A well-balanced operator splitting based stochastic Galerkin method for the one-dimensional Saint-Venant system with uncertainty

Alina Chertock*, Shi Jin†, and Alexander Kurganov ‡

December 15, 2015

Abstract

We introduce an operating splitting based stochastic Galerkin method for the Saint-Venant system of shallow water equations with random inputs. The method uses a generalized polynomial chaos approximation in the stochastic Galerkin framework (referred to as the gPC-SG method). It is well-known that such approximations for nonlinear hyperbolic systems do not necessarily yield globally hyperbolic systems: their Jacobians may contain complex eigenvalues and thus trigger instabilities and ill-posedness. In this paper, we overcome this difficulty by extending the gPC-SG method from [A. Chertock, S. Jin and A. Kurganov, An operator splitting based stochastic Galerkin method for the one-dimensional compressible Euler equations with uncertainty, Submitted], where it was developed for the Euler equations of gas dynamics. The main idea is to split the underlying system into a linear hyperbolic system and a nonlinear degenerated hyperbolic system, which can be successively solved as scalar conservation laws with variable coefficients and source terms. The gPC-SG method, when applied to each of these subsystems, results in globally hyperbolic systems. We then utilize the second-order semi-discrete central-upwind scheme based well-balanced scheme in order to exactly preserve the “lake at rest” steady states. This is a very important feature of the proposed method since the lack of balance between the flux and source terms may cause large artificial waves to emerge in the case the method is used on a coarse grid, while using sufficiently fine mesh may be impractical. The performance of the designed gPC-SG method is illustrated on a number of numerical examples.

Key words. Uncertainty quantification, hyperbolic systems of balance laws, Saint-Venant system of shallow water equations, stochastic Galerkin methods, polynomial chaos, random input.

AMS subject classifications. 35R60, 35L65, 35Q31, 65M08, 76M12, 65M60, 65M70, 86-08, 86A05.

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA (chertock@math.ncsu.edu).
†Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, USA (jin@math.wisc.edu) and Department of Mathematics, Institute of Natural Sciences, MOE-LESC and SHL-MAC, Shanghai Jiao Tong University, Shanghai 200240, China.
‡Department of Mathematics, Tulane University, New Orleans, LA 70118, USA (kurganov@math.tulane.edu).
1 Introduction

In this paper, we consider the Saint-Venant system of shallow water equations, which was derived back in 1871 in [4], but is still widely used in many practical applications ranging from modeling water flows in rivers, lakes and coastal areas to oceanography and atmospheric sciences. We study the Saint-Venant system with uncertainties, which can appear in the source terms, initial or boundary data due to empirical approximations or measuring errors. Quantifying these uncertainties is important for many applications since it helps to conduct sensitivity analysis and to provide guidance for improving the models [1][7][8][24][28][29].

In the one-dimensional (1-D) case, the studied system reads as

\[
\begin{cases}
    h_t + q_x = 0, \\
    q_t + \left( hu^2 + \frac{g}{2} h^2 \right)_x = -ghB_x,
\end{cases}
\]  

(1.1)

where \( x \) is the spatial variable, \( t \) is the time, \( h(x,t,z) \) is the water depth, \( u(x,t,z) \) is the mean velocity, \( q := hu \) is the discharge, \( g \) is the gravitational constant, and \( B(x,z) \) is the bottom topography function. Note that \( h, u, q \) and \( B \) depend on the random variable \( z \). We consider the case when uncertainty can arise from the bottom and/or initial data:

\[
h(x,0,z) = h_0(x,z), \quad q(x,0,z) = q_0(x,z).
\]  

(1.2)

The shallow water system (1.1) is hyperbolic since the Jacobian of its flux has two distinct eigenvalues \( \lambda_{\pm} = u \pm \sqrt{gh} \). It is also easy to show that (1.1) admits a family of smooth steady-state solutions,

\[
q(x,z) = C_1(z), \quad \frac{u(x,z)^2}{2} + g(h(x,z) + B(x,z)) = C_2(z),
\]  

(1.3)

among which are the so-called “lake at rest” equilibria satisfying

\[
q(x,z) \equiv 0, \quad h(x,z) + B(x,z) = C(z)
\]  

(1.4)

are the most physically relevant since in many practical situations the water waves can be viewed as, in fact, small perturbations of the “lake at rest” state. Starting from [12], numerical methods that are capable of exactly preserving the steady-states solutions (1.3) (or at least the “lake at rest” states (1.4)) at the discrete level are usually referred to as well-balanced.

Our goal is to develop a well-balanced generalized polynomial chaos stochastic Galerkin (gPC-SG) method for the Saint-Venant system (1.1) by extending the gPC-SG method recently developed in [3] for the Euler equations of gas dynamics. In general, gPC-SG methods belong to the class of intrusive polynomial chaos methods. These methods use the Galerkin approximation, which results in a system of deterministic equations, solving which will give the stochastic moments of the solution of the original uncertain problem.

Existing gPC-SG methods have been successfully applied to many physical and engineering problems, where spectral convergence can be observed if the underlying solution is sufficiently smooth. The application of the gPC-SG approach to nonlinear hyperbolic systems, nevertheless,
encounters major difficulties. For linear hyperbolic systems and scalar hyperbolic conservation laws, the gPC-SG approximation yields a system which is always hyperbolic, thus gPC-SG methods are viable methods for uncertainty quantification for such equations [38]. The gPC-SG approximation also remains hyperbolic if the original system is symmetric, which motivated its use in [13] for the Hamilton-Jacobi equations with uncertainty, since the gradient of the Hamilton-Jacobi equations give a symmetric hyperbolic system. However, when applied to nonlinear hyperbolic systems, gPC-SG methods result in systems which are not necessarily globally hyperbolic, since their Jacobian matrices may contain complex eigenvalues. Such examples were given in [5]. There have been some literatures that address this issue. In [5], for the compressible Euler equations, by using the gPC approximation for the entropy variables one guarantees the hyperbolicity, since under the entropy variables the system is symmetric hyperbolic. In this approach, one needs to change from entropy to conservative variables by solving a minimization problem at every mesh point and time step, which can be computationally costly. Another approach is built upon the Roe variables, which locally linearizes the underlying PDE system, as proposed in [31]; see also [32, 36]. This approach is restricted to systems which admit the Roe linearization and requires switching between the Roe and original variables at every grid point and every time step by solving nonlinear algebraic systems using the trust-region-dogleg algorithm.

In [3], we studied the Euler equations of gas dynamics and obtained the hyperbolic systems for the gPC coefficients in the following manner. We first used the operator splitting approach to split the Euler system into the three simpler subsystems: one of them is linear and the other two are degenerate nonlinear systems, which, in fact, reduce to scalar hyperbolic equations. We then applied the gPC-SG method to each of the subproblems and the resulting system for the gPC coefficients is guaranteed to be hyperbolic since it is well-known that for linear hyperbolic systems and nonlinear scalar equations, gPC-SG methods yield globally hyperbolic discretizations. In the current work, we extend this approach to the system of shallow water equations (1.1).

In gPC-SG methods, the solution of the underlying system is sought in terms of orthogonal polynomial series [9, 40], whose coefficients satisfy deterministic (systems of) time-dependent PDEs. Since in our approach the obtained gPC-SG system of equations for the coefficients is guaranteed to be hyperbolic, we follow [3] and numerically solve it by a semi-discrete second-order central-upwind scheme, which has been originally introduced in [17,18,20] as a Riemann-problem-solver-free black box solver for general multidimensional hyperbolic systems of conservation laws and then extended to the Saint-Venant system of shallow water equations in [16,19]. One of the advantages of the central-upwind schemes is their direct applicability to systems with complicated (generalized) Riemann problem solutions including the systems for the gPC coefficients arising in the process of application of gPC-SG methods.

This paper differs from [3] in two aspects. First, the operator splitting framework presented in [3] is problem specific. The splitting for the shallow-water equations is indeed different from that for the Euler equations. Second, the issue of well-balancedness, which is essential for shallow-water equations with (possibly discontinuous) bottom topography, is not addressed in [3].

An alternative, the non-intrusive approach called the stochastic collocation method [39], for shallow water equations with uncertainty was introduced in [28,29]. Some comparisons between the stochastic Galerkin method and the stochastic collocation method for the diffusion equation were given in [6]. A stochastic well-balanced gPC-SG method for scalar hyperbolic balance law
was given in [15].

The rest of the paper is organized as follows. In Section 2, we briefly review the gPC-SG method for a general hyperbolic system of balance laws. In Section 3, we describe a splitting approach, which is validated in Section 3.1 where we present the modified equation analysis to provide some guidance about the stability property of the proposed splitting approach, and in Sections 3.3, where we numerically solve the deterministic Saint-Venant system using the well-balanced discretization presented in Section 3.2. Then, in Section 4, we explain how the gPC-SG method can be directly applied to each one of the subsystems and finally, in Section 5, we demonstrate the performance of the resulting gPC-SG method on a number of numerical examples with random inputs.

2 The gPC-SG method – an overview

In this section, we briefly describe the gPC-SG method for the hyperbolic system of balance laws

$$U_t + F(U, x, z)_x = R(U, x, z), \quad x \in \mathbb{R}, \; t > 0, \; z \in \Omega \subset \mathbb{R}^d,$$

(2.1)

where $U(x, t, z)$ is the unknown vector function, $F(U, x, z)$ is the flux, and $R(U, x, z)$ is the source term. In the gPC expansion, the solution of the system of stochastic PDEs (2.1) is sought in terms of an orthogonal polynomial series in $z$ ([40]):

$$U(x, t, z) \approx U_N(x, t, z) = \sum_{i=0}^{M-1} \hat{U}_i(x, t) \Phi_i(z), \quad M = \binom{d + N}{d},$$

(2.2)

where $\{\Phi_m(z)\}$ are $d$-variate orthonormal polynomials of degree up to $N \geq 1$ from $\mathbb{P}_N^d$ satisfying

$$\int_{\Omega} \Phi_i(z) \Phi_\ell(z) \mu(z) \, dz = \delta_{i\ell}, \quad 0 \leq i, \ell \leq M - 1, \; M = \dim(\mathbb{P}_N^d).$$

(2.3)

Here, $\mu(z)$ is the probability density function of $z$ and $\delta_{i\ell}$ is the Kronecker symbol. The choice of the orthogonal polynomials depends on the distribution function of $z$. For example, a Gaussian distribution defines the Hermite polynomials; a uniform distribution defines the Legendre polynomials, etc. Note that when the random dimension $d > 1$, $\{\Phi_i(z)\}$ are multidimensional polynomials of degree up to $N$ of $z$. An ordering scheme for multiple index is required to re-order the polynomials into a single index $i$ in (2.2). Typically, the graded lexicographic order is used; see, e.g., [38, Section 5.2].

For the PDE system with random inputs (2.1), the gPC-SG method seeks to satisfy the governing equations in a weak form by ensuring that the residual is orthogonal to the gPC polynomial space. Substituting the approximation $U_N$ from (2.2) into the governing system (2.1) and using the Galerkin projection yield

$$(\hat{U}_i)_t + (\hat{F}_i)_x = \hat{R}_i, \quad 0 \leq i \leq M - 1,$$

(2.4)
where
\[
\hat{F}_i = \int_{\Omega} F\left(\sum_{j=0}^{M-1} \hat{U}_j(x,t)\Phi_j(z), x, z\right)\Phi_i(z)\mu(z) \, dz,
\]
\[
\hat{R}_i = \int_{\Omega} R\left(\sum_{j=0}^{M-1} \hat{U}_i(x,t)\Phi_i(z), x, z\right)\Phi_i(z)\mu(z) \, dz,
\]
\[0 \leq i \leq M - 1. \tag{2.5}\]

This is a system of deterministic equations for the expansion coefficients \(\hat{U}_i, \ i = 0, \ldots, M - 1\).

In most cases, the equations in (2.4) are coupled.

For linear or symmetric nonlinear hyperbolic systems as well as in the nonlinear scalar case, the gPC system (2.4) is hyperbolic. However, if the nonlinear hyperbolic system (2.1) is not symmetric (as in the case of the Saint-Venant system (1.1) studied here), the system (2.4) is not always hyperbolic since its Jacobian may encounter complex eigenvalues. In the latter case, the initial and initial-boundary value problems for (2.4) are ill-posed and the gPC-SG method can be unstable. In what follows, we shall present a way to overcome this difficulty by introducing an operator splitting approach for the system (1.1), which will guarantee that the gPC-SG discretization of each of the split subsystems always results in a globally hyperbolic system for the expansion coefficients.

### 3 An operator splitting

We start by splitting the system (1.1) into the following two subsystems:

\[
\begin{aligned}
&\begin{cases}
ht + qx = 0, \\
qt + a^2hx = 0,
\end{cases} \tag{3.1}
\end{aligned}
\]
and

\[
\begin{aligned}
&\begin{cases}
ht = 0, \\
qt + \left[\frac{q^2}{h} + \frac{g}{2}h^2 - a^2h\right]_x = -ghB_x.
\end{cases} \tag{3.2}
\end{aligned}
\]

The first subsystem (3.1) is a linear hyperbolic system whose Jacobian has two distinct real eigenvalues \(\lambda_{\pm} = \pm a\), where the parameter \(a > 0\) is to be chosen to satisfy the following subcharacteristic condition:

\[-a \leq u - \sqrt{gh} < u + \sqrt{gh} \leq a. \tag{3.3}\]

The second system (3.2) is essentially a scalar Burgers equation for \(q\) with variable coefficient and source term since \(h\) remains constant in time in (3.2). Furthermore, we choose \(a\) to be larger than the characteristic speed for in the second equation of the second subsystem (3.2). This yields the following choice:

\[a = \sup_u \left\{ \max \left(\frac{|u|}{2|\sqrt{gh}|}, 2|u|\right) \right\}.\]

It is easy to check that under the subcharacteristic condition (3.3) each of the subsystem in the above splitting is strictly hyperbolic. Therefore, after the gPC-SG approximation, each of
the above systems gives a globally hyperbolic system for the gPC coefficients.

### 3.1 A modified equation analysis

To gain some understanding of how a shock capturing scheme based on this splitting could yield a numerically stable discretization, we perform a modified equation analysis of the following version of the first-order Lax-Friedrichs scheme applied to equations (3.1) and (3.2) using the first-order splitting approach (A.2) (see Appendix A), in the case of flat bottom topography ($B = 0$):

\[
\begin{align*}
\frac{h_j^n - h_j^{n+1}}{\Delta t} + \frac{q_j^{n+1} - q_j^n}{2\Delta x} &= \frac{a}{2\Delta x} (h_{j+1}^n - 2h_j^n + h_{j-1}^n), \\
\frac{q_j^n - q_j^{n+1}}{\Delta t} + a^2 \frac{h_j^{n+1} - h_j^n}{2\Delta x} &= \frac{a}{2\Delta x} (q_{j+1}^n - 2q_j^n + q_{j-1}^n), \\
\frac{h_j^{n+1}}{\Delta t} &= h_j^*, \\
\frac{q_j^{n+1} - q_j^*}{\Delta t} + \frac{1}{2} \left[ \frac{(q_j^{n+1})^2}{h_j^{n+1}} + \frac{g}{2} (h_j^{n+1})^2 - a^2 h_j^{n+1} \right] - \frac{1}{2} \left[ \frac{(q_j^{n-1})^2}{h_j^{n-1}} + \frac{g}{2} (h_j^{n-1})^2 - a^2 h_j^{n-1} \right] &= \frac{b}{2\Delta x} (q_j^{n+1} - 2q_j^n + q_j^{n-1}).
\end{align*}
\]

Here, $\Delta x$ is the mesh size and $\Delta t$ is the timestep, the point values of $h$ and $q$ are computed at the grid points ($x_j = j\Delta x$, $t^n = n\Delta t$), and $b \approx 2||u||_{\infty}$. The superscripts $n$ and $n+1$ indicate the quantities at times $t^n$ and $t^{n+1}$, respectively, while * denotes an intermediate state obtained at the end of the first splitting substep.

The corresponding semi-discrete (discrete in time and continuous in space) modified equations are

\[
\begin{align*}
\frac{h^n - h^n}{\Delta t} + g^n x &= a \frac{\Delta x h^n}{2}, \\
\frac{q^n}{\Delta t} + a^2 h^n &= a \frac{\Delta x q^n}{2},
\end{align*}
\]

and

\[
\begin{align*}
\frac{h^{n+1}}{\Delta t} &= h^*, \\
\frac{q^{n+1} - q^*}{\Delta t} + \frac{1}{2} \left[ \frac{(q^*)^2}{h^*} + \frac{g}{2} (h^*)^2 - a^2 h^* \right] x &= \frac{b}{2\Delta x} q^{n+1} - 2q^n + q^{n-1}.
\end{align*}
\]

We now eliminate $h^*$ and $q^*$ and drop the superscript $n$ for all of the quantities computed at time level $t^n$ to obtain

\[
\frac{q^{n+1} - q}{\Delta t} + a^2 h_x - \frac{a}{2} \Delta x q_{xx}
\]

\[
+ \left[ \frac{g}{2} \left( h - \Delta t q_x + \frac{a}{2} \Delta t \Delta x h_{xx} \right)^2 - a^2 \left( h - \Delta t q_x + \frac{a}{2} \Delta t \Delta x h_{xx} \right) \right] x
\]

\[
+ \left[ \frac{g}{2} \left( h - \Delta t q_x + \frac{a}{2} \Delta t \Delta x h_{xx} \right)^2 - a^2 \left( h - \Delta t q_x + \frac{a}{2} \Delta t \Delta x h_{xx} \right) \right] x
\]

\[
= \frac{b}{2\Delta x} q^{n+1} - 2q^n + q^{n-1}.
\]
\[
\frac{q^{n+1} - q}{\Delta t} - \frac{a^2}{2} \Delta x = \frac{b}{2} \Delta x \left( q - a^2 \Delta t h_x + \frac{a}{2} \Delta t \Delta q_{xx} \right)_{xx}. \tag{3.4}
\]

Keeping only the \(O(1)\), \(O(\Delta x)\) and \(O(\Delta t)\) terms on the right-hand side of (3.4) and dropping the higher-order terms result in

\[
\frac{q^{n+1} - q}{\Delta t} - \frac{a^2}{2} \Delta x_{xx} + \left[ \frac{q^2 - 2a^2 \Delta t q h_x}{h - \Delta t q_x} + \frac{g}{2} (h^2 - 2\Delta t q_x) \right]_{x} + a^2 \Delta t q_{xx} = \frac{b}{2} \Delta x q_{xx}.
\]

The last equation can be further simplified to give

\[
\frac{q^{n+1} - q}{\Delta t} + \left[ \frac{q^2}{h} + \frac{g h^2}{2} \right]_{x} = \Delta t \left( 2a^2 (u h)_x - (u^2 q_x)_x + g (h q_x)_x - a^2 q_{xx} \right) + \frac{\Delta x}{2} (a + b) q_{xx}.
\]

Next, we Taylor expand \(q^{n+1}\) about \(t^n\) and use (1.1) to express all of the time derivatives in terms of the corresponding space derivatives. A necessary condition for the method to be stable is that the coefficient of \(q_{xx}\) is positive, which is guaranteed if the following timestep restriction is satisfied:

\[
\sup \left( \frac{5u^2 - gh + 2a^2}{a + b} \right) \frac{\Delta t}{\Delta x} \leq 1.
\]

Notice that the CFL condition for the Lax-Friedrichs method applied to the original unsplit system (1.1) is given by

\[
\sup(|u| + \sqrt{gh}) \frac{\Delta t}{\Delta x} \leq 1,
\]

and since \(b \approx 2\|u\|_{\infty}\), one has

\[
\frac{5u^2 - gh + 2a^2}{a + b} \geq \sup(|u| + \sqrt{gh}),
\]

and thus we conclude that a smaller \(\Delta t\) is required for the numerical solution of the split system to be stable compared to the unsplit case.

### 3.2 A well-balanced discretization

While there have been many numerical methods available in the literature for solving linear hyperbolic systems like (3.1) and nonlinear scalar equations like (3.2) (see, e.g., [22] and references therein), here the additional difficulty is associated with the fact that the resulting gPC-SG approximation should not only lead to a globally hyperbolic system, but also be well-balanced. In order to be able to exactly preserve the “lake at rest” steady state solution (1.4) at each substep of the splitting algorithm, we adopt the well balanced strategy from [16] according to which one needs to work with the equilibrium variables, the water surface \(w := h + B\) and discharge \(q\), rather than with the conservative ones \((h\) and \(q)). Following [16], we rewrite the shallow water system (1.1) in the equivalent form:

\[
\begin{align*}
\begin{cases}
{w_t + q_x} = 0, \\
q_t + \left[ \psi + g \left( \frac{1}{2} w^2 - w B \right) \right]_x = -gw B_x,
\end{cases}
\end{align*}
\tag{3.5}
\]
where we use the notation

$$\psi := \frac{q^2}{w - B}. \quad (3.6)$$

The corresponding subsystems then become:

$$\begin{cases} 
    w_t + q_x = 0, \\
    q_t + a^2 w_x = 0, 
\end{cases} \quad (3.7)$$

and

$$\begin{cases} 
    w_t = 0, \\
    q_t + \left[ \psi + g \left( \frac{1}{2} w^2 - wB \right) - a^2 w \right]_x = -gwB_x. 
\end{cases} \quad (3.8)$$

### 3.3 Numerical validation of the operator splitting

Before proceeding further with deriving the gPC-SG method for the systems (3.7) and (3.8), we check the validity of the hyperbolic splitting on the purely deterministic shallow water model, in which neither the bottom topography function nor the initial data depend on a random variable $z$. To this end, we numerically solve the system (3.5), (3.6) using two different methods.

We first directly apply the second-order semi-discrete central-upwind scheme from [19] (briefly described in Appendix B in the context of general hyperbolic systems of balance laws) to the original system (3.5), which can be written in the vector form (2.1) with

$$U_j = (w_j, q_j)^T, \quad F_j = (q_j, \psi + g \left( \frac{1}{2} w_j^2 - w_jB \right)), \quad R_j = \left( 0, -gwB_x \right)^T.$$ 

The semi-discrete central-upwind discretization of the system (3.5) is thus obtained from (B.7), (B.8) with the cell averages $U_j = (w_j, q_j)^T$ defined in (B.1), the point values $U_{j+\frac{1}{2}} = (w_{j+\frac{1}{2}}, q_{j+\frac{1}{2}})^T$ computed according to (B.6) and the one-sided local speeds of propagation (B.5) given by

$$a^+_{j+\frac{1}{2}} = \max \left\{ u^-_{j+\frac{1}{2}} + \sqrt{g(w_{j+\frac{1}{2}} - B_{j+\frac{1}{2}})}, \ u^+_{j+\frac{1}{2}} + \sqrt{g(w_{j+\frac{1}{2}} - B_{j+\frac{1}{2}})}, \ 0 \right\},$$

$$a^-_{j+\frac{1}{2}} = \min \left\{ u^-_{j+\frac{1}{2}} - \sqrt{g(w_{j+\frac{1}{2}} - B_{j+\frac{1}{2}})}, \ u^+_{j+\frac{1}{2}} - \sqrt{g(w_{j+\frac{1}{2}} - B_{j+\frac{1}{2}})}, \ 0 \right\},$$

where

$$u^+_{j+\frac{1}{2}} = \frac{q_{j+\frac{1}{2}}}{w_{j+\frac{1}{2}} - B_{j+\frac{1}{2}}}, \quad B_{j+\frac{1}{2}} = B(x_{j+\frac{1}{2}}). \quad (3.9)$$

As it has been mentioned above, the shallow water system (3.5) admits steady-state solutions and therefore it is necessary to solve it by a well-balanced scheme. This can be achieved by approximating the second component in the source term $R$ by a special quadrature, which guarantees the well-balancedness of the resulting scheme. Following [16, 19], we compute the cell averages $\overline{R}_j$ as

$$\overline{R}_j = \left( 0, -g w_j \frac{B_{j+\frac{1}{2}} - B_{j-\frac{1}{2}}}{\Delta x} \right)^T.$$ 

(3.10)
For the splitting approach, we implement the second-order Strang operator splitting method (A.3); see Appendix A. Each of the subsystems (3.7) and (3.8) is again solved using the second-order semi-discrete central-upwind scheme described in Appendix B. The first system (3.7) is a linear system of conservation laws of the form (2.1) with

$$U = \begin{pmatrix} w \\ q \end{pmatrix}, \quad F = \begin{pmatrix} q \\ a^2 w \end{pmatrix}, \quad R = 0,$$

(3.11)

for which the central-upwind scheme (B.7), (B.8) applied with $a^+_j = -a^-_{j+\frac{1}{2}} = a \forall j$ reads

$$\frac{d}{dt} U_j = -\frac{H_{j+\frac{1}{2}} - H_{j-\frac{1}{2}}}{\Delta x}, \quad H_{j+\frac{1}{2}} = \frac{F(U^-_{j+\frac{1}{2}}) - F(U^+_{j+\frac{1}{2}})}{2} - \frac{a}{2} \left[ U^+_{j+\frac{1}{2}} - U^-_{j+\frac{1}{2}} \right].$$

(3.12)

The second system (3.8) is actually reduced to a scalar hyperbolic balance law (2.1) for $q$ with

$$U = q, \quad F = \psi + g \left( \frac{1}{2} w^2 - wB \right), \quad R = -gwBx,$$

(3.13)

which is also solved by the central-upwind scheme (B.7), (B.8) with the local speeds

$$a^+_j = \max \left\{ 2u^+_j, 2u^+_j, 0 \right\} \quad \text{and} \quad a^-_j = \min \left\{ 2u^-_j, 2u^-_j, 0 \right\},$$

and the well-balanced source term quadrature $R_j$ taken as the second component in (3.10).

Once the two numerical methods are constructed, we run numerical simulations for a dam break over the nonflat bottom topography

$$B(x) = \begin{cases} 0.1 + 0.125(\cos(5\pi x) + 1), & x < 0, \\ 0.1, & x > 0. \end{cases}$$

(3.14)

The gravitational constant $g = 1$ and the initial data are

$$w(x, 0) = \begin{cases} 1, & x < 0, \\ 0.5, & x > 0, \quad q(x, 0) = 0. \end{cases}$$

(3.15)

The problem (1.1), (3.14), (3.15) is numerically solved by both unsplit and split approaches. The results computed on a uniform grid with $\Delta x = 1/400$ at time $t = 0.8$ are plotted in Figure 3.1 which shows a good agreement between the two solutions.

This example indicates that the splitting approach results in a good shock capturing scheme in the deterministic setting, helping to build the confidence for the stochastic problem.
Figure 3.1: Shallow water equations: $w$ (left) and $q$ (right) computed by both the unsplit and split approaches.

4 The gPC-SG approximation

We now derive a gPC-SG scheme for the systems (3.7) and (3.8). To this end, we seek polynomial approximations of $w, q, p$ and $B$ in the following form:

$$w_N(x, t, z) = \sum_{i=0}^{N} \hat{w}_i(x, t) \Phi_i(z), \quad q_N(x, t, z) = \sum_{i=0}^{N} \hat{q}_i(x, t) \Phi_i(z),$$
$$\psi_N(x, t, z) = \sum_{i=0}^{N} \hat{\psi}_i(x, t) \Phi_i(z), \quad B_N(x, z) = \sum_{i=0}^{N} \hat{B}_i(x) \Phi_i(z).$$

Substituting (4.1) into both (3.7) and (3.8) and conducting the Galerkin projection yield the following equations for the gPC coefficients $\{\hat{w}_i\}$ and $\{\hat{q}_i\}$ for all $i = 0, \ldots, N$:

$$\begin{cases}
(\hat{w}_i)_t + (\hat{q}_i)_x = 0, \\
(\hat{q}_i)_t + a^2(\hat{w}_i)_x = 0,
\end{cases}$$

and

$$\begin{cases}
(\hat{w}_i)_t = 0, \\
(\hat{q}_i)_t + \left[ \hat{\psi}_i + g \sum_{k,\ell=0}^{N} \left( \frac{1}{2} \hat{w}_k \hat{w}_\ell - \hat{w}_k \hat{B}_\ell \right) S_{k\ell i} - a^2 \hat{w}_i \right]_x = -g \sum_{k,\ell=0}^{N} \hat{w}_k (B_\ell)_x S_{k\ell i},
\end{cases}$$

where

$$S_{k\ell i} = \int_{\Omega} \Phi_k(z) \Phi_\ell(z) \Phi_i(z) \mu(z) \, dz \quad \text{for all} \quad k, \ell, i = 0, \ldots, N.$$
As in, for example, [25], the gPC coefficients \( \hat{\psi_i} \) needed in (4.3) can be computed by applying the gPC-SG approximation of the relation (3.6) written as \((w - B)\psi = q^2\):

\[
\sum_{k,\ell=0}^{N} \hat{\psi}_k (\hat{w}_k - \hat{B}_\ell) S_{ik\ell} = \sum_{k,\ell=0}^{N} \hat{q}_k \hat{q}_\ell S_{ik\ell}, \quad i = 0, \ldots, N, \tag{4.5}
\]

and once \( \{\hat{q}_i\}, \{\hat{w}_i\} \) and \( \{\hat{B}_i\} \) are given, one can solve the linear system (4.5) to obtain \( \{\hat{\psi}_i\}\).

**Remark 4.1** When applying the gPC-SG to the relation \((w - B)\psi = q^2\), for the exact value of \(w - B = h > 0\), one would obtain the following linear system for \( \{\hat{\psi}_i\} \) for all \( i = 0, \ldots N \):

\[
\sum_{k=0}^{N} \hat{\psi}_k \int_{\Omega} h(z) \Phi_i(z) \Phi_k(z) \mu(z) \, dz = \sum_{k,\ell=0}^{N} \hat{q}_k \hat{q}_\ell S_{ik\ell}. \tag{4.6}
\]

For \( h > 0 \), the coefficient matrix on the left-hand side of the linear system (4.6) is symmetric positive definite; see, e.g., [44]. Thus, the system (4.6) is invertible. The linear system (4.5) is obtained by approximating \( h = w - B \approx \sum_{\ell=0}^{N} (\hat{w}_\ell - \hat{B}_\ell) \Phi_\ell \). Thus, the entries in the coefficient matrix in (4.5) are just the gPC approximation of those in (4.6), and they are spectrally close to each other for smooth solutions. In the general case, there is a lack of analysis on the invertibility of the linear system (4.5), but we have never encountered any problems in our numerical simulations.

For the spatial discretizations of the systems (4.2) and (4.3), we again apply the second-order semi-discrete central-upwind scheme described in Appendix B. Here, for each \( i = 0, \ldots, N \) in (4.2), we have a system of conservation laws of the form (2.1) with

\[
U_i = \begin{pmatrix} \hat{w}_i \\ \hat{q}_i \end{pmatrix}, \quad F_i = \begin{pmatrix} \hat{q}_i \\ a^2 \hat{w}_i \end{pmatrix}, \quad R_i = 0,
\]

and thus apply the central upwind scheme (B.7), (B.8) with \( a_{j+\frac{1}{2}}^+ = -a_{j-\frac{1}{2}}^- = a \) \( \forall j \) to each one of the above \( N + 1 \) systems. The resulting schemes for each \( i = 0, \ldots, N \) read

\[
\frac{d}{dt} (U_i)_j = -\frac{(H_i)_{j+\frac{1}{2}} - (H_i)_{j-\frac{1}{2}}}{\Delta x}, \quad (H_i)_{j+\frac{1}{2}} = \frac{F_i((U_i)^+_{j+\frac{1}{2}}) - F_i((U_i)^-_{j+\frac{1}{2}})}{2} - \frac{a}{2} \left[ (U_i)^+_{j+\frac{1}{2}} - (U_i)^-_{j+\frac{1}{2}} \right].
\]

The second system (4.3) is now a set of \( N \) decoupled scalar hyperbolic equations for \( \{\hat{q}_i\} \), which still can be put into the form (2.1) with

\[
U_i = \hat{q}_i, \quad F_i = \hat{\psi}_i + g \sum_{k,\ell=0}^{N} \left( \frac{1}{2} \hat{w}_k \hat{w}_k - \hat{w}_k \hat{B}_\ell \right) S_{ik\ell} - a^2 \hat{w}_i, \quad R_i = -g \sum_{k,\ell=0}^{N} \hat{w}_k (\hat{B}_\ell) x S_{ik\ell},
\]

each of which is solved by the central-upwind scheme (B.7), (B.8) with the local speeds estimated using the leading (zeroth) terms in the gPC expansions (4.1) as

\[
a^{+}_{j+\frac{1}{2}} = \max \left\{ 2\hat{u}_{j+\frac{1}{2}}, 2\hat{u}_{j+\frac{1}{2}}^+, 0 \right\}, \quad a^{-}_{j+\frac{1}{2}} = \min \left\{ 2\hat{u}_{j+\frac{1}{2}}, 2\hat{u}_{j+\frac{1}{2}}^+, 0 \right\}.
\]
Here, we have used the following notation:

\[ \hat{u}_{j+\frac{1}{2}} := \frac{(\hat{q}_0)_{j+\frac{1}{2}} \pm (\hat{w}_0)_{j+\frac{1}{2}} - (\hat{B}_0)_{j+\frac{1}{2}}}{2}. \]

**Well-balanced property.** It should be observed that the resulting central-upwind scheme for the gPC systems (4.2), (4.3) is well-balanced in the sense that it is capable of exactly preserving the “lake at rest” steady states (1.4). Indeed, for the state (1.4) one has \( \hat{w}_i \equiv \text{Const}, \hat{q}_i \equiv 0 \) and \( \hat{\psi}_i \equiv 0 \) for all \( i = 0, \ldots, N \). One can easily show that substituting these constant quantities into the central-upwind scheme for (4.2) and (4.3) results in the following system of ODEs for each \( i = 0, \ldots, N \) and each finite-volume cell \( I_j \):

\[
\begin{align*}
\frac{d(\hat{w}_i)_j}{dt} &= -\frac{\hat{q}_i - \hat{q}_i}{\Delta x} = 0, \\
\frac{d(\hat{q}_i)_j}{dt} &= -a^2 \frac{\hat{w}_i - \hat{w}_i}{\Delta x} = 0,
\end{align*}
\]

and

\[
\begin{align*}
\frac{d(\hat{w}_i)_j}{dt} &= 0, \\
\frac{d(\hat{q}_i)_j}{dt} &= -\frac{\hat{w}_i - \hat{w}_i}{\Delta x} + a^2 \frac{\hat{w}_i - \hat{w}_i}{\Delta x} - g \sum_{k,\ell=0}^N \frac{\hat{w}_k \hat{w}_\ell S_{k\ell i} - \hat{w}_k \hat{w}_\ell S_{k\ell i}}{\Delta x} \\
&\quad + g \sum_{k,\ell=0}^N \frac{\hat{B}_\ell (\hat{B}_\ell)_{j+\frac{1}{2}} - (\hat{B}_\ell)_{j+\frac{1}{2}} S_{k\ell i} - g \sum_{k,\ell=0}^N \frac{\hat{B}_\ell (\hat{B}_\ell)_{j+\frac{1}{2}} - (\hat{B}_\ell)_{j+\frac{1}{2}} S_{k\ell i}}{\Delta x} = 0,
\end{align*}
\]

respectively. This proves that the steady state (1.4) is exactly preserved by the designed central-upwind scheme.

**5 Numerical examples**

In this section, we illustrate the performance of the proposed gPC-SG method on a number of numerical examples. The randomness/uncertainty can enter the problem through either the initial conditions (Example 1) or bottom topography (Example 2). For simplicity, we will always assume a 1-D random variable \( z \) obeying the uniform distribution on \([-1, 1]\), thus the Legendre polynomials are used as the gPC basis. The mean and standard deviation of the computed solution \( U \), which are shown in the figures below, are given by

\[
E[U] = \hat{U}_0, \quad \sigma[U] = \sum_{i=1}^N (\hat{U}_i)^2,
\]

where \( \hat{U}_i, i = 0, \ldots, N \) are the computed gPC coefficients of \( U \).

In all of the examples, the second-order semi-discrete central-upwind scheme (B.7), (B.8) was implemented for the spatial discretization, the piecewise linear interpolants were constructed using the minmod limiter (B.3), (B.4) with \( \theta = 1.3 \), and the arising systems of ODEs were...
numerically integrated by the three-stage third-order strong stability-preserving (SSP) Runge-Kutta method \cite{10,11}.

**Example 1 (A perturbed water surface).** In the first example, we numerically solve the system \((1.1)\) with \(g = 1\) and the following deterministic bottom topography (see Figure (5.1)):

\[
B(x) = \begin{cases} 
0.25(\cos(5\pi(x + 0.35)) + 1), & -0.55 < x < -0.15, \\
0.125(\cos(10\pi(x - 0.35)) + 1), & 0.25 < x < 0.45, \\
0, & \text{otherwise}.
\end{cases}
\]

The initial data (also shown in Figure (5.1)) corresponds to a random perturbation with maximum amplitude \(\delta = 0.001\) of the stationary steady-state solution \(w \equiv 1, u \equiv 0\). The system is considered in the computational domain \([-1, 1]\) with the outflow boundary conditions imposed at both end points.

![Figure 5.1: Example 1: Initial state and bottom topography. The water surface plot is not to scale.](image)

The perturbation, initially located at \([-0.05, 0.05]\), splits into two pulses moving in opposite directions. When the pulses interact with the nonflat part of the bottom, a complicated surface wave structure is developed. This is illustrated in Figures 5.2 and 5.3 where we plot the mean and standard deviation of the water surface \(w\) (left figure) and discharge \(q\) (right figure) computed at time \(t = 0.8\) using the well-balanced gPC-SG method on different uniform spatial grids with \(\Delta x = 1/100, 1/400\) and \(1/1600\) and with \(N = 8\) (the highest degree in the gPC expansion).

In order to demonstrate the importance of the well-balanced property, we show in Figure 5.4 the mean of the solution, \(w\) (left) and \(q\) (right) computed by the proposed gPC-SG method with the spatial discretization performed using a non-well-balanced version of the second-order semi-discrete central-upwind scheme \((B.7, B.8)\). The latter is obtained by replacing the well-balanced quadrature \((3.10)\) with the midpoint one,

\[
\overline{R}_j = (0, -g\overline{w}_j B_x(x_j))^T.
\]

As one can see, the solution computed by the non-well-balanced scheme using a coarse mesh
with $\Delta x = 1/100$ contains unphysical oscillations of the magnitude comparable to the size of the physical waves (those captured by the well-balanced scheme, see Figure 5.2). Moreover, the magnitude of the oscillations decreases quite slowly when the mesh is refined, which confirms that the well-balanced property is essential. We note that the importance of using well-balanced gPC-SG for hyperbolic balance laws with random inputs was addressed in [15].

**Example 2 – A perturbed bottom topography.** In this section, we consider the same example as in Section 3.3, but with uncertain bottom topography. Namely, we solve the shallow water system (1.1) with $g = 1$, perturbed function (3.14), namely,

$$B(x, z) = \begin{cases} 
0.125(\cos(5\pi x) + 1) + 0.1z, & x < 0.1, \\
0.1z, & \text{otherwise}, 
\end{cases}$$
and subject to the same deterministic Riemann initial data as in (3.15).

The problem is numerically solved by the well-balanced gPC-method. The results (the mean and standard deviation of the water surface \(w\) and discharge \(q\)) computed on uniform grids with \(\Delta x = 1/100, 1/400\) and \(N = 8\) at time \(t = 0.8\) are plotted in Figures 5.5 and 5.6, which show a good agreement of the mean with the results shown in Figure 3.1. We have also performed the same numerical experiments using a higher-order polynomial expansion in \(z\) with \(N = 16\). The obtained results (not reported here for brevity) are almost identical to the \(N = 8\) results.

Figure 5.5: Example 2: Mean of \(w\) (left) and \(q\) (right) computed by the well-balanced gPC-SG method on different grids. The solid lines in the left figure represent the bottom topography with perturbation.
In this section, we give an account of the operator splitting approach, which can be briefly described as follows. Consider the system (2.1) and assume that it can be split into two subsystems:

\[ U_t + F_1(U)_x = 0, \quad U_t + F_2(U)_x = 0, \tag{A.1} \]

where \( F = F_1 + F_2 \) and \( S_1, S_2 \) denote the exact solution operators associated with the corresponding subsystems.

Let us assume that the solution of the original system (2.1) is available at time \( t \). We then introduce a (small) time step \( \Delta t \) and evolve the solution of (2.1) from \( t \) to \( t + \Delta t \) in 2 substeps, which result in the following approximate solution at time \( t + \Delta t \):

\[ U(x, t + \Delta t, z) = S_2(\Delta t)S_1(\Delta t)U(x, t, z). \tag{A.2} \]

In general, if all solutions involved in the splitting algorithm (A.1), (A.2) are smooth, the method is first-order accurate; see, e.g., [26,27,33].

Higher-order operator splitting algorithms can be derived by considering additional substeps.
For instance, one time step of the second-order Strang splitting method \cite{26,27,33} consists of 3 substeps:

\[
U(x, t + \Delta t, z) = S_1(\Delta t/2)S_2(\Delta t)S_1(\Delta t/2)U(x, t, z). \tag{A.3}
\]

We also refer the readers to \cite{14,21,34,42}, where higher-order operator splitting methods can be found.

In practice, the exact solution operators $S_1$ and $S_2$ are to be replaced by their numerical approximations. Note that if the obtained subproblems are of a different nature, they can be solved by different numerical methods — this is one of the main advantages of the operator splitting technique.

## B The semi-discrete second-order central-upwind scheme

In this section, we briefly describe the central-upwind schemes for 1-D hyperbolic systems of balance laws. These schemes are Godunov-type finite-volume methods. For their complete description and derivation, we refer the reader to \cite{16,19}.

We consider a 1-D system of balance laws (2.1) and divide the computational domain into the uniform cells $I_j = [x_{j-1/2}, x_{j+1/2}]$ of size $|I_j| = \Delta x$ centered at points $x_j = j\Delta x, \; j = j_L, \ldots, j_R$. We assume that at certain time $t$ the cell averages of the computed solution,

\[
\left( U_j(t) = \frac{1}{\Delta x} \int_{I_j} U(x,t) \, dx \right. \tag{B.1}
\]

are available. From here on we suppress the time-dependence of all of the indexed quantities in order to shorten the notation.

We use the cell averages (B.1) to reconstruct a non-oscillatory piecewise linear polynomial

\[
\tilde{U}(x) = U_j + (U_x)_j(x - x_j), \quad x_{j-1/2} < x < x_{j+1/2}, \quad \forall j, \tag{B.2}
\]

whose slopes $(U_x)_j$ are obtained using a nonlinear limiter, say, the generalized minmod one (see, e.g., \cite{23,30,35,37}):

\[
(U_x)_j = \text{minmod} \left( \theta \frac{U_{j+1} - U_j}{\Delta x}, \frac{U_{j+1} - U_{j-1}}{2\Delta x}, \theta \frac{U_j - U_{j-1}}{\Delta x} \right), \tag{B.3}
\]

where the minmod function is defined by

\[
\text{minmod}(z_1, z_2, \ldots) := \begin{cases} 
\min(z_1, z_2, \ldots), & \text{if } z_i > 0 \; \forall i, \\
\max(z_1, z_2, \ldots), & \text{if } z_i < 0 \; \forall i, \\
0, & \text{otherwise},
\end{cases} \tag{B.4}
\]

and the parameter $\theta \in [1,2]$ controls the amount of numerical dissipation: The larger the $\theta$ the smaller the numerical dissipation.

The global solution $\tilde{U}(x)$ is, in general, discontinuous at the interface points $\{x_{j+1/2}\}$. The discontinuities propagate with right- and left-sided local speeds, which, for example, can be esti-
mated by the smallest and largest eigenvalues $\lambda_{\text{min}} < \ldots < \lambda_{\text{max}}$ of the Jacobian $\frac{\partial F}{\partial U}$. Namely,

$$
a_{j+\frac{1}{2}}^+ = \max \left\{ \lambda_{\text{max}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^- \right) \right), \lambda_{\text{max}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^+ \right) \right), 0 \right\},
$$

$$
a_{j+\frac{1}{2}}^- = \min \left\{ \lambda_{\text{min}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^- \right) \right), \lambda_{\text{min}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^+ \right) \right), 0 \right\},
$$

where

$$U_{j+\frac{1}{2}}^+ = U_j + \left( U_x \right)_j \frac{\Delta x}{2}, \quad U_{j+\frac{1}{2}}^- = U_j - \left( U_x \right)_j \frac{\Delta x}{2}
$$

are the corresponding right and left values of the reconstruction (B.2).

Using the above construction we obtain the semi-discrete central-upwind scheme (see, e.g., [17, 18]):

$$
\frac{d}{dt} U_j = - \frac{H_{j+\frac{1}{2}} - H_{j-\frac{1}{2}}}{\Delta x} + R_j,
$$

where the numerical fluxes $H_{j+\frac{1}{2}}$ are given by

$$
H_{j+\frac{1}{2}} := \frac{a_{j+\frac{1}{2}}^+ F(U_{j+\frac{1}{2}}^-) - a_{j+\frac{1}{2}}^- F(U_{j+\frac{1}{2}}^+)}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} + \frac{a_{j+\frac{1}{2}}^+ a_{j+\frac{1}{2}}^-}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} \left[ U_{j+\frac{1}{2}}^+ - U_{j+\frac{1}{2}}^- \right]
$$

with $a_{j+\frac{1}{2}}^\pm$ and $U_{j+\frac{1}{2}}^\pm$ computed according to (B.5) and (B.6), respectively.

**Remark 2.1** If the $i$th component of $F$ is identically zero, then we replace the numerical flux (B.8) for this component with

$$
H_{j+\frac{1}{2}}^{(i)} = 0, \quad \forall j.
$$

The cell average of the source term $R_j$ in (B.7) is approximated by an appropriate quadrature for $\frac{1}{\Delta x} \int_{t_i}^t R(U(x,t), x, t) \, dx$. The choice of the quadrature rule is very important especially in cases when the underlying balance law admits steady state solutions. In such cases, one needs to design a so-called well-balanced scheme — the scheme that exactly preserves the appropriate steady states on the discrete level. We refer the reader to [2, 16, 19] and references therein for a review of some well-balanced methods.

**Remark 2.2** The semi-discretization (B.7) is a system of time dependent ODEs, which should be solved by a sufficiently accurate and stable ODE solver.

**References**


