



# An adaptive quadrature-based moment closure

Jonas Kusch<sup>1</sup> · Martin Frank<sup>1</sup>

Published online: 15 October 2019  
© Indian Institute of Technology Madras 2019

**Abstract** Methods to numerically quantify uncertainties in hyperbolic equations can be divided into intrusive and non-intrusive techniques. Standard intrusive methods such as Stochastic Galerkin yield oscillatory solutions in the vicinity of shocks and require a new implementation. The more advanced Intrusive Polynomial Moment (IPM) method necessitates a costly solution reconstruction, but promises bounds on oscillatory over- and undershoots. Non-intrusive methods such as Stochastic Collocation (SC) can suffer from aliasing errors, and their black-box nature comes at the cost of losing control over the time evolution of the solution. In this paper, we derive an intrusive method, which adaptively switches between SC and IPM updates by locally refining the quadrature set on which the solution is calculated. The IPM reconstruction of the solution is performed on the quadrature set and uses suitable basis vectors, which reduces numerical costs and allows non-oscillatory reconstructions. We test the method on Burger's equation, where we obtain non-oscillating solution approximations fulfilling the maximum principle.

**Keywords** Conservation laws · Hyperbolic · Oscillations · Aliasing · Collocation · Intrusive · Closure

## 1 Introduction

Hyperbolic equations play an important role in various research areas, ranging from the Euler equations in fluid mechanics to the magnetohydrodynamics (MHD) equations in plasma physics. The general form of a scalar, one-dimensional hyperbolic equation is

$$\partial_t u(t, x) + \partial_x f(u(t, x)) = 0, \quad \text{with } u(t = 0, x) = u_{IC}(x), \quad (1)$$

where  $u : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}$  is the unknown solution which depends on time  $t \in \mathbb{R}^+$  and the spatial variable  $x \in \mathcal{D} \subset \mathbb{R}$ . The physical flux is given by  $f : \mathbb{R} \rightarrow \mathbb{R}$ . To allow discontinuous solutions, the notion of weak solutions is used. A weak solution fulfills

$$\int_{\mathbb{R}^+ \times \mathbb{R}} u \phi_t + f(u) \phi_x \, dx dt + \int_{\mathbb{R}} u_{IC}(x) \phi(0, x) \, dx = 0, \quad \forall \phi \in C_0^1(\mathbb{R}^2), \quad (2)$$

meaning that  $u$  does not need to be differentiable in the spatial variable and time. Since condition (2) allows multiple possible solutions, the concept of entropy solutions has been introduced to pick a unique physically meaningful weak solution. A convex function  $s : \mathbb{R} \rightarrow \mathbb{R}$  is called an entropy function for (1) if the integrability condition  $s'(u)f'(u) = F'(u)$  holds. If a weak solution fulfills

$$\partial_t s(u(t, x)) + \partial_x F(u(t, x)) \leq 0$$

for all entropy functions  $s$ , it is called an entropy solution. For further details, see [1, Chapter 3.8]. Important properties of the entropy solution  $u$  are

✉ Jonas Kusch  
jonas.kusch@kit.edu

Martin Frank  
martin.frank@kit.edu

<sup>1</sup> Karlsruhe Institute of Technology, Karlsruhe, Germany

$$\|u(t, \cdot) - v(t, \cdot)\|_{L^1(\mathcal{D})} \leq \|u_{IC} - v_{IC}\|_{L^1(\mathcal{D})}, \tag{3a}$$

$$\|u(t, \cdot)\|_{L^\infty(\mathcal{D})} \leq \|u_{IC}\|_{L^\infty(\mathcal{D})}, \tag{3b}$$

where  $v$  is an entropy solution to (1) with initial condition  $v_{IC}$  [2, Chapter 2.4]. Property (3a), known as  $L^1$  stability, guarantees boundedness of oscillations, playing a key role in the construction of numerical schemes for solving scalar problems (1) as this ensures convergence to a unique, physically meaningful solution [see (1, Chapter 15.3)]. Property (3b) is known as the maximum principle and ensures bounds on the solution. Numerical schemes which are constructed to satisfy these bounds can be found in [3–7].

To incorporate the fact that solutions to (1) are in practice influenced by uncertainties, various methods to calculate moments of the solutions such as expected value and variance have been constructed. Uncertainties can enter through model parameters as well as initial or boundary conditions. Since uncertain parameters can be expressed by uncertain initial conditions [see (8)], this type of problem is studied in the following. The general form of a scalar hyperbolic conservation law with uncertain initial condition is

$$\begin{aligned} \partial_t u(t, x, \xi) + \partial_x f(u(t, x, \xi)) &= 0, \\ \text{with } u(t = 0, x, \xi) &= u_{IC}(x, \xi), \end{aligned} \tag{4}$$

where  $\xi \in \Theta$  is a random variable with the probability density function  $f_\Xi(\xi)$ . To avoid a high-dimensional solution, one aims at calculating the moments of  $u$  for a given set of basis functions  $\varphi_i(\xi)$  where  $i \in \{0, 1, \dots, N\}$  is the polynomial order. A common choice of the basis functions are so-called chaos polynomials [9], which for smooth data guarantee spectral convergence toward the exact solution when using the moments as coefficients of a polynomial approximation, see [10]. The moments are defined to be

$$\hat{u}_i(t, x) := \int_{\Theta} u(t, x, \xi) \varphi_i(\xi) f_\Xi(\xi) d\xi \quad \text{for } i = 0, \dots, N. \tag{5}$$

Methods to calculate these moments can be divided into two classes, namely intrusive and non-intrusive methods. A standard non-intrusive method is the so-called Stochastic Collocation (SC) method, see [11–14]. It makes use of a quadrature rule to calculate the moments, meaning that

$$\hat{u}_i(t, x) \approx \sum_{k=1}^{N_q} u(t, x, \xi_k) \varphi_i(\xi_k) f_\Xi(\xi_k) w_k, \tag{6}$$

where the solution  $u$  needs to be calculated at  $N_q$  quadrature points  $\xi_k$  with the corresponding quadrature weights  $w_k$ . Since the solution evaluated at a fixed

quadrature point  $\xi_k$  is deterministic, the time evolution of the moments can be described by standard deterministic solvers applied to equation (4) for fixed  $\xi$ . Hence, the deterministic function  $u(t, x, \xi_k)$  in (6) is given by the solution of

$$\begin{aligned} \partial_t u(t, x, \xi_k) + \partial_x f(u(t, x, \xi_k)) &= 0, \\ \text{with } u(t = 0, x, \xi_k) &= u_{IC}(x, \xi_k). \end{aligned}$$

Given a numerical scheme to solve the deterministic problem (1), the implementation of Stochastic Collocation can be treated as a black-box approach. Furthermore, the resulting computer program is embarrassingly parallel, allowing an efficient computation of the SC solution. The error of this solution is given by the classical quadrature error. As a consequence, the gPC approximation can suffer from aliasing effects, see [15, Chapter 2]. Furthermore, the black-box approach which facilitates the implementation comes at the cost of lacking control over the solution at intermediate time steps.

Intrusive methods aim at solving the moment system, which describes the exact time evolution of the moments. This system can be derived by multiplying (4) with the basis functions  $\varphi_i$  for  $i = 0, \dots, N$  as well as the probability density  $f_\Xi$  and integrating over  $\xi$ . Using the vector notation  $\hat{\mathbf{u}} = (\hat{u}_0, \dots, \hat{u}_N)^T$  and  $\boldsymbol{\varphi} = (\varphi_0, \dots, \varphi_N)^T$ , the resulting moment system becomes

$$\partial_t \hat{\mathbf{u}}(t, x) + \partial_x \int_{\Theta} f(u(t, x, \xi)) \boldsymbol{\varphi}(\xi) f_\Xi(\xi) d\xi = \mathbf{0}, \tag{7a}$$

$$\hat{\mathbf{u}}(t = 0, x) = \int_{\Theta} u_{IC}(x, \xi) \boldsymbol{\varphi}(\xi) f_\Xi(\xi) d\xi. \tag{7b}$$

This system describes the exact time evolution of the first  $N + 1$  moments. However, to actually solve the moment system (7), one needs to find a closure, as the flux term depends on the unknown solution  $u$ , whereas the only known variables are the first  $N + 1$  moments. Hence, one needs to find a function

$$\mathcal{U}(\hat{u}_0, \dots, \hat{u}_N) \approx u \tag{8}$$

to be able to describe the time evolution of the first  $N + 1$  moments via the moment system (7). Plugging the closure (8) into the moment system (7) yields the closed moment system

$$\partial_t \hat{\mathbf{u}}(t, x) + \partial_x \int_{\Theta} f(\mathcal{U}(\hat{\mathbf{u}})) \boldsymbol{\varphi}(\xi) f_\Xi(\xi) d\xi = \mathbf{0}, \tag{9}$$

which can now be solved as the  $N + 1$  equations describe the time evolution of  $N + 1$  moments. The closure  $\mathcal{U}$  should be constructed such that the so-called moment constraint

$$\hat{\mathbf{u}} \stackrel{\dagger}{=} \int_{\Theta} \mathcal{U}(\hat{\mathbf{u}}) \boldsymbol{\varphi}(\xi) f_{\Xi}(\xi) d\xi$$

is fulfilled while guaranteeing a hyperbolic moment system. A well-known closure which satisfies these two desirable properties for scalar equations is the Stochastic Galerkin (SG) method [16]

$$\mathcal{U}(\hat{\mathbf{u}}) = \sum_{i=0}^N \hat{u}_i \varphi_i. \tag{10}$$

If all integrals appearing in the moment system (9) can be computed exactly, the method does not suffer from interpolation errors. Consequently, we replace the aliasing error of SC methods with a closure error. Smooth problems can nicely be approximated using the SG method, see, for example, [17]. Spectral convergence has been proven in [18] for kinetic equations and in [8] for Burgers’ equation assuming sufficiently smooth solutions. However, solutions of hyperbolic equations tend to develop discontinuities in the physical as well as the random space. Therefore, the Stochastic Galerkin method generally converges slowly since the closure (10) suffers from Gibbs phenomena. The oscillatory over- and undershoots furthermore violate important properties of the entropy solution as given in (3), hence violating fundamental analytic findings for hyperbolic equations. Furthermore, the implementation of the SG method requires writing an entirely new code since the deterministic solver cannot be reused. All these shortcomings of Stochastic Galerkin indicate that the Stochastic Galerkin method is not suited for closing the moment system of hyperbolic problems. A non-oscillatory closure which fulfills the maximum principle is the Intrusive Polynomial Moment (IPM) closure [19]. For applications and further properties of the IPM method, see [8, 19–24]. The IPM closure is constructed such that

$$\mathcal{U}(\hat{\mathbf{u}}) = \arg \min_u \int_{\Theta} s(u) f_{\Xi} d\xi \text{ s.t. } \hat{\mathbf{u}} = \int_{\Theta} u \boldsymbol{\varphi}_{\Xi} d\xi,$$

where  $s(u)$  is a convex entropy for the deterministic problem (1). To calculate the closure, one makes use of the adjoint approach, which lets us rewrite the closure problem as

$$\mathcal{U}(\hat{\mathbf{u}}) = (s')^{-1}(\hat{\boldsymbol{\lambda}}^T \boldsymbol{\varphi}), \tag{11a}$$

$$\hat{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda} \in \mathbb{R}^{N+1}} \left( \int_{\Theta} s^*(\boldsymbol{\lambda}^T \boldsymbol{\varphi}) f_{\Xi} d\xi - \boldsymbol{\lambda}^T \hat{\mathbf{u}} \right). \tag{11b}$$

Here, the Legendre transformation of the entropy  $s^*(\boldsymbol{\lambda}^T \boldsymbol{\varphi})$  is used. For more details on the adjoint approach, see [25, Chapter 5]. The need to solve an optimization problem with  $N + 1$  parameters to calculate the IPM closure (11) makes the IPM method numerically expensive, especially

if a large number of moments are needed. However, nice approximation properties, the hyperbolicity of the IPM moment system and the maximum principle make IPM an attractive method for hyperbolic problems.

In this paper, we connect intrusive and non-intrusive methods to come up with a strategy to combine benefits of both approaches, i.e., saving computational costs while achieving a good solution approximation. We review the connection of collocation and intrusive methods: When discretizing the moment system (7) with an inaccurate quadrature rule, the computed solution will match the collocation result. To avoid the resulting aliasing errors in regions with high uncertainty, the quadrature is locally refined, meaning that we switch to intrusive methods. Consequently, a closure needs to be used. The arising computationally expensive calculation of a non-oscillatory closure on the refined quadrature set is performed by introducing a set of basis vectors spanning the solution at the quadrature points. By choosing basis functions lying in the null space of the moment constraint, the numerical reconstruction costs can be reduced. Computing the closure on the quadrature points allows choosing reconstructions with minimal total variation. The resulting method is intrusive; however, when switching to the collocation solution, we circumvent the expensive reconstruction step of the standard intrusive method.

The paper is structured as follows: After the introduction in Sect. 1, we investigate the numerical approximation of integrals showing up in the moment system in Sect. 2. This investigation shows a connection between intrusive and non-intrusive methods. In Sect. 3, we introduce the quadrature-based closure (QBC) approach. The implementation of the closure as well as the quadrature refinement is discussed in Sect. 4. Section 5 shows the connection between the QBC and the IPM method. In Sect. 6, we present numerical results, and Sect. 7 summarizes our findings and gives an outlook on future work.

## 2 Transition from stochastic collocation to Stochastic Galerkin

We first investigate the moment system (7) discretized in the random variable  $\xi$ . This gives a connection between collocation and intrusive methods, which we use to discuss potential error sources of collocation methods. Furthermore, we use this link to later adaptively switch between a collocation and an intrusive method. Note that this connection is well known in various areas such as Discontinuous Galerkin (DG) (see, for example, [26]) and is reviewed in this section for the sake of completeness.

Let us start by looking at the moment system (7), when making use of the definition of the moments (5), which gives

$$\partial_t \int_{\Theta} u(t, x, \xi) \varphi_i(\xi) f_{\Xi}(\xi) d\xi + \partial_x \int_{\Theta} f(u(t, x, \xi)) \varphi_i(\xi) f_{\Xi}(\xi) d\xi = 0$$

for  $i = 0, \dots, N$ . This system describes the exact time evolution of the first  $N + 1$  moments. In order to calculate the time evolution numerically, one needs to discretize the moment system. For now,  $x$  and  $t$  remain continuous and we only discretize the integral evaluations with the help of a quadrature rule

$$\int_{\Theta} g(\xi) f_{\Xi}(\xi) d\xi \approx \sum_{k=1}^{N_q} w_k g(\xi_k) f_{\Xi}(\xi_k),$$

where  $\xi_k$  is the  $k$ th quadrature point with the corresponding quadrature weight  $w_k$ . Approximating the terms of the moment system with this quadrature yields

$$\begin{aligned} \partial_t \sum_{k=1}^{N_q} w_k u(t, x, \xi_k) \varphi_i(\xi_k) f_{\Xi}(\xi_k) \\ + \partial_x \sum_{k=1}^{N_q} w_k f(u(t, x, \xi_k)) \varphi_i(\xi_k) f_{\Xi}(\xi_k) = 0. \end{aligned} \tag{12}$$

Now we must choose a suitable number of quadrature points  $N_q$  to capture the correct time evolution of the moments. First, we see what happens if  $N_q$  equals the number of moments, meaning that  $N_q = N + 1$ . This appears to be a striking choice, since it means that especially the time evolution equations for the higher moments are approximated poorly. Nevertheless, we discuss this case because it yields the Stochastic Collocation solution:

**Proposition 1** Assume a quadrature rule with  $N + 1$  quadrature points being exact for polynomials  $p \in \mathbb{P}^{2N}$  under the integral density  $f_{\Xi}$  exists, i.e.,

$$\int_{\Theta} p(\xi) f_{\Xi}(\xi) d\xi = \sum_{k=1}^{N+1} w_k p(\xi_k) f_{\Xi}(\xi_k). \tag{13}$$

When making use of this quadrature rule to discretize the exact moment system (7), the resulting time evolution of the moments is equivalent to the evolution of the Stochastic Collocation method (6).

*Proof* Using the quadrature rule (13) to discretize the exact moment system (7) yields the discrete moment system

$$\begin{aligned} \partial_t \sum_{k=1}^{N+1} w_k u(t, x, \xi_k) \varphi_i(\xi_k) f_{\Xi}(\xi_k) \\ + \partial_x \sum_{k=1}^{N+1} w_k f(u(t, x, \xi_k)) \varphi_i(\xi_k) f_{\Xi}(\xi_k) = 0. \end{aligned} \tag{14}$$

Making use of the matrix  $A \in \mathbb{R}^{(N+1) \times (N+1)}$  with  $a_{ik} := w_k \varphi_i(\xi_k) f_{\Xi}(\xi_k)$ , the discrete moment system can be rewritten as

$$\sum_{k=1}^{N+1} a_{ik} (\partial_t u(t, x, \xi_k) + \partial_x f(u(t, x, \xi_k))) = 0. \tag{15}$$

The inverse of  $A$  exists and is given by  $A^{-1} = (\varphi_i(\xi_j))_{i,j=0,\dots,N}$ , since

$$(AA^{-1})_{ij} = \sum_{k=1}^{N+1} \varphi_i(\xi_k) \varphi_j(\xi_k) f_{\Xi}(\xi_k) w_k \stackrel{(13)}{=} \int_{\Theta} \varphi_i(\xi) \varphi_j(\xi) f_{\Xi}(\xi) d\xi = \delta_{ij}.$$

Multiplication of the discrete moment system (15) with  $A^{-1}$  decouples this system leading to

$$\partial_t u(t, x, \xi_k) + \partial_x f(u(t, x, \xi_k)) = 0 \text{ for } k = 1, \dots, N + 1, \tag{16}$$

which is the Stochastic Collocation (SC) method.  $\square$

*Remark 1* The extension to multidimensional problems is straightforward when using tensorized grids. However, this strategy cannot be applied for sparse grids.

This result shows that the Stochastic Collocation solution can be seen as the solution of the moment system when making use of an inaccurate quadrature rule. Consequently, especially the time evolution of high-order moments is described poorly due to integration errors when evaluating the physical flux.

Our goal is to improve the time evolution of the moment system in regions with high uncertainty. To gain further accuracy, one needs to add more quadrature points to the discrete moment system (12). Again defining a matrix  $A \in \mathbb{R}^{(N+1) \times N_q}$  with  $a_{ik} := w_k \varphi_i(\xi_k) f_{\Xi}(\xi_k)$ , the discrete moment system (12) can be rewritten as

$$\sum_{k=1}^{N_q} a_{ik} (\partial_t u(t, x, \xi_k) + \partial_x f(u(t, x, \xi_k))) = 0. \tag{17}$$

In this case, the  $N + 1$  equations of the moment system no longer fully determine the time evolution of the solution  $u$ , due to the fact that one needs to know this solution at  $N_q > N + 1$  quadrature points. Hence, to obtain a better integral approximation of the flux, we need to feed additional information to the moment equations (i.e., define a closure). In order to actually achieve an improved result compared to collocation, this information should not violate the moment constraint and at the same time fulfill properties of the entropy solution (3). Before specifying the choice of the remaining degrees of freedom, we derive a method for choosing the degrees of freedom in  $u(t, x, \xi_k)$  for  $k = 1, \dots, N_q$  without violating the moment constraint.

*Remark 2* The inexact integration of the flux function when using collocation methods also arises in DG methods when applying nodal instead of modal schemes: Nodal DG schemes can be seen as collocation methods. They suffer from quadrature errors of the physical fluxes, which can lead to poor results as well as stability issues. A strategy to avoid these problems has been presented in [26], where the flux evaluation uses the solution values belonging to a lower-order polynomial (i.e., high-order coefficients are set to zero), guaranteeing an exact integration. The approach taken in this paper is different since we intend to keep all moments fixed but add quadrature points to obtain an improved accuracy of integral approximations.

### 3 Quadrature-based closure approach

The moment constraint computed with the chosen quadrature rule (13) can be rewritten as

$$\hat{\mathbf{u}}(t, x) = \sum_{k=1}^{N_q} w_k \boldsymbol{\varphi}(\xi_k) f_{\Xi}(\xi_k) \mathbf{u}(t, x, \xi_k) = \mathbf{A}\mathbf{u}(t, x), \quad (18)$$

where  $\mathbf{u} := (u(t, x, \xi_1), \dots, u(t, x, \xi_{N_q}))^T$  is the solution evaluated at the quadrature points. Now the vector  $\mathbf{u}$  needs to be chosen such that the moment constraint (18) is fulfilled.

The main idea of the quadrature-based closure is to find a suitable basis to span the solution evaluated at the quadrature points. This facilitates the task of adding admissible information to the moment system. For this, we calculate the kernel of  $\mathbf{A}$ , which is given by

$$\text{kern}(\mathbf{A}) = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{N_q-N-1}).$$

Choosing the vectors  $\mathbf{v}_i$  to represent the solution, one needs  $N + 1$  additional linearly independent vectors  $\mathbf{b}_0, \dots, \mathbf{b}_N$  to span the entire solution space  $\mathbb{R}^{N_q}$ .<sup>1</sup> The solution is now given by

$$\mathbf{u}(t, x) = \sum_{i=1}^{N_q-N-1} \alpha_i(t, x) \mathbf{v}_i + \sum_{i=0}^N \beta_i(t, x) \mathbf{b}_i, \quad (19)$$

and hence we shifted the problem of choosing  $\mathbf{u}$  to choosing the coefficient vectors  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . Luckily, the coefficient vector  $\boldsymbol{\beta}$  is directly given by the moment constraint: Plugging the solution reconstruction (19) into the moment constraint (18) leads to

<sup>1</sup> Since  $\mathbf{A}$  is a priori known, all basis vectors can be computed before running the program.

$$\mathbf{A}\mathbf{u}(t, x) = \sum_{i=0}^N \mathbf{A}\mathbf{b}_i \beta_i(t, x) = \mathbf{C}\boldsymbol{\beta}(t, x) \stackrel{!}{=} \hat{\mathbf{u}}(t, x) \quad (20)$$

with  $c_{ij} = \sum_l a_{il} b_{lj}$ , where  $b_{lj} := (\mathbf{b}_j)_l$ . Hence, the coefficients  $\boldsymbol{\beta}$  are directly given by the moment vector  $\hat{\mathbf{u}}$ . The coefficients  $\boldsymbol{\alpha}$  do not influence the moments of the solution and can therefore be picked freely. These coefficients resemble the additional information we need to give to the moment system in order to apply an improved quadrature rule. Plugging  $\mathbf{u}$  into the discretized moment system (12) gives

$$\begin{aligned} \partial_t \hat{u}_i + \partial_x \sum_{k=1}^{N_q} a_{ik} f(u_k(t, x)) &= 0, \\ \mathbf{u}(t, x) &= \sum_{i=1}^{N_q-N-1} \alpha_i(t, x) \mathbf{v}_i + \sum_{i=0}^N (\mathbf{C}^{-1} \hat{\mathbf{u}})_i \mathbf{b}_i. \end{aligned}$$

It can be seen that we do not have an equation describing the time evolution of  $\alpha_i$ . Clearly, the choice of  $\boldsymbol{\alpha}$  is important since it influences the time evolution of  $\hat{\mathbf{u}}$ . This is due to the fact that the vectors  $\mathbf{v}_i$  span the kernel of the moment constraint, but do not lie in the null space of the flux term. The degrees of freedom we obtain by adding more quadrature points to the moment system to improve the Collocation solution can now easily be picked by choosing values for  $\boldsymbol{\alpha}$ . Our first choice yields the Stochastic Galerkin solution.

**Proposition 2** *Representing the solution by (19) while choosing the degrees of freedom  $\boldsymbol{\alpha}$  such that the solution minimizes*

$$\frac{1}{2} \sum_{k=1}^{N_q} w_k u_k^2 f_{\Xi}(\xi_k), \quad (21)$$

yields the Stochastic Galerkin method.

*Proof* We need to pick  $\boldsymbol{\alpha}$  such that we minimize the discrete  $L^2$  norm (21), hence

$$\partial_{\alpha_j} \frac{1}{2} \sum_k (u_k)^2 w_k \stackrel{!}{=} 0,$$

meaning that

$$\sum_{k=1}^{N_q} \left( \sum_{i=1}^{N_q-N-1} \alpha_i(t, x) v_{ki} + \sum_{i=0}^N \beta_i(t, x) b_{ki} \right) w_k v_{jk} \stackrel{!}{=} 0.$$

Hence, one needs to choose  $\boldsymbol{\alpha}$  such that

$$\sum_{i=1}^{N_q-N-1} \alpha_i(t, x) v_{ki} + \sum_{i=0}^N \beta_i(t, x) b_{ki} \in \text{kern}(\mathbf{V}) = \text{span}(\mathbf{b}),$$

therefore  $\alpha_i = 0$ . It remains to show that

$$u_k = (\mathbf{B}\boldsymbol{\beta})_k = (\mathbf{B}\mathbf{C}^{-1}\hat{\mathbf{u}})_k \stackrel{!}{=} \boldsymbol{\varphi}(\xi_k)^T \hat{\mathbf{u}} = u_{SG}(\xi_k),$$

meaning that we must show  $(\mathbf{B}\mathbf{C}^{-1})_{ki} \stackrel{!}{=} \varphi_i(\xi_k)$ . For this, we note that  $\mathbf{B}$  can be written as

$$b_{ki} = \sum_{n=0}^{N+1} \varphi_n(\xi_k) l_{ni},$$

where  $(l_{ni})_{n,i} =: \mathbf{L} \in \mathbb{R}^{(N+1) \times (N+1)}$  is a transformation matrix. Hence,

$$\begin{aligned} c_{ji} &= \sum_{k=1}^{N_q} a_{jk} b_{ki} = \sum_{n=0}^{N+1} \sum_{k=1}^{N_q} w_k \varphi_j(\xi_k) f_{\Xi}(\xi_k) l_{in} \varphi_n(\xi_k) \\ &= \sum_{n=0}^{N+1} l_{ni} \delta_{nj} = l_{ji}. \end{aligned}$$

Defining  $\mathbf{T} \in \mathbb{R}^{(N+1) \times N_q}$  with  $t_{kn} = \varphi_n(\xi_k)$  yields

$$\mathbf{B}\mathbf{C}^{-1} = \mathbf{T}\mathbf{L}\mathbf{L}^{-1} = \mathbf{T}.$$

Hence  $(\mathbf{B}\mathbf{C}^{-1})_{ki} = t_{ki} = \varphi_i(\xi_k)$ , which concludes the proof.  $\square$

*Remark 3* Note that for linear problems (e.g.,  $f(u) = \xi u$ ),  $N + 1$  quadrature points suffice to discretize the SG moment system in  $\xi$ . Therefore,  $\boldsymbol{\alpha}$  has dimension zero, meaning that the SG and collocation yield the same result.

The choice of picking the degrees of freedom such that one minimizes the weighted  $L^2$  norm of the solution is convenient in the context of spectral convergence for smooth solutions. However, it does not fit to the theory of hyperbolic equations, which is based on  $L^1$  properties. Particularly, in non-smooth regimes, the minimization of the  $L^2$  norm leads to oscillatory approximations violating the maximum principle, see Sect. 6.2. Therefore, we choose the resulting degrees of freedom in accordance with the  $L^1$  stability property (3a). When including the random variable  $\xi$ , this stability result becomes

$$\begin{aligned} &\int_{\mathcal{D}} \int_{\Theta} |u(t, x, \xi) - v(t, x, \xi)| f_{\Xi}(\xi) d\xi dx \\ &\leq \int_{\mathcal{D}} \int_{\Theta} |u_{IC}(x, \xi) - v_{IC}(x, \xi)| f_{\Xi}(\xi) d\xi dx, \end{aligned} \tag{22}$$

where  $u$  and  $v$  are entropy solutions of (4) with two different initial conditions  $u_{IC}, v_{IC}$ . One chooses  $v_{IC}(x, \xi) = u_{IC}(x, \xi + h)$ , multiplies (22) with  $1/h$ , and lets  $h$  go to zero to get

$$\int_{\mathcal{D}} \int_{\Theta} |\partial_{\xi} u(t, x, \xi)| f_{\Xi}(\xi) d\xi dx \leq \int_{\mathcal{D}} \int_{\Theta} |\partial_{\xi} u_{IC}(x, \xi)| f_{\Xi}(\xi) d\xi dx,$$

which shows that oscillations w.r.t.  $\xi$  are bounded by oscillations of the initial condition. Therefore,  $\boldsymbol{\alpha}$  is now

picked such that the reconstruction has minimal oscillations: The total variation in  $\xi$  is denoted by

$$\text{TV}(u) := \int_{\Theta} |\partial_{\xi} u(t, x, \xi)| f_{\Xi}(\xi) d\xi.$$

Since we are only interested in the solution on a finite set of quadrature points, the discrete total variation

$$\text{TV}_{\Delta}(\mathbf{u}) = \sum_{k=1}^{N_q} w_k f_{\Xi}(\xi_k) |u_{k+1} - u_k|$$

is used. The coefficient vector  $\boldsymbol{\alpha}$  is now picked such that  $\text{TV}_{\Delta}(\mathbf{u})$  is minimized. The full QBC moment system is then given by

$$\partial_t \hat{u}_i + \partial_x \sum_{k=1}^{N_q} a_{ik} f(u_k(t, x)) = 0, \tag{23a}$$

$$\mathbf{u}(t, x) = \sum_{i=1}^{N_q-N-1} \hat{u}_i(t, x) \mathbf{v}_i + \sum_{i=0}^N (\mathbf{C}^{-1} \hat{\mathbf{u}})_i \mathbf{b}_i, \tag{23b}$$

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \text{TV}_{\Delta}(\mathbf{u}). \tag{23c}$$

To demonstrate the non-oscillatory behavior of the chosen QBC reconstruction (23b) and (23c), we approximate a shock which is shown in Fig. 1: After computing the moments of this shock, the solution is reconstructed on the quadrature points when using the discrete  $L^2$  norm as well as the total variation. The reconstructions match the first ten moments at 15 quadrature points, i.e.,  $\boldsymbol{\alpha} \in \mathbb{R}^5$  needs to be determined. A non-oscillatory discrete solution which nicely represents the shock on the quadrature points is obtained when choosing  $\boldsymbol{\alpha}$  such that the solution minimizes the discrete total variation. When minimizing the  $L^2$  norm, the solution at the quadrature points oscillates, destroying the maximum principle and leading to a poor approximation. Comparing the solution to the continuous Stochastic Galerkin reconstruction  $u_{SG} = \hat{\mathbf{u}}^T \boldsymbol{\varphi}$  shows that the QBC

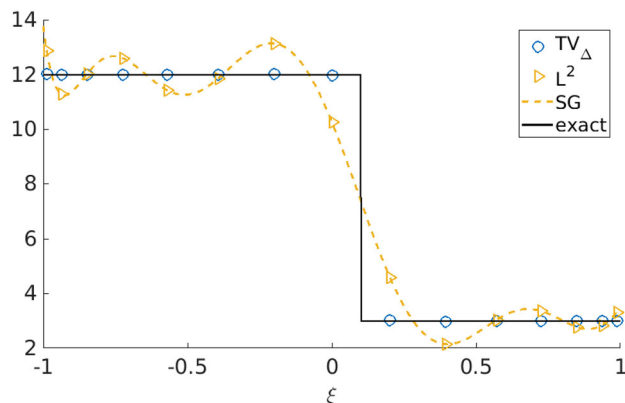


Fig. 1 Reconstructions of QBC

reconstruction ansatz when minimizing the  $L^2$  norm resembles the Stochastic Galerkin solution evaluated at the quadrature points as shown in Proposition 2.

As discussed in Sect. 2, non-intrusive methods yield a poor approximation of the exact moment system. Therefore, our goal is to add quadrature points in regions with high uncertainty. The QBC allows adding these quadrature points without violating the moment constraint or introducing oscillations. In the following section, we discuss the implementation of the refinement process. Furthermore, we point out how the QBC moment system (23) can be implemented efficiently.

### 4 Implementation and refinement

#### 4.1 Implementation

The implementation of the QBC moment system (23) does not require writing an entirely new program. In fact, most of the code of a given finite-volume (FV) implementation for solving the deterministic problem (1) can be reused.

The key idea of FV solvers is to divide the spatial domain into cells. The numerical solution is now a collection of intermediate cell values at discrete time steps  $t_n$

$$u_j^n \simeq \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(t_n, x) dx.$$

Since only the discrete solution for  $n = 0$  is known, a time update formula  $H$ , updating the solution by a time step of  $\Delta t$ , needs to be derived. The FV implementation now calls  $H$  repeatedly until the end time  $t_{\text{end}} = N_t \Delta t$  is reached. A common choice for the update formula is

$$u_j^{n+1} = H(u_{j-1}^n, u_j^n, u_{j+1}^n)$$

where

$$H(u_{j-1}^n, u_j^n, u_{j+1}^n) := u_j^n - \frac{\Delta t}{\Delta x} \left( g(u_j^n, u_{j+1}^n) - g(u_{j-1}^n, u_j^n) \right). \tag{24}$$

The function  $g$  is the numerical flux, which needs to be consistent with the physical flux, i.e.,  $g(u, u) = f(u)$ . For better readability, we assume that the numerical fluxes only depend on two states. The extension to higher order stencils is straightforward.

Our goal is to calculate the solution of the QBC moment system (23). To solve this system, we again use a finite-volume approach, meaning that we represent the moments by

$$\hat{u}_j^n := \mathbf{A} \mathbf{u}_j^n, \tag{25}$$

where  $\mathbf{u}_j^n = (u_{j,1}^n, \dots, u_{j,N_q}^n)^T$  with

$$u_{j,k}^n \simeq \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(t_n, x, \xi_k) dx.$$

One strategy to evolve the moment vector  $\hat{u}_j^n$  in time according to (23) is to reconstruct the solution at every quadrature point, meaning that we compute  $\mathbf{u}$ . On the discrete level, the reconstructed solution is given by

$$\mathbf{u}_j^n = \sum_{i=1}^{N_q-N-1} \hat{\alpha}_{j,i}^n v_i + \sum_{i=0}^N \left( \mathbf{C}^{-1} \hat{\mathbf{u}}_j^n \right)_i \mathbf{b}_i \tag{26}$$

where  $\hat{\alpha}_j^n$  is given by

$$\hat{\alpha}_j^n = \arg \min_{\alpha} \text{TV}_{\Delta}(\mathbf{u}_j^n).$$

After that, the reconstructed solution in every cell at each quadrature point is evolved in time using the time update function of the FV implementation

$$u_{j,k}^{n+1} = H(u_{j-1,k}^n, u_{j,k}^n, u_{j+1,k}^n). \tag{27}$$

The time-updated moments can then be calculated by (25). This process is repeated until  $t_{\text{end}}$  is reached. The extension of a given finite-volume implementation to the QBC scheme can be achieved according to Fig. 2. We use  $\mathbf{u}_{\Delta}^n$  to denote the field containing the solution at all spatial cells and quadrature points for time step  $t_n$ .

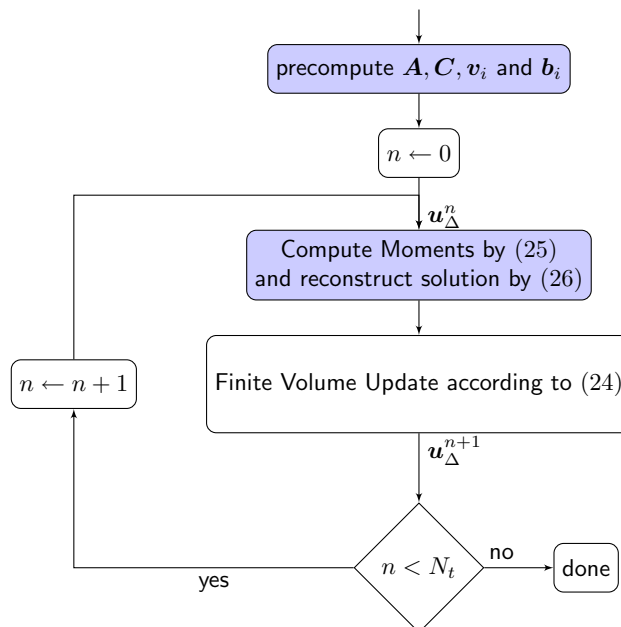


Fig. 2 Integration of the QBC scheme into the finite-volume cycle

### 4.2 Refinement

One can now think about which quadrature should be chosen. For this, we would like to take advantage of the fact that intrusive methods know the behavior of the solution in every time step. In contrast to the non-intrusive black-box approach, adapting the quadrature is possible: If a cell  $I^*$  with ‘high uncertainty’ is detected at spatial position  $x_i$  and time step  $t_m$ , we are now able to refine the quadrature locally by performing the reconstruction step (26) in  $I^*$  with a more accurate quadrature, which we denote by

$$\int_{\Theta} u(\xi) d\xi \approx \sum_{l=1}^{\bar{N}_q} \bar{w}_l u(\bar{\xi}_l),$$

where  $\bar{N}_q > N_q$ . For ease of notation, the solution at the new quadrature points is denoted by

$$\bar{u}_{i,j}^n \simeq \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(t_n, x_j, \bar{\xi}_l) dx.$$

Defining the matrix  $A \in \mathbb{R}^{N+1 \times \bar{N}_q}$  with  $\bar{a}_{il} = \bar{w}_l \varphi_i(\bar{\xi}_l) f_{\Xi}(\bar{\xi}_l)$ , we compute  $\mathbf{v}, \mathbf{b}$  and  $\mathbf{C}$  analogous to before making use of  $A$  instead of  $\mathbf{A}$ . The reconstruction on the finer quadrature set given a moment vector  $\hat{\mathbf{u}}_j^n$  becomes

$$\mathbf{u}_j^n = \sum_{i=1}^{\bar{N}_q - N - 1} \bar{\alpha}_{j,i}^n \mathbf{v}_i + \sum_{i=0}^N (\mathbf{C}^{-1} \hat{\mathbf{u}}_j^n)_i \mathbf{b}_i \tag{28}$$

with

$$\boldsymbol{\alpha} = \arg \min_{\boldsymbol{\alpha}} \text{TV}_{\Delta}(\mathbf{u}).$$

Assume that we wish to perform a time update after each cell has been reconstructed by either the standard or the refined quadrature rule. We now update the solution in the refined cell  $I^*$  in time by

$$\bar{u}_{i,l}^{m+1} = H(\bar{u}_{i-1,l}^m, \bar{u}_{i,l}^m, \bar{u}_{i+1,l}^m) \tag{29}$$

for  $l = 1, \dots, \bar{N}_q$ . Here, one notices that solution values of the neighboring cells of  $I^*$  also need to be known on the refined quadrature points. Therefore, cells which have neighbors with a fine as well as a standard reconstruction need to be reconstructed on both the fine and the standard quadrature set. Using Clenshaw–Curtis quadrature sets

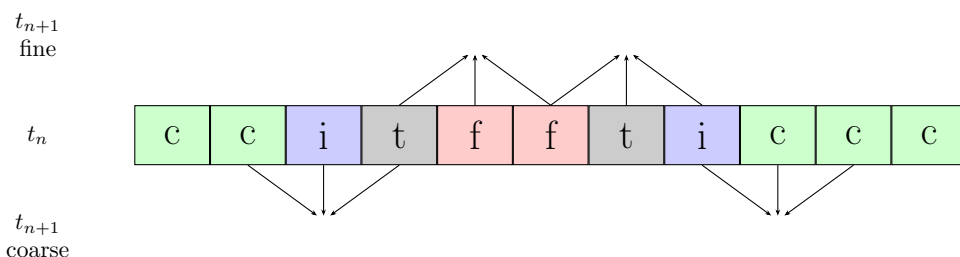
allows reconstructing these interface cells only for the fine quadrature rule since the quadrature points of different refinement levels are nested. In the following, we present the QBC algorithm with refinement for non-nested quadrature rules.

Assume that a spatial grid as depicted in Fig. 3 is given and on each cell we know the moment vector. The cell types are then identified by choosing a function which indicates high uncertainty. Here, we use the highest order moments to identify non-smooth regimes in the uncertain domain similar to [27]. Note that this strategy has also been applied for DG methods [28]. To avoid unwanted artifacts from even or odd functions, we take the last two moments and divide by the solution’s squared  $L^2$  norm as smoothness indicator. If the indicator in a given cell lies above a specified value  $\eta$ , i.e.,

$$\frac{|\hat{u}_N^n| + |\hat{u}_{N-1}^n|}{\mathbf{u}^T \mathbf{u}} > \eta \tag{30}$$

we indicate it as a *fine* cell. Cells which are not *fine*, but have a neighboring *fine* cell are *transition* cells. The cells which have a *transition* neighbor and are not fine are called *interface* cells. All remaining cells are *coarse*. In our example in Fig. 3, the two red cells have a variance lying above a specified value and are therefore *fine* (f). The neighboring cells are identified as *transition* (t), since they have *fine* neighbors, *interface* (i), since they are not fine and have a *transition* neighbor and *coarse* (c). In a next step, the reconstructions need to be computed: We do this on the fine quadrature grid if the cells are *fine* and on the coarse grid if they are *coarse*. For *interface* and *transition* cells, we compute both a coarse and a fine reconstruction. The next task is to evolve the reconstructions in time. One needs to perform a coarse time update for *coarse* and *interface* cells and a fine time update for the remaining cells, which are exemplarily illustrated once for each cell type in Fig. 3. The coarse time update stencil of (27) is depicted by downward pointing, and the fine stencil of (29) is depicted by upward pointing arrows. By looking at the dependencies of the stencils, one sees why it is necessary to reconstruct both the fine and the coarse solution for *interface* and *transition* cells. We then compute the time-updated moments using  $\mathbf{A} \mathbf{u}_j^{n+1}$  on *coarse* and *interface* cells and

Fig. 3 QBC with refinement





$Au_j^{n+1}$  on *transition* and *fine* cells. The entire process is then repeated until the end time is reached.

We use the proposed refinement strategy to locally switch between collocation and standard QBC, i.e., the coarse quadrature set has  $N + 1$  quadrature points. If a cell has high uncertainty, we add quadrature points by our refinement strategy, to improve the time evolution approximation of the moment system. We refer to this method as adaptive QBC. Note that switching to collocation updates does not mean the proposed method is non-intrusive. However, we can circumvent the expensive computation of a closure (i.e., picking  $\alpha \in \mathbb{R}^{N_q - N - 1}$  by solving (23c)), since  $N_q - N - 1 = 0$ .

### 5 Interpretation as IPM closure

The derived closure can be interpreted as an Intrusive Polynomial Moment (IPM) method. This method represents the solution by the ansatz  $\mathcal{U}_{ME}$  which minimizes the convex entropy  $\int_{\Theta} s(u)f_{\Xi}(\xi) d\xi$  under the moment constraints  $\hat{u} = \int_{\Theta} u\varphi_{\Xi}(\xi) d\xi$ , i.e.,

$$\begin{aligned} \mathcal{U}_{ME}(\hat{u}) &= \arg \min_u \int_{\Theta} s(u)f_{\Xi}(\xi) d\xi \\ &\text{subject to } \hat{u} = \int_{\Theta} u\varphi_{\Xi}(\xi) d\xi. \end{aligned} \tag{31}$$

A standard computation using duality converts this problem into an unconstrained finite-dimensional problem. Indeed, using the Lagrangian  $L$ , we have

$$\begin{aligned} \max_{\lambda} \min_u L(u, \lambda) &= \\ \max_{\lambda} \min_u \left( \int_{\Theta} s(u)f_{\Xi}(\xi) d\xi + \lambda^T \left( \hat{u} - \int_{\Theta} u\varphi_{\Xi}(\xi) d\xi \right) \right) \\ &= \min_{\lambda} \left( \int_{\Theta} s^*(\lambda^T \varphi) f_{\Xi}(\xi) d\xi - \lambda^T \hat{u} \right). \end{aligned} \tag{32}$$

The Legendre transform  $s^*$  of the entropy fulfills

$$\int_{\Theta} s^*(\lambda^T \varphi) f_{\Xi}(\xi) d\xi := \int_{\Theta} -s(u_{ME}(\lambda^T \varphi)) + u_{ME}(\lambda^T \varphi) \lambda^T \varphi f_{\Xi}(\xi) d\xi$$

with  $u_{ME}(\lambda^T \varphi) = (s')^{-1}(\lambda^T \varphi)$ . Calculating the closure by solving an infinite-dimensional constrained optimization problem has now been reduced to finding the dual variables  $\lambda$  by solving the dual problem (32). This problem is finite-dimensional as well as unconstrained. Thus, we have

$$\mathcal{U}(\hat{u}) = u_{ME}(\hat{\lambda}^T \varphi)$$

$$\text{with } \hat{\lambda} = \arg \min_{\lambda} \left( \int_{\Theta} s^*(\lambda^T \varphi) f_{\Xi}(\xi) d\xi - \lambda^T \hat{u} \right)$$

Inserting the IPM closure into (7) leads to the closed moment system

$$\partial_t \hat{u} + \partial_x \int_{\Theta} f(\mathcal{U}_{ME}(\hat{u})) \varphi f_{\Xi}(\xi) d\xi = 0,$$

$$\hat{u}(0, x) = \int_{\Theta} u_{IC}(x, \cdot) \varphi f_{\Xi}(\xi) d\xi.$$

This is the system of equations of the IPM method. The solution to the IPM system has nice properties. First, it dissipates the entropy

$$S(t) := \int_{\mathcal{D}} \int_{\Theta} s(\mathcal{U}_{ME}(\hat{u}(t, x))) f_{\Xi}(\xi) d\xi dx.$$

Second, the IPM system is hyperbolic provided that the entropy density  $s$  is strictly convex. Additionally, the IPM method fulfills a maximum principle in the case of scalar problems. However, these two advantages come at the cost of solving an optimization problem with  $N + 1$  unknowns at each time step in every spatial cell. In the following, we interpret the quadrature-based closure approach as an efficient strategy to solve the IPM closure if the number of quadrature points is only slightly larger than the number of moments: First, let us revisit the constraint optimization problem (31). The strategy of the dual ansatz is to parametrize all possible solutions which are potential minimizers of the given entropy. To determine the parameters such that the solution fulfills the moment constraint, the dual problem is solved. We take the approach of first ensuring that the moment constraint is fulfilled and after that, we minimize the entropy: In the case of orthogonal basis functions, one way to construct an ansatz which fulfills the moment constraint is

$$\mathcal{U}(\hat{u}) = \sum_{i=0}^N \hat{u}_i \varphi_i + \sum_{l=N+1}^M \alpha_l \varphi_l. \tag{34}$$

Note that we need to truncate the series by choosing a finite  $M > N$ . Consequently, a truncation error is introduced, meaning that we cannot recover the IPM solution. However, we can avoid introducing a truncation error by first discretizing the integrals of the primal problem. In this case, the solution only needs to be reconstructed on a finite set of quadrature points. Again using  $u_k := u(t, x, \xi_k)$ , one obtains

$$\begin{aligned} \mathcal{U}(\hat{\mathbf{u}}) &= \arg \min_{\mathbf{u} \in \mathbb{R}^{N_q}} \sum_{k=1}^{N_q} w_k s(u_k) f_{\Xi}(\zeta_k) \text{ subject to } \hat{\mathbf{u}} \\ &= \sum_{k=1}^{N_q} w_k u_k \varphi(\zeta_k) f_{\Xi}(\zeta_k). \end{aligned}$$

Now the task of finding a parameterization which fulfills the constraint is easy: We again span the solution vector  $\mathbf{u}$  as in (19), i.e.,

$$\mathbf{u}(t, x) = \sum_{i=1}^{N_q-N-1} \alpha_i(t, x) \mathbf{v}_i + \sum_{i=0}^N \beta_i(t, x) \mathbf{b}_i.$$

The coefficient vector  $\boldsymbol{\beta}$  is uniquely defined by the moment constraint, see (20). Hence, we have a parametrization of all solutions fulfilling the moment constraint, where the number of parameters  $\alpha_i$  is finite, namely  $N_q - N - 1$ . We then choose these parameters such that the discrete entropy

$$\sum_{k=1}^{N_q} w_k s(u_k) f_{\Xi}(\zeta_k)$$

is minimized. In our case, we choose this entropy to be the total variation. The main difference to the IPM method is that we construct the closure on the discretized level, enabling us to find a finite-dimensional parameterization of all admissible reconstructions from which we pick the one with the smallest entropy. Consequently, the problem of finding  $N + 1$  dual variables  $\hat{\boldsymbol{\lambda}}$  has been shifted to finding  $N_q - N - 1$  reconstruction parameters  $\boldsymbol{\alpha}$ . Hence, the computational costs decrease if  $N_q - N - 1 < N + 1$ , i.e., in the transition between collocation and intrusive methods.

### 6 Results

In the following, we present results of the QBC method and compare them to results obtained with the SG method. To compare these different strategies, we follow [19] and solve the uncertain Burgers’ equation

$$\partial_t u(t, x, \xi) + \partial_x \frac{u(t, x, \xi)^2}{2} = 0, \tag{35a}$$

$$u(t = 0, x, \xi) = u_{IC}(x, \xi) \tag{35b}$$

on  $x \in [0, 3]$  with the uncertain initial condition

$$u_{IC}(x, \xi) := \begin{cases} u_L, & \text{if } x < x_0 + \sigma \xi \\ u_L + \frac{u_R - u_L}{x_0 - x_1} (x_0 + \sigma \xi - x), & \text{if } x \in [x_0 + \sigma \xi, x_1 + \sigma \xi] \\ u_R, & \text{else} \end{cases} \tag{36}$$

The initial condition is a forming shock with uncertain position. We choose  $\xi$  to be uniformly distributed in  $\Theta = [-1, 1]$ . To ensure good integral approximations on  $\Theta$ , we use a Gauss–Legendre quadrature rule. Parameters of the initial condition and the numerical scheme can be found in the following table:

$N_x = 500$	Number of spatial cells
$t_{\text{end}} = 0.11$	End time
$x_0 = 0.5, x_1 = 1.5, u_L = 12, u_R = 1, \sigma = 0.3$	Parameters of initial condition (36)
$N + 1 = 15$	Number of moments
$\eta = 5 \times 10^{-4}$	Barrier (30) for refinement

The basis vectors  $\mathbf{v}_i$  of the QBC are calculated by performing a singular value decomposition of  $\mathbf{A}$ , and the vectors  $\mathbf{b}_i$  are chosen as

$$(\mathbf{b}_i)_k = \varphi_i(\zeta_k)$$

for  $i = 0, \dots, N$ .

#### 6.1 Convergence of expected value

In the following, we study the effects of varying the number of quadrature points  $N_q$  to study the  $L^1$  error of the expected value by

$$e := \int_0^3 |\hat{u}_0(t_{\text{end}}, x) - \hat{u}_{0,\text{ex}}(t_{\text{end}}, x)| dx,$$

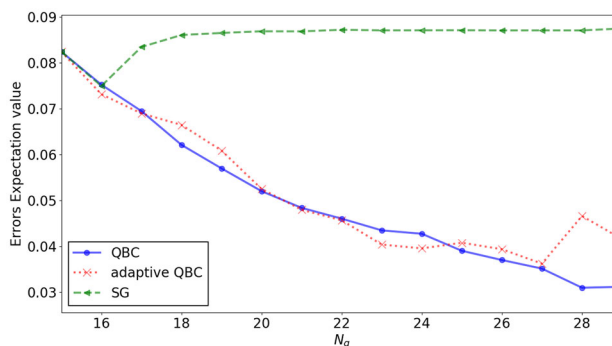
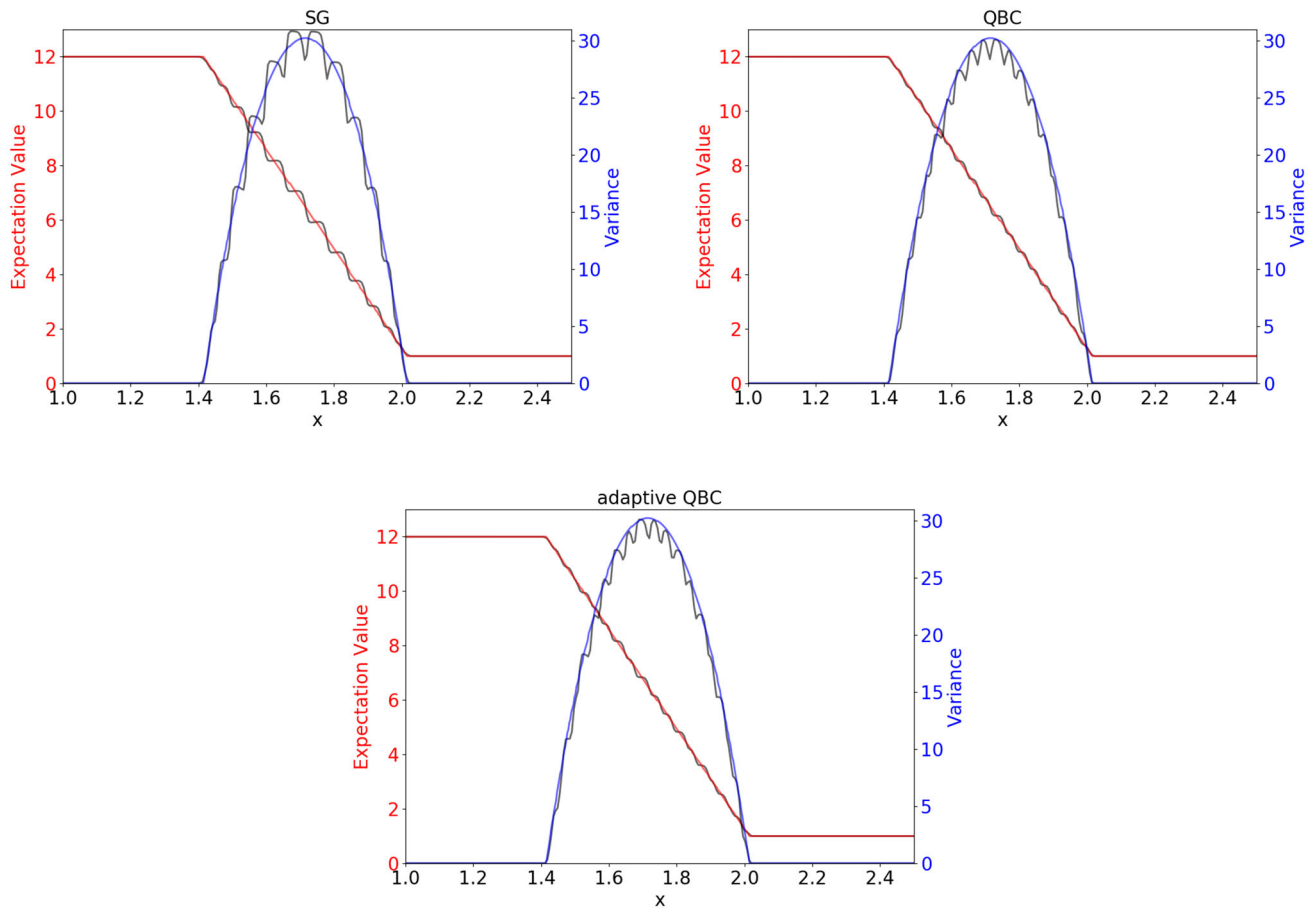


Fig. 4  $L^1$  error of the expected value. The adaptive QBC switches between Collocation (15 quadrature points) and  $N_q$  on the x-axis



**Fig. 5** Expected value and variance for ten moments with 25 quadrature points. The adaptive QBC switches between collocation (15 quadrature points) and 25 quadrature points. The exact expectation value is plotted in red, and the exact variance is plotted in blue

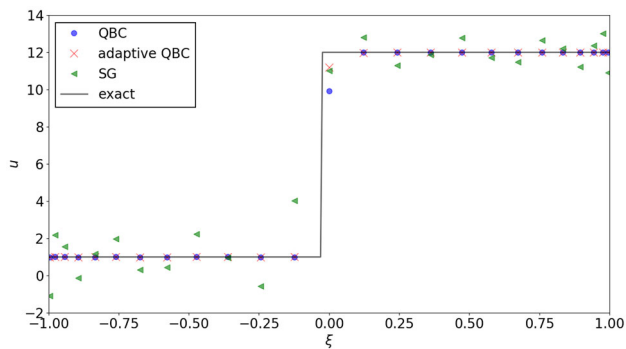
where  $\hat{u}_{0,\text{ex}}(t_{\text{end}}, x)$  is the expected value (which is equal to the zeroth moment) of the exact solution at the final time  $t_{\text{end}}$ . As discussed in Sect. 2, choosing  $N_q > N + 1$  necessitates feeding information to the collocation moment system. The QBC picks these degrees of freedom to minimize the total variation, whereas the SG method minimizes the  $L^2$  norm. When applying the adaptive QBC, we can switch between different quadrature rules to locally refine the set of integration points. In the following, the number of quadrature points on the coarse level  $N_q^c$  equals the number of moments, i.e.,  $N_q^c = N + 1 = 15$ . Hence, we perform collocation updates on the coarse level. We switch to intrusive methods whenever the smoothness indicator lies above a value of  $5 \times 10^{-4}$ . The number of quadrature points on the fine level  $N_q$  is then varied to study the error of the expected value.

The resulting errors for different  $N_q$  are shown in Fig. 4. If  $N_q = 15$ , all three methods resemble Stochastic Collocation, which is why they all have the same error. The methods differ in the choice of degrees of freedom when adding quadrature points. Taking a look at the SG method,

improving the numerical integration has little effect. Moreover, the resulting error will start to grow after having added one quadrature point, leaving the SG method with a poorer approximation as the collocation result. In contrast to this, the QBC steadily decreases the error when adding quadrature points. Hence, the choice of degrees of freedom helps in improving the approximation of the integrals appearing in the moment system. The adaptive QBC shows a similar error decrease. Locally switching to collocation whenever the solution shows low uncertainty decreases the computational costs while leading to a slightly increased error.

## 6.2 Comparison expectation value and variance

In the following, we fix the number of quadrature points at 25. The adaptive QBC switches between collocation as well as QBC with 25 quadrature points. We start by looking at the expectation value and the variance in Fig. 5. The SG method shows a step-like approximation of the expectation value, whereas QBC and adaptive QBC show



**Fig. 6** Solution at fixed spatial position  $x^* = 2.1$

only small variations from the exact solution. Furthermore, one observes that the expected values computed with QBC and its adaptive version differ only slightly. Consequently, the poor approximation of integrals in the moment system in cells with low smoothness indicator does not destroy the accuracy achieved with the standard QBC. A similar result can be found when investigating the variance: While the variance of SG oscillates heavily, both QBC and adaptive QBC show mitigated oscillations.

Lastly, we take a look at the reconstruction of the solution at a fixed spatial position  $x^* = 2.1$  for the final time  $t_{\text{end}}$ . The exact solution is a reversed shock from  $u_R$  to  $u_L$ . Note that since we use a quadrature rule with  $N_q = 25$ , we only need to know the solution at the corresponding 25 quadrature points. A continuous reconstruction of the SG method can be computed by interpolating the given solution points with Legendre polynomials. In the QBC case, a continuous reconstruction can be calculated with the adjoint approach of IPM (11). Since our interest lies in computing the solution's moments, we skip this step. Figure 6 shows that the SG method has great difficulties when representing the reversed shock, because this leads to oscillations. The solution violates the maximum principle and shows high total variation in  $\xi$ . Both the QBC and adaptive QBC yield a reconstruction which does not violate the bounds of the initial condition, namely  $u_R$  and  $u_L$ . The solution nicely captures the exact shock structure. Also note that the adaptive QBC makes use of 25 instead of 15 quadrature points at the shock position, meaning that the spatial cell at  $x^*$  is identified as a *fine* cell. Consequently, the shock position is nicely localized.

## 7 Conclusion and outlook

In this paper, we have presented a method to locally add quadrature points to the collocation solution in regions with high uncertainty, allowing a transition from non-intrusive

to intrusive methods. Refining the quadrature set requires choosing resulting degrees of freedom, which are described by the kernel of the moment constraint. Using this kernel as basis functions for our solution led to an efficient way of choosing the added degrees of freedom. It turned out that the choice of these parameters is crucial. By choosing the parameters such that the  $L^2$  norm of the solution is minimized, we added characteristics which contradict analytic findings of hyperbolic equations. The resulting Stochastic Galerkin solution showed bad approximation results such as oscillations and violations of the maximum principle. Nice approximation results which fulfill the maximum principle were achieved by the quadrature-based closure (QBC) method, which uses the degrees of freedom to minimize the total variation. Adaptive refinement of the quadrature set allows switching between non-intrusive and the derived intrusive QBC method. The adaptive QBC yields an efficient computation of moments and reconstructions, which closely agree with the standard QBC results.

In future work, we plan to use the presented refinement strategy solely on collocation methods. Refining the quadrature set by a reconstruction allows adding collocation points without having to recalculate the entire solution. Furthermore, we wish to investigate different functions indicating high uncertainty such as shock indicators to refine the solution. Additionally, more than one refinement levels should be added. To avoid having to introduce new *transition* and *interface* cells especially for high-order spatial discretizations, one should use a DG implementation or nested quadrature grids. In addition to that, the refinement strategy should be tested on sparse quadrature grids for high-dimensional problems.

**Funding** This work was supported by the German Research Foundation (DFG) under Grant FR 2841/6-1.

## References

1. LeVeque, R.J.: Numerical Methods for Conservation Laws. Basel, Birkhauser (1992)
2. Risebro, H.H., Henrik, N.: Front Tracking for Hyperbolic Conservation Laws, vol. 152. Springer, Berlin (2015)
3. Bell, J.B., Dawson, C.N., Shubin, G.R.: An unsplit, higher order godunov method for scalar conservation laws in multiple dimensions. *J. Comput. Phys.* **74**(1), 1–24 (1988)
4. Colella, P.: Multidimensional upwind methods for hyperbolic conservation laws. *J. Comput. Phys.* **87**(1), 171–200 (1990)
5. Liu, X.-D.: A maximum principle satisfying modification of triangle based adaptive stencils for the solution of scalar hyperbolic conservation laws. *SIAM J. Numer. Anal.* **30**(3), 701–716 (1993)
6. Zhang, X., Shu, C.W.: Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws:

- survey and new developments. *Proc. R. Soc. Math. Phys. Eng. Sci.* **467**(2134), 2752–2776 (2011)
7. Guermond, J.-L., Nazarov, M., Popov, B., Yang, Y.: A second-order maximum principle preserving lagrange finite element technique for nonlinear scalar conservation equations. *SIAM J. Numer. Anal.* **52**(4), 2163–2182 (2014)
  8. Després, B., Poëtte, G., Lucor, D.: Robust uncertainty propagation in systems of conservation laws with the entropy closure method. In: *Uncertainty Quantification in Computational Fluid Dynamics*, pp. 105–149. Springer (2013)
  9. Wiener, N.: The homogeneous chaos. *Am. J. Math.* **60**(4), 897–936 (1938)
  10. Canuto, C., Quarteroni, A.: Approximation results for orthogonal polynomials in sobolev spaces. *Math. Comput.* **38**(157), 67–86 (1982)
  11. Tatang, M.A.: Direct incorporation of uncertainty in chemical and environmental engineering systems, Ph.D. thesis, Massachusetts Institute of Technology (1995)
  12. Mathelin, L., Hussaini, M.Y., Zang, T.A.: Stochastic approaches to uncertainty quantification in cfd simulations. *Numer. Algorithms* **38**(1–3), 209–236 (2005)
  13. Xiu, D., Hesthaven, J.S.: High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.* **27**(3), 1118–1139 (2005)
  14. Babuška, Ivo, Nobile, Fabio, Tempone, Raul: A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.* **45**(3), 1005–1034 (2007)
  15. Canuto, C., Hussaini, M.Y., Quarteroni, A., Zang, T.A.: *Spectral Methods*. Springer, Berlin (2006)
  16. Ghanem, R.G., Spanos, P.D.: *Stochastic Finite Elements: A Spectral Approach*. Courier Corporation, North Chelmsford (2003)
  17. Gottlieb, D., Xiu, D.: Galerkin method for wave equations with uncertain coefficients. *Commun. Comput. Phys.* **3**(2), 505–518 (2008)
  18. Jin, S., Liu, J.-G., Ma, Z.: Uniform spectral convergence of the stochastic galerkin method for the linear transport equations with random inputs in diffusive regime and a micro-macro decomposition-based asymptotic-preserving method. *Res. Math. Sci.* **4**(1), 15 (2017)
  19. Poëtte, G., Després, B., Lucor, D.: Uncertainty quantification for systems of conservation laws. *J. Comput. Phys.* **228**(7), 2443–2467 (2009)
  20. Poëtte, G., Després, B., Lucor, D.: Treatment of uncertain material interfaces in compressible flows. *Comput. Methods Appl. Mech. Eng.* **200**(1), 284–308 (2011)
  21. Kusch, J., Alldredge, G.W., Frank, M.: Maximum-principle-satisfying second-order intrusive polynomial moment scheme. *SMAI J. Comput. Math.* **5**, 23–51 (2019)
  22. Schlachter, L., Schneider, F.: A hyperbolicity-preserving stochastic galerkin approximation for uncertain hyperbolic systems of equations. *J. Comput. Phys.* **375**, 80–98 (2018)
  23. Kusch, J., Frank, M.: Intrusive methods in uncertainty quantification and their connection to kinetic theory. *Int. J. Adv. Eng. Sci. Appl. Math.* **10**(1), 54–69 (2018)
  24. Kusch, J., McClarren, R.G., Frank, M.: Filtered stochastic galerkin methods for hyperbolic equations. (2018) arXiv preprint [arXiv:1808.00819](https://arxiv.org/abs/1808.00819)
  25. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
  26. Gassner, G.J., Beck, A.D.: On the accuracy of high-order discretizations for underresolved turbulence simulations. *Theor. Comput. Fluid Dyn.* **27**(3–4), 221–237 (2013)
  27. Kröker, I., Rohde, C.: Finite volume schemes for hyperbolic balance laws with multiplicative noise. *Appl. Numer. Math.* **62**(4), 441–456 (2012)
  28. Persson, P.O., Peraire, J.: Sub-cell shock capturing for discontinuous Galerkin methods. In: *44th AIAA Aerospace Sciences Meeting and Exhibit*, p. 112. (2006)

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.