**ORIGINAL PAPER**

Check for
updates

# Efficient Iterative Arbitrary High-Order Methods: an Adaptive Bridge Between Low and High Order

Lorenzo Micalizzi[1] · Davide Torlo[2] · Walter Boscheri[3]

## Abstract

We propose a new paradigm for designing efficient $p$-adaptive arbitrary high-order methods. We consider arbitrary high-order iterative schemes that gain one order of accuracy at each iteration and we modify them to match the accuracy achieved in a specific iteration with the discretization accuracy of the same iteration. Apart from the computational advantage, the newly modified methods allow to naturally perform the $p$-adaptivity, stopping the iterations when appropriate conditions are met. Moreover, the modification is very easy to be included in an existing implementation of an arbitrary high-order iterative scheme and it does not ruin the possibility of parallelization, if this was achievable by the original method. An application to the Arbitrary DERivative (ADER) method for hyperbolic Partial Differential Equations (PDEs) is presented here. We explain how such a framework can be interpreted as an arbitrary high-order iterative scheme, by recasting it as a Deferred Correction (DeC) method, and how to easily modify it to obtain a more efficient formulation, in which a local *a posteriori* limiter can be naturally integrated leading to the $p$-adaptivity and structure-preserving properties. Finally, the novel approach is extensively tested against classical benchmarks for compressible gas dynamics to show the robustness and the computational efficiency.

Lorenzo Micalizzi, Davide Torlo, and Walter Boscheri have contributed equally to this work.

✉ Walter Boscheri
    walter.boscheri@unfe.it

    Lorenzo Micalizzi
    lorenzo.micalizzi@math.uzh.ch

    Davide Torlo
    davide.torlo@sissa.it

1   Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland

2   SISSA mathLab, SISSA, via Bonomea 265, Trieste 34136, Italy

3   Dipartimento di Matematica e Informatica, University of Ferrara, via Machiavelli 30, Ferrara 44121, Italy

# 1 Introduction

In recent years, the need for a very accurate description of physical phenomena in the context of advanced technological applications has determined an increasing interest toward large-scale simulations. To reduce their enormous computational cost and to make them more accessible, several strategies have been proposed, among which:

- parallelization, leading to a reduction of the computational time proportional to the number of employed processors with excellent scaling properties [39, 52, 55, 68, 73, 81, 82];
- structure-preserving schemes, to preserve physical properties at the discrete level without excessive mesh refinements, e.g., positivity preserving schemes [27, 53, 60, 66, 67], well-balanced schemes [12, 22, 25–28, 47, 59, 64, 85], TVD or maximum principle preserving schemes [10, 48, 49, 86], and entropy conservative/dissipating schemes [4–7, 23, 24, 40, 44, 45, 56–58, 65, 69, 83, 84];
- high-order methods, which guarantee higher accuracy for coarser meshes and shorter computational times, on smooth problems, as they are able to catch complicated physical structures that low-order methods struggle to obtain, e.g., finite-element-based methods [1, 3, 8, 54, 57, 70, 71], finite volume methods [10, 12, 22, 64, 75, 85], and discontinuous Galerkin (DG) methods [17, 19, 23, 35, 44–46, 51].

However, high-order methods are, for the moment, mostly relegated to academic contexts. The main reason is given by the fact that concrete applications are characterized by shocks, which are well known to reduce the accuracy to first order, disregarding for the formal order of accuracy. Further, in the presence of shocks, high-order schemes are more subjected to instabilities. Users are, therefore, comprehensibly unwilling to pay extra costs in terms of the complexity of the numerical method and its implementation, if the effort is not rewarded with the initially expected advantages. Indeed, one can observe that in the case of non-regular solutions; usually, the shocks do not cover the whole computational domain but rather some lower dimensional manifolds. Therefore, a possibility to use high-order methods at their best is the adoption of extra procedures to be implemented, e.g., limiters, *a posteriori* correction techniques, blenders with low-order schemes, or adaptive strategies relying on shock detectors. However, such procedures require a relevant interference with the basic implementation as they are not naturally embedded in the original method at the theoretical level and, if their introduction is not performed in a careful way, the additional cost associated with them may be comparable to the computational gain given by the high-order feature.

Here, generalizing the idea introduced in [61], we propose a new arbitrary high-order formulation naturally allowing for order adaptivity, namely the so-called $p$-adaptivity. The formulation relies on an underlying arbitrary high-order iterative scheme, which is easily modified in a suitable efficient way. Arbitrary high-order iterative methods are characterized by iterative procedures involving an approximated solution to a certain problem, whose order of accuracy increases by one at each iteration, converging toward the solution of a background high-order scheme. The idea is to modify the generic iteration in such a way that the order of accuracy of the discretization proper to the iteration itself matches

the order of accuracy achieved in that specific iteration, hence reducing the computational cost. The number of iterations is chosen equal to the aimed order of accuracy, as already done in [8, 38, 50, 61, 63]. On the contrary, in other works, the iterations are stopped when a prescribed tolerance is reached [14, 21, 34, 42, 79]. This is most of the times unnecessary for explicit methods, as the accuracy of the underlying discretization is not as accurate as that tolerance.

The modification we propose in this work results in several advantages: for a fixed final order, we get a substantial drop in the computational cost with respect to the traditional approach as the low-order iterations are performed with low-order structures which are computationally cheaper. Moreover, in this framework, it is straightforward to limit the achieved order on the fly, stopping the computation at a certain iteration if specified criteria are not met. This last aspect allows to overcome the typical drawback of *a posteriori* Multi-dimensional Optimal Order Detection (MOOD) techniques [9, 17, 18, 29, 31, 32], in which, if the high-order scheme produces a solution which is not valid according to some physical or numerical criteria, low-order solutions must be recomputed with their associated computational cost after the high-order solution has been already computed.

Apart from the advantages and the many possible applications, a remarkable aspect which is worth underlying is given by the fact that, if one already has an implementation of an arbitrary high-order iterative scheme, then the introduction of the proposed modification is straightforward. Furthermore, the modification does not prevent the possible parallelization of the original code.

In this paper, we discuss the application to an Arbitrary DERivative (ADER) framework for hyperbolic Partial Differential Equations (PDEs), proposed originally by Millington et al. [62, 78]. The approach is validated on challenging benchmarks, showing the arbitrary high-order character and the optimal performance in the context of the adaptivity. Furthermore, the efficiently designed *a posteriori* limiter, which drives the *p*-adaptivity, allows to provably preserve physical properties. In this work, we will use it to preserve the positivity of some quantities associated with the numerical solution. The resulting numerical schemes are of high order of accuracy with one-step time discretization and making use of general polygonal cells in space.

Summarizing, the main contributions of this work are the following: the generalization of the idea introduced in [61] into an abstract and rigorous framework, its application to the ADER context for hyperbolic PDEs with increasing reconstruction degree along the iterations of the space-time predictor solution, and the design of an efficient structure-preserving *a posteriori* limiter.

The work is organized as follows. In Sect. 2, we describe the general idea in the specific framework of Deferred Correction (DeC) methods. In Sect. 3, we present the ADER method for hyperbolic PDEs with the DG spatial discretization and we explain how it can be interpreted as an iterative arbitrary high-order method. Section 4 is devoted to the description of the proposed modifications for the ADER framework to obtain new efficient *p*-adaptive schemes. The new methods are validated against several challenging benchmarks in Sect. 5, demonstrating the accuracy and the robustness of the novel approach. Moreover, the computational advantages with respect to the original formulation are experimentally shown. Finally, conclusions and future perspectives are reported in Sect. 6.

## 2 New Efficient Iterative Arbitrary High-Order Methods

Iterative arbitrary high-order methods are numerical methods characterized by an iteration process converging to the solution of an underlying arbitrary high-order scheme. Here, we focus on particular iterative arbitrary high-order methods, for which the order of accuracy with respect to the limit solution increases by one at each iteration. Examples of such methods are the DeC [2, 8, 38, 50, 63] and the ADER schemes [16, 34, 35, 51], which have been broadly used in the context of the numerical solution of hyperbolic systems of PDEs. The current use of such methods consists in fixing an underlying high-order scheme and performing the iteration process until convergence or, more efficiently, until the desired accuracy is reached, i.e., with a number of iterations exactly equal to the order of the method.

We propose here a new simple modification of the aforementioned framework which allows the computational cost of the original methods to be reduced, designing novel schemes with a natural adaptive character. In particular, let us consider a general iterative arbitrary high-order method of order $P$, whose generic iteration is denoted by $\mathcal{M}_P$. Then, a simple sketch of the method is given by

$$\underline{u}^{(0)} \xrightarrow{\mathcal{M}_P} \underline{u}^{(1)} \xrightarrow{\mathcal{M}_P} \underline{u}^{(2)} \xrightarrow{\mathcal{M}_P} \cdots \xrightarrow{\mathcal{M}_P} \underline{u}^{(P-1)} \xrightarrow{\mathcal{M}_P} \underline{u}^{(P)}, \tag{1}$$

where $\underline{u}^{(p)}$ is the result of the $p$th iteration and $P$ iterations have been considered to achieve the optimal accuracy with the minimal number of iterations. The proposed modification consists in replacing the generic $p$th iteration of the method of order $P$ with the iteration $\mathcal{M}_p$, that is the iteration associated to the same method but with order $p$ which is in general cheaper but, nevertheless, accurate enough to get order $p$. The sketch of the modified method reads

$$\underline{u}^{(0)} \xrightarrow{\mathcal{M}_1} \underline{u}^{(1)} \xrightarrow{\mathcal{M}_2} \underline{u}^{(2)} \xrightarrow{\mathcal{M}_3} \cdots \xrightarrow{\mathcal{M}_{P-1}} \underline{u}^{(P-1)} \xrightarrow{\mathcal{M}_P} \underline{u}^{(P)}. \tag{2}$$

The formal order of accuracy is not spoiled as $\mathcal{M}_p$, the new $p$th iteration, is still sufficiently accurate to provide $\underline{u}^{(p)}$ of order $p$ starting by $\underline{u}^{(p-1)}$ of order $p-1$. The technical details of changing the iteration structures at each iteration depend on the underlying iterative arbitrary high-order method under consideration. For example, an intermediate embedding process like an interpolation may be needed to project $\underline{u}^{(p-1)}$ onto the same space of $\underline{u}^{(p)}$ in order to perform the $p$th iteration. The modification to pass from the original formulation (1) to the new one (2) is minimal, and its inclusion in an existing implementation of an iterative arbitrary high-order method is straightforward. The modified methods are in general cheaper than the original ones, as the computational cost related to $\mathcal{M}_p$ for $p < P$ is smaller than the one related to $\mathcal{M}_P$. Moreover, differently from what happens in the original framework, increasing the number of iterations always determines an increase in the order of accuracy without any saturation. Therefore, in principle, it is possible not to fix *a priori* the final order, but instead to continue the iterations until a certain tolerance is matched. This may provide a valid strategy for engineering applications.

We focus now on a particular family of iterative arbitrary high-order methods, and we discuss their modification to comply with the *p*-adaptivity setting proposed in this work.

## 2.1 DeC Methods

The DeC is an abstract procedure that can be exploited to design arbitrary high-order iterative methods for differential problems. In particular, the formulation presented in [2] relies on the definition of two operators $\mathcal{L}_\Delta^1, \mathcal{L}_\Delta^2 : X \to Y$ associated with a given problem, dependent on a parameter $\Delta$, acting between two normed vector spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$. The operator $\mathcal{L}_\Delta^2$ is a high-order nonlinear implicit operator that we would like to solve, i.e., to find $\underline{u}_\Delta \in X$, such that $\mathcal{L}_\Delta^2(\underline{u}_\Delta) = \mathbf{0}_Y$, to get a high-order approximation of the solution to the original problem. Nevertheless, due to its implicit nature, the operator $\mathcal{L}_\Delta^2$ is difficult to be solved. On the other hand, the operator $\mathcal{L}_\Delta^1$ is a low-order explicit operator, for which it is easy to find $\underline{\widetilde{u}} \in X$, such that $\mathcal{L}_\Delta^1(\underline{\widetilde{u}}) = \underline{z}$ for $\underline{z} \in Y$. Due to its simplicity, it would be desirable to solve $\mathcal{L}_\Delta^1$, rather than $\mathcal{L}_\Delta^2$. However, the resulting solution would not be accurate enough for our purposes.

The following theorem allows to approximate $\underline{u}_\Delta$ arbitrarily well by a simple explicit iterative procedure, which is much cheaper than the direct solution of the operator $\mathcal{L}_\Delta^2$.

**Theorem 1** (DeC) *Let the operators $\mathcal{L}_\Delta^1$ and $\mathcal{L}_\Delta^2$ fulfill the following properties.*

   (i)   *Existence of a unique solution to $\mathcal{L}_\Delta^2$*
            *$\exists! \, \underline{u}_\Delta \in X$ solution of $\mathcal{L}_\Delta^2$, such that $\mathcal{L}_\Delta^2(\underline{u}_\Delta) = \mathbf{0}_Y$.*
  (ii)  *Coercivity-like property of $\mathcal{L}_\Delta^1$*
            *$\exists \, \alpha_1 \geqslant 0$ independent of $\Delta$ s.t.*

$$\left\| \mathcal{L}_\Delta^1(\underline{v}) - \mathcal{L}_\Delta^1(\underline{w}) \right\|_Y \geqslant \alpha_1 \left\| \underline{v} - \underline{w} \right\|_X, \ \forall \underline{v}, \underline{w} \in X. \tag{3}$$

 (iii)  *Lipschitz-continuity-like property of $\mathcal{L}_\Delta^1 - \mathcal{L}_\Delta^2$*
            *$\exists \, \alpha_2 \geqslant 0$ independent of $\Delta$ s.t.*

$$\left\| \left[ \mathcal{L}_\Delta^1(\underline{v}) - \mathcal{L}_\Delta^2(\underline{v}) \right] - \left[ \mathcal{L}_\Delta^1(\underline{w}) - \mathcal{L}_\Delta^2(\underline{w}) \right] \right\|_Y \leqslant \alpha_2 \Delta \left\| \underline{v} - \underline{w} \right\|_X, \ \forall \underline{v}, \underline{w} \in X. \tag{4}$$

*Given a $\underline{u}^{(0)} \in X$, define recursively the sequence of vectors $\underline{u}^{(p)}$ as the solution of*

$$\mathcal{L}_\Delta^1(\underline{u}^{(p)}) := \mathcal{L}_\Delta^1(\underline{u}^{(p-1)}) - \mathcal{L}_\Delta^2(\underline{u}^{(p-1)}), \quad p \geqslant 1. \tag{5}$$

*Then, the following error estimate holds*:

$$\left\| \underline{u}^{(p)} - \underline{u}_\Delta \right\|_X \leqslant \left( \Delta \frac{\alpha_2}{\alpha_1} \right)^p \left\| \underline{u}^{(0)} - \underline{u}_\Delta \right\|_X, \quad \forall p \in \mathbb{N}. \tag{6}$$

**Proof** Using the coercivity-like property of $\mathcal{L}_\Delta^1$ (3), the definition of the DeC iteration (5), the fact that $\mathcal{L}_\Delta^2(\underline{u}_\Delta) = \mathbf{0}_Y$, and the Lipschitz-continuity-like property of $\mathcal{L}_\Delta^1 - \mathcal{L}_\Delta^2$ (4), we get

$$
\begin{aligned}
\left\| \underline{\boldsymbol{u}}^{(p)} - \underline{\boldsymbol{u}}_\Delta \right\|_X &\leqslant \frac{1}{\alpha_1} \left\| \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}^{(p)}) - \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}_\Delta) \right\|_Y \\
&= \frac{1}{\alpha_1} \left\| \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}_\Delta^2 (\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}_\Delta) \right\|_Y \\
&= \frac{1}{\alpha_1} \left\| \left[ \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}^{(p-1)}) - \mathcal{L}_\Delta^2 (\underline{\boldsymbol{u}}^{(p-1)}) \right] - \left[ \mathcal{L}_\Delta^1 (\underline{\boldsymbol{u}}_\Delta) - \mathcal{L}_\Delta^2 (\underline{\boldsymbol{u}}_\Delta) \right] \right\|_Y \\
&\leqslant \Delta \frac{\alpha_2}{\alpha_1} \left\| \underline{\boldsymbol{u}}^{(p-1)} - \underline{\boldsymbol{u}}_\Delta \right\|_X.
\end{aligned}
\tag{7}
$$

Applying recursively the previous inequality, we obtain the desired result.

For $\Delta$ small enough, the sequence of vectors $\underline{\boldsymbol{u}}^{(p)}$ converges to $\underline{\boldsymbol{u}}_\Delta$ independently of the initial vector $\underline{\boldsymbol{u}}^{(0)}$. At each iteration of the DeC procedure (5), the computation of $\underline{\boldsymbol{u}}^{(p)}$ is straightforward by our assumptions on the operator $\mathcal{L}_\Delta^1$, since $\underline{\boldsymbol{u}}^{(p-1)}$ is known and it is possible to explicitly compute the right-hand side. Furthermore, thanks to the accuracy estimate (6), at each iteration, one order of accuracy is gained with respect to $\underline{\boldsymbol{u}}_\Delta$.

**Remark 1** (On "over-resolving" the operator $\mathcal{L}_\Delta^2$ and on the number of iterations $P$) Usually, we are not strictly interested in the solution $\underline{\boldsymbol{u}}_\Delta$ of the operator $\mathcal{L}_\Delta^2$, but rather on the analytical solution $\underline{\boldsymbol{u}}_{\mathrm{ex}}$ of the underlying problem. If $S$ is the order of accuracy of $\underline{\boldsymbol{u}}_\Delta$, in general, it suffices to approximate $\underline{\boldsymbol{u}}_\Delta$ with the $S$th-order accuracy. This consideration allows to bound the number of iterations, saving computational time, without necessarily getting the convergence toward the solution to the operator $\mathcal{L}_\Delta^2$. In particular, thanks to Theorem 1, if $\underline{\boldsymbol{u}}^{(0)}$ is an $O(\Delta)$-approximation of $\underline{\boldsymbol{u}}_{\mathrm{ex}}$, then $\left\| \underline{\boldsymbol{u}}^{(P)} - \underline{\boldsymbol{u}}_{\mathrm{ex}} \right\| = O(\Delta^{1+\min(P,S)})$ leading to an order of accuracy equal to $\min(P, S)$. Hence, the optimal choice is $P = S$. Any further iteration will not increase the order of accuracy of the method with respect to $\underline{\boldsymbol{u}}_{\mathrm{ex}}$ but only with respect to $\underline{\boldsymbol{u}}_\Delta$.

**Remark 2** (On explicit and implicit DeC methods) The DeC philosophy is based on having a simple iterative procedure allowing to obtain a high-order approximation of the solution of a given problem, which would have been difficult to compute directly. In the applications, we will focus on an explicit setting by considering explicit operators $\mathcal{L}_\Delta^1$, leading to an updated formula (5) with an explicit character. However, within the described framework, one could easily switch to an implicit setting by selecting an implicit low-order operator $\mathcal{L}_\Delta^1$. In that case, one obtains an iterative procedure that is not explicit, yet, much simpler than the direct solution of $\mathcal{L}_\Delta^2$. In applications to ordinary differential equations (ODEs) and PDEs, implicit formulations [8, 17, 20] allow to achieve better stability properties and less time step restrictions. The theoretical framework presented in this work applies both to explicit and implicit settings.

## 2.2 New Efficient DeC Methods

Here, we will discuss, at the theoretical level, an efficient modification for DeC methods. It is based on the replacement of the operators $\mathcal{L}_\Delta^1$ and $\mathcal{L}_\Delta^2$ by iteration-specific operators

$\mathcal{L}_\Delta^{1,(p)}$ and $\mathcal{L}_\Delta^{2,(p)}$ to strictly obtain the order $p$ at the $p$th iteration. In particular, we prove the following result.

**Theorem 2** *Consider a problem with an exact solution $\underline{u}_{ex} \in Z$. Then, take some normed spaces $(X^{(p)}, \|\cdot\|_{X^{(p)}})$ for $p \in \mathbb{N}$ and $(Y^{(p)}, \|\cdot\|_{Y^{(p)}})$ for $p \geqslant 1$. For every $p \geqslant 1$, consider also two operators $\mathcal{L}_\Delta^{1,(p)}, \mathcal{L}_\Delta^{2,(p)} : X^{(p)} \to Y^{(p)}$ dependent on the same parameter $\Delta$ and fulfilling the properties of Theorem 1 for some $\alpha_1^{(p)}, \alpha_2^{(p)} > 0$ and $\underline{u}_\Delta^{(p)} \in X^{(p)}$. Furthermore, let us assume that for any $p \in \mathbb{N}$, there exists an embedding operator $\mathcal{E}^{(p)} : X^{(p)} \to X^{(p+1)}$, associating with each $\underline{u}^{(p)} \in X^{(p)}$ an approximation $\underline{u}^{*(p)} := \mathcal{E}^{(p)}(\underline{u}^{(p)}) \in X^{(p+1)}$, and some projection $\Pi^{(p)} : Z \to X^{(p)}$, associating with $\underline{u}_{ex}$ an approximation $\underline{u}_{ex}^{(p)} := \Pi^{(p)}(\underline{u}_{ex}) \in X^{(p)}$.*

*Given $\underline{u}^{(0)} \in X^{(0)}$, we consider the new DeC method whose general $p$th iteration is given by*

$$\begin{cases} \underline{u}^{*(p-1)} := \mathcal{E}^{(p-1)}(\underline{u}^{(p-1)}), \\ \mathcal{L}_\Delta^{1,(p)}(\underline{u}^{(p)}) := \mathcal{L}_\Delta^{1,(p)}(\underline{u}^{*(p-1)}) - \mathcal{L}_\Delta^{2,(p)}(\underline{u}^{*(p-1)}). \end{cases} \tag{8}$$

*Suppose that the following properties hold*:

(i) *the accuracy of $\underline{u}_\Delta^{(p)}$ with respect to $\underline{u}_{ex}^{(p)}$*

$$\left\| \underline{u}_\Delta^{(p)} - \underline{u}_{ex}^{(p)} \right\|_{X^{(p)}} = O(\Delta^{p+1}), \quad p \geqslant 1; \tag{9}$$

(ii) *the accuracy of the embedding $\mathcal{E}^{(p)}$*

$$\left\| \underline{u}^{*(p)} - \underline{u}_{ex}^{(p+1)} \right\|_{X^{(p+1)}} \leqslant C \left\| \underline{u}^{(p)} - \underline{u}_{ex}^{(p)} \right\|_{X^{(p)}}, \quad \forall p \in \mathbb{N} \tag{10}$$

*for some constant $C$ independent on $\Delta$;*

(iii) *the accuracy of $\underline{u}^{(0)}$*

$$\left\| \underline{u}^{(0)} - \underline{u}_{ex}^{(0)} \right\|_{X^{(0)}} = O(\Delta). \tag{11}$$

*Then, it follows that*

$$\left\| \underline{u}^{(p)} - \underline{u}_{ex}^{(p)} \right\|_{X^{(p)}} = O(\Delta^{p+1}), \quad \forall p \in \mathbb{N}. \tag{12}$$

**Proof** The proof is based on the induction. The base case for $p = 0$ is trivially given by assumption (11). Let us now focus on the induction step. We assume that (12) holds for a specific $p$ and we will prove it for $p + 1$. By the triangular inequality, we have

$$\left\| \underline{u}^{(p+1)} - \underline{u}_{ex}^{(p+1)} \right\|_{X^{(p+1)}} \leqslant \left\| \underline{u}^{(p+1)} - \underline{u}_\Delta^{(p+1)} \right\|_{X^{(p+1)}} + \left\| \underline{u}_\Delta^{(p+1)} - \underline{u}_{ex}^{(p+1)} \right\|_{X^{(p+1)}}. \tag{13}$$

The second term at the right-hand side is $O(\Delta^{p+2})$ for (9), and hence, let us focus on the first term. By the proof of Theorem 1 concerning the original methods, we have that

$$\left\|\underline{\boldsymbol{u}}^{(p+1)} - \underline{\boldsymbol{u}}_{\Delta}^{(p+1)}\right\|_{X^{(p+1)}} \leqslant \Delta \frac{\alpha_2^{(p+1)}}{\alpha_1^{(p+1)}} \left\|\underline{\boldsymbol{u}}^{*(p)} - \underline{\boldsymbol{u}}_{\Delta}^{(p+1)}\right\|_{X^{(p+1)}}, \tag{14}$$

which, applying the triangular inequality, gives

$$
\begin{aligned}
&\left\|\underline{\boldsymbol{u}}^{(p+1)} - \underline{\boldsymbol{u}}_{\Delta}^{(p+1)}\right\|_{X^{(p+1)}} \\
&\leqslant \Delta \frac{\alpha_2^{(p+1)}}{\alpha_1^{(p+1)}} \left( \left\|\underline{\boldsymbol{u}}^{*(p)} - \underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p+1)}\right\|_{X^{(p+1)}} + \left\|\underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p+1)} - \underline{\boldsymbol{u}}_{\Delta}^{(p+1)}\right\|_{X^{(p+1)}} \right).
\end{aligned}
\tag{15}
$$

Again, due to (9), the second term in parenthesis at the right-hand side is $O(\Delta^{p+2})$; therefore, we focus on the first term. Due to the assumption on the accuracy of the embedding (10) and to the induction hypothesis, we have

$$\left\|\underline{\boldsymbol{u}}^{*(p)} - \underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p+1)}\right\|_{X^{(p+1)}} \leqslant C \left\|\underline{\boldsymbol{u}}^{(p)} - \underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p)}\right\|_{X^{(p)}} = O(\Delta^{p+1}). \tag{16}$$

Hence

$$\Delta \frac{\alpha_2^{(p+1)}}{\alpha_1^{(p+1)}} \left\|\underline{\boldsymbol{u}}^{*(p)} - \underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p+1)}\right\|_{X^{(p+1)}} = O(\Delta^{p+2}), \tag{17}$$

which completes the proof.

Let us notice that, in the previous theorem, the accuracy estimate is always referred to a projection of the exact solution and not to a fixed high-order approximation. Hence, the order of accuracy is formally not bounded and we can approximate $\underline{\boldsymbol{u}}_{\mathrm{ex}}$ arbitrarily well. In particular, if $\underline{\boldsymbol{u}}_{\mathrm{ex}}^{(p)}$ yields an approximation of $\underline{\boldsymbol{u}}_{\mathrm{ex}}$ which is $O(\Delta^{p+1})$ accurate, thanks to Theorem 2, also the approximation associated with $\underline{\boldsymbol{u}}^{(p)}$ will have the same accuracy with respect to $\underline{\boldsymbol{u}}_{\mathrm{ex}}$.

## 3 ADER-Discontinuous Galerkin (DG) Scheme

The ADER methods are various techniques to obtain arbitrary high-order methods for differential problems. Even though the first ADER [82] was based on the Cauchy-Kovalevskaya theorem; nowadays, it is mainly known as a technique that exploits the weak formulation of the original problem to obtain high-order discretization forms that are solved iteratively [16, 34, 50]. In this section, we will present a formulation for hyperbolic PDEs in combination with a DG space discretization, and we will show how it can be interpreted as an arbitrary high-order iterative method in the previously presented DeC framework. We will also describe in a final subsection the ADER-$\mathbb{P}_N\mathbb{P}_M$ variant of the method, still recastable as the DeC scheme, which allows for applications to finite volume (FV) formulations, as well.

## 3.1 Numerical Method

We want to approximate the analytical solution $\boldsymbol{u} : \overline{\Omega} \times \mathbb{R}_0^+ \to \mathbb{R}^Q$ of the following $Q$-dimensional hyperbolic PDE:

$$\frac{\partial}{\partial t}\boldsymbol{u}(\boldsymbol{x}, t) + \mathrm{div}_{\boldsymbol{x}} \boldsymbol{F}(\boldsymbol{u}(\boldsymbol{x}, t)) = \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}, t)), \quad (\boldsymbol{x}, t) \in \Omega \times \mathbb{R}_0^+, \tag{18}$$

supplemented with suitable initial and boundary conditions, where $\Omega \subseteq \mathbb{R}^D$ is a bounded $D$-dimensional space domain, $\boldsymbol{F} : \mathbb{R}^Q \to \mathbb{R}^{Q \times D}$ is the flux tensor, and $\boldsymbol{S} : \overline{\Omega} \times \mathbb{R}^Q \to \mathbb{R}^Q$ is the source function. To shorten the notation, let us define $\boldsymbol{E}(\boldsymbol{u}, \boldsymbol{x}) := \mathrm{div}_{\boldsymbol{x}} \boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u})$, the time evolution operator of the PDE up to the minus sign, so that (18) becomes

$$\frac{\partial}{\partial t}\boldsymbol{u}(\boldsymbol{x}, t) + \boldsymbol{E}(\boldsymbol{u}(\boldsymbol{x}, t), \boldsymbol{x}) = \boldsymbol{0}, \quad (\boldsymbol{x}, t) \in \Omega \times \mathbb{R}_0^+. \tag{19}$$

Let us focus on a generic time step $[t_n, t_{n+1}]$ with $\Delta t := t_{n+1} - t_n$. The goal is to find an approximation $\boldsymbol{u}_{n+1}(\boldsymbol{x}) \approx \boldsymbol{u}(\boldsymbol{x}, t_{n+1})$ of the analytical solution in $\overline{\Omega}$ at time $t_{n+1}$ by knowing an approximation $\boldsymbol{u}_n(\boldsymbol{x}) \approx \boldsymbol{u}(\boldsymbol{x}, t_n)$ at time $t_n$.

In particular, for any $n$, we adopt a classical DG space discretization for $\boldsymbol{u}_n(\boldsymbol{x}) \approx \boldsymbol{u}(\boldsymbol{x}, t_n)$: we consider a tessellation $\mathcal{T}_h$ of $\overline{\Omega}$ made of non-overlapping convex polytopals $K$ with a mesh parameter $h$, and we consider $\boldsymbol{u}_n(\boldsymbol{x})$ in a space of discontinuous piecewise polynomial functions of degree $M$, i.e., $(V_M)^Q$ with $V_M := \{g \in L^2(\Omega) \text{ s.t. } g|_K \in \mathbb{P}_M(K)\}$, yielding an $(M + 1)$th order of accuracy approximation space. Therefore, locally in each element $K$, we can consider the following representation of the approximated solution with a local basis $\{\varphi_i(\boldsymbol{x})\}_{i=1,\cdots,I}$ of $\mathbb{P}_M(K)$:

$$\boldsymbol{u}_n(\boldsymbol{x}) := \sum_{i=1}^{I} \boldsymbol{c}_i^n \varphi_i(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in K, \tag{20}$$

where $I$ is the number of local basis functions and the label $K$ on the coefficients $\boldsymbol{c}_i^n$, and on the basis functions, $\varphi_i$ is omitted to lighten the notation as we will consider computations in a single generic element, in the sequel.

The ADER method, applied to this context, is based on the weak formulation of the governing equations (19) in space-time and it is characterized by two steps: an iterative local space-time predictor and a final corrector step, described hereafter. Before entering the details, it is useful to briefly describe the role of such steps for ADER methods. The predictor, based on a local explicit iterative procedure, is used to compute local high-order polynomial approximations of the solution in space-time control volumes without considering any communication between different cells. The solution obtained in this step is high-order accurate but not stable, as no upwinding has been taken into account in its computation. This prediction is later used in the corrector step to provide a global explicit update of the numerical solution, allowing communication between neighboring cells through numerical fluxes, which introduce the necessary upwinding and numerical dissipation to achieve the stability. A stability study of the method can be found in [34].

### 3.1.1 Local Space-Time Predictor

The purpose of this step is to find a high-order approximation of the solution in each space-time control volume $C_K = K \times [t_n, t_{n+1}]$. In this step, no communication between the cells

happens. We consider the weak formulation of (18) over $C_K$, obtained through the multiplication by a smooth test function $\vartheta : C_K \to \mathbb{R}$, the integration over $C_K$ and subsequent integration by parts in time:

$$
\int_K \left[ \boldsymbol{u}(\boldsymbol{x}, t_{n+1})\vartheta(\boldsymbol{x}, t_{n+1}) - \boldsymbol{u}(\boldsymbol{x}, t_n)\vartheta(\boldsymbol{x}, t_n) \right] \mathrm{d}\boldsymbol{x} - \int_{C_K} \boldsymbol{u}(\boldsymbol{x}, t)\frac{\partial}{\partial t}\vartheta(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t
$$
$$
+ \int_{C_K} \boldsymbol{E}(\boldsymbol{u}(\boldsymbol{x}, t), \boldsymbol{x})\vartheta(\boldsymbol{x}, t)\,\mathrm{d}\boldsymbol{x}\mathrm{d}t = \boldsymbol{0}. \tag{21}
$$

Next, we project it onto a finite-dimensional space spanned by the tensor product of the previously introduced local spatial basis $\left\{ \varphi_i(\boldsymbol{x}) \right\}_{i=1,\cdots,I}$ and a temporal basis $\{\psi^m(t)\}_{m=0,\cdots,M}$ over $[t_n, t_{n+1}]$ guaranteeing $(M + 1)$th order of accuracy. As an example for the latter, one can think to a Lagrangian basis of degree $M$ or a truncated Taylor series up to the $M$th-degree term to obtain an approximation of order $M + 1$. Here, we will use modal basis functions both for space and time, although such choice is not mandatory. The basis functions are explicitly described in Appendix A. As usual in the literature, we assume the basis functions to be normalized in such a way that their maximum absolute value over $C_K$ is an $O(1)$.

In particular, we consider the local discretization of $\boldsymbol{u}_h$ in $C_K$

$$
\boldsymbol{u}_h(\boldsymbol{x}, t) := \sum_{i=1}^I \sum_{m=0}^M \boldsymbol{u}_i^m \varphi_i(\boldsymbol{x})\psi^m(t) = \sum_{\ell=1}^L \boldsymbol{u}^\ell \vartheta^\ell(\boldsymbol{x}, t), \quad \forall (\boldsymbol{x}, t) \in C_K, \tag{22}
$$

in which, to shorten the notation, we have denoted by $\{\vartheta^\ell(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L}$ a permutation of the basis functions $\{\varphi_i(\boldsymbol{x})\psi^m(t)\}_{\substack{i=1,\ldots,I \\ m=0,\ldots,M}}$ and by $\boldsymbol{u}^\ell$ the corresponding coefficients $\boldsymbol{u}_i^m$, where implicitly we defined a bijection that gives $\ell = \ell(i, m)$. Finally, we consider the projection of (21) on the space-time DG functional space generated by $\{\vartheta^\ell(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L}$, that is

$$
\sum_{\ell=1}^L \left[ \int_K \vartheta^\ell(\boldsymbol{x}, t_{n+1})\vartheta^j(\boldsymbol{x}, t_{n+1})\mathrm{d}\boldsymbol{x} - \int_{C_K} \vartheta^\ell(\boldsymbol{x}, t)\frac{\partial}{\partial t}\vartheta^j(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t \right] \boldsymbol{u}^\ell
$$
$$
- \int_K \boldsymbol{u}_n(\boldsymbol{x})\vartheta^j(\boldsymbol{x}, t_n)\mathrm{d}\boldsymbol{x} + \int_{C_K} \boldsymbol{E}(\boldsymbol{u}_h(\boldsymbol{x}, t), \boldsymbol{x})\vartheta^j(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t = \boldsymbol{0} \tag{23}
$$

for any $j = 1, \cdots, L$. This is a nonlinear system in the unknowns $\boldsymbol{u}^\ell$, whose solution yields the $(M + 1)$th-order accurate approximation (22) of the analytical solution. Let us notice that $\boldsymbol{u}_n(\boldsymbol{x})$ in (23) is known by assumption and the related integral involving such function can thus be computed. Notice that, to obtain a fully local formulation, the divergence theorem in space has not been applied. On the other hand, the integration by parts in time has been performed to introduce a causality effect and a dependency on the initial information at time $t_n$.

Now, it is possible to recast each local system (23) in a matrix-vector formulation writing

$$
B\underline{\boldsymbol{u}} - \underline{\boldsymbol{r}} + \widetilde{\boldsymbol{\phi}}(\underline{\boldsymbol{u}}) = \boldsymbol{0}, \tag{24}
$$

where the matrix $B$ and the vectors $\underline{\boldsymbol{u}}$, $\underline{\boldsymbol{r}}$, and $\widetilde{\boldsymbol{\phi}}$ are given by

$$\begin{cases} B_{j,\ell} := \int_K \vartheta^\ell(\boldsymbol{x}, t_{n+1})\vartheta^j(\boldsymbol{x}, t_{n+1})\mathrm{d}\boldsymbol{x} - \int_{C_K} \vartheta^\ell(\boldsymbol{x}, t)\frac{\partial}{\partial t}\vartheta^j(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t, \\[2mm] \underline{\boldsymbol{u}} := \begin{pmatrix} \boldsymbol{u}^1 \\ \vdots \\ \boldsymbol{u}^L \end{pmatrix}, \quad \underline{\boldsymbol{r}} := \begin{pmatrix} \int_K \boldsymbol{u}_n(\boldsymbol{x})\vartheta^1(\boldsymbol{x}, t_n)\mathrm{d}\boldsymbol{x} \\ \vdots \\ \int_K \boldsymbol{u}_n(\boldsymbol{x})\vartheta^L(\boldsymbol{x}, t_n)\mathrm{d}\boldsymbol{x} \end{pmatrix}, \quad \widetilde{\underline{\boldsymbol{\phi}}}(\underline{\boldsymbol{u}}) := \begin{pmatrix} \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_1(\underline{\boldsymbol{u}}, t)\mathrm{d}t \\ \vdots \\ \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_L(\underline{\boldsymbol{u}}, t)\mathrm{d}t \end{pmatrix} \end{cases} \tag{25}$$

with $\boldsymbol{\phi}_j(\underline{\boldsymbol{u}}, t) := \int_K \boldsymbol{E}(\boldsymbol{u}_h(\boldsymbol{x}, t), \boldsymbol{x})\vartheta^j(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}$. Let us observe that $\underline{\boldsymbol{r}}$ is constant and explicitly computable as $\boldsymbol{u}_n(\boldsymbol{x})$ is known.

**Remark 3** (On the matrix $B$) The definition of the matrix $B$ is referred to a scalar PDE; it must be block-expanded for a vectorial problem. Let us notice that the elements of the matrix $B$ are $O(h^D)$ due to the integral over $K$ and to the normalization assumed on the basis functions. The integral in time on the second term of $B_{j,\ell}$ is balanced by the derivative in time on $\vartheta^j$.

Concerning the well-posedness and the solution of the nonlinear system (24), we can prove that for $\Delta t$ small enough, it admits a unique solution which can be obtained through the iterative procedure

$$\underline{\boldsymbol{u}}^{(p)} = B^{-1}\left[\underline{\boldsymbol{r}} - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{\boldsymbol{u}}^{(p-1)})\right], \tag{26}$$

which converges unconditionally to the solution of the system, for any initial vector $\underline{\boldsymbol{u}}^{(0)}$. To do that, let us first prove the following useful lemma.

**Lemma 1** (Lipschitz-continuity-like property of $\widetilde{\underline{\boldsymbol{\phi}}}$) *Under smoothness assumptions, the function $\widetilde{\underline{\boldsymbol{\phi}}}(\cdot)$ is such that*

$$\left\|\widetilde{\underline{\boldsymbol{\phi}}}(\underline{\boldsymbol{v}}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{\boldsymbol{w}})\right\|_\infty \leqslant \Delta t|K|C_{\mathrm{Lip}}\|\underline{\boldsymbol{v}} - \underline{\boldsymbol{w}}\|_\infty, \tag{27}$$

*where $\|\cdot\|_\infty$ is the infinity norm over $\mathbb{R}^{L \times Q}$ and $C_{\mathrm{Lip}}$ is a constant independent of $\Delta t$ and of the element $K$.*

**Proof** By a direct computation of the generic $j$th component of the left-hand side of (27), recalling the definition of the functions $\boldsymbol{\phi}_j$, through basic analysis, we get

$$\left\|\int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{\boldsymbol{v}}, t)\mathrm{d}t - \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{\boldsymbol{w}}, t)\mathrm{d}t\right\|_{\infty,Q} \leqslant \int_{t_n}^{t_{n+1}} \left\|\boldsymbol{\phi}_j(\underline{\boldsymbol{v}}, t) - \boldsymbol{\phi}_j(\underline{\boldsymbol{w}}, t)\right\|_{\infty,Q} \mathrm{d}t$$
$$\leqslant \int_{C_K} \left\|\boldsymbol{E}(\boldsymbol{v}_h(\boldsymbol{x}, t), \boldsymbol{x}) - \boldsymbol{E}(\boldsymbol{w}_h(\boldsymbol{x}, t), \boldsymbol{x})\right\|_{\infty,Q}\left|\vartheta^j(\boldsymbol{x}, t)\right|\mathrm{d}\boldsymbol{x}\mathrm{d}t, \tag{28}$$

where $\|\cdot\|_{\infty,Q}$ is the infinity norm over $\mathbb{R}^Q$, $\boldsymbol{v}_h(\boldsymbol{x}, t) := \sum_{\ell=1}^L \boldsymbol{v}^\ell \vartheta^\ell(\boldsymbol{x}, t)$, and $\boldsymbol{w}_h(\boldsymbol{x}, t) := \sum_{\ell=1}^L \boldsymbol{w}^\ell \vartheta^\ell(\boldsymbol{x}, t)$ for any $(\boldsymbol{x}, t) \in C_K$. For regular data, we can assume that the following Lipschitz-continuity property holds:

$$\left\| \boldsymbol{E}(v_h(\boldsymbol{x},t),\boldsymbol{x}) - \boldsymbol{E}(w_h(\boldsymbol{x},t),\boldsymbol{x}) \right\|_{\infty,Q} \leqslant C_0 \left\| \underline{v} - \underline{w} \right\|_\infty, \tag{29}$$

where $C_0$ is a constant independent of $\Delta t$ and of the element $K$, leading to

$$\left\| \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{v},t)\mathrm{d}t - \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{w},t)\mathrm{d}t \right\|_{\infty,Q} \leqslant C_0 \left\| \underline{v} - \underline{w} \right\|_\infty \int_{C_K} \left| \vartheta^j(\boldsymbol{x},t) \right| \mathrm{d}\boldsymbol{x}\mathrm{d}t. \tag{30}$$

The space-time basis functions $\vartheta^j$ are bounded in absolute value by a constant $C_\vartheta$ independent of $\Delta t$ and $K$, yielding

$$\left\| \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{v},t)\mathrm{d}t - \int_{t_n}^{t_{n+1}} \boldsymbol{\phi}_j(\underline{w},t)\mathrm{d}t \right\|_{\infty,Q} \leqslant C_0 C_\vartheta \left\| \underline{v} - \underline{w} \right\|_\infty \Delta t |K|, \tag{31}$$

which, setting $C_{\mathrm{Lip}} := C_0 C_\vartheta$ and taking the maximum over $j = 1, \cdots, L$ at the left-hand side, is the thesis.

A straightforward consequence of the previous result is the following corollary.

**Corollary 1** (Lipschitz-continuity-like property of $B^{-1}\widetilde{\boldsymbol{\phi}}$) *Under the assumptions of the previous lemma, it holds*

$$\left\| B^{-1}\left[ \widetilde{\underline{\boldsymbol{\phi}}}(\underline{v}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{w}) \right] \right\|_\infty \leqslant \Delta t \widetilde{C}_{\mathrm{Lip}} \left\| \underline{v} - \underline{w} \right\|_\infty, \tag{32}$$

*where $\widetilde{C}_{\mathrm{Lip}}$ is a constant independent of $\Delta t$ and of the element $K$.*

**Proof** By basic linear algebra, we have

$$\left\| B^{-1}\left[ \widetilde{\underline{\boldsymbol{\phi}}}(\underline{v}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{w}) \right] \right\|_\infty \leqslant \left\| B^{-1} \right\|_\infty \left\| \widetilde{\underline{\boldsymbol{\phi}}}(\underline{v}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{w}) \right\|_\infty, \tag{33}$$

where the infinity norm applied to $B^{-1}$ is the matrix norm induced by the related vector norm. As observed in Remark 3, $B$ is an $O(h^D)$ and, hence, its inverse is an $O(h^{-D})$, leading to $\left\| B^{-1} \right\|_\infty \leqslant C_B h^{-D}$ for some constant $C_B$ independent of the specific element $K$ and of $\Delta t$. Using this fact, in combination with the result of Lemma 1, we obtain

$$\left\| B^{-1}\left[ \widetilde{\underline{\boldsymbol{\phi}}}(\underline{v}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{w}) \right] \right\|_\infty \leqslant C_B h^{-D} \Delta t |K| C_{\mathrm{Lip}} \left\| \underline{v} - \underline{w} \right\|_\infty. \tag{34}$$

By observing that for a regular mesh $|K| \leqslant C_\tau h^D$ for some constant $C_\tau$ independent of $K$, we get the thesis

$$\left\| B^{-1}\left[ \widetilde{\underline{\boldsymbol{\phi}}}(\underline{v}) - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{w}) \right] \right\|_\infty \leqslant \Delta t C_B C_\tau C_{\mathrm{Lip}} \left\| \underline{v} - \underline{w} \right\|_\infty \tag{35}$$

for $\widetilde{C}_{\mathrm{Lip}} := C_B C_\tau C_{\mathrm{Lip}}$.

This allows us to prove the existence and uniqueness of the solution of (24).

**Proposition 1** (Well-posedness and solution of the nonlinear system) *Because $\Delta t$ is small enough, the nonlinear system* (24) *has a unique solution, which is the limit of* (26) *Because $p \rightarrow +\infty$.*

**Proof** We define the map $\mathcal{J} : \mathbb{R}^{L \times Q} \rightarrow \mathbb{R}^{L \times Q}$ as $\mathcal{J}(\underline{u}) := B^{-1}\left[\underline{r} - \underline{\widetilde{\phi}}(\underline{u})\right]$. It is immediate to verify that a fixed point of $\mathcal{J}$ (if any) is also a solution of (24) and viceversa. Due to the fact that $\mathbb{R}^{L \times Q}$ is finite dimensional, if we are able to prove that $\mathcal{J}$ is a contraction, by the Banach fixed-point theorem, we know that there exists a unique fixed point and that this can be obtained as the limit of the iterative procedure $\underline{u}^{(p)} := \mathcal{J}(\underline{u}^{(p-1)})$, which is equivalent to (26). We will now show that, for $\Delta t$ small enough, $\mathcal{J}$ is indeed a contraction. In fact, by a direct computation, it holds

$$\left\|\mathcal{J}(\underline{v}) - \mathcal{J}(\underline{w})\right\|_{\infty} = \left\|B^{-1}\left[\underline{\widetilde{\phi}}(\underline{v}) - \underline{\widetilde{\phi}}(\underline{w})\right]\right\|_{\infty} \tag{36}$$

and, applying Corollary 1 on the Lipschitz-continuity-like property of $B^{-1}\underline{\widetilde{\phi}}$, we retrieve the thesis

$$\left\|\mathcal{J}(\underline{v}) - \mathcal{J}(\underline{w})\right\|_{\infty} = \left\|B^{-1}\left[\underline{\widetilde{\phi}}(\underline{v}) - \underline{\widetilde{\phi}}(\underline{w})\right]\right\|_{\infty} \leqslant \Delta t \widetilde{C}_{\text{Lip}} \|\underline{v} - \underline{w}\|_{\infty} \tag{37}$$

for $\Delta t < \frac{1}{\widetilde{C}_{\text{Lip}}}$.

All the local approximations, obtained by solving the nonlinear system in each control volume $C_K$, constitute a global $(M + 1)$th-order accurate approximation of the analytical solution. It is piecewise polynomial in each $C_K$ and discontinuous across the faces of $C_K$ shared with other control volumes and we denote it, by an abuse of notation, as $\boldsymbol{u}_h$.

**Remark 4** (On the computational efficiency) In several works, the nonlinear system (24) is solved by carrying the iterative process (26) until a convergence criterion is met up to a certain tolerance [16, 34]. This leads to a waste of resources as the underlying discretization error of the system (24) with respect to the analytical solution of the PDE is of order $M + 1$, and hence, smaller tolerances are in general unnecessary. In this context, it is possible to obtain an $(M + 1)$th-order accurate approximation of the solution of (24) by performing exactly $M + 1$ iterations. More details on this will be explained in Sect. 3.2.

### 3.1.2 Final Corrector Step

From the predictor step, we have in each control volume $C_K$ a local $(M + 1)$th-order accurate approximation $\boldsymbol{u}_h$ of the analytical solution in the form (22), which has been computed without considering any sort of communication between the neighboring elements. In the final corrector step, we exploit such approximation to finally get $\boldsymbol{u}_{n+1}(\boldsymbol{x})$, taking into account the coupling between the elements. In particular, we consider again a weak formulation of (18) in $C_K$, but this time, we use a spatial-only test function $\varphi(\boldsymbol{x})$ and we apply the divergence theorem in space thus getting

$$\int_K \boldsymbol{u}(\boldsymbol{x}, t_{n+1})\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} - \int_K \boldsymbol{u}(\boldsymbol{x}, t_n)\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$
$$+ \int_{t_n}^{t_{n+1}} \int_{\partial K} \varphi(\boldsymbol{x})\boldsymbol{F}(\boldsymbol{u}(\boldsymbol{x}, t)) \cdot \boldsymbol{v}(\boldsymbol{x})\mathrm{d}\boldsymbol{\sigma}\mathrm{d}t \tag{38}$$
$$- \int_{C_K} \boldsymbol{F}(\boldsymbol{u}(\boldsymbol{x}, t)) \cdot \nabla_{\boldsymbol{x}}\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_{C_K} \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}, t))\varphi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t = \boldsymbol{0},$$

where $\boldsymbol{v}(\boldsymbol{x})$ is the outward pointing normal to $\partial K$. The divergence theorem in space provides the desired coupling between the neighboring cells, because the solution $\boldsymbol{u}_h$, computed locally in each control volume $C_K$ through the predictor, is discontinuous across the boundaries $\partial K$ and, thus, a numerical flux $\widehat{\boldsymbol{F}}$ is needed to compute the flux at the cell interfaces $\partial K$. We can use either a simple and robust local Lax-Friedrichs scheme [74], or a less dissipative Osher numerical flux function [36]. At the discrete level, recalling the adopted discretization (20) for $\boldsymbol{u}_n(\boldsymbol{x})$ for any $n$, we get for each control volume $C_K$

$$\sum_{i=1}^{I} \int_K \varphi_i(\boldsymbol{x})\varphi_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x}(\boldsymbol{c}_i^{n+1} - \boldsymbol{c}_i^n)$$
$$+ \int_{t_n}^{t_{n+1}} \int_{\partial K} \varphi_j(\boldsymbol{x})\widehat{\boldsymbol{F}}(\boldsymbol{u}_h|_K(\boldsymbol{x}, t), \boldsymbol{u}_h|_{K^+}(\boldsymbol{x}, t)) \cdot \boldsymbol{v}(\boldsymbol{x})\mathrm{d}\boldsymbol{\sigma}\mathrm{d}t \tag{39}$$
$$- \int_{C_K} \boldsymbol{F}(\boldsymbol{u}_h(\boldsymbol{x}, t)) \cdot \nabla_{\boldsymbol{x}}\varphi_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t - \int_{C_K} \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}_h(\boldsymbol{x}, t))\varphi_j(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t = \boldsymbol{0}$$

for every $j = 1, \cdots, I$, with $K^+$ being the neighboring cell of $K$ sharing $\partial K$ at a certain point $\boldsymbol{x}$. Again, we remark that this step has a global character due to the computation of the numerical fluxes, but it is explicit as $\boldsymbol{u}_h$ has been obtained in the predictor step. Let us notice that the linear systems involved in the corrector are local and even smaller than the predictor ones, thus readily invertible.

By solving the linear system (39) with respect to the coefficients $\boldsymbol{c}_i^{n+1}$ in each element $K$, we get the final solution $\boldsymbol{u}_{n+1}(\boldsymbol{x}) = \sum_{i=1}^{I} \boldsymbol{c}_i^{n+1}\varphi_i(\boldsymbol{x})$ for any $\boldsymbol{x} \in K$ which is an $(M + 1)$-th-order accurate approximation of $\boldsymbol{u}(\boldsymbol{x}, t_{n+1})$.

### 3.2 ADER-DG as DeC

It is possible to interpret the ADER-DG predictor step as a DeC procedure. We set $\Delta := \Delta t$ and, from the local space-time nonlinear system (24), we define the high-order nonlinear operator $\mathcal{L}_\Delta^2 : \mathbb{R}^{L \times Q} \to \mathbb{R}^{L \times Q}$ as

$$\mathcal{L}_\Delta^2(\underline{\boldsymbol{u}}) := \underline{\boldsymbol{u}} - B^{-1}\left[\underline{\boldsymbol{r}} - \widetilde{\underline{\boldsymbol{\phi}}}(\underline{\boldsymbol{u}})\right]. \tag{40}$$

Since solving the operator $\mathcal{L}_\Delta^2$ is equivalent to solve the system (24), we have already discussed the $(M + 1)$th order of accuracy of its solution.

The low-order operator $\mathcal{L}_\Delta^1 : \mathbb{R}^{L \times Q} \to \mathbb{R}^{L \times Q}$ is, instead, defined as

$$\mathcal{L}^1_\Delta(\underline{u}) := \underline{u} - B^{-1}\left[\underline{r} - \widetilde{\underline{\phi}}(\underline{u}_0)\right], \tag{41}$$

with $\underline{u}_0$ being a vector of local space-time representation coefficients with respect to the basis $\{\vartheta^\ell(\boldsymbol{x},t)\}_{\ell=1,\cdots,L}$, yielding an $O(\Delta t)$-approximation of the analytical solution in $C_K$. As an example, the vector $\underline{u}_0$ can be chosen, such that $\sum_{\ell=1}^{L} \boldsymbol{u}_0^\ell \vartheta^\ell(\boldsymbol{x},t) = \boldsymbol{u}_n(\boldsymbol{x})$ for all $t \in [t_n, t_{n+1}]$. In practice, this definition is merely formal as the related terms will cancel out in the iteration and in all the needed proofs. It can be shown that the local reconstruction of the PDE solution induced by the coefficients obtained by solving $\mathcal{L}^1_\Delta$ is first order accurate with respect to the analytical solution. Furthermore, let us observe how the problem $\mathcal{L}^1_\Delta(\widetilde{\underline{u}}) = \underline{z}$ for some given $\underline{z} \in \mathbb{R}^{L\times Q}$ can be easily solved by explicitly isolating $\widetilde{\underline{u}}$.

In the following, we will prove that the operators that we have defined respect the three properties needed to apply Theorem 1, but first let us characterize the related DeC iterative procedure.

### 3.2.1 Iterative ADER-DG-DeC Procedure

If we characterize the iterative procedure (5) in the ADER context with the operators (41) and (40), by a direct computation, we get

$$\begin{aligned}
\underline{u}^{(p)} - B^{-1}\left[\underline{r} - \widetilde{\underline{\phi}}(\underline{u}_0)\right] = {}&\underline{u}^{(p-1)} - B^{-1}\left[\underline{r} - \widetilde{\underline{\phi}}(\underline{u}_0)\right] \\
&- \underline{u}^{(p-1)} + B^{-1}\left[\underline{r} - \widetilde{\underline{\phi}}(\underline{u}^{(p-1)})\right],
\end{aligned} \tag{42}$$

which reduces to

$$\underline{u}^{(p)} = B^{-1}\left[\underline{r} - \widetilde{\underline{\phi}}(\underline{u}^{(p-1)})\right]. \tag{43}$$

This is nothing but the fixed-point iteration (26). The advantage of having put it into a DeC formulation is that, in this context, we have at our disposal an estimate for the accuracy of $\underline{u}^{(p)}$ obtained at the generic iteration $p$ given by (6). In particular, according to Remark 1, if $\underline{u}^{(0)}$ yields an $O(\Delta t)$-approximation of the analytical solution, we have that the optimal number of iterations to achieve the formal accuracy is given by $P = M + 1$. A natural choice of the initial vector is thus $\underline{u}^{(0)} := \underline{u}_0$.

### 3.2.2 Proof of the Properties of the Operators $\mathcal{L}^1_\Delta, \mathcal{L}^2_\Delta$

We have that the operators $\mathcal{L}^1_\Delta, \mathcal{L}^2_\Delta$ fulfill the hypotheses of Theorem 1 as stated in the next theorem.

**Theorem 3** (ADER-DG is DeC) *The operators* $\mathcal{L}^1_\Delta, \mathcal{L}^2_\Delta : \mathbb{R}^{L\times Q} \to \mathbb{R}^{L\times Q}$, *defined, respectively, in* (41) *and* (40), *fulfill the three hypotheses of Theorem* 1.

**Proof**

(i) **Existence of a unique solution to $\mathcal{L}_\Delta^2$**

This property has been already proved in Proposition 1, since solving the operator $\mathcal{L}_\Delta^2$ is equivalent to solve the nonlinear system (24).

(ii) **Coercivity-like property of $\mathcal{L}_\Delta^1$**

We consider the infinity norm over $\mathbb{R}^{L\times Q}$ and two general vectors $\underline{v}, \underline{w} \in \mathbb{R}^{L\times Q}$. The proof of this property is immediate, because, by a direct computation, we have

$$\left\| \mathcal{L}_\Delta^1(\underline{v}) - \mathcal{L}_\Delta^1(\underline{w}) \right\|_\infty = \left\| \underline{v} - \underline{w} \right\|_\infty, \tag{44}$$

and, thus, (3) holds with $\alpha_1 = 1$.

(iii) **Lipschitz-continuity-like property of $\mathcal{L}_\Delta^1 - \mathcal{L}_\Delta^2$**

The proof of this property is based on Corollary 1. A direct computation leads to the thesis

$$\left\| \left[ \mathcal{L}_\Delta^1(\underline{v}) - \mathcal{L}_\Delta^2(\underline{v}) \right] - \left[ \mathcal{L}_\Delta^1(\underline{w}) - \mathcal{L}_\Delta^2(\underline{w}) \right] \right\|_\infty \tag{45}$$

$$= \left\| B^{-1} \left[ \widetilde{\underline{\phi}}(\underline{v}) - \widetilde{\underline{\phi}}(\underline{w}) \right] \right\|_\infty \leqslant \Delta t \widetilde{C}_{\text{Lip}} \left\| \underline{v} - \underline{w} \right\|_\infty, \tag{46}$$

where in (46), we applied Corollary 1.

### 3.3 ADER-$\mathbb{P}_N\mathbb{P}_M$ and ADER-FV

Other formulations of ADER are available in the literature, in particular ADER-$\mathbb{P}_N\mathbb{P}_M$ [13, 33, 37, 43] is a generalization of the ADER-DG formulation. The ADER-$\mathbb{P}_N\mathbb{P}_M$ method is based on adopting, for the discretization $\boldsymbol{u}_n(\boldsymbol{x})$ of the solution at the time $t_n$, different local basis functions $\{\lambda_r\}_{r=1,\cdots,R}$ spanning a space of discontinuous piecewise polynomial functions of degree $N \leqslant M$, i.e., $V_N$ with $V_N := \{g \in L^2(\Omega) \text{ s.t. } g|_K \in \mathbb{P}_N(K)\}$, yielding the reconstruction

$$\boldsymbol{u}_n(\boldsymbol{x}) := \sum_{r=1}^R \boldsymbol{c}_r^n \lambda_r(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in K. \tag{47}$$

Then, the scheme is formally identical to the one described before. In the predictor (23), the same $M$th degree local spatial bases $\{\varphi_i(\boldsymbol{x})\}_{i=1,\cdots,I}$ and temporal bases $\{\psi^m(t)\}_{m=0,\cdots,M}$ are considered, yielding a local reconstruction $\boldsymbol{u}_h(\boldsymbol{x}, t)$ in each $C_K$ guaranteeing $(M+1)$th order of accuracy. The corrector is also identical to the one previously described (39), up to the replacement of the basis functions $\varphi_i$ with the basis functions $\lambda_r$.

The only difference with respect to the original formulation is given by the fact that, if $N < M$, a suitable $M$th-degree polynomial reconstruction $\widetilde{\boldsymbol{u}}_n(\boldsymbol{x})$ has to be considered in place of $\boldsymbol{u}_n(\boldsymbol{x})$ for the computation of the related integral over $K$ in the predictor (23) to guarantee $(M+1)$th order of accuracy. Usually, $\widetilde{\boldsymbol{u}}_n(\boldsymbol{x})$ is retrieved via a WENO or CWENO reconstruction [13, 37, 43].

Let us observe that if $N = M$ and the basis $\{\lambda_r(\boldsymbol{x})\}_{r=1,\cdots,R}$ coincides with $\{\varphi_i(\boldsymbol{x})\}_{i=1,\cdots,I}$, then the ADER-$\mathbb{P}_N\mathbb{P}_M$ scheme reduces exactly to the ADER-DG previously introduced.

On the other hand, the scheme obtained for $N = 0$, i.e., with a piecewise constant approximation of $\boldsymbol{u}_n(\boldsymbol{x})$ over $\overline{\Omega}$, is the ADER-FV scheme. One can observe that the corrector, in such a case, corresponds to an explicit $(M + 1)$th-order accurate FV step. All the schemes obtained for $0 < N < M$ are alternatives which vary between these two schemes.

Finally, let us notice that the predictor of such methods, being formally unchanged with respect to the original formulation, can also be seen as a DeC method in which one order is achieved at each iteration until $M + 1$.

# 4 New Efficient ADER Schemes

In this section, we will explain how to apply the novel modification to the described ADER framework. We will first introduce the efficient ADER-DG-u, obtained by simply matching the order of the space-time reconstruction in each predictor iteration with the order of accuracy achieved in the same iteration, without spoiling the original order of accuracy. Afterward, we will explain how such $p$-adaptivity strategy can be exploited to prescribe structure preservation by introducing the DOOM approach. We will focus on the ADER-DG scheme, bearing in mind that the same modifications hold true for ADER-FV and ADER-$\mathbb{P}_N\mathbb{P}_M$ schemes as well.

## 4.1 Modification of ADER-DG (ADER-DG-u)

We propose to change the predictor of ADER-DG by increasing the polynomial degree of the reconstruction of the numerical solution at each iteration $p$ according to the order of accuracy achieved in that specific iteration. In particular, we define for any $p$ the general local basis $\{\vartheta^{\ell,(p)}(\boldsymbol{x},t)\}_{\ell=1,\cdots,L^{(p)}}$ given by the tensor product of space basis functions $\varphi_i^{(p)}(\boldsymbol{x})$ and time basis functions $\psi^{m,(p)}(t)$ of degree $p$. We also define the functional spaces generated by these bases as $X^{(p)} := \left(\mathrm{span}\{\vartheta^{\ell,(p)}(\boldsymbol{x},t)\}_{\ell=1,\cdots,L^{(p)}}\right)^Q$.

**Remark 5** (On the spaces $X^{(p)}$) According to our definitions of the operators (41) and (40), formally, the spaces $X^{(p)}$ in the ADER context should be spaces of coefficients of the discrete solution. However, by definition, such spaces are in bijection with the functional spaces of $Q$-dimensional polynomials whose scalar components are spanned by the bases $\{\vartheta^{\ell,(p)}(\boldsymbol{x},t)\}_{\ell=1,\cdots,L^{(p)}}$. Since, in this context, referring to the polynomial degree of the numerical solution in each step of the process provides a clearer overview of the method, as an abuse of notation, we denote directly $X^{(p)}$ as the functional space associated to the corresponding coefficients, bearing in mind the aforementioned bijection.

Then, the main procedure at the iteration $p$ passes from the space-time representation coefficients $\underline{\boldsymbol{u}}^{(p-1)}$ with respect to the basis of $X^{(p-1)}$, $(p - 1)$th-order accurate with respect to the analytical solution, to $\underline{\boldsymbol{u}}^{(p)}$ in $X^{(p)}$ with accuracy $p$. To perform this step, we first use an embedding $\mathcal{E}^{(p-1)} : X^{(p-1)} \to X^{(p)}$, for example, an interpolation or an $L^2$-projection, to pass to $\underline{\boldsymbol{u}}^{*(p-1)} = \mathcal{E}^{(p-1)}(\underline{\boldsymbol{u}}^{(p-1)}) \in X^{(p)}$. This embedding should not spoil the accuracy of the reconstructed solution.

At this point, a simple iteration of the standard method (43), with structures (25) associated with the basis $\{\vartheta^{\ell,(p)}(\boldsymbol{x},t)\}_{\ell=1,\cdots,L^{(p)}}$, results in $\underline{\boldsymbol{u}}^{(p)}$ and the related $p$th-order accurate reconstruction in $C_K$. These structures read as

$$\begin{cases} \mathcal{L}_\Delta^{2,(p)}(\underline{\pmb u}) := \underline{\pmb u} - \left(B^{(p)}\right)^{-1}\left[\underline{\pmb r}^{(p)} - \underline{\widetilde{\pmb \phi}}^{(p)}(\underline{\pmb u})\right], \\ \mathcal{L}_\Delta^{1,(p)}(\underline{\pmb u}) := \underline{\pmb u} - \left(B^{(p)}\right)^{-1}\left[\underline{\pmb r}^{(p)} - \underline{\widetilde{\pmb \phi}}^{(p)}(\underline{\pmb u}_0^{(p)})\right] \end{cases} \tag{48}$$

with

$$\begin{cases} B_{j,\ell}^{(p)} := \int_K \vartheta^{\ell,(p)}(\pmb x, t_{n+1})\vartheta^{j,(p)}(\pmb x, t_{n+1})\mathrm{d}\pmb x - \int_{C_K} \vartheta^{\ell,(p)}(\pmb x, t)\frac{\partial}{\partial t}\vartheta^{j,(p)}(\pmb x, t)\mathrm{d}\pmb x\mathrm{d}t, \\ \underline{\pmb u} := \begin{pmatrix} \pmb u^1 \\ \vdots \\ \pmb u^{L^{(p)}} \end{pmatrix}, \quad \underline{\pmb r}^{(p)} := \begin{pmatrix} \int_K \pmb u_n(\pmb x)\vartheta^{1,(p)}(\pmb x, t_n)\mathrm{d}\pmb x \\ \vdots \\ \int_K \pmb u_n(\pmb x)\vartheta^{L^{(p)},(p)}(\pmb x, t_n)\mathrm{d}\pmb x \end{pmatrix}, \\ \underline{\widetilde{\pmb \phi}}^{(p)}(\underline{\pmb u}) := \begin{pmatrix} \int_{t_n}^{t_{n+1}} \pmb \phi_1^{(p)}(\underline{\pmb u}, t)\mathrm{d}t \\ \vdots \\ \int_{t_n}^{t_{n+1}} \pmb \phi_{L^{(p)}}^{(p)}(\underline{\pmb u}, t)\mathrm{d}t \end{pmatrix}, \end{cases} \tag{49}$$

where $\pmb \phi_j^{(p)}(\underline{\pmb u}, t) := \int_K E(\pmb u_h(\pmb x, t), \pmb x)\vartheta^{j,(p)}(\pmb x, t)\mathrm{d}\pmb x$, with $\pmb u_h(\pmb x, t) = \sum_{\ell=1}^{L^{(p)}} \pmb u^\ell \vartheta^{\ell,(p)}(\pmb x, t)$ for any

$(\pmb x, t) \in C_K$ and $\underline{\pmb u}_0^{(p)}$ some local coefficients extrapolated from the initial datum $\pmb u_n(\pmb x)$ in $K$, yielding an $O(\Delta t)$-approximation of the analytical solution to our PDE in $C_K$. Again, the definition of $\underline{\pmb u}_0^{(p)}$ is merely formal as it cancels in the iterations.

The resulting modified ADER-DG-u iterative procedure reads

$$\begin{cases} \underline{\pmb u}^{(0)} = \underline{\pmb u}_0^{(0)}, \\ \begin{cases} \underline{\pmb u}^{*(p-1)} = \mathcal{E}^{(p-1)}(\underline{\pmb u}^{(p-1)}), \\ \underline{\pmb u}^{(p)} = \left(B^{(p)}\right)^{-1}\left[\underline{\pmb r}^{(p)} - \underline{\widetilde{\pmb \phi}}^{(p)}(\underline{\pmb u}^{*(p-1)})\right], \end{cases} p \geqslant 1. \end{cases} \tag{50}$$

We remark that the superscript $(p)$ in the definition of the structures in (48) and (49) is simply referred to the iteration, the modified method is still fully explicit, and all the terms at the right-hand side of the iteration formula (50) can be explicitly computed.

The accuracy evolves as follows throughout the procedure. We start with $\underline{\pmb u}^{(0)}$ associated with a piecewise constant $O(\Delta t)$-approximation of the solution to the PDE in $C_K$ and we perform the embedding in $X^{(1)}$ to get $\underline{\pmb u}^{*(0)}$, still $O(\Delta t)$-accurate. Performing the first iteration via $\mathcal{L}_\Delta^{1,(1)}, \mathcal{L}_\Delta^{2,(1)}$, we get $\underline{\pmb u}^{(1)}$ yielding an $O(\Delta t^2)$-approximation of the solution in $C_K$. We continue iteratively with $\underline{\pmb u}^{(p-1)}$ associated with a $(p-1)$th-order accurate approximation of the solution in $X^{(p-1)}$ spanned by polynomial bases of degree $p-1$, that is embedded in $X^{(p)}$ obtaining $\underline{\pmb u}^{*(p-1)}$ with the same accuracy $p-1$. This allows to compute $\underline{\pmb u}^{(p)}$ via a DeC iteration with $\mathcal{L}_\Delta^{1,(p)}, \mathcal{L}_\Delta^{2,(p)}$ achieving $p$th order of accuracy.

**Table 1** Increasing degrees of polynomial spaces $X^{(p)}$ varying the iteration

| Space | $X^{(0)}$ | $X^{(1)}$ | $X^{(2)}$ | $X^{(3)}$ | $\cdots$ | $X^{(M-1)}$ | $X^{(M)}$ | $X^{(M+1)}$ |
|---|---|---|---|---|---|---|---|---|
| Polynomial degree | 0 | 1 | 2 | 3 | $\cdots$ | $M-1$ | $M$ | $M$ |

**Remark 6** (On the accuracy of the interpolation) One must notice that the discretization in the space $X^{(p)}$, corresponding to the tensor product of polynomials of degree $p$ in space and in time, allows in general a maximal order of accuracy $p + 1$ with respect to the analytical solution of the PDE, corresponding to an error $O(\Delta t^{p+2})$. On the other hand, the embedding $\mathcal{E}^{(p)} : X^{(p)} \to X^{(p+1)}$ can be simply realized by interpolation, i.e., by evaluating the reconstruction associated with $\underline{u}^{(p)} \in X^{(p)}$ in the nodal values defining the $X^{(p+1)}$ basis functions, when nodal bases are employed, or by an $L^2$-projection, when more general bases are considered. In both cases, this operation can preserve at most the accuracy of order $p$, hence introducing an error of $O(\Delta t^{p+1})$. Therefore, the embedding must be performed before saturating the accuracy associated with the current polynomial basis to avoid the consequent degradation of the order.

Due to the previous remark, if the final polynomial degree of the bases in space and in time is fixed to $M$, it is convenient to perform $M$ iterations in the form (50) to get $\underline{u}^{(M)}$, associated with the desired final discretization, plus a final iteration in the same space $X^{(M+1)} = X^{(M)}$ with the same structures as the ones used in the $M$th iteration to saturate the accuracy related to such discretization getting thus $\underline{u}^{(M+1)}$ yielding $(M + 1)$th order of accuracy.

The degrees of the bases of the spaces $X^{(p)}$, assuming a fixed final polynomial degree equal to $M$, are summarized in Table 1 and the procedure is displayed in the following sketch:

$$\underline{u}^{(0)} \xrightarrow{\mathcal{E}^{(0)}} \underline{u}^{*(0)} \xrightarrow[\mathcal{L}_\Delta^{1,(1)}]{\mathcal{L}_\Delta^{2,(1)}} \underline{u}^{(1)} \xrightarrow{\mathcal{E}^{(1)}} \underline{u}^{*(1)} \xrightarrow[\mathcal{L}_\Delta^{1,(2)}]{\mathcal{L}_\Delta^{2,(2)}} \underline{u}^{(2)} \xrightarrow{\mathcal{E}^{(2)}} \cdots \xrightarrow[\mathcal{L}_\Delta^{1,(M)}]{\mathcal{L}_\Delta^{2,(M)}} \underline{u}^{(M)} \xrightarrow[\mathcal{L}_\Delta^{1,(M)}]{\mathcal{L}_\Delta^{2,(M)}} \underline{u}^{(M+1)}.$$

$$O(\Delta t) \qquad O(\Delta t) O(\Delta t^2) \qquad O(\Delta t^2) O(\Delta t^3) \qquad\qquad O(\Delta t^{M+1}) O(\Delta t^{M+2})$$

Finally, always assuming a final polynomial degree equal to $M$ in the predictor, the corrector step (39) is normally performed with the $(M + 1)$th-order accurate discretization given by the $M$th-degree local polynomial basis functions $\varphi_i^{(M)}$ used in the two last predictor iterations.

This leads, in each element $K$, to the local approximation $\boldsymbol{u}_{n+1}(\boldsymbol{x}) = \sum_{i=1}^{I} c_i^{n+1} \varphi_i^{(M)}(\boldsymbol{x})$ which is

$(M + 1)$th-order accurate. The computational advantage of the modified method with respect to the original formulation is clear: all the iterations but the last two are performed with matrix and vector structures which are smaller, implying the solution of smaller systems. Also, the space-time discretization of $\boldsymbol{E}(\boldsymbol{u}_h, \boldsymbol{x}) = \text{div}_{\boldsymbol{x}} \boldsymbol{F}(\boldsymbol{u}_h(\boldsymbol{x}, t)) - \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}_h(\boldsymbol{x}, t))$ and the orders of the quadrature formulas used in the low-order iterations can be suitably chosen to decrease the related computational cost. The only extra cost can come from the embedding between the spaces, which, for interpolations, can be easily recast as products by precomputable interpolation matrices, which can, therefore, be efficiently performed.

In this work, we assume modal bases in space and time. This further increases the computational advantage as, in such context, the higher order mode is easily introduced by adding zero components to $\underline{u}^{(p)}$, thus getting $\underline{u}^{*(p)} = (\underline{u}^{(p)}, \boldsymbol{0})^{\mathrm{T}}$ without any other effort.

We denote this scheme by ADER-DG-u, referring to the $\alpha$DeCu schemes introduced in [61] with a similar technique, where u denotes the quantity that has been embedded.

**Remark 7** (Galerkin projection) In the specific context of these new modified ADER-DG methods, the embedding procedure between the spaces $X^{(p-1)}$ and $X^{(p)}$ could be replaced by a Galerkin projection onto $X^{(p)}$. Namely, in (50), one could skip the interpolation procedure

and directly consider $\widetilde{\underline{\boldsymbol{\phi}}}^{(p)}(\underline{\boldsymbol{u}}^{(p-1)})$ which is defined in each $j$th component by the integral over $[t_n, t_{n+1}]$ of

$$\boldsymbol{\phi}_j^{(p)}(\underline{\boldsymbol{u}}^{(p-1)}, t) := \int_K \boldsymbol{E}(\boldsymbol{u}_h^{(p-1)}(\boldsymbol{x}, t), \boldsymbol{x}) \vartheta^{j,(p)}(\boldsymbol{x}, t) \mathrm{d}\boldsymbol{x} \tag{51}$$

with $\boldsymbol{u}_h^{(p-1)}(\boldsymbol{x}, t) = \displaystyle\sum_{\ell=1}^{L^{(p-1)}} \boldsymbol{u}^{\ell,(p-1)} \vartheta^{\ell,(p-1)}(\boldsymbol{x}, t)$. This mismatch between the spaces of the explicit term and of the test functions permits the evolution to the next space $X^{(p)}$. This is particularly convenient as there would be no interpolation, whose cost is, however, negligible with respect to the rest of the scheme. For modal bases, the two approaches are equivalent.

**Remark 8** (Space-time accuracy) Since the ADER schemes are one-step fully discrete predictor-corrector methods on space-time control volumes $C_K$, the order of accuracy in space and time is simultaneously evolved; thus, if the iterative solution $\underline{\boldsymbol{u}}^{(p)}$ is of order $O(\Delta t^{p+1})$ in time, it is also accurate $O(h^{p+1})$ in space, assuming a suitable CFL condition linking $\Delta t$ and $h$. This is omitted to lighten the notation.

As already remarked, since the predictor of the ADER-$\mathbb{P}_N\mathbb{P}_M$ is identical to the one of the standard ADER-DG, we can analogously introduce the ADER-$\mathbb{P}_N\mathbb{P}_M$ -u and, as a particular case, the ADER-FV-u methods in a straightforward way.

**Remark 9** (On the memory and computational differences of ADER-DG-u) The implementation of ADER-DG and ADER-DG-u can be performed in various ways. The evolution structures of the predictor might be either pre-computed at the beginning of the simulation or on-the-fly at each time iteration (this is necessary for Lagrangian codes on moving meshes). We make use of polygonal meshes, but on triangular meshes, all operators can be pre-computed with minimal storage on a reference element.

For ADER-DG-u with respect to ADER-DG, there might be an increase in cost and storage for the different iterative structures, e.g., $B^{(p)}$, only if nodal basis functions are used. In the considered case, where modal polynomial basis functions are adopted, the iterative structures are simply constituted of slices of the highest order structures, that would be anyway computed for ADER-DG. Hence, there is no extra computation nor memory storage to be considered. In the modal case, these costs can be reduced by a pre-computation of all the operators and, in case of triangular meshes, with computations only on a reference element.

## 4.2 DOOM Limiter Based on Adaptivity

In the context of the novel schemes, it is very natural to introduce a limiter that guarantees some structural properties of the solution. The limiter will be denoted by Discrete Optimally increasing Order Method (DOOM), as it will stop the iteration process in the predictor at an optimal value.

We consider the ADER-FV-u scheme, to inherit the robustness of FV formulations and the far less-restrictive CFL constraints which are suitable for large-scale

simulations, and we introduce an adaptive criterion. We fix a final number of iterations $P = M + 1$, corresponding to $(M + 1)$th order of accuracy, and we perform the local predictor iterations as prescribed in the context of the ADER-FV-u scheme, but, in contrast with the standard method, we check for the non-violation of some physical constraints (for example the positivity of density and pressure in hydrodynamics) of the computed solution. If at iteration $p$, with $1 \leqslant p \leqslant M + 1$, the computed $\underline{\boldsymbol{u}}^{(p)}$ does not fulfill some of the mentioned constraints, the solution is rejected and $\underline{\boldsymbol{u}}^{(p-1)}$ is assumed to be the output of the iterative procedure for the correction step. Let us notice that, in the worst case, considering $\underline{\boldsymbol{u}}^{(0)}$ in the correction step in a given region of $\Omega$ leads to a standard first-order Godunov scheme which is, indeed, reliable. A sketch of the limiter is displayed in Algorithm 1, in which $\overline{\boldsymbol{u}}_n$ represents the local constant value of $\boldsymbol{u}_n(\boldsymbol{x})$ in a cell $K$ in the ADER-FV-u (and ADER-FV) context.

---

**Algorithm 1** DOOM limiter for ADER-FV-u on a cell $K$

---

**Require:** $\overline{\boldsymbol{u}}_n$

  $\underline{\boldsymbol{u}}^{(0)} = \overline{\boldsymbol{u}}_n$

  **for** $p = 1, \cdots, M + 1$ **do**

    $\underline{\boldsymbol{u}}^{*(p-1)} = \mathcal{E}^{(p-1)}(\underline{\boldsymbol{u}}^{(p-1)})$

    $\underline{\boldsymbol{u}}^{(p)} = \left(B^{(p)}\right)^{-1} \left[\underline{\boldsymbol{r}}^{(p)} - \widetilde{\underline{\boldsymbol{\phi}}}^{(p)}(\underline{\boldsymbol{u}}^{*(p-1)})\right]$

    **if** $\underline{\boldsymbol{u}}^{(p)}$ does not meet the criteria **then**

      `return` $\underline{\boldsymbol{u}}^{(p-1)}$

    **end if**

  **end for**

  `return` $\underline{\boldsymbol{u}}^{(M+1)}$

---

The strategy may remind the *a posteriori* MOOD technique [9, 18, 29, 31] with some fundamental differences. The low-order acceptable solution $\underline{\boldsymbol{u}}^{(p-1)}$ has been computed before $\underline{\boldsymbol{u}}^{(p)}$, as it was a necessary step toward the increase of an order of accuracy. Moreover, the order of accuracy is automatically pushed as much as possible without violating the physical constraints: in fact, $\underline{\boldsymbol{u}}^{(p)}$, possibly rejected, is computed if and only if $\underline{\boldsymbol{u}}^{(p-1)}$ was reliable. This avoids the risk of an over-diffusion in having the safe low-order scheme guaranteeing an accuracy lower than the one actually achievable. Therefore, it is then possible to preserve some physical properties through this procedure as explained in the following proposition.

**Proposition 2** (ADER-FV-u with DOOM property) *Suppose that the FV scheme preserves a property $\mathscr{P}$. Suppose that the property $\mathscr{P}$ is checked in the DOOM admissibility criteria. Then, the ADER-FV-u with the DOOM limiter preserves the property $\mathscr{P}$.*

**Proof** In the ADER-FV-u case, $\boldsymbol{u}_n$ is locally represented through a single constant basis function $\lambda_1 \equiv 1$. By induction on the time step $n$, suppose that the value of $\boldsymbol{u}_n$ in each cell $K$, i.e., $\boldsymbol{c}_1^n := \overline{\boldsymbol{u}}_n$ of (47), verifies the property $\mathscr{P}$. Then, independently of the WENO reconstruction of $\boldsymbol{u}_n$ used to obtain the polynomial $\widetilde{\boldsymbol{u}}_n(\boldsymbol{x})$ for computing the related integrals in (49) with the desired accuracy, $\underline{\boldsymbol{u}}^{(0)} = \boldsymbol{c}_1^n$ still is the local value of the original $\boldsymbol{u}_n$ and fulfills the property $\mathscr{P}$. Then, during the ADER-FV-u DOOM procedure, $\underline{\boldsymbol{u}}^{(p)}$ are kept in the

iterations only if property $\mathscr{P}$ is fulfilled. Hence, property $\mathscr{P}$ holds for the final predictor $\boldsymbol{u}_h$ and, in the corrector step (39), we perform the FV method using $\boldsymbol{u}_h$. Therefore, $\boldsymbol{u}_{n+1}$ still fulfills property $\mathscr{P}$.

At the moment, this procedure does not guarantee to preserve the property for ADER-$\mathbb{P}_N\mathbb{P}_M$ with $N > 0$, due to the fact that the corrector in such a case is not an explicit FV step. However, the authors are working on new structure-preserving strategies for such schemes. Moreover, in the simulations of this work, only positivity of density and pressure is checked with the DOOM limiter, but other properties like discrete local maximum principle or entropy inequalities [4, 49, 56, 57] can be ensured.

**Remark 10** (Other applications of the *p*-adaptivity) The adaptive nature of the novel methods can be exploited also for other applications. In particular, the approach can approximate the exact solution with arbitrary precision as $p \to +\infty$ and it is not constrained to a maximum degree $M$ and the related approximation accuracy. Within this framework, it is easy to design efficient arbitrary high-order adaptive schemes, as in [61] in a DeC context for ODEs, choosing the stopping criterion for the iterations in accordance with the iteration error. Also the *hp*-adaptivity can be introduced in this framework. As soon as the DOOM limiter requires low-order steps, it is possible to locally use the *h*-adaptivity to recover for the lost accuracy. These applications are already object of study of the authors, but they will not be treated in this work.

## 5 Numerical Results

In this section, we will report the numerical results of several tests performed to validate the accuracy and the robustness of the novel ADER-DG-u and ADER-FV-u methods, i.e., ADER-$\mathbb{P}_N\mathbb{P}_M$ -u, respectively, with $N = M$ and $N = 0$. To quantify the obtained speedup in terms of computational time, they will be compared with the state-of-the-art ADER-DG and ADER-FV methods [16], characterized by a fixed polynomial degree along the whole iterative procedure and a convergence criterion $\left\| \underline{\boldsymbol{u}}^{(p)} - \underline{\boldsymbol{u}}^{(p-1)} \right\|_\infty < \texttt{tol}$ to stop the predictor iterations (26), where here we assume $\texttt{tol} = 10^{-12}$.

We adopt the following notation: for each method, we explicitly specify the formal order of accuracy. Therefore, ADER-DG($M + 1$) and ADER-FV($M + 1$) represent the original methods with predictor spatial and temporal basis functions of degree $M$ guaranteeing $(M + 1)$th order of accuracy. In the context of ADER-DG-u($M + 1$) and ADER-FV-u($M + 1$), instead, $M$ is the final degree of the predictor spatial and temporal basis functions at the end of the iteration process still leading to accuracy $M + 1$.

We will focus on the Euler and compressible Navier-Stokes equations. The Euler equations are a system of hyperbolic PDEs in the form (18) given by

$$\boldsymbol{u} = \begin{pmatrix} \rho \\ \boldsymbol{q} \\ E \end{pmatrix}, \quad \boldsymbol{F}(\boldsymbol{u}) = \begin{pmatrix} \boldsymbol{q} \\ \rho \boldsymbol{v} \otimes \boldsymbol{v} + p\mathbb{I} \\ \boldsymbol{v}(E + p) \end{pmatrix}, \quad \boldsymbol{S}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{0}, \tag{52}$$

where $\rho$ is the density, $\boldsymbol{q} \in \mathbb{R}^D$ the momentum, $E$ the energy, $p$ the pressure, $\boldsymbol{v} = \frac{\boldsymbol{q}}{\rho} \in \mathbb{R}^D$ the velocity of the flow, and $\mathbb{I} \in \mathbb{R}^{D \times D}$ is the identity matrix. The system is completed by specifying the closure equation of state $E = \frac{p}{\gamma-1} + \rho \frac{\|\boldsymbol{v}\|_2^2}{2}$, where $\gamma = \frac{c_p}{c_v}$ is the adiabatic coef-

ficient defined as the ratio between the specific heats at constant pressure and volume and is here assumed to be $\gamma = 1.4$.

The more general compressible Navier-Stokes equations are obtained by keeping the viscosity effects into account and, for ideal gases, are defined by

$$u = \begin{pmatrix} \rho \\ q \\ E \end{pmatrix}, \quad F(u) = \begin{pmatrix} q \\ \rho v \otimes v + \sigma(u, \nabla_x u) \\ v(E\mathbb{I} + \sigma(u, \nabla_x u)) - \kappa \nabla_x T \end{pmatrix}, \quad S(x, u) = \mathbf{0}, \tag{53}$$

where $\sigma(u, \nabla_x u)$ denotes the stress tensor, $\kappa$ is the heat conduction coefficient, and $T$ represents the temperature, while the other terms have the same meaning as in the context of the Euler equations. In particular, the stress tensor $\sigma(u, \nabla_x u)$ is given, under the Stokes hypothesis, by

$$\sigma(u, \nabla_x u) = \left( p + \frac{2}{3} \mu \operatorname{div}_x v \right) \mathbb{I} - \mu (\nabla_x v + \nabla_x v^{\mathrm{T}}), \tag{54}$$

with $p$ being the pressure of the fluid and $\mu$ the dynamic viscosity that we assume to be constant. The heat conduction coefficient $\kappa$ is linked to the viscosity coefficient through the Prandtl number $Pr$ with the following law:

$$\kappa = \frac{\mu \gamma c_v}{Pr}, \tag{55}$$

where again $\gamma = \frac{c_p}{c_v}$. A thermal and a caloric equation of state are needed for the closure of (53). For an ideal gas, those are

$$\frac{p}{\rho} = RT, \quad \frac{e}{\rho} = c_v T \tag{56}$$

with $R$ being the specific gas constant and $e = E - \rho \frac{\|v\|_2^2}{2}$ the internal energy.

We will consider two-dimensional (2D) problems, and hence, $v := (u, v)^{\mathrm{T}}$.

If not stated otherwise, the CFL number is set to CFL = 0.5, and the time step is computed according to an explicit stability condition which is given by

$$\Delta t \leqslant \text{CFL} \; \frac{\min\limits_{K \in \tau_h} h_K}{(2N + 1) \max\limits_{K \in \tau_h} \left( \max\limits_{x \in K} \|\lambda\|_\infty + 2 \max\limits_{x \in K} \|\lambda_v\|_\infty \frac{2N+1}{h_K} \right)}, \tag{57}$$

where $N$ represents the degree of the chosen polynomial representation, while $\lambda = \left( \|v\|_2 - \sqrt{\gamma \frac{p}{\rho}}, \|v\|_2, \|v\|_2 + \sqrt{\gamma \frac{p}{\rho}} \right)$ are the convective eigenvalues of the Euler system, and the viscous eigenvalues $\lambda_v$ are given in [35]. The characteristic mesh size of the cell $h_K$ is given by the square root of its surface in two dimensions. If not stated differently, the local Lax-Friedrichs numerical flux function [74] is used in the corrector step (39).

For more challenging tests, in which the density is close to zero, we will activate the DOOM limiter checking for the positivity of the density and pressure in the quadrature points and that no NaN appears in the solution. This limiter will be used only with the ADER-FV-u technique, which provably guarantees the preservation of the positivity of these quantities.
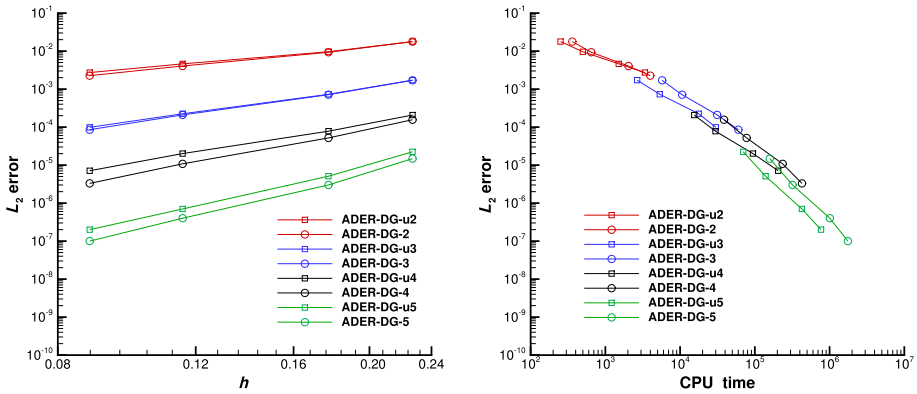
**Fig. 1** Comparison between ADER-DG-u and ADER-DG schemes from second up to fifth order of accuracy. Left: dependency of the error norm on the mesh size. Right: dependency of the error norm on the CPU time

**Table 2** Numerical convergence results for the compressible Euler equations using both ADER-DG-u and ADER-DG schemes from second up to fifth order of accuracy in space and time. The errors are measured in the $L_2$ norm and refer to the variable $\rho$ (density) at time $t_f = 1$. The absolute CPU time of each simulation is reported in seconds

| $h(\Omega)$ | ADER-DG-u | | | ADER-DG | | |
|---|---|---|---|---|---|---|
| | $\rho_{L_2}$ | $O(\rho_{L_2})$ | CPU time/s | $\rho_{L_2}$ | $O(\rho_{L_2})$ | CPU time/s |
| | Order of accuracy: $O(2)$ | | | | | |
| 2.270E−01 | 1.781E−02 | – | 2.511E+02 | 1.775E−02 | – | 3.616E+02 |
| 1.773E−01 | 9.625E−03 | 2.49 | 4.997E+02 | 9.322E−03 | 2.61 | 6.472E+02 |
| 1.155E−01 | 4.614E−03 | 1.71 | 1.509E+03 | 4.055E−03 | 1.94 | 2.039E+03 |
| 8.786E−02 | 2.723E−03 | 1.93 | 3.387E+03 | 2.262E−03 | 2.14 | 3.989E+03 |
| | Order of accuracy: $O(3)$ | | | | | |
| 2.270E−01 | 1.719E−03 | – | 2.664E+03 | 1.704E−03 | – | 5.750E+03 |
| 1.773E−01 | 7.301E−04 | 3.46 | 5.346E+03 | 7.121E−04 | 3.53 | 1.065E+04 |
| 1.155E−01 | 2.247E−04 | 2.75 | 1.773E+04 | 2.095E−04 | 2.85 | 3.133E+04 |
| 8.786E−02 | 9.871E−05 | 3.01 | 3.010E+04 | 8.542E−05 | 3.29 | 6.020E+04 |
| | Order of accuracy: $O(4)$ | | | | | |
| 2.270E−01 | 2.076E−04 | – | 1.547E+04 | 1.563E−04 | – | 3.868E+04 |
| 1.773E−01 | 7.803E−05 | 3.96 | 2.975E+04 | 5.195E−05 | 4.46 | 7.766E+04 |
| 1.155E−01 | 2.013E−05 | 3.16 | 9.427E+04 | 1.085E−05 | 3.65 | 2.354E+05 |
| 8.786E−02 | 7.139E−06 | 3.80 | 2.054E+05 | 3.332E−06 | 4.32 | 4.270E+05 |
| | Order of accuracy: $O(5)$ | | | | | |
| 2.270E−01 | 2.238E−05 | – | 6.993E+04 | 1.475E−05 | – | 3.171E+05 |
| 1.773E−01 | 5.080E−06 | 6.00 | 1.393E+05 | 3.002E−06 | 6.44 | 6.390E+05 |
| 1.155E−01 | 7.405E−07 | 4.49 | 4.261E+05 | 4.180E−07 | 4.60 | 2.015E+06 |
| 8.786E−02 | 2.154E−07 | 4.52 | 7.691E+05 | 1.228E−07 | 4.48 | 3.516E+06 |

**Fig. 2** Speedup of the ADER-DG-u schemes compared to the ADER-DG methods depending on the mesh size for different orders
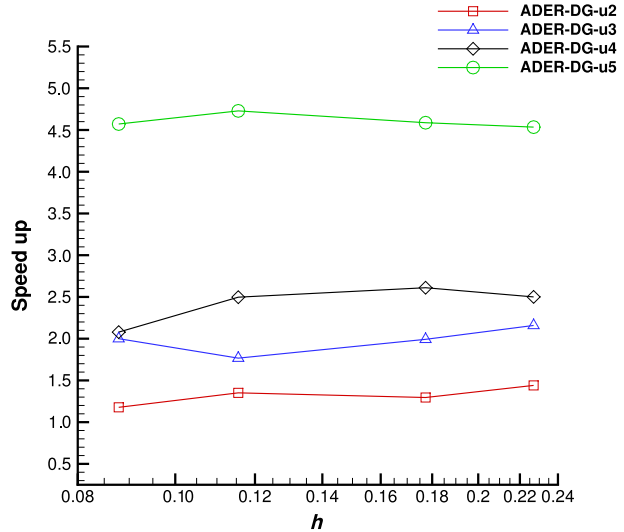


## 5.1 Numerical Convergence Studies

To test the accuracy of the method, we perform a convergence test on a smooth isentropic vortex [76] for the compressible Euler equations. The computational domain is $\Omega = [0, 10]^2$ with periodic boundary conditions, and it is tessellated by a polygonal mesh. The vortex is centered at the initial time in $\boldsymbol{x}_c = (x_c, y_c)^{\mathrm{T}} = (5, 5)^{\mathrm{T}}$ and moves with a background speed of $\boldsymbol{v}_\infty = (u_\infty, v_\infty)^{\mathrm{T}} = (1, 1)^{\mathrm{T}}$. The initial position of the vortex, in a generic point $\boldsymbol{x} = (x, y)^{\mathrm{T}}$, can be described using the radial coordinate $r := \left\| \boldsymbol{x} - \boldsymbol{x}_c \right\|_2$ as

$$
\begin{cases}
\rho(\boldsymbol{x}, 0) = (1 + \delta T)^{\frac{1}{\gamma-1}}, \\
\boldsymbol{v}(\boldsymbol{x}, 0) = \boldsymbol{v}_\infty + \frac{\epsilon}{2\pi} e^{\frac{1-r^2}{2}} \begin{pmatrix} -(y - y_c) \\ (x - x_c) \end{pmatrix}, \quad \delta T = -\frac{(\gamma - 1)\epsilon^2}{8\gamma\pi} e^{1-r^2} \\
p(\boldsymbol{x}, 0) = (1 + \delta T)^{\frac{\gamma}{\gamma-1}}
\end{cases}
\tag{58}
$$

with $T$ denoting the fluid temperature. The exact solution is obtained as $\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}(\boldsymbol{x} - \boldsymbol{v}_\infty t, 0)$. We run the simulation until the final time $t_f = 1$ using the Osher-type numerical flux function [36].

In Fig. 1, we can observe on the left the errors of the ADER-DG and ADER-DG-u methods for different mesh sizes. All the methods achieve the formal order of accuracy. As expected, the ADER-DG-u has slightly larger errors with respect to the original ADER-DG method, as the first iterations of the predictors are done with lower order accurate operators. Nevertheless, the final error is quite comparable with the ADER-DG one and, looking at the right figure, we observe that the computational time required by ADER-DG-u for such simulations is much less (for high-order methods, it is around half) than the one required by the competitor. The slight increase in error is hugely beaten by the computational advantage of the new ADER-DG-u schemes. Indeed, the Pareto front on the right figure is only composed by ADER-DG-u points. The results are quantitatively reported in Table 2. We observe that the computed orders of accuracy are very close to the expected ones. Convergence analyses

**Table 3** Initial conditions for Riemann problems

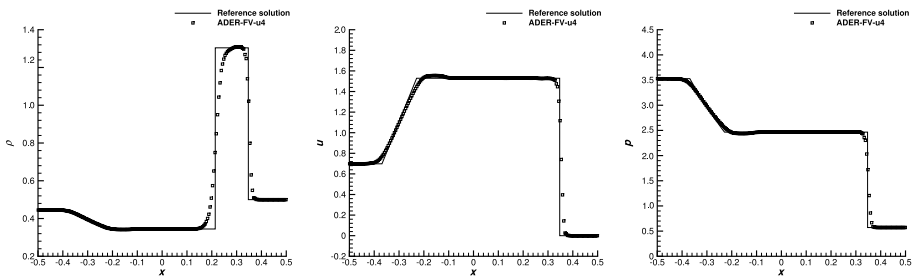| Test | $\rho_L$ | $u_L$ | $p_L$ | $\rho_R$ | $u_R$ | $p_R$ | $t_f$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.445 | 0.698 | 3.528 | 0.5 | 0 | 0.571 | 0.14 |
| 2 | 1 | 2 | 0.1 | 1 | −2 | 0.1 | 0.8 |
| 3 | 1 | −2 | 0.4 | 1 | 2 | 0.4 | 0.15 |
| 4 | 1 | 0 | 1 000 | 1 | 0 | 100 | 0.012 |



**Fig. 3** Lax shock tube problem (RP1) at the final time $t_f = 0.14$. Comparison of density, velocity, and pressure versus the reference solution for the ADER-FV-u4 scheme

with vortex-type solutions are often subjected to some loss of order of accuracy as explained in [72, 77]. In our case, this may also be due to an imprecise choice of the mesh parameter for our polygonal meshes: we consider the maximum internal diameter of the polygons, but for some meshes, this choice might not well represent the characteristic size of the cells. The same convergence trends have been observed also in [16].

Finally, Fig. 2 depicts the speedup achieved by the novel adaptive schemes compared against the classical formulation of iterative methods, namely ADER-DG-u versus ADER-DG. As the order of accuracy increases, the speedup becomes higher obtaining efficient schemes which are up to $\approx 4.5$ times faster than the classical methods. Let us notice that the formal order of accuracy is still maintained, while getting a remarkable gain in the computational efficiency.

## 5.2 Riemann Problems

In this section, we will show the results of the ADER-FV-u4 scheme, i.e., with $M = 3$, for some Riemann problems. The computational domain is the box $\Omega = [-0.5, 0.5] \times [-0.05, 0.05]$ with periodic boundary conditions in the $y$ direction and Dirichlet boundaries imposed at $x = \pm 0.5$. We use an unstructured polygonal mesh made of $N_h = 2\,226$ control volumes of characteristic mesh size of $h \approx \frac{1}{100}$. Despite the one-dimensional setting of the test case, we underline that the preservation of symmetry of the solution is not trivial on unstructured meshes, where no cell boundaries are in principle aligned with the main flow velocity. We solve again the Euler equations (52) with the initial conditions given, as a function of the $x$ coordinate only, by

$$\boldsymbol{u}(x, 0) = \begin{cases} \boldsymbol{u}_L, & \text{if } x < 0, \\ \boldsymbol{u}_R, & \text{else,} \end{cases} \tag{59}$$
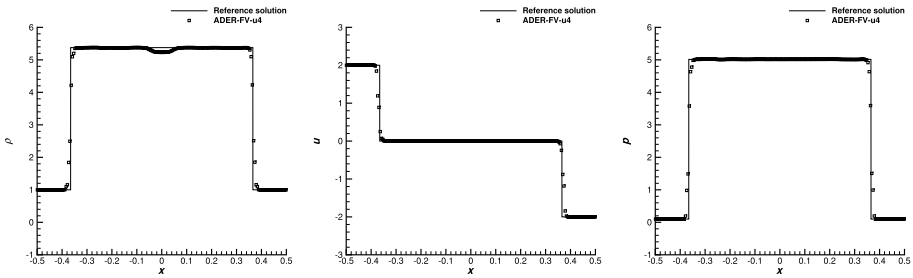
**Fig. 4** Colliding shock test (RP2) at the final time $t_f = 0.8$. Comparison of density, velocity, and pressure versus the reference solution for the ADER-FV-u4 scheme
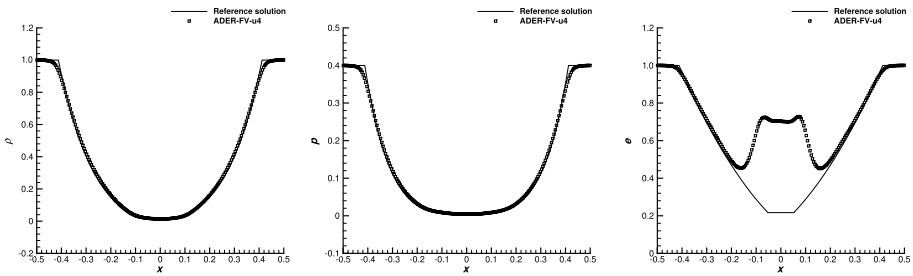


**Fig. 5** Double rarefaction test (RP3) at the final time $t_f = 0.15$. Comparison of density, pressure, and internal energy versus the reference solution for the ADER-FV-u4 scheme

where the values of $\boldsymbol{u}_L$ and $\boldsymbol{u}_R$ and the final times for the different tests are taken from [80] and they can be found in Table 3. The velocity along the $y$-direction is set to be $v = 0$ for all the tests.

The DOOM limiter is here active checking for the positivity of density and pressure and avoiding NaN. These tests are very challenging and not all the numerical methods can stably perform on them. In particular, shocks are often not well captured or numerical oscillations appear around them and it is common that negative density or pressure values appear in the simulations, making the code crash.

The first test (RP1) is the classical Lax shock tube problem. The initial discontinuity develops into a rarefaction wave, a contact discontinuity, and a shock. In Fig. 3, we observe that the ADER-FV-u does not exhibit any oscillations around the shock and that exactly catches the speed of the discontinuities.

The second test (RP2) consists of a colliding shock test. The initial discontinuity in the velocity gives rise to two shocks traveling outside the domain. This test creates a very high density and pressure region in the middle of the domain. As it can be seen in Fig. 4, the new ADER-FV-u with DOOM limiter is able to perfectly capture the shock behavior within few cells without over/under-shootings at the sides of the shocks.

The next problem (RP3) is the one presented as Test 2 in [80, Section 4.3.3]. It is a double rarefaction waves which leads to very low pressure and density areas at the center of the domain. In Fig. 5, we can appreciate the capability of the scheme of maintaining positive quantities for these variables, thanks to the DOOM limiter which, at the beginning of the simulation, ensures positivity preservation in the predictor. The mismatching of the
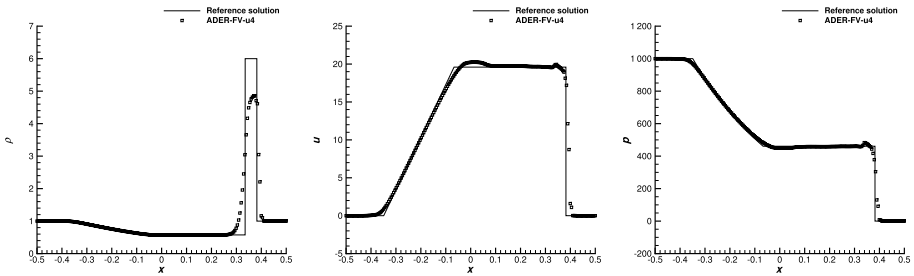
**Fig. 6** Test RP4 at the final time $t_f = 0.012$. Comparison of velocity and pressure versus the reference solution for the ADER-FV-u4 scheme

internal energy distribution is essentially due to the excessive numerical dissipation of the scheme, which could be reduced by introducing entropy preserving techniques [23, 44, 57].

The last Riemann problem (RP4) is the one presented as Test 3 in [80, Section 4.3.3]. It is a very severe test problem, and it consists of a rarefaction, a contact discontinuity, and a shock. The results obtained in Fig. 6 are in agreement with the reference solution and the smearing around the contact discontinuity is comparable to other high-order FV schemes with similar resolution.

### 5.3 Viscous Shock Profile

Now, we consider an isolated viscous shock that is traveling through a medium at rest with a shock Mach number $M_s > 1$ [11, 15, 16, 20, 33]; thus, we solve the compressible Navier-Stokes equations (53). The analytical solution and the details to compute it can be found in [11], where the stationary shock wave at a Prandtl number $Pr = 0.75$ is resolved with constant viscosity. The computational domain is $\Omega = [0, 1] \times [0, 0.2]$, which is discretized by $N_h = 1\,120$ Voronoi elements. On the left side of the domain, a constant inflow velocity is prescribed, while outflow boundary conditions are assumed at the right of the domain. Periodic boundary conditions are, instead, assigned to the top/bottom boundaries. The initial condition consists of a shock wave centered at $x = 0.25$ propagating at Mach $M_s = 2$ from left to right with a Reynolds number $Re = 100$; thus, the viscosity coefficient is set to $\mu = 2 \cdot 10^{-2}$. The upstream shock state is defined, such that the adiabatic sound speed is $c_0 = 1$. The final time of the simulation is $t_f = 0.2$ with the shock front located at $x = 0.65$.

We run the simulations with ADER-DG-u(4). Since the solution is smooth, nothing is checked along the DOOM procedure. Qualitatively, we see in Fig. 7 that there is an excellent agreement between the numerical solution and the analytical one. We underline that this test case allows all terms contained in the Navier-Stokes system to be properly checked, since advection, thermal conduction, and viscous stresses are present.

### 5.4 2D Taylor-Green Vortex

A classical test case for the incompressible Navier-Stokes equations is the Taylor-Green vortex problem. In two dimensions, the exact solution is known and it is given on the domain $\Omega = [0, 2\pi]^2$ with periodic boundary conditions by

$$\begin{cases} u(\boldsymbol{x}, t) = \sin(x)\cos(y)e^{-2vt}, \\ v(\boldsymbol{x}, t) = -\cos(x)\sin(y)e^{-2vt}, \\ p(\boldsymbol{x}, t) = C + \dfrac{1}{4}(\cos(2x)\cos(2y))e^{-4vt} \end{cases} \qquad (60)$$

with $v = \frac{\mu}{\rho}$ the kinematic viscosity and $\mu = 10^{-2}$. In this test, we also validate the quality of the scheme in a low Mach regime. Hence, the additive constant for the pressure is chosen as $C = \frac{100}{\gamma}$ and the density is set at the beginning as $\rho(\boldsymbol{x}, 0) \equiv 1$. For this test, heat conduction is neglected, i.e., $\kappa = 0$. The mesh is discretized by $N_h = 2\,916$ cells and the final time is set at $t_f = 1$. We use the ADER-DG-u method with order 4 for this simulation without checks in the DOOM procedure. The results are depicted in Fig. 8, which are compared against the analytical solutions, obtaining an excellent matching.

We also compare the numerical results with the ADER-DG scheme, and the errors are reported in Table 4 as well as the computational time. We observe that the errors are almost the same for both methods, while the novel ADER-DG-u scheme is 2.5 times faster than the classical ADER-DG.

## 5.5 Compressible Mixing Layer

Finally, we test the novel ADER-DG-u4 on the unsteady compressible mixing layer studied in [30]. The 2D computational domain is the rectangular box $\Omega = [-200, 200] \times [-50, 50]$, and a total number of $N_h = 15\,723$ polygonal Voronoi cells compose the computational mesh. The initial condition of the flow is given by two fluid layers moving with different velocities along the $x$-direction, that is

$$\begin{cases} \rho(\boldsymbol{x}, 0) = \rho_0 = 1, \\ \boldsymbol{v}(\boldsymbol{x}, 0) = \boldsymbol{v}_0 = \begin{pmatrix} \frac{1}{8}\tanh(2y) + \frac{3}{8} \\ 0 \end{pmatrix}, \\ p(\boldsymbol{x}, 0) = p_0 = \dfrac{1}{\gamma}. \end{cases} \qquad (61)$$

The free stream velocities are imposed as boundary conditions in the $y$-direction, and thus, we set $u_{+\infty} = 0.5$ and $u_{-\infty} = 0.25$ for $y \to +\infty$ and $y \to -\infty$, respectively. Along the $x$-direction, the right side is simply assigned an outflow boundary, whereas the left side is given a time-dependent inflow boundary condition with a perturbation $\delta(y, t)$

$$\begin{cases} \rho(0, y, t) = \rho_0 + 0.05\,\delta(y, t), \\ \boldsymbol{v}(0, y, t) = \boldsymbol{v}_0 + \begin{pmatrix} 1.0 \\ 0.6 \end{pmatrix}\delta(y, t), \\ p(0, y, t) = p_0 + 0.2\,\delta(y, t). \end{cases} \qquad (62)$$

The function $\delta(y, t)$ is given by

$$
\delta(y,t) = -10^{-3} \exp(-0.25y^2) \cdot \left[ \cos(\omega t) + \cos\left(\frac{1}{2}\omega t - 0.028\right) \right.
$$
$$
\left. + \cos\left(\frac{1}{4}\omega t + 0.141\right) + \cos\left(\frac{1}{8}\omega t + 0.391\right) \right]
$$
(63)

with the fundamental frequency of the mixing layer $\omega = 0.314\,787\,6$. The compressible Navier-Stokes equations are considered with the viscosity coefficient $\mu = 10^{-3}$ and no heat conduction ($\kappa = 0$). The final time is $t_f = 1\,596.8$ and the DG solution is depicted in Fig. 9 at three different output times. The vorticity of the flow field is shown, demonstrating the capability of the novel methods to capture the complex vortical structures generated by the perturbation assigned at the inflow of the channel.

## 6 Conclusions and Further Developments

To sum up, generalizing the idea proposed in [61], we have introduced a new framework for the construction of efficient $p$-adaptive arbitrary high-order methods, based on the modification of underlying arbitrary high-order iterative schemes. Specifically, the accuracy of the discretization is progressively increased with the number of iterations, gaining one order of accuracy at each iteration. Given an implementation of an iterative arbitrary high-order method, the novel technique is easy to include and it gives a remarkable advantage in terms of computational costs. Moreover, in this context, the $p$-adaptivity can be achieved very naturally inserting some criteria to stop the iterations. We showed an application to ADER-DG, designing the new efficient ADER-$\mathbb{P}_N\mathbb{P}_M$-u methods. In particular, in the ADER-FV-u context ($N = 0$), we have proposed DOOM, an *a posteriori* limiter, that is able to preserve the physical properties of the solution (i.e., positivity of density and pressure) obtaining the maximum admissible order of accuracy that guarantees these physical constraints to be respected. In this framework, there is a huge advantage with respect to similar *a posteriori* limiters, e.g., MOOD [29], as DOOM is waste-free, i.e., all the computations are useful either for increasing the order of accuracy or for detecting a troubled state. In the numerical tests, we have solved Euler and compressible Navier-Stokes equations very robustly, provably keeping the positivity of density and pressure, and with computational costs up to four times smaller than the original method.

We believe that the proposed framework is very versatile and can improve many arbitrary high-order methods on different sides: reducing the computational costs, easily providing the $p$-adaptivity in a very efficient and natural way without wasting computed solutions, and helping obtaining structure-preserving solutions. The authors are currently working on the application of the novel framework to obtain: adaptive methods that converge to the analytical solution up to a given tolerance and $hp$-adaptive methods introducing local mesh refinements in non-smooth regions. We also aim at further investigations on the implicit version of the schemes obtained by choosing a low order implicit operator $\mathcal{L}^1_\Delta$ in the DeC formulation and on structure-preserving.

## Appendix A DG Modal Taylor Basis Functions

The basis functions $\left\{ \vartheta^\ell(\boldsymbol{x},t) \right\}_{\ell=1,\cdots,L}$ used to span the predictor polynomial spaces in this work are modal Taylor basis functions. As already said, they are the tensor products of space basis functions $\left\{ \varphi_i(\boldsymbol{x}) \right\}_{i=1,\cdots,I}$ and time basis functions $\{ \psi^m(t) \}_{m=0,\cdots,M}$.
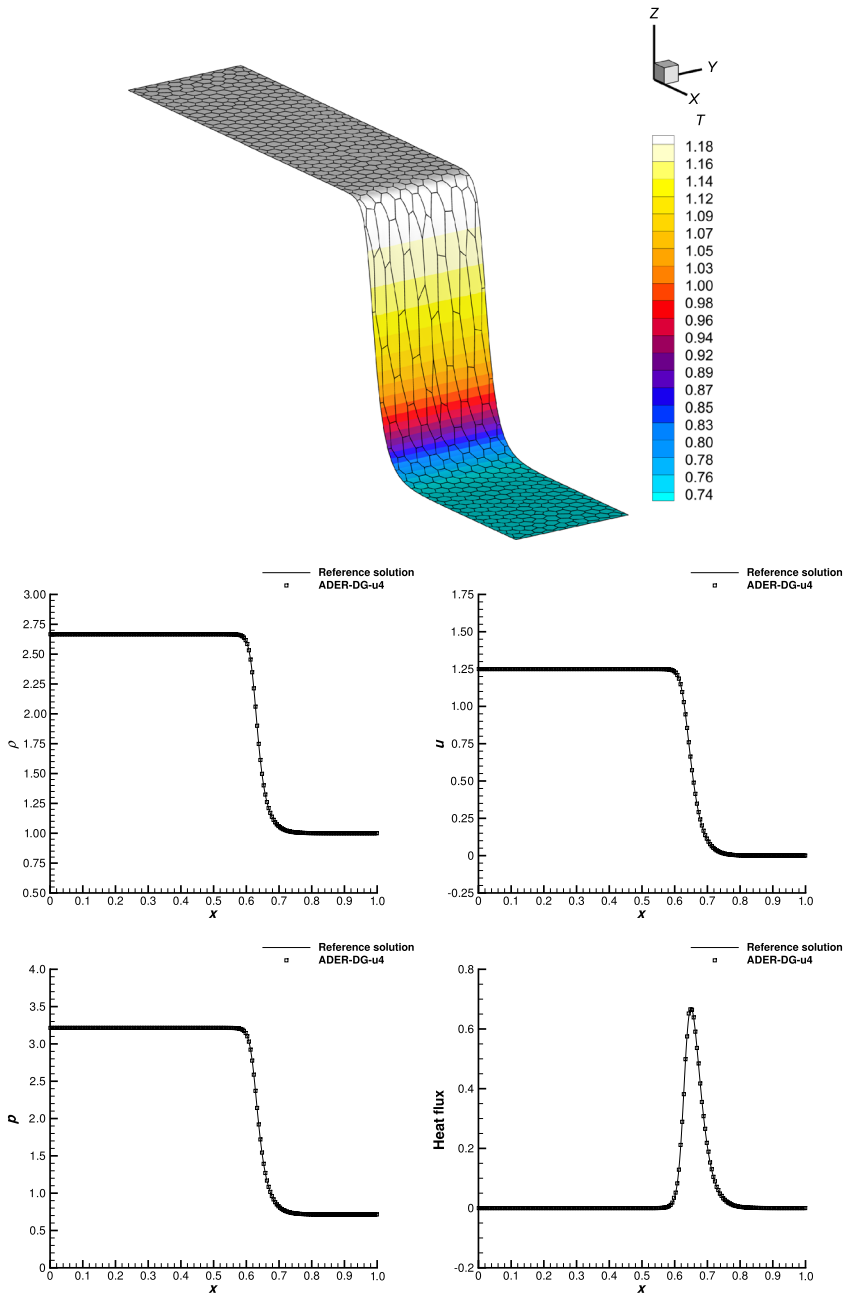
**Fig. 7** Viscous shock profile with a shock Mach number $M_s = 2$ and a Prandtl number $Pr = 0.75$ at time $t_f = 0.2$. Top panel: Voronoi tessellation and temperature distribution along the $z$-axis. The fourth-order numerical solution with the ADER-DG-u scheme compared against the reference solution for the density, horizontal velocity, pressure, and heat flux (from middle left to bottom right panel): in particular, we show a one-dimensional cut of 200 equidistant points along the $x$-direction at $y = 0.1$
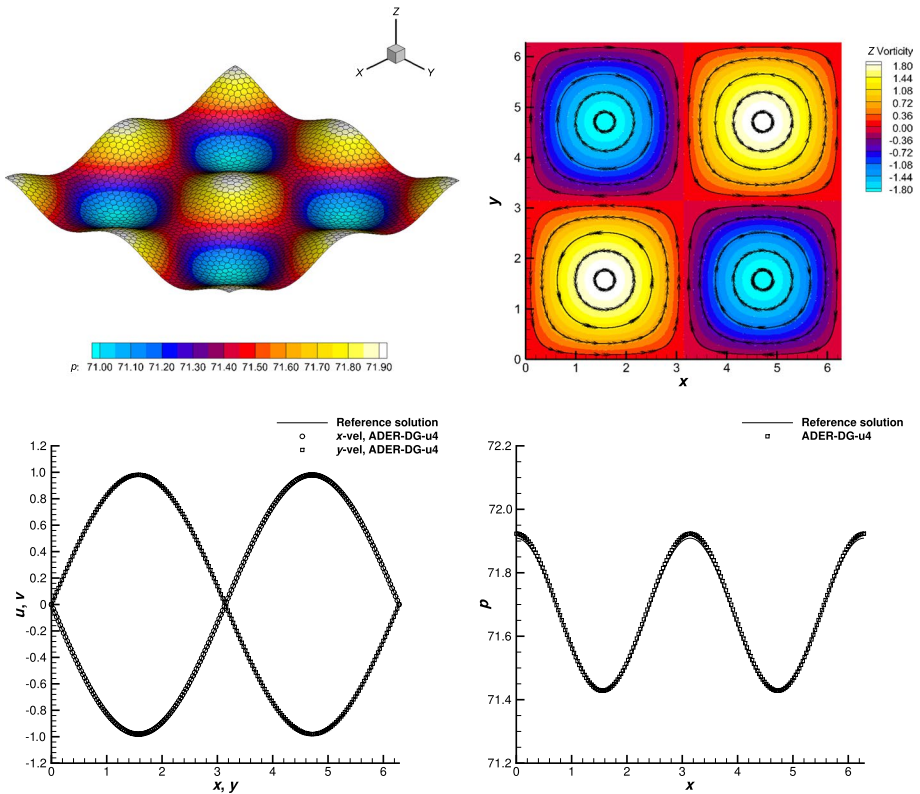
**Fig. 8** 2D Taylor-Green vortex at time $t_f = 1$ with viscosity $\mu = 10^{-2}$. The exact solution of the Navier-Stokes equations and the fourth-order numerical solution with the ADER-DG-u4 scheme. Top: mesh configuration with pressure distribution (left) and $z$-vorticity with stream traces (right). Bottom: one-dimensional cut of 200 equidistant points along the $x$-axis and the $y$-axis for the velocity components $u$ and $v$ (left) and for the pressure $p$ (right)

**Table 4** Error analysis for the Taylor-Green vortex using both ADER-DG-u and ADER-DG schemes with fourth order of accuracy in space and time. The errors are measured in $L_2$ and $L_\infty$ norms and refer to the variables $\rho$ (density) and horizontal velocity $u$ at the final time $t_f = 1$. The computational time measured in seconds is also reported

| Scheme | Density ($\rho$) | | Velocity ($u$) | | CPU time |
| --- | --- | --- | --- | --- | --- |
| | $L_2$ | $L_\infty$ | $L_2$ | $L_\infty$ | /s |
| ADER-DG-u4 | 8.950E−03 | 3.305E−03 | 1.604E−03 | 6.112E−04 | 1.154E+04 |
| ADER-DG4 | 8.950E−03 | 3.305E−03 | 1.604E−03 | 6.112E−04 | 2.706E+04 |

The spatial basis functions $\left\{\varphi_i(\boldsymbol{x})\right\}_{i=1,\cdots,I}$ of degree at most $M$ in $D$ dimensions are defined locally for each element $K$. Denoting by $\boldsymbol{x}_K$ the barycenter of the element, they can be defined as
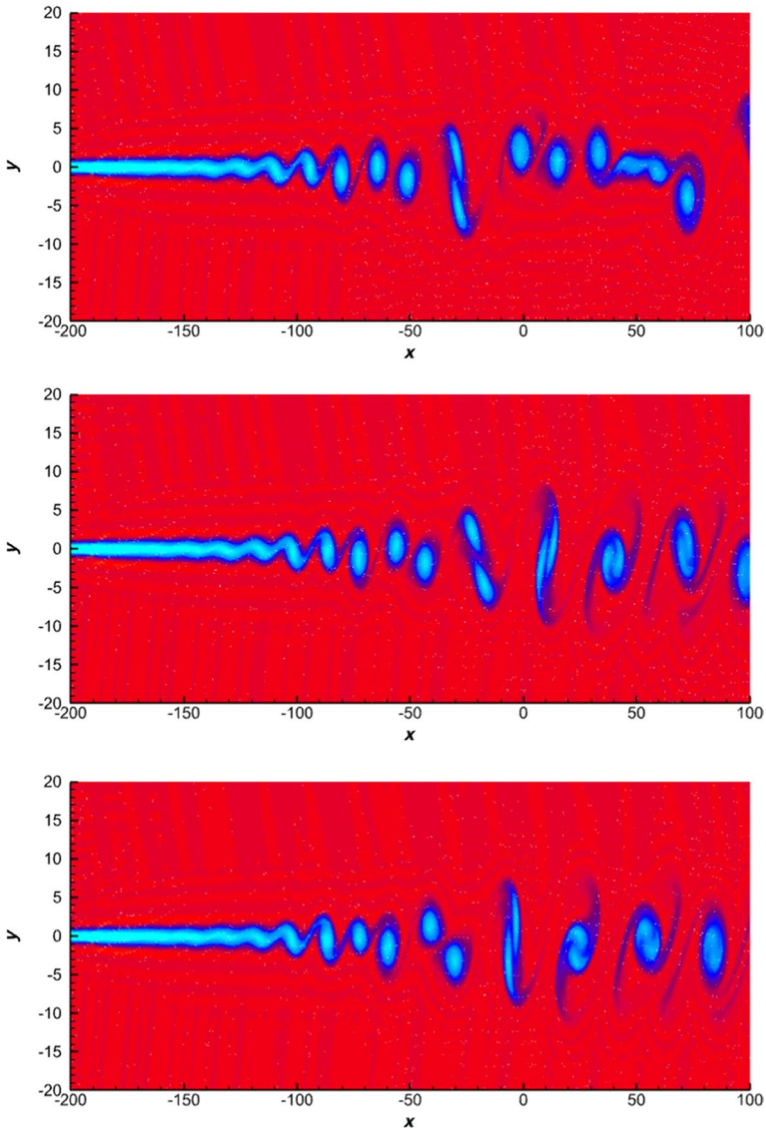
**Fig. 9** Compressible mixing layer at time $t = 500$, $t = 1\,000$, and $t = 1\,596.8$ (from top to bottom row). Fourth-order numerical results with ADER-DG-u for $z$-vorticity. Fifty-one contour levels in the range $[-0.12, 0.12]$ have been used for plotting the vorticity distribution on the sub-domain $[-200, 100] \times [-20, 20]$

$$\varphi_\alpha(\boldsymbol{x}) := \prod_{d=1}^{D} \frac{(x_d - x_{K,d})^{\alpha_d}}{\alpha_d! h_K^{\alpha_d}}, \quad 0 \leqslant |\alpha| \leqslant M, \tag{A1}$$

where $h_K = \sqrt[D]{|K|}$ is the characteristic mesh size of the element $K$, used to rescale the functions to agree with the Taylor expansion terms, while $\alpha$ is a $D$-dimensional multi-index

with $|\alpha| = \sum_{d=1}^{D} \alpha_d$. In practice, we identify the multi-index $\alpha$ as a single index $i = 1, \cdots, I$ with $I = \binom{M+D}{D}$ via a bijection giving $i = i(\alpha)$.

The time basis functions $\{\psi^m(t)\}_{m=0,\cdots,M}$ of degree at most $M$ are defined in a similar fashion, but with respect to a scalar argument only, over $[t_n, t_{n+1}]$

$$\psi^m(t) := \frac{(t - t_n)^m}{m! \Delta t^m}, \quad 0 \leqslant m \leqslant M. \tag{A2}$$

Finally, the tensor product between the two functional spaces gives the space-time basis functions $\{\vartheta^\ell(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L}$. For full reproducibility, we specify that accuracy $M + 1$ has been achieved selecting space-time basis functions up to degree $M$ only. Also, for the spaces $\{\vartheta^{\ell,(p)}(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L^{(p)}}$, we have used basis functions up to degree $p$ only.

**Remark A1** (On the ordering of the space-time basis functions) The novel approach is based on the adoption of iteration-specific bases $\{\vartheta^{\ell,(p)}(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L^{(p)}}$ of increasing degree. For modal bases, the introduction of higher order modes is simply performed by considering higher order terms in the space-time polynomial expansion. Therefore, in the context of an efficient implementation of the new methods, it is particularly useful to directly define all the basis functions $\{\vartheta^{\ell,(M)}(\boldsymbol{x}, t)\}_{\ell=1,\cdots,L^{(M)}}$, up to an accuracy order $M + 1$, ordering them in an increasing polynomial order. By doing so, it is enough to change the final index from $L^{(p-1)}$ to $L^{(p)}$ to pass from $X^{(p-1)}$ to $X^{(p)}$ in all the iterations but the last one, which is performed without changing polynomial space to saturate the related accuracy.

## Declarations

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Compliance with Ethical Standards** On behalf of all authors, the corresponding author is available to collect documentation of compliance with ethical standards and send upon request.

# References

1. Abgrall, R.: Residual distribution schemes: current status and future trends. Comput. Fluids **35**(7), 641–669 (2006)
2. Abgrall, R.: High order schemes for hyperbolic problems using globally continuous approximation and avoiding mass matrices. J. Sci. Comput. **73**(2/3), 461–494 (2017). https://doi.org/10.1007/s10915-017-0498-4
3. Abgrall, R., Bacigaluppi, P., Tokareva, S.: High-order residual distribution scheme for the time-dependent Euler equations of fluid dynamics. Comput. Math. Appl. **78**(2), 274–297 (2019)
4. Abgrall, R., Le Mélédo, E., Öffner, P., Torlo, D.: Relaxation deferred correction methods and their applications to residual distribution schemes. SMAI J. Comput. Math. **8**, 125–160 (2022). https://doi.org/10.5802/smai-jcm.82
5. Abgrall, R., Lukácova-Medvid'ová, M., Öffner, P.: On the convergence of residual distribution schemes for the compressible Euler equations via dissipative weak solutions. arXiv:2207.11969 (2022)
6. Abgrall, R., Nordström, J., Öffner, P., Tokareva, S.: Analysis of the SBP-SAT stabilization for finite element methods part II: entropy stability. Commun. Appl. Math. Comput. **5**, 573–595 (2021)
7. Abgrall, R., Öffner, P., Ranocha, H.: Reinterpretation and extension of entropy correction terms for residual distribution and discontinuous Galerkin schemes: application to structure preserving discretization. J. Comput. Phys. **453**, 110955 (2022)
8. Abgrall, R., Torlo, D.: High order asymptotic preserving deferred correction implicit-explicit schemes for kinetic models. SIAM J. Sci. Comput. **42**(3), 816–845 (2020)
9. Bacigaluppi, P., Abgrall, R., Tokareva, S.: "A posteriori" limited high order and robust residual distribution schemes for transient simulations of fluid flows in gas dynamics. arXiv:1902.07773 (2019)
10. Balsara, D.S., Shu, C.-W.: Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. J. Comput. Phys. **160**(2), 405–452 (2000)
11. Becker, R.: Stosswelle und detonation. Zeitschrift für Physik **8**(1), 321–362 (1922)
12. Berberich, J.P., Chandrashekar, P., Klingenberg, C.: High order well-balanced finite volume methods for multi-dimensional systems of hyperbolic balance laws. Comput. Fluids **219**, 104858 (2021)
13. Boscheri, W., Balsara, D.S.: High order direct Arbitrary-Lagrangian-Eulerian (ALE) $P_N P_M$ schemes with WENO Adaptive-Order reconstruction on unstructured meshes. J. Comput. Phys. **398**, 108899 (2019)
14. Boscheri, W., Dumbser, M.: Arbitrary-Lagrangian-Eulerian one-step WENO finite volume schemes on unstructured triangular meshes. Commun. Comput. Phys. **14**, 1174–1206 (2013)
15. Boscheri, W., Dumbser, M.: Arbitrary-Lagrangian-Eulerian discontinuous Galerkin schemes with a posteriori subcell finite volume limiting on moving unstructured meshes. J. Comput. Phys. **346**, 449–479 (2017)
16. Boscheri, W., Dumbser, M., Gaburro, E.: Continuous finite element subgrid basis functions for discontinuous Galerkin schemes on unstructured polygonal Voronoi meshes. Commun. Comput. Phys. **32**, 259–298 (2022)
17. Boscheri, W., Loubère, R.: High order accurate direct Arbitrary-Lagrangian-Eulerian ADER-MOOD finite volume schemes for non-conservative hyperbolic systems with stiff source terms. Commun. Comput. Phys. **21**, 271–312 (2017)
18. Boscheri, W., Loubère, R., Dumbser, M.: Direct arbitrary-Lagrangian-Eulerian ADER-MOOD finite volume schemes for multidimensional hyperbolic conservation laws. J. Comput. Phys. **292**, 56–87 (2015)
19. Busto, S., Chiocchetti, S., Dumbser, M., Gaburro, E., Peshkov, I.: High order ADER schemes for continuum mechanics. Front. Phys. **8**, 32 (2020). https://doi.org/10.3389/fphy.2020.00032
20. Busto, S., Dumbser, M., Río-Martín, L.: Staggered semi-implicit hybrid finite volume/finite element schemes for turbulent and non-Newtonian flows. Mathematics **9**(22), 2972 (2021)
21. Busto, S., Toro, E.F., Vázquez-Cendón, M.E.: Design and analysis of ADER-type schemes for model advection-diffusion-reaction equations. J. Comput. Phys. **327**, 553–575 (2016)
22. Castro, M.J., Parés, C.: Well-balanced high-order finite volume methods for systems of balance laws. J. Sci. Comput. (2020). https://doi.org/10.1007/s10915-020-01149-5
23. Chen, T., Shu, C.-W.: Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws. J. Comput. Phys. **345**, 427–461 (2017)
24. Chen, T., Shu, C.-W.: Review of entropy stable discontinuous Galerkin methods for systems of conservation laws on unstructured simplex meshes. CSIAM Trans. Appl. Math. **1**, 1–52 (2020)

25. Cheng, Y., Chertock, A., Herty, M., Kurganov, A., Wu, T.: A new approach for designing moving-water equilibria preserving schemes for the shallow water equations. J. Sci. Comput. **80**(1), 538–554 (2019)
26. Chertock, A., Cui, S., Kurganov, A., Özcan, C.N., Tadmor, E.: Well-balanced schemes for the Euler equations with gravitation: conservative formulation using global fluxes. J. Comput. Phys. **358**, 36–52 (2018). https://doi.org/10.1016/j.jcp.2017.12.026
27. Ciallella, M., Micalizzi, L., Öffner, P., Torlo, D.: An arbitrary high order and positivity preserving method for the shallow water equations. Comput. Fluids **247**, 105630 (2022). https://doi.org/10.1016/j.compfluid.2022.105630
28. Ciallella, M., Torlo, D., Ricchiuto, M.: Arbitrary high order WENO finite volume scheme with flux globalization for moving equilibria preservation. arXiv:2205.13315 (2022)
29. Clain, S., Diot, S., Loubère, R.: A high-order finite volume method for systems of conservation laws-Multi-dimensional Optimal Order Detection (MOOD). J. Comput. Phys. **230**(10), 4028–4050 (2011)
30. Colonius, T., Lele, S., Moin, P.: Sound generation in a mixing layer. J. Fluid Mech. **330**, 375–409 (1997)
31. Diot, S., Clain, S., Loubère, R.: Improved detection criteria for the Multi-dimensional Optimal Order Detection (MOOD) on unstructured meshes with very high-order polynomials. Comput. Fluids **64**, 43–63 (2012)
32. Diot, S., Loubère, R., Clain, S.: The MOOD method in the three-dimensional case: very-high-order finite volume method for hyperbolic systems. Int. J. Numer. Methods Fluids **73**, 362–392 (2013)
33. Dumbser, M.: Arbitrary high order $P_N P_M$ schemes on unstructured meshes for the compressible Navier-Stokes equations. Comput. Fluids **39**(1), 60–76 (2010)
34. Dumbser, M., Balsara, D.S., Toro, E.F., Munz, C.-D.: A unified framework for the construction of one-step finite volume and discontinuous Galerkin schemes on unstructured meshes. J. Comput. Phys. **227**(18), 8209–8253 (2008)
35. Dumbser, M., Munz, C.-D.: ADER discontinuous Galerkin schemes for aeroacoustics. Comptes Rendus Mécanique **333**(9), 683–687 (2005)
36. Dumbser, M., Toro, E.F.: A simple extension of the Osher Riemann solver to non-conservative hyperbolic systems. J. Sci. Comput. **48**, 70–88 (2011)
37. Dumbser, M., Zanotti, O.: Very high order $P_N P_M$ schemes on unstructured meshes for the resistive relativistic MHD equations. J. Comput. Phys. **228**(18), 6991–7006 (2009)
38. Dutt, A., Greengard, L., Rokhlin, V.: Spectral deferred correction methods for ordinary differential equations. BIT **40**(2), 241–266 (2000). https://doi.org/10.1023/A:1022338906936
39. Farhat, C., Fezoui, L., Lanteri, S.: Two-dimensional viscous flow computations on the connection machine: unstructured meshes, upwind schemes and massively parallel computations. Comput. Methods Appl. Mech. Eng. **102**(1), 61–88 (1993). https://doi.org/10.1016/0045-7825(93)90141-J
40. Friedrich, L., Winters, A.R., Fernández, D.C.D.R., Gassner, G.J., Parsani, M., Carpenter, M.H.: An entropy stable *h/p* non-conforming discontinuous Galerkin method with the summation-by-parts property. J. Sci. Comput. **77**(2), 689–725 (2018)
41. Gaburro, E.: A unified framework for the solution of hyperbolic PDE systems using high order direct Arbitrary-Lagrangian-Eulerian schemes on moving unstructured meshes with topology change. Arch. Comput. Methods Eng. **28**(3), 1249–1321 (2021)
42. Gaburro, E., Boscheri, W., Chiocchetti, S., Klingenberg, C., Springel, V., Dumbser, M.: High order direct Arbitrary-Lagrangian-Eulerian schemes on moving Voronoi meshes with topology changes. J. Comput. Phys. **407**, 109167 (2020)
43. Gaburro, E., Dumbser, M.: A posteriori subcell finite volume limiter for general $P_N P_M$ schemes: applications from gasdynamics to relativistic magnetohydrodynamics. J. Sci. Comput. **86**(3), 1–41 (2021)
44. Gaburro, E., Öffner, P., Ricchiuto, M., Torlo, D.: High order entropy preserving ADER-DG scheme. Appl. Math. Comput. **440**, 127644 (2023). https://doi.org/10.1016/j.amc.2022.127644
45. Gassner, G.J., Winters, A.R., Kopriva, D.A.: A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations. Appl. Math. Comput. **272**, 291–308 (2016)
46. Glaubitz, J., Öffner, P.: Stable discretisations of high-order discontinuous Galerkin methods on equidistant and scattered points. Appl. Numer. Math. **151**, 98–118 (2020)
47. Gómez-Bueno, I., Boscarino, S., Castro, M.J., Parés, C., Russo, G.: Implicit and semi-implicit well-balanced finite-volume methods for systems of balance laws. Appl. Numer. Math. **184**, 18–48 (2023)
48. Gottlieb, S., Shu, C.-W.: Total variation diminishing Runge-Kutta schemes. Math. Comput. Am. Math. Soc. **67**(221), 73–85 (1998)
49. Hajduk, H.: Monolithic convex limiting in discontinuous Galerkin discretizations of hyperbolic conservation laws. Comput. Math. Appl. **87**, 120–138 (2021)

50. Han Veiga, M., Öffner, P., Torlo, D.: DeC and ADER: similarities, differences and a unified framework. J. Sci. Comput. **87**(1), 1–35 (2021)

51. Han Veiga, M., Velasco-Romero, D.A., Abgrall, R., Teyssier, R.: Capturing near-equilibrium solutions: a comparison between high-order discontinuous Galerkin methods and well-balanced schemes. Commun. Comput. Phys. **26**(1), 1–34 (2019). https://doi.org/10.4208/cicp.oa-2018-0071

52. Huang, D.Z., Avery, P., Farhat, C., Rabinovitch, J., Derkevorkian, A., Peterson, L.D.: Modeling, simulation and validation of supersonic parachute inflation dynamics during mars landing. In: AIAA Scitech 2020 Forum, pp. 0313 (2020)

53. Huang, J., Shu, C.-W.: Positivity-preserving time discretizations for production-destruction equations with applications to non-equilibrium flows. J. Sci. Comput. **78**(3), 1811–1839 (2019)

54. Jund, S., Salmon, S.: Arbitrary high-order finite element schemes and high-order mass lumping. Int. J. Appl. Math. Comput. Sci. **17**(3), 375–393 (2007)

55. Krais, N., Beck, A., Bolemann, T., Frank, H., Flad, D., Gassner, G., Hindenlang, F., Hoffmann, M., Kuhn, T., Sonntag, M., Munz, C.-D.: FLEXI: a high order discontinuous Galerkin framework for hyperbolic-parabolic conservation laws. Comput. Math. Appl. **81**, 186–219 (2021)

56. Kuzmin, D.: Entropy stabilization and property-preserving limiters for $\mathbb{P}_1$ discontinuous Galerkin discretizations of scalar hyperbolic problems. J. Numer. Math. **29**(4), 307–322 (2021)

57. Kuzmin, D., de Luna, M.Q.: Entropy conservation property and entropy stabilization of high-order continuous Galerkin approximations to scalar conservation laws. Comput. Fluids **213**, 104742 (2020)

58. Lukáčová-Medvid'ová, M., Öffner, P.: Convergence of discontinuous Galerkin schemes for the Euler equations via dissipative weak solutions. Appl. Math. Comput. **436**, 127508 (2023)

59. Mantri, Y., Noelle, S.: Well-balanced discontinuous Galerkin scheme for 2×2 hyperbolic balance law. J. Comput. Phys. **429**, 110011 (2021)

60. Meister, A., Ortleb, S.: On unconditionally positive implicit time integration for the DG scheme applied to shallow water flows. Int. J. Numer. Meth. Fluids **76**(2), 69–94 (2014)

61. Micalizzi, L., Torlo, D.: A new efficient explicit Deferred Correction framework: analysis and applications to hyperbolic PDEs and adaptivity (2022). https://doi.org/10.48550/arxiv.2210.02976

62. Millington, R., Toro, E., Nejad, L.: Arbitrary high order methods for conservation laws I: the one dimensional scalar case. PhD thesis, Manchester Metropolitan University, Department of Computing and Mathematics (June 1999)

63. Minion, M.L.: Semi-implicit spectral deferred correction methods for ordinary differential equations. Commun. Math. Sci. **1**(3), 471–500 (2003)

64. Noelle, S., Xing, Y., Shu, C.-W.: High-order well-balanced finite volume WENO schemes for shallow water equation with moving water. J. Comput. Phys. **226**(1), 29–58 (2007)

65. Öffner, P., Glaubitz, J., Ranocha, H.: Stability of correction procedure via reconstruction with summation-by-parts operators for Burgers' equation using a polynomial chaos approach. ESAIM: Mathematical Modelling and Numerical Analysis **52**(6), 2215–2245 (2018)

66. Öffner, P., Torlo, D.: Arbitrary high-order, conservative and positivity preserving Patankar-type deferred correction schemes. Appl. Numer. Math. **153**, 15–34 (2020)

67. Perthame, B., Shu, C.-W.: On positivity preserving finite volume schemes for Euler equations. Numer. Math. **73**(1), 119–130 (1996)

68. Rabenseifner, R., Hager, G., Jost, G.: Hybrid MPI/OpenMP parallel programming on clusters of multi-core SMP nodes. In: 2009 17th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, pp. 427–436 (2009)

69. Ranocha, H., Sayyari, M., Dalcin, L., Parsani, M., Ketcheson, D.I.: Relaxation Runge-Kutta methods: fully discrete explicit entropy-stable schemes for the compressible Euler and Navier-Stokes equations. SIAM J. Sci. Comput. **42**(2), 612–638 (2020)

70. Ricchiuto, M.: An explicit residual based approach for shallow water flows. J. Comput. Phys. **280**, 306–344 (2015)

71. Ricchiuto, M., Abgrall, R.: Explicit Runge-Kutta residual distribution schemes for time dependent problems: second order case. J. Comput. Phys. **229**(16), 5653–5691 (2010)

72. Ricchiuto, M., Torlo, D.: Analytical travelling vortex solutions of hyperbolic equations for validating very high order schemes (2021). https://doi.org/10.48550/ARXIV.2109.10183

73. Río-Martín, L., Busto, S., Dumbser, M.: A massively parallel hybrid finite volume/finite element scheme for computational fluid dynamics. Mathematics **9**(18), 2316 (2021)

74. Rusanov, V.V.: Calculation of interaction of non-steady shock waves with obstacles. J. Comput. Math. Phys. USSR **1**, 267–279 (1961)

75. Shu, C.-W.: Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In: Quarteroni, A. (ed) Advanced Numerical Approximation of Nonlinear

Hyperbolic Equations. Lecture Notes in Mathematics, vol. 1697, pp. 325–432. Springer, Berlin, Heidelberg (1998)

76. Shu, C.-W., Osher, S.: Efficient implementation of essentially non-oscillatory shock capturing schemes. J. Comput. Phys. **77**, 439–471 (1988)

77. Spiegel, S.C., Huynh, H., DeBonis, J.R.: A survey of the isentropic Euler vortex problem using high-order methods. In: 22nd AIAA Computational Fluid Dynamics Conference, AIAA 2015-2444 (2015)

78. Titarev, V.A., Toro, E.F.: ADER: arbitrary high order Godunov approach. J. Sci. Comput. **17**(1/2/3/4), 609–618 (2002)

79. Titarev, V.A., Toro, E.F.: ADER schemes for three-dimensional nonlinear hyperbolic systems. J. Comput. Phys. **204**, 715–736 (2005)

80. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics: a Practical Introduction. Springer, Berlin, Heidelberg (2009). https://books.google.com/books?id=SqEjX0um8o0C

81. Tsoutsanis, P., Antoniadis, A.F., Jenkins, K.W.: Improvement of the computational performance of a parallel unstructured WENO finite volume CFD code for Implicit Large Eddy Simulation. Comput. Fluids **173**, 157–170 (2018)

82. Tsoutsanis, P., Drikakis, D.: Addressing the challenges of implementation of high-order finite-volume schemes for atmospheric dynamics on unstructured meshes. In: Papadrakakis, M., Papadopoulos, V., Stefanou, G., Plevris, V. (eds.) ECCOMAS Congress 2016-Proceedings of the 7th European Congress on Computational Methods in Applied Sciences and Engineering, vol. 1, pp. 684–708. National Technical University of Athens, GRC (2016)

83. Wintermeyer, N., Winters, A.R., Gassner, G.J., Kopriva, D.A.: An entropy stable nodal discontinuous Galerkin method for the two dimensional shallow water equations on unstructured curvilinear meshes with discontinuous bathymetry. J. Comput. Phys. **340**, 200–242 (2017)

84. Winters, A.R., Gassner, G.J.: A comparison of two entropy stable discontinuous Galerkin spectral element approximations for the shallow water equations with non-constant topography. J. Comput. Phys. **301**, 357–376 (2015)

85. Xing, Y., Shu, C.-W.: High-order well-balanced finite difference WENO schemes for a class of hyperbolic systems with source terms. J. Sci. Comput. **27**(1), 477–494 (2006)

86. Zhang, X., Shu, C.-W.: Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: survey and new developments. Proc. R. Soci. A Math. Phys. Eng. Sci. **467**(2134), 2752–2776 (2011)