A generalized Active Flux method of arbitrarily high order in two dimensions

Wasilij Barsukow [a,*], Praveen Chandrashekar [b], Christian Klingenberg [c], Lisa Lechner [c]

[a] *Institut de Mathématiques de Bordeaux (IMB), CNRS UMR 5251, 351 Cours de la Libération, 3340, Talence, France*
[b] *Centre for Applicable Mathematics, Tata Institute of Fundamental Research, 560065, Bengaluru, India*
[c] *Institute of Mathematics, University of Würzburg, Emil-Fischer-Straße 40 Würzburg, 97074, Germany*

ARTICLE INFO

ABSTRACT

The Active Flux method can be seen as an extended finite volume method. The degrees of freedom of this method are cell averages, as in finite volume methods, and in addition shared point values at the cell interfaces, giving rise to a globally continuous reconstruction. Its classical version was introduced as a one-stage fully discrete, third-order method. Recently, a semi-discrete version of the Active Flux method was presented with various extensions to arbitrarily high order in *one* space dimension. In this paper we extend the semi-discrete Active Flux method on *two*-dimensional Cartesian grids to arbitrarily high order, by including moments as additional degrees of freedom (hybrid finite element–finite volume method). The stability of this method is studied for linear advection. For a fully discrete version, using an explicit Runge-Kutta method, a CFL restriction is derived. We end by presenting numerical examples for hyperbolic conservation laws.

1. Introduction

The solution of systems of hyperbolic conservation laws is required by many different physical problems from fluid dynamics. As most of these problems are fairly complex, numerical methods are needed to solve them. Finite volume methods evolve cell averages and maintain conservation in a discrete sense which is important for convergence to a weak solution. Godunov's method [1], for example, is a seminal approach to hyperbolic conservation laws. More recently, Roe and coauthors introduced the Active Flux method [2–4] inspired by van Leer's scheme V [5] for linear advection. The new method not only uses cell averages but also point values at the cell interfaces. This allows to reconstruct the solution piecewise parabolically in a globally continuous fashion. In van Leer [5], the point values at the cell interfaces were updated by tracing back the characteristics. This way, a method of third order for linear advection in one spatial dimension arose. Roe's one-stage fully discrete Active Flux method (which we will call the *classical Active Flux method*, for a review on Cartesian grids see e.g. Chudzik and Helzel [6]) extends this idea to certain nonlinear and/or multi-dimensional hyperbolic problems by replacing characteristic transport with suitable either exact or approximate evolution operators (e.g. Maeng [7], Fan [8], Barsukow et al. [9], Barsukow [10], Chudzik et al. [11]). Depending on the system of conservation laws finding such an evolution operator may be

quite challenging. Recently the authors in Abgrall [12], Abgrall and Barsukow [13,14], Abgrall et al. [15] proposed a semi-discrete formulation of the Active Flux method (which we will call the *generalized Active Flux method*) that allows for a more flexible approach to update the point values.

In Abgrall and Barsukow [13], several extensions of this method to arbitrary order in one spatial dimension are presented, while the multidimensional method in Abgrall et al. [15] was only third-order accurate. In this paper we extend this work to obtain a semi-discrete generalized Active Flux method of arbitrarily high order on Cartesian grids in two spatial dimensions. The design of the method is based on a hybrid finite element–finite volume method (see also Abgrall and Barsukow [14]) and also takes inspiration from the one-dimensional arbitrary-order Active Flux method with additional point values developed in Abgrall and Barsukow [13].

While this paper focuses on Cartesian grids, we want to also mention [16], where a semi-discrete Active Flux method on triangular meshes is presented and results for a third- and fourth-order accurate method are shown, using the cell average and point values at the cell interfaces. Furthermore, He [17] and Roe [18], Samani and Roe [19], Samani [20] consider extensions of the (classical) fully discrete Active Flux method to arbitrary order. The first approach uses additional point values. In the latter, they typically enrich the stencil using additional information

---

in the form of derivatives at the already defined point values at the cell interfaces (also called "Hermite Active Flux").

Here, we aim at using higher moments in the cells and additional point values at the cell interfaces. For efficiency, we especially aim at achieving arbitrarily high order with a minimal number of degrees of freedom while maintaining the compact stencil. Another possibility is a tensor-like extension of the hybrid finite element–finite volume method in Abgrall and Barsukow [14], which is also formally defined in Kern [21], but it uses more degrees of freedom than actually necessary for the required order of accuracy.

Using our approach, key features of Active Flux like global continuity of the reconstructed solution and a compact stencil of the method in space are retained. As the method is based on the semi-discrete approach, it has the potential to be more easily applicable to different systems of conservation laws than the classical Active Flux. For integration in time we use an explicit Runge-Kutta method. While classical Active Flux is a one-stage fully discrete method, the multiple stages of a Runge-Kutta method widen the effective stencil of the generalized Active Flux method. In Roe [18], this has been linked to a reduction of the maximal CFL number, compared to the classical Active Flux method. Yet, Barsukow et al. [22] shows encouraging similarities of the classical and generalized Active Flux method of third order on Cartesian grids for linear acoustics, such as stationarity preservation (see Barsukow et al. [9] for stationarity preservation of the classical Active Flux).

The paper is structured as follows: In Section 2 we briefly recall arbitrarily high-order one-dimensional Active Flux from Abgrall and Barsukow [13,14] and third-order two-dimensional Active Flux from Abgrall et al. [15]. Then, our high order method is defined in Section 3. Section 4 focuses on a stability analysis for linear advection of both the semi-discrete and the fully discrete method using RK3. Numerical results up to seventh order are presented for our high-order method in Section 5, including an example for the 2-d compressible Euler equations.

2. An overview of the semi-discrete Active Flux method

Consider the hyperbolic system of conservation laws

$$\partial_t q(t, \mathbf{x}) + \nabla \cdot \mathbf{f}(q(t, \mathbf{x})) = 0 \tag{1}$$

for $q : \mathbb{R}_0^+ \times \Omega \to \mathbb{R}^s$ on the domain $\Omega \subseteq \mathbb{R}^d$. Throughout this paper the focus is on the two-dimensional case $d = 2$ with $\mathbf{f} = (f^x, f^y)$, $f^x, f^y : \mathbb{R}^s \to \mathbb{R}^s$. The notation used here is similar to Abgrall et al. [15].

Furthermore, we restrict ourselves to rectangular domains $\Omega = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$ discretized with Cartesian grids with equidistant cell widths Δx and Δy in x- and y-direction. For a Cartesian grid of size $N_x \times N_y$ the grid cells will be denoted as $C_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ for $i = 0, \ldots, N_x - 1, j = 0, \ldots N_y - 1$ centered at $\mathbf{x}_{ij} = (x_i, y_j)$.

Next, we will give an overview of semi-discrete Active Flux methods, that we will use and extend in this paper.

2.1. A third-order method on two-dimensional Cartesian grids

In Abgrall et al. [15] a semi-discrete Active Flux method of third order on Cartesian grids was presented for the two-dimensional case. As for the classical Active Flux method on Cartesian grids (see e.g. Barsukow et al. [9], Helzel et al. [23]), the degrees of freedom for each cell C_{ij} are the cell average

$$\bar{q}_{ij}(t) = \frac{1}{\Delta x \Delta y} \int_{C_{ij}} q(t, \mathbf{x}) d\mathbf{x} \tag{2}$$

and the nodal, vertical and horizontal point values

$$q_{i+\frac{1}{2}, j+\frac{1}{2}}(t) = q\left(t, x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}}\right),$$

$$q_{i, j+\frac{1}{2}}(t) = q\left(t, x_i, y_{j+\frac{1}{2}}\right), \quad q_{i+\frac{1}{2}, j}(t) = q\left(t, x_{i+\frac{1}{2}}, y_j\right) \tag{3}$$

at the cell interfaces. All point values are shared, i.e. each cell has access to eight point values at $\mathbf{x}_p$, $p \in \left\{(i \pm \frac{1}{2}, j \pm \frac{1}{2}), (i \pm \frac{1}{2}, j), (i, j \pm \frac{1}{2})\right\}$ (see Fig. 1(a)).

The reconstruction on cell C_{ij} is the biparabolic polynomial $q_{ij,\text{recon}}$: $[-\frac{\Delta x}{2}, \frac{\Delta x}{2}] \times [-\frac{\Delta y}{2}, \frac{\Delta y}{2}] \to \mathbb{R}^s$ (see Barsukow et al. [9], Abgrall et al. [15]) defined on a reference cell with $\mathbf{x}_{\text{ref}} := \mathbf{x} - \mathbf{x}_{ij}$ and $\mathbf{x}_{\text{ref},p} := \mathbf{x}_p - \mathbf{x}_{ij}$ which has to satisfy

$$\frac{1}{\Delta x \Delta y} \int_{-\frac{\Delta y}{2}}^{\frac{\Delta y}{2}} \int_{-\frac{\Delta x}{2}}^{\frac{\Delta x}{2}} q_{ij,\text{recon}}(\mathbf{x}_{\text{ref}}) dx_{\text{ref}} dy_{\text{ref}} = \bar{q}_{ij},$$

$$q_{ij,\text{recon}}(\mathbf{x}_{\text{ref},p}) = q_p \quad \forall p \in \left\{\left(i \pm \frac{1}{2}, j \pm \frac{1}{2}\right), \left(i \pm \frac{1}{2}, j\right), \left(i, j \pm \frac{1}{2}\right)\right\}.$$

It can also be written in terms of shape functions, see e.g. Barsukow et al. [22]. This gives rise to a globally continuous reconstruction q_{recon} on Ω (see Fig. 1(b)) with

$$q_{\text{recon}}|_{C_{ij}}(\mathbf{x}) := q_{ij,\text{recon}}(\mathbf{x}_{\text{ref}}). \tag{4}$$

For the update of the point values $q_p, p \in \left\{\left(i + \frac{1}{2}, j + \frac{1}{2}\right), \left(i + \frac{1}{2}, j\right), \left(i, j + \frac{1}{2}\right)\right\}$ the quasi-linear form of (1) at $\mathbf{x}_p$

$$\frac{d}{dt} q_p(t) + Df^x(q_p(t)) \partial_x q|_p(t) + Df^y(q_p(t)) \partial_y q|_p(t) = 0$$

is considered. In order to introduce stabilization via upwinding, the Jacobians $J^x = Df^x(q_p), J^y = Df^y(q_p)$ are split according to the sign of the wave speeds

$$(J^x)^\pm := T^x \text{diag}(\lambda_{x,1}^\pm, \ldots, \lambda_{x,s}^\pm)(T^x)^{-1},$$

$$\lambda_{x,k}^+ := \max(0, \lambda_{x,k}), \quad \lambda_{x,k}^- := \min(0, \lambda_{x,k}) \quad \forall k = 1, \ldots, s.$$

The derivatives $\partial_x q|_p$ are approximated with finite difference formulas in upwind direction $(D_x^\pm)_p q$ and $(D_y^\pm)_p q$ derived from the reconstruction q_{recon}. In directions tangential to the cell interface for the horizontal and vertical edge point, the derivative of the reconstruction is continuous and no upwinding is included.

The update of the cell averages is obtained by integrating the conservation law (1) over the cell C_{ij} and applying the divergence theorem

$$\frac{d}{dt} \bar{q}_{ij}(t) + \frac{1}{\Delta x \Delta y} \int_{\partial C_{ij}} \mathbf{f}(q(t, \mathbf{x})) \cdot \mathbf{n} dS = 0.$$

The reconstructed solution is continuous at the cell interfaces and the flux terms over the cell interfaces are numerically integrated with a Gauss-Lobatto quadrature with three points, which coincide with the point values at the cell interfaces.

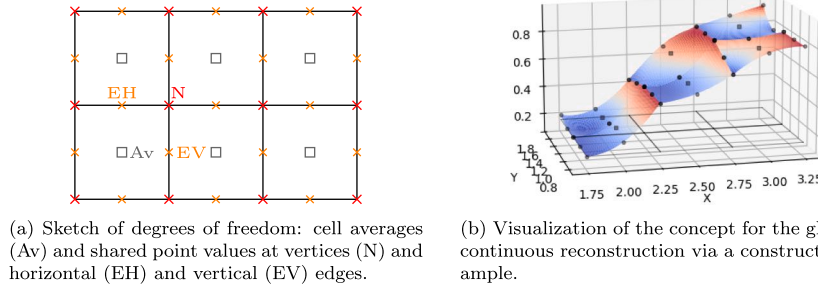### 2.2. Methods of arbitrary order in one dimension

In Abgrall and Barsukow [13], different methods to extend semi-discrete Active Flux to arbitrarily high order in one dimension were suggested. Here, two of these concepts are reviewed, and a variation of the second method is presented.

2.2.1. A method with higher moments

In this paper, we will mainly focus on the hybrid finite element–finite volume method as also defined in Abgrall and Barsukow [14, Definition 4.1 (Method A)]. For a method of order $N + 1$ on a compact stencil, $N + 1$ degrees of freedom accessible to a cell are needed to approximate a polynomial of degree N. In addition to the traditional degrees of freedom, i.e. the cell averages and the point values at the cell interfaces, this method uses higher moments

$$q_i^{(k)}(t) := A_k \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(t, x) b_k(x - x_i) dx \qquad k \in \mathbb{N}.$$

The test functions for the moments are $b_k : [-\frac{\Delta x}{2}, \frac{\Delta x}{2}] \to \mathbb{R}, b_k(x) = x^k$ with the normalization factors $A_k = \frac{(k+1)2^k}{\Delta x^{k+1}}$. This allows to obtain more

(a) Sketch of degrees of freedom: cell averages (Av) and shared point values at vertices (N) and horizontal (EH) and vertical (EV) edges.

(b) Visualization of the concept for the globally continuous reconstruction via a constructed example.

Fig. 1. Degrees of freedom and reconstruction for a third-order Active Flux method on two-dimensional Cartesian grids.

information on a compact stencil. The update of the cell average $\bar{q}_i = q_i^{(0)}$ and higher moments is given as

$$\frac{\mathrm{d}}{\mathrm{d}t} q_i^{(k)}(t) = -A_k \left(f(q_{i+\frac{1}{2}}(t)) b_k\left(\frac{\Delta x}{2}\right) - f(q_{i-\frac{1}{2}}(t)) b_k\left(-\frac{\Delta x}{2}\right) \right)$$

$$+ A_k \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(q(t,x)) b_k'(x - x_i) \mathrm{d}x \quad \forall k = 0, \ldots, N-2$$

for a method of order $N + 1$. The update of the point values $q_{i+\frac{1}{2}} = q(x_{i+\frac{1}{2}})$ uses a Jacobian splitting of $Df(q)$ to include upwinding

$$\frac{\mathrm{d}}{\mathrm{d}t} q_{i+\frac{1}{2}}(t) = -Df(q_{i+\frac{1}{2}})^+ (D_x^+)_{i+\frac{1}{2}} q - Df(q_{i+\frac{1}{2}})^- (D_x^-)_{i+\frac{1}{2}} q.$$

The polynomial reconstructions $q_{i,\mathrm{recon}} : [-\frac{\Delta x}{2}, \frac{\Delta x}{2}] \to \mathbb{R}^s$ of degree $N \geq 2$ are used to find a finite difference approximation for $(D_x^\pm)_{i+\frac{1}{2}} q$.

2.2.2. A method with additional point values

Another possibility to achieve arbitrarily high order of the Active Flux method is to introduce further point values $q_{i,\xi_a} = q(x_i + \Delta x \xi_a), \xi_a \in (-\frac{1}{2}, \frac{1}{2})$ within each cell in addition to the cell averages and the point values at the cell interfaces $q_{i\pm\frac{1}{2}}$. This method will especially be relevant for the discussion of stability in Section 4. To achieve a method of order $N + 1$, $N - 2$ additional point values are needed and similarly to Section 2.2.1, a polynomial reconstruction can be defined on the cells C_i (except in the case of a symmetric distribution of an odd number of point values). In Abgrall and Barsukow [13] this method has been introduced using evolution operators for the point value updates.

Before proceeding with the arbitrarily high-order two-dimensional method, we propose a variation of the above method here. Since it is a challenge to find exact or high-order approximate evolution updates for the point values for general systems of conservation laws and having in mind the point value updates of the method reviewed in Section 2.2.1, it seems straightforward to write a variation of the method from Abgrall and Barsukow [13], for $N + 1 \geq 4$, as the following semi-discretization:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \bar{q}_i(t) = -\dfrac{f(q_{i+\frac{1}{2}}(t)) - f(q_{i-\frac{1}{2}}(t))}{\Delta x} \\ \frac{\mathrm{d}}{\mathrm{d}t} q_{i+\frac{1}{2}}(t) = -F\Big(q_{i-\frac{1}{2}}(t), q_{i,\xi_0}(t), \ldots, q_{i,\xi_{N-3}}(t), q_{i+\frac{1}{2}}(t), \\ \qquad\qquad\qquad q_{i+1,\xi_0}(t), \ldots, q_{i+1,\xi_{N-3}}(t), q_{i+\frac{3}{2}}(t) \Big) \\ \frac{\mathrm{d}}{\mathrm{d}t} q_{i,\xi_a}(t) = -\tilde{F}_a\Big(q_{i-\frac{1}{2}}(t), q_{i,\xi_0}(t), \ldots, q_{i,\xi_{N-3}}(t), q_{i+\frac{1}{2}}(t) \Big) \\ \qquad\qquad\qquad \forall a \in \{0, \ldots, N-3\} \end{cases}$$

with consistent approximations $F, \tilde{F}_a$ of $\partial_x f(q)$ at $x_{i+\frac{1}{2}}$ and $x_i + \Delta x \xi_a$, respectively. As in Section 2.2.1, a Jacobian splitting can be considered and a finite difference formula can be derived from the reconstructed polynomial. For the point values at the cell interfaces it is natural to use the reconstruction of the cell in upwind direction such that $(D_x^+)_{i+\frac{1}{2}} q :=$

$\frac{\mathrm{d}}{\mathrm{d}x_{\mathrm{ref}}} q_{i,\mathrm{recon}}(x_{\mathrm{ref}})|_{x_{\mathrm{ref}}=\frac{\Delta x}{2}}, (D_x^-)_{i+\frac{1}{2}} q := \frac{\mathrm{d}}{\mathrm{d}x_{\mathrm{ref}}} q_{i+1,\mathrm{recon}}(x_{\mathrm{ref}})|_{x_{\mathrm{ref}}=-\frac{\Delta x}{2}}$. For the

point values within the cell, the reconstruction is uniquely defined and $(D_x)_{i,\xi_a} q := \frac{\mathrm{d}}{\mathrm{d}x_{\mathrm{ref}}} q_{i,\mathrm{recon}}(x_{\mathrm{ref}})|_{x_{\mathrm{ref}}=\Delta x \xi_a}$. This resembles the approximation of the derivative tangential to a cell interface in Section 2.1. The version described in this Section will serve as inspiration in Section 4.

3. A semi-discrete Active Flux method of arbitrarily high order on two-dimensional Cartesian grids

In the following Section we discuss the extension of the semi-discrete, third-order Active Flux method (as recapitulated in Section 2.1) to arbitrary order of accuracy $N + 1$. It will be constructed in such a way that

- the reconstructed numerical solution q_{recon} is globally continuous,
- it uses a compact stencil with a minimal number of degrees of freedom per cell, and
- the cell average is always included among the degrees of freedom.

As for the third-order method, the shared point values at the cell interfaces will guarantee that the reconstruction is globally continuous and the inclusion of the cell averages will ensure that the resulting method is conservative. To gain sufficient information for the reconstruction on a compact stencil, higher moments are defined. Hence, this work extends the hybrid finite element–finite volume approach (see Section 2.2.1) from one spatial dimension to two spatial dimensions on Cartesian grids. The way the point values are distributed on the edge has an influence on stability as shall be discussed in Section 4.1. There, similar considerations are made as for the Active Flux method with additional point values (see Section 2.2.2).

Remark: This approach to extend the generalized Active Flux method is only one possibility. It uses point values at the cell interfaces analogously to the third order method in 2-d and moments in the interior of the cell, extending the 1-d method from Section 2.2.1. Using moments in the interior gives a natural way to minimize the number of degrees of freedom. Another possibility can be a tensor-like extension of the 1-d method with moments which will also briefly be considered here. Further extensions can be thought of. For example the method described in Section 2.2.2 could be extended. A similar method considering point values only is the face-upwinded spectral element (FUSE) method [24] on quadrilateral meshes.

3.1. Reference element for Cartesian grids

From now on, we define the reference element as $C^{\mathrm{ref}} := \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$ and the transformation $\Phi_{ij} : C_{ij} \to C^{\mathrm{ref}}$

$$\Phi_{ij}(x, y) = \begin{pmatrix} \frac{x - x_i}{\Delta x} \\ \frac{y - y_j}{\Delta y} \end{pmatrix} \tag{5}$$

with its inverse $\Phi_{ij}^{-1} : C^{\mathrm{ref}} \to C_{ij}$

$$\Phi_{ij}^{-1}(x_{\mathrm{ref}}, y_{\mathrm{ref}}) = \begin{pmatrix} x_i + \Delta x x_{\mathrm{ref}} \\ y_j + \Delta y y_{\mathrm{ref}} \end{pmatrix}. \tag{6}$$
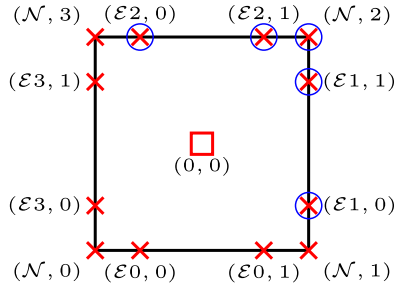
Fig. 2. Superscripts specifying the degrees of freedom for a cell C_{ij} for GenAF($M_\triangledown$) of fourth order (a generalized Active Flux method with minimal number of degrees of freedom, see Notation 2). The circled point values mark the points *belonging* to a cell.

3.2. Degrees of freedom

The degrees of freedom used for the generalized Active Flux method of order $N + 1$ are shared point values at the cell interfaces and moments. The notation we introduce here uses a subscripted index ij to identify the cell and a superscripted index to specify the degree of freedom. An example for the superscripts can be seen in Fig. 2.

As described above, the point values at the cell interfaces are chosen such that the reconstruction q_{recon} is globally continuous on the domain Ω and maximal use of these shared degrees of freedom is made. $N + 1$ shared point values at each cell interface are required to define a unique polynomial of degree N along each cell interface and achieve $q_{\text{recon}} \in C^0(\Omega, \mathbb{R}^s)$. First, the four nodes (= vertices) are included. This leaves $N - 1$ points per edge, i.e. $4(N - 1)$ edge points in total. For cell C_{ij} the point values are defined as

$$q_{ij}^{(p)}(t) := q\big(t, \mathbf{x}_{ij}^{(p)}\big) \quad \forall p \in I_{\mathcal{P}} \tag{7}$$

with $\mathbf{x}_{ij}^{(p)} \in \partial C_{ij}$. The index set

$$I_{\mathcal{P}} := I_{\mathcal{N}} \cup I_{\mathcal{E}} \tag{8}$$

defines all points *accessible* to the cell. These are

- the nodes $\{\mathbf{x}_{ij}^{(n)}\}_{n \in I_{\mathcal{N}}}$, $I_{\mathcal{N}} := \{(\mathcal{N}, 0), \dots, (\mathcal{N}, 3)\}$ located at

$$(x_{i \pm \frac{1}{2}}, y_{j \pm \frac{1}{2}})$$

and assembled counterclockwise starting with the lower left node and

- the edge points $\{\mathbf{x}_{ij}^{(e,a)}\}_{(e,a) \in I_{\mathcal{E}}}$, $I_{\mathcal{E}} := \{(e, a) | e = \mathcal{E}0, \dots, \mathcal{E}3, a = 0, \dots, N - 2\}$ with the edges assembled counterclockwise. The horizontal edge points with $e = 0, 2$ are located at

$$(x_{i + \xi_a}, y_{j \pm \frac{1}{2}}) = (x_i + \xi_a \Delta x, y_{j \pm \frac{1}{2}})$$

and the vertical edge points with $e = 1, 3$ at

$$(x_{i \pm \frac{1}{2}}, y_{j + \xi_a}) = (x_{i \pm \frac{1}{2}}, y_j + \xi_a \Delta y)$$

with $\xi_a \in (-\frac{1}{2}, \frac{1}{2})$. The choice of ξ_a is discussed in Section 4.1.

Since the point values are shared between adjacent cells it is possible to define the points *belonging* to a cell, e.g. chosen as the upper right node $n = 2$ as well as the edge points on the upper horizontal $e = \mathcal{E}2$ and the right vertical edge $e = \mathcal{E}1$, assembled in the index set

$$\widetilde{I_{\mathcal{P}}} := \{(\mathcal{N}, 2)\} \cup \{(e, a) | e = \mathcal{E}1, \mathcal{E}2, a = 0, \dots, N - 2\}. \tag{9}$$

Thus, on average only 1 node and $2(N - 1)$ edge points have to be updated per cell.

The moments of order $m = k + l$, $k, l \geq 0$ for cell C_{ij} are defined as

$$q_{ij}^{(k,l)}(t) := A_{k,l} \int_{C_{ij}} b_{k,l}(\Phi_{ij}(\mathbf{x})) q(t, \mathbf{x}) d\mathbf{x} \quad \forall (k, l) \in I_{\mathcal{M}} \tag{10}$$

with the test functions $(b_{k,l})_{k,l \in \mathbb{N}_0}$

$$b_{k,l} : \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \times \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \to \mathbb{R}, \quad b_{k,l}(x, y) = x^k y^l \tag{11}$$

and a corresponding normalization factor

$$A_{k,l} = \frac{(k + 1)2^k (l + 1)2^l}{\Delta x \Delta y}. \tag{12}$$

Thus, there are $m + 1$ moments of order m and the cell averages $\bar{q}_{ij}$ are equal to the moments $q_{ij}^{(0,0)}$, i.e. $m = 0$.

Beyond the requirement of including the average among the degrees of freedom, there are different possibilities to choose the moments. The two choices considered here are

$$I_{\mathcal{M}} = I_{\mathcal{M}, \square} := \{(k, l) | k, l \in \mathbb{N}_0, 0 \leq k, l \leq N - 2\} \quad \text{(tensor-like)} \tag{13}$$

$$I_{\mathcal{M}} = I_{\mathcal{M}, \triangle} := \{(k, l) | k, l \in \mathbb{N}_0, 0 \leq k + l \leq \max\{0, N - 4\}\} \quad \text{(triangle-like)} \tag{14}$$

for $N \geq 2$. They entail different reconstruction spaces which will be discussed in the following Section.

The point values and moments define all our degrees of freedom and the index set of the nodal and modal degrees of freedom accessible to each cell is denoted by

$$I_{\text{dof}} := I_{\mathcal{P}} \cup I_{\mathcal{M}}. \tag{15}$$

and analogously we denote by

$$\widetilde{I_{\text{dof}}} := \widetilde{I_{\mathcal{P}}} \cup I_{\mathcal{M}} \tag{16}$$

the index set of all degrees of freedom belonging to a cell. Fig. 3 shows some examples for the reference elements of spatial order $N + 1$ using the triangle-like moments (14).

The updates of the degrees of freedom are discussed in the Sections 3.4 and 3.5.

Remark: Finite elements using nodal and modal degrees of freedom are also called hybrid finite elements (see e.g. Ern and Guermond [25, Chapter 6.3.3, Chapter 7.6]).

3.3. Reconstruction

Next, the spatial reconstruction $q_{\text{recon}} : \Omega_h \to \mathbb{R}^s$, $\Omega_h := \cup_{i,j} C_{ij}$, is considered (for 1-d it reduces to the reconstruction in Abgrall and Barsukow [14], and for third order in 2-d to the one in Barsukow et al. [9]). The polynomial solution reconstruction on cell C_{ij}

$$q_{\text{recon}}|_{C_{ij}} : C_{ij} \to \mathbb{R}^s \qquad q_{\text{recon}}|_{C_{ij}} \in (P^{\text{recon}})^s \tag{17}$$

has to fulfill (7) and (10), and for the method to be of order $N + 1$ in space it must at least lie in the bivariate polynomial space

$$P^N := \text{span}\{x^m y^n | m, n \in \mathbb{N}_0, 0 \leq m + n \leq N\} \tag{18}$$

of maximal total degree N, for each element, i.e. $P^N \subset P^{\text{recon}}$. Here, we use all degrees of freedom $q_{ij}^{(r)}, r \in I_{\text{dof}}$ accessible to the cell to define (17). Clearly, a necessary condition for unisolvence is that the dimension $|P^{\text{recon}}|$ of P^{recon} matches the number of degrees of freedom. Below, two possible choices for the reconstruction space are presented. Later on, they are used to define a finite element for the generalized Active Flux method.

3.3.1. Tensor-like reconstruction space

As a natural extension from the one-dimensional case (see Abgrall and Barsukow [13,14]) the bi-polynomial space

$$P^{N,N} := \text{span}\{x^m y^n | m, n \in \{0, \cdots, N\}\} \tag{19}$$

can be considered for P^{recon}, similar to tensor-product finite elements (see e.g. Ern and Guermond [25, Chapter 6.4]).

This choice satisfies $P^N \subset P^{\text{recon}}$ and requires $|P^{\text{recon}}| = (N + 1)^2$ degrees of freedom. As discussed in Section 3.2, $4N$ of these must be point values which leaves $(N - 1)^2$ modal degrees of freedom within each
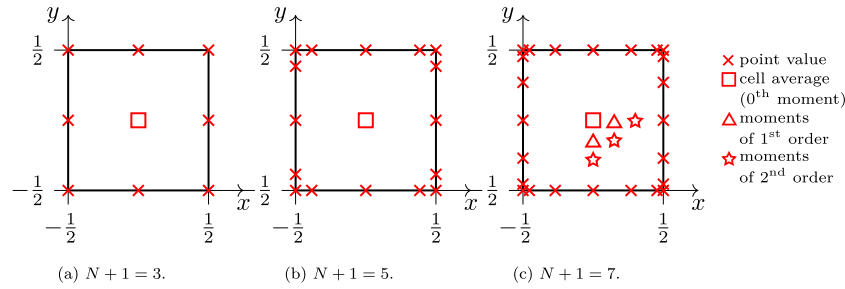
Fig. 3. Reference elements C^{ref} for GenAF($M_{\triangledown}$) of order $N+1$ (a generalized Active Flux method with minimal number of degrees of freedom, see Notation 2).

Table 1

Monomial basis for serendipity space S^N. The circled elements extend P^N. This is used to define a minimal basis for P^{recon}, see (20). The bold lines separate the basis elements required for the edges from the triangle-shaped remainder that ensures unisolvence of the moments.

1	x	x^2	$\dots$	x^{N-2}	x^{N-1}	x^N
y	xy	$x^2 y$	$\dots$	$x^{N-2}y$	$x^{N-1}y$	$\boxed{x^N y}$
y^2	xy^2	$x^2 y^2$	$\dots$	$x^{N-2}y^2$	$-$	$-$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$-$	$\ddots$	$\vdots$
y^{N-2}	xy^{N-2}	$x^2 y^{N-2}$	$-$	$\ddots$	$-$	$-$
y^{N-1}	xy^{N-1}	$-$	$\ddots$	$-$	$-$	$-$
y^N	$\boxed{xy^N}$	$-$	$\dots$	$-$	$-$	$-$

Table 2

Number of degrees of freedom for GenAF($M_{\triangledown}$).

order $N+1$	$\|P^{\mathrm{recon}}_{\min}\|$ $\dfrac{(N+1)(N+2)}{2}+2$	#PVs $4N$	#Moments $\max\left\{1, \frac{1}{2}(N-3)(N-2)\right\}$
3	8 ⊞+1	8	☐1
4	12 ⊞+1	12	☐1
5	17	16	1
6	23	20	3
7	30	24	6

cell, i.e. the cell average and the higher moments. Thus, the tensor-like reconstruction space is associated to the tensor-like choice (13) of moments spanning $P^{N-2,N-2}$. This choice reduces to the reconstruction space for the third-order Active Flux method as first introduced in Barsukow et al. [9], Helzel et al. [23], see also Section 2.1.

For the resulting method we introduce the following notation:

Notation 1. *The* tensor-like *generalized Active Flux method on two-dimensional Cartesian grids with $P^{\mathrm{recon}} = P^{N,N}$ shall be denoted by* GenAF($M_{\square}$).

3.3.2. Reconstruction space with minimal number of degrees of freedom

High-order methods need to evolve many degrees of freedom. We thus aim at finding a computationally more efficient choice for P^{recon} with only the minimal possible number of degrees of freedom needed to achieve order $N+1$ on a Cartesian cell.

Setting $P^{\mathrm{recon}} = P^N$ turns out not to be sufficient for unisolvence, as discussed below. Instead we define

$$P^{\mathrm{recon}} := P^{\mathrm{recon}}_{\min} := \begin{cases} S^N \oplus \mathrm{span}\{x^2 y^2\} & \text{for } N = 2,3 \\ S^N & \text{for } N \geq 4. \end{cases} \tag{20}$$

using the so called serendipity space (see also Table 1)

$$S^N := P^N \oplus \mathrm{span}\{x^N y, xy^N\} \tag{21}$$

known from serendipity finite elements (see e.g. Ern and Guermond [25, Chapter 6.4.3] and Arnold and Awanou [26]), which are based on the idea of reducing the degrees of freedom from the tensor-product finite element.

First, let us consider the case $N \geq 4$ for the reconstruction space in (20) with $|S^N| = \frac{1}{2}(N+1)(N+2) + 2 = \{8, 12, 17, 23, \dots\}_{N \geq 2}$ degrees of freedom. There must be $4N$ point values (see Section 3.2). This leaves us with $\frac{1}{2}(N-3)(N-2) = |P^{N-4}|$ moments, that we will thus take with respect to the basis of P^{N-4}. For $N = 2,3$ with $|S^N| = 4N$, there is an exception. In order to include the cell average, S^N needs to be extended. In both cases $\{x^2 y^2\}$ is a suitable choice for the additional basis element. This choice of moments corresponds to the triangle-like choice (14) and

coincides with the tensor-like method for $N+1 = 3$. An overview of the number of degrees of freedom can be found in Table 2.

As mentioned above, the choice $P^{\mathrm{recon}} = P^N$ (for $N \geq 5$) turned out not to be sufficient for unisolvence of the method on Cartesian cells, which was verified using MATHEMATICA [27] for $N = 5, 6$, where we chose $I_{\mathcal{M}} = I_{\mathcal{M},\triangle} \setminus \{(N-4,0),(0,N-4)\}$. This is because two basis elements of the highest power N in x and y each are needed to be able to uniquely define the polynomials up to degree N on the two horizontal and vertical edges, respectively. For $N = 2,3,4$ it was also necessary to extend P^N in a suitable manner due to our requirements specified at the beginning of Section 3. For $N = 2,3$, we chose $S^N \oplus \{x^2 y^2\}$ as defined in (20) which we deemed a natural and symmetric choice lying within $P^{N,N}$. For $N = 4$, we tried two versions: First, $P^N \oplus \{x^3 y^2, x^2 y^3\}$, which we found insufficient to achieve unisolvence and second, we tested S^N as in (20). Now, consider the minimal choice of degrees of freedom with the basis elements spanning $P^{\mathrm{recon}}_{\min}$. Then, one can think of associating the nodal degrees of freedom (nodes and edge points) on the horizontal edges with the basis elements $1, x, \dots, x^N$ and $y, xy, \dots, x^N y$ on an edge each, and respectively, $1, y, \dots, y^N$ and $x, xy, \dots, xy^N$ with the points on the vertical edges. The moments can then be associated with the basis elements in (14), which form a triangle in Table 1 (indicated by the bold lines). As shown with the help of MATHEMATICA [27] for up to order 7, the basis given in (20) for $P^{\mathrm{recon}}_{\min}$ is unisolvent.

Referring to the choice of moments the following notation is introduced:

Notation 2. *The generalized Active Flux method on two-dimensional Cartesian grids with* minimal number of degrees of freedom *using $P^{\mathrm{recon}}_{\min}$ given in (20) shall be denoted with* GenAF($M_{\triangledown}$).

3.3.3. A hybrid finite element

Next, the hybrid finite element as defined in Abgrall and Barsukow [14, Definition 3.1] is extended to a Cartesian element in two spatial dimensions. We also include the interpolation operator and the reconstruction and follow Ern and Guermond [25, Definitions 5.2, 5.5, 5.7, 5.11, Proposition 5.12] and Abgrall and Barsukow [14]:

Definition 1. *Let $d = 2$ and $N \geq 2$. Given an index set $I_{\mathrm{dof}} := I_P \cup I_{\mathcal{M}}$ with a set of points $\{\mathbf{x}_p\}_{p \in I_P}$, here $\{\mathbf{x}_p = (x_p, y_p)|x_p, y_p \in \{\pm\frac{1}{2}\}$ or $(x_p \in \{\xi_a\}_a, y_p \in \{\pm\frac{1}{2}\})$ or $(x_p \in \{\pm\frac{1}{2}\}, y_p \in \{\xi_a\}_a)$ with $\xi_a \in (-\frac{1}{2}, \frac{1}{2})$ $\forall a = 0, \dots, N-2\}$, and a (e.g. monomial) basis $\{b_{k,l}\}_{k,l \in I_{\mathcal{M}}}$ for either*
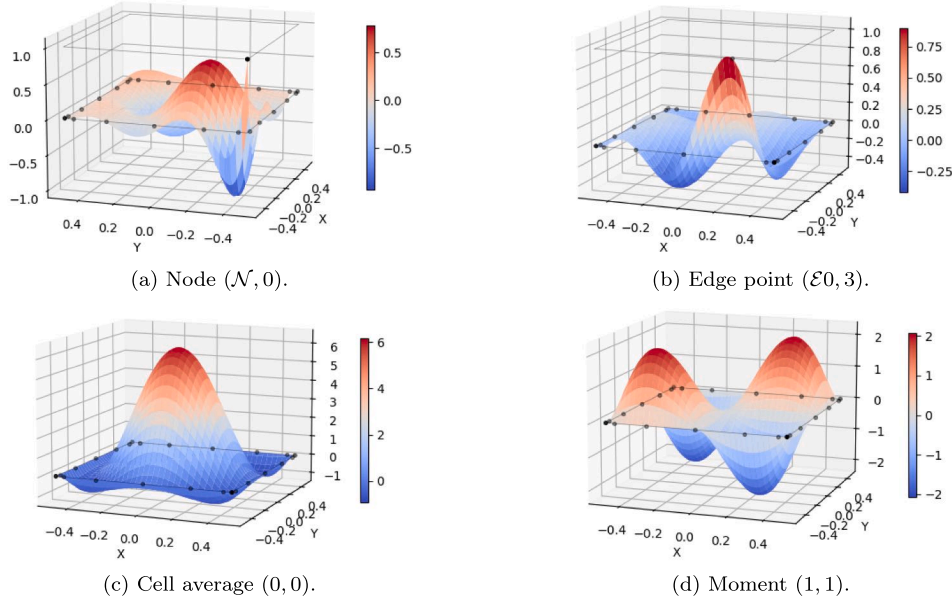
(a) Node $(\mathcal{N}, 0)$.

(b) Edge point $(\mathcal{E}0, 3)$.

(c) Cell average $(0, 0)$.

(d) Moment $(1, 1)$.

Fig. 4. Shape functions for GenAF($M_{\triangledown}$) of 7$^{\text{th}}$ order.

- $P^{N-2,N-2}$ ($I_{\mathcal{M}}$ as in (13)), or
- the polynomial vector space P^M with $M = \max\{0, N-4\}$ ($I_{\mathcal{M}}$ as in (14)).

  Then, the finite element (K, P, Σ) is

- $K = C^{\text{ref}} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]$,
- P given either by
  - $P^{\text{recon}} = P^{N,N}$, or
  - $P^{\text{recon}} = S^N \oplus \text{span}\{x^2 y^2\}$ for $N \leq 3$ and $P^{\text{recon}} = S^N$ for $N > 3$
  and
- $\Sigma = \{\sigma_i\}_{i \in I_{\text{dof}}}$ the set of degrees of freedom with $\sigma_i : P \to \mathbb{R}$

$\sigma_p(v) := v(\mathbf{x}_p) \quad \forall p \in I_P$

$\sigma_{k,l}(v) := A_{k,l}^{\text{ref}} \int_K b_{k,l}(\mathbf{x}) v(\mathbf{x}) d\mathbf{x} \quad \forall (k,l) \in I_{\mathcal{M}}$

for all $v \in P$ and with the normalization factors $A_{k,l}^{\text{ref}} = (k+1)2^k(l+1)2^l$ with respect to K.

We define the shape functions $B_r : K \to \mathbb{R}$ such that

$$\sigma_r(B_s) = \delta_{rs} \quad \forall r, s \in I_{\text{dof}} \tag{22}$$

which form the canonical basis of P with respect to the finite element. Fig. 4 shows examples for the shape functions of the GenAF($M_{\triangledown}$) method of order $N + 1 = 7$.

The interpolation operator corresponding to the finite element is defined to interpolate real-valued functions over K in P. We consider the space $V := L^1(K, \mathbb{R})$. With $P \subset V$, the elements in $\{\sigma_r\}_{r \in I_{\text{dof}}}$ – spanning the space of linear forms $\hom(P, \mathbb{R})$ – can naturally be extended to $\hom(V, \mathbb{R})$. We denote them by $\{\tilde{\sigma}_r\}_{r \in I_{\text{dof}}}$. Then, the interpolation operator $I_K : V \to P$ is defined as

$$I_K(v)(\mathbf{x}) := \sum_{r \in I_{\text{dof}}} \tilde{\sigma}_r(v) B_r(\mathbf{x}) \quad \forall \mathbf{x} \in K. \tag{23}$$

Finally, the reconstruction $R : \mathbb{R}^{|I_{\text{dof}}| \cdot s} \to P^s$,

$$R\left((q_r)_{r \in I_{\text{dof}}}\right)(\mathbf{x}) := \sum_{r \in I_{\text{dof}}} q_r B_r(\mathbf{x}) \quad \forall \mathbf{x} \in K \tag{24}$$

maps the degrees of freedom $q_r \in I_{\text{dof}}$ to the polynomial space P^s. This gives the reconstruction (17):

$$q_{\text{recon}}|_{C_{ij}} \circ \Phi_{ij}^{-1} = R\left(\left(q_{ij}^{(r)}\right)_{r \in I_{\text{dof}}}\right) \in (P^{\text{recon}})^s \tag{25}$$

$$\Leftrightarrow \quad q_{\text{recon}}|_{C_{ij}}(\mathbf{x}) = \sum_{r \in I_{\text{dof}}} q_{ij}^{(r)} B_r(\Phi_{ij}(\mathbf{x})). \tag{26}$$

3.4. Update of point values

For the update of the point values it is possible to consider a non-conservative formulation of (1), which also allows to include stabilizing upwinding. In spirit we follow Abgrall and Barsukow [13,14], Abgrall et al. [15]. The quasi-linear form for the system of conservation laws (1) is

$$\partial_t q(t, \mathbf{x}) + Df^x(q(t, \mathbf{x})) \partial_x q(t, \mathbf{x}) + Df^y(q(t, \mathbf{x})) \partial_y q(t, \mathbf{x}) = 0, \tag{27}$$

which is considered at $\mathbf{x}_{ij}^{(p)}$ to derive the point value updates. Here, the Jacobians of the fluxes f^x, f^y are given, since they can be calculated directly, whereas the derivatives of q with respect to x and y have to be approximated. We intend to use different derivative approximations $(D_x)_{ij}^{(p)} \simeq \partial_x, (D_y)_{ij}^{(p)} \simeq \partial_y$ for each upwind direction. To this end, the diagonalized Jacobians

$$Df^x(q_{ij}^{(p)}) = T^x \text{diag}(\lambda_{x,1}, \ldots, \lambda_{x,s})(T^x)^{-1}, \tag{28}$$

$$Df^y(q_{ij}^{(p)}) = T^y \text{diag}(\lambda_{y,1}, \ldots, \lambda_{y,s})(T^y)^{-1} \tag{29}$$

are split according to the positive and negative wave speeds to incorporate upwinding:

$$Df^x(q_{ij}^{(p)})^+ := T^x \text{diag}(\lambda_{x,1}^+, \ldots, \lambda_{x,s}^+)(T^x)^{-1},$$

$$Df^x(q_{ij}^{(p)})^- := T^x \text{diag}(\lambda_{x,1}^-, \ldots, \lambda_{x,s}^-)(T^x)^{-1},$$

$$Df^y(q_{ij}^{(p)})^+ := T^y \text{diag}(\lambda_{y,1}^+, \ldots, \lambda_{y,s}^+)(T^y)^{-1},$$

$$Df^y(q_{ij}^{(p)})^- := T^y \text{diag}(\lambda_{y,1}^-, \ldots, \lambda_{y,s}^-)(T^y)^{-1},$$

where

$$\lambda_{*,i}^+ := \max(0, \lambda_{*,i}), \qquad \lambda_{*,i}^- := \min(0, \lambda_{*,i}), \quad *= x, y.$$

This yields

$$\frac{d}{dt} q_{ij}^{(p)}(t) = -Df^x(q_{ij}^{(p)})^+ (D_x^+)_{ij}^{(p)} q - Df^x(q_{ij}^{(p)})^- (D_x^-)_{ij}^{(p)} q$$
$$- Df^y(q_{ij}^{(p)})^+ (D_y^+)_{ij}^{(p)} q - Df^y(q_{ij}^{(p)})^- (D_y^-)_{ij}^{(p)} q. \tag{30}$$

The upwinded finite difference formulas $(D_x^\pm)_{ij}^{(p)}, (D_y^\pm)_{ij}^{(p)}$ are derived with the help of the reconstructions $q_{\text{recon}}|_{C_{ij}}$, which are differentiable on each

cell C_{ij} (see also Abgrall et al. [15], Barsukow et al. [22] for third order):

$$(D_x^\pm)_{ij}^{(p)} q := \partial_x q_{\text{recon}}\big|_{C^{\text{upw}}}(\mathbf{x})\big|_{\mathbf{x}_{ij}^{(p)}}, \tag{31}$$

$$(D_y^\pm)_{ij}^{(p)} q := \partial_y q_{\text{recon}}\big|_{C^{\text{upw}}}(\mathbf{x})\big|_{\mathbf{x}_{ij}^{(p)}}. \tag{32}$$

The upwind cell C^{upw} is the cell adjacent to $\mathbf{x}_{ij}^{(p)}$ from the corresponding "$\pm$"-direction for x or y, i.e. at the node $\mathbf{x}_{ij}^{(p)} = (x_{i+\frac{1}{2}}, y_{j+\frac{1}{2}})$

$$(D_x^+)_{ij}^{(p)} q := \partial_x q_{\text{recon}}\big|_{C_{ij}}(x, y_{j+\frac{1}{2}})|_{x=x_{i+\frac{1}{2}}},$$

$$(D_x^-)_{ij}^{(p)} q := \partial_x q_{\text{recon}}\big|_{C_{i+1,j}}(x, y_{j+\frac{1}{2}})|_{x=x_{i-\frac{1}{2}}},$$

$$(D_y^+)_{ij}^{(p)} q := \partial_y q_{\text{recon}}\big|_{C_{ij}}(x_{i+\frac{1}{2}}, y)|_{y=y_{j+\frac{1}{2}}},$$

$$(D_y^-)_{ij}^{(p)} q := \partial_y q_{\text{recon}}\big|_{C_{i,j+1}}(x_{i+\frac{1}{2}}, y)|_{y=y_{j-\frac{1}{2}}},$$

and at the horizontal edge points $(x_{i+\xi_a}, y_{j+\frac{1}{2}})$

$$(D_x)_{ij}^{(p)} q = (D_x^+)_{ij}^{(p)} q = (D_x^-)_{ij}^{(p)} = \partial_x q_{\text{recon}}\big|_{C_{ij}}(x, y_{j+\frac{1}{2}})|_{x=x_{i+\xi_a}},$$

$$(D_y^+)_{ij}^{(p)} q := \partial_y q_{\text{recon}}\big|_{C_{ij}}(x, y)|_{(x,y)=(x_{i+\xi_a}, y_{j+\frac{1}{2}})},$$

$$(D_y^-)_{ij}^{(p)} q := \partial_y q_{\text{recon}}\big|_{C_{i,j+1}}(x, y)|_{(x,y)=(x_{i+\xi_a}, y_{j+\frac{1}{2}})}.$$

The derivatives for the vertical edge points $(x_{i+\frac{1}{2}}, y_{j+\xi_a})$ are defined analogously. Considering definition (26) for the reconstruction $q_{\text{recon}}\big|_{C_{ij}}$ the partial derivatives in x- and y-direction are given as

$$\partial_x q_{\text{recon}}\big|_{C_{ij}}(x, y) = \sum_{r \in I_{\text{dof}}} q_{ij}^{(r)} \nabla_\Phi B_r(\Phi_{ij}(x, y)) \cdot \partial_x \Phi_{ij}(x, y), \tag{33}$$

$$\partial_y q_{\text{recon}}\big|_{C_{ij}}(x, y) = \sum_{r \in I_{\text{dof}}} q_{ij}^{(r)} \nabla_\Phi B_r(\Phi_{ij}(x, y)) \cdot \partial_y \Phi_{ij}(x, y). \tag{34}$$

An example for the thus derived finite difference formulas for (31) and (32) can be seen in Fig. 5.

3.5. Update of moments

In the following, an update procedure is derived for the moments, extending [13–15]. In particular, a conservative update of the cell averages is obtained, which is key to ensure convergence to the weak solution of the conservation law. To this end, the weak formulation of (1)

$$\frac{d}{dt} \int_C v(\mathbf{x}) q(t, \mathbf{x}) dC + \int_{\partial C} v(\mathbf{x}) \mathbf{f}(q(t, \mathbf{x})) \cdot \mathbf{n} dS$$

$$- \int_C \nabla v(\mathbf{x}) \cdot \mathbf{f}(q(t, \mathbf{x})) d\mathbf{x} = 0 \tag{35}$$

on a domain $C \subset \mathbb{R}^d$, is considered. To find the update for the cell average and higher moments on cell C_{ij}, Eq. (35) with $C := C_{ij}$ is multiplied by the normalization factor (12) and the test function v is set to (11). This yields the semi-discrete update formula for the moments of order $m = k + l$

$$\frac{d}{dt} q_{ij}^{(k,l)}(t) = -A_{k,l} \left(\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} b_{k,l}(\Phi_{ij}(x_{i+\frac{1}{2}}, y)) f^x(q(t, x_{i+\frac{1}{2}}, y)) \right.$$

$$- b_{k,l}(\Phi_{ij}(x_{i-\frac{1}{2}}, y)) f^x(q(t, x_{i-\frac{1}{2}}, y)) dy$$

$$- \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \partial_x b_{k,l}(\Phi_{ij}(x, y)) f^x(q(t, x, y)) dx dy$$

$$+ \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} b_{k,l}(\Phi_{ij}(x, y_{j+\frac{1}{2}})) f^y(q(t, x, y_{j+\frac{1}{2}}))$$

$$- b_{k,l}(\Phi_{ij}(x, y_{j-\frac{1}{2}})) f^y(q(t, x, y_{j-\frac{1}{2}})) dx$$

$$\left. - \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \partial_y b_{k,l}(\Phi_{ij}(x, y)) f^y(q(t, x, y)) dx dy \right), \tag{36}$$

where the integrals are approximated with the help of a numerical quadrature formula. Here, we choose either a Gauss-Lobatto or a Gauss-Legendre quadrature for the integrals. The necessary values at the quadrature points (both on the cell boundaries and within the cells) are obtained by evaluating the reconstruction (26). For now, the quadrature formulas are chosen to be sufficient for polynomials of at least degree $N + \max(k, l)$.

Remark: The nodal degrees of freedom at the cell interfaces (nodes and edge points) could be included to calculate the quadratures. Yet, along the edge there are typically more point values than actually necessary for the quadrature alone. Here, we try to avoid this overintegration and the potentially larger computational effort. Furthermore, for the bulk integral, the reconstruction needs to be evaluated at quadrature points anyway as the pointwise degrees of freedom are located on cell interfaces only.

3.6. Method definition

Concluding, a definition is given of the generalized Active Flux method of arbitrarily high order on Cartesian grids in 2-d.

Definition 2. *Given $N \geq 2$, the* generalized Active Flux method *of order $N + 1$ on 2-d Cartesian grids is the semi-discretization of* (1)

$$\begin{cases} \dfrac{d}{dt} q_{ij}^{(k,l)}(t) = -A_{k,l} \displaystyle\int_{\partial C_{ij}} b_{k,l}(\Phi_{ij}(\mathbf{x})) \mathbf{f}(q(t, \mathbf{x})) \cdot \mathbf{n}_{ij} dS \\ \qquad\qquad + A_{k,l} \displaystyle\int_{C_{ij}} \nabla_{\mathbf{x}} b_{k,l}(\Phi_{ij}(\mathbf{x})) \cdot \mathbf{f}(q(t, \mathbf{x})) d\mathbf{x} \quad \forall (k, l) \in I_{\mathcal{M}} \\ \dfrac{d}{dt} q_{ij}^{(p)}(t) = -F_p\left(\left(q_{i_p, j_p}^{(r)} \right)_{(i_p, j_p) \in \sigma_p, r \in I_{\text{dof}}} \right) \quad \forall p \in \widetilde{I_P} \end{cases}$$

with I_{dof}, $I_{\mathcal{M}}$, $\{b_{k,l}\}$ and $A_{k,l}^{\text{ref}}$ given as in Definition 1, $A_{k,l} = \dfrac{A_{k,l}^{\text{ref}}}{\Delta x \Delta y}$ and the outward pointing unit normal $\mathbf{n}_{ij}$ on ∂C_{ij}. The set $\{\mathbf{x}_{ij}^{(p)}\}_{p \in \widetilde{I_P}}$ contains point values belonging to C_{ij} as it is given by (9). *F_p is a consistent approximation of $\partial_x f^x(q) + \partial_y f^y(q)$ at point $\mathbf{x}_{ij}^{(p)}$ where $\sigma_p := \{(i, j) \mid \mathbf{x}_{ij}^{(p)} \in \partial C_{ij}\}$ defines the index set of the cell neighbours of $\mathbf{x}_{ij}^{(p)}$.*

Here, we choose F_p as given in (30) with finite difference formulas (31), (32) derived in Section 3.4.

4. Stability analysis

Stability results for the generalized Active Flux method introduced in the previous Section are presented for linear advection in two spatial dimensions. First, the semi-discrete generalized Active Flux method is studied with the help of an eigenvalue spectrum analysis. We observe that the stability of the method depends on the location of the edge points, and this analysis is used to determine a stable distribution. Second, a generalized Active Flux method using a third-order Runge-Kutta method for the discretization in time is considered to determine a stability bound for the fully discretized system. The methods used for the stability analysis are for example described in Lomax et al. [28].

4.1. Semi-discrete Active Flux method

The generalized Active Flux method of order $N + 1$ uses moments and point values at the cell interfaces as degrees of freedom. In the one-dimensional case, the choice of the points is straightforward at the cell interfaces. The two-dimensional case allows more flexibility: out of $N + 1$ points $\mathbf{x}_{ij}^{(p)}$ along each cell interface, two are located at the nodes, (see Section 3.2), but the question how the $N - 1$ edge points are distributed remains.

Some first numerical tests with a uniform distribution of the edge points implied that this choice does not lead to a stable method. In Abgrall and Barsukow [13] a similar finding was observed for the arbitrarily high order one-dimensional Active Flux method with additional
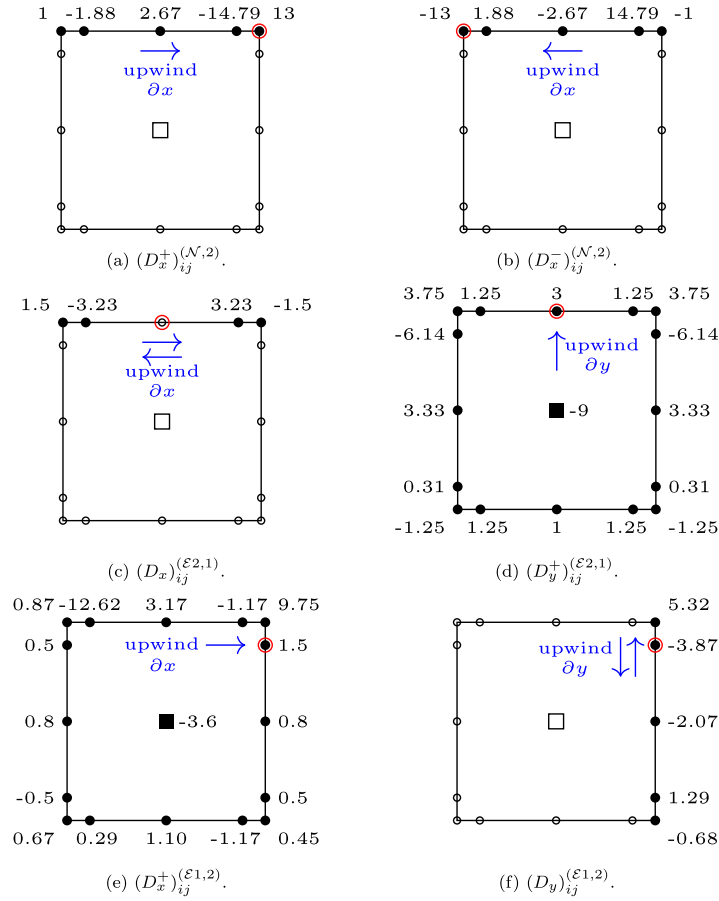
Fig. 5. Example of finite difference formulas for GenAF($M_\triangledown$) of fifth order at locations of different degrees of freedom, marked by the additional circle (see (31), (32); coefficients rounded to two decimal points). Upwind direction: "→".

point values, reviewed in Section 2.2.2. There, the additional points set within the cells had to be moved closer to the cell interfaces to gain stability. Inspired by this result we would like to find a similar way to distribute the edge points for the generalized Active Flux method in 2-d. Here, the focus is on symmetric distributions of the points with respect to the edge midpoint. For a fourth and fifth order method this leaves one free parameter for the edge point distribution and two for a sixth and seventh order method: Considering the edge $[-\frac{1}{2}, \frac{1}{2}]$ in x- or y-direction the edge points are

- $\{-\xi, \xi\}$ for $N + 1 = 4$ or $\{-\xi, 0, \xi\}$ for $N + 1 = 5$, with $\xi \in (0, \frac{1}{2})$ and
- $\{-\xi_1, -\xi_0, \xi_0, \xi_1\}$ for $N + 1 = 6$ or $\{-\xi_1, -\xi_0, 0, \xi_0, \xi_1\}$ for $N + 1 = 7$ with $\xi_0, \xi_1 \in (0, \frac{1}{2})$, $\xi_0 < \xi_1$

While it might still be viable to analyze many different setups for a fourth or fifth order method, for higher orders $N + 1$ with $\lfloor \frac{N-1}{2} \rfloor$ free parameters for the edge point distribution this is not feasible. Hence, the following stability analysis is confined to a few ways to distribute the edge points that we consider to be a sensible choice. In particular we will try Gauss and Gauss-Lobatto points. Even with Gauss points, we still include point values at the nodes in the set of our degrees of freedom. Although this choice is inspired by quadrature points, this choice is generally independent of the actual quadrature, see also Section 3.5. Also the number of point values is fixed with respect to the order of the method (see Section 3.2).

Next, the eigenvalue spectrum of the semi-discrete generalized Active Flux method is analyzed for scalar linear advection in two spatial dimensions

$$\partial_t q(t, \mathbf{x}) + \nabla \cdot (\mathbf{a}q(t, \mathbf{x})) = 0 \tag{37}$$

with a constant advection speed $\mathbf{a} = (a_x, a_y) := (\cos \theta, \sin \theta)^T$, $\theta = [0, 2\pi)$ and periodic boundary conditions. Thereto, the vector $\mathbf{q} = (q_i)_{i \in \mathbf{I}_{\text{dofs}}}$ with $\mathbf{I}_{\text{dofs}} = \bigcup_{i,j} (\widetilde{I_{\text{dof}}})_{ij}$ assembling all degrees of freedom of the discretized domain Ω_h is defined and the semi-discrete generalized Active Flux method for (37) is re-written as the linear ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{q}(t) = A\mathbf{q}(t), \tag{38}$$

where A assembles the right hand side updates of the method. A necessary condition for stability of the linear ODE system (38) is

$$\text{Re}(\lambda_i) \leq 0 \qquad \forall \lambda_i \in \sigma(A). \tag{39}$$

This condition becomes sufficient, if the algebraic multiplicity of all λ_i with $\text{Re}(\lambda_i) = 0$ equals their geometric multiplicity.

In the following, the GenAF($M_\triangledown$) method of orders $N + 1 = 3, \ldots, 7$ is studied with Gauss, Gauss-Lobatto and uniformly distributed points on the edges. To cover a range of different advection directions, $\theta \in [0, \frac{\pi}{2}]$ is chosen in increments of $\frac{1}{32}$-th, i.e. $\theta = 0, \frac{\pi}{64}, \ldots, \frac{\pi}{2}$, for the Gauss points and $\frac{1}{4}$-th for Gauss-Lobatto and uniformly distributed points. The restriction of θ to $[0, \frac{\pi}{2}]$ is possible due to symmetry properties of the method. The setup for the grid sizes is $N_x = N_y \in \{3, 5, 10\}$. The results use a Gauss integration sufficient for the calculation of the moments for linear advection, i.e. for moments of order m it is chosen to be sufficient for polynomials of at least degree $N + \max(k, l)$.
Remark: Some further tests with a higher accuracy or different type of quadrature yielded comparable results.

For the setups described above, the stability of (38) is analyzed numerically due to the size of the update matrices A (e.g. for seventh
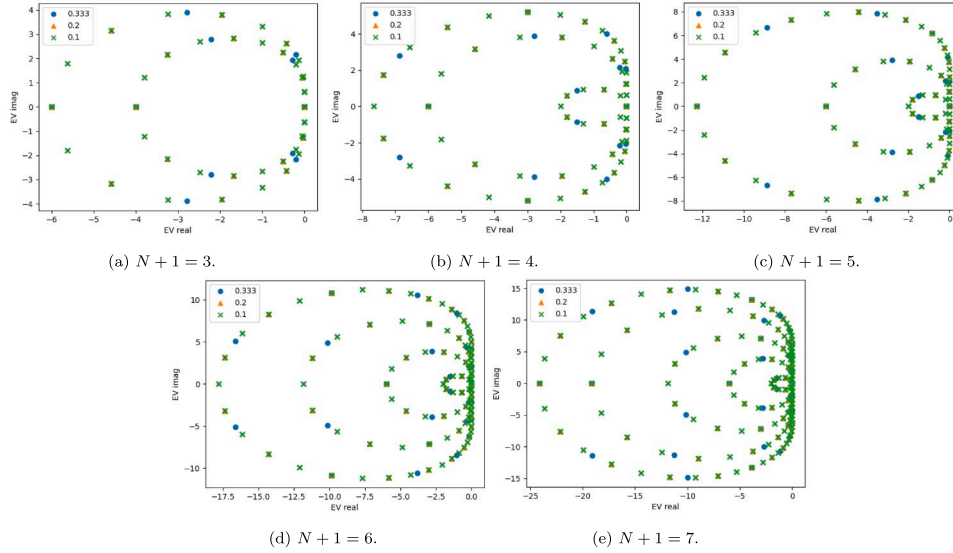
Fig. 6. Scaled $\lambda_i \in \sigma(A)$, i.e. $h\lambda_i$, for GenAF($M_{\triangledown}$) of order $N + 1$ with Gauss edge point distribution. $\theta = 0$ and $h = \Delta x = \Delta y = 0.1, 0.2, \frac{1}{3}$.

order, 10×10 grid: 1700×1700).[1] We obtain the following results: only the method with Gauss edge point distribution is stable for all tested orders $N + 1 = 3, \ldots, 7$ and corresponding setups. From the computer-aided analysis one finds that $\mathrm{Re}(\lambda_i) < \epsilon$, i.e. the eigenvalues have a non-postive real part up to a tolerance ϵ, and all λ_i with $|\mathrm{Re}(\lambda_i)| < \epsilon$ are at least semisimple for all i with tolerance $\epsilon = 5 \cdot 10^{-13}$ for $N + 1 = 3, 4, 5$, $\epsilon = 1 \cdot 10^{-12}$ for $N + 1 = 6$ and $\epsilon = 5 \cdot 10^{-12}$ for $N + 1 = 7$ for the Gauss edge point distribution. In Fig. 6 an example is shown for $N + 1 = 3, \ldots, 7$ with $\theta = 0$ and the three grid sizes. It can also be seen that the eigenvalues scale with the grid size.

The methods with uniform and with Gauss-Lobatto edge point distribution result in an unstable system (38) for orders $N + 1 = 4, \ldots, 7$. For $N + 1 = 5, 6, 7$ the tests show that there exist eigenvalues with positive real part for all tested θ. For $N + 1 = 4$, eigenvalues with positive real part are observed for $\theta = 0, \frac{\pi}{2}$. For $N + 1 = 3$ the edge point distribution coincides with the Gauss distribution.

In this way, we are led to conclude that Gauss points are a suitable choice for the distribution of the edge points that yield a stable method and we will use this distribution further on. Despite the limitations of the analysis, which has been conducted for orders $N + 1 = 3, \ldots, 7$ and linear advection only, we are optimistic that these results will carry over to higher orders. Furthermore, we apply the Gauss edge point distribution to linear and non-linear systems of conservation laws in our numerical examples and have not observed instabilities.

Remark: The semi-discrete GenAF($M_{\square}$) method of orders $N + 1 = 4$ and 5 was also studied for Gauss, Gauss-Lobatto and uniformly distributed edge points. The setup used for the advection directions is $\theta \in [0, \frac{\pi}{2}], \theta = 0, \frac{\pi}{16}, \ldots, \frac{\pi}{2}$ and the grid sizes are $N_x = N_y = 3, 5, 10$. Again, only the Gauss edge point distribution yields a stable system (38) for all setups. For Gauss-Lobatto and uniformly distributed points, positive eigenvalues are found for $\theta = 0, \frac{\pi}{2}$.

4.2. Fully discrete Active Flux method

The semi-discrete formulation of the method (see Definition 2) allows to choose a time discretization. Similarly as in Abgrall and Bar-

sukow [13,14] we focus on a strong stability preserving Runge-Kutta method of order three (SSP-RK3).

Based on the approach in Section 4.1 we analyze the stability of a fully discretized generalized Active Flux method with Gauss edge point distribution. This allows us to derive a CFL condition for linear advection (37). First, the ODE system (38) of the semi-discrete method for linear advection with periodic boundary conditions is diagonalized. This is possible, because the dimension of the sum of the eigenspaces for all eigenvalues of A is equal to $|\mathbf{I}_{\mathrm{dofs}}|$, which was computationally checked up to a tolerance $\epsilon = 5 \cdot 10^{-13}$ for $N + 1 = 3, 4, 5$, $\epsilon = 1 \cdot 10^{-12}$ for $N + 1 = 6$ and $\epsilon = 5 \cdot 10^{-12}$ for $N + 1 = 7$. The diagonalized system for $\hat{\mathbf{q}} = R^{-1}\mathbf{q}$ is given as

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathbf{q}}(t) = \Lambda\hat{\mathbf{q}}(t) \tag{40}$$

with $A = R\Lambda R^{-1}$, $\Lambda = \mathrm{diag}((\lambda_i)_{i\in\mathbf{I}_{\mathrm{dofs}}})$. Applying the RK3 method with a time step $\Delta t = t_{n+1} - t_n$ to (40) yields the fully discretized system

$$\hat{\mathbf{q}}_i^{n+1} = G(\lambda_i \Delta t)\hat{\mathbf{q}}_i^n \qquad \forall i \in \mathbf{I}_{\mathrm{dofs}} \tag{41}$$

with $G(z) := 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3$ and the stability domain $S = \{z \in \mathbb{C} | |G(z)| \le 1\}$. From this it is possible to find a maximal time step $\Delta t_{\max}(\theta)$, dependent on the advection direction, such that $\lambda_i \Delta t_{\max}(\theta) \in S$ for all $i \in \mathbf{I}_{\mathrm{dofs}}$. For our computer-based study, we increment the time step by 10^{-4} for $N + 1 = 3, 4, 5$ and $5 \cdot 10^{-5}$ for $N + 1 = 6, 7$ to approximate $\Delta t_{\max}$. Fig. 7 shows an example for the GenAF($M_{\triangledown}$) method of spatial order $N + 1 = 3, \ldots, 7$ for $N_x = N_y = 10$. The scaled eigenvalues $\lambda_i \Delta t_{\max}(\frac{\pi}{4})$ are plotted for the approximated maximal time step that still allows them to fit inside the stability domain for RK3 up to a tolerance ϵ.

The stable regions for the fully discretized method of spatial order $N + 1 = 3, \ldots, 7$ are plotted as function of the CFLs $\mathrm{cfl}_x = \frac{a_x \Delta t}{\Delta x}$ and $\mathrm{cfl}_y = \frac{a_y \Delta t}{\Delta y}$ in x- and y-direction 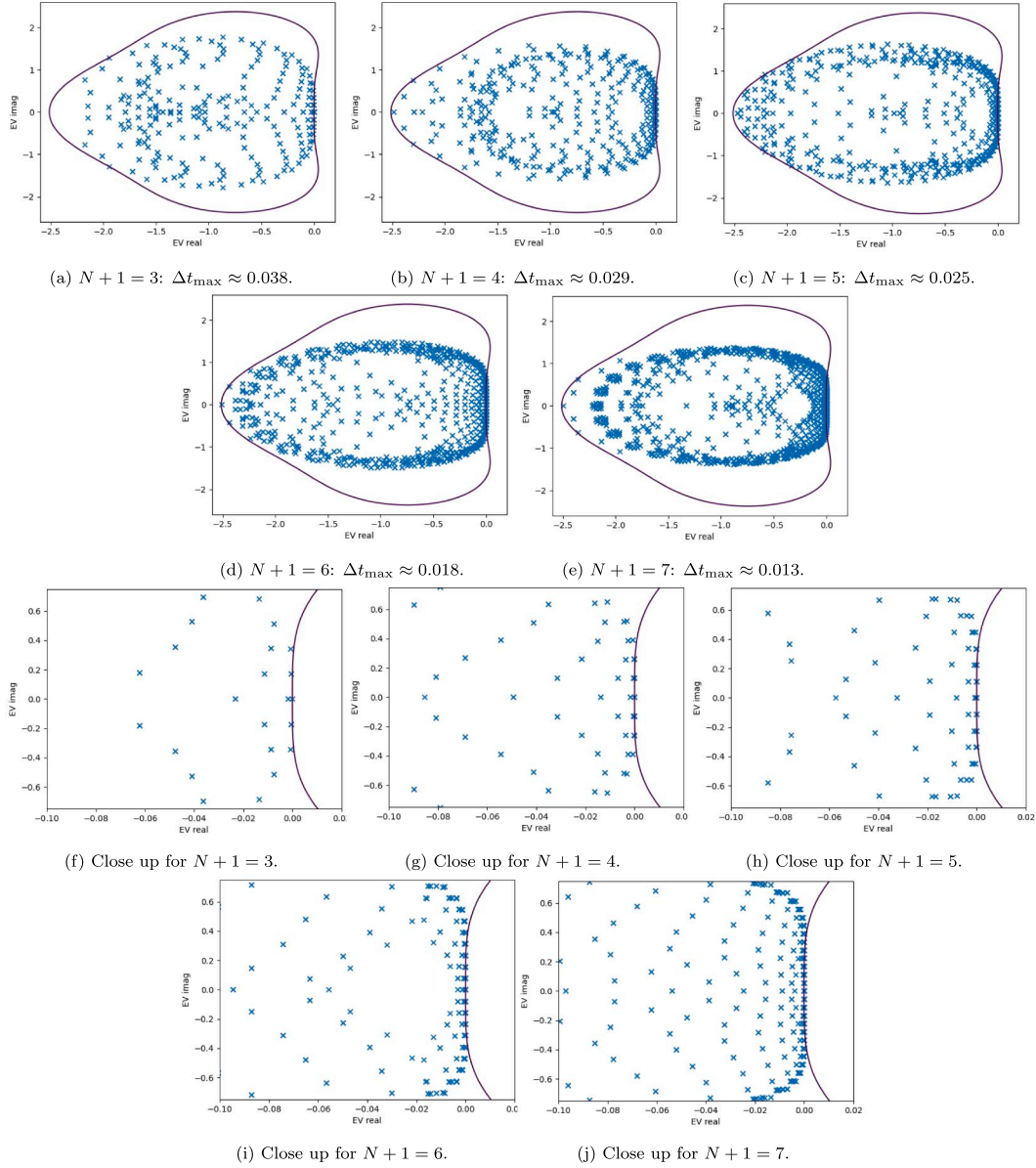for $h = \Delta x = \Delta y = 0.1$ in Fig. 8. One observes that the time step depends not only on the magnitude, but also the direction of the advection, and that the nature of this dependence changes as the order increases. This is studied below in more detail. In order to compare the CFL constraints for different orders of accuracy in a simple way, we consider the maximal CFL at $a = a_x = a_y$

$$C_{\mathrm{cfl}} := \frac{|a|\Delta t_{\max}}{h} \tag{42}$$

given in Table 3 for the GenAF($M_{\triangledown}$) method discretized with SSP-RK3 in time, $h = 0.1$.

For other grid sizes h the results are comparable, because the eigenvalues λ_i scale with $\frac{1}{h}$ (see Section 4.1).

---

[1] The matrix is assembled using an implementation of the method in C++ including the package EIGEN [29]. The actual analysis is done in Python using the NUMPY package [30] (v2.3) with its linear algebra library (LINALG). For plotting we rely on the MATPLOTLIB [31] (v3.10).

(a) $N + 1 = 3$: $\Delta t_{\max} \approx 0.038$. (b) $N + 1 = 4$: $\Delta t_{\max} \approx 0.029$. (c) $N + 1 = 5$: $\Delta t_{\max} \approx 0.025$.

(d) $N + 1 = 6$: $\Delta t_{\max} \approx 0.018$. (e) $N + 1 = 7$: $\Delta t_{\max} \approx 0.013$.

(f) Close up for $N + 1 = 3$. (g) Close up for $N + 1 = 4$. (h) Close up for $N + 1 = 5$.

(i) Close up for $N + 1 = 6$. (j) Close up for $N + 1 = 7$.

Fig. 7. Stability domain S_{RK3} (—) and scaled eigenvalues $\lambda_i \Delta t_{\max}$ ($\times$) for GenAF($M_\triangledown$) of order $N + 1$ in space for linear advection with $\theta = \frac{\pi}{4}$, and $\Delta x = \Delta y = 0.1$. Including close ups around origin.

Table 3
C_{cfl} for GenAF($M_\triangledown$) of order $N + 1$ in space for linear advection and RK3 discretization in time. (Two significant digits of C_{cfl} are shown for $h = 0.1$).

order ($N + 1$)	3	4	5	6	7
C_{cfl} (RK3)	0.27	0.20	0.17	0.12	0.088

The stability domain for the third-order method (with $M_\triangledown = M_\square$) can approximately be described by

$$S_{\text{circ}} = \left\{ (\text{cfl}_x, \text{cfl}_y) | \text{cfl}_x^2 + \text{cfl}_y^2 \leq r_{\text{cfl}}^2 \right\} \tag{43}$$

with $r_{\text{cfl}} := \sqrt{2} C_{\text{cfl}}$ and gives the following bound for the time step size:

$$\Delta t \leq \frac{r_{\text{cfl}} \min\{\Delta x, \Delta y\}}{\sqrt{a_x^2 + a_y^2}}. \tag{44}$$

For $N + 1 = 4$ this is still a useful approximation while for higher spatial orders of the method the domain seems more restricted. Here,

$$S_{\text{sq}} = \left\{ (\text{cfl}_x, \text{cfl}_y) | \max\{\text{cfl}_x, \text{cfl}_y\} \leq C_{\text{cfl}} \right\} \tag{45}$$

could be considered which gives the bound

$$\Delta t \leq \frac{C_{\text{cfl}} \min\{\Delta x, \Delta y\}}{\max\{a_x, a_y\}}. \tag{46}$$

Remark: For $N + 1 = 5$, GenAF($M_\triangledown$) almost yields a square. Looking more closely at the results in Fig. 8 this already shows for $N + 1 = 4$, and for $N + 1 = 6, 7$ the corner starts to round off again. Comparing these results one can notice, that for GenAF($M_\triangledown$) with $N + 1 = 4, 5$ only point values at the cell interfaces are added as further degrees of freedom, seemingly favouring the maximal advection speed. Including higher moments for $N + 1 \geq 6$ inside the cell decreases this effect.

Remark: The GenAF($M_\square$) method with Gauss edge point distribution and RK3 for the time discretization was analyzed for $N + 1 = 4, 5$. The stable regions $S_{\text{GenAF}(M_\square)}|_{N+1=3,4,5}$ can be approximated by Eq. (43), see Fig. 9. The bounds r_{cfl} are found to be ~ 0.2 ($C_{\text{cfl}} \approx 0.14$) for $N + 1 = 4$, and ~ 0.125 ($C_{\text{cfl}} \approx 0.089$) for $N + 1 = 5$. Although C_{cfl} is smaller than with
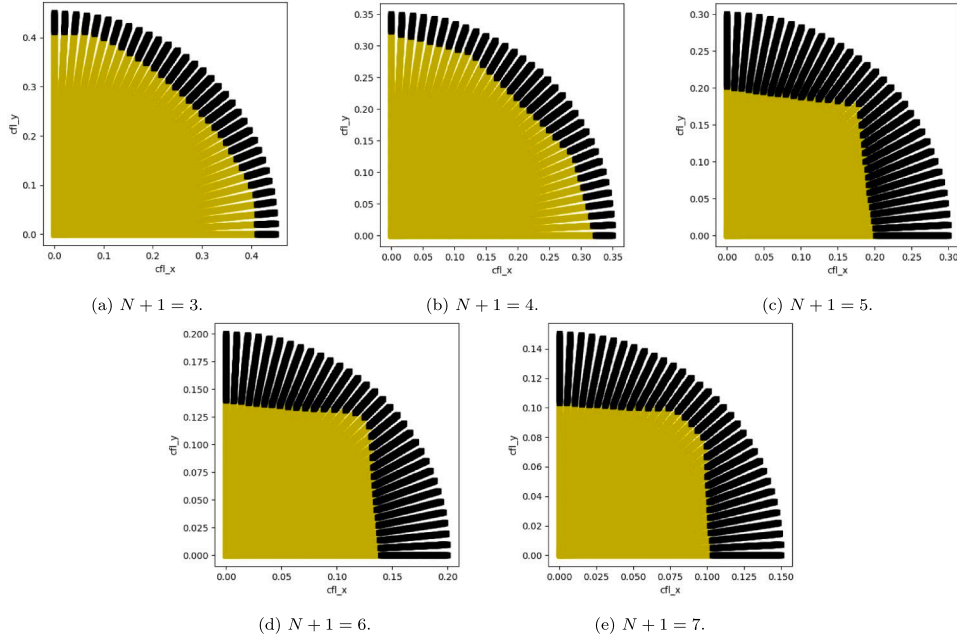
(a) $N + 1 = 3$. (b) $N + 1 = 4$. (c) $N + 1 = 5$.

(d) $N + 1 = 6$. (e) $N + 1 = 7$.

Fig. 8. Stable regions (▨) for GenAF($M_{\triangledown}$) of order $N + 1$ in space for linear advection. $\Delta x = \Delta y = 0.1$ and RK3 time discretization. The yellow (▨) and black (■) domains depict the tested area.

(a) $N + 1 = 4$. (b) $N + 1 = 5$.
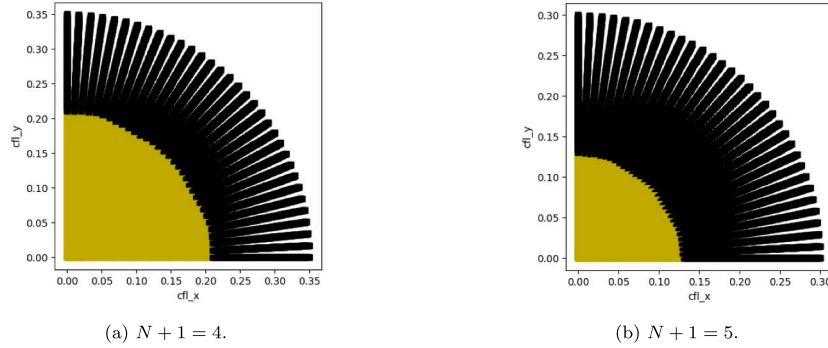
Fig. 9. Stable regions (▨) for GenAF($M_{\square}$) of order $N + 1$ in space for linear advection. $\Delta x = \Delta y = 0.1$ and RK3 time discretization. The yellow (▨) and black (■) domains depict the tested area.

$M_{\triangledown}$, the stability domain with $M_{\square}$ seems to be less dependent on the advection direction. This fits with the observation above, considering the tensor-like added degrees of freedom.

Comparing the maximal CFL numbers to the one-dimensional case in Abgrall and Barsukow [14, Table 1 (Method A, RK3)] it can be observed that for the method in 2-d with $M_{\square}$ and RK3, the maximal CFL $\approx r_{\text{cfl}}$ for linear advection in x- or y-direction is approximately the same for the tested $N + 1 = 3, 4, 5$.

In Barsukow et al. [22] a similar result for the generalized Active Flux method of third-order for linear acoustics in 2-d was indicated. A stability bound of cfl $= \frac{|c|\Delta t}{\min\{\Delta x, \Delta y\}} < 0.28$ was derived using Fourier analysis.

5. Numerical examples

Numerical results for the generalized Active Flux method GenAF($M_{\triangledown}$) for orders 3 to 7 on two-dimensional Cartesian grids with a SSP-RK3 time discretization are shown. In the following we concentrate on smooth problems. More complex problems with e.g. discontinuities will need an appropriate limiting strategy, see e.g. Duan et al. [32] for limiting of the third order Active Flux method. For higher orders this shall be part of future work.

5.1. Linear advection

The first example considers linear advection $\partial_t q(t, \mathbf{x}) + \partial_x q(t, \mathbf{x}) + \partial_y q(t, \mathbf{x}) = 0$ on $[0, 1]^2$ with periodic boundary conditions where initial data in the shape of a cone of radius $r_{\max} = 0.2$

$$
q(0, \mathbf{x}) = \begin{cases} 1 - \dfrac{r}{r_{\max}} & \text{for } r < r_{\max}, \\ 0 & \text{otherwise} \end{cases}
$$

with $r = \sqrt{(x - 0.5)^2 + (y - 0.5)^2}$ is being advected diagonally through the domain. Fig. 10 shows a cross section of the solution at $t = 5$ on a grid with 101×101 cells for spatial orders $N + 1 = 3$ to 7. A cfl $= C_{\text{cfl}}$ as shown in Table 3 (except for $N + 1 = 7$: cfl $= 0.085$) is used to calculate the time step. One observes that the apex of the cone is better approximated the higher the order of the method.

The second example considers smooth initial data (in resemblance to Abgrall and Barsukow [13, Section 3.5.1])

$$
q(0, \mathbf{x}) = 0.8 + \exp\left(-\left(\frac{x-0.5}{0.05}\right)^2 - \left(\frac{y-0.5}{0.05}\right)^2\right)
$$

that are used to show the convergence order of the method with $N + 1 = 3, \ldots, 7$ at $t = 0.1$. To recover the spatial convergence order $N + 1$ with a
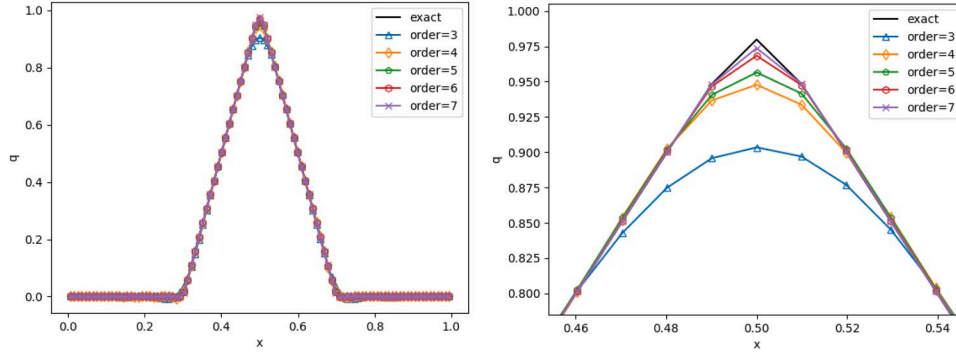
Fig. 10. Example of diagonally advected cone. Cross section at $y = 0.5$ of cell average on grid 101×101 at $t = 5$ for GenAF(M_∇) of order $N + 1 = 3, \ldots, 7$ (left) and close up (right).
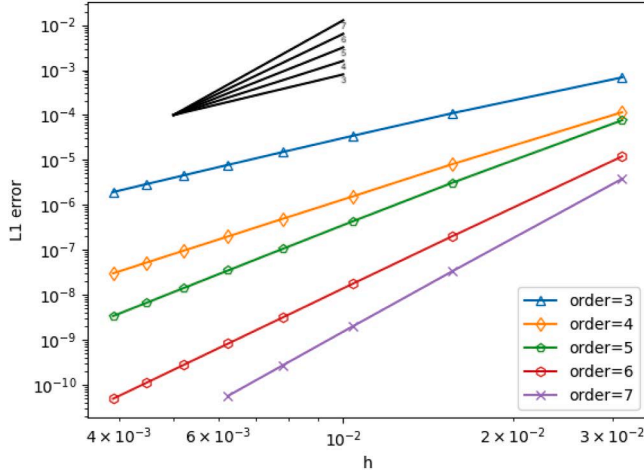

Fig. 11. Convergence of the cell averages in L^1-error for linear advection for GenAF(M_∇) of order $N + 1 = 3, \ldots, 7$ and RK3 time discretization.
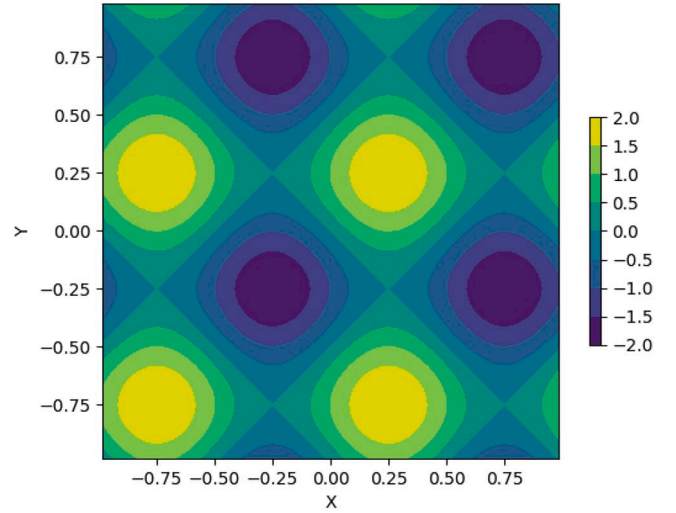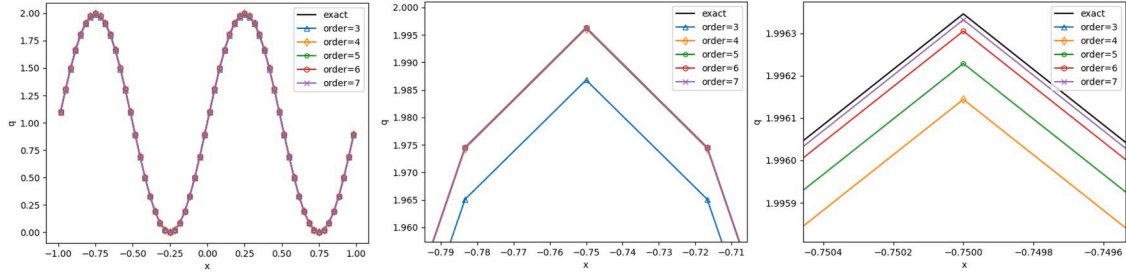

Fig. 12. Example for acoustics with sine wave in pressure. Exact solution for the cell average $p^{(0,0)}$ of pressure on a 60×60 grid at $t = 5$. *Remark:* The cell average values are approximated by numerical integration.

RK3 time discretization $\Delta t^3 = \mathcal{O}(h^{N+1})$ is needed, $h = \Delta x = \Delta y$. For the cell sizes $\{h_i\}_i$ an adaptive $\mathrm{cfl}(h_i) = \mathrm{cfl}(h_1)(\frac{h_i}{h_1})^{\frac{N-2}{3}}$ is used starting with $h_1 = \frac{1}{32}$ and the $\mathrm{cfl}(h_1) = C_{\mathrm{cfl}}$ as above. In Fig. 11 and the corresponding Table 4 the convergence of the error of the cell averages in the L^1-norm is shown and a convergence order $\mathcal{O}(h^{N+1})$ is observed.

5.2. Acoustic equations

The following example for the acoustic equations with $c > 0$

$$\partial_t p + c \nabla \cdot \mathbf{v} = 0$$
$$\partial_t \mathbf{v} + c \nabla p = 0$$

on $[-1, 1]^2$ considers initial conditions with a sine wave in pressure and zero initial velocity as suggested in Eymann and Roe [4] from Lukáčová-Medvid'ová et al. [33]

$$p(0, \mathbf{x}) = \frac{1}{c}(\sin(2\pi x) + \sin(2\pi y)),$$
$$\mathbf{v}(0, \mathbf{x}) = (0, 0)^\mathrm{T}$$

with periodic boundary conditions, for which the exact solution is given as

$$p(t, \mathbf{x}) = \frac{1}{c}\cos(2\pi ct)(\sin(2\pi x) + \sin(2\pi y)),$$
$$\mathbf{v}(t, \mathbf{x}) = \frac{1}{c}\sin(2\pi ct)(\cos(2\pi x), \cos(2\pi y))^\mathrm{T}.$$

We consider $c = 1$. The solution at $t = 5$ where the exact solution matches the initial condition is computed (see Fig. 12). The grid size

Table 4
L^1-error $e_{\mathrm{L}1}$ and corresponding experimental order of convergence (EOC) for GenAF(M_∇) of order $N + 1$ of the cell averages for linear advection and RK3 time discretization.

$N+1$	3		4		5	
h	$e_{\mathrm{L}1}$	EOC	$e_{\mathrm{L}1}$	EOC	$e_{\mathrm{L}1}$	EOC
0.03125	$6.87 \cdot 10^{-4}$	–	$1.15 \cdot 10^{-4}$	–	$7.65 \cdot 10^{-5}$	–
0.015625	$1.10 \cdot 10^{-4}$	2.65	$8.06 \cdot 10^{-6}$	3.84	$3.10 \cdot 10^{-6}$	4.62
0.0104167	$3.46 \cdot 10^{-5}$	2.84	$1.55 \cdot 10^{-6}$	4.07	$4.33 \cdot 10^{-7}$	4.86
0.0078125	$1.50 \cdot 10^{-5}$	2.91	$4.89 \cdot 10^{-7}$	4.01	$1.05 \cdot 10^{-7}$	4.94
0.00625	$7.76 \cdot 10^{-6}$	2.95	$1.98 \cdot 10^{-7}$	4.05	$3.46 \cdot 10^{-8}$	4.96
0.00520833	$4.52 \cdot 10^{-6}$	2.96	$9.50 \cdot 10^{-8}$	4.03	$1.40 \cdot 10^{-8}$	4.97
0.00446429	$2.86 \cdot 10^{-6}$	2.97	$5.11 \cdot 10^{-8}$	4.03	$6.49 \cdot 10^{-9}$	4.98
0.00390625	$1.92 \cdot 10^{-6}$	2.98	$2.98 \cdot 10^{-8}$	4.02	$3.34 \cdot 10^{-9}$	4.98

| $N+1$	6		7	
h	$e_{\mathrm{L}1}$	EOC	$e_{\mathrm{L}1}$	EOC
0.03125	$1.20 \cdot 10^{-5}$	–	$3.79 \cdot 10^{-6}$	–
0.015625	$2.01 \cdot 10^{-7}$	5.90	$3.33 \cdot 10^{-8}$	6.83
0.0104167	$1.77 \cdot 10^{-8}$	5.99	$1.99 \cdot 10^{-9}$	6.95
0.0078125	$3.11 \cdot 10^{-9}$	6.05	$2.67 \cdot 10^{-10}$	6.98
0.00625	$8.13 \cdot 10^{-10}$	6.01	$5.60 \cdot 10^{-11}$	7.00
0.00520833	$2.72 \cdot 10^{-10}$	6.01	–	–
0.00446429	$1.07 \cdot 10^{-10}$	6.03	–	–
0.00390625	$4.81 \cdot 10^{-11}$	6.01	–	–

Fig. 13. Example for acoustics with sine wave in pressure. Cross section at $y = 0.25$ of $p^{(0,0)}$ on grid 60×60 at $t = 5$ for GenAF(M_{∇}) of order $N + 1 = 3, \ldots, 7$ (left) and two close ups (middle, right) zooming in at the left maximum (observe the different scales of the x-axes).
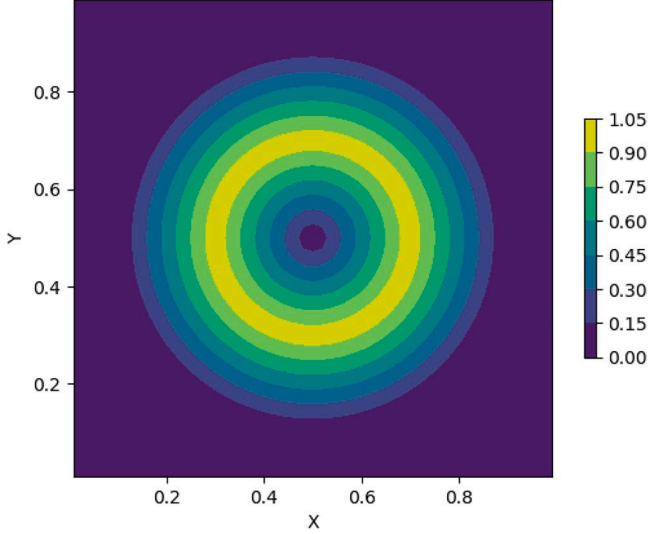
Fig. 14. Gresho vortex for Euler equations. Exact solution for the norm of the cell average of the momentum $\|(\rho \mathbf{v})^{(0,0)}\|_2$ on a 51×51 grid at $t = 1$. *Remark:* The cell average values are approximated by numerical integration.

is $N_x = N_y = 60$ and the time step $\Delta t = \mathrm{cfl} \frac{\min\{\Delta x, \Delta y\}}{c}$ is computed with $\mathrm{cfl} = C_{\mathrm{cfl}}$, a CFL bound obtained for linear advection, see Table 3 (except for $N + 1 = 7$: $\mathrm{cfl} = 0.085$). As for linear advection, a better approximation of the exact solution is observed for higher orders of the generalized Active Flux method. In particular, a cross section through a maximum of the pressure sine wave is shown in Fig. 13.
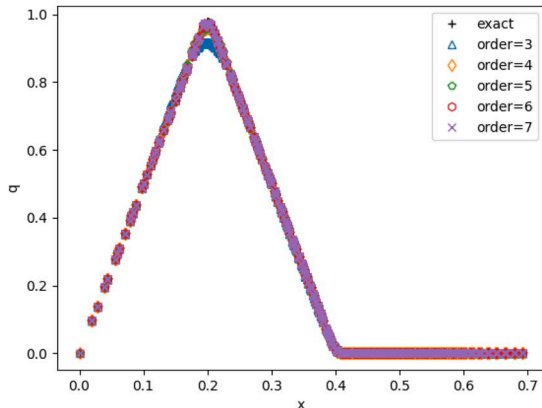
5.3. Euler equations

The generalized Active Flux method can also be applied to non-linear problems like the compressible Euler equations

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0$$
$$\partial_t (\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v} + p \mathbb{I}) = 0$$
$$\partial_t E + \nabla \cdot ((E + p)\mathbf{v}) = 0$$

In this paper an ideal polytropic gas is considered with $E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho\|\mathbf{v}\|_2^2$ and $\gamma = 1.4$.

A Gresho vortex (see Gresho and Chan [34], Barsukow et al. [35]) on $[0, 1]^2$

$$\rho(0, \mathbf{x}) = 1$$

$$\mathbf{v}(0, \mathbf{x}) = \begin{cases} \frac{5r}{r}(-y - 0.5, x - 0.5)^{\mathrm{T}} & r < 0.2 \\ \frac{2 - 5r}{r}(-y - 0.5, x - 0.5)^{\mathrm{T}} & 0.2 \leq r < 0.4 \\ 0 & \text{else} \end{cases}$$

$$p(0, \mathbf{x}) = \begin{cases} \frac{1}{\gamma M^2} - \frac{1}{2} + \frac{(5r)^2}{2} & r < 0.2 \\ \frac{1}{\gamma M^2} - \frac{1}{2} + 4\ln(5r) + 4 - 20r + \frac{(5r)^2}{2} & 0.2 \leq r < 0.4 \\ \frac{1}{\gamma M^2} - \frac{1}{2} + 4\ln 2 - 2 & \text{else} \end{cases}$$

for $M = 0.1$ with $r = \sqrt{(x - 0.5)^2 + (y - 0.5)^2}$ and periodic boundary conditions is computed at $t = 1$ on a grid of 51×51 cells (see Fig. 14). The time step is calculated with the help of the CFL bound from Table 3 (except for $N + 1 = 7$: $\mathrm{cfl} = 0.085$). Fig. 15 shows the radial plot of the cell average for the norm of the momentum $\|(\rho \mathbf{v})^{(0,0)}\|_2$ for the generalized Active Flux method of order $3, \ldots, 7$.
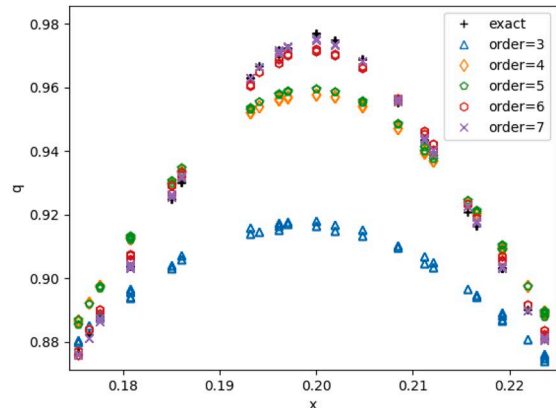
Fig. 15. Radial plot of $\|(\rho \mathbf{v})^{(0,0)}\|_2$ for Gresho vortex on 51×51 grid at $t = 1$ for GenAF(M_{∇}) of order $N + 1 = 3, \ldots, 7$ (left) and close up (right).

6. Conclusion and outlook

This paper presents a generalized Active Flux method on two-dimensional Cartesian grids of arbitrarily high order using higher moments in addition to the cell averages and point values at the cell interfaces. It is extending a semi-discrete hybrid finite element–finite volume version of Active Flux from 1-d. We focused on a method with a serendipity-like, hybrid finite element using a minimal number of degrees of freedom on the Cartesian cells to keep the computational cost minimal. An alternative is a tensor-like extension of the element from 1-d to 2-d. The point values at the cell interfaces include the nodes and edge points, such that a globally continuous reconstruction is obtained. Their update uses a Jacobian splitting of the non-conservative formulation of the system and finite difference formulas for the approximation of the derivatives of the solution. These are derived using the reconstruction of the solution on the cells. The moments are updated based on the weak formulation with the integrals approximated by quadrature formulas.

The eigenvalue spectrum of the semi-discrete method was analyzed for linear advection up to order 7. We find that the edge points at the cell interfaces cannot be distributed arbitrarily. We have identified, at least for orders up to 7, Gauss points as a suitable choice to achieve a stable setup of the semi-discrete method. For the time discretization we relied on a SSP-RK3 method and found CFL bounds of the method up to order 7 for linear advection. Numerical examples confirm the theoretical convergence order and demonstrate how the high-order methods allow to better resolve features of the solution.

Further studies to better understand the effect of suitable edge point distributions on the stability of the method will be interesting. In future work, it will also be important to develop a suitable limiting for the method. To this end, a flux vector splitting for the point value update, which was recently presented in Duan et al. [32] for third order, could be considered. Further work shall also consider an extension to 3-d as well as applications to simulations of turbulent convection and wave propagation, which require very high orders of accuracy.

In memoriam

This paper is dedicated to the memory of Prof. Arturo Hidalgo López (*July 3rd 1966 - †August 26th 2024) of the Universidad Politecnica de Madrid, organizer of HONOM 2019 and active participant in many other editions of HONOM. Our thoughts and wishes go to his wife Lourdes and his sister María Jesús, whom he left behind.

CRediT authorship contribution statement

Wasilij Barsukow: Software, Conceptualization, Writing – review & editing; **Praveen Chandrashekar:** Formal analysis, Writing – review & editing; **Christian Klingenberg:** Funding acquisition, Resources, Conceptualization, Project administration, Supervision, Writing – review & editing; **Lisa Lechner:** Formal analysis, Visualization, Software, Writing – original draft, Conceptualization.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] Godunov SK. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Matematicheskii Sbornik 1959;47(89-3):271–306.

[2] Eymann TA, Roe PL. In: 49th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition. American Institute of Aeronautics and Astronautics; 2011, p. 49. https://doi.org/10.2514/6.2011-382

[3] Eymann T, Roe P. Active Flux schemes for systems. In: 20th AIAA computational fluid dynamics conference. American Institute of Aeronautics and Astronautics; 2011. https://doi.org/10.2514/6.2011-3840

[4] Eymann TA, Roe PL. In: 21st AIAA computational fluid dynamics conference. American Institute of Aeronautics and Astronautics; 2013. https://doi.org/10.2514/6.2013-2940

[5] van Leer B. Towards the ultimate conservative difference scheme. iv. a new approach to numerical convection. J Comput Phys 1977;23(3):276–99. https://doi.org/10.1016/0021-9991(77)90095-X

[6] Chudzik E, Helzel C. Finite volumes for complex applications X-volume 1, elliptic and parabolic problems; vol. 432 of *Cham*. Springer; 2023. . Springer Proceedings in Mathematics & Statistics. https://doi.org/10.1007/978-3-031-40864-9_6

[7] Maeng J. Tech. Rep.; University of Michigan; 2017. Ph.D. thesis; https://deepblue.lib.umich.edu/handle/2027.42/138695.

[8] Fan D. Tech. Rep.; University of Michigan; 2017. Ph.D. thesis; https://deepblue.lib.umich.edu/handle/2027.42/140800.

[9] Barsukow W, Hohm J, Klingenberg C, Roe PL. The Active Flux scheme on Cartesian grids and its low Mach number limit. J Sci Comput 2019;81(1):594–622. https://doi.org/10.1007/s10915-019-01031-z

[10] Barsukow W. The Active Flux scheme for nonlinear problems. J Sci Comput 2020;86(1):3. https://doi.org/10.1007/s10915-020-01381-z

[11] Chudzik E, Helzel C, Lukáčová-Medvid'ová M. Active Flux methods for hyperbolic systems using the method of bicharacteristics. J Sci Comput 2024;99(1):16. https://doi.org/10.1007/s10915-024-02462-z

[12] Abgrall R. A combination of residual distribution and the Active Flux formulations or a new class of schemes that can combine several writings of the same hyperbolic problem: application to the 1d euler equations. Commun Appl Math Comput 2023;5(1):370–402. https://doi.org/10.1007/s42967-021-00175-w

[13] Abgrall R, Barsukow W. Extensions of Active Flux to arbitrary order of accuracy. ESAIM: mathematical modelling and numerical analysis (ESAIM: M2AN; vol. 57. 2023. https://doi.org/10.1051/m2an/2023004

[14] Abgrall R, Barsukow W. A hybrid finite element-finite volume method for conservation laws. Appl Math Comput 2023;447:127846. https://doi.org/10.1016/j.amc.2023.127846

[15] Abgrall R, Barsukow W, Klingenberg C. A semi-discrete Active Flux method for the euler equations on cartesian grids. J Sci Comput 2025;102(2):36. https://doi.org/10.1007/s10915-024-02749-1

[16] Abgrall R, Lin J, Liu Y. Active Flux for triangular meshes for compressible flows problems. Beijing J Pure Appl Math 2025;2(1):1–33. https://doi.org/10.4310/BPAM.250414223907

[17] He F. Towards a new-generation numerical scheme for the compressible navier-stokes equations with the Active Flux method. Tech. Rep.; Applied and Interdisciplinary Mathematics, University of Michigan; 2021. Ph.D. thesis. https://doi.org/10.7302/2732

[18] Roe P Designing Cfd methods for bandwidth-a physical approach. Comput Fluid 2021;214:104774. . https://doi.org/10.1016/j.compfluid.2020.104774

[19] Samani I, Roe PL. Acoustics on a coarse grid. In: AIAA SCITECH 2023 forum. American Institute of Aeronautics and Astronautics; 2023. https://doi.org/10.2514/6.2023-1156

[20] Samani I. Active Flux methods with gradient degrees of freedom. Tech. Rep.; University of Michigan; 2024. Ph.D. thesis. https://doi.org/10.7302/22838

[21] Kern J. On a stationarity preserving multi-dimensional finite volume scheme for conservation laws. Tech. Rep.; Faculty of Mathematics and Computer Science, Julius-Maximilians-Universität Würzburg; 2024. Master's thesis.

[22] Barsukow W, Kern J, Klingenberg C, Lechner L. Analysis of the multi-dimensional semi-discrete Active Flux method using the fourier transform. 2025. https://doi.org/10.1007/s42967-025-00512-3

[23] Helzel C, Kerkmann D, Scandurra L. A new ader method inspired by the Active Flux method. J Sci Comput 2019;80(3):1463–97. https://doi.org/10.1007/s10915-019-00988-1

[24] Pan Y, Persson P-O. A face-upwinded spectral element method. J Comput Phys 2024;503:112825. https://doi.org/10.1016/j.jcp.2024.112825

[25] Ern A, Guermond J-L. Finite elements I: approximation and interpolation vol. 72 of Texts in applied mathematics, Cham. Springer; 2021. https://doi.org/10.1007/978-3-030-56341-7

[26] Arnold DN, Awanou G. The serendipity family of finite elements. Foundat Comput Math, 2011;11:337–44. https://doi.org/10.1208/s10208-011-9087-3

[27] Mathematica, Version 13.3, Champaign, IL. Wolfram Research, Inc; 2023.

[28] Lomax H, Pulliam TH, Zingg DW. Fundamentals of computational fluid dynamics, scientific computation. Berlin, Heidelberg; Springer; 2001. https://doi.org/10.1007/978-3-662-04654-8

[29] Guennebaud G, Jacob B, et al. Eigen v3 2010. http://eigen.tuxfamily.org.

[30] Harris CR, Millman KJ, Walt S J VD, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with numpy. Nature 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2

[31] Hunter JD. Matplotlib: a 2d graphics environment. Comput Sci Eng 2007;9(3):90–5. https://doi.org/10.1109/MCSE.2007.55

[32] Duan J, Barsukow W, Klingenberg C. Active Flux methods for hyperbolic conservation laws-flux vector splitting and bound-preservation. SIAM J Sci Comput 2025;47(2):A811–A837. https://doi.org/10.1137/24M1658887

[33] Lukáčová-Medvid'ová M, Morton KW, Warnecke G. Evolution galerkin methods for hyperbolic systems in two space dimensions. Math Comput 2000;69(232):1355–85. https://doi.org/10.1090/s0025-5718-00-01228-x

[34] Gresho PM, Chan ST. On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. part 2: implementation. Int J Numer Method Fluid 1990;11(5):621–59. https://doi.org/10.1002/fld.1650110510

[35] Barsukow W, Edelmann P VF, Klingenberg C, Miczek F, Röpke FK. A numerical scheme for the compressible low-mach number regime of ideal fluid dynamics. J Sci Comput 2017;72(2):623–46. https://doi.org/10.1007/s10915-017-0372-4