An operator splitting based stochastic Galerkin method for the one-dimensional compressible Euler equations with uncertainty

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

Abstract

We introduce an operator splitting based stochastic Galerkin method for the one-dimensional compressible Euler 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 system of hyperbolic conservation laws do not necessarily yield globally hyperbolic systems: the Jacobian may contain complex eigenvalues and thus trigger instabilities and ill-posedness. In this paper, we propose to split the underlying system into a linear hyperbolic system and two effectively scalar linear or nonlinear hyperbolic equations with variable coefficients and source terms. The gPC-SG method, when applied to each of these subsystems, results in globally hyperbolic systems. The performance of the new gPC-SG method is illustrated with a number of numerical examples with uncertainties from the initial data or equation of state.

Key words. Uncertainty quantification, compressible Euler equations, stochastic Galerkin methods, polynomial chaos, random input, central-upwind schemes, operator splitting.

AMS subject classifications. 35R60, 35L65, 35Q31, 65M08, 76M12, 65M60, 65M70.

1 Introduction

The compressible Euler equations is a system of nonlinear hyperbolic systems of conservation laws. Its equation of state is usually empirical, thus may contain uncertainties. Uncertainties can also 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.

In recent years, one has seen increasing activities and many advances in solving partial differential equations with uncertainties. Among the most popular methods for such problems is the polynomial chaos or generalized polynomial chaos (gPC) approach [8,10,12]. There are two types of polynomial chaos methods: intrusive and non-intrusive ones. The intrusive methods

∗Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA (cher-
tock@math.ncsu.edu).
†Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, USA (jin@math.wisc.edu) and Institute of Natural Sciences, Department of Mathematics, 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).
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. The non-intrusive methods solve the original problem at selected sampling points, thus one can use the same code as the deterministic problem, and then use interpolation and quadrature rules to numerically evaluate the statistical moments [26, 27, 41]. A comparison of computational costs of the two approaches for the diffusion equation with random coefficients was carried out in [7], showing a lower computational cost for the stochastic Galerkin method. The stochastic Galerkin method also has theoretical advantages since it is based on the Galerkin framework. In this paper, we develop a new generalized polynomial chaos stochastic Galerkin (gPC-SG) method for the one-dimensional compressible Euler equations. Its extension to multidimensional Euler equations is straightforward.

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 of conservation and balance laws, nevertheless, encounters major difficulties. For linear hyperbolic systems and scalar hyperbolic conservation laws, the gPC-SG approximation yields a system for the gPC coefficients which is always hyperbolic, thus gPC-SG methods are viable methods for uncertainty quantification (UQ) for such equations [40]. The gPC-SG approximation also remains hyperbolic if the original system is symmetric, see, for instance, [12], where the gPC-SG method has been applied to a symmetric hyperbolic system obtained by differentiating the Hamilton-Jacobi equation in space. However, when applied to general nonlinear (non-symmetric) hyperbolic systems, gPC-SG methods result in systems for the gPC coefficients, which are not necessarily globally hyperbolic [6] since their Jacobian matrices may contain complex eigenvalues. This is similar to the issue of losing hyperbolicity in Grad’s thirteen moment closure of the Boltzmann equation [11]. Consequently, extra efforts are needed in order to obtain well-behaved discrete systems. One approach is to use gPC approximations for entropy variables [31]. In this method, 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 expensive for large scale problems. Another simpler approach is to use the Roe variables as it was proposed in [29]; see also [30]. Using a Roe-type solver for symmetrizable hyperbolic systems was also suggested in [36]. These methods are restricted to systems which admit the Roe linearization and require 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 this paper, we present a splitting strategy for the one-dimensional (1-D) compressible Euler equations such that the gPC-SG approximation of each of the split subsystems always results in a globally hyperbolic system for the corresponding gPC coefficients. This strategy is generic, although the nonlinear system studied in this paper is specific (in [5] we will propose a similar splitting for the Saint-Venant system of shallow water equations). Our main idea is to split the underlying hyperbolic system into a linear hyperbolic system and linear or nonlinear scalar equations with variable coefficients and source terms. It is well-known that for linear hyperbolic systems and scalar conservation laws, gPC-SG methods yield globally hyperbolic systems. Thus, our splitting approach provides a reliable and stable gPC-SG method, which is simple and generic, and hence attractive for practical applications.

In gPC-SG methods, the solution of the underlying system is sought in terms of orthogonal
polynomial series \cite{42}, 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 numerically solve it by a finite-volume Godunov-type method. In particular, we use a semi-discrete second-order central-upwind scheme, which has been originally introduced in \cite{16,17,19} as a Riemann-problem-solver-free black box solver for general multidimensional hyperbolic systems of conservation laws. One of the advantages of the central-upwind schemes is their direct applicability to systems with complicated (generalized) Riemann problem solutions including the cases of conservation laws with variable coefficients.

Here, we shall not try to address the numerical difficulties of dealing with discontinuities in the random space using the gPC approximation, which may trigger the Gibbs phenomenon in the random space. It is well-known that in this case one should use piecewise approximations such as, for example, a multi-element gPC \cite{13,38,39}, a wavelet basis \cite{20,21} and essentially non-oscillatory (ENO) or weighted essentially non-oscillatory (WENO) interpolations \cite{1}, which can sometimes be coupled with local subspace recovery techniques \cite{4} or multiresolution approaches \cite{2,35}.

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 conservation laws. In Section 3 we describe the proposed splitting approach for the compressible Euler equations, and numerically demonstrate that, for deterministic problems, such a splitting provides a simply and viable framework for deriving shock capturing schemes. The gPC-SG approximations of each of the resulting split subsystems are presented in Section 4. Finally, in Section 5, we provide several numerical examples to illustrate the performance of the proposed method. The paper is concluded in Section 6.

2 The gPC-SG method – an overview

In this section, we briefly describe the gPC-SG method for the 1-D hyperbolic system of conservation laws

$$U_t + F(U, x, z)_x = 0, \quad x \in \mathbb{R}, \ t > 0, \ z \in \Omega \subset \mathbb{R}^d,$$  \hspace{1cm} (2.1)

where the unknown vector function $U = U(x, t, z)$ depends on the spatial variable $x$, time $t$ and random variable $z$. In the gPC expansion, the solution of the system of stochastic PDEs (2.1) is sought in terms of an orthogonal polynomial series in $z$ (see, e.g., \cite{42}):

$$U(x, t, z) \approx U_N(x, t, z) = \sum_{i=0}^{M-1} \tilde{U}_i(x, t) \Phi_i(z), \quad M = \binom{d+N}{d},$$  \hspace{1cm} (2.2)

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

$$\int_{\Omega} \Phi_i(z) \Phi_\ell(z) \mu(z) dz = \delta_{i\ell}, \quad 0 \leq i, \ell \leq M - 1, \ M = \dim(\mathbb{P}_N^d).$$  \hspace{1cm} (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., [40, 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

\[
(U_i)_t + (F_i)_x = 0, \quad 0 \leq i \leq M - 1,
\]

where

\[
F_i = \int_{\Omega} \left( \sum_{j=0}^{M-1} \hat{U}_j(x,t)\Phi_j(z), x, z \right) \Phi_i(z) \mu(z) \, dz, \quad 0 \leq i \leq M - 1.
\]

This is a system of deterministic equations for the expansion coefficients $U_i$, $i = 0, \ldots, M - 1$. In most cases, the equations in (2.4) are coupled.

For linear hyperbolic systems, (2.4) is a vectorized version of the original system (2.1) and thus it remains hyperbolic. For nonlinear symmetric hyperbolic systems—including scalar hyperbolic conservation laws—the Jacobian of the flux is a symmetric matrix and thus the gPC system (2.4) is hyperbolic as well. However, if the nonlinear hyperbolic system (2.1) is not symmetric, (2.4) is not always hyperbolic since its Jacobian may encounter complex eigenvalues, see such examples in [6]. 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 the next section, we shall present a way to overcome this difficulty by introducing operator splittings for the compressible Euler equations, which can be written as system (2.1), that will guarantee that the gPC-SG discretization of each of the split subsystems always results in a globally hyperbolic system for the expansion coefficients. Although in this paper we present such a splitting for the compressible Euler equations only, our splitting strategy is generic, and a similar splitting for the Saint-Venant system of shallow water equations will be presented in a subsequent paper [5].

3 A splitting of the compressible Euler equations

In this section, we consider the 1-D compressible Euler equations:

\[
\begin{align*}
\rho_t + m_x &= 0, \\
m_t + (\rho u^2 + p)_x &= 0, \\
E_t + (u(E + p))_x &= 0,
\end{align*}
\]

where $\rho$ is the density, $u$ is the velocity, $m := \rho u$ is the momentum, $p$ is the pressure, and $E$ is the total energy. For a polytropic gas, the equation of state is given by

\[
p = (\gamma - 1) \left( E - \frac{1}{2} \rho u^2 \right)
\]
with $\gamma$ being the specific heat ratio.

We assume here that the initial data and the equation of state may depend on a random variable $z$, that is,

$$
\rho(x,0,z) = \rho_0(x,z), \quad u(x,0,z) = u_0(x,z), \quad p(x,0,z) = p_0(x,z), \quad \gamma = \gamma(z). \quad (3.3)
$$

Uncertainty may also arise from boundary data and other terms (such as source terms, a case treated in [5]).

The Euler equations is a hyperbolic system that has three distinct characteristic speeds $u, u\pm c$ with $c := \sqrt{\gamma p/\rho}$ being the speed of sound, but as it has been mentioned, a direct application of the gPC-SG method to the system (3.1) may fail due to the loss of hyperbolicity after the gPC-SG discretization. In order to overcome this difficulty we introduce an operator splitting technique. The proposed splitting consists of a linear system and two nonlinear and linear scalar equations with variable coefficients and thus the gPC-SG approximation is guaranteed to maintain the hyperbolicity for each of the subsystems.

### 3.1 An operator splitting

Our main idea is to split the Euler system (3.1) into the following three subsystems:

$$
\begin{aligned}
\rho_t + m_x &= 0, \\
m_t + ((\gamma - 1)E + am)_x &= 0, \\
E_t - aE_x &= 0,
\end{aligned}
$$

(3.4)

and

$$
\begin{aligned}
\rho_t &= 0, \\
m_t + \left(\frac{3 - \gamma}{2} \cdot \frac{m^2}{\rho} - am\right)_x &= 0, \\
E_t &= 0,
\end{aligned}
$$

(3.5)

and

$$
\begin{aligned}
\rho_t &= 0, \\
m_t &= 0, \\
E_t + \left(\frac{m}{\rho} \cdot \left[\gamma E - \frac{\gamma - 1}{2} \cdot \frac{m^2}{\rho}\right] + aE\right)_x &= 0.
\end{aligned}
$$

(3.6)

The first system (3.4) is linear hyperbolic with three distinct characteristic speeds $0, \pm a$. The parameter $a$ will be chosen in such a way that the eigenvalues of the Jacobian of the first subsystem (3.4) satisfy the following subcharacteristic condition:

$$
-|a| \leq u - c < u + c \leq |a|. \quad (3.7)
$$

Furthermore, we choose $a$ such that convection coefficients (and consequently the characteristic speeds $(3 - \gamma)u$ for (3.5) and $\gamma u$ for (3.6)) in the second and third equations of the second and third subsystems (3.5) and (3.6), respectively, do not change sign. (When they change signs they
could bring analytical and numerical difficulties; see, e.g., [3]). This yields the following choice:

\[ a = \pm \sup_u \left\{ \max \left( |u| + c, |u| (3 - \gamma) \right) \right\} . \]

Here, the sign of \( a \) is not crucial, and we alternate it at each splitting step as the solution is numerically evolved in time. One can also select the sign of \( a \) in the beginning of each splitting step randomly.

It is easy to check that under the subcharacteristic condition (3.7) each of the subsystem in the above splitting is strictly hyperbolic (their Jacobian matrices of the fluxes are diagonalizable in the real space). The second system (3.5) is essentially a scalar Burgers equation for \( m \) with variable coefficient and source term, since \( \rho \) remains constant in time in (3.5). In the third system (3.6), \( \rho \) and \( m \) do not change in time and thus the last equation in (3.6) for \( E \) is again a scalar (and, in fact, linear) hyperbolic equation with variable coefficients and a source term. Therefore, after the gPC-SG approximation, each of the above systems gives a globally hyperbolic system for the gPC coefficients.

### 3.2 Numerical validation of the operating splitting

Before proceeding further with the derivation of the gPC-SG method for the systems (3.4)–(3.6), it is helpful to check the performance of the hyperbolic splitting described in the previous section on a purely deterministic Euler equations, in which no data depend on a random variable \( z \). To this end, we numerically solve the system (3.1) using two different methods.

We first directly apply the second-order semi-discrete central-upwind scheme from [17] (also briefly described in Appendix A for general systems of conservation/balance laws) to the original system (3.1), which can be written in the vector form (2.1) with

\[
U = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad F = \begin{pmatrix} m \\ \rho u^2 + p \\ u(E + p) \end{pmatrix}.
\]

The semi-discrete central-upwind discretization of the system (3.1) is thus obtained from (A.7), (A.8) with the cell averages \( \overline{U}_j = (\overline{\rho}_j, \overline{m}_j, \overline{E}_j)^T \) defined in (A.1), the point values \( U_{j+\frac{1}{2}}^\pm = (\rho_{j+\frac{1}{2}}^\pm, m_{j+\frac{1}{2}}^\pm, E_{j+\frac{1}{2}}^\pm)^T \) computed according to (A.6) and the one-sided local speeds of propagation (A.5) given by

\[
a_{j+\frac{1}{2}}^+ = \max \left\{ \frac{u_{j+\frac{1}{2}}^- + c_{j+\frac{1}{2}}^-}{\rho_{j+\frac{1}{2}}^-}, \frac{u_{j+\frac{1}{2}}^+ + c_{j+\frac{1}{2}}^+}{\rho_{j+\frac{1}{2}}^+}, 0 \right\}, 
\]

\[
a_{j+\frac{1}{2}}^- = \min \left\{ \frac{u_{j+\frac{1}{2}}^- - c_{j+\frac{1}{2}}^-}{\rho_{j+\frac{1}{2}}^-}, \frac{u_{j+\frac{1}{2}}^+ - c_{j+\frac{1}{2}}^+}{\rho_{j+\frac{1}{2}}^+}, 0 \right\},
\]

where

\[
u_{j+\frac{1}{2}}^\pm = \frac{m_{j+\frac{1}{2}}^\pm}{\rho_{j+\frac{1}{2}}^\pm}, \quad c_{j+\frac{1}{2}}^\pm = \sqrt{\frac{\gamma p_{j+\frac{1}{2}}^\pm}{\rho_{j+\frac{1}{2}}^\pm}}.
\]

For the splitting approach, we use the second-order Strang operator splitting method (B.3) with \( K = 3 \) (see Appendix B). Each of the subsystems (3.4), (3.5) and (3.6) is again solved using the second-order semi-discrete central-upwind scheme described in Appendix A. The first
system (3.4) is a linear system of conservation laws of the form (2.1) with

\[ U = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad F = \begin{pmatrix} m \\ (\gamma - 1)E + am \\ -aE \end{pmatrix}, \tag{3.9} \]

for which the central-upwind scheme (A.7), (A.8) applied with \( a^+_{j+\frac12} = -a^-_{j+\frac12} = a \) \( \forall j \) reads

\[ \frac{d}{dt} U_j = -\frac{H_{j+\frac12} - H_{j-\frac12}}{\Delta x}, \quad H_{j+\frac12} = \frac{F(U^-_{j+\frac12}) + F(U^+_{j+\frac12})}{2} - \frac{a}{2} \left[ U^+_{j+\frac12} - U^-_{j+\frac12} \right]. \tag{3.10} \]

The second system (3.5) is actually reduced to a scalar hyperbolic conservation law for \( m \) with

\[ U = m, \quad F = \frac{3 - \gamma}{2} \cdot \frac{m^2}{\rho} - am, \tag{3.11} \]

and the third system (3.6) is reduced to a linear equation for \( E \) with

\[ U = E, \quad F = \frac{m}{\rho} \left[ \gamma E - \frac{\gamma - 1}{2} \cdot \frac{m^2}{\rho} \right] + aE. \tag{3.12} \]

Both scalar equations (3.11) and (3.12) are numerically solved by the central-upwind scheme (A.7), (A.8) with \( m^\pm_{j+\frac12} \) and \( m^\pm_{j+\frac12} \) computed according to (A.6) and the local speeds \( a^\pm_{j+\frac12} \) taken exactly the same as in (3.8).

Once the two numerical approaches (the unsplit and split ones) are designed, we test them on the Sod shock tube problem, in which the system (3.1) is considered subject to the following initial conditions:

\[ \rho_0(x) = \begin{cases} 1, & x < 0.5, \\ 0.125, & x > 0.5, \end{cases} \quad u_0(x) \equiv 0, \quad p_0(x) = \begin{cases} 1, & x < 0.5, \\ 0.1, & x > 0.5. \end{cases} \tag{3.13} \]

The solutions computed on a uniform grid with \( \Delta x = 1/400 \) at time \( t = 0.1644 \), are plotted in Figure 3.1, which shows a good agreement between the two approaches.

This example shows that our splitting provides a simple and useful way to build a shock capturing scheme for the (deterministic) compressible Euler equations, providing support to use this approach for the Euler equations with uncertainty.

### 4 The gPC-SG approximation

We define the gPC expansions of \( \rho, m, E \) and \( \gamma \) in the following form:

\[
\rho_N(x,t,z) = \sum_{i=0}^{N} \hat{\rho}_i(x,t) \Phi_i(z), \quad m_N(x,t,z) = \sum_{i=0}^{N} \hat{m}_i(x,t) \Phi_i(z), \quad E_N(x,t,z) = \sum_{i=0}^{N} \hat{E}_i(x,t) \Phi_i(z), \quad \gamma_N(z) = \sum_{i=0}^{N} \hat{\gamma}_i \Phi_i(z), \tag{4.1}
\]
and use them to derive a gPC-SG scheme for the systems (3.4)–(3.6). To this end, we substitute (4.1) into each of the split subsystems (3.4)–(3.6) and after implementing the Galerkin projection we obtain the corresponding three systems for the gPC coefficients:

\[
\begin{align*}
(\hat{\rho}_i)_t + (\hat{m}_i)_x &= 0, \\
(\hat{m}_i)_t + a(\hat{m}_i)_x + \sum_{k,\ell=0}^{N} \hat{\gamma}_k (\hat{E}_\ell)_x S_{k\ell i} &= 0, \\
(\hat{E}_i)_t - a(\hat{E}_i)_x &= 0, \\
(\hat{\rho}_i)_t &= 0, \\
(\hat{m}_i)_t + \sum_{k,\ell=0}^{N} \hat{\gamma}_k (\hat{\psi}_\ell)_x S_{k\ell i} - a(\hat{m}_i)_x &= 0, \\
(\hat{E}_i)_t &= 0,
\end{align*}
\]

(4.2) (4.3)
\begin{align}
\begin{cases}
(\hat{\rho}_i)_t & = 0, \\
(\hat{m}_i)_t & = 0, \\
(\hat{E}_i)_t + a(\hat{E}_i)_x + \sum_{k,\ell=0}^N (\hat{\psi}_k \hat{E}_\ell)_x S_{k\ell i} - \sum_{k,\ell=0}^N (\hat{\psi}_k^\dag \hat{\psi}_\ell)_x S_{k\ell i} &= 0,
\end{cases}
\end{align}

(4.4)

where \( S_{k\ell i} \) is given by

\[ S_{k\ell i} = \int_\Omega \Phi_k(z) \Phi_\ell(z) \Phi_i(z) \mu(z) \, dz \text{ for all } k,\ell,i = 0,\ldots,N. \]

(4.5)

In the above equations, we have used the following definition:

\[ \gamma_N(z) - 1 = \sum_{i=0}^N \hat{\gamma}_i \Phi_i(z), \quad \frac{3 - \gamma_N(z)}{2} = \sum_{i=0}^N \frac{3}{\hat{\gamma}_i} \Phi_i(z), \quad \left( \frac{m^2}{\hat{\rho}} \right)_N (x,t,z) = \sum_{i=1}^N \hat{\psi}_i(x,t) \Phi_i(z), \]

\[ \frac{\hat{\gamma} m}{\rho} (x,t,z) = \sum_{i=1}^N \hat{\psi}_i(x,t) \Phi_i(z), \quad \frac{\left( \gamma - 1 \right) m}{\rho} (x,t,z) = \sum_{i=1}^N \hat{\psi}_i(x,t) \Phi_i(z). \]

The gPC coefficients \( \hat{\psi}_i \) needed in (4.3) and (4.4) can be computed by applying the gPC-SG approximation to the relation \( \rho \psi = m^2 \):

\[ \sum_{k,\ell=0}^N \hat{\psi}_k \hat{\rho}_\ell S_{k\ell i} = \sum_{k,\ell=0}^N \hat{m}_k \hat{m}_\ell S_{k\ell i}, \quad i = 0,\ldots,N, \]

(4.6)

and once \( \{ \hat{\rho}_i \} \) and \( \{ \hat{m}_i \} \) are given, one can solve the linear system (4.6) to obtain \( \{ \hat{\psi}_i \} \). Similarly, the gPC coefficients \( \{ \hat{\psi}_i \} \) and \( \{ \hat{\psi}_i \} \) that appear in (4.4) can be obtained by solving the linear systems

\[ \sum_{k,\ell=0}^N \hat{\psi}_k \hat{\rho}_\ell S_{k\ell i} = \sum_{k,\ell=0}^N \hat{\gamma}_k \hat{m}_\ell S_{k\ell i}, \quad i = 0,\ldots,N, \]

(4.7)

and

\[ \sum_{k,\ell=0}^N \hat{\psi}_k \hat{m}_\ell S_{k\ell i} = \sum_{k,\ell=0}^N \hat{\psi}_k \hat{m}_\ell S_{k\ell i}, \quad i = 0,\ldots,N, \]

(4.8)

respectively.

**Remark 4.1** We are not able to prove the invertibility of the linear systems (4.6)–(4.8). Here, we argue that they are likely invertible under certain positivity and smoothness assumptions. Take (4.6) as an example. When applying the gPC-SG approximation to the relation \( \rho \psi = m^2 \) for the exact value of \( \rho > 0 \), one would obtain the following linear system for \( \{ \hat{\psi}_i \} \):

\[ \sum_{k=0}^N \hat{\psi}_k \int_\Omega \rho(z) \Phi_i(z) \Phi_k(z) \mu(z) \, dz = \sum_{k,\ell=0}^N \hat{m}_k \hat{m}_\ell S_{k\ell i}, \quad i = 0,\ldots,N. \]

(4.9)

For \( \rho > 0 \), the coefficient matrix on the left-hand side of the linear system (4.9) is symmetric positive definite (see, e.g., [43]). Thus, the system (4.9) is invertible. The linear system (4.6)
is obtained by approximating $\rho \approx \rho_N = \sum_{\ell=0}^N \hat{\rho}_\ell$. Thus, the entries in the coefficient matrix in (4.6) are just the gPC approximation of those in (4.9), and they are spectrally close to each other for smooth solutions so the corresponding coefficient matrix in (4.6) is likely invertible. In particular, (4.6) is invertible if $\rho_N > 0$. In the general case, there is no guarantee of the positivity of $\rho_N$ and there is a lack of analysis on the invertibility of the linear systems (4.6)–(4.8). Nevertheless, we have never encountered any problems in our numerical simulations.

For the spatial discretizations of the systems (4.2)–(4.4), we again apply the second-order semi-discrete central-upwind scheme described in Appendix A. Here, for each $i = 0, \ldots, N$ in (4.2), we have a linear hyperbolic system of the form (2.1) with

$$ U_i = \begin{pmatrix} \hat{\rho}_i \\ m_i \\ E_i \end{pmatrix}, \quad F_i = \begin{pmatrix} m_i \\ \sum_{k,\ell=0}^N \hat{\gamma}_k(\hat{E}_\ell)S_{k\ell i} \\ -aE_i \end{pmatrix}, $$

and thus apply the semi-discrete scheme (3.10) to each one of the $N+1$ equations.

The second system (4.3) is a set of $N+1$ decoupled scalar hyperbolic conservation laws for $\hat{m}_i$, which still can be put into the form (2.1) with

$$ U_i = \hat{m}_i, \quad F_i = \sum_{k,\ell=0}^N \hat{\gamma}_k \hat{\psi}_\ell S_{k\ell i} - a\hat{m}_i, $$

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

$$ a_{j+\frac{1}{2}}^+ = \max \left\{ \hat{u}_{j+\frac{1}{2}}^- + \hat{c}_{j+\frac{1}{2}}^-, \hat{u}_{j+\frac{1}{2}}^+ + \hat{c}_{j+\frac{1}{2}}^+, 0 \right\}, $$

$$ a_{j+\frac{1}{2}}^- = \min \left\{ \hat{u}_{j+\frac{1}{2}}^- - \hat{c}_{j+\frac{1}{2}}^-, \hat{u}_{j+\frac{1}{2}}^+ - \hat{c}_{j+\frac{1}{2}}^+, 0 \right\}. \quad (4.10) $$

Here, we have used the following notation:

$$ \hat{u}_{j+\frac{1}{2}}^\pm := \frac{(\hat{m}_0)_{j+\frac{1}{2}}^\pm}{(\hat{m}_0)_{j+\frac{1}{2}}^\pm}, \quad \hat{c}_{j+\frac{1}{2}}^\pm := \frac{\hat{\gamma}_0 \hat{p}_{j+\frac{1}{2}}^\pm}{\hat{m}_0_{j+\frac{1}{2}}^\pm}, \quad \hat{\rho}_{j+\frac{1}{2}}^\pm := \hat{\gamma}_0 \left[ (\hat{E}_0)_{j+\frac{1}{2}}^\pm - \frac{(\hat{m}_0)_{j+\frac{1}{2}}^\pm (\hat{v})_{j+\frac{1}{2}}^\pm}{2} \right]. $$

Finally, the third system (4.4) is a set of $N+1$ decoupled scalar linear hyperbolic equations for $\hat{E}_i$ with

$$ U_i = \hat{E}_i, \quad F_i = a\hat{E}_i + \sum_{k,\ell=0}^N \hat{\gamma}_k \hat{\psi}_\ell S_{k\ell i} - \sum_{k,\ell=0}^N \hat{\gamma}_k \hat{\psi}_\ell S_{k\ell i}, $$

each of which is solved by the central-upwind scheme (A.7), (A.8) with the local speeds given by (4.10).

**Remark 4.2** In the case of deterministic $\gamma \neq \gamma(z)$, the gPC-SG approximation becomes substantially simpler. We omit the details for the sake of brevity.
5 Numerical experiments

In this section, we illustrate the performance of the new gPC-SG method on a number of numerical examples. Throughout this section, we consider the Sod shock tube problem \((3.1), (3.13)\) with the randomness/uncertainty, which can enter through either the initial data (Example 1), adiabatic coefficient \(\gamma\) (Example 2) or initial interface location (Example 3). 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 highest degree in the gPC expansion is \(N = 8\)). We would like to point out that the mathematical formulation and numerical methods work for \(z\) in any dimension. The mean and standard deviation of the computed solution \(U\), which are shown in the Figures below, are given by

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

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

In all of the examples below, the second-order semi-discrete central-upwind scheme \((A.7), (A.8)\) was implemented for the spatial discretization on uniform grids, the piecewise linear interpolants were constructed using the minmod limiter \((A.3), (A.4)\) with \(\theta = 1.3\), and the arising systems of ODEs were numerically integrated by the third-order strong stability-preserving (SSP) Runge-Kutta method \([9][10]\). All of the experiments were conducted in the computation domain \([0, 1]\) with nonreflecting boundary conditions and the results will be shown at time \(t = 0.1644\).

Example 1 – Perturbed Initial Data. We first consider the Sod shock tube problem with perturbed initial data, that is, the system \((3.1)\) with \(\gamma = 1.4\) and subject to the following initial condition:

\[
\rho_0(x, z) = \begin{cases} 
1 + 0.1z, & x < 0.5, \\
0.125, & x > 0.5,
\end{cases} \quad u_0(x) \equiv 0, \quad p_0(x) = \begin{cases} 
1, & x < 0.5, \\
0.1, & x > 0.5,
\end{cases}
\]

and solve it numerically by the gPC-SG method. The mean and standard deviation of \(\rho, m\) and \(E\) computed on the two different grids with \(\Delta x = 1/200\) and \(\Delta x = 1/800\), are plotted in Figures 5.1. The results demonstrate a good agreement (everywhere except for the left part of the rarefaction area) with the deterministic solution shown in Figure 3.1: the time evolution of the right shock and the contact wave is correctly captured by the proposed gPC-SG approach.

Example 2 – Perturbed \(\gamma\). In the next example, we analyze the effect of uncertainty in the adiabatic coefficient \(\gamma\). To this end, we consider the original Sod shock tube problem \((3.1), (3.13)\) with random coefficient \(\gamma\), that is,

\[
\gamma(z) = 1.4 + 0.1z,
\]

and solve it numerically by the gPC-SG method. A similar problem was considered in \([35]\). The mean and standard deviation of \(\rho, m\) and \(E\) computed on the two different grids with \(\Delta x = 1/200\) and \(\Delta x = 1/800\), are plotted in Figure 5.2. In this case, the uncertainty in the sound velocity
Figure 5.1: Example 1: Mean (left) and standard deviation (right) of $\rho$ (top row), $m$ (middle row) and $E$ (bottom row) computed by the gPC-SG method on different grids.

affects the propagation of the shock, contact discontinuity and rarefaction wave. Although the behavior of the computed solution is similar to the purely deterministic case (Figure 3.1), the positions of the discontinuities are different. The spreading of the location for both the shock and contact discontinuity is clearly visible, while the impact of the uncertain sound speed on the
(smooth) rarefaction wave is less pronounced.

Figure 5.2: Example 2: Mean (left) and standard deviation (right) of $\rho$ (top row), $m$ (middle row) and $E$ (bottom row) computed by the gPC-SG method on different grids.

Example 3 – A Perturbed Interface. In the last experiment, we consider the Sod shock tube problem with uncertainty carried on the initial interface position between the light and
heavy fluids. We consider the system (3.1) with $\gamma = 1.4$ and subject to the following initial condition:

$$
\rho_0(x, z) = \begin{cases} 
1, & x < 0.5 + 0.05z, \\
0.125, & x > 0.5 + 0.05z,
\end{cases} 
\quad u_0(x) \equiv 0, 
\quad p_0(x) = \begin{cases} 
1, & x < 0.5 + 0.05z, \\
0.1, & x > 0.5 + 0.05z.
\end{cases}
$$

This problem was considered in [31].

As before, we numerically solve the problem numerically by the gPC-SG method and plot the mean and standard deviation of $\rho$, $m$ and $E$ computed on the two different grids with $\Delta x = 1/200$ and $\Delta x = 1/800$, see Figure 5.3. As one can see, the computed means are in a good agreement with the deterministic solution (Figure 3.1): the corresponding positions of the three waves are quite well captured while of course affected by the uncertainty. The locations of the three sharp fronts corresponding to the shock, contact discontinuity and rarefaction fan, coincide with regions of large standard deviation.

6 Conclusion

In this paper, we have studied generalized polynomial chaos stochastic Galerkin (gPC-SG) methods for the 1-D compressible Euler equations with random inputs. It is well known that a direct application of gPC-SG methods to nonlinear hyperbolic system of conservation laws results in nonlinear systems of PDEs for the gPC coefficients. These systems are not necessarily globally hyperbolic and this may trigger instability. We have developed a new approach, which guarantees that the obtained gPC systems are hyperbolic. Our method is based on a special operator splitting: we split the Euler system into the linear system and two scalar equations with variable coefficients, for which the corresponding gPC systems are known to be hyperbolic.

We have conducted several numerical experiments with uncertainties introduced in either the initial data or equation of state. The obtained results demonstrate that our gPC-SG method is capable of accurately capturing both the mean and standard deviation of the studied solutions. However, more work needs to be done to enhance the proposed method and make it robust so that it would be able to handle problems with waves containing strong discontinuities in the random space. One also needs to establish a solid theoretical foundation for this method as well as generalize it for other hyperbolic systems of conservation and balance laws and for the multidimensional problems. These tasks stay beyond the scope of this paper and are left for future development.

Acknowledgments

The work of A. Chertock was supported in part by the NSF grants DMS-1115682, DMS-1216974 and DMS-1521051 and the NSF RNMS "Ki-net" Grant DMS-1107444. The work of S. Jin was partially supported by the NSF grants DMS-1522184 and DMS-1107291: RNMS "KI-Net", by the NSFC grant 91330203, and by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. The work of A. Kurganov was supported in part by the NSF grants DMS-1115718, DMS-1216957 and DMS-1521009.
Figure 5.3: Example 3: Mean (left) and standard deviation (right) of $\rho$ (top row), $m$ (middle row) and $E$ (bottom row) computed by the gPC-SG method on different grids.
A The semi-discrete second-order central-upwind scheme

In this section, we briefly describe the central-upwind schemes for 1-D hyperbolic systems of conservation and balance laws. These schemes are Godunov-type finite-volume methods. For their complete description and derivation, we refer the reader to [15–18].

We consider a 1-D system of conservation/balance laws (2.1) and divide the computational domain into the uniform cells

\[ C_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \]

of size \( |C_j| = \Delta x \) centered at points \( x_j = j\Delta x, \quad j = j_L, \ldots, j_R \). We assume that at certain time \( t \) the cell averages of the computed solution,

\[ \bar{U}_j(t) = \frac{1}{\Delta x} \int_{C_j} U(x,t) \, dx \quad (A.1) \]

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 (A.1) to reconstruct a non-oscillatory piecewise linear polynomial,

\[ \tilde{U}(x) = \bar{U}_j + (U_x)_j(x - x_j), \quad x_{j-\frac{1}{2}} < x < x_{j+\frac{1}{2}}, \quad \forall j, \quad (A.2) \]

where the slopes \( (U_x)_j \) are obtained using a nonlinear limiter, say, the generalized minmod one (see, e.g., [23,28,34,37]):

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

with the minmod function is defined by

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

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 \( \{x_{j+\frac{1}{2}}\} \). The discontinuities propagate with right- and left-sided local speeds, which, for example, can be estimated by the smallest and largest eigenvalues \( \lambda_{\text{min}} < \ldots < \lambda_{\text{max}} \) of the Jacobian \( \frac{\partial F}{\partial U} \). Namely,

\[ a_{j+\frac{1}{2}}^+ = \max\left\{ \lambda_{\text{max}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^- \right) \right), \lambda_{\text{max}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^+ \right) \right), 0 \right\}, \quad (A.5) \]

\[ a_{j+\frac{1}{2}}^- = \min\left\{ \lambda_{\text{min}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^- \right) \right), \lambda_{\text{min}} \left( \frac{\partial F}{\partial U} \left( U_{j+\frac{1}{2}}^+ \right) \right), 0 \right\}, \quad (A.5) \]

where

\[ U_{j+\frac{1}{2}}^+ = \bar{U}_j + (U_x)_j \frac{\Delta x}{2}, \quad U_{j+\frac{1}{2}}^- = \bar{U}_j - (U_x)_j \frac{\Delta x}{2} \quad (A.6) \]

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

Using the above construction we obtain the semi-discrete central-upwind scheme (see, e.g.,
\[ \frac{d}{dt} U_j = -\frac{H_{j+\frac{1}{2}} - H_{j-\frac{1}{2}}}{\Delta x}, \quad (A.7) \]

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

\[ H_{j+\frac{1}{2}} := \frac{a_{j+\frac{1}{2}}^+ F(U_{j+\frac{1}{2}}^+) - a_{j+\frac{1}{2}}^- F(U_{j+\frac{1}{2}}^-)}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} + \frac{a_{j+\frac{1}{2}}^+ a_{j+\frac{1}{2}}^-}{a_{j+\frac{1}{2}}^+ - a_{j+\frac{1}{2}}^-} \left[ (U_{j+\frac{1}{2}}^+ - U_{j+\frac{1}{2}}^-) \right], \quad (A.8) \]

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

**Remark 1.1** If the \( i \)-th component of \( F \) is identically zero, then we replace the numerical flux (A.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 (A.7) is approximated by an appropriate quadrature for \( \frac{1}{\Delta x} \int_{C_j} R(U(x, t), x, t) \, dx \). 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 [15, 18] and the references therein for a review of some well-balanced methods.

**Remark 1.2** The semi-discretization (A.7) is a system of time dependent ODEs, which should be solved by a sufficiently accurate and stable ODE solver.

### B The Operator Splitting

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 several subsystems:

\[ U_t + F_1(U)_x = 0, \quad \ldots, \quad U_t + F_K(U)_x = 0, \quad (B.1) \]

where \( F = F_1 + \cdots + F_K \) and \( S_1, \ldots, S_K \) 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 \( K \) substeps, which result in the following approximate solution at time \( t + \Delta t \):

\[ U(x, t + \Delta t, z) = S_K(\Delta t) \cdots S_1(\Delta t) U(x, t, z). \quad (B.2) \]

In general, if all solutions involved in the splitting algorithm (B.1), (B.2) are smooth, the method is first-order accurate (see, e.g., [24, 25, 32]).

Higher-order operator splitting algorithms can be derived by considering additional substeps. For instance, one time step of the second-order Strang splitting method [24, 25, 32] consists of
$2K - 1$ substeps:

$$U(x, t + \Delta t, z) = S_1(\Delta t/2) \cdots S_{K-1}(\Delta t/2) S_K(\Delta t) S_{K-1}(\Delta t/2) \cdots S_1(\Delta t/2) U(x, t, z). \quad (B.3)$$

All splitting algorithms, presented in this paper, are based on the Strang splitting algorithm \[B.3\]. We also refer the readers to [14,22,33,44], where higher-order operator splitting methods can be found.

In practice, the exact solution operators $S_1, \ldots, S_K$ 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.

References


