A Roe variable based chaos method for the Euler equations under uncertainty

Per Pettersson*†, Gianluca Iaccarino* and Jan Nordström†‡

* Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA.
† Department of Information Technology, Uppsala University, P.O. Box 337, SE-75105 Uppsala, Sweden.
‡ Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden.

Abstract

The Euler equations subject to uncertainty in the input parameters are investigated via the stochastic Galerkin approach. We present a new fully intrusive method based on a variable transformation of the continuous equations. Roe variables are employed to get quadratic dependence in the flux function and a well-defined Roe average matrix that can be determined without matrix inversion.

In previous formulations based on generalized chaos expansion of the physical variables, the need to introduce stochastic expansions of inverse quantities, or square-roots of stochastic quantities of interest, adds to the number of possible different ways to approximate the original stochastic problem. We present a method where no auxiliary quantities are needed, resulting in an unambiguous problem formulation.

The Roe formulation saves computational cost compared to the formulation based on expansion of conservative variables. Moreover, the Roe formulation is more robust and can handle cases of supersonic flow, for which the conservative variable formulation leads to instability. For more extreme cases, where the global Legendre polynomials poorly approximate discontinuities in stochastic space, we use the localized Haar wavelet basis.

1 Introduction

Generalized chaos expansions are frequently used to represent uncertain quantities in numerical solutions of PDEs with uncertainty in e.g. initial- and boundary data,
These quantities of interest are expressed as generalized Fourier series, where orthogonal polynomials (e.g. Legendre or Hermite polynomials) are commonly used as basis functionals. Building on results due to Wiener [1], this is the generalized polynomial chaos method introduced by Ghanem and Spanos [2] which is well suited for smooth problems.

Spectral convergence of the generalized polynomial chaos expansion is observed when the solutions are sufficiently regular and continuous [3], but for general nonlinear conservation laws - such as in fluid dynamics problems - the convergence is usually less favorable. Spectral expansion representations are still of interest for these problems because of the potential efficiency with respect to brute force sampling methods, but special attention must be devoted to the numerical methodology used. For some problems with steep gradients in the stochastic dimensions, polynomial chaos expansions fail entirely to capture the solution [4]. Global methods can still give superior overall performance, for instance Padé approximation methods based on rational function approximation [5], and hierarchical wavelet methods that are global methods with localized support of each resolution level [6, 7].

Alternative approaches to generalized chaos methods have also been presented in the literature. Abgrall et. al. [8, 9] developed a semi-intrusive method based on a finite-volume like reconstruction technique of the discretized stochastic space. A deterministic problem is obtained by taking conditional expectations given a stochastic subcell, over which ENO constructions are used to reconstruct the fluxes in the stochastic dimensions. This makes it particularly suitable for non-smooth probability distributions.

In many nonlinear applications of the stochastic Galerkin method, truncation of the generalized chaos expansion leads to non-unique formulations of the systems of equations. For instance, cubic products between stochastic quantities $a, b$ and $c$, are represented as products of truncated approximations $\tilde{a}, \tilde{b}$ and $\tilde{c}$, but the pseudospectral multiplication operator $\ast$, to be explicitly defined in a later section, is not associative, i.e. $(\tilde{a} \ast \tilde{b}) \ast \tilde{c} \neq \tilde{a} \ast (\tilde{b} \ast \tilde{c})$. Similar problems are investigated in more detail in [10]. The need to introduce stochastic expansions of inverse quantities, or square-roots of stochastic quantities of interest, adds to the number of possible different ways to approximate the original stochastic problem. This leads to ambiguity of the problem formulation. We present a method where this ambiguity is avoided since no auxiliary quantities are needed.

Intrusive generalized chaos methods for nonlinear conservation laws have been investigated in e.g. [11], where a reduced-cost Roe solver with entropy corrector was presented, and in [12] with different localized representations of uncertainty in initial functions and problem coefficients. In previous works we investigated well-posedness and stability for an intrusive formulation of Burgers’ equation with uncertainty in [13], and imposition of uncertain boundary conditions in [14].

Poëtte et. al. [15] used a nonlinear projection method to bound the oscillations close to stochastic discontinuities by polynomial chaos expansion of the entropy variables obtained from a transformation of the conservative variables. Each time step is complemented by a functional minimization to obtain the entropy variables needed to update the solution vector. The method we will present here may appear similar
at first sight, but it relies on a different kind of variable transformation and not on kinetic theory considerations. We do not suggest a variable transformation for general conservation laws, but a formulation that specifically targets the solution of the Euler equations with uncertainty in the variables. It is in fact less complicated than a direct polynomial chaos expansion of the conservative variables.

Variables appearing in the Euler equations are represented as generalized chaos series, using a set of basis functionals that can be either global or localized. The stochastic Galerkin system is obtained by projection of the stochastic Euler equations on the generalized chaos basis functionals. We consider the Sod test case subject to uncertainty in the downstream density, and uncertain diaphragm location, respectively. We employ a MUSCL reconstruction in space [16] and a fourth-order Runge-Kutta method for the time integration. Both a Roe flux based on a stochastic Galerkin Roe average matrix, and a more robust HLL flux, will be employed. A Roe average matrix for the numerical Roe flux is derived and proven to fulfill the conditions of Roe under certain circumstances.

In our method, the system of equations is reformulated using Roe variables so that only quadratic terms occur. No fourth-order tensors need to be approximated or calculated, resulting in increased accuracy and/or reduced computational cost. Moreover, there is no need to compute chaos expansion for any inverse quantities, such as the inverse density. The Roe variable expansion provides a simple and unambiguous formulation of the Euler equations. For brevity of notation, we will refer to our expansion method as the Roe expansion, and the method based on expansion of the conservative variables as the conservative expansion.

In section 2, we present the stochastic framework we use for representing stochastic quantities to be used in the stochastic Galerkin formulations of the Euler equations in conservative and Roe variables, which are introduced in 3. For each one of two test cases, two sets of basis functions are used to represent uncertainty: Legendre polynomials and Haar wavelets. The Legendre basis exactly represents the input uncertainty in our first test case, but it leads to oscillations around the discontinuity in stochastic space. The Haar wavelets do not represent the input uncertainty exactly in either test case, but are more robust to discontinuities. In section 4, the numerical flux functions are introduced. We derive and prove properties of the solvers that are necessary to capture essential dynamics of hyperbolic problems. We use a Roe flux suitable for situations where the system eigenvalues can be accurately estimated, and a more robust HLL flux based on the fastest signal speeds only. Numerical results are presented in section 5, where the previously developed techniques are evaluated.

2 Representation of uncertainty

Let \((\Omega, \mathcal{A}, \mathcal{P})\) be a probability space with event space \(\Omega\), and probability measure \(\mathcal{P}\) defined on the \(\sigma\)-field \(\mathcal{A}\) of subsets of \(\Omega\). Let \(\xi = \{\xi_j(\omega)\}_{j=1}^N\) be a set of \(N\) independent and identically distributed random variables for \(\omega \in \Omega\). Consider a generalized chaos basis \(\{\Psi_i(\xi)\}_{i=0}^\infty\) spanning the space of second order (i.e. finite variance) random processes on this probability space. The basis functionals are assumed to be orthonormal,
i.e. they satisfy
\[ \langle \Psi_i, \Psi_j \rangle = \delta_{ij}, \tag{1} \]
where the inner product between two functionals \( a(\xi(\omega)) \) and \( b(\xi(\omega)) \) is defined by
\[ \langle a(\xi(\omega))b(\xi(\omega)) \rangle = \int_{\Omega} a(\xi(\omega))b(\xi(\omega))dP(\omega). \]

Any second order random field \( u(x,t,\xi) \) can be expressed as
\[ u(x,t,\xi) = \sum_{i=0}^{\infty} u_i(x,t)\Psi_i(\xi). \tag{2} \]

Independent of the choice of basis \( \{\Psi_i\}_{i=0}^{\infty} \), we can express the mean and variance of \( u(x,t,\xi) \) as
\[ E(u(x,t,\xi)) = u_0(x,t), \quad \text{Var}(u(x,t,\xi)) = \sum_{i=1}^{\infty} u_i(x,t)^2, \]
respectively. For practical purposes, (2) is truncated to a finite number \( P \) terms, and we set
\[ u(x,t,\xi) \approx u^P(x,t,\xi) = \sum_{i=0}^{P-1} u_i(x,t)\Psi_i(\xi). \tag{3} \]

### 2.1 Global basis: Legendre polynomials

For smooth problems, the generalized polynomial chaos expansion is attractive due to its spectral convergence. For non-smooth problems such as nonlinear hyperbolic problems, discontinuities will emerge in finite time in stochastic space and polynomial chaos expansions tend to result in Gibbs oscillations. In the worst case, polynomial chaos may fail entirely to capture essential features of the solution [6].

Let \( \{\Psi_i\}_{i=0}^{\infty} \) be the set of Legendre polynomials that are defined on \([-1,1]\) and orthogonal with respect to the uniform measure. The normalized Legendre polynomials can be defined recursively from
\[ \Psi_{j+1}(\xi) = \sqrt{2j + 3} \left( \frac{\sqrt{2j + 1}}{j + 1} \xi \Psi_j(\xi) - \frac{j}{(j+1)\sqrt{2j-1}} \Psi_{j-1}(\xi) \right), \]
\[ \Psi_0 = 1, \Psi_1 = \sqrt{3}\xi. \]

Double products are readily computed from (1), and higher-order products are pre-computed using numerical integration.

The truncated chaos series (3) may result in solutions that are unphysical. An extreme example is when a strictly positive quantity, say density, with uncertainty within a bounded range is represented by a polynomial expansion with infinite range, for instance Hermite polynomials of standard Gaussian variables. The Hermite series expansion converges to the true density with bounded range in the limit \( P \to \infty \), but for a given order of expansion, say \( P = 2 \), the representation \( \rho = \rho_0 + \rho_1 H_1(\xi) \)
results in negative density with non-zero probability since the Hermite polynomial \( H_1 \) takes arbitrarily large negative values. Similar problems may be encountered also for polynomial representations with bounded support. Polynomial reconstruction of a discontinuity in stochastic space leads to Gibbs oscillations that may yield negative values of an approximation of a solution that is close to zero but strictly positive by definition.

### 2.2 Localized basis: Haar wavelet expansion

An alternative to polynomial expansions for non-smooth and oscillatory problems is generalized chaos based on a localization or discretization of the stochastic space \([17, 18]\). Methods based on stochastic discretization include adaptive stochastic multi-elements \([19]\) and stochastic simplex collocation \([20]\). The robust properties of discretized stochastic space can also be obtained by globally defined wavelets, see \([6]\). In this paper, piecewise constant Haar wavelets will be used.

Haar wavelets are defined hierarchically on different resolution levels, representing successively finer features of the solution with increasing resolution. They have non-overlapping support within each resolution level, and in this sense they are localized. Still, the Haar basis is global due to the overlapping support of wavelets belonging to different resolution levels. Haar wavelets do not exhibit spectral convergence, but avoid the Gibbs phenomenon.

Consider the mother wavelet function defined by

\[
\psi^W(y) = \begin{cases} 
1 & 0 \leq y < \frac{1}{2} \\
-1 & \frac{1}{2} \leq y < 1 \\
0 & \text{otherwise},
\end{cases}
\]

Based on (4) we get the wavelet family

\[
\psi^W_{j,k}(y) = 2^{j/2}\psi^W(2^j y - k), \quad j = 0, 1, \ldots; \quad k = 0, \ldots, 2^j - 1,
\]

Given the probability measure of the stochastic variable \( \xi \) with cumulative distribution function \( F_\xi(\xi_0) = P(\omega : \xi(\omega) \leq \xi_0) \), define the basis functions

\[
W_{j,k}(\xi) = \psi^W_{j,k}(F_\xi(\xi))
\]

Adding the basis function \( W_0(y) = 1 \) in \( y \in [0, 1] \) and concatenating the indices \( j \) and \( k \) into \( i = 2^j + k \) so that \( W_i(\xi) = \psi^W_{n,k}(F_\xi(\xi)) \), we can represent any random variable \( u(x, t, \xi) \) with finite variance as

\[
u(x, t, \xi) = \sum_{i=0}^{\infty} u_i(x, t)W_i(\xi),
\]

which is of the form (2).
3 Euler equations with input uncertainty

Consider the 1D Euler equations, in non-dimensional form given by

$$u_t + f(u)_x = 0,$$

where the solution and flux vector are given by

$$u = \begin{bmatrix} \rho \\ \rho v \\ E \end{bmatrix}, \quad f = \begin{bmatrix} \rho v \\ \rho v^2 + p \\ (E + p) v \end{bmatrix},$$

where \(\rho\) is density, \(v\) velocity, \(E\) total energy and pressure \(p\). A perfect gas equation of state is assumed, and energy and pressure are related by

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2,$$

where \(\gamma\) is the ratio of the specific heats. For the numerical method, we will need the flux Jacobian, given by

$$\nabla f = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2}(\gamma - 3)v^2 & (3 - \gamma)v & \gamma - 1 \\ \frac{1}{2}(\gamma - 1)v^3 - vH & H - (\gamma - 1)v^2 & \gamma v \end{bmatrix},$$

with the total enthalpy \(H = (E + p)/\rho\).

We scale the variables to get the dimensionless variables \(\rho = \rho^* / \rho_L^*,\ E = E^* / (\gamma p_L^*),\ p = p^* / (\gamma p_L^*),\ \text{and}\ v = v^* / a_L^*\) where \(a_L^* = (\gamma p_L^*/\rho_L^*)^{1/2}\).

3.1 Formulation in Roe variables

Roe [21] introduced the variables

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \rho^{1/2} \\ \rho^{1/2} v \\ \rho^{1/2} H \end{bmatrix}.$$

The flux and the conservative variables are given by

$$\hat{f}(w) = \begin{bmatrix} \gamma - 1 \gamma w_1 w_2 + \gamma w_3 w_2^2 \\ w_1 w_2 \\ w_2 w_3 \end{bmatrix}, \quad u = \hat{g}(w) = \begin{bmatrix} w_1^2 \\ w_1 w_2 \gamma + \gamma - 1 w_2^2 \\ \frac{w_1 w_3}{\gamma} + \frac{\gamma - 1}{2\gamma} w_2^2 \end{bmatrix}.$$

Then

$$\hat{g}(w)_t + \hat{f}_x(w) = 0$$

is equivalent to (5). The flux Jacobian in the Roe variables is given by

$$\nabla \hat{f} = \begin{bmatrix} \frac{w_2}{\gamma - 1} & \frac{w_1}{\gamma - 1} w_3 & 0 \\ \frac{\gamma + 1}{\gamma} w_1 & \frac{\gamma + 1}{\gamma} w_2 & 0 \\ 0 & w_3 & w_2 \end{bmatrix}.$$
3.2 Stochastic Galerkin formulation of the Euler equations

Define the pseudo-spectral product $u * v$ of order $P$ by

$$\langle u \ast v \rangle_k = \sum_{i=0}^{P-1} \sum_{j=0}^{P-1} u_i v_j \langle \Psi_i \Psi_j \Psi_k \rangle, \quad k = 0, \ldots, P - 1,$$

where

$$\langle \Psi_i \Psi_j \Psi_k \rangle = \int_{\Omega} \Psi_i(\xi) \Psi_j(\xi) \Psi_k(\xi) d\xi.$$

Alternatively, using matrix notation, we can write the spectral product as

$$[A(u)]_{jk} = \sum_{i=0}^{P-1} u_i \langle \Psi_i \Psi_j \Psi_k \rangle.$$

We will need the pseudo-spectral inverse $q^{-*}$, defined as the solution of $q * q^{-*} = 1$, and the pseudo-spectral square root, defined as the solution $q^{*/2}$ of $q^{*/2} * q^{*/2} = q$, where the spectral expansion of the quantity of interest $q$ is assumed to be known. For more details, see [22].

The Euler equations represented by the conservative formulation (5) can be written as an augmented system, after stochastic Galerkin projection,

$$u_t^P + f^P(u^P)_x = 0,$$

where

$$u^P = \begin{bmatrix} u_1^P \\ u_2^P \\ u_3^P \end{bmatrix} = \begin{bmatrix} (u_1)_0, \ldots, (u_1)_{P-1} \\ (u_2)_0, \ldots, (u_2)_{P-1} \\ (u_3)_0, \ldots, (u_3)_{P-1} \end{bmatrix}^T,$$

$$f^P(u^P) = \begin{bmatrix} (u_1^P)^{-*} u_2^P \ast u_2^P + p^P \\ (u_3^P + p^P) \ast u_2^P \ast (u_1^P)^{-*} \end{bmatrix}.$$

with $p^P = (\gamma - 1)(u_3^P - (u_1^P)^{-*} u_2^P \ast u_2^P / 2)$. The cubic products of (8) are approximated by the application of two third-order tensors, instead of one fourth-order tensor. That is, we replace $(a \ast b \ast c)_l = \sum_{ijk} \langle \Psi_i \Psi_j \Psi_k a_i b_j c_k \rangle$ by the approximation $(a \ast b \ast c)_l \approx (a \ast b \ast c)_l$.

For the Roe variable formulation, the stochastic Galerkin projection of (6) gives

$$\hat{g}^P(w^P)_t + \hat{f}^P(w^P)_x = 0,$$

where

$$\hat{g}^P(w^P) = \begin{bmatrix} w_1^P \ast w_1^P \\ w_1^P \ast w_2^P \\ \frac{w_1^P + w_2^P}{\gamma} + \gamma - 1 \frac{w_2^P}{2\gamma} w_2^P \ast w_2^P \end{bmatrix}, \quad \hat{f}^P(w^P) = \begin{bmatrix} \gamma - 1 w_1^P \ast w_2^P + \frac{\gamma + 1}{2\gamma} w_2^P \ast w_2^P \\ w_2^P \ast w_3^P \end{bmatrix}.$$

As $P \to \infty$, the formulations (8) and (9), as well as any other consistent formulation, are equivalent. However, $P$ is assumed to be small ($< 20$), and truncation and conditioning of the system matrices will play an important role for the accuracy of the solution.
4 Numerical method

4.1 Expansion of conservative variables

The semi-discretized form of (5) is given by

\[
\frac{dU_j}{dt} + \frac{F_{j+1/2} - F_{j-1/2}}{\Delta x} = 0 \quad (10)
\]

where \(F_{j+1/2}\) denotes the numerical flux function evaluated at the interface between cells \(j\) and \(j+1\).

For the MUSCL scheme, we take the numerical flux

\[
F_{j+1/2} = \frac{1}{2} \left( f^P(U^L_{j+1/2}) + f^P(U^R_{j+1/2}) \right) + \frac{1}{2} |\hat{J}_{PC}^{j+1/2}| \left( U^L_{j+1/2} - U^R_{j+1/2} \right), \quad (11)
\]

where the Roe average \(\hat{J}_{PC}^{j+1/2} = \nabla F(U_j, U_{j+1})\) is the pseudo-spectral generalization of the standard Roe average of the deterministic Euler equations, i.e.

\[
\hat{J}_{PC}(\tau, \Pi) = \begin{bmatrix}
0_{P \times P} & I_{P \times P} & 0_{P \times P} \\
\frac{1}{2}(\gamma - 3)A(\tau)^2 & (3 - \gamma)A(\tau) & (\gamma - 1)I_{P \times P} \\
\frac{1}{2}(\gamma - 1)A(\tau)^3 - A(\tau)A(\Pi) & A(\Pi) - (\gamma - 1)A(\tau)^2 & \gamma A(\tau)
\end{bmatrix}
\]

with

\[
\tau = (\rho^{-s/2}_L + \rho^{-s/2}_R) * (\rho^{s/2}_L * v_L + \rho^{s/2}_R * v_R),
\]

and

\[
\Pi = (\rho^{s/2}_L * H_L + \rho^{s/2}_R * H_R) * (\rho^{-s/2}_L + \rho^{-s/2}_R).
\]

The computation of \(\tau\) and \(\Pi\) require the spectral square root \(\rho^{s/2}\) and its inverse, that are computed solving a nonlinear and a linear system, respectively.

Further details about the formulation of the Roe average matrix are given in [12]. The scheme is a direct generalization of the deterministic MUSCL scheme. Flux limiters are applied componentwise to all generalized chaos coefficients in sharp regions. For a more detailed description of the MUSCL scheme and flux limiters, see e.g. [23], and for application to the stochastic Burgers’ equation [24].

4.2 Expansion of Roe’s variables

Let \(U\) and \(W\) denote the spatial discretizations of \(u\) and \(w\), respectively. The semi-discretized form of (9) is given by (10) with the numerical flux function

\[
F_{j+1/2} = \frac{1}{2} \left( f^P(W^L_{j+1/2}) + f^P(W^R_{j+1/2}) \right) + \frac{1}{2} |\hat{J}_{PC}^{j+1/2}| \left( W^L_{j+1/2} - W^R_{j+1/2} \right), \quad (12)
\]

with \(\hat{J}_{PC} = \hat{J}_{PC}(W)\) being the Roe matrix for the stochastic Galerkin formulation of the Euler equations in Roe’s variables, to be derived below.

Each time step provides the update of the solution vector \(U\), from which we can solve for \(W\) to be used in the update of the numerical flux. This involves solving the nonlinear system

\[
A(W_1)W_1 = U_1 \quad (13)
\]
for $W_1$, and then using $W_1$ to solve the linear $(P \times P)$-systems
\[ A(W_1)W_2 = U_2 \]
for $W_2$, and
\[ A(W_1)W_3 = \gamma U_3 - \frac{\gamma - 1}{2} A(W_2)W_2 \]
for $W_3$.

The system (13) is solved iteratively with a trust-region-dogleg algorithm\(^1\). Starting with the value of the previous time-step as initial guess, few iterations are required (typically 2-3). The same method is used to solve for spectral square roots for the conservative variable formulation.

We consider cases where $A(W_1)$ is sufficiently positive definite to avoid ill-conditioning or singularity. This corresponds loosely to avoid problem setups where the density is negative with non-zero probability. However, as will be discussed below, the occurrence of nearly-zero eigenvalues is also dependent on the choice of basis functionals, and not only on the parameter choices in the initial-boundary-value problem.

### 4.3 Stochastic Galerkin Roe average matrix for Roe variables

The Roe average matrix $\tilde{J}^{PC}$ is given as a function of the Roe variables $w = (w_1, w_2, w_3)^T$, where each $w_i$ is a vector of generalized chaos coefficients. It is designed to satisfy the following properties:

(i) $\tilde{J}^{PC}(w^L, w^R) \rightarrow \partial F/\partial w(w)$ as $w^L, w^R \rightarrow w$.

(ii) $\tilde{J}^{PC}(w^L, w^R) \times (w^L - w^R) = F^L - F^R, \forall w^L, w^R$

(iii) The eigenvectors of $\tilde{J}^{PC}$ are linearly independent.

In the standard approach introduced by Roe and commonly used for deterministic calculations, the conservative variables are mapped to the $w$ variables, which are then averaged.

In the deterministic approach, we have
\[ F^L - F^R = \tilde{J}(w^L, w^R) \times (w^L - w^R), \quad (14) \]
where
\[
\tilde{J}(w^L, w^R) = \begin{bmatrix}
\overline{w_2} & \overline{w_1} & 0 \\
\frac{\gamma - 1}{\gamma} \overline{w_3} & \frac{\gamma + 1}{\gamma} \overline{w_2} & \frac{\gamma - 1}{\gamma} \overline{w_1} \\
0 & \overline{w_3} & \overline{w_2}
\end{bmatrix}.
\]

Overbars denote arithmetic averages of assumed left and right values of a variable, i.e.
\[ \overline{w_j} = \frac{w^L_j + w^R_j}{2}, \quad j = 1, 2, 3. \]

\(^1\)This is the default algorithm for `fsolve` in Matlab. For more details, see [25].
It is a straightforward extension of the analysis by Roe in [21] to show properties (i) and (ii) for the Roe variables, without mapping to the conservative variables. There exists an eigenvalue decomposition

\[ \tilde{J} = VDV^{-1}, \]  

where

\[ V(w_1, w_2, w_3) = \begin{bmatrix} \frac{w_1 w_2 - \sqrt{w_2^2 + 8w_1 w_3 \gamma (\gamma - 1)}}{2\gamma w_3} & \frac{w_1 w_2 + \sqrt{w_2^2 + 8w_1 w_3 \gamma (\gamma - 1)}}{2\gamma w_3} & -\frac{w_1}{w_3} \\ \frac{w_2}{1} & \frac{w_2}{1} & 1 \end{bmatrix}, \]  

and with eigenvalues

\[ D(w_1, w_2, w_3) = \begin{bmatrix} \frac{w_2 (1 + 2\gamma) - \sqrt{8w_1 w_3 \gamma (\gamma - 1) + w_2^2}}{2\gamma} & 0 & 0 \\ 0 & \frac{w_2 (1 + 2\gamma) + \sqrt{8w_1 w_3 \gamma (\gamma - 1) + w_2^2}}{2\gamma} & 0 \\ 0 & 0 & w_2 \end{bmatrix}. \]

The eigenvalues of \( \tilde{J} \) are distinct, so property (iii) is also satisfied.

Now consider the stochastic Galerkin formulation, i.e. assume that the \( w_i \)'s are vectors of generalized chaos coefficients. The stochastic Galerkin Roe average matrix \( \tilde{J}^{PC} \) for the Roe variables formulation is a generalization of the mapping (14), i.e. of the matrix \( \tilde{J} \). We define

\[ \tilde{J}^{PC}(w^L, w^R) = \begin{bmatrix} A(w_2) & A(w_1) & 0_{P \times P} \\ \frac{\gamma - 1}{\gamma} A(w_3) & \frac{\gamma + 1}{\gamma} A(w_2) & \frac{\gamma - 1}{\gamma} A(w_1) \\ 0_{P \times P} & A(w_3) & A(w_2) \end{bmatrix}, \]  

where the submatrix \( A(w_j) \) is given by (7). The mapping defined by (17) satisfies the properties (i) and (ii). To prove (iii), we will need the following proposition.

**Proposition 1.** Let \( A(w_j) \) \((j = 1, 2, 3)\) be defined by (7) and \( A(w_j) = X\Lambda_j X^T \) be an eigenvalue decomposition with constant eigenvector matrix \( X \) and assume that \( \Lambda_1 \) and \( \Lambda_3 \) are non-singular. Then the stochastic Galerkin Roe average matrix \( \tilde{J}^{PC} \) has an eigenvalue decomposition \( \tilde{J}^{PC} = W\tilde{\Lambda}^{PC} W^{-1} \) with a complete set of eigenvectors.

**Proof.** We will use the Kronecker product \( \otimes \), defined for two matrices \( B \) (of size \( m \times n \)) and \( C \) by

\[ B \otimes C = \begin{bmatrix} b_{11} C & \ldots & b_{1n} C \\ \vdots & \ddots & \vdots \\ b_{m1} C & \ldots & b_{mn} C \end{bmatrix}. \]

The eigenvalue decompositions of each \( P \times P \) matrix block of (17) have the same eigenvector matrix \( X \), hence we can write

\[ \tilde{J}^{PC} = (I_3 \otimes X)\tilde{\Lambda}(I_3 \otimes X^T) \]  

(18)
where

\[
\hat{J} = \begin{bmatrix}
\frac{1}{\gamma+1} \Lambda_3 & \frac{1}{\gamma+1} \Lambda_2 & 0_{P \times P} \\
\frac{1}{\gamma+1} \Lambda_3 & \frac{1}{\gamma+1} \Lambda_2 & 0_{P \times P} \\
0_{P \times P} & 0_{P \times P} & \Lambda_2
\end{bmatrix}.
\]

By assumption, \(I_3 \otimes X\) is non-singular, and it remains to be shown that \(\hat{J}\) has distinct eigenvectors. Let

\[
S = \text{diag}(\Lambda_1 \Lambda_3^{-1}, \sqrt{(\gamma - 1)/\gamma \Lambda_1^{1/2} \Lambda_3^{-1/2}}, I_P).
\]

By assumption, \(\Lambda_1\) and \(\Lambda_3\) are invertible, so \(S\) and \(S^{-1}\) exist. We have

\[
J^S \equiv S^{-1} \hat{J} S = \begin{bmatrix}
\frac{1}{\gamma+1} \Lambda_3 & \frac{1}{\gamma+1} \Lambda_2 & 0_{P \times P} \\
\frac{1}{\gamma+1} \Lambda_3 & \frac{1}{\gamma+1} \Lambda_2 & 0_{P \times P} \\
0_{P \times P} & 0_{P \times P} & \Lambda_2
\end{bmatrix}.
\]

Clearly, \(J^S\) is symmetric and has the same eigenvalues as \(\hat{J}\) and \(\tilde{J}^{PC}\). Hence, \(J^S\) has an eigenvalue decomposition \(J^S = Y \tilde{\Lambda}^{PC} Y^T\). Then,

\[
\hat{J} = SY \tilde{\Lambda}^{PC} Y^T S^{-1} = SY \tilde{\Lambda}^{PC} (SY)^{-1}.
\]

Combining (18) and (19), we get

\[
\tilde{J}^{PC} = [(I_3 \otimes X)SY] \tilde{\Lambda}^{PC} [(I_3 \otimes X)SY]^{-1}.
\]

Setting \(W = (I_3 \otimes X)SY\), we get the eigenvalue decomposition \(\tilde{J}^{PC} = W \tilde{\Lambda}^{PC} W^{-1}\). By assumption, \(S\) and \(Y\) are non-singular, and we have that

\[
det(W) = det((I_3 \otimes X)SY) \neq 0,
\]

which proves that \(W\) is non-singular, and thus \(\tilde{J}^{PC}\) has a complete set of eigenvectors.

Under the conditions of proposition 4.1, Roe’s condition (iii) is satisfied. It is true for certain basis functionals assuming moderate uncertainty, but it can not be guaranteed for every case, and it certainly does not hold for pathological cases with e.g. negative density. The assumption of constant eigenvectors of the matrix \(A\) holds for Haar wavelets, for all \(P = 2^n\), with \(n \in \mathbb{N}\). See the appendix for a proof sketch. The requirement of non-singularity of \(\Lambda_1, \Lambda_3\) is not very restrictive since it amounts only to exclude unphysical behavior, for instance naturally positive quantities taking negative values with non-zero probability.

If the Roe average matrix of the flux (eq. (11) or (12)) becomes near-singular, the matrices in the eigenvalue decomposition

\[
|\tilde{J}^{PC}| = W |\Lambda| W^{-1},
\]

become ill-conditioned. This results in inaccurate or even unstable solutions. This is in particular a problem for the conservative variable formulation, with condition numbers over \(10^{16}\) for short times (e.g. \(10 - 100\) time steps). The corresponding figure for Roe variables is an upper bound on the condition number of about 200.
4.4 HLL Riemann solver

As a more robust but possibly less accurate alternative to the MUSCL-Roe solver, we use the HLL Riemann solver introduced by [26]. Instead of computing the Roe average matrix needed for the Roe fluxes (11) and (12), only the fastest signal velocities need to be estimated for the HLL solver. These signal velocities $S_L$ and $S_R$ are the estimated maximum and minimum eigenvalues of the Jacobian of the flux. In the conservative variable formulation the Jacobian is defined by the matrix $\partial F/\partial u$, and for the Roe variable formulation it is replaced by $\partial F/\partial w$. Then, the HLL flux is defined by

$$F_{HLL} = \begin{cases} f_L & \text{if } 0 \leq S_L \\ \frac{S_R f_L - S_L f_R + S_L S_R (U_R - U_L)}{S_R - S_L} & \text{if } S_L \leq 0 \leq S_R \\ f_R & \text{if } 0 \geq S_R \end{cases} .$$

In general, obtaining accurate eigenvalue estimates may be computationally costly. However, for the Haar wavelet expansion, we have explicit expressions for the system eigenvalues due to the constant eigenvector matrix of $A$. The HLL-flux is used both for the Roe variable formulation and the conservative variable formulation.

5 Numerical results

We introduce two test cases; case 1 with an initial function that can be exactly represented by two Legendre polynomials, and case 2 with slow initial decay in both Legendre polynomials and Haar wavelets. Qualitative results are presented to indicate what behavior we can expect for the convergence of the Legendre polynomials and Haar wavelet basis, respectively. After that, the relation between accuracy and the combination of stochastic representation and numerical flux is investigated. Robustness with respect to more extreme cases (density close to zero leading to high Mach number) is demonstrated for the Roe variable formulation. Finally, we perform a comparative study of the computational time for the formulation in conservative variables and the formulation in Roe variables.

5.1 Initial conditions

We consider (5) with two different initial functions on the domain $[0, 1]$. Since the analytical solution of Sod’s test case is known for any fixed value of the input parameters, the exact stochastic solution can be formulated as a function of the stochastic input $\xi$. Exact statistics can be computed by numerical integration over $\xi$. As case 1, assume that the density of the right state is subject to uncertainty, and all other quantities are deterministic at $t = 0$. The initial condition for (5) is given by

$$u(x, t = 0, \xi) = \begin{cases} u_L = (1, 0, 2.5)^T & x < 0.5 \\ u_R = (0.2 + \sigma \xi, 0, 0.25)^T & x > 0.5 \end{cases}$$

where we assume $\xi \in \mathcal{U}[-1, 1]$ and the scaling parameter $\sigma = 0.1$. This is a simple initial condition in the sense that the first two Legendre polynomials are sufficient
to represent the initial function exactly. As case no 2, we consider (5) subject to uncertainty in the initial shock location. Let

\[
\begin{align*}
  u(x, t = 0, \xi) = \begin{cases} 
    u_L = (1, 0, 2.5)^T & x < 0.5 + \sigma \xi \\
    u_R = (0.125, 0, 0.25)^T & x > 0.5 + \sigma \xi
  \end{cases}
\end{align*}
\]

where we assume \( \xi \in U[-1, 1] \) and the scaling parameter \( \sigma = 0.05 \). For case 2, exact representation of the initial function requires an infinite number of expansion terms in the Legendre polynomial basis. Figure 1 depicts the shock tube setup for the two cases, with dashed lines denoting uncertain parameters. We will also investigate a supersonic version of case 2, where the right state density is significantly reduced to obtain a strong shock.

### 5.2 Initial conditions and resolution requirements

For case 2, it should be noted that although the initial shock position can be exactly described by the first two terms of the Legendre polynomial chaos expansion, this is not the case for the initial state variables. In fact, the polynomial chaos expansions for the density, momentum and energy, decay only slowly with the number of expansion terms. Thus, unless a reasonably large number of expansion terms are retained, the stochastic Galerkin solution of case 2 will not be accurate even for small times.

The Legendre coefficients at small times display an oscillating behavior that becomes sharper with the order of the coefficients. The wavelet coefficients exhibit peaks that get sharper with the resolution level, and require a fine mesh. Figure 2 shows the initial Legendre coefficients and the initial Haar wavelets for case 2. The numerical method has a tendency to smear the chaos coefficients, resulting in under-predicting the variance. The increasing cost of using a larger number of basis functionals is further increased by the need for a finer mesh to resolve the solution modes.

### 5.3 Convergence of generalized chaos expansions

#### 5.3.1 Legendre polynomial expansion

Figure 3 shows the computed density (left) and exact density (right) at time \( t = 0.15 \) for case 1, plotted in the \( x - \xi \)-plane. The computed solution captures essential features of

---

**Figure 1:** Schematic representation of the initial setup for case 1 (left) and case 2 (right).
the exact solution, but the global Legendre polynomials cause oscillations downstream of the shock. A similar behavior is observed for case 2, as shown in figure 4.

5.3.2 Haar wavelet expansion

Figure 5 shows the density of case 1 at $t = 0.15$ computed with Haar basis functionals (left) and the true solution (right). There are no oscillations downstream, as in the Legendre polynomials case. The 8 ‘plateaus’ seen in the figure corresponds to the 8 basis functionals, and serves to illustrate the nature of the convergence of the wavelet chaos expansion. When the order of wavelet chaos expansion increases, the number of plateaus increases, and the solution converges to the exact solution. Figure 6 illustrates a similar behavior for case 2.

From figures 3 - 6 it is clear that the accuracy of the solution will strongly depend on the statistics of interest, i.e. the integrands and the stochastic limits of integration. Integration over the stochastic dimension in a region containing the oscillations seen in polynomial expansions is not likely to yield an accurate result, whereas quantities of interest depending only on integration over a smooth region is likely to be close to the true solution.

5.4 Accuracy

Some combinations of variables of chaos expansion and numerical flux functions are prone to fail in accurately capturing statistics of interest. Figure 7 shows computed variances with the MUSCL-Roe solver and the HLL solver, Roe variables and conservative variables, respectively. The conservative variable system is not accurately
Figure 3: Case 1. Density as a function of $x$ and $\xi$ at $t = 0.15$.

Figure 4: Case 2. Density as a function of $x$ and $\xi$ at $t = 0.15$. 

(a) $P=8$ Legendre polynomials  
(b) Exact solution
Figure 5: Density as a function of $x$ and $\xi$ at $t = 0.15$.

Figure 6: Density as a function of $x$ and $\xi$ at $t = 0.15$.  

16
solved using the MUSCL-Roe flux, whose eigenvalue decomposition becomes very ill-conditioned. The variance peaks are not well captured, as can be seen in figure 7. A more robust numerical flux function is needed for these variables.

5.5 Robustness

The stochastic Galerkin method applied to the Roe variables gives a more robust method than the conservative variables formulation. Figure 8 shows the Mach number for a modified supersonic version of case 2 with stronger shocks, $\rho_L = 1$, $\rho_R = 0.0039$ for 8 basis wavelets. For this problem, the conservative variable formulation leads to an unstable method. Thus, the Roe variable formulation is more suitable for problems where robustness is an issue.

Legendre polynomials are not suitable for this problem. As seen in figure 4, the solution is oscillatory in the right state close to the shock. If the right state density is small, as in this supersonic case, such oscillations cause the density to be very close to zero, or even negative. This leads to an unphysical solution and breakdown of the numerical method.

5.6 Computational cost

For stochastic Galerkin systems built on the same order of polynomial chaos, the size of the resulting numerical problem is the same for Legendre polynomials and Haar wavelets. Hence, the computational cost is similar for Haar wavelets and Legendre polynomials of the same order of generalized chaos. Although the different bases could result in properties that make them very different in e.g. the number of iterations required to solve the nonlinear matrix problems, numerical experiments yield very similar computational costs for the cases tested.

In order to compare the computational cost of the Roe variable expansion with that of the conservative expansion, a similar experimental setup is used for both methods. Sufficiently small test cases are run in order not to exceed the cache limit which would impair the simulation time for fine meshes and bias the result. The relatively simple test case 1 for short simulation times is used to avoid excessive computational time in the solution steps involving iterations for solving nonlinear systems.

In the experiments, the same time step has been used for the different variable expansions, since the stability limit is very similar. Table 1 displays the relative CPU time of the two different variable expansions for increasing number of Haar wavelets. The higher computational cost for the conservative variable formulation is in part due to the need to compute inverse quantities and cubic spectral products. The Roe variable formulation only requires the solution of the nonlinear system for the square root of the density and quadratic flux function evaluations. The cost of the computation of the inverse density, could potentially be reduced with a more efficient optimization method. However, the nonlinear system for the square root of the density in the Roe variable formulation would most likely benefit equally much from a more efficient optimization algorithm. Thus, table 1 gives an indication about the relative performance of the problem formulations although the simulation times could all be reduced by optimizing.
Figure 7: Variances for Roe variables and conservative variables. Case 1, solution obtained with MUSCL-Roe solver and HLL solver at $t = 0.05$, $P = 16$ Haar wavelets.
the codes. The relative benefit of the Roe variable expansion decreases with the order of wavelet expansion which is due to the increasing cost of forming spectral products that dominates the total cost for high-order expansions. The MUSCL-Roe solver requires the calculation of spectral inverse quantities not needed for the HLL solver. This explains the higher relative cost of conservative variables compared to Roe variables using the MUSCL-Roe solver.

<table>
<thead>
<tr>
<th>Solver</th>
<th>$P = 2$</th>
<th>$P = 4$</th>
<th>$P = 8$</th>
<th>$P = 16$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLL</td>
<td>10.59</td>
<td>9.83</td>
<td>8.24</td>
<td>7.75</td>
</tr>
<tr>
<td>MUSCL</td>
<td>29.28</td>
<td>28.57</td>
<td>20.53</td>
<td>10.72</td>
</tr>
</tbody>
</table>

Table 1: Relative simulation time $\frac{T_{\text{cons}}}{T_{\text{Roe}}}$ using conservative variables relative to using Roe variables.

6 Conclusions

An intrusive formulation of the stochastic Euler equations based on Roe variables is presented. A Roe average matrix for the standard MUSCL-Roe scheme with Roe variables is derived, and we prove that it satisfies the conditions stated by Roe under certain conditions. A more robust HLL flux for the Roe variables is also introduced.

The Roe variable formulation is robust for supersonic problems where the conservative variable formulation fails, but only for localized basis functionals of the generalized chaos representation. For global Legendre polynomials, the discontinuities in stochastic space lead to oscillations and unphysical behavior of the solution and numerical instability. Wavelet functionals are more robust in this respect, and do not yield oscillations
around discontinuities in stochastic space.

The Roe variable formulation leads to speedup compared to the conservative variable formulation. This is most prominent for the MUSCL-Roe solver. The relative speedup decreases with the order of generalized chaos since the total computational cost for high-order expansions is no longer dominated by spectral inversion and square root calculations. Instead, the main cost lies in the formation of spectral product matrices.

We demonstrate the need for robust flux functions by presenting cases where the standard MUSCL-Roe flux fails to capture the solution, but for which the more robust HLL flux captures the essential features of the stochastic solution. The design of a robust numerical method is also highly dependent on the choice of the stochastic basis. The Haar wavelets are not only more robust than Legendre polynomials for representation of discontinuities in stochastic space, but also admit the proof of existence of a Roe matrix and more specifically the hyperbolicity of the stochastic Galerkin formulation. This implies that the truncated problem mimics the original problem - a desirable feature.

If the representation of the initial function has not converged, the solution at future times can not be accurate. Case 2 illustrates the need to find a representation of uncertainty with fast decay of the coefficients of the generalized chaos expansion. An alternative to more accurate representation of the input uncertainty is to combine the intrusive Roe variable formulation presented here with multi-element methods, for instance in the manner presented by [12].

Acknowledgements

This work is supported by King Abdullah University of Science and Technology (KAUST).

Appendix

**Proposition 1.** The matrix $A$ defined by (7) for Haar wavelets $\{\Psi_j\}_{j=0}^{P-1}$ has constant eigenvectors for all $P = 2^n$, $n \in \mathbb{N}$.

*Sketch of proof.* We will use induction on the order $P$ of wavelet chaos to show that the matrix $A$ has constant eigenvectors for all orders $P$. In order to do this, we will need certain features of the structure of $A$. We can express $A^{2P}$ in terms of the matrix $A^P$. Two properties of the triple product $\langle \Psi_i \Psi_j \Psi_k \rangle$ will be used to prove that $A$ indeed has the matrix structure presented.

**Property 1:** Let $i \in \{0, \ldots, P - 1\}$, $j \in \{P, \ldots, 2P - 1\}$ and let $j'$ and $j''$ be the progenies of $j$. Then

$$\langle \Psi_i \Psi_j^2 \rangle = \langle \Psi_i \Psi_{j'}^2 \rangle = \langle \Psi_i \Psi_{j''}^2 \rangle.$$

**Property 2:** Consider the indices $i \in \{P, \ldots, 2P - 1\}$, $j \in \{2P, \ldots, 4P - 1\}$. Then

$$\langle \Psi_i \Psi_j^2 \rangle = \begin{cases} P^{1/2} & \text{if } j \text{ first progeny of } i \\ -P^{1/2} & \text{if } j \text{ second progeny of } i \\ 0 & \text{otherwise} \end{cases}.$$
As induction hypothesis, we assume that given $A_P$ for some $P = 2^n$, $n \in \mathbb{N}$, the
next order of triple product matrix $A_{2P}$ can be written

$$A_{2P} = \begin{bmatrix}
A_P & X_PM_P \\
M_PX_P^T & \Lambda
\end{bmatrix}$$

where $X_P$ is the matrix of constant eigenvectors of $A_P$ satisfying $\|X_P\|_2^2 = P$, $M_P = \text{diag}(w_P, ..., w_{2P-1})$ and $\Lambda$ is diagonal and contains the eigenvalues of $A_P$. Then, we have that

$$\begin{bmatrix}
A_P & X_PM_P \\
M_PX_P^T & \Lambda
\end{bmatrix} \begin{bmatrix}
X_P \\
\pm P^{1/2} I
\end{bmatrix} = \begin{bmatrix}
X_P \Lambda \pm P^{1/2} XM_P \\
PM_P \pm P^{1/2} \Lambda
\end{bmatrix} = \begin{bmatrix}
X_P \\
\pm P^{1/2} I
\end{bmatrix} (\Lambda \pm P^{1/2} M),$$

so the eigenvalues and eigenvectors of $A_{2P}$ are given by $\Lambda \pm P^{1/2} M$ and $[X_P, \pm P^{1/2} I]^T$, respectively. For the next order of expansion, $4P$, we have

$$A_{4P} = \begin{bmatrix}
A_P & X_PM_P \\
M_PX_P^T & \Lambda
\end{bmatrix} \begin{bmatrix}
X_P \otimes [1,1] \\
P^{1/2} I \otimes [1,1]
\end{bmatrix}M_{2P} = \begin{bmatrix}
X_P \otimes [1,1] \\
P^{1/2} I \otimes [1,-1]
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix} (\Lambda \otimes I_2 + P^{1/2} M_P \otimes M_{2P}).$$

To see that this is indeed the structure of $A_{4P}$, note that any non-zero matrix entry not already present in $A_{2P}$, can be deduced using properties 1 and 2, and scaling the rows/columns by multiplication by the diagonal matrix $M_{2P}$. The structure of $A_{4P}$ follows from the construction of the Haar wavelet basis, but we do not give a proof here.

One can verify that $A_{4P}$ given by (20) has the eigenvectors and eigenvalues

$$X_{4P} = \begin{bmatrix}
X_P \otimes [1,1] \\
P^{1/2} I \otimes [1,-1]
\end{bmatrix} \pm (2P)^{1/2} I_{2P},$$

$$\Lambda_{4P} = \Lambda \otimes I_2 + P^{1/2} M_P \otimes \begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix} \pm (2P)^{1/2} M_{2P},$$

so the eigenvectors are constant (but the eigenvalues are variable in the coefficients $(w_i)_j$ through $M_P$ and $M_{2P}$). The base cases $P = 1$, $P = 2$, can easily be verified, so by induction $A_P$ has constant eigenvectors for all $P = 2^n$, $n \in \mathbb{N}$.

References


