Journal of Computational Physics 345 (2017) 224–244

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

A stochastic Galerkin method for first-order quasilinear hyperbolic systems with uncertainty

Kailiang Wu [a], Huazhong Tang [a,b,*], Dongbin Xiu [c]

[a] HEDPS, CAPT & LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, PR China
[b] School of Mathematics and Computational Science, Xiangtan University, Xiangtan 411105, Hunan Province, PR China
[c] Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 16 January 2016
Received in revised form 14 May 2017
Accepted 15 May 2017
Available online 19 May 2017

Keywords:
Uncertainty quantification
Quasilinear hyperbolic system
Stochastic Galerkin methods
Generalized polynomial chaos
Symmetrically hyperbolic
Operator splitting

ABSTRACT

This paper is concerned with generalized polynomial chaos (gPC) approximation for first-order quasilinear hyperbolic systems with uncertainty. The one-dimensional (1D) hyperbolic system is first symmetrized with the aid of left eigenvector matrix of the Jacobian matrix. Then the gPC stochastic Galerkin method is applied to derive a provably symmetrically hyperbolic equations for the gPC expansion coefficients. The resulting deterministic gPC Galerkin system is discretized by a path-conservative finite volume WENO scheme in space and a third-order total variation diminishing Runge–Kutta method in time. The method is further extended to two-dimensional (2D) quasilinear hyperbolic system with uncertainty, where the symmetric hyperbolicity of the one-dimensional gPC Galerkin system is carried over via an operator splitting technique. Several numerical experiments are conducted to demonstrate the accuracy and effectiveness of the proposed gPC stochastic Galerkin method.

1. Introduction

This paper is concerned with uncertainty quantification (UQ) of quasilinear hyperbolic system. UQ has received increasing attention in recent years and found its use in many problems. One of the most widely used UQ methods is generalized polynomial chaos (gPC) [44]. As an extension of the classical polynomial chaos [15], gPC approximates the uncertain solutions as a (truncated) generalized Fourier series by utilizing orthogonal polynomials, and the unknown expansion coefficient functions can be computed by an intrusive or a non-intrusive method. The intrusive methods typically employ stochastic Galerkin (SG) projection, which results in a larger coupled deterministic system of equations for the gPC coefficients. The non-intrusive methods are often of stochastic collocation (SC) type. They solve the original problem at some sampling points of the random variables, and then evaluate the gPC coefficients by using the polynomial interpolation or numerical quadrature, e.g. [32,43,4,27]. For a review of the methods, see [42].

Although the gPC-SG method has been successfully applied to a large variety of problems, its applications to quasilinear hyperbolic system is still quite limited. This is mainly due to a lack of theoretical understanding of the resulting deterministic gPC-SG system. For linear and scalar hyperbolic equations, the resulting gPC-SG systems are still hyperbolic, see cf. [9,16,31,21]. However, for a general quasilinear hyperbolic system, the resulting gPC-SG system may be not globally

* Corresponding author at: School of Mathematical Sciences, Peking University, Beijing 100871, PR China. Fax: +86 10 62751801.
E-mail addresses: wukl@pku.edu.cn (K. Wu), hztang@math.pku.edu.cn (H. Tang), xiu.16@osu.edu (D. Xiu).

http://dx.doi.org/10.1016/j.jcp.2017.05.027
0021-9991/

复特征值

hyperbolic [13]. The lack of hyperbolicity means that the Jacobian matrix may contain complex eigenvalues, which lead to ill-posedness of the initial or boundary problem and instability of the numerical computations. Recently, some efforts were made to obtain well-behaved gPC-SG system for several specific quasilinear hyperbolic systems. Després et al. used the gPC approximation to the entropy variables instead of the conservative variables in the Euler equations [13], and proved that the resulting gPC-SG system is hyperbolic, based on the fact that the Euler equations can be reformulated in a symmetrically hyperbolic form in term of the entropy variables. This method, however, can not be extended to a general quasilinear hyperbolic system without a convex entropy pair or a non-symmetrically hyperbolic system. An approach using the Roe variables was proposed for the Euler equations in [30]. Although effective, its extension to general systems is limited, due to the Roe linearization. More recently, a class of operator splitting based SG methods were developed for the Euler equations [10] and the Saint-Venant system [11]. The idea is to split the underlying system into several subsystems, and the gPC-SG method for each of these subsystems may result in globally hyperbolic gPC-SG system. However, such splitting is problem dependent and difficult to extended to a general quasilinear hyperbolic system. On the other hand, there exist certain non-stochastic Galerkin methods for UQ of hyperbolic conservation laws, see e.g. [26,1]. The readers are referred to [3] for a recent review of UQ for hyperbolic systems of conservation (balance) laws.

How to derive a hyperbolic (deterministic) gPC-SG system for a general quasilinear hyperbolic system with uncertainty is still an open problem and is considered in this paper. The major contribution of this paper is the development of a gPC Galerkin approach that results in symmetrically hyperbolic system of equations for the gPC coefficients, for a general quasilinear hyperbolic system, such as the Euler equations, in the 1D space. The key ingredient of the method is the symmetrization of the 1D hyperbolic system via the left eigenvector matrix of its Jacobian matrix. The symmetric form of the system is then discretized and approximated by the gPC Galerkin approach. It is then proven that the resulting larger gPC-SG system is symmetrically hyperbolic. The symmetric hyperbolicity of the gPC-SG system is an important property and allows one to further conduct proper numerical discretizations. In this paper a fifth-order accurate, path-conservative, finite volume WENO scheme is used in space, and a third-order accurate, total variation diminishing, explicit Runge–Kutta method is used in time. Obviously, one is free to use other proper spatial and temporal discretization. For multi-dimensional problems, an operator splitting technique is employed to take advantage of the hyperbolicity of the one-dimensional gPC-SG systems.

The rest of the paper is organized as follows. Section 2 presents the gPC-SG method for 1D first-order quasilinear hyperbolic system with uncertainty, including the discretization in random space in Subsection 2.2, the spatial discretization in Subsection 2.3, and the time discretization in Subsection 2.4. Section 3 extends the proposed gPC-SG method to multi-dimensional case. Section 4 conducts several 1D and 2D numerical experiments to demonstrate the performance and accuracy of the proposed gPC-SG method. Concluding remarks are presented in Section 5.

2. One-dimensional gPC-SG method

This section considers the gPC-SG method of a general 1D quasilinear hyperbolic system

$$\frac{\partial}{\partial t}\boldsymbol{U}(x,t,\boldsymbol{\xi})+\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial x}\boldsymbol{U}(x,t,\boldsymbol{\xi})=\boldsymbol{0},\quad x\in\Omega\subseteq\mathbb{R},\ t>0, \tag{2.1}$$

where $\boldsymbol{U}(x,t,\boldsymbol{\xi})\in\mathbb{R}^N$ is the unknown, $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$ is a $N\times N$ real matrix, and $\boldsymbol{\xi}\in\Theta\subset\mathbb{R}^d$ denotes the random variables that parameterize the uncertain coefficients or initial conditions of the given problem. The system (2.1) is assumed to hyperbolic, which implies that the matrix $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$ is real diagonalizable for all admissible state $\boldsymbol{U}\in\mathcal{G}$ and almost everywhere $\boldsymbol{\xi}\in\Theta$, where $\mathcal{G}$ denotes the set of admissible states.

The solutions of nonlinear hyperbolic systems generally involve discontinuities. It is necessary to define the weak solution of (2.1). In some special cases where the matrix $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$ is exactly the Jacobian matrix of a flux function $\boldsymbol{F}(\boldsymbol{U};\boldsymbol{\xi})$, the system (2.1) can be rewritten into a divergence form

$$\frac{\partial}{\partial t}\boldsymbol{U}(x,t,\boldsymbol{\xi})+\frac{\partial}{\partial x}\boldsymbol{F}(\boldsymbol{U};\boldsymbol{\xi})=\boldsymbol{0}. \tag{2.2}$$

For such special cases, the notion of solutions in the sense of distributions can be used to define the weak solution of (2.2) in terms of Rankine–Hugoniot conditions. Otherwise, the DLM theory [12] should be adopted to define the weak solution of (2.1) associated with a family of paths $\boldsymbol{\Psi}(s,\boldsymbol{U}^-,\boldsymbol{U}^+)$. Here $\boldsymbol{\Psi}$ is a locally Lipschitz map from $[0,1]\times\mathcal{G}^2\mapsto\mathcal{G}$, such that $\boldsymbol{\Psi}(0,\boldsymbol{U}^-,\boldsymbol{U}^+)=\boldsymbol{U}^-$ and $\boldsymbol{\Psi}(1,\boldsymbol{U}^-,\boldsymbol{U}^+)=\boldsymbol{U}^+$. It is known that a different choice of paths may lead to a different weak solution of (2.1). Particularly, for a chosen $\boldsymbol{\Psi}$, across a discontinuity with speed σ the weak solution should satisfy the generalized Rankine–Hugoniot condition

$$\int_0^1\Big(\sigma\boldsymbol{I}-\boldsymbol{A}\left(\boldsymbol{\Psi}(s,\boldsymbol{U}^-,\boldsymbol{U}^+);\boldsymbol{\xi}\right)\Big)\frac{\partial\boldsymbol{\Psi}}{\partial s}(s,\boldsymbol{U}^-,\boldsymbol{U}^+)ds=0,$$

where $\boldsymbol{U}^-$ and $\boldsymbol{U}^+$ are the left and right limits of the solution at the discontinuity, respectively. The readers are referred to [24,12,28,8] for more details.

2.1. Symmetrization

Let the invertible matrix $\boldsymbol{L}(\boldsymbol{U};\boldsymbol{\xi})$ be the left eigenvector matrix of the matrix $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$, then

$$\boldsymbol{A} = \boldsymbol{L}^{-1}\boldsymbol{\Lambda}(\boldsymbol{U},\boldsymbol{\xi})\boldsymbol{L}, \quad \text{a.e. } \boldsymbol{\xi}\in\Theta,$$

where the diagonal matrix $\boldsymbol{\Lambda}(\boldsymbol{U},\boldsymbol{\xi}) = \mathrm{diag}\{\lambda_1(\boldsymbol{U},\boldsymbol{\xi}),\cdots,\lambda_N(\boldsymbol{U},\boldsymbol{\xi})\}$, and $\lambda_\ell(\boldsymbol{U},\boldsymbol{\xi})$, $\ell = 1,\cdots,N$, are the N eigenvalues of $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$.

Multiplying (2.1) from the left by the positive-definite matrix

$$\boldsymbol{A}_0(\boldsymbol{U};\boldsymbol{\xi}) := \boldsymbol{L}^\top\boldsymbol{L},$$

把A0左乘在原系统

one obtains a symmetric system

$$\boldsymbol{A}_0(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial t}\boldsymbol{U}(x,t,\boldsymbol{\xi}) + \boldsymbol{A}_1(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial x}\boldsymbol{U}(x,t,\boldsymbol{\xi}) = \boldsymbol{0}, \tag{2.3}$$

构造对称双曲系统：A0对称正定，A1对称
让Galerkin投影稳定

where

$$\boldsymbol{A}_1(\boldsymbol{U};\boldsymbol{\xi}) := \boldsymbol{L}^\top\boldsymbol{\Lambda}\boldsymbol{L}$$

is a real symmetric matrix and $\boldsymbol{L}^\top$ is the transpose of $\boldsymbol{L}$. The system (2.3) is the starting point of our stochastic Galerkin method for the system (2.1).

Remark 2.1. The symmetric form of (2.1) is not unique, that is, it is not limited to (2.3). In fact, one can similarly derive a more general symmetric form

$$(\boldsymbol{D}\boldsymbol{L})^\top(\boldsymbol{D}\boldsymbol{L})\frac{\partial}{\partial t}\boldsymbol{U}(x,t,\boldsymbol{\xi}) + (\boldsymbol{D}\boldsymbol{L})^\top(\boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}^{-1})(\boldsymbol{D}\boldsymbol{L})\frac{\partial}{\partial x}\boldsymbol{U}(x,t,\boldsymbol{\xi}) = \boldsymbol{0}, \tag{2.4}$$

where $\boldsymbol{D}$ can be any nonsingular diagonal matrix. It is interesting to investigate the choice of $\boldsymbol{D}$ and its effect on the numerical solutions.

2.2. Discretization in random space

Let $\{\phi_i(\boldsymbol{\xi})\}_{i\in\mathbb{N}\cup\{0\}}$ represent a complete set of polynomials in d variables and are orthogonal in the sense that

$$\int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi}) = \delta_{ij},$$

正交多项式

where $\mu(\boldsymbol{\xi})$ is the probability distribution function of $\boldsymbol{\xi}$, and δ_{ij} is the Kronecker symbol. We seek a finite approximation

$$\boldsymbol{u}_M(x,t,\boldsymbol{\xi}) = \sum_{i=0}^{M}\big(\hat{\boldsymbol{u}}_i(x,t)\big)^\top\phi_i(\boldsymbol{\xi}) \in \mathcal{G}, \tag{2.5}$$

gPC系数

to the solution $\boldsymbol{U}(x,t,\boldsymbol{\xi})$, where $\hat{\boldsymbol{u}}_i(x,t)$ is row vector, $i = 0,\cdots,M$. In gPC-SG procedure, (2.5) is then substituted into (2.3) and the residual is forced to be orthogonal to $\mathrm{span}\{\phi_0(\boldsymbol{\xi}),\cdots,\phi_M(\boldsymbol{\xi})\}$. The final deterministic gPC system for the expansion coefficients is as follows

$$\hat{\boldsymbol{A}}_0\frac{\partial}{\partial t}\hat{\boldsymbol{U}}(x,t) + \hat{\boldsymbol{A}}_1\frac{\partial}{\partial x}\hat{\boldsymbol{U}}(x,t) = \boldsymbol{0}, \tag{2.6}$$

where $\hat{\boldsymbol{U}} := (\hat{\boldsymbol{u}}_0,\cdots,\hat{\boldsymbol{u}}_M)^\top \in \mathbb{R}^{(M+1)N}$ and the coefficient matrices $\hat{\boldsymbol{A}}_k \in \mathbb{R}^{(M+1)N\times(M+1)N}$, $k = 0, 1$, are of the form

$$\hat{\boldsymbol{A}}_k = \begin{pmatrix} \hat{\boldsymbol{A}}_{00}^{(k)} & \cdots & \hat{\boldsymbol{A}}_{0M}^{(k)} \\ \vdots & & \vdots \\ \hat{\boldsymbol{A}}_{M0}^{(k)} & \cdots & \hat{\boldsymbol{A}}_{MM}^{(k)} \end{pmatrix}.$$

Here the blocks or sub-matrices are defined by

$$\hat{\boldsymbol{A}}_{ij}^{(k)} = \int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_k(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi}) \in \mathbb{R}^{N\times N}, \quad i,j = 0,\cdots M.$$

- 现在我们把近似解 $\mathbf{u}_M(x,t,\xi)$ 代入该方程，会得到一个**残差（residual）**：

$$R(x,t,\xi) = A_0(\mathbf{u}_M;\xi)\frac{\partial \mathbf{u}_M}{\partial t} + A_1(\mathbf{u}_M;\xi)\frac{\partial \mathbf{u}_M}{\partial x}$$

这个残差一般不等于零（因为 $\mathbf{u}_M$ 只是近似解）。

- **伽辽金方法的核心思想**：让残差与所有基函数 $\phi_j(\xi)$ 正交，即：

$$\int_\Theta R(x,t,\xi)\cdot\phi_j(\xi)\,d\mu(\xi) = 0, \quad j = 0,\ldots,M$$

这相当于在随机空间中对残差做"最小二乘"或"投影"处理，使得近似解在统计意义上最优。

Theorem 2.1. *If $\boldsymbol{u}_M(x,t,\boldsymbol{\xi}) \in \mathcal{G}$, then $\hat{\boldsymbol{A}}_0$ is real symmetric and positive definite, and $\hat{\boldsymbol{A}}_1$ is real symmetric, and the gPC-SG system (2.6) is symmetrically hyperbolic.*

Proof. Both matrices $\hat{\boldsymbol{A}}_0$ and $\hat{\boldsymbol{A}}_1$ are real since $\boldsymbol{u}_M(x,t,\boldsymbol{\xi}) \in \mathcal{G}$. Let $\tilde{\boldsymbol{A}}_{ij}^{(k)} \in \mathbb{R}^{N\times N}$ for $i,j = 0,\cdots,M$ be the blocks of the matrix $\hat{\boldsymbol{A}}_k^\top$, then one has

$$\begin{aligned}\tilde{\boldsymbol{A}}_{ij}^{(k)} &= \big(\hat{\boldsymbol{A}}_{ji}^{(k)}\big)^\top = \left(\int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_k(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi})\right)^\top \\ &= \int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_k^\top(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi}) = \int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_k(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi}) = \hat{\boldsymbol{A}}_{ij}^{(k)},\end{aligned}$$

which implies that $\hat{\boldsymbol{A}}_k$ is symmetric, $k = 0, 1$.

To show that $\hat{\boldsymbol{A}}_0$ is positive definite, consider an arbitrary $\boldsymbol{z} := (\boldsymbol{z}_0,\cdots,\boldsymbol{z}_M)^\top \in \mathbb{R}^{(M+1)N\times 1}$ with $\boldsymbol{z}_i^\top \in \mathbb{R}^N$, $i = 0,\cdots,M$. Then,

$$\begin{aligned}\boldsymbol{z}^\top\hat{\boldsymbol{A}}_0\boldsymbol{z} &= \sum_{i=0}^M\sum_{j=0}^M \boldsymbol{z}_i\hat{\boldsymbol{A}}_{ij}^{(0)}\boldsymbol{z}_j^\top \\ &= \sum_{i=0}^M\sum_{j=0}^M \boldsymbol{z}_i\left(\int_\Theta \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_0(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\mathrm{d}\mu(\boldsymbol{\xi})\right)\boldsymbol{z}_j^\top \\ &= \sum_{i=0}^M\sum_{j=0}^M \int_\Theta \boldsymbol{z}_i\phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})\boldsymbol{A}_0(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\boldsymbol{z}_j^\top\mathrm{d}\mu(\boldsymbol{\xi}) \\ &= \sum_{i=0}^M\sum_{j=0}^M \int_\Theta \big(\phi_i(\boldsymbol{\xi})\boldsymbol{z}_i\big)\boldsymbol{A}_0(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\big(\phi_j(\boldsymbol{\xi})\boldsymbol{z}_j\big)^\top\mathrm{d}\mu(\boldsymbol{\xi}) \\ &= \int_\Theta \sum_{i=0}^M\sum_{j=0}^M \big(\phi_i(\boldsymbol{\xi})\boldsymbol{z}_i\big)\boldsymbol{A}_0(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\big(\phi_j(\boldsymbol{\xi})\boldsymbol{z}_j\big)^\top\mathrm{d}\mu(\boldsymbol{\xi}) \\ &= \int_\Theta \left(\sum_{i=0}^M \phi_i(\boldsymbol{\xi})\boldsymbol{z}_i\right)\boldsymbol{A}_0(\boldsymbol{u}_M(x,t,\boldsymbol{\xi});\boldsymbol{\xi})\left(\sum_{j=0}^M \phi_j(\boldsymbol{\xi})\boldsymbol{z}_j\right)^\top\mathrm{d}\mu(\boldsymbol{\xi}) \ge 0.\end{aligned}$$

If $\boldsymbol{z}^\top\hat{\boldsymbol{A}}_0\boldsymbol{z} = 0$, then one has

$$\sum_{i=0}^M \phi_i(\boldsymbol{\xi})\boldsymbol{z}_i^\top = \boldsymbol{0}, \quad \text{a.e.}$$

Since $\{\phi_i(\boldsymbol{\xi})\}_{i\in\mathbb{N}\cup\{0\}}$ are basis polynomials, thus $\boldsymbol{z}_i = \boldsymbol{0}$ for all $i = 0,\cdots,M$, i.e. $\boldsymbol{z} = \boldsymbol{0}$. The proof is completed. □

An obvious corollary of Theorem 2.1 is that $\hat{\boldsymbol{B}} := \hat{\boldsymbol{A}}_0^{-1}\hat{\boldsymbol{A}}_1$ is real diagonalizable. 构造的gPC系统是拟线性的、非守恒的

The gPC-SG system (2.6) is quasilinear and of nonconservative form. It generally cannot be reformulated into a divergence form, even if the governing system (2.1) has a divergence form (2.2). Therefore, the weak solution of (2.6) should be defined via the DLM theory [12] based on a family of paths. The choice of the family of paths is important, and also difficult, in general. The simplest option, suggested in [7,28], is the family of segments

$$\boldsymbol{\Psi}\big(s,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\big) = \hat{\boldsymbol{U}}^- + s\big(\hat{\boldsymbol{U}}^+ - \hat{\boldsymbol{U}}^-\big), \quad \text{在间断点} \tag{2.7}$$

which will be considered as an example in this paper. Certainly, other proper options can be explored. It is also possible to study the characteristic fields of (2.6) in some specific cases together with the physical backgrounds, so as to obtain a more proper family of paths.

2.3. Spatial discretization

The hyperbolicity of the gPC Galerkin system (2.6) allows one to further conduct proper numerical discretization. For the quasilinear hyperbolic systems involving nonconservative products, such as (2.1) or (2.6), a popular choice is the so-called path-conservative scheme, see e.g. [7,28,5,6]. It is based on the DLM theory [12] and a generalization of the usual concept of conservative methods for the system of conservation laws. Some discussions on the known limitations and deficiencies of path-conservative schemes can be found in [8,2].

This section uses the high-order accurate, path-conservative finite volume WENO scheme [28,41] to discretize the gPC-SG system (2.6) in the physical space. For the sake of convenience, the spatial domain Ω is divided into the uniform mesh $\left\{x_{j+\frac{1}{2}} = \left(j-\frac{1}{2}\right)\Delta x \in \Omega \,|\, j \in \mathbb{Z}\right\}$, where Δx denotes the spatial step-size. Multiplying (2.6) by $\hat{\boldsymbol{A}}_0^{-1}$ from the left, and using the higher-order, path-conservative, finite volume WENO scheme to (2.6) for spatial discretization may give

$$\begin{aligned}\frac{\mathrm{d}\overline{\hat{\boldsymbol{U}}}_j(t)}{\mathrm{d}t} =& -\frac{1}{\Delta x}\left(\hat{\boldsymbol{B}}^+_{j-\frac{1}{2}}\left(\hat{\boldsymbol{U}}^+_{j-\frac{1}{2}}(t)-\hat{\boldsymbol{U}}^-_{j-\frac{1}{2}}(t)\right)+\hat{\boldsymbol{B}}^-_{j+\frac{1}{2}}\left(\hat{\boldsymbol{U}}^+_{j+\frac{1}{2}}(t)-\hat{\boldsymbol{U}}^-_{j+\frac{1}{2}}(t)\right)\right)\\ &-\sum_{m=1}^{q}\omega_m\hat{\boldsymbol{B}}\left(\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G,t)\right)\frac{\partial\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}}{\partial x}(x_m^G,t) =: \mathcal{L}\left(\overline{\hat{\boldsymbol{U}}}(t);j\right),\end{aligned} \tag{2.8}$$

where $\overline{\hat{\boldsymbol{U}}}_j(t)$ denotes the cell-averaged approximation of $\hat{\boldsymbol{U}}(x,t)$ over the cell $I_j := \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$, $\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x,t)$ is a polynomial vector function approximating $\hat{\boldsymbol{U}}(x,t)$ in the cell I_j and obtained by using the WENO reconstruction from $\overline{\hat{\boldsymbol{U}}}_j(t)$, x_m^G and ω_m denote the mth Gauss–Lobatto node transformed into the interval I_j and the associated weight, respectively, and

$$\hat{\boldsymbol{U}}^+_{j-\frac{1}{2}}(t) := \hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_{j-\frac{1}{2}}+0,t), \quad \hat{\boldsymbol{U}}^-_{j+\frac{1}{2}}(t) := \hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_{j+\frac{1}{2}}-0,t).$$

Here the matrix $\hat{\boldsymbol{B}}^+_{j+\frac{1}{2}}$ (resp. $\hat{\boldsymbol{B}}^-_{j+\frac{1}{2}}$) has only non-positive (resp. non-negative) eigenvalues and is given by a suitable splitting of the intermediate matrix

$$\hat{\boldsymbol{B}}_\Psi\left(\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right) = \left(\sum_{m=1}^{\tilde{q}}\tilde{\omega}_m\hat{\boldsymbol{A}}_0\left(\boldsymbol{\Psi}\left(s_m,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)\right)\right)^{-1}\left(\sum_{m=1}^{\tilde{q}}\tilde{\omega}_m\hat{\boldsymbol{A}}_1\left(\boldsymbol{\Psi}\left(s_m,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)\right)\right), \tag{2.9}$$

with the path $\boldsymbol{\Psi}$ defined in (2.7) and $s_m \in [0,1]$, that is

$$\hat{\boldsymbol{B}}_\Psi\left(\hat{\boldsymbol{U}}^-_{j+\frac{1}{2}},\hat{\boldsymbol{U}}^+_{j+\frac{1}{2}}\right) = \hat{\boldsymbol{B}}^-_{j+\frac{1}{2}} + \hat{\boldsymbol{B}}^+_{j+\frac{1}{2}}.$$

Here, the Lax–Friedrichs type splitting is employed, i.e.,

$$\hat{\boldsymbol{B}}^\pm_{j+\frac{1}{2}} = \frac{1}{2}\left(\hat{\boldsymbol{B}}_\Psi\left(\hat{\boldsymbol{U}}^-_{j+\frac{1}{2}},\hat{\boldsymbol{U}}^+_{j+\frac{1}{2}}\right) \pm \alpha_{j+\frac{1}{2}}\boldsymbol{I}_{(M+1)N}\right), \tag{2.10}$$

where $\boldsymbol{I}_{(M+1)N}$ denotes the identity matrix of size $(M+1)N$, and the coefficient $\alpha_{j+\frac{1}{2}}$ is not less than the spectral radius of the intermediate matrix $\hat{\boldsymbol{B}}_\Psi$, which may be estimated by using the eigenvalues of the matrix $\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})$, see Theorem 2.2 given below.

Practically, the intermediate matrix $\hat{\boldsymbol{B}}_\Psi\left(\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)$ is an approximation of the matrix

$$\left(\int_0^1\hat{\boldsymbol{A}}_0\left(\boldsymbol{\Psi}\left(s,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)\right)\mathrm{d}s\right)^{-1}\left(\int_0^1\hat{\boldsymbol{A}}_1\left(\boldsymbol{\Psi}\left(s,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)\right)\mathrm{d}s\right),$$

by using the Gaussian quadrature along the integral path $\boldsymbol{\Psi}\left(s,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right) = \hat{\boldsymbol{U}}^- + s\left(\hat{\boldsymbol{U}}^+-\hat{\boldsymbol{U}}^-\right)$, $s\in[0,1]$, which is the segment between $\hat{\boldsymbol{U}}^-$ and $\hat{\boldsymbol{U}}^+$, and s_m and $\tilde{\omega}_m$ denote the mth Gaussian node transformed into the interval $[0,1]$ and the associated weight, respectively. It is not difficult to prove by using Theorem 2.1 that the intermediate matrix $\hat{\boldsymbol{B}}_\Psi\left(\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)$ is real diagonalizable if $\boldsymbol{\Psi}\left(s_m,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)$, $m=1,\cdots,\tilde{q}$, are admissible, i.e.

$$\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}) = \sum_{i=0}^{M}\left(\hat{\boldsymbol{u}}_i^{(m)}\phi_i(\boldsymbol{\xi})\right)^\top \in \mathcal{G}, \quad \forall\boldsymbol{\xi}\in\Theta, \tag{2.11}$$

where $\left(\hat{\boldsymbol{u}}_0^{(m)},\cdots,\hat{\boldsymbol{u}}_M^{(m)}\right)^\top := \boldsymbol{\Psi}\left(s_m,\hat{\boldsymbol{U}}^-,\hat{\boldsymbol{U}}^+\right)$.

Before estimating the upper bound of the spectral radius of $\hat{\boldsymbol{B}}_{\Psi}$, we first prove a lemma.

Lemma 2.1. *If $\mathcal{A}_0$ is a real symmetric and positive-definite matrix, and $\mathcal{A}_1$ is a real symmetric matrix, then $\lambda\mathcal{A}_0 \pm \mathcal{A}_1$ is positive semi-definite if and only if $\lambda \geq \varrho\left(\mathcal{A}_0^{-1}\mathcal{A}_1\right)$, where $\varrho\left(\mathcal{A}_0^{-1}\mathcal{A}_1\right)$ denotes the spectral radius of the matrix $\mathcal{A}_0^{-1}\mathcal{A}_1$, and λ is a real number.*

Proof. Because $\lambda\mathcal{A}_0 \pm \mathcal{A}_1 = \left(\mathcal{A}_0^{\frac{1}{2}}\right)^{\top}\left(\lambda \boldsymbol{I} \pm \mathcal{A}_0^{\frac{1}{2}}\mathcal{A}_0^{-1}\mathcal{A}_1\mathcal{A}_0^{-\frac{1}{2}}\right)\mathcal{A}_0^{\frac{1}{2}}$, where $\boldsymbol{I}$ is the identity matrix, $\lambda\mathcal{A}_0 \pm \mathcal{A}_1$ is congruent to $\lambda \boldsymbol{I} \pm \mathcal{A}_0^{\frac{1}{2}}\mathcal{A}_0^{-1}\mathcal{A}_1\mathcal{A}_0^{-\frac{1}{2}}$. The hypothesis and Theorem 2.1 imply that $\mathcal{A}_0^{-1}\mathcal{A}_1$ is real diagonalizable. Thus $\lambda\mathcal{A}_0 \pm \mathcal{A}_1$ is positive semi-definite if and only if $\lambda \boldsymbol{I} \pm \mathcal{A}_0^{\frac{1}{2}}\mathcal{A}_0^{-1}\mathcal{A}_1\mathcal{A}_0^{-\frac{1}{2}}$ is positive semi-definite, equivalently, $\lambda \geq \varrho\left(\mathcal{A}_0^{\frac{1}{2}}\mathcal{A}_0^{-1}\mathcal{A}_1\mathcal{A}_0^{-\frac{1}{2}}\right) = \varrho\left(\mathcal{A}_0^{-1}\mathcal{A}_1\right)$. The proof is completed. □

Theorem 2.2. *If*

$$\boldsymbol{\Psi}\left(s_m, \hat{\boldsymbol{U}}^{-}, \hat{\boldsymbol{U}}^{+}\right) \in \hat{\mathcal{G}}_M := \left\{ \hat{\boldsymbol{U}} := (\hat{\boldsymbol{u}}_0, \cdots, \hat{\boldsymbol{u}}_M)^{\top} \in \mathbb{R}^{(M+1)N} \;\middle|\; \sum_{i=0}^{M} \left(\hat{\boldsymbol{u}}_i \phi_i(\boldsymbol{\xi})\right)^{\top} \in \mathcal{G},\ \forall \boldsymbol{\xi} \in \Theta \right\},$$

for $m = 1, \cdots, \tilde{q}$, then the spectral radius of the intermediate matrix $\hat{\boldsymbol{B}}_{\Psi}$ satisfies

$$\alpha := \max_{\ell, m} \sup_{\boldsymbol{\xi} \in \Theta} \left\{ \left| \lambda_{\ell}\left(\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}); \boldsymbol{\xi}\right) \right| \right\} \geq \varrho\left(\hat{\boldsymbol{B}}_{\Psi}\right), \qquad \text{这里取得是当前界面的局部最大特征值} \tag{2.12}$$

where λ_{ℓ} is the ℓ-th eigenvalue of the matrix $\boldsymbol{A}(\boldsymbol{U}; \boldsymbol{\xi})$, $\ell = 1, 2, \cdots, N$.

Proof. Under the hypothesis, it is easy to show that $\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi})$ defined in (2.11) belongs to $\mathcal{G}$, for all $\boldsymbol{\xi} \in \Theta$. Thanks to Theorem 2.1, $\hat{\boldsymbol{A}}_0\left(\boldsymbol{\Psi}\left(s_m, \hat{\boldsymbol{U}}^{-}, \hat{\boldsymbol{U}}^{+}\right)\right)$ is real symmetric and positive definite, and $\hat{\boldsymbol{A}}_1\left(\boldsymbol{\Psi}(s_m, \hat{\boldsymbol{U}}^{-}, \hat{\boldsymbol{U}}^{+})\right)$ is real symmetric. Because

$$\hat{\boldsymbol{A}}_k^{\Psi} := \sum_{m=1}^{\tilde{q}} \tilde{\omega}_m \hat{\boldsymbol{A}}_k\left(\boldsymbol{\Psi}\left(s_m, \hat{\boldsymbol{U}}^{-}, \hat{\boldsymbol{U}}^{+}\right)\right),\ k = 0, 1,$$

are two convex combinations of $\hat{\boldsymbol{A}}_k\left(\boldsymbol{\Psi}\left(s_m, \hat{\boldsymbol{U}}^{-}, \hat{\boldsymbol{U}}^{+}\right)\right)$, $m = 1, \cdots, \tilde{q}$, $\hat{\boldsymbol{A}}_0^{\Psi}$ is real symmetric and positive definite, and $\hat{\boldsymbol{A}}_1^{\Psi}$ is real symmetric. Due to Lemma 2.1, the inequality (2.12) is equivalent to the positive semi-definiteness of the matrix $\alpha\hat{\boldsymbol{A}}_0^{\Psi} \pm \hat{\boldsymbol{A}}_1^{\Psi}$, which is a block matrix of the form

$$\alpha\hat{\boldsymbol{A}}_0^{\Psi} \pm \hat{\boldsymbol{A}}_1^{\Psi} = \begin{pmatrix} \boldsymbol{A}_{00}^{\Psi} & \cdots & \boldsymbol{A}_{0M}^{\Psi} \\ \vdots & & \vdots \\ \boldsymbol{A}_{M0}^{\Psi} & \cdots & \boldsymbol{A}_{MM}^{\Psi} \end{pmatrix},$$

where the blocks

$$\boldsymbol{A}_{ij}^{\Psi} = \sum_{m=1}^{\tilde{q}} \tilde{\omega}_m \int_{\Theta} \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi}) \left(\alpha \boldsymbol{A}_0(\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}); \boldsymbol{\xi}) \pm \boldsymbol{A}_1(\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}); \boldsymbol{\xi})\right) \mathrm{d}\mu(\boldsymbol{\xi}) \in \mathbb{R}^{N \times N}.$$

The matrix $\alpha \boldsymbol{A}_0(\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}); \boldsymbol{\xi}) \pm \boldsymbol{A}_1(\boldsymbol{u}_M^{(m)}(\boldsymbol{\xi}); \boldsymbol{\xi})$ is positive semi-definite for all $\boldsymbol{\xi} \in \Theta$ due to the definition of α and Lemma 2.1. Then, the positive semi-definiteness of $\alpha\hat{\boldsymbol{A}}_0^{\Psi} \pm \hat{\boldsymbol{A}}_1^{\Psi}$ can be concluded by following the proof of Theorem 2.1. The proof is completed. □

Theorem 2.2 indicates that the complicate calculation of the spectral radius of the large matrix $\hat{\boldsymbol{B}}_{\Psi}$ may be avoided in the Lax–Friedrichs type splitting (2.10) and the CFL condition for determining the time step-size in practical computations.

Remark 2.2. In our computations, the parameters q in (2.8) and $\tilde{q}$ in (2.9) are taken as 4 and 3, respectively. The function $\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x, t)$ is derived by using the Lagrange interpolation based on $\left\{\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G, t)\right\}_{m=1}^{q}$, which are obtained by the fifth-order accurate WENO reconstruction from $\{\overline{\boldsymbol{U}}_j(t)\}$. If $\overline{\boldsymbol{U}}_j(t) \in \hat{\mathcal{G}}_M$ but $\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G, t) \notin \hat{\mathcal{G}}_M$ for at least one Gauss–Lobatto point x_m^G, then $\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G, t)$ is limited by the limiting procedure

$$\widetilde{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G,t)=\theta\left(\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G,t)-\overline{\hat{\boldsymbol{U}}}_j(t)\right)+\overline{\hat{\boldsymbol{U}}}_j(t),\quad m=1,\cdots,q,$$

where $\theta=\min_{1\le m\le q}\{\theta_m\}$, and $\theta_m=1$ for $\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G,t)\in\hat{\mathcal{G}}_M$; otherwise, θ_m corresponds to the intersection point between the line segment $\left\{\boldsymbol{\Psi}\left(s,\overline{\hat{\boldsymbol{U}}}_j(t),\hat{\boldsymbol{U}}_j^{\mathrm{WENO}}(x_m^G,t)\right),s\in[0,1]\right\}$ and the boundary of $\hat{\mathcal{G}}_M$. If the admissible state set $\mathcal{G}$ is convex, e.g. for the Euler equations [45] or the relativistic hydrodynamical equations [38,39], then $\hat{\mathcal{G}}_M$ is also convex, and thus the intersection point is unique. However, for the system (2.1), the above gPC-SG method cannot generally preserve the property that $\overline{\hat{\boldsymbol{U}}}_j(t)\in\hat{\mathcal{G}}_M$. Example 4.4 in Section 4 will show that the solutions $\overline{\hat{\boldsymbol{U}}}_j(t)\notin\hat{\mathcal{G}}_M$ may appear for very few cells in few time steps. For this case, $\overline{\hat{\boldsymbol{U}}}_j(t)=:(\overline{\hat{\boldsymbol{u}}}_0,\cdots,\overline{\hat{\boldsymbol{u}}}_M)^\top$ may be further limited as $\widetilde{\overline{\hat{\boldsymbol{U}}}}_j(t;\tilde{\theta})=(\overline{\hat{\boldsymbol{u}}}_0,\tilde{\theta}\overline{\hat{\boldsymbol{u}}}_1,\cdots,\tilde{\theta}\overline{\hat{\boldsymbol{u}}}_M)^\top$, where $\tilde{\theta}$ corresponds to the intersection point between the line segment $\left\{\boldsymbol{\Psi}\left(s,\widetilde{\overline{\hat{\boldsymbol{U}}}}_j(t;0),\overline{\hat{\boldsymbol{U}}}_j(t)\right),s\in[0,1]\right\}$ and the boundary of $\hat{\mathcal{G}}_M$.

如果WENO重构的UjWENO(x,t)不在容许状态集内，可能导致数值不稳定

无需

线段与状态集的交点

交点唯一

后面的平均值也不一定在状态集内

均值不动？

2.4. Time discretization

The time derivatives in the semi-discrete system (2.8) can be approximated by the explicit total variation diminishing Runge–Kutta methods [34]. Here, the third-order accurate version is considered, as an example. It takes the following form

$$\begin{aligned}
\overline{\hat{\boldsymbol{U}}}_j^{*}&=\overline{\hat{\boldsymbol{U}}}_j^{n}+\Delta t_n\mathcal{L}\left(\overline{\hat{\boldsymbol{U}}}^{n};j\right),\\
\overline{\hat{\boldsymbol{U}}}_j^{**}&=\frac{3}{4}\overline{\hat{\boldsymbol{U}}}_j^{n}+\frac{1}{4}\left(\overline{\hat{\boldsymbol{U}}}_j^{*}+\Delta t_n\mathcal{L}\left(\overline{\hat{\boldsymbol{U}}}^{*};j\right)\right),\\
\overline{\hat{\boldsymbol{U}}}_j^{n+1}&=\frac{1}{3}\overline{\hat{\boldsymbol{U}}}_j^{n}+\frac{2}{3}\left(\overline{\hat{\boldsymbol{U}}}_j^{**}+\Delta t_n\mathcal{L}\left(\overline{\hat{\boldsymbol{U}}}^{**};j\right)\right),
\end{aligned}\tag{2.13}$$

where Δt_n denotes the time step-size. The complete 1D gPC-SG method can be expressed as

$$\overline{\hat{\boldsymbol{U}}}_j^{n+1}=\overline{\hat{\boldsymbol{U}}}_j^{n}+\mathcal{E}_{1\mathrm{D}}\left(\cdots,\overline{\hat{\boldsymbol{U}}}_{j-2}^{n},\overline{\hat{\boldsymbol{U}}}_{j-1}^{n},\overline{\hat{\boldsymbol{U}}}_j^{n},\overline{\hat{\boldsymbol{U}}}_{j+1}^{n},\overline{\hat{\boldsymbol{U}}}_{j+2}^{n},\cdots;\Delta t_n\right).\tag{2.14}$$

3. Two-dimensional gPC-SG method

This section extends the above gPC-SG method to a general two-dimensional (2D) quasilinear hyperbolic system

$$\frac{\partial}{\partial t}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})+\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial x}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})+\boldsymbol{G}(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial y}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})=\boldsymbol{0},\quad (x,y)\in\Omega,\ t>0.\tag{3.1}$$

不能用前面的方法对称化

If the eigenvector matrices of the matrices $\boldsymbol{A}$ and $\boldsymbol{G}$ are different, the general system (3.1) can not be symmetrized with the approach used in Section 2.1. To derive an effective gPC-SG method for the general system (3.1), the operator splitting technique is employed here, which in essence reduces the multi-dimensional problem to a sequence of augmented 1D problems. For example, the 2D system (3.1) is decomposed into two subsystems (the x- and y-split 2D systems) as follows

$$\frac{\partial}{\partial t}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})+\boldsymbol{A}(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial x}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})=\boldsymbol{0},\tag{3.2}$$

拆分为2个子问题

and

$$\frac{\partial}{\partial t}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})+\boldsymbol{G}(\boldsymbol{U};\boldsymbol{\xi})\frac{\partial}{\partial y}\boldsymbol{U}(x,y,t,\boldsymbol{\xi})=\boldsymbol{0}.\tag{3.3}$$

For the split sub-systems, the approach presented in Section 2 can be used to obtain a deterministic symmetrically hyperbolic system for the gPC expansion coefficients.

Assuming that the 2D spatial domain Ω is divided into the uniform rectangular mesh $\left\{\left(x_{j+\frac{1}{2}}=\left(j+\frac{1}{2}\right)\Delta x,y_{k+\frac{1}{2}}=\left(k+\frac{1}{2}\right)\Delta y\right)\in\Omega\,|\,j,k\in\mathbb{Z}\right\}$, and the cell-averaged approximation of the "initial" gPC expansion coefficients $\overline{\hat{\boldsymbol{U}}}_{jk}^{n}$ over the cell $[x_{j-\frac{1}{2}},x_{j+\frac{1}{2}}]\times[y_{k-\frac{1}{2}},y_{k+\frac{1}{2}}]$ at $t=t_n$ are given, then the cell-averaged approximation of the expansion coefficients at $t=t_n+\Delta t_n$ can approximately be calculated based on some higher-order accurate operator splitting method, e.g. a third-order accurate operator splitting method [35,36]

$$\overline{\hat{\boldsymbol{U}}}_{jk}^{n+1}=\mathcal{E}_x^{\tau_1}\mathcal{E}_y^{\tau_1+\tau_2}\mathcal{E}_x^{\tau_2}\mathcal{E}_y^{\tau_3}\mathcal{E}_x^{\tau_3+\tau_4}\mathcal{E}_y^{\tau_4}\overline{\hat{\boldsymbol{U}}}_{jk}^{n},\tag{3.4}$$

or

$$\overline{\hat{\boldsymbol{U}}}_{jk}^{n+1}=\mathcal{E}_y^{\tau_1}\mathcal{E}_x^{\tau_1+\tau_2}\mathcal{E}_y^{\tau_2}\mathcal{E}_x^{\tau_3}\mathcal{E}_y^{\tau_3+\tau_4}\mathcal{E}_x^{\tau_4}\overline{\hat{\boldsymbol{U}}}_{jk}^{n},\tag{3.5}$$

where $\mathcal{E}_x^\tau$ and $\mathcal{E}_y^\tau$ denote the 1D finite volume WENO scheme for the deterministic systems corresponding to the split systems (3.2) and (3.3), respectively, and the "time step-sizes" τ_i are

$$\tau_1 = \frac{2\Delta t_n}{5-\sqrt{13}+\sqrt{2(1+\sqrt{13})}},\quad \tau_2 = \frac{7+\sqrt{13}-\sqrt{2(1+\sqrt{13})}}{12}\Delta t_n,$$

$$\tau_3 = \frac{\tau_1^2}{\tau_2-\tau_1},\qquad \tau_4 = \Delta t_n - (\tau_1+\tau_2+\tau_3).$$

Taking the operator $\mathcal{E}_x^\tau$ for example, it can be expressed in term of the 1D evolution operator in (2.14) as

$$\mathcal{E}_x^\tau \overline{\hat{\boldsymbol{U}}}^n_{jk} = \overline{\hat{\boldsymbol{U}}}^n_{jk} + \mathcal{E}_{1D}\big(\cdots, \overline{\hat{\boldsymbol{U}}}^n_{j-2,k}, \overline{\hat{\boldsymbol{U}}}^n_{j-1,k}, \overline{\hat{\boldsymbol{U}}}^n_{j,k}, \overline{\hat{\boldsymbol{U}}}^n_{j+1,k}, \overline{\hat{\boldsymbol{U}}}^n_{j+2,k}, \cdots; \tau\big), \tag{3.6}$$

调用1维程序即可

for all k. Therefore, the implementation of the 2D gPC-SG method is easy and only needs to call the 1D code directly.

It is known that for complex genuinely multi-dimensional structures such as oblique shocks or circular expansions, the dimensional splitting approach may yield spurious local wave structure resolutions and is not always robust and accurate. Therefore, it is still an open issue to seek a genuinely multi-dimensional stochastic Galerkin method that preserves hyperbolicity, for general multi-dimensional quasilinear hyperbolic systems with uncertainty.

4. Numerical experiments

This section presents several numerical examples to verify the accuracy and effectiveness of the proposed gPC-SG methods. The quasilinear systems (2.1) and (3.1) are respectively taken as the 1D Euler equations

$$\frac{\partial \boldsymbol{U}}{\partial t} + \boldsymbol{A}(\boldsymbol{U})\frac{\partial \boldsymbol{U}}{\partial x} = 0, \tag{4.1}$$

with $\boldsymbol{A}(\boldsymbol{U}) = \partial \boldsymbol{F}/\partial \boldsymbol{U}$,

(密度，动量，总能量)

$$\boldsymbol{U} = (\rho, \rho u, E)^\top,\quad \boldsymbol{F} = (\rho u, \rho u^2 + p, uE + up)^\top,$$

and the 2D Euler equations

$$\frac{\partial \boldsymbol{U}}{\partial t} + \boldsymbol{A}(\boldsymbol{U})\frac{\partial \boldsymbol{U}}{\partial x} + \boldsymbol{G}(\boldsymbol{U})\frac{\partial \boldsymbol{U}}{\partial y} = 0, \tag{4.2}$$

with $\boldsymbol{A}(\boldsymbol{U}) = \partial \boldsymbol{F}_1/\partial \boldsymbol{U}$, $\boldsymbol{G}(\boldsymbol{U}) = \partial \boldsymbol{F}_2/\partial \boldsymbol{U}$,

$$\boldsymbol{U} = (\rho, \rho u, \rho v, E)^\top,\ \boldsymbol{F}_1 = (\rho u, \rho u^2 + p, \rho uv, uE + up)^\top,\ \boldsymbol{F}_2 = (\rho v, \rho uv, \rho v^2 + p, vE + vp)^\top.$$

Here ρ, u, v, p, and E denote the density, the velocity components in x- and y-directions, the pressure, and the total energy, respectively. In 1D, $E = \rho e + \frac{1}{2}\rho u^2$; and in 2D, $E = \rho e + \frac{1}{2}\rho(u^2+v^2)$, where e is the internal energy. An equation of state is needed to close the system (4.1) or (4.2). We focus on the case for the ideal gases, i.e.

$$p = (\Gamma - 1)\rho e, \tag{4.3}$$

绝热率，空气通常取1.4

where Γ is the adiabatic index.

The left eigenvector matrix $\boldsymbol{L}$ used to symmetrize the 1D and 2D split Euler systems is taken as the widely-used form in the literature, see e.g. [47]. In the numerical experiments, unless specified otherwise, uncertainty enters the problems through the initial or boundary conditions, or the adiabatic index in the equation of state. More specifically, the adiabatic index or the initial ρ, u, v or p or boundary condition depends on a 1D random variable ξ, which is assumed to obey the uniform distribution on $[-1, 1]$ for simplicity. The Legendre polynomials are taken as the gPC basis, and thus the mean and variance of the gPC solution $\boldsymbol{u}_M$ in (2.5) are respectively given by

$$\mathbb{E}[\boldsymbol{u}_M] = \hat{\boldsymbol{u}}_0,\quad \mathrm{Var}[\boldsymbol{u}_M] = \sum_{i=1}^{M} \hat{\boldsymbol{u}}_M^2,$$

稳定性条件

and the corresponding standard deviation $\sigma[\boldsymbol{u}_M] = \sqrt{\mathrm{Var}[\boldsymbol{u}_M]}$. Unless specifically stated, all computations will use the CFL number of 0.6 in the finite volume WENO schemes for the deterministic gPC-SG systems. Due to the complex nature of the solution in physical space, we focus more on the $d = 1$ case in random space, which allows us to fully resolve the solution in random space and to carefully examine the numerical properties of the method. Higher random dimensional examples are presented in Example 4.2. Extension to multiple random variables is straightforward and merely increases the simulation time significantly.

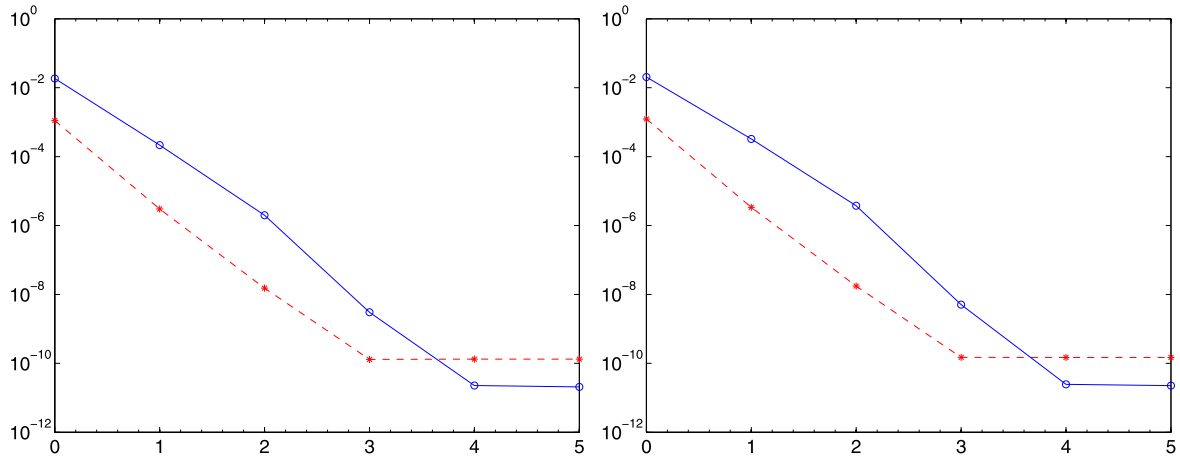

Fig. 4.1. Example 4.1: l^1-errors (left) and l^2-errors (right) at $t = 0.2$ in mean (dashed lines with symbols "$*$") and standard deviation (solid lines with symbols "$\circ$") of the density, with respect to gPC order M for the gPC-SG method with 320 uniform cells.

For the purpose of numerical validation, the reference solutions are computed using a stochastic collocation method. The collocation method takes the Gaussian points within the interval $[-1, 1]$ as the collocation points, and solves the deterministic problem for each collocation point by using the fifth-order accurate finite difference WENO scheme and the third-order Runge Kutta time discretization (2.13). First, for each Gaussian point ξ_ℓ^G, $\ell = 1, 2, \cdots, N_{\text{gs}}$, we fix the random parameter as $\xi = \xi_\ell^G$, and solve the deterministic Euler equations in *conservative form* to give the numerical solution denoted by $\boldsymbol{u}_{\text{ref}}(x, t, \xi_\ell^G)$ at spatial grid points. Then, the reference mean and standard deviation are computed by using the Gaussian quadrature formula over the 1D random space. If the 1D random variable ξ obeys the uniform distribution on $[-1, 1]$, the reference mean and variance are computed component-wisely by

$$\begin{aligned}
\mathbb{E}[\boldsymbol{u}_{\text{ref}}] &= \sum_{\ell=1}^{N_{\text{gs}}} \omega_\ell \boldsymbol{u}_{\text{ref}}(x, t, \xi_\ell^G) \approx \frac{1}{2}\int_{-1}^{1} \boldsymbol{U}(x, t, \xi) d\xi =: \mathbb{E}[\boldsymbol{U}], \\
\text{Var}[\boldsymbol{u}_{\text{ref}}] &= \sum_{\ell=1}^{N_{\text{gs}}} \omega_\ell \Big(\boldsymbol{u}_{\text{ref}}(x, t, \xi_\ell^G) - \mathbb{E}[\boldsymbol{u}_{\text{ref}}]\Big)^2 \approx \frac{1}{2}\int_{-1}^{1} \Big(\boldsymbol{U}(x, t, \xi) - \mathbb{E}[\boldsymbol{U}]\Big)^2 d\xi,
\end{aligned} \tag{4.4}$$

and the corresponding standard deviation is evaluated by

$$\sigma[\boldsymbol{u}_{\text{ref}}] = \sqrt{\text{Var}[\boldsymbol{u}_{\text{ref}}]}.$$

In (4.4), $\{\omega_\ell\}$ denote the associated Gaussian weights, and satisfy $\sum_{\ell=1}^{N_{\text{gs}}} \omega_\ell = 1$. Unless specifically stated, N_{gs} is taken as 40. Such collocation method is non-intrusive and has also been used for the validation purpose in the literature, see e.g. [20,21,19].

4.1. 1D case

Example 4.1 *(Smooth problem I).* This example is used to check the accuracy of the gPC-SG method for smooth solution

$$(\rho, u, p)(x, t, \xi) = \Big(1 + 0.2\sin\big(2\pi(x - (0.8 + 0.2\xi)t)\big), 0.8 + 0.2\xi, 1\Big),$$

with randomness, which describes a sine wave propagating periodically within the spatial domain $[0, 1]$ with uncertain velocity.

The spatial domain is divided into N_c uniform cells and the periodic boundary conditions are specified. The time step is taken as $\Delta t_n = \Delta x^{\frac{5}{3}}$ in order to realize fifth-order accuracy in time in the present case. The l^1- and l^2-errors in the mean and standard deviation of the density at $t = 0.2$ obtained by the gPC-SG method with $N_c = 320$ uniform cells and different gPC orders M are plotted in Fig. 4.1. The fast exponential convergence with respect to the order of gPC expansion is observed both in mean and standard deviation. The errors saturate at modest gPC orders, because the spatial and time discretization errors become dominant at this stage. Table 4.1 lists the l^1- and l^2-errors at $t = 0.2$ in the mean and standard deviation of the density and corresponding convergence rates for the gPC-SG method with $M = 4$ and different N_c. The results show that the convergence rate of fifth-order can be almost obtained in space and time.

Table 4.1
Example 4.1: l^1- and l^2-errors at $t = 0.2$ in the mean and standard deviation of the density and corresponding convergence rates for the gPC-SG method with $M = 4$.

N_c	Mean of ρ				Standard deviation of ρ			
	l^1 error	l^1 order	l^2 error	l^2 order	l^1 error	l^1 order	l^2 error	l^2 order
10	3.11e−3	–	3.52e−3	–	4.46e−4	–	5.29e−4	–
20	1.43e−4	4.45	1.66e−4	4.41	2.17e−5	4.36	2.63e−5	4.33
40	4.38e−6	5.02	5.10e−6	5.02	9.48e−7	4.51	1.01e−6	4.71
80	1.36e−7	5.01	1.54e−7	5.05	2.89e−8	5.04	3.10e−8	5.02
160	4.25e−9	5.00	4.74e−9	5.02	7.62e−10	5.24	8.26e−10	5.23
320	1.33e−10	5.00	1.48e−10	5.00	2.27e−11	5.07	2.45e−11	5.08

Table 4.2
Example 4.1: The estimated CPU time (seconds) for the gPC-SG method with $M = 4$.

N_c	10	20	40	80	160	320
total CPU time	0.19	1.12	6.91	43.54	276.29	1759.72
CPU time per time-step	0.019	0.037	0.073	0.146	0.293	0.588

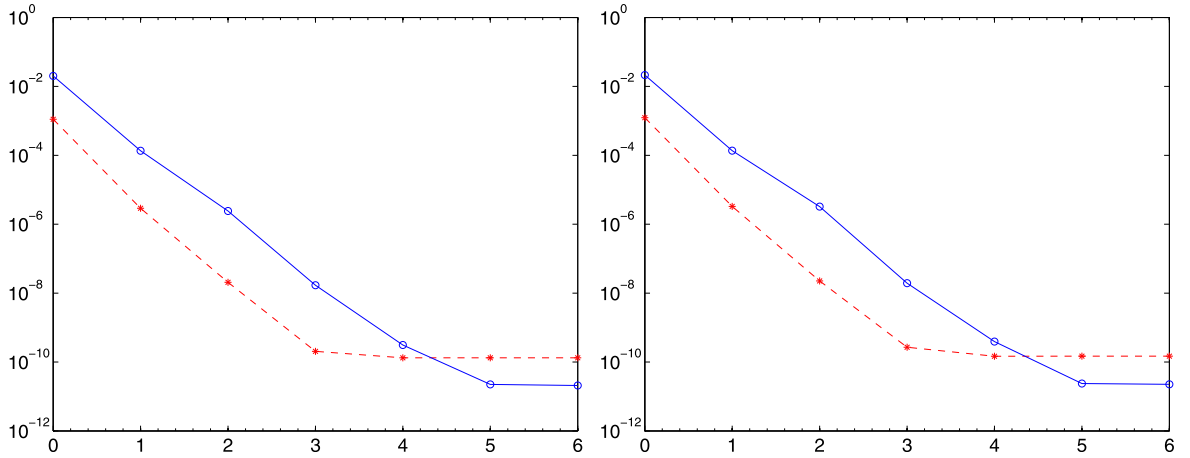

Fig. 4.2. Same as Fig. 4.1 except for Example 4.2.

As a reference information on the computational cost of the gPC-SG method, the estimated CPU time is listed in Table 4.2. This test is carried out on a computer equipped with the Intel Xeon E5-2697 v2 processor (2.70 GHz). Because our time step-sizes are taken as $\Delta x^{\frac{5}{3}}$, it is also necessary to show the averaged CPU time per time-step. As expected, the CPU time per time-step grows linearly with respect to the number of cells N_c.

Example 4.2 *(Smooth problem II).* This problem is similar to Example 4.1, except for containing another independent uncertainty in the initial density. More specifically, the initial data are

$$(\rho, u, p)(x, 0, \boldsymbol{\xi}) = \Big(1.0 + 0.2\big(1 + 0.1\sin(\pi\xi_1/4)\big)\sin(2\pi x), 0.8 + 0.2\xi_2, 1\Big),$$

where ξ_1 and ξ_2 independently obey the uniform distribution on $[-1, 1]$.

The spatial domain is divided into N_c uniform cells and the periodic boundary conditions are specified. The time step is taken as $\Delta t_n = \Delta x^{\frac{5}{3}}$ in order to realize fifth-order accuracy in time in the present case. The 2D Legendre polynomials of total degree up to M are used in the gPC expansion. Fig. 4.2 displays the l^1- and l^2-errors in the mean and standard deviation of the density at $t = 0.2$ obtained by the gPC-SG method with $N_c = 320$ uniform cells and different gPC orders M. Similar to Example 4.1, fast exponential convergence with respect to the order of gPC expansion is observed both in mean and standard deviation, before the errors saturate at modest gPC orders. Table 4.3 lists the l^1- and l^2-errors at $t = 0.2$ in the mean and standard deviation of the density and corresponding convergence rates for the gPC-SG method with $M = 6$ and different N_c. The results show again that the convergence rate of fifth-order can be almost obtained in space and time.

Example 4.3 *(Uncertain boundary condition problem).* Initially, the spatial domain $[0, 1]$ is filled with static fluid with unit density, pressure of 0.6 and adiabatic index of $\frac{5}{3}$. Waves are excited by a time periodic driver which acts at the left boundary $x = 0$, i.e.,

Table 4.3
Same as Table 4.1 except for Example 4.2 and $M = 6$.

N_c	Mean of ρ				Standard deviation of ρ			
	l^1 error	l^1 order	l^2 error	l^2 order	l^1 error	l^1 order	l^2 error	l^2 order
10	3.11e−3	–	3.52e−3	–	4.58e−4	–	5.40e−4	–
20	1.43e−4	4.45	1.66e−4	4.41	2.03e−5	4.49	2.58e−5	4.39
40	4.38e−6	5.02	5.10e−6	5.02	8.10e−7	4.65	8.89e−7	4.86
80	1.36e−7	5.01	1.54e−7	5.05	2.42e−8	5.07	2.56e−8	5.13
160	4.25e−9	5.00	4.74e−9	5.02	6.87e−10	5.14	7.34e−10	5.12
320	1.33e−10	5.00	1.48e−10	5.00	2.09e−11	5.04	2.25e−11	5.03

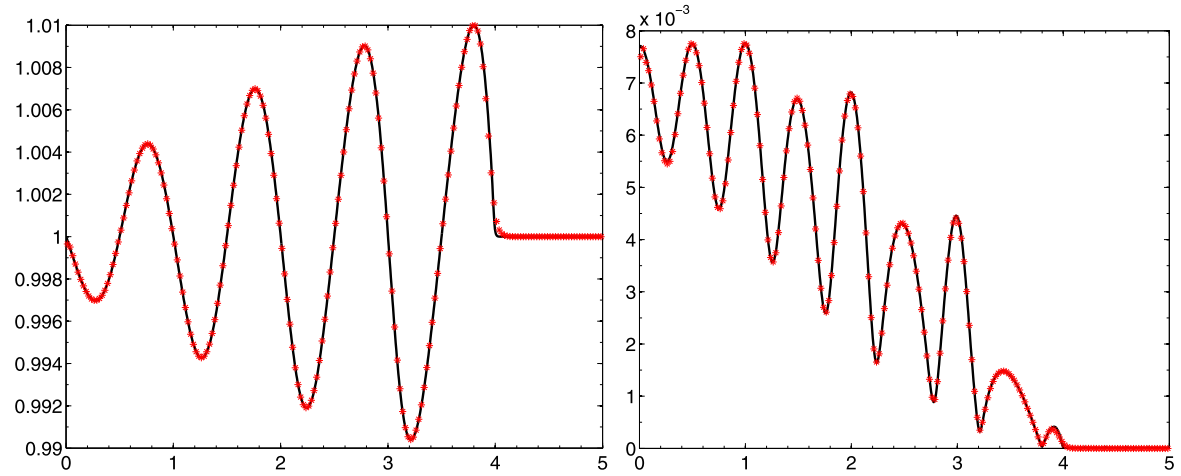

Fig. 4.3. Example 4.3: the mean and standard deviation of ρ at $t = 4$ by using the gPC-SG method ("$*$") with $M = 3$ and 200 uniform cells, and the solid lines represent the reference solutions given by a collocation method with 1000 uniform cells.

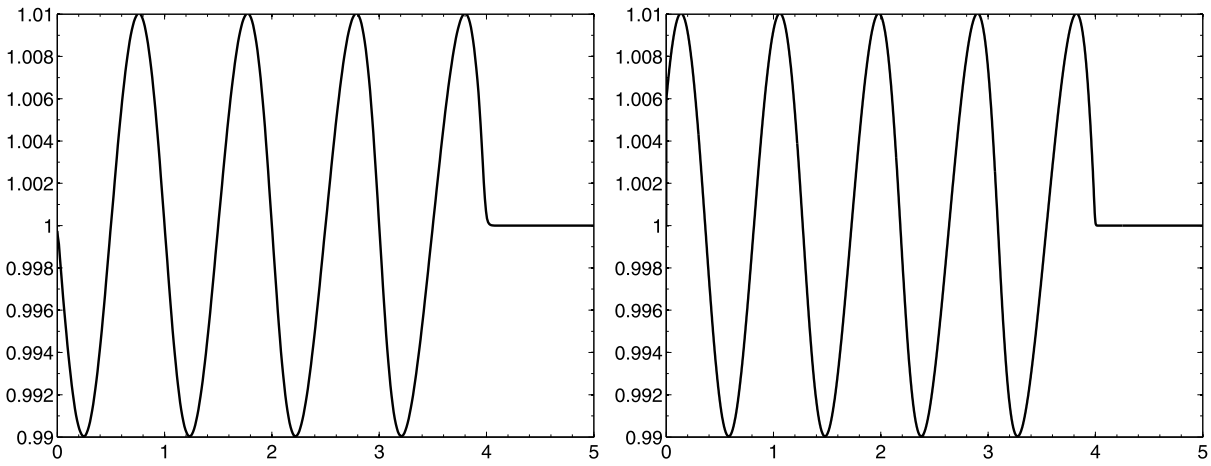

Fig. 4.4. Example 4.3: ρ at $t = 4$ for deterministic w cases: $w = 1$ (left), $w = 1.1$ (right).

$$(\rho, u, p)(0, t) = (1, 0.02\sin(2\pi wt), 0.6), \tag{4.5}$$

while outflow boundary condition is specified at the right $x = 5$. Similar problems are considered in [14,22]. Here, the parameter w describing the frequency of the waves contains uncertainty as follows $w(\xi) = 1 + 0.1\xi$. Fig. 4.3 gives mean and standard deviation of the density at time $t = 4$ by using the gPC-SG method with $M = 3$ and 200 uniform cells, where the solid lines represent the reference solutions given by a collocation method with 1000 uniform cells. Good agreements between the numerical solutions and the reference solutions can be observed in these results, but we can observe that the uncertainty in frequency influences the local peak and valley values of the mean of density. It is different from the results in Fig. 4.4, where the same peak and valley values are observed in the deterministic problems with different w.

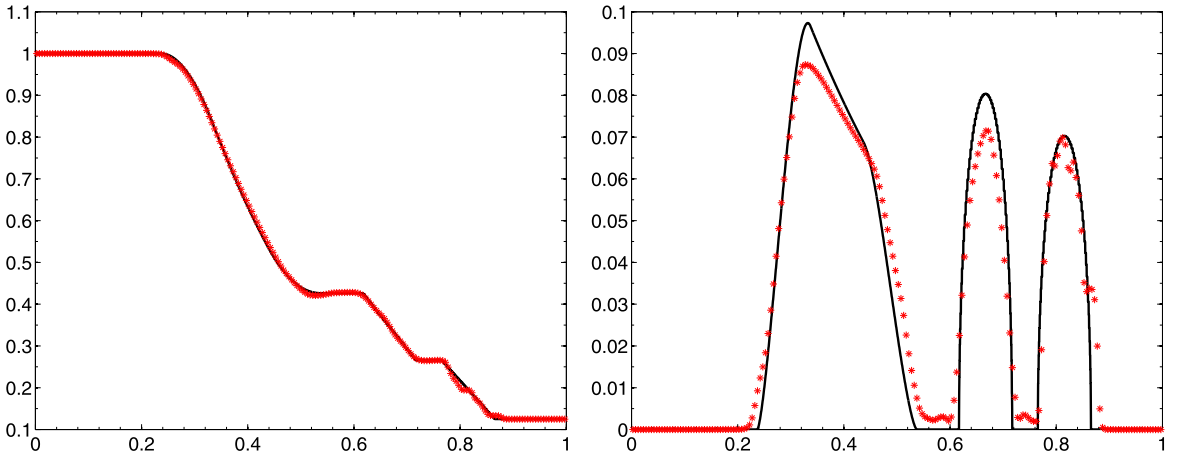

Fig. 4.5. The first configuration of Example 4.4: the mean and standard deviation of the density at time $t = 0.18$ by using the gPC-SG method ("$*$") with $M = 8$ and 200 uniform cells, and the solid lines represent the reference solutions.

Example 4.4 *(Sod problems with uncertainty).* This test considers two perturbed versions of the Sod shock tube problem with $\Gamma = 1.4$.

The first configuration is with uncertainty in the location of the initial discontinuity. The initial data are

$$(\rho, u, p)(x, 0, \xi) = \begin{cases} (1, 0, 1), & x < 0.5 + 0.05\xi, \\ (0.125, 0, 0.1), & x > 0.5 + 0.05\xi. \end{cases} \tag{4.6}$$

The same setup is considered in [29,13,30,10]. The gPC-SG method may easily fail in this test due to the appearance of negative density caused by the oscillations [29]. In our computations, numerical solutions with negative density and pressure in few cells are encountered in the first few steps, the positivity-preserving technique presented in Remark 2.2 is used (only in this test) to deal with such difficulty. Fig. 4.5 displays numerical means and standard deviations of the density at $t = 0.18$, which are obtained by using the gPC-SG method with 200 uniform cells and $M = 8$, where the solid lines represent the reference solutions given by the exact Riemann solver with 64 Gaussian points in the random space to evaluate the mean and standard deviation. Three sharp fronts with large standard deviation are observed, and their locations are corresponding to the rarefaction fan, contact discontinuity and shock from left to right. It is observed that the uncertain effect on the three types of waves are well captured, but in regions between the waves the computed standard deviation values are not much close to zero due to numerical errors. One can see small oscillations in the numerical standard deviation, which is also observed in [13]. This Gibbs phenomenon results from the discontinuity of the solution in random space, and can be improved by utilizing piecewise approximations in random space, such as multi-element gPC [37] or wavelet basis [25,30].

In the second configuration, the uncertainty appears in the amplitude of the initial discontinuity. Specifically, the initial data are

$$(\rho, u, p)(x, 0, \xi) = \begin{cases} (1 + 0.3\xi, 0, 1), & x < 0.5, \\ (0.125, 0, 0.1), & x > 0.5. \end{cases} \tag{4.7}$$

Figs. 4.6 and 4.7 respectively show the numerical means and standard deviations of the density at $t = 0.18$, which are obtained by using the gPC-SG method with 2000 uniform cells and different M. In those figures, the solid lines represent the reference solutions, which are evaluated at 10,000 uniform grid points by the exact Riemann solver, and with 64 Gaussian points in the random space to compute the mean and standard deviation. It is seen that, as M increases, the mean and standard deviation are resolved better. Oscillations (Gibbs phenomenon) are observed in the numerical standard deviations when the exact solution contains discontinuities in random space.

4.2. 2D case

The 2D Riemann problems are theoretically studied for the first time in [46]. Since then, they become the benchmark tests for verifying the accuracy and resolution of numerical schemes, see [33,23,17,40].

Four 2D Riemann problems in spatial domain $[0, 1] \times [0, 1]$ with uncertain initial data or adiabatic index of 2D Euler equations (4.2) are considered here.

Example 4.5 *(2D Riemann problems I and II).* The first two problems are two perturbed versions of a classic deterministic Riemann problem with initial conditions [23]

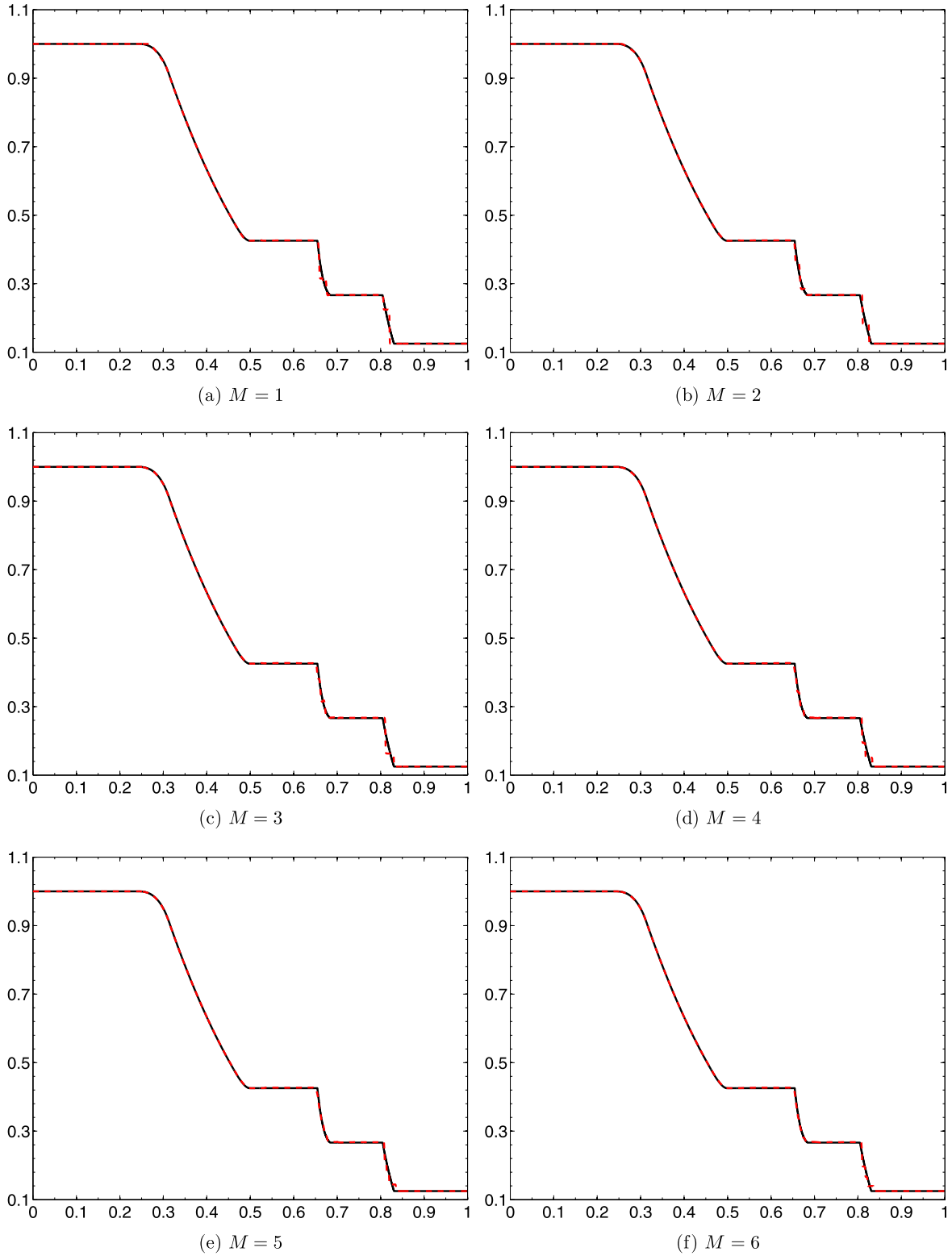

Fig. 4.6. The second configuration of Example 4.4: the mean of density at time $t = 0.18$ by using the gPC-SG method (dash lines) with different M and 2000 uniform cells, and the solid lines represent the reference solutions.

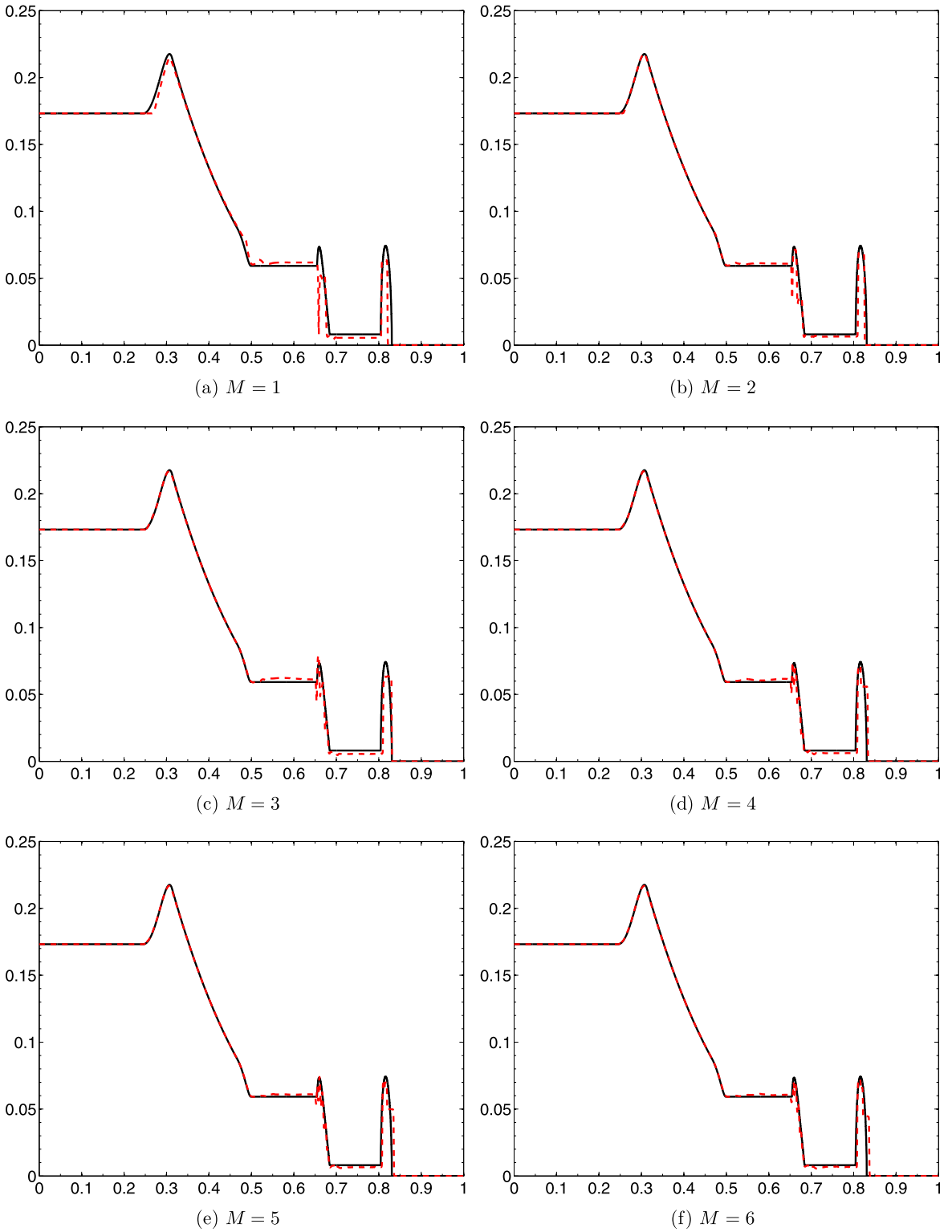

Fig. 4.7. Same as Fig. 4.6 except for the standard deviation of density.

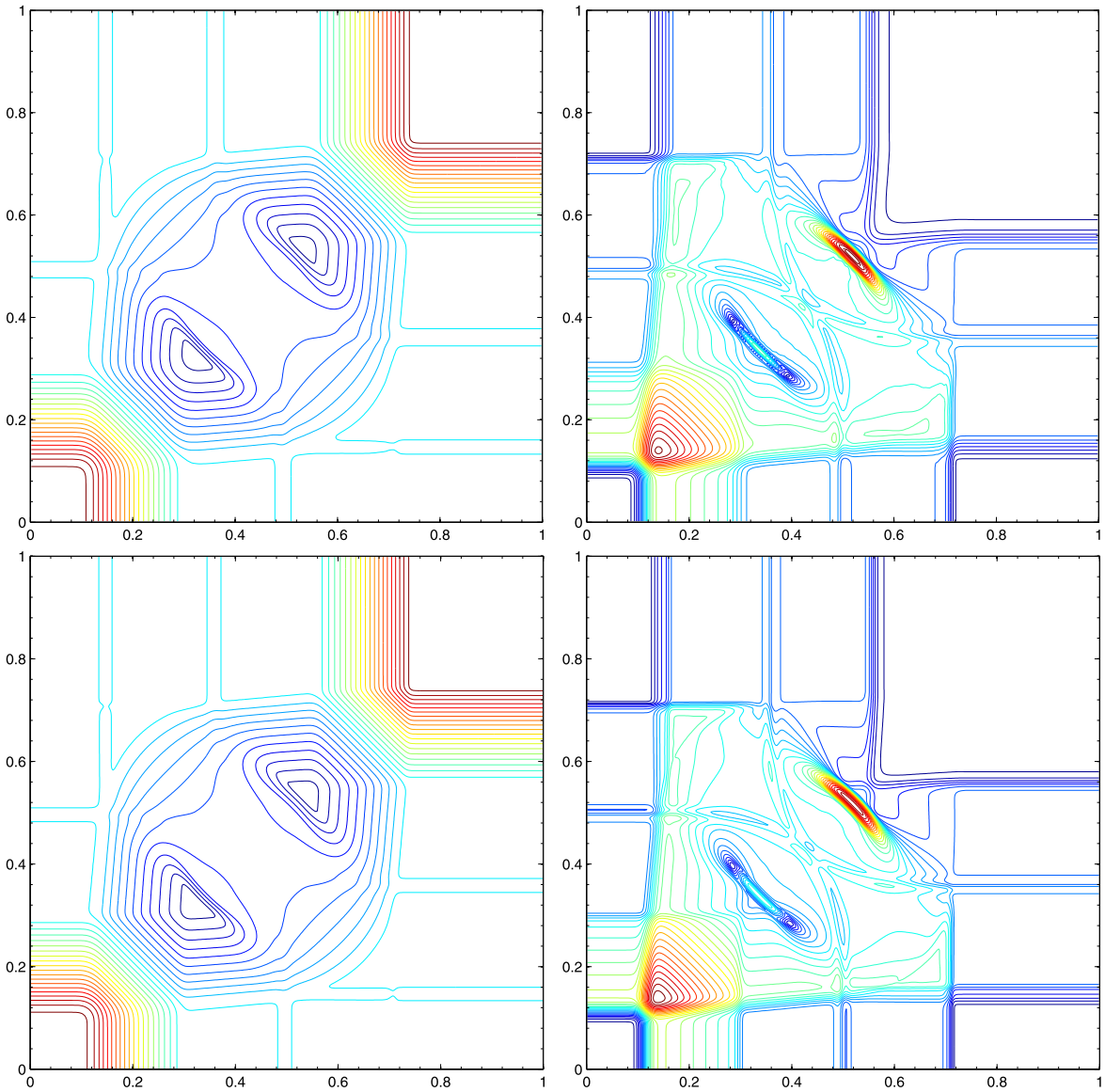

Fig. 4.8. The first configuration of Example 4.5: The contours of numerical mean (left) and standard deviation (right) of the density at $t = 0.2$ within the domain $[0, 1] \times [0, 1]$. 30 equally spaced contour lines are used. From top to bottom: gPC-SG method with $M = 4$ and 250×250 uniform cells, and collocation method with 400×400 uniform cells.

$$(\rho, u, v, p)(x, y, 0) = \begin{cases} (1, 0, 0, 1), & x > 0.5, y > 0.5, \\ (0.5197, -0.7259, 0, 0.4), & x < 0.5, y > 0.5, \\ (1, -0.7259, -0.7259, 1), & x < 0.5, y < 0.5, \\ (0.5197, 0, -0.7259, 0.4), & x > 0.5, y < 0.5, \end{cases} \tag{4.8}$$

which are about the interaction of four rarefaction waves.

In the first configuration, we take $\Gamma = 1.4$ and assume that the initial data -0.7259 of fluid velocity in (4.8) are perturbed to $-0.7259 + 0.1\xi$. The gPC-SG method is used to study the effect of this random inputs on the flow structure. Fig. 4.8 gives the contours of numerical mean and standard deviation of the density at time $t = 0.2$ by using the gPC-SG method with $M = 4$ and 250×250 uniform cells, while the reference solutions given by a collocation method with 400×400 uniform cells are also displayed. It can be seen that the mean and standard deviation of the density are correctly captured by the gPC-SG method. For a further comparison, the mean and standard deviation of the density are plotted along the line $y = x$, see Fig. 4.9, where the results by $M = 3$ are also presented. Those plots validate the above observation.

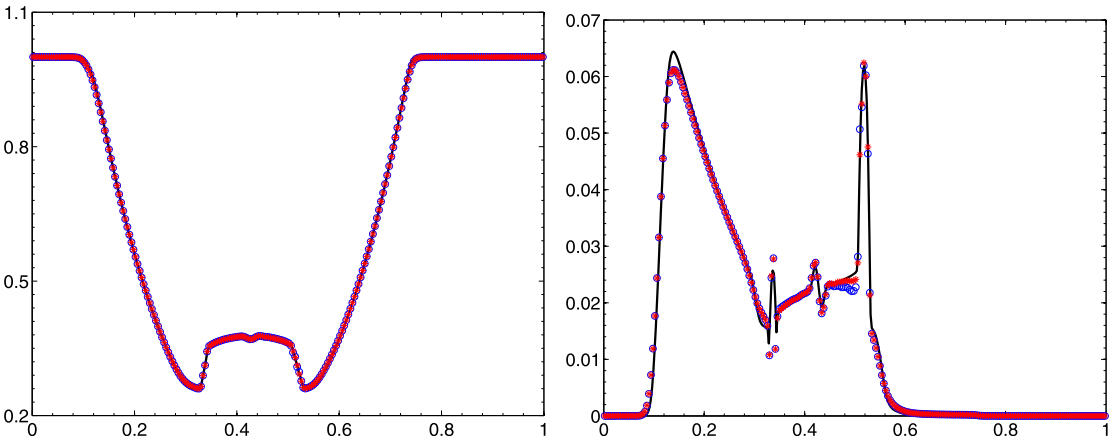

Fig. 4.9. The first configuration of Example 4.5: The mean (left) and standard deviation (right) of the density along the line $y = x$ within the scaled interval $[0, 1]$. Symbols "$\circ$" (resp. "$*$") denote the numerical results given by gPC-SG method with $M = 3$ (resp. $M = 4$) and 250×250 uniform cells, while the solid lines represent the numerical results of collocation method with 400×400 uniform cells.

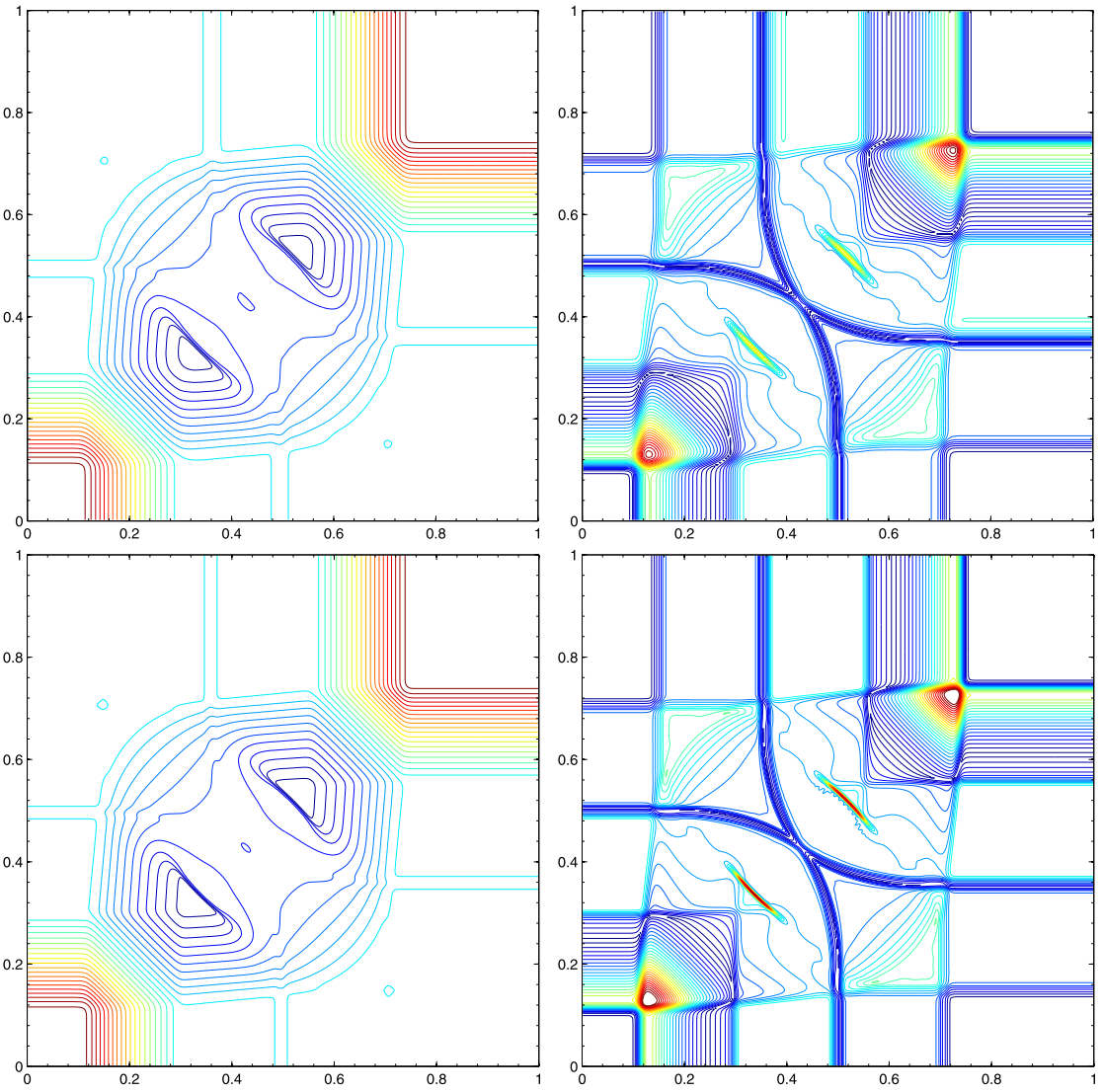

Fig. 4.10. Same as Fig. 4.8, except for the second configuration of Example 4.5.

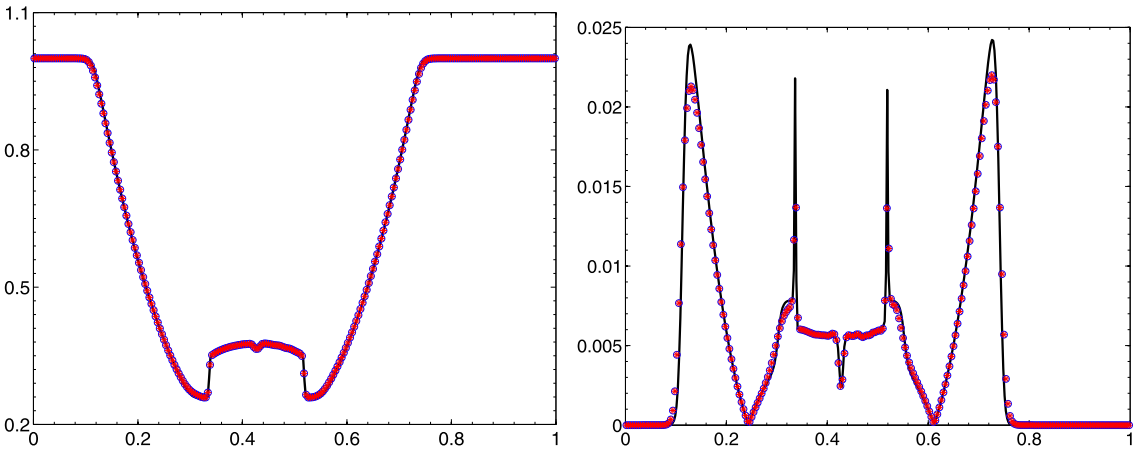

Fig. 4.11. Same as Fig. 4.9, except for the second configuration of Example 4.5.

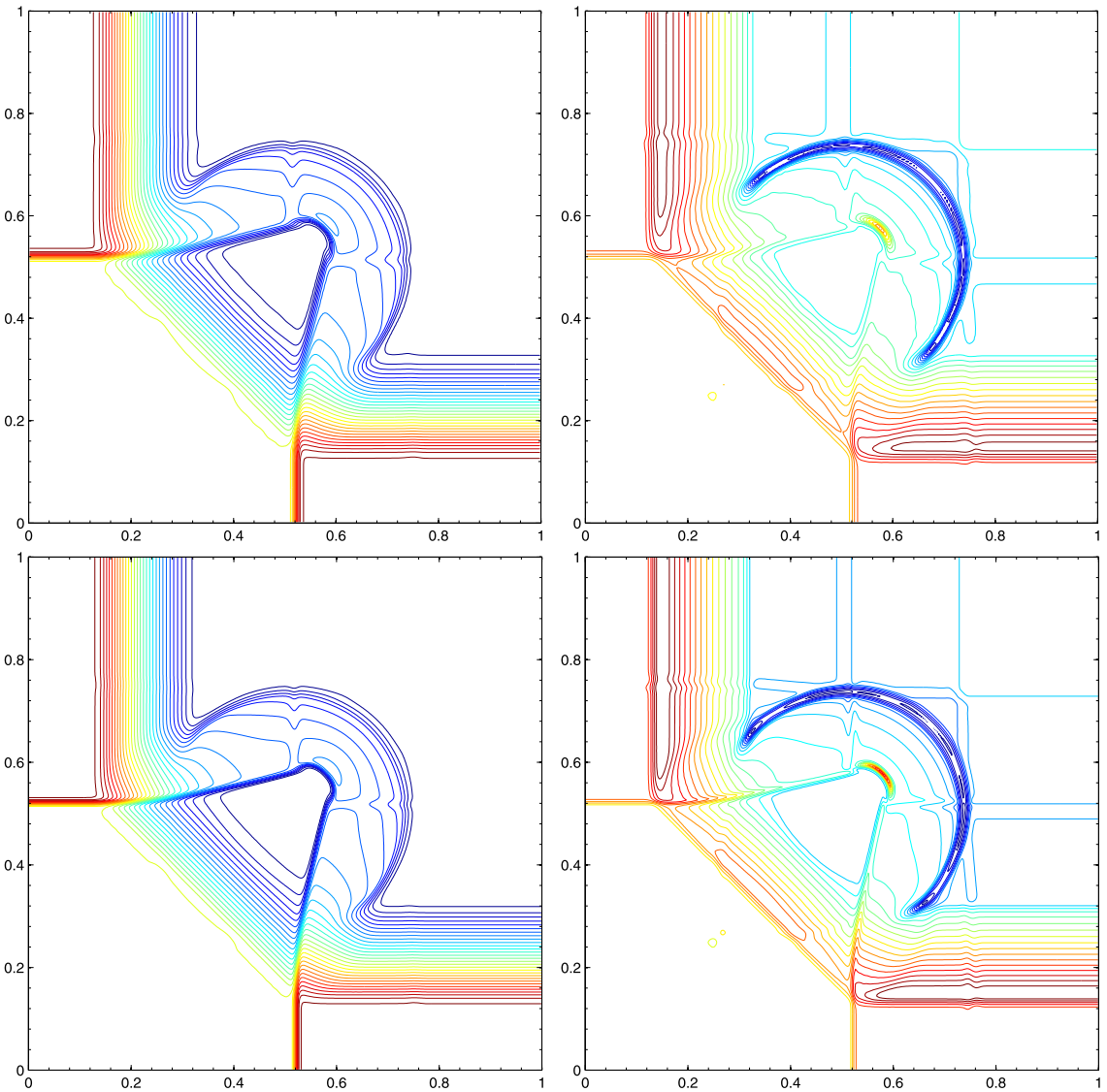

Fig. 4.12. Same as Fig. 4.8, except for the first configuration of Example 4.6 and 25 equally spaced contour lines.

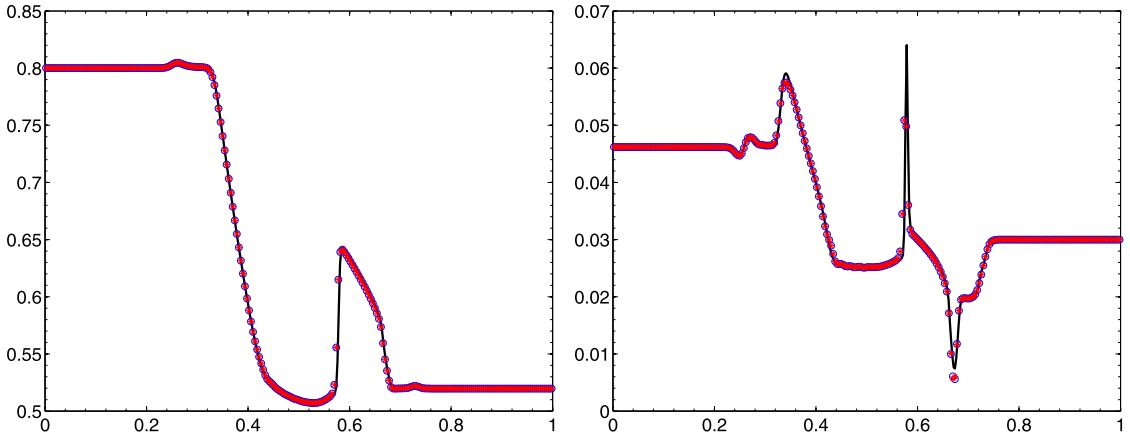

Fig. 4.13. Same as Fig. 4.9, except for the first configuration of Example 4.6.

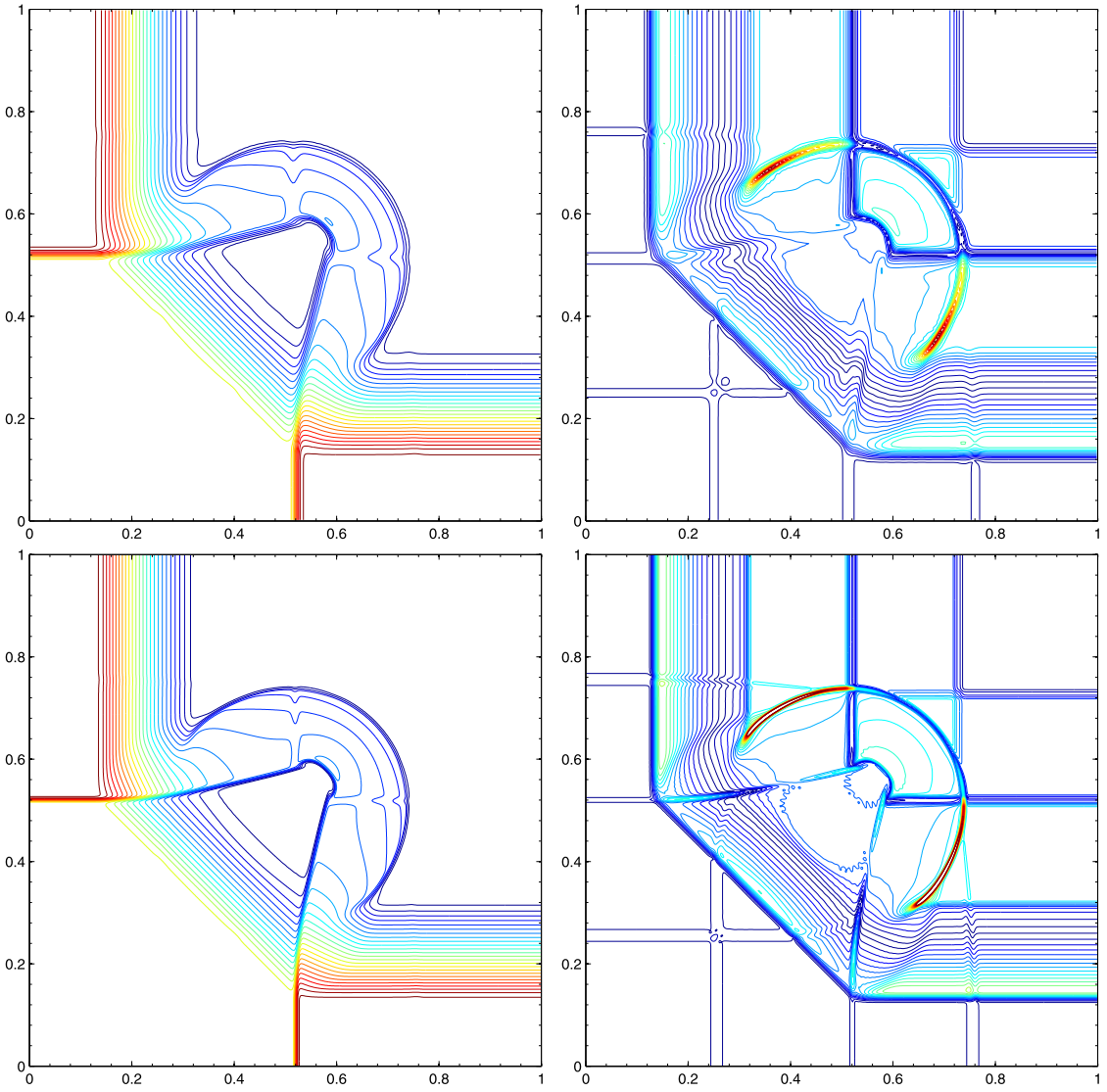

Fig. 4.14. Same as Fig. 4.8, except for the second configuration of Example 4.6, the collocation method using 640×640 uniform cells and $N_{gs} = 64$, and 25 equally spaced contour lines.

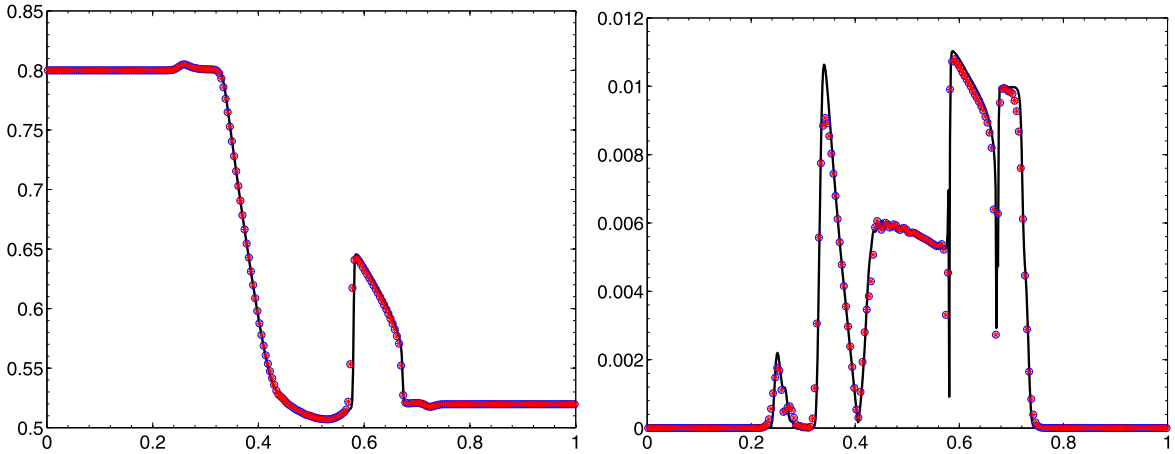

Fig. 4.15. Same as Fig. 4.9, except for the second configuration of Example 4.6 and the collocation method using 640×640 uniform cells and $N_{gs} = 64$.

The second configuration takes the certain initial data (4.8) with uncertain adiabatic index Γ, which satisfies

$$\Gamma(\xi) = 1.4 + 0.1\xi.$$

To analyze the effect of this uncertainty, the gPC-SG method is used to compute the numerical solutions with $M = 4$ and 250×250 uniform cells. The contours of numerical mean and standard deviation of the density at time $t = 0.2$ are displayed in Fig. 4.10, where the reference solutions are obtained by the collocation method with 400×400 uniform cells. It is seen that the mean of the density is similar to that in the first configuration, but the standard deviation is very different. Moreover, the gPC-SG method captures the flow structure and resolves the standard deviation with high resolution. Fig. 4.11 gives a further comparison of the mean and standard deviation of the density along the line $y = x$, and demonstrates good agreement between the numerical solutions of the gPC-SG method and the reference ones.

Example 4.6 *(2D Riemann problems III and IV).* The last two problems are two perturbed versions of a classic deterministic Riemann problem with initial conditions [23]

$$(\rho, u, v, p)(x, y, 0) = \begin{cases} (0.5197, 0.1, 0.1, 0.1, 0.4), & x > 0.5, y > 0.5, \\ (1, -0.6259, 0.1, 1), & x < 0.5, y > 0.5, \\ (0.8, 0.1, 0.1, 1), & x < 0.5, y < 0.5, \\ (1, 0.1, -0.6259, 1), & x > 0.5, y < 0.5, \end{cases} \tag{4.9}$$

which describe the interaction of two rarefaction waves and two contact discontinuities.

The first case takes $\Gamma = 1.4$ and assumes that the initial density $\rho(x, y, 0)$ given in (4.9) contains "ten percent" uncertainty, that is, it is perturbed to $(1 + 0.1\xi)\rho(x, y, 0)$. The gPC-SG method is used to study the effect of this random inputs on the flow structure. Fig. 4.12 displays the contours of numerical mean and standard deviation of the density at time $t = 0.2$ by using the gPC-SG method with $M = 4$ and 250×250 uniform cells, and the reference ones given by the collocation method with 400×400 uniform cells. Fig. 4.13 gives the mean and standard deviation of the density along the line $y = x$. We see that the results given by the gPC-SG method agree well with the reference solutions.

The second case considers certain initial data (4.9), and uncertain adiabatic index

$$\Gamma(\xi) = 1.4 + 0.1\xi.$$

The contours of numerical mean and standard deviation of the density at time $t = 0.2$ given by the gPC-SG method with $M = 4$ and 250×250 uniform cells are displayed in Fig. 4.14, where the reference solutions are given by the collocation method with 640×640 uniform cells and $N_{gs} = 64$. The numerical results exhibit the good performance of the proposed gPC-SG method in resolving 2D flow structures and quantifying the uncertainties. For a further comparison, the mean and standard deviation of the density are plotted along the line $y = x$, see Fig. 4.15. It can be seen clearly that the means and standard deviations obtained by the two methods are in good agreement.

5. Conclusions

In this paper, an effective gPC stochastic Galerkin method for the system of first-order quasilinear hyperbolic equations with uncertainty is proposed. The advantage of the gPC-SG method is that its Galerkin system of equations is proved to be symmetrically hyperbolic in 1D. This allows one to further conduct proper numerical discretizations in physical space

and time. For 2D, the method can be readily adopted via operator splitting. Several examples for the Euler equations in 1D and 2D, exhibiting complex structure in physical space, are presented to demonstrate the accuracy and effectiveness of the proposed gPC-SG method. The proposed gPC-SG method is built on a non-conservative form of the conservation law. This may induce theoretical and numerical issues associated with non-conservative form for hyperbolic PDEs [18,8,2]. Developing a hyperbolic gPC-SG method for a general system of hyperbolic equations in its conservative form is still an open issue. Moreover, it is very important and challenging to reduce the computational cost of the gPC-SG method in order to handle a number of stochastic parameters.

Acknowledgements

KLW and HZT were partially supported by the National Natural Science Foundation of China (Nos. 91330205, 11421101, 91630310). DX was partially supported by AFOSR (grant No. FA95501410022), DARPA (grant No. N660011524053), NSF (grant No. DMS-1418771).

References

[1] R. Abgrall, P.M. Congedo, A semi-intrusive deterministic approach to uncertainty quantification in non-linear fluid flow problems, J. Comput. Phys. 235 (2013) 828–845.
[2] R. Abgrall, S. Karni, A comment on the computation of non-conservative products, J. Comput. Phys. 229 (2010) 2759–2763.
[3] R. Abgrall, S. Mishra, Uncertainty quantification for hyperbolic systems of conservation laws, in: Rémi Abgrall, Chi-Wang Shu (Eds.), Handbook of Numerical Analysis, vol. 18, Elsevier, 2017, pp. 507–544.
[4] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM J. Numer. Anal. 45 (2007) 1005–1034.
[5] M.J. Castro, E.D. Fernández-Nieto, A.M. Ferreiro, J.A. García-Rodríguez, C. Parés, High order extensions of Roe schemes for two-dimensional nonconservative hyperbolic systems, J. Sci. Comput. 39 (2009) 67–114.
[6] M.J. Castro, U.S. Fjordholm, S. Mishra, C. Parés, Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems, SIAM J. Numer. Anal. 51 (2013) 1371–1391.
[7] M.J. Castro, J.M. Gallardo, C. Parés, High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Application to shallow-water systems, Math. Comput. 75 (2006) 1103–1134.
[8] M.J. Castro, P.G. LeFloch, M.L. Muñoz-Ruiz, C. Parés, Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes, J. Comput. Phys. 227 (2008) 8107–8129.
[9] Q.-Y. Chen, D. Gottlieb, J.S. Hesthaven, Uncertainty analysis for the steady-state flows in a dual throat nozzle, J. Comput. Phys. 204 (2005) 387–398.
[10] A. Chertock, S. Jin, A. Kurganov, An operator splitting based stochastic Galerkin method for the one-dimensional compressible Euler equations with uncertainty, preprint, 2015. Available at http://www.math.wisc.edu/~jin/research.html.
[11] A. Chertock, S. Jin, A. Kurganov, A well-balanced operator splitting based stochastic Galerkin method for the one-dimensional Saint-Venant system with uncertainty, preprint, 2015. Available at http://www.math.wisc.edu/~jin/research.html.
[12] G. Dal Maso, P.G. LeFloch, F. Murat, Definition and weak stability of nonconservative products, J. Math. Pures Appl. 74 (1995) 483–548.
[13] B. Després, G. Poëtte, D. Lucor, Robust uncertainty propagation in systems of conservation laws with the entropy closure method, in: Uncertainty Quantification in Computational Fluid Dynamics, in: Lect. Notes Comput. Sci. Eng., vol. 9, Springer, Heidelberg, 2013, pp. 105–149.
[14] F.G. Fuchs, A.D. McMurry, S. Mishra, N.H. Risebro, K. Waagan, High order well-balanced finite volume schemes for simulating wave propagation in stratified magnetic atmospheres, J. Comput. Phys. 229 (2010) 4033–4058.
[15] R.G. Ghanem, P. Spanos, Stochastic Finite Elements: a Spectral Approach, Springer-Verlag, 1991.
[16] D. Gottlieb, D. Xiu, Galerkin method for wave equations with uncertain coefficients, Commun. Comput. Phys. 3 (2008) 505–518.
[17] E. Han, J.Q. Li, H.Z. Tang, Accuracy of the adaptive GRP scheme and the simulation of 2-D Riemann problems for compressible Euler equations, Commun. Comput. Phys. 10 (2011) 577–606.
[18] T.Y. Hou, P.G. LeFloch, Why nonconservative schemes converge to wrong solutions: error analysis, Math. Comput. 62 (1994) 497–530.
[19] J. Hu, S. Jin, A stochastic Galerkin method for the Boltzmann equation with uncertainty, J. Comput. Phys. 315 (2016) 150–168.
[20] J. Hu, S. Jin, D. Xiu, A stochastic Galerkin method for Hamilton–Jacobi equations with uncertainty, SIAM J. Sci. Comput. 37 (2015) A2246–A2269.
[21] S. Jin, D. Xiu, X. Zhu, A well-balanced stochastic Galerkin method for scalar hyperbolic balance laws with random inputs, J. Sci. Comput. 67 (2016) 1198–1218.
[22] R. Käppeli, S. Mishra, Well-balanced schemes for the Euler equations with gravitation, J. Comput. Phys. 259 (2014) 199–219.
[23] P.D. Lax, X.D. Liu, Solution of two-dimensional Riemann problems of gas dynamics by positive schemes, SIAM J. Sci. Comput. 19 (1998) 319–340.
[24] P.G. LeFloch, Entropy weak solutions to nonlinear hyperbolic systems under nonconservative form, Commun. Partial Differ. Equ. 13 (1988) 669–727.
[25] O.P. Le Maitre, O.M. Knio, H.N. Najm, R.G. Ghanem, Uncertainty propagation using Wiener–Haar expansions, J. Comput. Phys. 197 (2004) 28–57.
[26] S. Mishra, C. Schwab, J. Šukys, Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions, J. Comput. Phys. 231 (2012) 3365–3388.
[27] F. Nobile, R. Tempone, C.G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (2008) 2309–2345.
[28] C. Parés, Numerical methods for nonconservative hyperbolic systems: a theoretical framework, SIAM J. Numer. Anal. 44 (2006) 300–321.
[29] G. Poëtte, B. Després, D. Lucor, Uncertainty quantification for systems of conservation laws, J. Comput. Phys. 228 (2009) 2443–2467.
[30] P. Pettersson, G. Iaccarino, J. Nordström, A stochastic Galerkin method for the Euler equations with Roe variable transformation, J. Comput. Phys. 257 (2014) 481–500.
[31] R. Pulch, D. Xiu, Generalised polynomial chaos for linear conservation laws, J. Sci. Comput. 51 (2012) 293–312.
[32] M.T. Reagan, H.N. Najm, R.G. Ghanem, O.M. Knio, Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection, Combust. Flame 132 (2003) 545–555.
[33] C.W. Schulz-Rinne, J.P. Collins, H.M. Glaz, Numerical solution of the Riemann problem for two-dimensional gas dynamics, SIAM J. Sci. Comput. 14 (1993) 1394–1414.
[34] C.-W. Shu, Total-variation-diminishing time discretizations, SIAM J. Sci. Stat. Comput. 9 (1988) 1073–1084.
[35] A.T. Sornborger, E.D. Stewart, Higher-order methods for simulations on quantum computers, Phys. Rev. A 60 (1999) 1956–1965.
[36] M. Thalhammer, M. Caliari, C. Neuhauser, High-order time-splitting Hermite and Fourier spectral methods, J. Comput. Phys. 228 (2009) 822–832.

[37] X. Wan, G.E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, J. Comput. Phys. 209 (2005) 617–642.
[38] K.L. Wu, H.Z. Tang, High-order accurate physical-constraints-preserving finite difference WENO schemes for special relativistic hydrodynamics, J. Comput. Phys. 298 (2015) 539–564.
[39] K.L. Wu, H.Z. Tang, Physical-constraint-preserving central discontinuous Galerkin methods for special relativistic hydrodynamics with a general equation of state, Astrophys. J. Suppl. Ser. 228 (2017) 3.
[40] K.L. Wu, Z.C. Yang, H.Z. Tang, A third-order accurate direct Eulerian GRP scheme for the Euler equations in gas dynamics, J. Comput. Phys. 264 (2014) 177–208.
[41] T. Xiong, C.-W. Shu, M.P. Zhang, WENO scheme with subcell resolution for computing nonconservative Euler equations with applications to one-dimensional compressible two-medium flows, J. Sci. Comput. 53 (2012) 222–247.
[42] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.
[43] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (2005) 1118–1139.
[44] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2002) 619–644.
[45] X. Zhang, C.-W. Shu, On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes, J. Comput. Phys. 229 (2010) 8918–8934.
[46] T. Zhang, Y.X. Zheng, Conjecture on the structure of solutions of the Riemann problem for two-dimensional gas dynamics systems, SIAM J. Math. Anal. 21 (1990) 593–630.
[47] J. Zhu, X. Zhong, C.-W. Shu, J. Qiu, Runge–Kutta discontinuous Galerkin method with a simple and compact Hermite WENO limiter, Commun. Comput. Phys. 19 (2016) 944–969.