\,ds\int_0^t k(s)\left\|u(s)-u_N(s)\right\|_{L^2}^2 ds\,, \tag{3.14}$$

so from the Gronwall inequality and the continuity of $\|u(t)-u_N(t)\|_{L^2}^2$ with respect to t we conclude that inequality (3.5) holds for $i=n+1$. □

In deriving a simpler estimation of $\|u(t)-u_N(t)\|_{L^2}^2$, we introduce the following lemma.

Lemma 3.1. *Let i_0 be an integer such that $i_0-1<T\le i_0$, and*

$$\tilde{c}_i=\begin{cases} 3\int_0^T k(s)\,ds\int_0^T k(s)\|\gamma(s,\omega)-\gamma_N(s,y)\|_{L^2}^2\,ds\,, & i=1\,,\\ \tilde{c}_1+\sum_{l=1}^{i-1}3\tilde{c}_l e(l)\int_l^{l+1}k(s)\,ds\int_0^T k(s)\,ds\,, & 1<i\le i_0-1\,,\\ \tilde{c}_1+\sum_{l=1}^{i_0-2}3\tilde{c}_l e(l)\int_l^{l+1}k(s)\,ds\int_0^T k(s)\,ds+3\tilde{c}_{i_0-1}e(i_0-1)\int_{i_0-1}^T k(s)\,ds\int_0^T k(s)\,ds\,, & i=i_0\,, \end{cases}$$

Then we have $c_1(t)\le\tilde{c}_1$ on $[0,T]$, $c_i(t)\le\tilde{c}_i$ on $(i-1,i]$ for $i=2,3,\cdots,i_0-1$, and $c_{i_0}(t)\le\tilde{c}_{i_0}$ on $(i_0-1,T]$, and that $\tilde{c}_{i_0}$ can be rewritten as

$$\tilde{c}_{i_0}=\left(1+3e(i_0-1)\int_{i_0-1}^T k(s)\,ds\int_0^T k(s)\,ds\right)\prod_{i=2}^{i_0-1}\left(1+3e(i-1)\int_{i-1}^i k(s)\,ds\int_0^T k(s)\,ds\right)\tilde{c}_1. \tag{3.15}$$

Proof. It is easy to prove that $c_i(t)\le\tilde{c}_i$ for $i\ge1$ by induction, so we need only prove Eq. (3.15). Indeed, it follows from the definition of $\tilde{c}_i$ that

$$\tilde{c}_i=\Big(1+3e(i-1)\int_{i-1}^i k(s)\,ds\int_0^T k(s)\,ds\Big)\tilde{c}_{i-1}\,,\quad \forall i:2\le i<i_0\,,$$

and

$$\tilde{c}_{i_0}=\Big(1+3e(i_0-1)\int_{i_0-1}^T k(s)\,ds\int_0^T k(s)\,ds\Big)\tilde{c}_{i_0-1}\,,$$

so Eq. (3.15) follows immediately from the above recursions. □

A combination of Theorem 3.1 and Lemma 3.1 produces the following conclusion.

Theorem 3.2. *Assume that the Lipschitz condition* (3.2) *holds. Then the error arising from the finite-dimensional noise assumption can be estimated from*

$$\left\|u(t)-u_N(t)\right\|_{L^2}^2\le\tilde{c}_{i_0}e(t),\quad t\in[0,T]. \tag{3.16}$$

Remark 3.1. Using (3.3) and Lemma 3.1, we can obtain $\lim_{N\to\infty}c_i(t)=0$ for $i\ge1$ and $\lim_{N\to\infty}\tilde{c}_{i_0}=0$, so the error arising from the finite-dimensional noise assumption tends to zero as N increases.

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

For the projection error, there is the following result.

Lemma 3.2 (cf. Ref. [16]). *Let* $\tilde{v} = \sum_{i=0}^{M-1} \mathbb{E}\{v\Psi_i\}\Psi_i$ *where* $v \in L^2(\Omega)$. *Then the following convergence property holds:*

$$\lim_{M\to\infty} \|v - \tilde{v}\|_{L^2(\Omega)} = 0. \tag{3.17}$$

In order to estimate the aliasing error, we define

$$\tilde{e}(t) = \exp\left\{4\int_0^T k(s)\,ds \int_0^t k(s)\,ds\right\},$$

$$w_i(t) = \begin{cases} 4\int_0^t k(s)\,ds \int_0^t k(s)\big(\|u_N(s) - \tilde{u}(s)\|_{L^2}^2 + \|u_N(\lfloor s\rfloor) - \tilde{u}(\lfloor s\rfloor)\|_{L^2}^2\big)\,ds, & i = 1, \\ w_1(t) + \sum_{l=1}^{i-2} 4w_l(l)\tilde{e}(l)\int_l^{l+1} k(s)\,ds \int_0^t k(s)\,ds \\ \quad + 4w_{i-1}(i-1)\tilde{e}(i-1)\int_{i-1}^t k(s)ds \int_0^t k(s)\,ds, & i \geq 2. \end{cases}$$

Theorem 3.3. *Assume that the Lipschitz condition* (3.2) *holds. Then the aliasing error can be estimated from*

$$\left\|\tilde{u}(t) - \breve{u}(t)\right\|_{L^2}^2 \leq w_i(t)\tilde{e}(t), \quad \forall t \in (i-1, i]. \tag{3.18}$$

Proof. It follows from (2.5) and (3.17) that

$$\sum_{i=0}^{\infty} \tilde{u}_i'(t)\Psi_i(y) = f\left(t, \gamma_N(t,y), \sum_{i=0}^{\infty} \tilde{u}_i(t)\Psi_i(y), \sum_{i=0}^{\infty} \tilde{u}_i(\lfloor t\rfloor)\Psi_i(y)\right), \tag{3.19}$$

so projecting onto each basis $\Psi_k(y)$ we have

$$\tilde{u}_k'(t) = \mathbb{E}\left\{f\left(t, \gamma_N(t,y), \tilde{u}(t) + \sum_{i=M}^{\infty} \tilde{u}_i(t)\Psi_i(y), \tilde{u}(\lfloor t\rfloor) + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor t\rfloor)\Psi_i(y)\right)\Psi_k(y)\right\}, \quad k = 0, 1, \cdots, M-1, \tag{3.20}$$

and on integrating both sides from 0 to t we then have

$$\tilde{u}_k(t) = \mathbb{E}\{\Psi_k(y)\varphi_0\} + \int_0^t \mathbb{E}\left\{f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s\rfloor) + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s\rfloor)\Psi_i(y)\right)\Psi_k(y)\right\} ds, \quad k = 0, 1, \cdots, M-1. \tag{3.21}$$

Now the exact solution of the system (2.7) can be written as

$$\breve{u}_k(t) = \mathbb{E}\{\Psi_k(y)\varphi_0\} + \int_0^t \mathbb{E}\left\{f\big(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s\rfloor)\big)\Psi_k(y)\right\} ds, \quad k = 0, 1, \cdots, M-1, \tag{3.22}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

so for $k = 0, 1, \cdots, M-1$ from Eqs. (3.21) and (3.22) we have that

$$\tilde{u}_k(t) - \breve{u}_k(t) = \int_0^t \mathbb{E}\left\{\left(f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s \rfloor)\right.\right.\right. \\ \left.\left.\left. + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s \rfloor)\Psi_i(y)\right) - f\left(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s \rfloor)\right)\right)\Psi_k(y)\right\} ds\,, \tag{3.23}$$

and invoking the Cauchy-Schwarz inequality we consequently obtain

$$\begin{aligned} &\sum_{k=0}^{M-1} \left(\tilde{u}_k(t) - \breve{u}_k(t)\right)^2 \\ &= \int_0^t \mathbb{E}\left\{\left(f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s \rfloor) + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s \rfloor)\Psi_i(y)\right)\right.\right. \\ &\quad \left.\left. - f\left(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s \rfloor)\right)\right) \sum_{k=0}^{M-1} \left(\tilde{u}_k(t) - \breve{u}_k(t)\right)\Psi_k(y)\right\} ds \\ &\le \int_0^t \left\| f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s \rfloor) + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s \rfloor)\Psi_i(y)\right)\right. \\ &\quad \left. - f\left(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s \rfloor)\right)\right\|_{L^2} \left\| \sum_{k=0}^{M-1} \left(\tilde{u}_k(t) - \breve{u}_k(t)\right)\Psi_k(y)\right\|_{L^2} ds\,. \end{aligned} \tag{3.24}$$

In addition, according to the definitions of $\tilde{u}(t)$ and $\breve{u}(t)$ we have

$$\left\|\tilde{u}(t) - \breve{u}(t)\right\|_{L^2}^2 = \left\| \sum_{i=0}^{M-1} \left(\tilde{u}_i(t) - \breve{u}_i(t)\right)\Psi_i(y)\right\|_{L^2}^2 = \sum_{i=0}^{M-1} \left(\tilde{u}_i(t) - \breve{u}_i(t)\right)^2, \tag{3.25}$$

hence from inequality (3.24) and Eq. (3.25) we have

$$\left\|\tilde{u}(t) - \breve{u}(t)\right\|_{L^2} \le \int_0^t \left\| f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s \rfloor) + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s \rfloor)\Psi_i(y)\right)\right. \\ \left. - f\left(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s \rfloor)\right)\right\|_{L^2} ds\,, \tag{3.26}$$

and hence on applying the Cauchy-Schwarz inequality that

$$\left\|\tilde{u}(t) - \breve{u}(t)\right\|_{L^2}^2 \le \int_0^t k(s)\, ds \int_0^t \frac{1}{k(s)} \left\| f\left(s, \gamma_N(s,y), \tilde{u}(s) + \sum_{i=M}^{\infty} \tilde{u}_i(s)\Psi_i(y), \tilde{u}(\lfloor s \rfloor)\right.\right. \\ \left.\left. + \sum_{i=M}^{\infty} \tilde{u}_i(\lfloor s \rfloor)\Psi_i(y)\right) - f\left(s, \gamma_N(s,y), \breve{u}(s), \breve{u}(\lfloor s \rfloor)\right)\right\|_{L^2}^2 ds\,. \tag{3.27}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

Further, from the Lipschitz condition (3.2) and the Parallelogram identity we obtain

$$\begin{aligned}
\left\|\tilde{u}(t)-\breve{u}(t)\right\|_{L^2}^2 \le & \int_0^t k(s)\,ds \int_0^t k(s)\Big(\left\|\tilde{u}(s)+\sum_{i=M}^{\infty}\tilde{u}_i(s)\Psi_i(y)-\breve{u}(s)\right\|_{L^2} \\
& + \left\|\tilde{u}(\lfloor s\rfloor)+\sum_{i=M}^{\infty}\tilde{u}_i(\lfloor s\rfloor)\Psi_i(y)-\breve{u}(\lfloor s\rfloor)\right\|_{L^2}\Big)^2 ds \\
\le & 2\int_0^t k(s)\,ds \int_0^t k(s)\Big(\left\|\tilde{u}(s)+\sum_{i=M}^{\infty}\tilde{u}_i(s)\Psi_i(y)-\breve{u}(s)\right\|_{L^2}^2 \\
& + \left\|\tilde{u}(\lfloor s\rfloor)+\sum_{i=M}^{\infty}\tilde{u}_i(\lfloor s\rfloor)\Psi_i(y)-\breve{u}(\lfloor s\rfloor)\right\|_{L^2}^2\Big) ds \\
\le & 4\int_0^t k(s)\,ds \int_0^t k(s)\Big(\left\|\tilde{u}(s)-\breve{u}(s)\right\|_{L^2}^2 + \left\|\sum_{i=M}^{\infty}\tilde{u}_i(s)\Psi_i(y)\right\|_{L^2}^2 \\
& + \left\|\tilde{u}(\lfloor s\rfloor)-\breve{u}(\lfloor s\rfloor)\right\|_{L^2}^2 + \left\|\sum_{i=M}^{\infty}\tilde{u}_i(\lfloor s\rfloor)\Psi_i(y)\right\|_{L^2}^2\Big) ds\,. \qquad (3.28)
\end{aligned}$$

Next, we prove inequality (3.18) by induction. When $t \in [0,1)$, we have

$$\int_0^t k(s)\left\|\tilde{u}(\lfloor s\rfloor)-\breve{u}(\lfloor s\rfloor)\right\|_{L^2}^2 ds = \int_0^t k(s)\left\|\tilde{u}(0)-\breve{u}(0)\right\|_{L^2}^2 ds = 0\,, \qquad (3.29)$$

so inequality (3.28) becomes

$$\left\|\tilde{u}(t)-\breve{u}(t)\right\|_{L^2}^2 \le w_1(t) + 4\int_0^T k(s)\,ds \int_0^t k(s)\left\|\tilde{u}(s)-\breve{u}(s)\right\|_{L^2}^2 ds\,, \qquad (3.30)$$

and on applying the Gronwall inequality we consequently have

$$\left\|\tilde{u}(t)-\breve{u}(t)\right\|_{L^2}^2 \le w_1(t)\tilde{e}(t), \quad \forall t \in [0,1)\,. \qquad (3.31)$$

Since $\|\tilde{u}(t)-\breve{u}(t)\|_{L^2}^2$ is continuous in t, the inequality (3.18) holds for $i=1$. Assume that (3.18) holds for $i: 1 \le i \le n$. When $t \in (n, n+1)$, we have

$$\begin{aligned}
\int_0^t k(s)\left\|\tilde{u}(\lfloor s\rfloor)-\breve{u}(\lfloor s\rfloor)\right\|_{L^2}^2 ds &= \sum_{l=1}^{n-1}\int_l^{l+1} k(s)\left\|\tilde{u}(l)-\breve{u}(l)\right\|_{L^2}^2 ds + \int_n^t k(s)\left\|\tilde{u}(n)-\breve{u}(n)\right\|_{L^2}^2 ds \\
&\le \sum_{l=1}^{n-1} w_l(l)\tilde{e}(l)\int_l^{l+1} k(s)\,ds + w_n(n)\tilde{e}(n)\int_n^t k(s)\,ds\,. \qquad (3.32)
\end{aligned}$$

Inserting (3.32) into (3.28) gives

$$\left\|\tilde{u}(t)-\breve{u}(t)\right\|_{L^2}^2 \le w_{n+1}(t) + 4\int_0^T k(s)\,ds \int_0^t k(s)\left\|\tilde{u}(s)-\breve{u}(s)\right\|_{L^2}^2 ds\,, \quad \forall t \in (n, n+1)\,, \qquad (3.33)$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

and again using the Gronwall inequality and the continuity of $\|\tilde{u}(t)-\breve{u}(t)\|_{L^2}^2$ for t we conclude that (3.18) holds for $i=n+1$. □

To obtain a simpler estimate of $\|\tilde{u}(t)-\breve{u}(t)\|_{L^2}^2$, we introduce the following lemma.

Lemma 3.3. *Let i_0 be an integer with $i_0-1<T\le i_0$, and*

$$\tilde{w}_i=\begin{cases} 4\int_0^T k(s)\,ds\int_0^T k(s)\big(\|u_N(s)-\tilde{u}(s)\|_{L^2}^2+\|u_N(\lfloor s\rfloor)-\tilde{u}(\lfloor s\rfloor)\|_{L^2}^2\big)\,ds\,, & i=1\,,\\ \tilde{w}_1+\sum_{l=1}^{i-1}4\tilde{w}_l\tilde{e}(l)\int_l^{l+1}k(s)ds\int_0^T k(s)\,ds\,, & 2\le i<i_0\,,\\ \tilde{w}_1+\sum_{l=1}^{i_0-2}4\tilde{w}_l\tilde{e}(l)\int_l^{l+1}k(s)\,ds\int_0^T k(s)\,ds\\ \quad +4\tilde{w}_{i_0-1}\tilde{e}(i_0-1)\int_{i_0-1}^T k(s)ds\int_0^T k(s)\,ds\,, & i=i_0\,. \end{cases} \tag{3.34}$$

Then $w_1(t)\le\tilde{w}_1$ on $[0,T]$, $w_i(t)\le\tilde{w}_i$ on $(i-1,i]$ for $i=2,3,\cdots,i_0-1$, $w_{i_0}(t)\le\tilde{w}_{i_0}$ on $(i_0-1,T]$, and

$$\tilde{w}_{i_0}=\left(1+4\tilde{e}(i_0-1)\int_{i_0-1}^T k(s)ds\int_0^T k(s)\,ds\right)\prod_{i=2}^{i_0-1}\left(1+4\tilde{e}(i-1)\int_{i-1}^i k(s)\,ds\int_0^T k(s)\,ds\right)\tilde{w}_1\,. \tag{3.35}$$

Proof. It follows from (3.34) that

$$\tilde{w}_i=\left(1+4\tilde{e}(i-1)\int_{i-1}^i k(s)\,ds\int_0^T k(s)\,ds\right)\tilde{w}_{i-1}\,,\quad \forall i:2\le i<i_0\,,$$

and

$$\tilde{w}_{i_0}=\left(1+4\tilde{e}(i_0-1)\int_{i_0-1}^T k(s)\,ds\int_0^T k(s)\,ds\right)\tilde{w}_{i_0-1}\,,$$

and the result follows immediately by induction. □

By combining Theorem 3.3 and Lemma 3.3, the simpler estimate of the aliasing error follows.

Theorem 3.4. *Assume that the Lipschitz condition* (3.2) *holds. Then the aliasing error can be estimated from*

$$\big\|\tilde{u}(t)-\breve{u}(t)\big\|_{L^2}^2\le\tilde{w}_{i_0}\tilde{e}(t)\,,\quad \forall t\in[0,T]\,. \tag{3.36}$$

Remark 3.2. Using Lemma 3.2 and the definition of $w_i(t)$ we obtain $\lim_{M\to\infty}w_i(t)=0$ for $i\ge1$, and from Lemma 3.3 we obtain $\lim_{M\to\infty}\tilde{w}_{i_0}=0$, hence the aliasing error tends to zero as M increases.

For the discretisation error, we make two assumptions:

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

(I) except at the points $\tilde{t} \in \{\tilde{t} : \ \tilde{t} = \lfloor t \rfloor, \ t \in [0,T]\}$, the solution $z(t)$ of (2.8) is sufficiently smooth on $[0,T]$; and

(II) the function $g : [0,T] \times \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^M$ in (2.8) satisfies the Lipchitz condition

$$\big\|g(t,\iota_1,\vartheta_1) - g(t,\iota_2,\vartheta_2)\big\| \le \kappa(t)\big(\|\iota_1-\iota_2\| + \|\vartheta_1-\vartheta_2\|\big), \quad \forall t \in [0,T], \iota_1,\iota_2,\vartheta_1,\vartheta_2 \in \mathbb{R}^M, \tag{3.37}$$

where $\|\cdot\|$ is a norm induced by the standard inner product on $\mathbb{R}^M$, and $\kappa(t) > 0$ is a given function with $\int_0^T \kappa(s)\,ds < \infty$.

Theorem 3.5. *Assume that the conditions* (I)-(II) *hold and the underlying RK method has local order q. Then the corresponding extended RK method* (2.9) *has convergent order q.*

Proof. We prove this theorem by mathematical induction. When $t \in [0,1)$, we have $\lfloor t \rfloor = 0$ such that (2.8) on $[0,1)$ becomes an ODE system. Since the underlying RK method has local order q, from Theorem 3.6 in Ref. [5] the extended RK method (2.9) satisfies

$$\big\|z_n - z(t_n)\big\| \le C_1 h^q, \quad 1 \le n \le m, \tag{3.38}$$

where $C_1 > 0$ is a constant, the stepsize $h = 1/m$ ($m \in \mathbb{Z}^+$), and the continuity of $z(t)$ is used for the case $n = m$.

Let us assume that when $t \in [0,r]$, where r is a positive integer with $1 \le r < T$, there exists a constant $C_2 > 0$ such that

$$\big\|z_n - z(t_n)\big\| \le C_2 h^q, \quad 1 \le n \le rm. \tag{3.39}$$

When $t \in (r, r+1)$, we have $\lfloor t \rfloor = r$ and hence (2.8) on $(r, r+1)$ becomes an ODE system

$$\begin{cases} z'(t) = g\big(t, z(t), z(r)\big), & t \in (r, r+1), \\ z(r) \text{ given}. \end{cases} \tag{3.40}$$

In the actual computation $z(r)$ is approximated by z_{rm} which is obtained by an extended RK method. Thus to solve (3.40), we solve the ODE system

$$\begin{cases} z_*'(t) = g\big(t, z_*(t), z_{rm}\big), & t \in (r, r+1), \\ z_*(r) = z_{rm}. \end{cases} \tag{3.41}$$

Since the underlying RK method has local order q, there exists a constant $C_3 > 0$ such that

$$\big\|z_n - z_*(t_n)\big\| \le C_3 h^q, \quad rm < n \le (r+1)m, \tag{3.42}$$

where the case $n = (r+1)m$ is proved given the continuity of $z_*(t)$. Moreover, it follows from (3.40) and (3.41) that

$$z(t) - z_*(t) = z(r) - z_{rm} + \int_r^t \Big(g\big(s, z(s), z(r)\big) - g\big(s, z_*(s), z_{rm}\big)\Big)\,ds, \tag{3.43}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

hence

$$\left\|z(t)-z_*(t)\right\| \le \left\|z(r)-z_{rm}\right\| + \left\| \int_r^t \left(g\big(s,z(s),z(r)\big) - g\big(s,z_*(s),z_{rm}\big)\right) ds \right\| . \tag{3.44}$$

From the Cauchy-Schwarz inequality we have

$$\begin{aligned}
&\left\| \int_r^t \left(g\big(s,z(s),z(r)\big) - g\big(s,z_*(s),z_{rm}\big)\right) ds \right\|^2 \\
&= \sum_{i=0}^{M-1} \left(\int_r^t \left(g_i\big(s,z(s),z(r)\big) - g_i\big(s,z_*(s),z_{rm}\big)\right) ds \right)^2 \\
&\le \sum_{i=0}^{M-1} \int_r^t \kappa(s)\, ds \int_r^t \frac{1}{\kappa(s)} \Big(g_i\big(s,z(s),z(r)\big) - g_i\big(s,z_*(s),z_{rm}\big)\Big)^2 ds \\
&= \int_r^t \kappa(s)\, ds \int_r^t \frac{1}{\kappa(s)} \left\| g\big(s,z(s),z(r)\big) - g\big(s,z_*(s),z_{rm}\big)\right\|^2 ds ,
\end{aligned} \tag{3.45}$$

and consequently from inequality (3.44) that

$$\begin{aligned}
&\left\|z(t)-z_*(t)\right\|^2 \\
&\le 2\left\|z(r)-z_{rm}\right\|^2 + 2\left\| \int_r^t \left(g\big(s,z(s),z(r)\big) - g\big(s,z_*(s),z_{rm}\big)\right) ds \right\|^2 \\
&\le 2\|z(r)-z_{rm}\|^2 + 2\int_r^t \kappa(s)\, ds \int_r^t \frac{1}{\kappa(s)} \left\| g(s,z(s),z(r)) - g(s,z_*(s),z_{rm})\right\|^2 ds .
\end{aligned} \tag{3.46}$$

Under the condition (3.37), this implies that

$$\begin{aligned}
&\left\|z(t)-z_*(t)\right\|^2 \\
&\le 2\left\|z(r)-z_{rm}\right\|^2 + 2\int_r^t \kappa(s)\, ds \int_r^t \kappa(s) \big(\|z(s)-z_*(s)\| + \|z(r)-z_{rm}\|\big)^2 ds \\
&\le \left(2 + 4\left(\int_r^{r+1} \kappa(s)\, ds \right)^2 \right) \left\|z(r)-z_{rm}\right\|^2 + 4\int_r^{r+1} \kappa(s)\, ds \int_r^t \kappa(s) \left\|z(s)-z_*(s)\right\|^2 ds ,
\end{aligned} \tag{3.47}$$

and hence on applying the Gronwall inequality we have

$$\left\|z(t)-z_*(t)\right\|^2 \le \left(2 + 4\left(\int_r^{r+1} \kappa(s)\, ds \right)^2 \right) \exp\left\{ 4\int_r^{r+1} \kappa(s)\, ds \int_r^t \kappa(s)\, ds \right\} \left\|z(r)-z_{rm}\right\|^2 . \tag{3.48}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

From the induction assumption (3.39) we also have $\|z(r)-z_{rm}\|^2 \le C_2^2 h^{2q}$, so inequality (3.48) implies that there exists a constant $C_4 > 0$ such that

$$\left\|z(t)-z_*(t)\right\| \le C_4 h^q\,, \quad t \in (r, r+1]\,, \tag{3.49}$$

where the continuities of $z(t)$ and $z_*(t)$ have been used for the case $t = r+1$. Combining inequalities (3.42) and (3.49), we conclude that there exists a constant $C_5 > 0$ such that

$$\left\|z_n - z(t_n)\right\| \le \left\|z_n - z_*(t_n)\right\| + \left\|z_*(t_n) - z(t_n)\right\| \le C_5 h^q, \quad rm < n \le (r+1)m \tag{3.50}$$

to complete the proof. □

Theorem 3.5 shows that there exists a constant $C_0 > 0$ such that the discretisation error may be estimated from

$$\left\|\check{u}(t_n)-\bar{u}(t_n)\right\|_{L^2}^2 = \left\|z_n - z(t_n)\right\|^2 \le C_0 h^{2q}\,, \quad n > 0\,, \tag{3.51}$$

whenever the conditions (I)-(II) hold and the underlying RK method has local order q.

Several ways to estimate the projection error have been considered — cf. for example Ref. [3], and particularly Ref. [16] where the expression given for estimating the projection error is

$$\left\|u_N(t)-\tilde{u}(t)\right\|_{L^2}^2 = \inf_{\psi(t,\cdot)\in\mathbb{P}_p^N(y)} \left\|u_N(t)-\psi(t,y)\right\|_{L^2}^2\,, \quad \forall u_N(t) \in L_\rho^2(\Gamma)\,, \tag{3.52}$$

involving

$$\mathbb{P}_p^N(y) = \left\{\psi : \Gamma \to \mathbb{R} \middle| \psi(y) = \sum_{|i|=i_1+i_2+\cdots+i_N\le p} c_i \Psi_i(y)\right\}, \quad y = (\xi_1, \xi_2, \cdots, \xi_N)\,,$$

$$L_\rho^2(\Gamma) = \left\{\varsigma : \Gamma \to \mathbb{R} \;\middle|\; \mathbb{E}(\varsigma^2(y)) = \int_\Gamma \varsigma^2(y)\rho(y)dy < \infty\right\}.$$

Consequently, we need not discuss the estimation of the projection error further here.

In summary, the error can be estimated by considering the error arising from the finite-dimensional noise assumption, the projection error, the aliasing error and the discretisation error. Improved error estimation for our gPCRK methods will probably require some more powerful analytical tools.

4. Numerical Example

In this section, we consider the following nonlinear RDEPCA to illustrate the computational effectiveness of the gPCRK approach.

Example 4.1.

$$\begin{cases} u'(t) = -\gamma(t,\omega)u(t) + u^2(\lfloor t \rfloor) - \exp\left\{-2\int_0^{\lfloor t \rfloor}\gamma(s,\omega)\,ds\right\},\ t \in [0,10],\ \omega \in \Omega\,, \\ u(0) = 1\,, \end{cases}$$

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

where $\gamma(t,\omega)$ is a Gaussian or uniform random field with mean-value 1, standard deviation 0.15, and Gaussian correlation function $\exp\{-4(t_1-t_2)^2/25\}$ with $t_1,t_2\in[0,10]$. The exact solution is $u(t)=\exp\{-\int_0^t \gamma(s,\omega)\,ds\}$.

The truncation term $\gamma_N(t,y)$ of $\gamma(t,y)$ can be obtained by the Fourier expansion with N terms, determined by $\varepsilon_\gamma=[\int_0^{10}\sigma_1(t)dt]/[\int_0^{10}\sigma_2(t)dt]$ where

$$\sigma_1(t)=\mathbb{E}\left\{\left(\gamma_N(t,y)-\mathbb{E}\{\gamma_N(t,y)\}\right)^2\right\},\qquad \sigma_2(t)=\mathbb{E}\left\{\left(\gamma(t,\omega)-\mathbb{E}\{\gamma(t,\omega)\}\right)^2\right\}.$$

For example, if we take $\varepsilon_\gamma=84.05\%$ or $\varepsilon_\gamma=95.11\%$, then we have the corresponding cases $N=5$ and $N=7$. When $\gamma(t,\omega)$ is a Gaussian or uniform random field, we employ Hermite or Legendre chaos, respectively. Finally, to avoid significant discretisation errors in the numerical computation, we use the fifth-order Radau IA method with stepsize $h=0.001$ to solve the corresponding deterministic DDEs.

To give some insight into the accuracy of the gPCRK method employed, we define several error measures for the random field expansion and the solutions $u_N(t)$ and $\bar{u}(t)$:

$$\varepsilon_1=\max_{0\le t\le 10}\left|\frac{\hat{u}(t)-u_I(t)}{u_I(t)}\right|,\qquad \varepsilon_2=\max_{0\le t\le 10}\left|\frac{\hat{\sigma}(t)-\sigma_I(t)}{\sigma_I(t)}\right|,\qquad \varepsilon_3=\max_{0\le t\le 10}\left|\frac{u_{II}(t)-u_I(t)}{u_I(t)}\right|,$$
$$\varepsilon_4=\max_{0\le t\le 10}\left|\frac{\sigma_{II}(t)-\sigma_I(t)}{\sigma_I(t)}\right|,\quad \varepsilon_f=\max_{0\le t\le 10}\|u(t)-u_N(t)\|_{L^2}^2,\quad \varepsilon_g=\max_{0\le t\le 10}\|u(t)-\bar{u}(t)\|_{L^2}^2,$$

where

$$\hat{u}(t)=\mathbb{E}\{u_N(t)\},\qquad \hat{\sigma}(t)=\mathbb{E}\left\{\left(u_N(t)-\hat{u}(t)\right)^2\right\},\quad u_I(t)=\mathbb{E}\{u(t)\},$$
$$\sigma_I(t)=\mathbb{E}\left\{\left(u(t)-u_I(t)\right)^2\right\},\quad u_{II}(t)=\mathbb{E}\{\bar{u}(t)\},\qquad \sigma_{II}(t)=\mathbb{E}\left\{\left(\bar{u}(t)-u_{II}(t)\right)^2\right\}.$$

When $\gamma(t,\omega)$ is a Gaussian random field, the errors ε_1, ε_2, ε_f, ε_3, ε_4 and ε_g with $N=5,7$ and different p (the highest order of the gPC basis) are displayed in Tables 1-4, showing that the gPCRK methods are effective for solving RDEPCA problems. Moreover, we also present the behaviour of the differences $\varepsilon_3-\varepsilon_1$, $\varepsilon_4-\varepsilon_2$ and $\varepsilon_g-\varepsilon_f$ in Fig. 1, which for a given N shows that ε_3, ε_4 and ε_g respectively converge to ε_1, ε_2 and ε_f as p increases. Thus when N is fixed and the discretisation error can be neglected, as p increases the error of a gPCRK method tends to the error arising from the finite-dimensional noise assumption. Similarly, when $\gamma(t,\omega)$ is a uniform random field, the effectiveness of the gPCRK methods and their error behaviour also can be observed in Tables 5-8 and Fig. 2, respectively.

Table 1: The errors ε_1, ε_2 and ε_f with different N when $\gamma(t,\omega)$ is a Gaussian random field.

	ε_1	ε_2	ε_f
$N=5$	6.514×10^{-3}	1.595×10^{-1}	4.609×10^{-4}
$N=7$	1.234×10^{-3}	4.893×10^{-2}	1.348×10^{-4}

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

Table 2: The errors ε_3 with different p, N when $\gamma(t,\omega)$ is a Gaussian random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	5.134×10^{-2}	9.172×10^{-3}	6.517×10^{-3}	6.514×10^{-3}
$N=7$	4.829×10^{-2}	4.222×10^{-3}	1.364×10^{-3}	1.234×10^{-3}

Table 3: The errors ε_4 with different p, N when $\gamma(t,\omega)$ is a Gaussian random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	6.603×10^{-1}	2.363×10^{-1}	1.595×10^{-1}	1.595×10^{-1}
$N=7$	6.567×10^{-1}	2.174×10^{-1}	5.155×10^{-2}	4.893×10^{-2}

Table 4: The errors ε_g with different p, N when $\gamma(t,\omega)$ is a Gaussian random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	4.840×10^{-4}	4.611×10^{-4}	4.609×10^{-4}	4.609×10^{-4}
$N=7$	1.660×10^{-4}	1.349×10^{-4}	1.348×10^{-4}	1.348×10^{-4}

Table 5: The errors ε_1, ε_2 and ε_f with different N when $\gamma(t,\omega)$ is a uniform random field.

	ε_1	ε_2	ε_f
$N=5$	6.507×10^{-3}	1.595×10^{-1}	4.605×10^{-4}
$N=7$	1.234×10^{-3}	4.893×10^{-2}	1.348×10^{-4}

Table 6: The errors ε_3 with different p, N when $\gamma(t,\omega)$ is a uniform random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	3.042×10^{-2}	6.604×10^{-3}	6.509×10^{-3}	6.507×10^{-3}
$N=7$	2.863×10^{-2}	2.224×10^{-3}	1.234×10^{-3}	1.234×10^{-3}

Table 7: The errors ε_4 with p, N when $\gamma(t,\omega)$ is a uniform random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	4.843×10^{-1}	1.595×10^{-1}	1.595×10^{-1}	1.595×10^{-1}
$N=7$	4.787×10^{-1}	6.839×10^{-2}	4.893×10^{-2}	4.893×10^{-2}

Table 8: The errors ε_g with different p, N when $\gamma(t,\omega)$ is a uniform random field.

	$p=1$	$p=2$	$p=3$	$p=4$
$N=5$	4.786×10^{-4}	4.606×10^{-4}	4.605×10^{-4}	4.605×10^{-4}
$N=7$	1.597×10^{-4}	1.348×10^{-4}	1.348×10^{-4}	1.348×10^{-4}

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

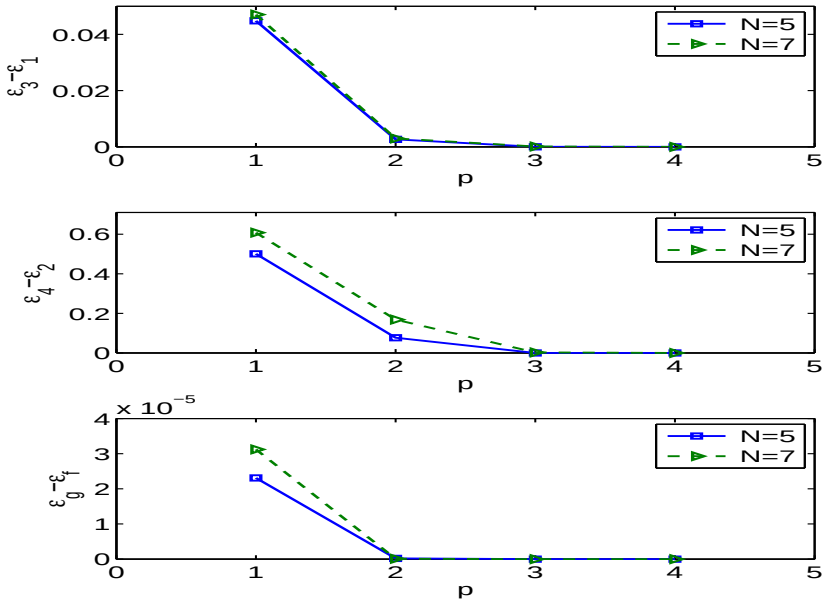


Figure 1: $\varepsilon_3-\varepsilon_1$, $\varepsilon_4-\varepsilon_2$ and $\varepsilon_g-\varepsilon_f$ with different p when $\gamma(t,\omega)$ is a Gaussian random field.

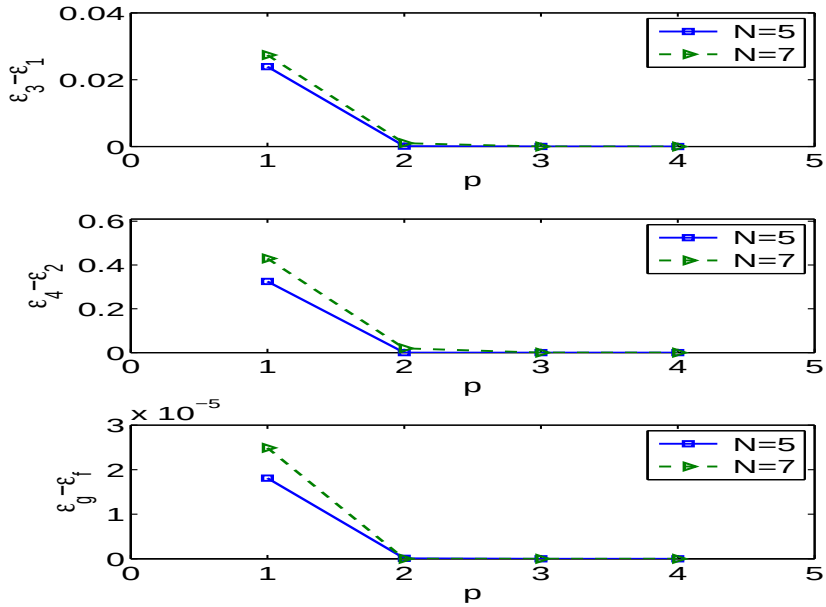


Figure 2: $\varepsilon_3-\varepsilon_1$, $\varepsilon_4-\varepsilon_2$ and $\varepsilon_g-\varepsilon_f$ with different p when $\gamma(t,\omega)$ is a uniform random field.

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

5. Concluding Remarks

We have proposed a class of numerical methods to solve the RDEPCA problem (1.2) — viz. gPCRK methods, obtained by combining generalised polynomial chaos with Runge-Kutta methods. The error of these gPCRK methods was analysed by considering the error arising from a finite-dimensional noise assumption, the projection error, the aliasing error and the discretisation error. A numerical example then illustrated the computational effectiveness of this approach.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 11501427 and 11571128). The authors would like to thank the referees for their helpful suggestions.

References

[1] I. Babuska, F. Nobile and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal. **45**, 1005-1034 (2007).
[2] J. Beck, F. Nobile, L. Tamellini and R. Tempone, *Implementation of optimal Galerkin and collocation approximations of PDEs with random coefficients*, ESIAM Proc. **33**, 10-21 (2011).
[3] O.G. Ernst, A. Mugler, H.J. Starkloff and E. Ullmann, *On the convergence of generalized polynomial chaos expansions*, ESIAM Math. Model. Numer. **46**, 317-339 (2012).
[4] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York (1991).
[5] E. Hairer, S. P. Nϕrsett and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, Springer, Berlin (1993).
[6] M. Liu, S. Ma and Z. Yang, *Stability analysis of Runge-Kutta methods for unbounded retarded differential equations with piecewise continuous arguments*, Appl. Math. Comput. **191**, 57-66 (2007).
[7] X. Liu and M. Liu, *Asymptotic stability of Runge-Kutta methods for nonlinear differential equations with piecewise continuous arguments*, J. Comput. Appl. Math. **280**, 265-274 (2015).
[8] A. Narayan and T. Zhou, *Stochastic collocation on unstructured multivariate meshes*, Commun. Comput. Phys. **18**, 1-36 (2015).
[9] W. Shi and C. Zhang, *Error analysis of generalized polynomial chaos for nonlinear random ordinary differential equations*, Appl. Numer. Math. **62**, 1954-1964 (2012).
[10] W. Shi and C. Zhang, *Generalized polynomial chaos for nonlinear random Pantograph equations*, Acta Math. Appl. Sin. English Ser. **32**, 685-700 (2016).
[11] T. Tang and T. Zhou, *Recent developments in high order numerical methods for uncertainty quantification*, Sci. Sin. Math. **45**, 891-928 (2015).
[12] W. Wang and S. Li, *Dissipativity of Runge-Kutta methods for neutral delay differential equations with piecewise constant delay*, Appl. Math. Lett. **21**, 983-991 (2008).
[13] W. Wang, *Stability of solutions of nonlinear neutral differential equations with piecewise constant delay and their discretisations*, Appl. Math. Comput. **219**, 4590-4600 (2013).
[14] J. Wiener, *Generalized Solutions of Functional Differential Equations*, World Scientific, Singapore (1993).

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press

[15] D. Xiu, *Fast numerical methods for stochastic computations: A review*, Commun. Comput. Phys. **5**, 242-272 (2009).
[16] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, New Jersey (2010).
[17] D. Xiu and G.E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. **24**, 619-644 (2002).
[18] T. Zhou, *A stochastic collocation method for delay differential equations with random input*, Adv. Appl. Math. Mech. **6**, 403-418 (2014).

https://doi.org/10.4208/eajam.150616.071216a Published online by Cambridge University Press