On the Eigenvalues of the ADER-WENO Galerkin Predictor

Haran Jackson

Cavendish Laboratory, JJ Thomson Ave, Cambridge, UK, CB3 0HE

Abstract

ADER-WENO methods represent an effective set of techniques for solving hyperbolic systems of PDEs. These systems may be non-conservative and non-homogeneous, and contain stiff source terms. The methods require a spatio-temporal reconstruction of the data in each spacetime cell, at each time step. This reconstruction is obtained as the root of a nonlinear system, resulting from the use of a Galerkin method. It is proved here that the eigenvalues of certain matrices appearing in these nonlinear systems are always 0, regardless of the number of spatial dimensions of the PDEs, or the chosen order of accuracy of the ADER-WENO method. This guarantees fast convergence to the Galerkin root for certain classes of PDEs.

Keywords: ADER-WENO, Galerkin, Eigenvalues, Convergence

1. Background

ADER-WENO methods have proved extremely useful in obtaining arbitrarily high-order solutions to problems involving hyperbolic systems of PDEs. For example, it has been demonstrated that for the same computational cost as a Runge-Kutta scheme of a certain order, one can obtain an ADER scheme of one higher order of accuracy (see Balsara et al. [1]). Additionally, Runge-Kutta schemes suffer from the presence of Butcher barriers (see Butcher [3]), limiting the order of temporal accuracy that one can comfortably achieve. There are no such limitations present in ADER-WENO schemes.

The cumbersome analytical derivation of the temporal derivatives of the solution required by the original ADER formulation (see Toro [10]) has been replaced by the use of a cell-wise local Galerkin predictor. The predictor can take either a discontinuous or a continuous form (see Dumbser et al. [4] and Balsara et al. [2], respectively). The Galerkin predictor is a high-order polynomial reconstruction of the data in both space and time, found as the root of a non-linear system.

It has been conjectured that the eigenvalues of certain matrices appearing in these non-linear systems are always zero, leading to desirable system properties for certain classes of PDEs. It is proved here that this is in deed the case for any number of spatial dimensions and any desired order of accuracy, for both the discontinuous and continuous Galerkin variants. This result is independent of the choice of reconstruction basis polynomials.

The Einstein summation convention is to be assumed throughout this paper.

2. The ADER-WENO Method

Take a non-homogeneous, non-conservative (and for simplicity, one-dimensional) hyperbolic system of the form:

$$\frac{\partial \mathbf{Q}}{\partial t} + \frac{\partial F(\mathbf{Q})}{\partial x} + B(\mathbf{Q}) \cdot \frac{\partial \mathbf{Q}}{\partial x} = S(\mathbf{Q})$$
 (1)

where Q is the vector of conserved variables, F is the conservative nonlinear flux, B is the matrix corresponding to the purely non-conservative component of the system, and S(Q) is the algebraic source vector.

Take the set of grid points $x_0 < x_1 < \ldots < x_K$ and define $\Delta x_i = x_{i+1} - x_i$. Take the time steps $t_0 < t_1 < \ldots$ while defining $\Delta t_n = t_{n+1} - t_n$. Following the formulations presented in Dumbser et al. [5, 6], Balsara et al. [2], the WENO method and Galerkin method produce at each time step t_n a local polynomial approximation to Q on each space-time cell $[x_i, x_{i+1}] \times [t_n, t_{n+1}]$.

Define the scaled space variable:

$$\chi^{i} = \frac{1}{\Delta x_{i}} \left(x - x_{i} \right) \tag{2}$$

Take a basis $\{\psi_0, ..., \psi_N\}$ of P_N and inner product $\langle \cdot, \cdot \rangle$. This basis can either be nodal $(\psi_i(\chi_j) = \delta_{ij})$ where $\{\chi_0, ..., \chi_N\}$ are a set of nodal points, such as the Gauss-Legendre abscissae), or modal (such as the Jacobi polynomials).

The WENO method (as used in Dumbser et al. [8]) produces an order-N polynomial reconstruction of the data at time t_n in cell $[x_i, x_{i+1}]$, using $\{\psi_0, \dots, \psi_N\}$ as a basis. This is denoted:

$$\mathbf{w}\left(x\right) = \mathbf{w}_{\gamma}\psi_{\gamma}\left(\chi^{i}\left(x\right)\right) \tag{3}$$

Email address: hj305@cam.ac.uk (Haran Jackson)

This spatial reconstruction at the start of the time step is to be used as initial data in the problem of finding the Galerkin predictor.

Now define the scaled time variable:

$$\tau^n = \frac{1}{\Delta t_n} \left(t - t_n \right) \tag{4}$$

Thus, (1) becomes:

$$\frac{\partial \mathbf{Q}}{\partial \tau^{n}} + \frac{\partial \mathbf{F}^{*}(\mathbf{Q})}{\partial \chi^{i}} + B^{*}(\mathbf{Q}) \cdot \frac{\partial \mathbf{Q}}{\partial \chi^{i}} = \mathbf{S}^{*}(\mathbf{Q})$$
 (5)

where

$$F^* = \frac{\Delta t_n}{\Delta x} F \qquad B^* = \frac{\Delta t_n}{\Delta x} B \qquad S^* = \Delta t_n S \tag{6}$$

The non-dimensionalization notation and spacetime cell indexing notation will be dropped for simplicity in what follows. Now define the set of spatio-temporal basis functions:

$$\{\theta_k(\chi,\tau)\} = \left\{ \psi_p(\chi) \, \psi_s(\tau) : 0 \le p, s \le N \right\} \tag{7}$$

Denoting the Galerkin predictor by q, take the following set of approximations:

$$\mathbf{Q} \approx \mathbf{q} = \theta_{\beta} \mathbf{q}_{\beta} \tag{8a}$$

$$F(Q) \approx \theta_{\beta} F_{\beta}$$
 (8b)

$$B(\mathbf{Q}) \cdot \frac{\partial \mathbf{Q}}{\partial \chi} \approx \theta_{\beta} \mathbf{B}_{\beta} \tag{8c}$$

$$S(Q) \approx \theta_{\beta} S_{\beta}$$
 (8d)

for some coefficients q_{β} , F_{β} , B_{β} , S_{β} .

If $\{\psi_0, ..., \psi_N\}$ is a nodal basis, the *nodal basis representation* may be used:

$$F_{\beta} = F(q_{\beta}) \tag{9a}$$

$$\boldsymbol{B}_{\beta} = B\left(\boldsymbol{q}_{\beta}\right) \cdot \left(\frac{\partial \theta_{\gamma}\left(\chi_{\beta}, \tau_{\beta}\right)}{\partial \chi} \boldsymbol{q}_{\gamma}\right) \tag{9b}$$

$$S_{\beta} = S(q_{\beta}) \tag{9c}$$

where $(\chi_{\beta}, \tau_{\beta})$ are the coordinates of the node corresponding to basis function θ_{β} .

If a modal basis is used, F_{β} , B_{β} , S_{β} may be found from the previous values of q_{β} in the iterative processes described below.

For functions $f(\chi, \tau) = f_{\chi}(\chi) f_{\tau}(\tau)$ and $g(\chi, \tau) = g_{\chi}(\chi) g_{\tau}(\tau)$, define the following integral operators:

$$[f,g]^t = f_\tau(t) g_\tau(t) \langle f_\chi, g_\chi \rangle$$
 (10a)

$$\{f,g\} = \langle f_{\tau}, g_{\tau} \rangle \langle f_{\chi}, g_{\chi} \rangle$$
 (10b)

Multiplying (5) by test function θ_{α} , using the polynomial approximations for Q, F, B, S, and integrating over space and time gives:

$$\left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \tau}\right\} \boldsymbol{q}_{\beta} = -\left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi}\right\} \boldsymbol{F}_{\beta} + \left\{\theta_{\alpha}, \theta_{\beta}\right\} \left(\boldsymbol{S}_{\beta} - \boldsymbol{B}_{\beta}\right)$$
(11)

2.1. The Discontinuous Galerkin Method

This method of computing the Galerkin predictor allows solutions to be discontinuous at temporal cell boundaries, and is also suitable for stiff source terms.

Integrating (11) by parts in time gives:

$$\left(\left[\theta_{\alpha}, \theta_{\beta} \right]^{1} - \left\{ \frac{\partial \theta_{\alpha}}{\partial \tau}, \theta_{\beta} \right\} \right) \boldsymbol{q}_{\beta} = \left[\theta_{\alpha}, \boldsymbol{w} \right]^{0} - \left\{ \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi} \right\} \boldsymbol{F}_{\beta} \qquad (12)$$

$$+ \left\{ \theta_{\alpha}, \theta_{\beta} \right\} \left(\boldsymbol{S}_{\beta} - \boldsymbol{B}_{\beta} \right)$$

where w is the reconstruction obtained at the start of the time step with the WENO method. Define the following:

$$U_{\alpha\beta} = \left[\theta_{\alpha}, \theta_{\beta}\right]^{1} - \left\{\frac{\partial \theta_{\alpha}}{\partial \tau}, \theta_{\beta}\right\}$$
 (13a)

$$V_{\alpha\beta} = \left\{ \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \nu} \right\} \tag{13b}$$

$$\mathbf{W}_{\alpha} = \left[\theta_{\alpha}, \psi_{\gamma}\right]^{0} \mathbf{w}_{\gamma} \tag{13c}$$

$$Z_{\alpha\beta} = \left\{ \theta_{\alpha}, \theta_{\beta} \right\} \tag{13d}$$

Thus:

$$U_{\alpha\beta}q_{\beta} = W_{\alpha} - V_{\alpha\beta}F_{\beta} + Z_{\alpha\beta}(S_{\beta} - B_{\beta})$$
 (14)

This nonlinear system in q_{β} is solved by a Newton method. The source terms must be solved implicitly if they are stiff. Note that W has no dependence on q.

2.2. The Continuous Galerkin Method

This method of computing the Galerkin predictor is not suitable for stiff source terms, but it provides substantial savings on computational cost and ensures continuity across temporal cell boundaries.

 $\{\psi_0, ..., \psi_N\}$ must be chosen in such a way that the first N+1 elements of $\{\theta_\beta\}$ have only a spatial dependence. The first N+1 elements of \boldsymbol{q} are then fixed by demanding continuity at $\tau=0$:

$$q(\chi, 0) = w(\chi) \tag{15}$$

where w is spatial the reconstruction obtained at the start of the time step with the WENO method.

For a given vector $\mathbf{v} \in \mathbb{R}^{(N+1)^2}$ and matrix $X \in M_{(N+1)^2,(N+1)^2}(\mathbb{R})$, let $\mathbf{v} = \begin{pmatrix} \mathbf{v^0}, \mathbf{v^1} \end{pmatrix}$ and $X = \begin{pmatrix} X^{00} & X^{01} \\ X^{10} & X^{11} \end{pmatrix}$ where $\mathbf{v^0}, X^{00}$ are the components relating solely to the first N+1 components of \mathbf{v} . We only need to find the latter components of \mathbf{q} , and thus, from (11), we have:

$$\left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \tau}\right\}^{11} \boldsymbol{q}_{\beta}^{1} = \left\{\theta_{\alpha}, \theta_{\beta}\right\}^{11} \left(\boldsymbol{S}_{\beta}^{1} - \boldsymbol{B}_{\beta}^{1}\right) - \left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi}\right\}^{11} \boldsymbol{F}_{\beta}^{1} \quad (16)$$

$$+ \left\{\theta_{\alpha}, \theta_{\beta}\right\}^{10} \left(\boldsymbol{S}_{\beta}^{0} - \boldsymbol{B}_{\beta}^{0}\right) - \left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi}\right\}^{10} \boldsymbol{F}_{\beta}^{0}$$

Define the following:

$$U_{\alpha\beta} = \left\{ \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \tau} \right\}^{11} \tag{17a}$$

$$V_{\alpha\beta} = \left\{ \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi} \right\}^{11} \tag{17b}$$

$$\boldsymbol{W}_{\alpha} = \left\{ \theta_{\alpha}, \theta_{\beta} \right\}^{10} \left(\boldsymbol{S}_{\beta} - \boldsymbol{B}_{\beta} \right)^{0} - \left\{ \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi} \right\}^{10} \boldsymbol{F}_{\beta}^{0}$$
 (17c)

$$Z_{\alpha\beta} = \left\{ \theta_{\alpha}, \theta_{\beta} \right\}^{11} \tag{17d}$$

Thus:

$$U_{\alpha\beta}q_{\beta}^{1} = W_{\alpha} - V_{\alpha\beta}F_{\beta}^{1} + Z_{\alpha\beta}\left(S_{\beta}^{1} - B_{\beta}^{1}\right)$$
 (18)

Note that, as with the discontinuous Galerkin method, W has no dependence on the degrees of freedom in q.

3. Conjecture

Extending the Galerkin method described in the previous section to three dimensions, the following system must be solved for q:

$$U_{\alpha\beta}\mathbf{q}_{\beta} = \mathbf{W}_{\alpha} - V_{\alpha\beta}^{1}\mathbf{F}_{\beta} - V_{\alpha\beta}^{2}\mathbf{G}_{\beta} - V_{\alpha\beta}^{3}\mathbf{H}_{\beta}$$
(19)
+ $Z_{\alpha\beta}\left(\mathbf{S}_{\beta} - \mathbf{B}_{\beta}\right)$

where now we have the 3 scaled spatial variables χ_1, χ_2, χ_3 and G, H are the flux components in the second and third spatial directions, respectively. In the case of the continuous Galerkin

method, it is assumed that (19) is to be solved for only the non-fixed degrees of freedom in q. The matrices $V_{\alpha\beta}^i$ are defined thus:

$$V_{\alpha\beta}^{i} = \left\langle \theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi_{i}} \right\rangle \tag{20}$$

For the discontinuous Galerkin method, W_{α} now takes the form:

$$\mathbf{W}_{\alpha} = \left[\theta_{\alpha}, \Psi_{\gamma}\right]^{0} \mathbf{w}_{\gamma} \tag{21}$$

where $\Psi_{\gamma}(\chi_1,\chi_2,\chi_3)$ is an element of the following set, enumerated by γ :

$$\{\psi_i(\chi_1)\psi_j(\chi_2)\psi_c(\chi_3): 0 \le i, j, k \le N\}$$
 (22)

For the continuous Galerkin method, W_{α} now takes the form:

$$\boldsymbol{W}_{\alpha} = \left\{\theta_{\alpha}, \theta_{\beta}\right\}^{10} \left(\boldsymbol{S}_{\beta} - \boldsymbol{B}_{\beta}\right)^{0}$$

$$- \left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi^{1}}\right\}^{10} \boldsymbol{F}_{\beta}^{0} - \left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi^{2}}\right\}^{10} \boldsymbol{G}_{\beta}^{0} - \left\{\theta_{\alpha}, \frac{\partial \theta_{\beta}}{\partial \chi^{2}}\right\}^{10} \boldsymbol{H}_{\beta}^{0}$$

$$(23)$$

Dumbser et al. [4] remark that for the continuous Galerkin case, the eigenvalues of $U^{-1}V^i$ are all 0 for $0 \le N \le 5$, for i = 1, 2, 3. Dumbser and Zanotti [7] state the same result for the discontinuous Galerkin case. This implies that in the conservative, homogeneous case (B = S = 0), owing to the Banach Fixed Point Theorem, existence and uniqueness of a solution are established, and convergence to this solution is guaranteed. As noted in Dumbser and Zanotti [7], in the linear case it is implied that the iterative procedure converges after at most N + 1 iterations.

In Dumbser et al. [4] it is conjectured that the result concerning the eigenvalues of $U^{-1}V^i$ holds for any N, and any number of spatial dimensions. A solution to this conjecture is provided here. For the theory in linear algebra required for this section, please consult a standard textbook on the subject, such as Nering [9].

3.1. The Discontinuous Galerkin Case

First, given the basis polynomials $\{\psi_0, \dots, \psi_N\}$, define the following matrices:

$$\aleph_{ij} = \left\langle \psi_i, \psi_j \right\rangle \tag{24a}$$

$$\beth_{ij} = \left\langle \psi_i, \psi_j' \right\rangle \tag{24b}$$

Note that \aleph is the Gram matrix, which by linear independence of $\{\psi_0,...,\psi_N\}$ is invertible. Note also that if $p \in P_N$ then $p = a_j\psi_j$ for some unique coefficient vector \boldsymbol{a} . Thus, taking inner

products with ψ_i , we have $\langle \psi_i, \psi_j \rangle a_j = \langle \psi_i, p \rangle$ for i = 0, ..., N. This produces the following result:

$$p = a_j \psi_j \Leftrightarrow \boldsymbol{a} = \boldsymbol{\aleph}^{-1} \boldsymbol{x}, \ x_i = \langle \psi_i, p \rangle$$
 (25)

Without loss of generality, take the ordering:

$$\alpha = \alpha_t (N+1)^3 + \alpha_x (N+1)^2 + \alpha_y (N+1) + \alpha_z$$
 (26)

where $0 \le \alpha_t, \alpha_x, \alpha_y, \alpha_z \le N$. Using the same ordering for β , we have:

$$U_{\alpha\beta} = \left(\psi_{\alpha_t}(1)\psi_{\beta_t}(1) - \beth_{\beta_t\alpha_t}\right) \cdot \aleph_{\alpha_y\beta_y} \cdot \aleph_{\alpha_y\beta_y} \cdot \aleph_{\alpha_z\beta_z} \tag{27a}$$

$$V_{\alpha\beta}^{1} = \aleph_{\alpha_{t}\beta_{t}} \cdot \beth_{\alpha_{x}\beta_{x}} \cdot \aleph_{\alpha_{y}\beta_{y}} \cdot \aleph_{\alpha_{z}\beta_{z}}$$
 (27b)

Thus:

$$U = C \otimes \aleph \otimes \aleph \otimes \aleph \tag{28a}$$

$$V^1 = \aleph \otimes \beth \otimes \aleph \otimes \aleph \tag{28b}$$

where $C_{ij} = \psi_i(1) \psi_j(1) - \beth_{ji}$. Thus:

$$U^{-1}V^{1} = (C^{-1}\aleph) \otimes (\aleph^{-1}\beth) \otimes I \otimes I \tag{29}$$

Therefore:

$$\left(U^{-1}V^{1}\right)^{k} = \left(C^{-1}\aleph\right)^{k} \otimes \left(\aleph^{-1}\beth\right)^{k} \otimes I \otimes I \tag{30}$$

A matrix X is nilpotent ($X^k = 0$ for some $k \in \mathbb{N}$) if and only if all its eigenvalues are 0. The conjecture will be proved if it is shown that $\left(\mathbf{8}^{-1}\mathbf{\beth}\right)^k = 0$ for some $k \in \mathbb{N}$, as this would imply that $\left(U^{-1}V^1\right)^k = 0$, and thus all eigenvalues of $U^{-1}V^1$ are 0.

Taking $\boldsymbol{a} \in \mathbb{R}^{N+1}$, define:

$$p = a_0 \psi_0 + \ldots + a_N \psi_N \in P_N \tag{31}$$

Note that:

$$(\exists \boldsymbol{a})_{i} = \langle \psi_{i}, \psi_{0}' \rangle a_{0} + \ldots + \langle \psi_{i}, \psi_{N}' \rangle a_{N} = \langle \psi_{i}, p' \rangle$$
 (32)

Thus, by (25):

$$\left(\aleph^{-1} \Box a\right) \psi_i = p' \tag{33}$$

By induction:

$$\left(\left(\aleph^{-1}\beth\right)^k a\right)_i \psi_i = p^{(k)} \tag{34}$$

for any $k \in \mathbb{N}$. As $p \in P_N$, $(\aleph^{-1} \beth)^{N+1} a = 0$. As a was chosen arbitrarily, $(\aleph^{-1} \beth)^{N+1} = 0$. Thus, the conjecture is solved.

This proof is easily adapted to show that $U^{-1}V^2$ and $U^{-1}V^3$ are nilpotent, and clearly extends to any number of spatial dimensions. No specific choice has been made for $N \in \mathbb{N}$ and thus the result holds in general.

3.2. The Continuous Galerkin Case

In addition to \aleph , \beth , we now define \aleph' , \beth' where each new matrix is equal to the original, with its first row and column removed (the row and column corresponding to the constant-term polynomial ψ_0). Take the following ordering:

$$\alpha = \alpha_t (N+1)^3 + \alpha_x (N+1)^2 + \alpha_y (N+1) + \alpha_z$$
 (35)

where now $0 \le \alpha_x, \alpha_y, \alpha_z \le N$ and $0 \le \alpha_t \le N - 1$. Using the same ordering for β , we now have:

$$U_{\alpha\beta} = \beth_{\alpha,\beta_t}' \cdot \aleph_{\alpha_x\beta_x} \cdot \aleph_{\alpha_y\beta_y} \cdot \aleph_{\alpha_z\beta_z}$$
 (36a)

$$V_{\alpha\beta}^{1} = \aleph_{\alpha,\beta_{t}}' \cdot \beth_{\alpha_{x}\beta_{x}} \cdot \aleph_{\alpha_{y}\beta_{y}} \cdot \aleph_{\alpha_{z}\beta_{z}}$$
 (36b)

The proof for the continuous case follows in the same manner as the proof for the discontinuous case, with:

$$U^{-1}V^{1} = \left(\left(\Box' \right)^{-1} \aleph' \right) \otimes \left(\aleph^{-1} \Box \right) \otimes I \otimes I \tag{37}$$

4. References

- Balsara, D. S., Meyer, C., Dumbser, M., Dub, H., Xu, Z., 2013. Efficient implementation of ader schemes for euler and magnetohydrodynamical flows on structured meshes - speed comparisons with runge-kutta methods. Journal of Computational Physics 235, 934–969.
- [2] Balsara, D. S., Rumpf, T., Dumbser, M., Munz, C.-D., 2009. Efficient, high accuracy ader-weno schemes for hydrodynamics and divergencefree magnetohydrodynamics. Journal of Computational Physics 228, 2480–2516.
- [3] Butcher, J. C., 2009. On fifth and sixth order explicit runge-kutta methods - order conditions and order barriers. Canadian Applied Mathematics Ouarterly 17.
- [4] Dumbser, M., Balsara, D. S., Toro, E. F., Munz, C. D., 2008. A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes. Journal of Computational Physics 227 (18), 8209–8253.
- [5] Dumbser, M., Castro, M., Pares, C., Toro, E. F., 2009. Ader schemes on unstructured meshes for nonconservative hyperbolic systems: Applications to geophysical flows. Computers & Fluids 38 (9), 1731–1748.
- [6] Dumbser, M., Hidalgo, A., Zanotti, O., 2014. High order space-time adaptive ader-weno finite volume schemes for non-conservative hyperbolic systems. Computer Methods in Applied Mechanics & Engineering 268, 359–387.
- [7] Dumbser, M., Zanotti, O., 2009. Very high order pnpm schemes on unstructured meshes for the resistive relativistic mhd equations. Journal of Computational Physics 228 (18), 6991–7006.
- [8] Dumbser, M., Zanotti, O., Hidalgo, A., Balsara, D. S., 2013. Ader-weno finite volume schemes with space-time adaptive mesh refinement. Journal of Computational Physics 248, 257–286.
- [9] Nering, E. D., 1970. Linear Algebra and Matrix Theory. John Wiley & Sons.
- [10] Toro, E. F., 2009. Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer.

5. Acknowledgments

I acknowledge financial support from the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science under grant EP/L015552/1.