



Research article

Structure preservation in high-order hybrid discretisations of potential-driven advection-diffusion: linear and nonlinear approaches[†]

Simon Lemaire^{1,*} and Julien Moatti^{1,2}

¹ Inria, Univ. Lille, CNRS, UMR 8524–Laboratoire Paul Painlevé, F-59000 Lille, France

² Institute of Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstr. 8–10, A-1040 Wien, Austria

[†] **This contribution is part of the Special Issue: Advancements in Polytopal Element Methods**
Guest Editors: Michele Botti; Franco Dassi; Lorenzo Mascotto; Ilario Mazzieri
Link: www.aimspress.com/mine/article/6538/special-articles

* **Correspondence:** Email: simon.lemaire@inria.fr.

Abstract: We are interested in the high-order approximation of anisotropic, potential-driven advection-diffusion models on general polytopal partitions. We study two hybrid schemes, both built upon the Hybrid High-Order technology. The first one hinges on exponential fitting and is linear, whereas the second is nonlinear. The existence of solutions is established for both schemes. Both schemes are also shown to possess a discrete entropy structure, ensuring that the long-time behaviour of discrete solutions mimics the PDE one. For the nonlinear scheme, the positivity of discrete solutions is a built-in feature. On the contrary, we display numerical evidence indicating that the linear scheme violates positivity, whatever the order. Finally, we verify numerically that the nonlinear scheme has optimal order of convergence, expected long-time behaviour, and that raising the polynomial degree results, also in the nonlinear case, in an efficiency gain.

Keywords: high-order methods; hybrid methods; polytopal meshes; structure-preserving schemes; potential-driven advection-diffusion; long-time behaviour; entropy methods

1. Motivations and context

We are interested in the polytopal discretisation of a linear potential-driven advection-diffusion equation using high-order schemes. Our goal is to compare an exponentially fitted linear method with a nonlinear approach. Let Ω be an open, bounded, connected polytopal subset of \mathbb{R}^d , $d \in \{2, 3\}$, with Lipschitz boundary. We consider the following anisotropic advection-diffusion problem with

homogeneous Neumann boundary conditions: find the density $u : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ solution to

$$\begin{cases} \partial_t u - \operatorname{div}(\Lambda(\nabla u + u\nabla\phi)) = 0, & \text{in } \mathbb{R}_+^* \times \Omega, \\ \Lambda(\nabla u + u\nabla\phi) \cdot n = 0, & \text{on } \mathbb{R}_+^* \times \partial\Omega, \\ u(0, \cdot) = u^0, & \text{in } \Omega, \end{cases} \quad (1.1)$$

where n is the unit normal vector to $\partial\Omega$ pointing outward from Ω . We assume that the data satisfy:

- (i) $\Lambda \in L^\infty(\Omega)^{d \times d}$ is a uniformly elliptic anisotropy tensor: there exists $\lambda_b > 0$ such that, for a.e. x in Ω , $\Lambda(x)\xi \cdot \xi \geq \lambda_b |\xi|^2$ for all $\xi \in \mathbb{R}^d$;
- (ii) $\phi \in W^{1,\infty}(\Omega)$ is a regular potential;
- (iii) $u^0 \in L^1(\Omega)$ is a non-negative initial datum, such that $\int_\Omega u^0 \log(u^0) < \infty$.

The solution to (1.1) enjoys some specific and well-known properties. First, the mass is conserved along time, i.e., for almost every $t > 0$,

$$\int_\Omega u(t) = \int_\Omega u^0 = M, \quad (1.2)$$

where $M > 0$ is the initial mass. Second, the solution is positive:

$$\text{for a.e. } t > 0, \quad u(t, \cdot) > 0 \text{ a.e. in } \Omega. \quad (1.3)$$

Last, the solution converges exponentially fast when $t \rightarrow \infty$ towards the thermal equilibrium, unique steady solution to (1.1), given by

$$u^\infty = \frac{M}{\int_\Omega e^{-\phi}} e^{-\phi}. \quad (1.4)$$

In order to compute a reliable numerical approximation of Problem (1.1), one should preserve at the discrete level the three above-listed structural properties. In practice, our final target application are drift-diffusion semiconductor models [37, 50] (and, in particular, anisotropic ones [36]). In these models, the electric potential ϕ driving the drift is one of the unknowns of the problem, alongside with the densities of charge carriers. It is solution to a Poisson equation. At the PDE level, the thermal equilibrium is defined as the density $u^\infty > 0$ for which the flux $\Lambda(\nabla u^\infty + u^\infty \nabla \phi)$ identically vanishes in Ω . This characterisation implies that the equilibrium quasi-Fermi potential $\log(u^\infty) + \phi$ shall be constant in Ω [43]. At the discrete level, this motivates the following definition.

Definition 1 (Preservation of the thermal equilibrium). *A numerical scheme for (1.1) preserves the thermal equilibrium if the corresponding discrete equilibrium quasi-Fermi potential is constant.*

Note that this definition implies that the discrete equilibrium density has to be positive. In semiconductor models discretisations, the potential ϕ is sought as an element of the discrete space. By Definition 1, preserving the thermal equilibrium then essentially requires to also seek $\log(u^\infty)$ as an element of the latter discrete space. For the schemes we study in this work, the precise meaning of Definition 1 will be made clear in Proposition 1 below.

In the realm of Two-Point Flux Approximation (TPFA) finite volume schemes, the so-called Scharfetter–Gummel fluxes [47] are precisely devised so as to preserve the thermal equilibrium. They

naturally lead to linear structure-preserving discretisations of the problem (see [14, 15]). However, TPFA methods can only be used on meshes satisfying orthogonality conditions (with respect to the inner product induced by Λ , in case Λ is symmetric), which essentially restricts their use to isotropic problems. On the other hand, a number of finite volume methods using auxiliary unknowns has been introduced within the past twenty years or so for the discretisation of anisotropic problems on general meshes. One can cite the Discrete Duality Finite Volume (DDFV) method [27, 38], with additional unknowns attached to a dual mesh, the Vertex Approximate Gradient (VAG) scheme [34], with auxiliary unknowns attached to the mesh vertices, or the Mimetic Finite Difference (MFD) and Hybrid Finite Volume (HFV) methods [9, 33], with auxiliary unknowns attached to the mesh faces. Such methods have proved to be relevant solutions to the anisotropy issue, but none of these linear schemes preserves the positivity of the solutions (see [28]). A possible alternative was proposed in [13], with the introduction and analysis of a nonlinear positivity-preserving VAG scheme. The design and analysis of this scheme, as well as of its DDFV and HFV counterparts of [11, 12, 16], leverage the entropy structure of Problem (1.1): there exists some physically motivated quantity, called entropy, which decays along time. Reproducing this structure at the discrete level is key to get stability, convergence, and accurate time asymptotics. Other approaches to positivity preservation on general meshes have been explored in the literature. Still in the realm of finite volume methods, one can cite the works [5, 29, 46, 48, 49]. As opposed to [13], in which the nonlinearities are introduced at the PDE level then discretised, the latter contributions introduce nonlinearities directly at the discrete level. These nonlinearities, unfortunately, often do not lend themselves to a PDE re-interpretation, making difficult to unravel the potential discrete entropy structures hidden behind. Arbitrary-order positivity-preserving (or, more generally, discrete maximum principle preserving) methods have also been studied in the literature. In the finite element context, one can cite the seminal works [17, 18] by Ciarlet, as well as the more recent contributions [2, 35] (see also [3] for a comprehensive survey). These approaches are, however, restricted to standard meshes. In addition, only algebraic positivity can usually be enforced, that is positivity of the degrees of freedom, but not of the (piecewise polynomial) functions themselves over the domain. Weak positivity enforcement has also been explored in the Discontinuous Galerkin (DG) framework in [41, 42]: therein, positivity is enforced on the cell averages of the piecewise polynomial solutions. Turning to pointwise positivity enforcement, let us mention in the DG context the interesting contribution [7]. Therein, a nonlinear scheme is introduced for the (reaction-diffusion) Fisher–KPP equation $\partial_t u - \Delta u = u(1 - u)$, in which the (positive) densities are defined as $u = e^\lambda$, with λ piecewise polynomial. This scheme is developed so as to preserve the entropy structure of the PDE model. Compared to the high-order DG schemes of [41, 42], the main improvement lies in the fact that the discrete solutions are positive everywhere. Such a feature allows for a complete analysis of the scheme, including existence, long-time behaviour, and convergence towards a semi-discretised solution. The analysis is based on the properties of a well-chosen stabilisation term, whose expression implies L^∞ -norms of the polynomial unknowns over the mesh faces. The results of [7], valid on simplicial meshes, have recently been extended (excluding the long-time behaviour) to polytopal meshes in [22], still in the DG context. Along the same lines, yet restricted to standard meshes, let us also cite the conforming space-time Galerkin discretisation of [8] for cross-diffusion systems.

From the above literature review, it is quite clear that the landscape in terms of positivity-preserving polytopal methods of arbitrarily high approximation order for advection-diffusion problems is relatively scarce. Speaking of pointwise positivity enforcement, the only existing contribution we

are aware of is [22], in the DG context, and for reaction-diffusion equations. In the present work, our aim is to study an arbitrary-order **hybrid** polytopal scheme for Problem (1.1), preserving the three structural properties (1.2)–(1.4) listed above. One expected advantage of hybrid methods over DG schemes is a reduction of the number of globally coupled unknowns in the linear systems to be solved at each iteration of the Newton algorithm, which should be all the more substantial that the order of approximation increases. Our (nonlinear) scheme has been briefly introduced, and a first numerical assessment performed, in [44]. Our goal in the present article is twofold. First, we want to provide our approach with theoretical foundations. Second, we aim to conduct an extensive numerical validation of our method (convergence orders, efficiency, positivity, large time), including a comparison in terms of structure preservation with a similar (in the spirit) high-order linear scheme. One could indeed expect, at least in practice, that the use of a method (even linear) with sufficiently high order (and thus accuracy), could already constitute in itself a solution to positivity violation issues. The two (linear and nonlinear) methods we consider are built upon the Hybrid High-Order (HHO) technology [25, 26], as natural extensions of the HFV schemes introduced in [16, 45]. The linear scheme hinges on the exponential fitting strategy [10]. The key idea is the linear change of unknown $u = e^{-\phi} \rho$, which allows one to reformulate (1.1) as an unconditionally coercive problem in the variable ρ . As a by-product of this reformulation, the scheme naturally preserves the thermal equilibrium. The nonlinear scheme relies on the nonlinear change of unknown $u = e^{\ell}$ (cf. [13]), which is designed so as to preserve the Boltzmann entropy structure of the PDE model and, as a by-product, the positivity of solutions, the thermal equilibrium, and the long-time asymptotics. For the sake of simplicity, both schemes rely on a mixed-order HHO space: given an integer $k \geq 0$, the methods hinge on face unknowns of polynomial degree k , and enriched cell unknowns of polynomial degree $k + 1$. The main interest of such a discrepancy in the degree between face and cell unknowns is a simplification in the design of the higher-order bulk reconstruction and of the stabilisation, resulting in turn in a simplification of the analysis. In the meantime, such a choice preserves optimal accuracy (order $k + 2$ in L^2 -norm) and frugality (the face unknowns, of degree k , are the only globally coupled unknowns). Since we are manipulating mixed-order spaces, following [20], we could also refer to our methods as HDG methods. However, we prefer naming them HHO methods for the two following reasons (cf. [19, Section 1.5.2]). First, HDG schemes are developed adopting a mixed-hybrid viewpoint, whereas we adopt here the primal HHO viewpoint. Second, the analysis of HDG methods usually hinges on specific (often simplex-based) projections, whereas our HHO analysis makes here a systematic use of L^2 -orthogonal projectors (well-defined on polytopal cells). In any case, the two schemes we introduce in this work are new in the HDG/HHO context. Our first theoretical results, stated in Propositions 2 and 3, concern the well-posedness and long-time behaviour of the (linear) exponential fitting scheme. Regarding the nonlinear scheme, we prove the existence of (positive) solutions in Theorem 1. These discrete solutions are further proved to converge (in large time) in Proposition 5 towards the discrete equilibrium of the scheme. Note that we could also have compared our nonlinear scheme with the linear HHO method for advection-diffusion of [24], which generalises to arbitrary approximation orders the HMM scheme of [4] (both introduced for general advection fields). We have not pursued further in this direction, this for two reasons. First, the stability of this scheme hinges on some coercivity assumptions which constrain the variety of potentials that can be considered, and second it does not preserve the thermal equilibrium (see [16] in the lowest-order case).

The rest of the article is organised as follows. In Section 2, we first introduce the discrete framework,

and describe the two schemes under consideration. Then, in Section 3, we discuss the main properties of the two schemes, and we provide some elements of analysis regarding the well-posedness and discrete long-time behaviours. Last, in Section 4, we discuss the implementation of the nonlinear scheme, and we assess the behaviour of the methods on various test-cases.

2. Discrete setting and schemes

The two numerical schemes we consider in this article are based on a backward Euler discretisation in time, with uniform time step $\Delta t > 0$. The time discretisation is thus defined as $(t^n)_{n \in \mathbb{N}}$, where $t^n = n\Delta t$. Note that it is straightforward to generalise the discussion below to a variable time step. We focus in this section on space discretisation.

2.1. Polytopal meshes and anisotropy tensor

In the vein of [23, Definition 1.4], we define a discretisation of Ω as a couple $\mathcal{D} = (\mathcal{M}, \mathcal{E})$, where:

- The mesh \mathcal{M} is a partition of Ω , i.e., \mathcal{M} is a finite collection of disjoint, open, Lipschitz polytopes $K \subset \Omega$ with $|K|_d > 0$ (the cells) such that $\overline{\Omega} = \bigcup_{K \in \mathcal{M}} \overline{K}$;
- The set \mathcal{E} is a partition of the mesh skeleton $\partial\mathcal{M} = \bigcup_{K \in \mathcal{M}} \partial K$, i.e., \mathcal{E} is a finite collection of disjoint, connected, relatively open subsets σ of $\overline{\Omega}$ with $|\sigma|_{d-1} > 0$ (the faces) such that $\partial\mathcal{M} = \bigcup_{\sigma \in \mathcal{E}} \overline{\sigma}$. It is assumed that, for all $\sigma \in \mathcal{E}$, σ is a Lipschitz polytopal subset of an affine hyperplane of \mathbb{R}^d . We assume that, for all $K \in \mathcal{M}$, there exists a subset \mathcal{E}_K of \mathcal{E} (the set of faces of the cell K) such that $\partial K = \bigcup_{\sigma \in \mathcal{E}_K} \overline{\sigma}$. Finally, we let $n_{K,\sigma} \in \mathbb{R}^d$ be the (constant) unit normal vector to $\sigma \in \mathcal{E}_K$ pointing outward from K .

The diameter of a subset $X \subset \overline{\Omega}$ is denoted by $h_X = \sup\{|x - y| \mid (x, y) \in X^2\}$, and we define the size of \mathcal{D} (the mesh size) as $h_{\mathcal{D}} = \max_{K \in \mathcal{M}} h_K$. For further use, we also introduce the smallest cell diameter $h_b = \min_{K \in \mathcal{M}} h_K$ of \mathcal{D} .

When studying asymptotic behaviours with respect to the mesh size, one has to adopt a measure of regularity for refined families of discretisations. We classically follow [23, Definition 1.9], in which regularity for a refined mesh family is quantified by a uniform (with respect to the mesh size) parameter $\theta \in (0, 1)$, called mesh regularity parameter. This parameter measures the chunkiness of the cells, but also the diameter ratio between the cells and their faces. In what follows, to avoid the proliferation of multiplicative constants, we write $a \lesssim b$ in place of $Ca \leq b$ if $C > 0$ only depends on Ω , on the mesh regularity parameter θ , and (if need be) on Λ , ϕ , and the underlying polynomial degree, but is independent of both $h_{\mathcal{D}}$ (and h_b) and Δt .

Remark 1 (Relaxation of the mesh regularity assumptions). *Upon replacing the scalings h_{σ} for $\sigma \in \mathcal{E}_K$ by h_K in the stabilisations/discrete norms below, the analysis performed in this work remains valid under the (much) less stringent mesh regularity assumptions of [30, Assumption 1] (cf. also [23, Definition 1.41]). Contrary to [23, Definition 1.9], these relaxed mesh regularity assumptions allow for small faces and cells with numerous faces, as they may appear in agglomeration-based meshing.*

Last, we make an additional regularity assumption on the anisotropy tensor. We assume that

$$\Lambda|_K \in W^{1,\infty}(K)^{d \times d}, \quad \forall K \in \mathcal{M}. \quad (2.1)$$

2.2. Discrete space and operators

For $q \in \mathbb{N}$, and X subset of $\overline{\Omega}$ of Hausdorff dimension $1 \leq l \leq d$, we let $\mathbb{P}^q(X)$ denote the vector space of l -variate polynomial functions $X \rightarrow \mathbb{R}$ of total degree at most q . We also define the $L^2(X)$ -orthogonal projector $\Pi_X^q : L^1(X) \rightarrow \mathbb{P}^q(X)$ such that, given any $v \in L^1(X)$, $\Pi_X^q(v)$ is the only element in $\mathbb{P}^q(X)$ satisfying

$$\int_X \Pi_X^q(v)z = \int_X vz, \quad \forall z \in \mathbb{P}^q(X).$$

Given any $K \in \mathcal{M}$, we also introduce the vector space $\mathbb{P}^q(K)^d$ of d -variate polynomial vector fields $K \rightarrow \mathbb{R}^d$ of total degree at most q , as well as the corresponding $L^2(K)^d$ -orthogonal projector (denoted as its scalar version) $\Pi_K^q : L^1(K)^d \rightarrow \mathbb{P}^q(K)^d$. For any $\sigma \in \mathcal{E}_K$ and $v \in W^{1,1}(K)$, we also introduce the shortcut notation

$$\Pi_\sigma^q(v) = \Pi_\sigma^q(v|_\sigma).$$

Let k be a given non-negative integer. We introduce the mixed-order HHO space (see [19,23]), with face unknowns of degree k and (enriched) cell unknowns of degree $k+1$:

$$\underline{V}_\mathcal{D}^k = \left\{ \underline{v}_\mathcal{D} = ((v_K)_{K \in \mathcal{M}}, (v_\sigma)_{\sigma \in \mathcal{E}}) \mid \begin{array}{l} \forall K \in \mathcal{M}, \quad v_K \in \mathbb{P}^{k+1}(K) \\ \forall \sigma \in \mathcal{E}, \quad v_\sigma \in \mathbb{P}^k(\sigma) \end{array} \right\}.$$

Given a cell $K \in \mathcal{M}$, we let

$$\underline{V}_K^k = \mathbb{P}^{k+1}(K) \times \left(\prod_{\sigma \in \mathcal{E}_K} \mathbb{P}^k(\sigma) \right)$$

be the restriction of $\underline{V}_\mathcal{D}^k$ to K , and for a generic discrete element $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}^k$, we denote by $\underline{v}_K = (v_K, (v_\sigma)_{\sigma \in \mathcal{E}_K}) \in \underline{V}_K^k$ its local restriction to the cell K . To any $\underline{v}_\mathcal{D} \in \underline{V}_\mathcal{D}^k$, we associate the two piecewise polynomial functions $v_\mathcal{M} : \Omega \rightarrow \mathbb{R}$ and $v_\mathcal{E} : \partial\mathcal{M} \rightarrow \mathbb{R}$ such that

$$v_{\mathcal{M}|K} = v_K \text{ for all } K \in \mathcal{M} \quad \text{and} \quad v_{\mathcal{E}|\sigma} = v_\sigma \text{ for all } \sigma \in \mathcal{E}.$$

We also let $\underline{1}_\mathcal{D} \in \underline{V}_\mathcal{D}^k$ be the discrete element such that $1_K = 1$ for all $K \in \mathcal{M}$ and $1_\sigma = 1$ for all $\sigma \in \mathcal{E}$. Last, given a cell $K \in \mathcal{M}$, we define the local interpolator $\underline{I}_K^k : W^{1,1}(K) \rightarrow \underline{V}_K^k$ such that, for any $v \in W^{1,1}(K)$,

$$\underline{I}_K^k(v) = \left(\Pi_K^{k+1}(v), (\Pi_\sigma^k(v))_{\sigma \in \mathcal{E}_K} \right).$$

Similarly, the global interpolator $\underline{I}_\mathcal{D}^k : W^{1,1}(\Omega) \rightarrow \underline{V}_\mathcal{D}^k$ is defined, for any $v \in W^{1,1}(\Omega)$, by

$$\underline{I}_\mathcal{D}^k(v) = \left((\Pi_K^{k+1}(v|_K))_{K \in \mathcal{M}}, (\Pi_\sigma^k(v|_\sigma))_{\sigma \in \mathcal{E}} \right).$$

As standard in the HHO context, locally to any cell $K \in \mathcal{M}$, we introduce a discrete gradient operator $G_K : \underline{V}_K^k \rightarrow \mathbb{P}^k(K)^d$ such that, for any $\underline{v}_K \in \underline{V}_K^k$, $G_K(\underline{v}_K) \in \mathbb{P}^k(K)^d$ satisfies

$$\int_K G_K(\underline{v}_K) \cdot \tau = - \int_K v_K \nabla \cdot \tau + \sum_{\sigma \in \mathcal{E}_K} \int_\sigma v_\sigma \tau \cdot n_{K,\sigma}, \quad \forall \tau \in \mathbb{P}^k(K)^d. \quad (2.2)$$

This operator is a consistent discrete counterpart of the gradient operator. It satisfies the following commutation property:

$$\forall v \in W^{1,1}(K), \quad G_K \circ \underline{I}_K^k(v) = \Pi_K^k(\nabla v).$$

Given a face $\sigma \in \mathcal{E}_K$, we also define the jump operator $J_{K,\sigma} : \underline{V}_K^k \rightarrow \mathbb{P}^k(\sigma)$ such that, for $\underline{v}_K \in \underline{V}_K^k$,

$$J_{K,\sigma}(\underline{v}_K) = \Pi_\sigma^k(v_K) - v_\sigma. \quad (2.3)$$

Based on the above ingredients, one can define an HHO counterpart of the local diffusion bilinear form $(z, v) \mapsto \int_K \Lambda \nabla z \cdot \nabla v$. We let $a_K : \underline{V}_K^k \times \underline{V}_K^k \rightarrow \mathbb{R}$ be the bilinear form such that

$$a_K : (\underline{z}_K, \underline{v}_K) \mapsto \int_K \Lambda G_K(\underline{z}_K) \cdot G_K(\underline{v}_K) + \sum_{\sigma \in \mathcal{E}_K} \frac{\Lambda_{K,\sigma}}{h_\sigma} \int_\sigma J_{K,\sigma}(\underline{z}_K) J_{K,\sigma}(\underline{v}_K), \quad (2.4)$$

where $\Lambda_{K,\sigma} = \|\Lambda|_K n_{K,\sigma} \cdot n_{K,\sigma}\|_{L^\infty(\sigma)}$ (recall the regularity assumption (2.1)). In the context of HDG methods, the linear stabilisation used in (2.4) is often called Lehrenfeld–Schöberl stabilisation, as it was first introduced in [39, 40]. Classically, one can then define a global bilinear form $a_{\mathcal{D}} : \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k \rightarrow \mathbb{R}$, discretisation of $(z, v) \mapsto \int_\Omega \Lambda \nabla z \cdot \nabla v$, by summing the local contributions:

$$a_{\mathcal{D}} : (\underline{z}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \sum_{K \in \mathcal{M}} a_K(\underline{z}_K, \underline{v}_K). \quad (2.5)$$

For analysis purposes, we introduce a discrete H^1 -like semi-norm on $\underline{V}_{\mathcal{D}}^k$. Given a cell $K \in \mathcal{M}$, we first let, for any $\underline{v}_K \in \underline{V}_K^k$,

$$|\underline{v}_K|_{1,K}^2 = \|\nabla v_K\|_{L^2(K)^d}^2 + \sum_{\sigma \in \mathcal{E}_K} \frac{1}{h_\sigma} \|v_K - v_\sigma\|_{L^2(\sigma)}^2.$$

Then, at the global level, for any $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, we define

$$|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2 = \sum_{K \in \mathcal{M}} |\underline{v}_K|_{1,K}^2. \quad (2.6)$$

Notice that $|\cdot|_{1,\mathcal{D}}$ is not a norm on $\underline{V}_{\mathcal{D}}^k$, but any $\underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ satisfying $|\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}} = 0$ is proportional to $\underline{1}_{\mathcal{D}}$. In particular, this implies that $|\cdot|_{1,\mathcal{D}}$ is a norm on the zero-mass subspace of $\underline{V}_{\mathcal{D}}^k$ defined by

$$\underline{V}_{\mathcal{D},0}^k = \left\{ \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k \mid \int_\Omega v_M = 0 \right\}.$$

Standard HHO analysis implies the following stability estimate:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}^2 \lesssim a_{\mathcal{D}}(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}), \quad (2.7)$$

where the multiplicative constant is proportional to λ_b . In particular, since $|\cdot|_{1,\mathcal{D}}$ is a norm on $\underline{V}_{\mathcal{D},0}^k$, this estimate implies that $a_{\mathcal{D}}$ is coercive on $\underline{V}_{\mathcal{D},0}^k$. Finally, we recall the following discrete Poincaré–Wirtinger inequality (cf. [23, Theorem 6.5, $p = q = 2$] in the equal-order case):

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D},0}^k, \quad \|v_M\|_{L^2(\Omega)} \lesssim |\underline{v}_{\mathcal{D}}|_{1,\mathcal{D}}. \quad (2.8)$$

2.3. Exponential fitting scheme

The construction extends the ideas from [16]. Our (linear) scheme hinges on the exponential fitting strategy [10]. In a nutshell, the exponential fitting approach is based on the following rewriting of the PDE flux: letting $\omega = e^{-\phi}$, and introducing the Slotboom variable $\rho = u/\omega$, one has

$$\Lambda(\nabla u + u\nabla\phi) = \omega\Lambda\nabla\rho, \quad (2.9)$$

which allows to transform an advection-diffusion equation in u into a purely diffusive, unconditionally coercive (by regularity of ϕ , ω is a.e. uniformly bounded away from zero) problem in ρ . At the discrete level, the problem is solved in the Slotboom variable, which is sought in \underline{V}_D^k . The discrete density is then defined mimicking the relation $u = \omega\rho$.

In view of (2.9), in order to define our exponential fitting (mixed-order) HHO scheme, we need to introduce a discrete counterpart of the bilinear form $(\rho, v) \mapsto \int_{\Omega} \omega\Lambda\nabla\rho \cdot \nabla v$. To do so, given $K \in \mathcal{M}$, and leveraging the definition (2.4) of a_K , we let $a_K^\omega : \underline{V}_K^k \times \underline{V}_K^k \rightarrow \mathbb{R}$ be such that

$$a_K^\omega : (\underline{\rho}_K, \underline{v}_K) \mapsto \int_K \omega\Lambda G_K(\underline{\rho}_K) \cdot G_K(\underline{v}_K) + \sum_{\sigma \in \mathcal{E}_K} \frac{\Lambda_{K,\sigma}^\omega}{h_\sigma} \int_\sigma J_{K,\sigma}(\underline{\rho}_K) J_{K,\sigma}(\underline{v}_K), \quad (2.10)$$

where $\Lambda_{K,\sigma}^\omega = \|\omega\Lambda|_K n_{K,\sigma} \cdot n_{K,\sigma}\|_{L^\infty(\sigma)}$. At the global level, as previously, we construct the bilinear form $a_D^\omega : \underline{V}_D^k \times \underline{V}_D^k \rightarrow \mathbb{R}$ by summing the local contributions:

$$a_D^\omega : (\underline{\rho}_D, \underline{v}_D) \mapsto \sum_{K \in \mathcal{M}} a_K^\omega(\underline{\rho}_K, \underline{v}_K). \quad (2.11)$$

We can now introduce the exponential fitting HHO scheme for Problem (1.1): find

$$(\underline{\rho}_D^n)_{n \geq 1} \in (\underline{V}_D^k)^{\mathbb{N}^*}$$

such that, for all $n \in \mathbb{N}$,

$$\left\{ \begin{array}{l} \sum_{K \in \mathcal{M}} \int_K \frac{u_K^{\omega,n+1} - u_K^{\omega,n}}{\Delta t} v_K + a_D^\omega(\underline{\rho}_D^{n+1}, \underline{v}_D) = 0, \quad \forall \underline{v}_D \in \underline{V}_D^k, \quad (2.12a) \\ u_K^{\omega,n+1} = \omega|_K \rho_K^{n+1}, \quad \forall K \in \mathcal{M}, \quad (2.12b) \\ u_K^{\omega,0} = u_K^0, \quad \forall K \in \mathcal{M}. \quad (2.12c) \end{array} \right.$$

For any solution $(\underline{\rho}_D^n)_{n \geq 1}$ to (2.12), we define a sequence of corresponding densities $(\underline{u}_D^{\omega,n})_{n \geq 1}$ as follows. To the discrete Slotboom variable $\underline{\rho}_D \in \underline{V}_D^k$ we associate the discrete density

$$\underline{u}_D^\omega = ((u_K^\omega)_{K \in \mathcal{M}}, (u_\sigma^\omega)_{\sigma \in \mathcal{E}}),$$

defined, consistently with (2.12b), as the collection of (a priori non-polynomial) functions

$$u_K^\omega = \omega|_K \rho_K \quad \text{for all } K \in \mathcal{M} \quad \text{and} \quad u_\sigma^\omega = \omega|_\sigma \rho_\sigma \quad \text{for all } \sigma \in \mathcal{E}. \quad (2.13)$$

The non-polynomial nature of the components of \underline{u}_D^ω is, here and in what follows, emphasised by the use of Gothic fonts. Finally, to any discrete density \underline{u}_D^ω , we associate the two piecewise smooth functions $u_M^\omega : \Omega \rightarrow \mathbb{R}$ and $u_\mathcal{E}^\omega : \partial\mathcal{M} \rightarrow \mathbb{R}$ such that

$$u_{M|K}^\omega = u_K^\omega \quad \text{for all } K \in \mathcal{M} \quad \text{and} \quad u_{\mathcal{E}|\sigma}^\omega = u_\sigma^\omega \quad \text{for all } \sigma \in \mathcal{E}.$$

Remark that $u_M^\omega = \omega\rho_M$ in Ω and $u_\mathcal{E}^\omega = \omega|_{\partial\mathcal{M}}\rho_\mathcal{E}$ on $\partial\mathcal{M}$.

Remark 2 (Non-polynomial definition). *Here, we choose to define a discrete density with (a priori) non-polynomial components. One could also think of defining a density with polynomial components, by multiplying (component by component) $\underline{\rho}_{\mathcal{D}}$ by $\underline{I}_{\mathcal{D}}^k(\omega)$. This is how the solution to the low-order HFV exponential fitting scheme of [16] was defined (in that case, both cell/face unknowns were constants).*

Remark 3 (Initial condition). *Notice that we do not define $u_{\mathcal{M}}^{\omega,0}$ in the same way as $u_{\mathcal{M}}^{\omega,n+1}$. Indeed, we directly use in the definition (2.12c) the initial datum u^0 . This choice is motivated by the following observation. Define $\rho^0 = u^0/\omega$, and let $u_K^{\omega,0} = \omega|_K \Pi_K^{k+1}(\rho|_K^0)$ for all $K \in \mathcal{M}$ in place of (2.12c). Then, the resulting discrete solution's mass is $M^\omega = \sum_{K \in \mathcal{M}} \int_K u_K^{\omega,0} \neq M$ in general.*

Recall that $\int_{\Omega} u^0 = M > 0$. Let $\underline{\rho}_{\mathcal{D}}^\infty \in \underline{V}_{\mathcal{D}}^k$ be defined as

$$\underline{\rho}_{\mathcal{D}}^\infty = c_l^M \underline{1}_{\mathcal{D}}, \quad \text{with } c_l^M = \frac{M}{\int_{\Omega} e^{-\phi}} > 0.$$

One can easily check that $\underline{\rho}_{\mathcal{D}}^\infty$ is the only steady solution to the exponential fitting scheme (2.12). Based on (2.13), $\underline{\rho}_{\mathcal{D}}^\infty$ is associated to the discrete equilibrium density $\underline{u}_{\mathcal{D}}^{\omega,\infty}$ such that

$$u_{\mathcal{M}}^{\omega,\infty} = c_l^M e^{-\phi} \text{ in } \Omega \quad \text{and} \quad u_{\mathcal{E}}^{\omega,\infty} = c_l^M e^{-\phi|_{\partial\mathcal{M}}} \text{ on } \partial\mathcal{M}. \quad (2.14)$$

It follows that the reconstructed discrete equilibrium density $u_{\mathcal{M}}^{\omega,\infty}$ (always) coincides with the thermal equilibrium (1.4) in Ω . Such a striking property is, however, to be tempered by Remark 4 below. Following Remark 2, notice that if we had adopted instead a polynomial definition for the discrete densities, we would have obtained that $\underline{u}_{\mathcal{D}}^{\omega,\infty} = \underline{I}_{\mathcal{D}}^k(u^\infty) \in \underline{V}_{\mathcal{D}}^k$, as was the case for the low-order exponential fitting HFV method of [16] (in that case, both cell/face unknowns were constants). A drawback of such a definition, compared to (2.13), is that the components of $\underline{I}_{\mathcal{D}}^k(u^\infty)$ are not necessarily positive functions (note that this issue does not exist in the low- and equal-order HFV case).

Remark 4 (Alternative scheme definition). *Let $\underline{\phi}_{\mathcal{D}} = \underline{I}_{\mathcal{D}}^k(\phi) \in \underline{V}_{\mathcal{D}}^k$. Another definition of the exponential fitting HHO scheme consists in replacing ω by $e^{-\phi_{\mathcal{M}}}$ in the expressions of both a_K^ω (see (2.10)) and $u_K^{\omega,n+1}$ (see (2.12b)). Then, in place of (2.13), the following definition of discrete densities is adopted:*

$$u_K^\omega = e^{-\phi_K} \rho_K \text{ for all } K \in \mathcal{M} \quad \text{and} \quad u_\sigma^\omega = e^{-\phi_\sigma} \rho_\sigma \text{ for all } \sigma \in \mathcal{E}. \quad (2.15)$$

Such a scheme is somewhat closer to what one would encounter in the context of semiconductor models, since ϕ would be unknown, and sought, at the discrete level, in $\underline{V}_{\mathcal{D}}^k$. In this case, the discrete equilibrium density $\underline{u}_{\mathcal{D}}^{\omega,\infty}$ would satisfy, in place of (2.14), the same kind of relations as (2.22) and (2.23) below. Here, we rather choose to exploit the full knowledge we have of the potential ϕ to define the scheme.

2.4. Nonlinear scheme

The construction extends the ideas from [16, 45]. Our nonlinear scheme relies on a nonlinear reformulation of Problem (1.1) [13]. To do so, we introduce the logarithm potential $\ell = \log(u)$ and the quasi-Fermi potential $w = \ell + \phi$. At least formally, if u is positive, one has the following relation on the PDE flux:

$$\Lambda(\nabla u + u\nabla\phi) = u\Lambda\nabla(\log(u) + \phi) = e^\ell \Lambda\nabla w. \quad (2.16)$$

We choose to discretise the potentials as piecewise polynomials, i.e., we approximate ℓ and w as discrete unknowns in $\underline{V}_{\mathcal{D}}^k$. Then, mimicking the relation $u = e^\ell$, each discrete density component is defined as the exponential of a polynomial, thus ensuring its positivity.

In view of (2.16), in order to define our nonlinear HHO scheme, we shall introduce a discrete counterpart of the map $(\ell; w, v) \mapsto \int_{\Omega} e^\ell \Lambda \nabla w \cdot \nabla v$. Locally to any cell $K \in \mathcal{M}$, this discrete counterpart is built as the sum of a consistent (2.17a) and a stabilising (2.17b) contributions: for all $\underline{\ell}_K, \underline{w}_K, \underline{v}_K \in \underline{V}_K^k$, we let

$$C_K(\underline{\ell}_K; \underline{w}_K, \underline{v}_K) = \int_K e^{\underline{\ell}_K} \Lambda G_K(\underline{w}_K) \cdot G_K(\underline{v}_K), \quad (2.17a)$$

$$S_K(\underline{\ell}_K; \underline{w}_K, \underline{v}_K) = \sum_{\sigma \in \mathcal{E}_K} \frac{\Lambda_{K,\sigma}}{h_\sigma} \int_{\sigma} \frac{e^{\Pi_{\sigma}^k(\underline{\ell}_K)} + e^{\underline{\ell}_\sigma}}{2} J_{K,\sigma}(\underline{w}_K) J_{K,\sigma}(\underline{v}_K). \quad (2.17b)$$

We then introduce the local map $\mathcal{T}_K : \underline{V}_K^k \times \underline{V}_K^k \times \underline{V}_K^k \rightarrow \mathbb{R}$ such that

$$\mathcal{T}_K : (\underline{\ell}_K; \underline{w}_K, \underline{v}_K) \mapsto C_K(\underline{\ell}_K; \underline{w}_K, \underline{v}_K) + S_K(\underline{\ell}_K; \underline{w}_K, \underline{v}_K) + \varepsilon h_K^{k+2} a_K(\underline{w}_K, \underline{v}_K), \quad (2.18)$$

where ε is a non-negative parameter and a_K is the bilinear form defined by (2.4). At the global level, we finally define the map $\mathcal{T}_{\mathcal{D}} : \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k \times \underline{V}_{\mathcal{D}}^k \rightarrow \mathbb{R}$ by summing the local contributions:

$$\mathcal{T}_{\mathcal{D}} : (\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \sum_{K \in \mathcal{M}} \mathcal{T}_K(\underline{\ell}_K; \underline{w}_K, \underline{v}_K). \quad (2.19)$$

Remark 5 (Parameter ε). *The map $\mathcal{T}_{\mathcal{D}}$ is to be understood as a discretisation of $(\ell; w, v) \mapsto \int_{\Omega} (e^\ell + \varepsilon) \Lambda \nabla w \cdot \nabla v$, with ε of magnitude $\varepsilon h_{\mathcal{D}}^{k+2}$. At the theoretical level, this ε -perturbation of the model is necessary, at the moment, to show the existence result of Theorem 1. From a more practical viewpoint, the sensitivity of the method with respect to ε is not completely understood yet. First numerical experiments tend to show that the choices $\varepsilon = 1$ and $\varepsilon = 0$ produce essentially similar results. Concerning the choice of scaling factor h_K^{k+2} in (2.18), it seems to yield in practice (when $\varepsilon > 0$) the expected orders of convergence. The influence of the ε -term will be further investigated in future works.*

Let $\underline{\phi}_{\mathcal{D}} = \underline{I}_{\mathcal{D}}^k(\phi) \in \underline{V}_{\mathcal{D}}^k$. We can now introduce our nonlinear HHO scheme for Problem (1.1): find

$$\left(\underline{\ell}_{\mathcal{D}}^n \right)_{n \geq 1} \in \left(\underline{V}_{\mathcal{D}}^k \right)^{\mathbb{N}^*}$$

such that, for all $n \in \mathbb{N}$,

$$\left\{ \begin{array}{l} \sum_{K \in \mathcal{M}} \int_K \frac{u_K^{n+1} - u_K^n}{\Delta t} v_K + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{n+1}; \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = 0, \quad \forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad (2.20a) \\ u_K^{n+1} = e^{\underline{\ell}_K^{n+1}}, \quad \forall K \in \mathcal{M}, \quad (2.20b) \\ u_K^0 = u_{|K}^0, \quad \forall K \in \mathcal{M}. \quad (2.20c) \end{array} \right.$$

For any solution $\left(\underline{\ell}_{\mathcal{D}}^n \right)_{n \geq 1}$ to (2.20), we define a sequence of corresponding positive densities $\left(\underline{u}_{\mathcal{D}}^n \right)_{n \geq 1}$ as follows. To the discrete logarithm potential $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ we associate the discrete density

$$\underline{u}_{\mathcal{D}} = ((u_K)_{K \in \mathcal{M}}, (u_\sigma)_{\sigma \in \mathcal{E}}),$$

defined, consistently with (2.20b), as the collection of positive (non-polynomial) functions

$$u_K = e^{\ell_K} \text{ for all } K \in \mathcal{M} \quad \text{and} \quad u_\sigma = e^{\ell_\sigma} \text{ for all } \sigma \in \mathcal{E}. \quad (2.21)$$

The non-polynomial nature of the components of $\underline{u}_{\mathcal{D}}$ is, here also, emphasised by the use of Gothic fonts. Finally, to any discrete density $\underline{u}_{\mathcal{D}}$, we associate the two (positive) piecewise smooth functions $u_{\mathcal{M}} : \Omega \rightarrow \mathbb{R}^*$ and $u_{\mathcal{E}} : \partial\mathcal{M} \rightarrow \mathbb{R}^*$ such that

$$u_{\mathcal{M}|K} = u_K \text{ for all } K \in \mathcal{M} \quad \text{and} \quad u_{\mathcal{E}|\sigma} = u_\sigma \text{ for all } \sigma \in \mathcal{E}.$$

Remark that $u_{\mathcal{M}} = e^{\ell_{\mathcal{M}}}$ in Ω and $u_{\mathcal{E}} = e^{\ell_{\mathcal{E}}}$ on $\partial\mathcal{M}$.

Remark 6 (Initial condition). *Remark that we do not define $u_{\mathcal{M}}^0$ in the same way as $u_{\mathcal{M}}^{n+1}$. We indeed directly use in the definition (2.20c) the initial datum u^0 . This strategy allows one to circumvent the definition of some $\ell_{\mathcal{M}}^0$, cell interpolate of $\ell^0 = \log(u^0)$, the latter quantity being undefined in regions where u^0 vanishes. The question of defining an initial discrete logarithm potential remains however a major difficulty when it comes to numerical implementation, since it is needed for the initialisation of the Newton method when $n = 0$ (see Section 4.1.2).*

Recall that $\int_{\Omega} u^0 = M > 0$. Let $\underline{\ell}_{\mathcal{D}}^{\infty} \in \underline{V}_{\mathcal{D}}^k$ be defined as

$$\underline{\ell}_{\mathcal{D}}^{\infty} = \log(c_{nl}^M) \underline{1}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}}, \quad \text{with } c_{nl}^M = \frac{M}{\int_{\Omega} e^{-\phi_M}} > 0.$$

It can be easily checked that $\underline{\ell}_{\mathcal{D}}^{\infty}$ is the only steady solution to the nonlinear scheme (2.20). Based on (2.21), $\underline{\ell}_{\mathcal{D}}^{\infty}$ is associated to the discrete equilibrium density $\underline{u}_{\mathcal{D}}^{\infty}$ such that

$$u_{\mathcal{M}}^{\infty} = c_{nl}^M e^{-\phi_M} \text{ in } \Omega \quad \text{and} \quad u_{\mathcal{E}}^{\infty} = c_{nl}^M e^{-\phi_{\mathcal{E}}} \text{ on } \partial\mathcal{M}. \quad (2.22)$$

In a sense we make clear just below, the discrete equilibrium density $\underline{u}_{\mathcal{D}}^{\infty}$ is a reasonable approximation of the thermal equilibrium u^{∞} defined by (1.4):

$$\log(u_K^{\infty}) - \Pi_K^{k+1}(\log(u^{\infty})) = \log\left(\int_{\Omega} e^{-\phi}\right) - \log\left(\int_{\Omega} e^{-\phi_M}\right), \quad \forall K \in \mathcal{M}, \quad (2.23a)$$

$$\log(u_{\sigma}^{\infty}) - \Pi_{\sigma}^k(\log(u^{\infty})) = \log\left(\int_{\Omega} e^{-\phi}\right) - \log\left(\int_{\Omega} e^{-\phi_M}\right), \quad \forall \sigma \in \mathcal{E}. \quad (2.23b)$$

It follows that the reconstructed discrete equilibrium density $u_{\mathcal{M}}^{\infty}$ satisfies: if $\phi|_K \in \mathbb{P}^{k+1}(K)$ for all $K \in \mathcal{M}$, then $u_{\mathcal{M}}^{\infty} = u^{\infty}$ in Ω . Remark that the discrete equilibrium density $\underline{u}_{\mathcal{D}}^{\infty}$ is not equal to $\underline{I}_{\mathcal{D}}^k(u^{\infty}) \in \underline{V}_{\mathcal{D}}^k$. This is in contrast with what held true for the low-order nonlinear HFV method of [16] (in that case, both cell/face unknowns were constants). This can be explained by the choice of discretisation for ϕ , which was taken as $-\log(\underline{\omega}_{\mathcal{D}})$ (with $\underline{\omega}_{\mathcal{D}}$ HFV interpolate of ω) in [16] in place of $\underline{\phi}_{\mathcal{D}}$ here (the latter choice is inspired by [45] in the context of semiconductors). Remark that, in practice, $-\log(\underline{\omega}_{\mathcal{D}})$ and $\underline{\phi}_{\mathcal{D}}$ may coincide if the integrals are approximated using an evaluation at the barycenter. This was the case in the numerical experiments of [16].

3. Main features of the schemes

We present in this section some theoretical results about the two schemes introduced above. We focus, in particular, on the existence (and stability) of solutions, as well as on questions related to their long-time behaviour. The results presented below generalise those obtained in [16] in the low-order HFV context. In particular, the analysis strongly hinges on the entropy structure of both schemes.

Remark 7 (Lowest-order versions of the schemes ($k = 0$)). *Note that the lowest-order versions of the two schemes introduced above do not coincide with the exponential fitting and nonlinear HFV schemes of [16]. Indeed, the lowest-order versions of the methods (2.12) and (2.20) make use of (enriched) affine cell unknowns, whereas HFV schemes use constants. Also, whereas the nonlinear HFV method is built upon a stable discrete gradient operator (defined on a pyramidal submesh), the present nonlinear scheme is defined following the standard HHO philosophy of splitting consistency and (nonlinear) stabilisation. Therefore, the results presented here are new, even for $k = 0$.*

Before presenting individual results for each scheme, let us stress that both schemes exhibit a similar important property: the preservation of the thermal equilibrium.

Proposition 1 (Preservation of the thermal equilibrium). *The alternative (fully discrete) exponential fitting scheme of Remark 4 and the nonlinear scheme (2.20) preserve the thermal equilibrium in the sense of Definition 1. More precisely,*

- $\log(\underline{u}_{\mathcal{D}}^{\omega, \infty}) + \underline{\phi}_{\mathcal{D}}$ is proportional to $\underline{1}_{\mathcal{D}}$;
- $\underline{\ell}_{\mathcal{D}}^{\infty} + \underline{\phi}_{\mathcal{D}}$ is proportional to $\underline{1}_{\mathcal{D}}$.

For the original exponential fitting scheme (2.12), there holds that $\log(u_{\mathcal{M}}^{\omega, \infty}) + \phi$ over Ω and $\log(u_{\mathcal{E}}^{\omega, \infty}) + \phi|_{\partial\mathcal{M}}$ over $\partial\mathcal{M}$ are equal to the same constant.

3.1. Exponential fitting scheme

We present here the main properties of the exponential fitting HHO scheme (2.12), and give detailed proofs of the results. As a preliminary remark, note that since $\phi \in W^{1, \infty}(\Omega)$,

$$\|\phi\|_{L^{\infty}(\Omega)} \lesssim 1. \quad (3.1)$$

As a consequence, the tensor $\omega\Lambda$ is a.e. uniformly elliptic. Recalling (2.7), this implies the following stability estimate:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad |\underline{v}_{\mathcal{D}}|_{1, \mathcal{D}}^2 \lesssim a_{\mathcal{D}}^{\omega}(\underline{v}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}), \quad (3.2)$$

where the multiplicative constant is proportional to λ_b . We first state a well-posedness result, which is mainly a consequence of the previous stability estimate.

Proposition 2 (Well-posedness of the exponential fitting scheme). *The exponential fitting scheme (2.12) admits a unique solution $(\underline{\rho}_{\mathcal{D}}^n)_{n \geq 1}$. Moreover, the corresponding discrete densities $(\underline{u}_{\mathcal{D}}^{\omega, n})_{n \geq 1}$ have a mass equal to M :*

$$\forall n \geq 1, \quad \int_{\Omega} u_{\mathcal{M}}^{\omega, n} = \int_{\Omega} u^0 = M. \quad (3.3)$$

Proof. Let $n \geq 0$, and assume that $u_M^{\omega,n}$ is defined. We want to show that the system (2.12a)-(2.12b) admits a unique solution. To do so, we first define, for any $\underline{\rho}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$,

$$\|\underline{\rho}_{\mathcal{D}}\|_{\omega,\Delta t,\mathcal{D}}^2 = \Delta t |\underline{\rho}_{\mathcal{D}}|_{1,\mathcal{D}}^2 + \int_{\Omega} \omega \rho_{\mathcal{M}}^2.$$

Since $|\cdot|_{1,\mathcal{D}}$ is a semi-norm on $\underline{V}_{\mathcal{D}}^k$ with zero set spanned by $\underline{1}_{\mathcal{D}}$, it follows that the map $\|\cdot\|_{\omega,\Delta t,\mathcal{D}}$ defines a norm on $\underline{V}_{\mathcal{D}}^k$. Thus, by (3.2), the bilinear form $A_{\mathcal{D}}^{\omega} : (\underline{\rho}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) \mapsto \int_{\Omega} \omega \rho_{\mathcal{M}} v_{\mathcal{M}} + \Delta t a_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}, \underline{v}_{\mathcal{D}})$ satisfies the following coercivity property:

$$\forall \underline{\rho}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad \|\underline{\rho}_{\mathcal{D}}\|_{\omega,\Delta t,\mathcal{D}}^2 \lesssim A_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}, \underline{\rho}_{\mathcal{D}}).$$

By the Lax-Milgram lemma, the system (2.12a)-(2.12b) therefore admits a unique solution $\underline{\rho}_{\mathcal{D}}^{n+1}$ in $\underline{V}_{\mathcal{D}}^k$, from which one can uniquely define $\underline{u}_{\mathcal{D}}^{\omega,n+1}$ by (2.13). To prove mass conservation, we just test (2.12a) by $\underline{1}_{\mathcal{D}}$ to get

$$\int_{\Omega} \frac{u_{\mathcal{M}}^{\omega,n+1} - u_{\mathcal{M}}^{\omega,n}}{\Delta t} = 0.$$

We conclude by noticing that $\int_{\Omega} u_{\mathcal{M}}^{\omega,0} = \int_{\Omega} u^0$ according to (2.12c). \square

We now state our main result about the exponential fitting scheme, which ensures that the solution to (2.12) has similar long-time behaviour as the PDE solution. As usual with the entropy method, the main idea is to get a control of the entropy by its dissipation. Here, such an estimate is a consequence of the discrete Poincaré inequality (2.8).

Proposition 3 (Long-time behaviour of the exponential fitting scheme). *Assume that $u^0 \in L^2(\Omega)$. Let $(\underline{\rho}_{\mathcal{D}}^n)_{n \geq 1}$ be the solution to the exponential fitting scheme (2.12). Then, the following discrete entropy relation holds true:*

$$\forall n \in \mathbb{N}, \quad \frac{\mathbb{E}_{\omega}^{n+1} - \mathbb{E}_{\omega}^n}{\Delta t} + \mathbb{D}_{\omega}^{n+1} \leq 0, \quad (3.4)$$

where the discrete quadratic entropy is defined as

$$\mathbb{E}_{\omega}^n = \frac{1}{2} \int_{\Omega} \omega (\rho_{\mathcal{M}}^n - \rho_{\mathcal{M}}^{\infty})^2$$

with $\rho_{\mathcal{M}}^0$ defined (with a slight abuse in notation, since $\rho_{\mathcal{M}}^0$ is not piecewise polynomial) by $\rho_{\mathcal{M}}^0 = \frac{u^0}{\omega}$, and the discrete dissipation is given by

$$\mathbb{D}_{\omega}^n = a_{\mathcal{D}}^{\omega}(\underline{\rho}_{\mathcal{D}}^n - \underline{\rho}_{\mathcal{D}}^{\infty}, \underline{\rho}_{\mathcal{D}}^n - \underline{\rho}_{\mathcal{D}}^{\infty}), \quad \forall n \geq 1.$$

As a consequence, the reconstructed discrete density converges exponentially fast in time towards the reconstructed discrete equilibrium density: there exists a positive constant ν_{ω} , independent of both $h_{\mathcal{D}}$ and Δt , such that

$$\forall n \in \mathbb{N}, \quad \|u_{\mathcal{M}}^{\omega,n} - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)} \lesssim (1 + \nu_{\omega} \Delta t)^{-\frac{n}{2}} \|u^0 - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)}. \quad (3.5)$$

Proof. Let $n \in \mathbb{N}$. By convexity of $x \mapsto x^2$ on \mathbb{R} , one has

$$\mathbb{E}_\omega^{n+1} - \mathbb{E}_\omega^n = \frac{1}{2} \int_\Omega \omega \left((\rho_M^{n+1} - \rho_M^\infty)^2 - (\rho_M^n - \rho_M^\infty)^2 \right) \leq \int_\Omega \omega (\rho_M^{n+1} - \rho_M^n) (\rho_M^{n+1} - \rho_M^\infty).$$

Therefore, testing (2.12a) against $\underline{\rho}_D^{n+1} - \underline{\rho}_D^\infty \in \underline{V}_D^k$, we get

$$\frac{\mathbb{E}_\omega^{n+1} - \mathbb{E}_\omega^n}{\Delta t} \leq -a_D^\omega(\underline{\rho}_D^{n+1}, \underline{\rho}_D^{n+1} - \underline{\rho}_D^\infty).$$

Note that this estimate holds true also for $n = 0$ (using the definition of ρ_M^0). On the other hand, by the expression of $\underline{\rho}_D^\infty$ (proportional to $\underline{1}_D$), $a_D^\omega(\underline{\rho}_D^\infty, \underline{\rho}_D^{n+1} - \underline{\rho}_D^\infty) = 0$, hence by bilinearity of a_D^ω ,

$$\mathbb{D}_\omega^{n+1} = a_D^\omega(\underline{\rho}_D^{n+1}, \underline{\rho}_D^{n+1} - \underline{\rho}_D^\infty),$$

which yields the entropy relation (3.4). To get the exponential decay, one needs to compare \mathbb{D}_ω^{n+1} with \mathbb{E}_ω^{n+1} . To do so, we let $\underline{v}_D = \underline{\rho}_D^{n+1} - \underline{\rho}_D^\infty \in \underline{V}_D^k$, and we define the probability measure $d\mu = \frac{c_l^M}{M} \omega dx$ on Ω . We define $\langle v_M \rangle_\mu$ as the mass of v_M for the measure $d\mu$, i.e.,

$$\langle v_M \rangle_\mu = \int_\Omega v_M d\mu.$$

The definition (2.13), and the mass preservation identity (3.3), imply that

$$\langle v_M \rangle_\mu = \int_\Omega v_M d\mu = \frac{c_l^M}{M} \int_\Omega \omega v_M = \frac{c_l^M}{M} \int_\Omega (u_M^{\omega, n+1} - u_M^{\omega, \infty}) = c_l^M \frac{M - M}{M} = 0.$$

Therefore, letting $\langle v_M \rangle = \frac{1}{|\Omega|_d} \int_\Omega v_M$, and applying [11, Lemma 5.2, $q = 2$], we get

$$\frac{c_l^M}{M} \int_\Omega \omega v_M^2 = \int_\Omega (v_M - \langle v_M \rangle_\mu)^2 d\mu \leq 4 \int_\Omega (v_M - \langle v_M \rangle)^2 d\mu.$$

Using the definition of $d\mu$, and the bound (3.1), yields

$$\int_\Omega \omega v_M^2 \lesssim \int_\Omega (v_M - \langle v_M \rangle)^2 = \|v_M - \langle v_M \rangle\|_{L^2(\Omega)}^2.$$

By definition of $\langle v_M \rangle$, one has $\underline{v}_D - \langle v_M \rangle \underline{1}_D \in \underline{V}_{D,0}^k$, so we can apply the discrete Poincaré–Wirtinger inequality (2.8) to infer that

$$\|v_M - \langle v_M \rangle\|_{L^2(\Omega)}^2 \lesssim |\underline{v}_D - \langle v_M \rangle \underline{1}_D|_{1,D}^2 = |\underline{v}_D|_{1,D}^2.$$

Combining the two previous estimates, we get

$$\mathbb{E}_\omega^{n+1} = \frac{1}{2} \int_\Omega \omega v_M^2 \lesssim |\underline{v}_D|_{1,D}^2.$$

Now, one can use the stability estimate (3.2) to infer that

$$|\underline{v}_D|_{1,D}^2 \lesssim a_D^\omega(\underline{v}_D, \underline{v}_D) = \mathbb{D}_\omega^{n+1}.$$

Therefore, combining the last two estimates, one infers the existence of $\nu_\omega > 0$, independent of both $h_{\mathcal{D}}$ and Δt , such that the following relation between the entropy and its dissipation holds true:

$$\nu_\omega \mathbb{E}_\omega^{n+1} \leq \mathbb{D}_\omega^{n+1}.$$

Plugging this estimate into the entropy relation (3.4), we deduce that

$$(1 + \nu_\omega \Delta t) \mathbb{E}_\omega^{n+1} \leq \mathbb{E}_\omega^n.$$

This implies the exponential decay of the entropy:

$$\forall n \geq 0, \quad \mathbb{E}_\omega^n \leq (1 + \nu_\omega \Delta t)^{-n} \mathbb{E}_\omega^0.$$

To conclude, we just use (2.13) and (3.1) to infer that

$$\|u_{\mathcal{M}}^{\omega,n} - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)}^2 \lesssim \mathbb{E}_\omega^n \lesssim \|u_{\mathcal{M}}^{\omega,n} - u_{\mathcal{M}}^{\omega,\infty}\|_{L^2(\Omega)}^2,$$

which, combined with the fact that $u_{\mathcal{M}}^{\omega,0} = u^0$, finally yields (3.5). \square

Remark 8 (Regularity of the initial datum and topology of the convergence). *Notice that in Proposition 3 we have made the extra assumption that $u^0 \in L^2(\Omega)$. The long-time analysis of the exponentially fitted model indeed relies on the decay of the quadratic entropy (in the Slotboom variable)*

$$\mathbb{E}_\omega(t) = \frac{1}{2} \int_{\Omega} \omega(\rho(t) - \rho^\infty)^2.$$

In order to guarantee that the initial quadratic entropy is finite, assuming that the initial datum is in $L^2(\Omega)$ is a safe choice. At the end, as a reminiscence of the linearity of the model, the exponential fitting approach gives a convergence (in time) result in the L^2 -topology (in space). In contrast, the nonlinear approach will yield convergence in a weaker norm (typically L^1), but can be used to deal with less regular initial data, which are in $L \log(L)$ only.

3.2. Nonlinear scheme

We present here some results regarding the analysis of the nonlinear HHO method (2.20). Since we deal with a nonlinear scheme, unlike the exponential fitting scheme, the question of the existence of solutions is the main difficulty here. As often for this type of method, we start by establishing some a priori estimates. For the purpose of analysis, given a discrete logarithm potential $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, we associate a discrete quasi-Fermi potential $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ defined by

$$\underline{w}_{\mathcal{D}} = \underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}} - \log(c_{nl}^M) \underline{1}_{\mathcal{D}}, \quad (3.6)$$

where we recall that $c_{nl}^M = M / \int_{\Omega} e^{-\phi_M}$. By (2.21) and (2.22), one has

$$w_{\mathcal{M}} = \log\left(\frac{u_{\mathcal{M}}}{u_{\mathcal{M}}^\infty}\right) \text{ in } \Omega \quad \text{and} \quad w_{\mathcal{E}} = \log\left(\frac{u_{\mathcal{E}}}{u_{\mathcal{E}}^\infty}\right) \text{ on } \partial\mathcal{M}.$$

Note that, on the other hand, for any $\underline{\ell}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, we have

$$\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}) = \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}), \quad (3.7)$$

since $(\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}) - \underline{w}_{\mathcal{D}}$ is proportional to $\underline{1}_{\mathcal{D}}$. Similarly to previous works on nonlinear HFV schemes for semiconductor models [45], the discrete quasi-Fermi potentials are the key variables to perform the analysis of the method. As a last remark, notice that since $\phi \in H^1(\Omega)$, by boundedness of the interpolator (cf. [23, Proposition 5.3]),

$$|\underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}} \lesssim 1. \quad (3.8)$$

Let us now state some fundamental a priori relations. As for the exponential fitting scheme, the discrete entropy structure of the nonlinear scheme mainly results from the convexity of the entropy.

Proposition 4 (Fundamental a priori relations). *Let $(\underline{\ell}_{\mathcal{D}}^n)_{n \geq 1}$ be a given solution to the nonlinear scheme (2.20), and $(\underline{u}_{\mathcal{D}}^n)_{n \geq 1}$ be the corresponding discrete density. Then, the following a priori relations hold true:*

(i) *the mass is preserved along time:*

$$\forall n \geq 1, \quad \int_{\Omega} u_{\mathcal{M}}^n = \int_{\Omega} u^0 = M; \quad (3.9)$$

(ii) *a discrete entropy/dissipation relation is satisfied:*

$$\forall n \in \mathbb{N}, \quad \frac{\mathbb{E}^{n+1} - \mathbb{E}^n}{\Delta t} + \mathbb{D}^{n+1} \leq 0, \quad (3.10)$$

where the discrete entropy and dissipation are non-negative quantities defined by

$$\mathbb{E}^n = \int_{\Omega} u_{\mathcal{M}}^{\infty} \Phi_1 \left(\frac{u_{\mathcal{M}}^n}{u_{\mathcal{M}}^{\infty}} \right) \quad \text{and} \quad \mathbb{D}^n = \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^n; \underline{w}_{\mathcal{D}}^n, \underline{w}_{\mathcal{D}}^n) \quad \text{for } n \geq 1,$$

with $\Phi_1 : s \mapsto s \log(s) - s + 1$ (and $\Phi_1(0) = 1$).

Proof. Let $n \geq 0$. Using $\underline{1}_{\mathcal{D}}$ as a test function in (2.20a), we get that the mass is conserved:

$$\int_{\Omega} u_{\mathcal{M}}^{n+1} = \int_{\Omega} u_{\mathcal{M}}^n.$$

Therefore, by (2.20c), we infer (3.9). To establish the entropy relation, we first use the convexity of Φ_1 , which yields

$$\mathbb{E}^{n+1} - \mathbb{E}^n \leq \int_{\Omega} u_{\mathcal{M}}^{\infty} \Phi_1' \left(\frac{u_{\mathcal{M}}^{n+1}}{u_{\mathcal{M}}^{\infty}} \right) \frac{u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n}{u_{\mathcal{M}}^{\infty}}.$$

Then, since $w_{\mathcal{M}}^{n+1} = \log \left(\frac{u_{\mathcal{M}}^{n+1}}{u_{\mathcal{M}}^{\infty}} \right)$ and $\Phi_1' = \log$, one has

$$\mathbb{E}^{n+1} - \mathbb{E}^n \leq \int_{\Omega} w_{\mathcal{M}}^{n+1} (u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n). \quad (3.11)$$

On the other hand, testing (2.20a) with $\underline{w}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^k$, and using (3.7), we get

$$\int_{\Omega} w_{\mathcal{M}}^{n+1} (u_{\mathcal{M}}^{n+1} - u_{\mathcal{M}}^n) = -\Delta t \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{n+1}; \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}^{n+1}) = -\Delta t \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{n+1}; \underline{w}_{\mathcal{D}}^{n+1}, \underline{w}_{\mathcal{D}}^{n+1}),$$

which finally yields (3.10) by definition of the discrete dissipation. \square

Remark that since $u_M^0 = u^0$, and $u^0 \geq 0$ in Ω , $u^0 \in L^1(\Omega)$, and $\int_{\Omega} u^0 \log(u^0) < \infty$, one has $\mathbb{E}^0 < \infty$. Note finally that the previous results hold true for any $\varepsilon \geq 0$ in (2.18).

In the rest of this section, we focus on the existence of solutions and on their long-time behaviour. We henceforth assume that $\varepsilon > 0$. The proofs for both results rely on a discrete a priori estimate, which is obtained by means of a high-order counterpart of [16, Lemma 2]. In order to perform the analysis, we first introduce an inner product $\langle \cdot, \cdot \rangle$ on $\underline{V}_{\mathcal{D}}^k$:

$$\forall \underline{z}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad \langle \underline{z}_{\mathcal{D}}, \underline{v}_{\mathcal{D}} \rangle = \sum_{K \in \mathcal{M}} \left(\int_K z_K v_K + \sum_{\sigma \in \mathcal{E}_K} h_{\sigma} \int_{\sigma} (z_K - z_{\sigma})(v_K - v_{\sigma}) \right).$$

We denote by $\|\cdot\|$ the corresponding Euclidean norm:

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad \|\underline{v}_{\mathcal{D}}\|^2 = \sum_{K \in \mathcal{M}} \left(\|v_K\|_{L^2(K)}^2 + \sum_{\sigma \in \mathcal{E}_K} h_{\sigma} \|v_K - v_{\sigma}\|_{L^2(\sigma)}^2 \right).$$

Lemma 1 (Discrete boundedness by mass and energy semi-norm). *Let $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, and assume that there exist $C_{\sharp} > 0$ and $M_{\sharp} \geq M_b > 0$ such that*

$$M_b \leq \int_{\Omega} e^{\ell_{\mathcal{M}}} \leq M_{\sharp} \quad \text{and} \quad |\underline{\ell}_{\mathcal{D}}|_{1,\mathcal{D}} \leq C_{\sharp}. \quad (3.12)$$

Then, there exists a positive constant C , only depending on M_b , M_{\sharp} , C_{\sharp} , Ω , θ , k and h_b such that

$$\|\underline{\ell}_{\mathcal{D}}\| \leq C.$$

Proof. Let us first remark that

$$\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} h_{\sigma} \|\ell_K - \ell_{\sigma}\|_{L^2(\sigma)}^2 \leq h_{\mathcal{D}}^2 \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \frac{1}{h_{\sigma}} \|\ell_K - \ell_{\sigma}\|_{L^2(\sigma)}^2 \leq h_{\mathcal{D}}^2 |\underline{\ell}_{\mathcal{D}}|_{1,\mathcal{D}}^2 \leq \text{diam}(\Omega)^2 C_{\sharp}^2.$$

Hence, to estimate $\|\underline{\ell}_{\mathcal{D}}\|$, all that remains to bound is $\|\ell_{\mathcal{M}}\|_{L^2(\Omega)}$. Recalling the notation $\langle z \rangle = \frac{1}{|\Omega|_d} \int_{\Omega} z$, and applying the discrete Poincaré–Wirtinger inequality (2.8), it holds

$$\|\ell_{\mathcal{M}} - \langle \ell_{\mathcal{M}} \rangle\|_{L^2(\Omega)} \leq C_{PW} |\underline{\ell}_{\mathcal{D}}|_{1,\mathcal{D}} \leq C_{PW} C_{\sharp}, \quad (3.13)$$

with $C_{PW} > 0$ only depending on Ω , θ and k . Thus, by the triangle inequality, we infer

$$\|\ell_{\mathcal{M}}\|_{L^2(\Omega)} \leq \|\ell_{\mathcal{M}} - \langle \ell_{\mathcal{M}} \rangle\|_{L^2(\Omega)} + \|\langle \ell_{\mathcal{M}} \rangle\|_{L^2(\Omega)} \leq C_{PW} C_{\sharp} + |\Omega|_d^{1/2} |\langle \ell_{\mathcal{M}} \rangle|,$$

and we are only left with estimating $|\langle \ell_{\mathcal{M}} \rangle|$. We proceed in two steps, showing first an upper bound on $\langle \ell_{\mathcal{M}} \rangle$, and then a lower bound. Applying Jensen's inequality, the upper bound can be readily obtained:

$$e^{\langle \ell_{\mathcal{M}} \rangle} \leq \langle e^{\ell_{\mathcal{M}}} \rangle \leq \frac{M_{\sharp}}{|\Omega|_d},$$

which yields $\langle \ell_{\mathcal{M}} \rangle \leq \log \left(\frac{M_{\sharp}}{|\Omega|_d} \right)$. To prove the lower bound, we start from (3.13), and we use local reverse Lebesgue embedding (cf. [23, Lemmas 1.25 and 1.12]). This yields

$$\|\ell_{\mathcal{M}} - \langle \ell_{\mathcal{M}} \rangle\|_{L^{\infty}(\Omega)} \leq C_{RL} h_b^{-d/2} \|\ell_{\mathcal{M}} - \langle \ell_{\mathcal{M}} \rangle\|_{L^2(\Omega)} \leq C_{PW} C_{\sharp} C_{RL} h_b^{-d/2},$$

where $C_{RL} > 0$ only depends on d, θ and k . Then, remarking that

$$e^{\ell_M} = e^{\langle \ell_M \rangle} e^{(\ell_M - \langle \ell_M \rangle)} \leq e^{\langle \ell_M \rangle} e^{C_{PW} C_{\#} C_{RL} h_b^{-d/2}},$$

and integrating over Ω , we get

$$\int_{\Omega} e^{\ell_M} \leq e^{\langle \ell_M \rangle} |\Omega|_d e^{C_{PW} C_{\#} C_{RL} h_b^{-d/2}}.$$

Now, using the lower bound on $\int_{\Omega} e^{\ell_M}$, and taking the logarithm, we finally infer that

$$\log \left(\frac{M_b}{|\Omega|_d e^{C_{PW} C_{\#} C_{RL} h_b^{-d/2}}} \right) \leq \langle \ell_M \rangle.$$

This concludes the proof. \square

We now state the existence result, which holds true for positive ε . The proof adopts the methodology developed in [16] in the (nonlinear) HFV context.

Theorem 1 (Existence of solutions to the nonlinear scheme (2.20)). *Assume that the stabilisation parameter ε in (2.18) is positive. Then, there exists at least one solution $(\underline{\ell}_{\mathcal{D}}^n)_{n \geq 1}$ to the scheme (2.20). The corresponding discrete densities $(\underline{u}_{\mathcal{D}}^n)_{n \geq 1}$, defined by (2.21), are positive.*

Proof. The proof proceeds by induction. Let $n \in \mathbb{N}$, and assume that $u_{\mathcal{M}}^n$ is well defined, following (2.20b) (if $n \geq 1$) or (2.20c) (if $n = 0$). We now prove the existence of a solution $\underline{\ell}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^k$ to (2.20a). For convenience, instead of looking for the discrete logarithm potential, we will equivalently seek for the discrete quasi-Fermi potential $\underline{w}_{\mathcal{D}}^{n+1} = \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}} - \log(c_{nl}^M) \underline{1}_{\mathcal{D}}$ (cf. (3.6)).

First, notice that, given any $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$, and corresponding discrete logarithm potential $\underline{\ell}_{\mathcal{D}}$ (through (3.6)) and discrete density $\underline{u}_{\mathcal{D}}$ (through (2.21)), the map

$$\underline{v}_{\mathcal{D}} \mapsto \int_{\Omega} \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}})$$

is a bounded linear form on $\underline{V}_{\mathcal{D}}^k$. Therefore, by the Riesz–Fréchet representation theorem, there exists a unique element $\underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) \in \underline{V}_{\mathcal{D}}^k$ such that

$$\forall \underline{v}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k, \quad \langle \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}), \underline{v}_{\mathcal{D}} \rangle = \int_{\Omega} \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} v_{\mathcal{M}} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{v}_{\mathcal{D}}).$$

Remark that $\underline{w}_{\mathcal{D}} \mapsto \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}})$ is a continuous (nonlinear) map of $\underline{V}_{\mathcal{D}}^k$. Note also that, for any discrete quasi-Fermi potential $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ such that $\underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) = \underline{0}_{\mathcal{D}}$, by (3.7), the corresponding discrete logarithm potential $\underline{\ell}_{\mathcal{D}}$ solves (2.20a). Our aim from now on is thus to show that $\underline{\mathcal{G}}_{\mathcal{D}}$ does vanish on $\underline{V}_{\mathcal{D}}^k$.

To this purpose, we introduce a regularisation of $\underline{\mathcal{G}}_{\mathcal{D}}$: given any $\mu > 0$, we let

$$\underline{\mathcal{G}}_{\mathcal{D}}^{\mu} : \underline{V}_{\mathcal{D}}^k \rightarrow \underline{V}_{\mathcal{D}}^k; \quad \underline{w}_{\mathcal{D}} \mapsto \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}) + \mu \underline{w}_{\mathcal{D}}.$$

By definition of $\underline{\mathcal{G}}_{\mathcal{D}}$, one has

$$\begin{aligned} \langle \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \rangle &= \langle \underline{\mathcal{G}}_{\mathcal{D}}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \rangle + \mu \|\underline{w}_{\mathcal{D}}\|^2 \\ &= \int_{\Omega} \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}) + \mu \|\underline{w}_{\mathcal{D}}\|^2. \end{aligned}$$

As already shown in the proof of Proposition 4 (cf. (3.11)), by convexity of Φ_1 , one has

$$\int_{\Omega} \frac{u_{\mathcal{M}} - u_{\mathcal{M}}^n}{\Delta t} w_{\mathcal{M}} \geq \frac{\mathbb{E}(w_{\mathcal{D}}) - \mathbb{E}^n}{\Delta t},$$

where the discrete entropies are defined by

$$\mathbb{E}(w_{\mathcal{D}}) = \int_{\Omega} u_{\mathcal{M}}^{\infty} \Phi_1 \left(\frac{u_{\mathcal{M}}}{u_{\mathcal{M}}^{\infty}} \right) \quad \text{and} \quad \mathbb{E}^n = \int_{\Omega} u_{\mathcal{M}}^{\infty} \Phi_1 \left(\frac{u_{\mathcal{M}}^n}{u_{\mathcal{M}}^{\infty}} \right).$$

As already mentioned, since Φ_1 is a non-negative function, these two quantities are non-negative. Note that it may occur that $\mathbb{E}^n = 0$ (which is equivalent to $u_{\mathcal{M}}^n = u_{\mathcal{M}}^{\infty}$ in Ω for $n \geq 1$, or $u^0 = u_{\mathcal{M}}^{\infty}$ in Ω), in which case $\underline{\ell}_{\mathcal{D}} = \underline{\ell}_{\mathcal{D}}^{\infty}$ is the unique solution to (2.20a) (uniqueness follows from the entropy relation (3.10)). In the following, we therefore assume that $\mathbb{E}^n > 0$. The previous identities, and the non-negativity of the dissipation and entropy, imply that

$$\begin{aligned} \langle \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \rangle &\geq \frac{\mathbb{E}(w_{\mathcal{D}}) - \mathbb{E}^n}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}; \underline{w}_{\mathcal{D}}, \underline{w}_{\mathcal{D}}) + \mu \|\underline{w}_{\mathcal{D}}\|^2 \\ &\geq \mu \|\underline{w}_{\mathcal{D}}\|^2 - \frac{\mathbb{E}^n}{\Delta t}. \end{aligned} \tag{3.14}$$

Letting $r = \sqrt{\frac{\mathbb{E}^n}{\mu \Delta t}} > 0$, one has that $\langle \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}), \underline{w}_{\mathcal{D}} \rangle \geq 0$ for all $\underline{w}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ such that $\|\underline{w}_{\mathcal{D}}\| = r$. Therefore, according to [16, Lemma 1] (cf. also [32, Section 9.1]), which is a by-product of Brouwer's fixed-point theorem, there exists $\underline{w}_{\mathcal{D}}^{\mu} \in \underline{V}_{\mathcal{D}}^k$ such that

$$\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}^{\mu}) = \underline{0}_{\mathcal{D}} \quad \text{and} \quad \|\underline{w}_{\mathcal{D}}^{\mu}\| \leq \sqrt{\frac{\mathbb{E}^n}{\mu \Delta t}}. \tag{3.15}$$

Now, plugging $\underline{w}_{\mathcal{D}}^{\mu}$ in (3.14), and using that $\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(\underline{w}_{\mathcal{D}}^{\mu}) = \underline{0}_{\mathcal{D}}$, we get

$$\frac{\mathbb{E}(w_{\mathcal{D}}^{\mu})}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}; \underline{w}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}) + \mu \|\underline{w}_{\mathcal{D}}^{\mu}\|^2 \leq \frac{\mathbb{E}^n}{\Delta t},$$

so that $\mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}; \underline{w}_{\mathcal{D}}^{\mu}, \underline{w}_{\mathcal{D}}^{\mu}) \leq \frac{\mathbb{E}^n}{\Delta t}$. Thus, recalling the definition (2.18)-(2.19) of $\mathcal{T}_{\mathcal{D}}$, as well as the stability estimate (2.7) for $a_{\mathcal{D}}$, we infer that

$$\varepsilon h_b^{k+2} |\underline{w}_{\mathcal{D}}^{\mu}|_{1,\mathcal{D}}^2 \lesssim \frac{\mathbb{E}^n}{\Delta t}.$$

On the one hand, by (3.6) and the estimate (3.8) on $|\underline{\phi}|_{1,\mathcal{D}}$, it holds

$$|\underline{\ell}_{\mathcal{D}}^{\mu}|_{1,\mathcal{D}} = |\underline{w}_{\mathcal{D}}^{\mu} - \underline{\phi}|_{1,\mathcal{D}} \lesssim \sqrt{\frac{\mathbb{E}^n}{\varepsilon h_b^{k+2} \Delta t}} + 1. \tag{3.16}$$

On the other hand, by definition of $\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}$, one first infers that

$$\begin{aligned} 0 &= \langle \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(w_{\mathcal{D}}^{\mu}), \underline{1}_{\mathcal{D}} \rangle = \int_{\Omega} \frac{u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^n}{\Delta t} + \mathcal{T}_{\mathcal{D}}(\underline{\ell}_{\mathcal{D}}^{\mu}; w_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}}) + \mu \langle w_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}} \rangle \\ &= \int_{\Omega} \frac{u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^n}{\Delta t} + \mu \langle w_{\mathcal{D}}^{\mu}, \underline{1}_{\mathcal{D}} \rangle. \end{aligned}$$

Second, using the Cauchy-Schwarz inequality, followed by the bound (3.15) on $\|w_{\mathcal{D}}^{\mu}\|$, one gets

$$\left| \int_{\Omega} (u_{\mathcal{M}}^{\mu} - u_{\mathcal{M}}^n) \right| \leq \mu \Delta t \|w_{\mathcal{D}}^{\mu}\| \|\underline{1}_{\mathcal{D}}\| \leq \sqrt{\mu} \sqrt{\Delta t \mathbb{E}^n |\Omega|_d},$$

where we have also used that $\|\underline{1}_{\mathcal{D}}\| = |\Omega|_d^{1/2}$. Thus, letting $M^n = \int_{\Omega} u_{\mathcal{M}}^n > 0$ (recall that $\int_{\Omega} u_{\mathcal{M}}^0 = M > 0$), and $\mu^n = \frac{(M^n)^2}{4\Delta t \mathbb{E}^n |\Omega|_d} > 0$, for all $0 < \mu \leq \mu^n$ one has

$$\frac{M^n}{2} \leq \int_{\Omega} e^{\ell_{\mathcal{M}}^{\mu}} \leq \frac{3M^n}{2}. \quad (3.17)$$

Leveraging (3.16) and (3.17), one can eventually apply Lemma 1 with $M_b = \frac{M^n}{2}$, $M_{\#} = \frac{3M^n}{2}$, and $C_{\#}$ proportional to $\sqrt{\frac{\mathbb{E}^n}{\varepsilon h_b^{k+2} \Delta t}} + 1$ (note that these three constants do not depend on μ): there exists a constant $C > 0$, independent of μ , such that

$$\forall \mu \in (0, \mu^n], \quad \|\underline{\ell}_{\mathcal{D}}^{\mu}\| \leq C.$$

Then, by compactness, there exists $\underline{\ell}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^k$ such that, up to extraction (not relabelled), $\underline{\ell}_{\mathcal{D}}^{\mu} \rightarrow \underline{\ell}_{\mathcal{D}}^{n+1}$ when $\mu \rightarrow 0$. On the other hand, $\underline{\mathcal{G}}_{\mathcal{D}}^{\mu}$ tends to $\underline{\mathcal{G}}_{\mathcal{D}}$ as μ tends to 0. Therefore, letting

$$w_{\mathcal{D}}^{n+1} = \underline{\ell}_{\mathcal{D}}^{n+1} + \underline{\phi}_{\mathcal{D}} - \log(c_{nl}^M) \underline{1}_{\mathcal{D}},$$

we have $\underline{0}_{\mathcal{D}} = \underline{\mathcal{G}}_{\mathcal{D}}^{\mu}(w_{\mathcal{D}}^{\mu}) \rightarrow \underline{\mathcal{G}}_{\mathcal{D}}(w_{\mathcal{D}}^{n+1})$ as $\mu \rightarrow 0$, which implies that

$$\underline{\mathcal{G}}_{\mathcal{D}}(w_{\mathcal{D}}^{n+1}) = \underline{0}_{\mathcal{D}}.$$

It follows that $\underline{\ell}_{\mathcal{D}}^{n+1}$ is a solution to (2.20a). \square

Remark 9 (Uniqueness of the solution). *As for the low-order nonlinear VAG, DDFV and HFV schemes of [12, 13, 16], the uniqueness of the solution to (2.20) is still an open question. A possible approach to show such a result could be to consider the relative discrete entropy of a solution with respect to another solution, and show that this quantity vanishes.*

Last, we study the long-time behaviour of the nonlinear HHO scheme.

Proposition 5 (Long-time behaviour of the nonlinear scheme). *Assume that the stabilisation parameter ε in (2.18) is positive, and let $(\underline{\ell}_{\mathcal{D}}^n)_{n \geq 1}$ be a solution to the scheme (2.20). Then, the discrete solution converges in time towards the discrete equilibrium logarithm potential:*

$$\underline{\ell}_{\mathcal{D}}^n \xrightarrow{n \rightarrow \infty} \underline{\ell}_{\mathcal{D}}^{\infty} \text{ in } \underline{V}_{\mathcal{D}}^k. \quad (3.18)$$

Consequently, the corresponding reconstructed discrete density $(u_{\mathcal{M}}^n)_{n \geq 1}$ converges to $u_{\mathcal{M}}^{\infty}$ in $L^{\infty}(\Omega)$.

Proof. First, remark that owing to the entropy relation (3.10), one has

$$\sum_{n \geq 1} \mathbb{D}^n \leq \sum_{n \geq 1} \frac{\mathbb{E}^{n-1} - \mathbb{E}^n}{\Delta t} \leq \frac{\mathbb{E}^0}{\Delta t}.$$

Thus, according to the definition of the discrete dissipation \mathbb{D}^n , alongside with the definition (2.18)-(2.19) of $\mathcal{T}_{\mathcal{D}}$, and the stability estimate (2.7) for $a_{\mathcal{D}}$, we infer that

$$\sum_{n \geq 1} |w_{\mathcal{D}}^n|_{1,\mathcal{D}}^2 \lesssim \frac{\mathbb{E}^0}{\varepsilon h_b^{k+2} \Delta t}.$$

This implies, in particular, that

$$\forall n \geq 1, \quad |w_{\mathcal{D}}^n|_{1,\mathcal{D}} \lesssim \sqrt{\frac{\mathbb{E}^0}{\varepsilon h_b^{k+2} \Delta t}} \quad \text{and} \quad |w_{\mathcal{D}}^n|_{1,\mathcal{D}} \xrightarrow{n \rightarrow \infty} 0. \quad (3.19)$$

Let $n \geq 1$. By (3.6) and (3.8), one has

$$|\underline{\ell}_{\mathcal{D}}^n|_{1,\mathcal{D}} \lesssim \sqrt{\frac{\mathbb{E}^0}{\varepsilon h_b^{k+2} \Delta t}} + 1.$$

On the other hand, by the mass preservation (3.9), we have $\int_{\Omega} e^{\ell_{\mathcal{M}}} = M > 0$. Therefore, one can apply Lemma 1, and infer the existence of a positive constant C (which is independent of n) such that

$$\forall n \geq 1, \quad \|\underline{\ell}_{\mathcal{D}}^n\| \leq C. \quad (3.20)$$

It follows, by compactness, that there exists $\underline{\ell}_{\mathcal{D}} \in \underline{V}_{\mathcal{D}}^k$ such that, up to extraction (not relabelled),

$$\lim_{n \rightarrow \infty} \underline{\ell}_{\mathcal{D}}^n = \underline{\ell}_{\mathcal{D}} \quad \text{in } \underline{V}_{\mathcal{D}}^k.$$

By (3.6), (3.19), and continuity of $|\cdot|_{1,\mathcal{D}}$ on $\underline{V}_{\mathcal{D}}^k$, we infer that

$$|\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}}|_{1,\mathcal{D}} = 0.$$

This means that there exists $a \in \mathbb{R}$ such that $\underline{\ell}_{\mathcal{D}} + \underline{\phi}_{\mathcal{D}} = a \underline{1}_{\mathcal{D}}$. By mass preservation, we get

$$M = \int_{\Omega} e^{\ell_{\mathcal{M}}} \xrightarrow{n \rightarrow \infty} \int_{\Omega} e^{\ell_{\mathcal{M}}},$$

so that $a = \log(c_{nl}^M)$, which implies that

$$\underline{\ell}_{\mathcal{D}} = \log(c_{nl}^M) \underline{1}_{\mathcal{D}} - \underline{\phi}_{\mathcal{D}} = \underline{\ell}_{\mathcal{D}}^{\infty}.$$

By uniqueness of the limit, we finally infer the convergence of the whole sequence $(\underline{\ell}_{\mathcal{D}}^n)_{n \geq 1}$ towards $\underline{\ell}_{\mathcal{D}}^{\infty}$ in $\underline{V}_{\mathcal{D}}^k$. This implies, in particular, that $\ell_{\mathcal{M}}^n \xrightarrow{n \rightarrow \infty} \ell_{\mathcal{M}}^{\infty}$ in $L^{\infty}(\Omega)$, by norm equivalence in finite-dimensional vector spaces. Then, by the mean value theorem, we deduce that

$$\|u_{\mathcal{M}}^n - u_{\mathcal{M}}^{\infty}\|_{L^{\infty}(\Omega)} = \|e^{\ell_{\mathcal{M}}^n} - e^{\ell_{\mathcal{M}}^{\infty}}\|_{L^{\infty}(\Omega)} \leq e^{\max(\|\ell_{\mathcal{M}}^n\|_{L^{\infty}(\Omega)}, \|\ell_{\mathcal{M}}^{\infty}\|_{L^{\infty}(\Omega)})} \|\ell_{\mathcal{M}}^n - \ell_{\mathcal{M}}^{\infty}\|_{L^{\infty}(\Omega)},$$

which implies, by uniform boundedness (in n) of $(\ell_{\mathcal{M}}^n)_{n \geq 1}$, the convergence of the reconstructed discrete density in $L^{\infty}(\Omega)$. \square

Remark 10 (Non-uniformity of the bounds). *Note that the estimate (3.20) on the solution to (2.20) is not uniform with respect to the discretisation parameters $h_{\mathcal{D}}$ and Δt , nor with respect to the stabilisation parameter ε . Indeed, having a closer look to the dependencies of the corresponding upper bound (using Lemma 1), one realises that it blows up as soon as either $h_{\mathcal{D}}$, Δt or ε tends to zero.*

Remark 11 (Convergence to equilibrium). *Notice that the time-asymptotic result of Proposition 5 is relatively weaker than the one of Proposition 3 in the exponential fitting context. For the latter result, the convergence to equilibrium is shown to be exponentially fast, and the decay rate ν_{ω} uniform with respect to the discretisation parameters. The numerical results of [44] and Section 4.4 indicate that, also for the nonlinear scheme, the convergence is expected to be exponential, with seemingly uniform (and close to the PDE model one) decay rate. At the theoretical level, to prove exponential convergence to equilibrium, one has to establish a control of the discrete entropy by the discrete dissipation. Adapting the arguments from [45, Theorem 3], such a control can actually be established in the present context, but leads to a non-uniform decay rate (also depending on ε), and to a final result still only valid in the case $\varepsilon > 0$. In order to showcase a uniform (and ε -independent) decay rate, and establish a result also valid in the case $\varepsilon = 0$, a (high-order) uniform discrete Logarithmic-Sobolev inequality needs to be available (cf. [16] in the low-order HFV context). This is the subject of ongoing research.*

4. Numerical results

In this section, we extensively assess the high-order nonlinear scheme (2.20). We study positivity preservation, convergence, efficiency (accuracy vs. computational cost), and long-time behaviour. We also compare it, in terms of structure preservation, with the linear high-order exponential fitting scheme (2.12). All the test-cases considered below are set in the two-dimensional domain $\Omega = (0, 1)^2$, and are (except for the last one) taken from [16], to which we refer for more detailed descriptions. Given a (face) degree $k \geq 0$, the nonlinear scheme (2.20) will be referred to as `nlhho_k`, whereas the exponential fitting one (2.12) as `expf_k`. For the nonlinear scheme, we will always use below the value $\varepsilon = 1$ for the parameter ε in (2.18). However, in some situations, we will compare the two values $\varepsilon = 1$ and $\varepsilon = 0$. The nonlinear scheme with $\varepsilon = 0$ will then be denoted `nlhho_k_0`.

4.1. Implementation

All numerical tests presented below have been run on a laptop equipped with an Intel Core i7-9850H processor clocked at 2.60GHz and 32Gb of RAM. Our HHO implementation makes use of monomial basis functions for both the cell and face unknowns. Such a choice is known to introduce numerical instabilities for large values of k , we thus restrict our study to $k \leq 3$. The use of orthonormal basis functions, which is expected to improve on this situation (in particular for the convergence of the Newton algorithm in the nonlinear case), shall be studied in future works. We use quadrature formulas based on the Dunavant rules [31] (after subtessellation). To cope with non-polynomial integrands, we employ quadrature formulas of order $2k + 5$. We performed a few tests (not reported here) with higher-order formulas, and did not observe any significant changes. Last, the local computations are performed sequentially. One could expect a significant gain in terms of performances parallelising the latter. We discuss below some important implementation aspects for both schemes.

4.1.1. Exponential fitting scheme

For the linear exponential fitting scheme, the implementation follows the classical HHO strategy for linear diffusion problems. We directly solve for the discrete Slotboom variable $(\underline{\rho}_{\mathcal{D}}^n)_{n \geq 1}$. As standard for skeletal methods, we do not solve the full linear system, but first perform static condensation, which allows one to locally eliminate the cell unknowns. Since the scheme relies on the same LHS matrix at each time step, we perform once and for all an LU decomposition of the matrix at the beginning of the computation. At each time step, the solution is then inexpensive (the RHS has to be updated, but only through a matrix-vector product).

We do not address in this work the main questions which were highlighted in [16, Section 5.1.2] in the low-order HFV context, about the (harmonic) averaging of ω (which is related to the choice of quadrature formulas for the high-order scheme), and the preconditioning of the system (which was equivalent, in the simple HFV context, to choose to solve the system in the density variable). These aspects shall be investigated in future works. Nonetheless, in view of the results obtained in [16] for the HFV exponential fitting scheme, we believe these potential improvements will have no effect on the positivity violation issues.

4.1.2. Nonlinear scheme

The numerical scheme (2.20) requires to solve a nonlinear system of equations at each time step. For $n \in \mathbb{N}$, one wants to find $\underline{\ell}_{\mathcal{D}}^{n+1} \in \underline{V}_{\mathcal{D}}^k$ solution to (2.20): this scheme can be written as the equation

$$\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t}(\underline{\ell}_{\mathcal{D}}^{n+1}) = \underline{0}_{\mathcal{D}},$$

with $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t} : \underline{V}_{\mathcal{D}}^k \rightarrow \underline{V}_{\mathcal{D}}^k$ smooth (nonlinear) vector field. Numerically, to find a zero of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\Delta t}$, we use a Newton method.

In practice, the use of a naive method without any adaptation proves not to be enough to compute a solution in general. In order to get a robust implementation, which can handle various data and meshes, one has to deploy a few techniques. For further use, we let $\|\underline{\ell}_{\mathcal{D}}\|_{l^\infty}$ denote the l^∞ -norm of the coefficients of $\underline{\ell}_{\mathcal{D}}$ in the (cell and face) polynomial bases. The map $\|\cdot\|_{l^\infty}$ is a norm on $\underline{V}_{\mathcal{D}}^k$, which is easily (and at very low cost) computable in practice. To fix the ideas and the notation, the Newton method is defined as follows: given an initialisation $\underline{\ell}_{\mathcal{D},0} \in \underline{V}_{\mathcal{D}}^k$, and a time step $0 < \delta t \leq \Delta t$, one defines a sequence $(\underline{\ell}_{\mathcal{D},i})_{i \geq 0}$ of elements of $\underline{V}_{\mathcal{D}}^k$ such that

$$J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i+1} - \underline{\ell}_{\mathcal{D},i}) = -\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i}), \quad (4.1)$$

where $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}$ is the vector field associated to the nonlinear scheme (2.20) with time step δt instead of Δt , and $J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}$ is the differential (Jacobian in practice) of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}$ at $\underline{\ell}_{\mathcal{D},i}$. Note that, in practice, we do not solve this linear system, but perform static condensation in order to (locally) eliminate the cell unknowns. The resulting linear system is called ‘‘condensed system’’ in what follows. We discuss below the main tricks deployed to reach robustness in the implementation of the Newton algorithm.

Stopping criterion. We define the relative norm of the residual r_{i+1} , and the norm of the objective function g_i , as

$$r_{i+1} = \frac{\|\underline{\ell}_{\mathcal{D},i+1} - \underline{\ell}_{\mathcal{D},i}\|_{l^\infty}}{\|\underline{\ell}_{\mathcal{D},i}\|_{l^\infty}} \quad \text{and} \quad g_i = \|\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i})\|_{l^\infty}.$$

We consider that the Newton method has converged when either

$$(r_{i+1} \leq 0.1 \times \text{tol}) \quad \text{or} \quad (r_{i+1} \leq \text{tol} \text{ and } g_i \leq \text{tol}),$$

with $\text{tol} = 5 \cdot 10^{-10}$, in which case we set $\underline{\ell}_{\mathcal{D}}^{n+1} = \underline{\ell}_{\mathcal{D},i+1}$. On the other hand, if this criterion is not met at $i = 50$, the method is considered as non-convergent (and we then proceed with a time step reduction, see below). In practice, for the tests collected in this article, we never reached $i = 50$, either because the method converged, or because of a loop break (see below).

Loop break for unreasonably large $\underline{\ell}_{\mathcal{D}}$. The computations of $\underline{\mathcal{G}}_{\mathcal{D}}^{n,\delta t}(\underline{\ell}_{\mathcal{D},i})$ and $J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}$ imply punctual evaluations of $e^{\ell_{K,i}}$ (for $K \in \mathcal{M}$) and $e^{\ell_{\sigma,i}}$ (for $\sigma \in \mathcal{E}$) in the quadrature formulas. Such computations can lead to severe numerical issues if the values at the quadrature nodes are too large. Therefore, we declare that $\underline{\ell}_{\mathcal{D},i}$ is unreasonably large for the computations if there exists a cell quadrature node $x_{K,q} \in \overline{K}$, or a face quadrature node $x_{\sigma,q} \in \overline{\sigma}$, such that

$$|\ell_{K,i}(x_{K,q})| \geq 100 \quad \text{or} \quad |\ell_{\sigma,i}(x_{\sigma,q})| \geq 100.$$

In such a case, the method is immediately considered as non-convergent, and we proceed with a time step reduction (see below). Note that the choice of the value 100 allows one to compute densities $\underline{u}_{\mathcal{D}}$ over a range from 10^{-43} to 10^{43} , and hence should not be a significant restriction in practice. In the numerical simulations presented below, the use of this loop-breaking procedure is absolutely necessary in order to avoid the evaluation of too large quantities, leading to some ‘‘explosion’’ of the method and crash of the code. Moreover, we also operate a loop break if the linear solver does not perform a successful resolution of the condensed linear system associated to (4.1), which corresponds to situations for which either $J_{\underline{\ell}_{\mathcal{D},i}}^{n,\delta t}$ or its condensed counterpart are not invertible. Such situations occur in practice, essentially on very coarse meshes.

Adaptive time stepping. The previous strategies can lead to a solution failure for some given time step δt . If the Newton method did not converge, we try to compute the solution for a smaller time step $\delta t/2$. On the other hand, if the method did converge, we use for the subsequent time step the larger value $2\delta t$. The maximal time step allowed is the initial one, denoted by Δt . In practice, the scheme may perform numerous time step reductions at the beginning (early times) of the computation.

Initialisation by truncation and filtration. As for any Newton method, the question of the initialisation is fundamental in order to get a robust implementation. It appears that, for $n \geq 1$, the natural initialisation $\underline{\ell}_{\mathcal{D},0} = \underline{\ell}_{\mathcal{D}}^n$ is satisfactory when used with the adaptative time stepping strategy. However, for $n = 0$, such a choice is not possible, since $\underline{\ell}_{\mathcal{D}}^0$ does not exist in general if u^0 vanishes locally or is too small (cf. Remark 6). A first way of tackling this problem is to define a truncated initial logarithm potential $\tilde{\ell}^0$ as

$$\tilde{\ell}^0 = \log(\max(u^0, 10^{-8})),$$

and to initialise the Newton method with $\underline{\tilde{\ell}}_{\mathcal{D}}^0 = \underline{I}_{\mathcal{D}}^k(\tilde{\ell}^0) \in \underline{V}_{\mathcal{D}}^k$, provided one can give a sense to the face components. In fact, such a strategy entails another limitation: $\underline{\tilde{\ell}}_{\mathcal{D}}^0$ exhibits strong oscillations in the regions where the truncation is performed (this is also true when u^0 is discontinuous over Ω ,

as in Section 4.2). These oscillations usually make the method diverge, even with extremely small time step. Therefore, we eventually initialise the method with a “filtered” (non-oscillating) discrete logarithm potential $\underline{\ell}_{\mathcal{D}}^0 \in \underline{V}_{\mathcal{D}}^k$, which corresponds to a zero-order polynomial projection of $\tilde{\ell}^0$:

$$\ell_K^0 = \Pi_K^0(\tilde{\ell}_{|K}^0) \quad \forall K \in \mathcal{M} \quad \text{and} \quad \ell_{\sigma}^0 = \Pi_{\sigma}^0(\tilde{\ell}_{|\sigma}^0) \quad \forall \sigma \in \mathcal{E},$$

still provided one can give a sense to the face components. In practice, using $\underline{\ell}_{\mathcal{D},0} = \underline{\ell}_{\mathcal{D}}^0$ as the first initialisation (when $n = 0$) yields convergent Newton methods for all tests presented below. The use of filtered initial discrete data seems particularly crucial for high-order schemes ($k \geq 1$). For the lowest-order version of the scheme ($k = 0$), the use of $\tilde{\ell}_{\mathcal{D}}^0$ as a first initialisation (for $n = 0$) often yields convergent Newton methods (up to time step reduction).

Of course, the chosen values for the stopping criterion and the thresholds are arbitrary and could be modified. However, the set of values advocated here makes the scheme robust enough so as to be capable of computing solutions for all the test-cases in this article.

Remark 12 (Potentials vs. densities). *One of the main differences between the present nonlinear scheme and the low-order HFV ones from [16, 45] lies in the fact that we use here the potential ℓ as our (piecewise polynomial) unknown, whereas the density u was used in the low-order schemes. Notice that, in the present context, choosing u as the main variable would require to give a discrete meaning to $\nabla \log(u)$, which is not obvious for the following reason: polynomials of degree ≥ 1 are not stable by the log function. As a by-product of seeking for a potential, our stopping criterion only provides information on ℓ , while we are eventually interested in the corresponding density. Moreover, our criterion only takes into account the coefficients of the polynomials (through the use of the norm $\|\cdot\|_{l^{\infty}}$), but such a measure does not give much information about the effective behaviour of the unknowns. A more relevant stopping criterion could be to consider the residual in terms of densities*

$$\frac{\|u_{\mathcal{M},i+1} - u_{\mathcal{M},i}\|_{L^2(\Omega)}}{\|u_{\mathcal{M},i}\|_{L^2(\Omega)}}$$

(and analogous definition for the face unknowns) in order to ensure a satisfying accuracy on u . However, the main drawback of such a criterion is its evaluation cost. In this work, we thus chose to use instead a purely algebraic stopping criterion on ℓ , whose cost is marginal. The testing of other stopping criteria will be the subject of future investigations. Last, for the HFV schemes of [16, 45], the following loop-breaking strategy was used: when the computed density had almost-zero (or even non-positive) components, one performed a time step reduction. Here, such a situation cannot occur, since ℓ is authorised to take any real value, but this apparent latitude on the potential is in fact pernicious. Indeed, situations in which ℓ takes negative values with large magnitude are actually the counterpart of an almost-zero u for the HFV schemes. Like their counterpart, they lead to divergent Newton methods. The main difficulty then lies in the design of a relevant criterion in order to avoid these situations.

4.2. Positivity

This first section is dedicated to assessing discrete positivity preservation. For the test considered here, we set the advective potential and the anisotropy tensor to

$$\phi(x_1, x_2) = -\left((x_1 - 0.4)^2 + (x_2 - 0.6)^2\right) \quad \text{and} \quad \Lambda = \begin{pmatrix} 0.8 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the initial datum, we take $u^0 = 10^{-3} \mathbb{1}_B + \mathbb{1}_{\Omega \setminus B}$, where B is the Euclidean ball

$$B = \{(x_1, x_2) \in \mathbb{R}^2 \mid (x_1 - 0.5)^2 + (x_2 - 0.5)^2 \leq 0.2^2\}.$$

These data ensure that the solution u is positive on $\mathbb{R}_+ \times \Omega$. We perform the simulation on the time interval $[0, 5 \cdot 10^{-4}]$ with $\Delta t = 10^{-5}$, on a (fine) tilted hexagonal-dominant mesh featuring 4192 cells and 12512 edges. The computed discrete densities are denoted by $(\underline{u}_{\mathcal{D}}^n)_{1 \leq n \leq N_f}$ and $(\underline{u}_{\mathcal{D}}^{\omega, n})_{1 \leq n \leq 50}$. Remark that the situation $N_f > 50$ may occur if the nonlinear scheme has to perform time step adaptation.

In Table 1, we collect the minimal values reached by the discrete solutions. The values of `mincellA` (for “average”) are defined by

$$\min \left\{ \frac{1}{|K|_d} \int_K u_K^n \mid K \in \mathcal{M}, 1 \leq n \leq N_f \right\} \quad \text{and} \quad \min \left\{ \frac{1}{|K|_d} \int_K u_K^{\omega, n} \mid K \in \mathcal{M}, 1 \leq n \leq 50 \right\},$$

for, respectively, the nonlinear scheme and the exponential fitting scheme. The values of `mincellQN` are the minimal values taken by the densities at the cell quadrature nodes. Analogous definitions hold for the faces. The values of `#resol` correspond to the number of linear systems solved during the computation. Note that the size of these systems depends on the value of k , so it is not a relevant information to compare the cost of the schemes for different values of k . Last, `walltime` is the total time (in s) needed to compute the discrete solution (it includes the pre-computation steps, such as the computation of the matrices representing G_K).

Table 1. Positivity of discrete solutions.

scheme	walltime	#resol	mincellA	minfaceA	mincellQN	minfaceQN
<code>nlhho_0</code>	7.17e+01	224	1.00e-03	1.01e-03	2.41e-06	1.01e-03
<code>nlhho_1</code>	4.13e+02	248	6.65e-04	2.05e-05	1.78e-04	3.57e-08
<code>nlhho_2</code>	1.45e+03	251	9.50e-04	5.99e-04	2.67e-07	1.06e-05
<code>nlhho_3</code>	3.87e+03	254	9.85e-04	8.58e-04	1.10e-05	1.79e-05
<code>expf_0</code>	5.66e-01	50	1.02e-03	1.89e-03	-3.78e-01	1.89e-03
<code>expf_1</code>	2.23e+00	50	-1.29e-02	-2.40e-01	-4.91e-01	-3.71e-01
<code>expf_2</code>	6.34e+00	50	-6.14e-03	-1.02e-01	-5.08e-01	-5.35e-01
<code>expf_3</code>	1.53e+01	50	-3.24e-04	-1.02e-02	-5.52e-01	-4.05e-01

Recall that the exponential fitting scheme is linear (with corresponding matrix not depending on time), whence its extremely low cost compared to the nonlinear scheme. Note, however, that when an LU decomposition is unaffordable and an iterative solver has to be used instead, `nlhho_k` is approximately “only” five times more costly than `expf_k`. The results of Table 1 first indicate that, as expected, all nonlinear schemes preserve the positivity of the discrete solution. On the other hand, none of the linear schemes preserves positivity on the whole domain Ω . In fact, except `expf_0`, all linear schemes studied here do not even preserve the average positivity on each cell, in the sense that there exists $K_0 \in \mathcal{M}$ and n_0 integer such that

$$\int_{K_0} u_{K_0}^{\omega, n_0} < 0.$$

Moreover, it is interesting to note that the positivity violation peak (which can be approximated by `|mincellQN|` and `|minfaceQN|`) increases as k increases, whereas in average (values of `|mincellA|` and `|minfaceA|`) the lack of positivity becomes smaller as the order increases.

At this stage, it is worth pointing out the fact that quantifying the negativity of the solution is much more difficult for high-order schemes, since it is not possible to “count” the number of negative values (which correspond to the degrees of freedom for low-order schemes). While the `mincellQN` value gives information about the minimum value reached on the whole domain, it does not give any indication about the measure of the set $\{x \in \bar{\Omega} \mid u_M^{\omega,n}(x) < 0\}$ where the discrete cell unknown takes negative values. The same remark applies to `mincellA`. As an attempt to provide an idea of the size of this set, we display in Table 2 the number of cells with negative average over the whole simulation, defined as the cardinal of the set

$$\left\{ (K, n) \in \mathcal{M} \times \llbracket 1; 50 \rrbracket \mid \int_K u_K^{\omega,n} < 0 \right\}.$$

These data reveal that, excluding `expf_0` which performs quite well on this particular test, the higher the order, the smaller the size of the negative-average set.

Table 2. Number of negative cell averages.

scheme	#cells with negative average
<code>expf_0</code>	0
<code>expf_1</code>	824
<code>expf_2</code>	136
<code>expf_3</code>	1

The previous observations seem to indicate a competition between two phenomena for linear methods. As k increases, the accuracy is improved, and therefore the discrete solution becomes closer to the exact one. Hence, in average, high-order schemes compute solutions with smaller area of negativity, and lesser positivity violation. However, high values of k induce larger oscillations for the polynomial solution: the computed solution takes negative values on smaller sets, but the (pointwise) undershoots become bigger as k increases. At the end, it seems that there is no hope to get a positive discrete solution on the whole domain Ω with a linear method.

Remark 13 (An accuracy criterion taking into account positivity). *The previous observations suggest that, for applications in which preserving the positivity of the solution is an essential feature, the accuracy of the scheme should not simply be defined as an L^p -distance between the reconstructed discrete solution u_M and the exact one u . We believe that a relevant criterion in order to take into account both “classical accuracy” (distance between u_M and u) and positivity is to look at the relative Boltzmann entropy (or other kinds of relative Φ -entropies as defined in [6]) with respect to the exact solution, that is*

$$\text{Err}(u_M) = \int_{\Omega} u \Phi_1\left(\frac{u_M}{u}\right), \quad (4.2)$$

where $\Phi_1(s) = s \log(s) - s + 1$ for $s > 0$, $\Phi_1(0) = 1$, and $\Phi_1(s)$ takes large values for $s < 0$. The interest of such a definition is twofold. First, the negativity of u_M is penalised. Second, if u_M is positive and $\int_{\Omega} u_M = \int_{\Omega} u$ (which is the case in practice for problems with homogeneous Neumann boundary conditions), by Csiszár–Kullback inequality (see e.g., [11, Lemma 5.6]), one has

$$\|u_M - u\|_{L^1(\Omega)} \leq \sqrt{2\|u\|_{L^1(\Omega)} \text{Err}(u_M)}.$$

4.3. Convergence and efficiency

We here study the convergence as $(h_{\mathcal{D}}, \Delta t) \rightarrow (0, 0)$ of the nonlinear scheme for different values of the polynomial degree k . We consider a test-case with known exact solution. We set the advective potential and anisotropy tensor to

$$\phi(x_1, x_2) = -x_1 \quad \text{and} \quad \Lambda = \begin{pmatrix} l_{x_1} & 0 \\ 0 & 1 \end{pmatrix}$$

for $l_{x_1} > 0$. The exact solution is then given by

$$u(t, x_1, x_2) = C_1 e^{-\alpha t + \frac{x_1}{2}} (2\pi \cos(\pi x_1) + \sin(\pi x_1)) + 2C_1 \pi e^{x_1 - \frac{1}{2}}, \quad (4.3)$$

where $C_1 > 0$ and $\alpha = l_{x_1} \left(\frac{1}{4} + \pi^2\right)$. Note that u^0 vanishes on $\{x_1 = 1\}$, but for any $t > 0$, $u(t, \cdot) > 0$. Here, our experiments are performed using $l_{x_1} = 1$ and $C_1 = 10^{-1}$.

We compute the discrete solutions on the time interval $[0, 0.1]$, and we denote by $(\underline{u}_{\mathcal{D}}^n)_{1 \leq n \leq N_f}$ the corresponding discrete densities. We monitor the relative $L_t^2(L_x^2)$ -error and $L_t^2(H_x^1)$ -error on the solution, respectively defined by

$$\sqrt{\frac{\sum_{n=1}^{N_f} \delta t^n \|\underline{u}_{\mathcal{M}}^n - u(t^n, \cdot)\|_{L^2(\Omega)}^2}{\sum_{n=1}^{N_f} \delta t^n \|u(t^n, \cdot)\|_{L^2(\Omega)}^2}}$$

and

$$\sqrt{\frac{\sum_{n=1}^{N_f} \delta t^n \|\mathfrak{G}_{\mathcal{M}}(\underline{u}_{\mathcal{D}}^n) - \nabla u(t^n, \cdot)\|_{L^2(\Omega)^d}^2}{\sum_{n=1}^{N_f} \delta t^n \|\nabla u(t^n, \cdot)\|_{L^2(\Omega)^d}^2}},$$

where $\delta t^n = t^n - t^{n-1}$, and $\sum_{1 \leq n \leq N_f} \delta t^n = 0.1$. The discrete gradient of the densities $\mathfrak{G}_{\mathcal{M}}(\underline{u}_{\mathcal{D}}^n)$ is defined as follows. For all $K \in \mathcal{M}$, $(\mathfrak{G}_{\mathcal{M}}(\underline{u}_{\mathcal{D}}^n))|_K = \mathfrak{G}_K(\underline{u}_K^n)$, where $\mathfrak{G}_K(\underline{u}_K^n)$ is a smooth vector field on K defined by mimicking at the discrete level the relation $\nabla u = e^\ell \nabla \ell$:

$$\mathfrak{G}_K(\underline{u}_K^n) = e^{\ell_K} G_K(\underline{\ell}_K^n) \quad \text{in } K. \quad (4.4)$$

Notice that, with the chosen error measures, we do not take into account the time $t = 0$. We perform our simulations on a triangular mesh family $(\mathcal{D}_i)_{1 \leq i \leq 5}$, such that $h_{\mathcal{D}_i}/h_{\mathcal{D}_{i+1}} = 2$. Since the time discretisation is of order one, in $L_t^\infty(L_x^2)$ -norm, we expect the error to decrease as

$$\text{Error} \leq C_T \Delta t + C_S(k) h_{\mathcal{D}_i}^{k+2},$$

where $C_T, C_S(k) > 0$ are multiplicative constants respectively related to time and space discretisations, with $k \mapsto C_S(k)$ decreasing. We have $h_{\mathcal{D}_i} = h_{\mathcal{D}_1}/2^{i-1}$, so to balance the time and space contributions of the error upper bound, we need to take

$$\Delta t(i, k) \sim \frac{\Delta t(1, k)}{2^{(i-1)(k+2)}},$$

where $\Delta t(1, k) = \frac{C_S(k)}{C_T} h_{\mathcal{D}_1}^{k+2}$. For the values of k we consider, we assume (and we verify in practice that it is relevant) that $\Delta t(1, k) \geq 0.05/2^{k+2}$ (see [1] for a theoretical study of $k \mapsto C_S(k)$ in the HHO context). Thus, for given i and k , we define our (maximal) time step as

$$\Delta t(i, k) = \frac{0.05}{2^{i(k+2)}}.$$

On Figure 1, we plot the relative errors as functions of the mesh size $h_{\mathcal{D}}$ for $k \in \{0, 1, 2, 3\}$. For completeness, we also include the scheme `nlhho_2_0` (i.e., with $\varepsilon = 0$ in (2.18) for $k = 2$) in our comparison. First, we observe that `nlhho_2` and `nlhho_2_0` have the same behaviour (the two plots are superimposed). Tests with other values of k , not shown here, indicate that the influence of ε (0 or 1) on the accuracy of the scheme is not noticeable. Second, we see that, as one could expect, the method `nlhho_k` converges at order $k + 2$ in $L^2_t(L^2_x)$ -norm. In the $L^2_t(H^1_x)$ -norm, if the expected convergence order of $k + 1$ is attained for $k = 0$ and $k = 1$, then some sort of saturation appears for $k = 2$ and $k = 3$. Since this saturation does not show up in $L^2_t(L^2_x)$ -norm, we suspect this might be due to our definition (4.4) of the discrete density gradient. Indeed, remark that, at the discrete level, the chain rule is violated, thus (4.4) is not exactly a discrete version of ∇e^ℓ .

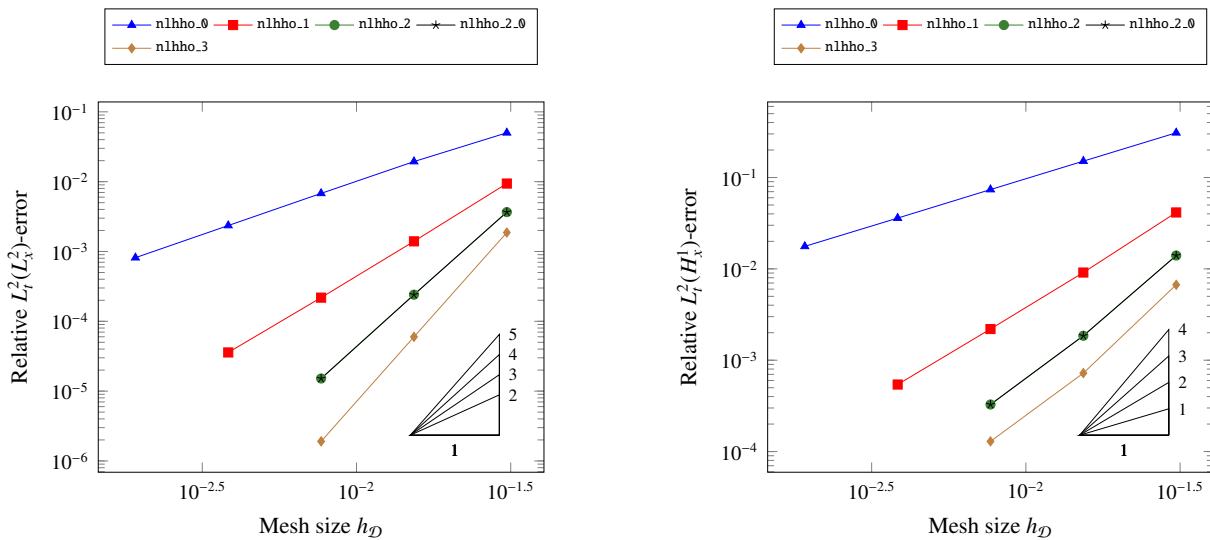


Figure 1. Accuracy vs. mesh size. Relative errors on triangular meshes.

We now study efficiency, that is to say accuracy for a given computational cost. On Figure 2, we plot the relative errors as functions of the simulation walltime (in s). Here again, the results for the schemes `nlhho_2` and `nlhho_2_0` are superimposed. It is quite remarkable to observe that, even with a low-order discretisation in time, significant efficiency gains can be obtained using a larger value of k , at least for values of $k \leq 2$. The gain is expected to be even larger after parallelising the local computations. Of course, the use of higher-order time-stepping methods should also lead to significant gains of efficiency. This will be investigated in future works.

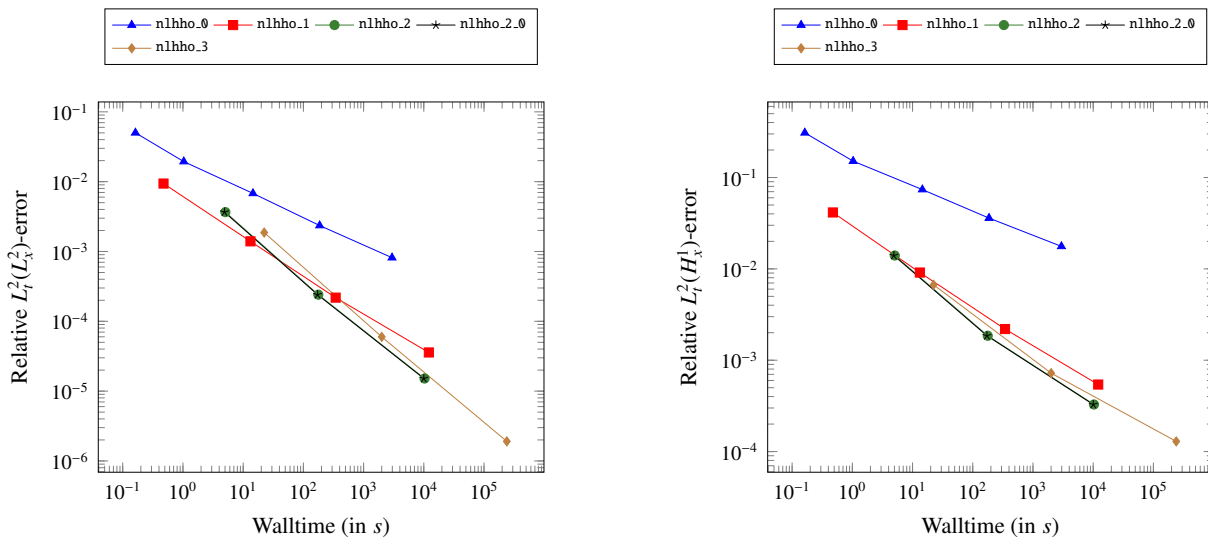


Figure 2. Accuracy vs. computational cost. Relative errors on triangular meshes.

Remark 14 (High-order schemes in time and space). *The extension of the nonlinear scheme (2.20) to arbitrary orders in time and space is a rather natural goal in order to achieve optimal efficiency. However, even with a time discretisation of order 2 (like for example Crank–Nicolson, which is perhaps the most natural extension to backward Euler), there is currently no successful approach retaining the discrete entropy structure. Since this structure is the cornerstone of the analysis (including the existence of solutions), it is of utmost importance to preserve it. Some numerical investigations on nonlinear entropic TPFA schemes for diffusive problems with BDF2 time discretisation have been performed in [21, Chapter 3], and indicate that such a time discretisation could also be a good candidate, even in regard of long-time behaviour (see [21, Chapter 3.4.4]). An alternative approach is to consider space-time methods, as in [8] in the context of conforming Galerkin discretisations of cross-diffusion systems. The extension of space-time techniques to polytopal grids is currently an active research area.*

For completeness, we finally perform simulations on distorted quadrangular meshes, and display the relative $L_t^2(L_x^2)$ -errors on Figure 3. Note that we use the same time step definition as for the previous simulations, whereas the initial mesh is coarser. As expected, the behaviour of the schemes is not strongly impacted by the mesh geometry, and `nlhho_k` converges at order $k + 2$ in $L_t^2(L_x^2)$ -norm. When it comes to efficiency, increasing the value of k leads to better accuracy for fixed computational cost, but the efficiency gain saturates for $k \geq 2$. It is also worth noting that on the coarsest mesh, `nlhho_3` has to perform more time step reductions than the other schemes, because at some iterations the linear solver is unable to perform LU decomposition. These time step reductions do not only occur at the beginning of the simulation, and are probably related to the bad conditioning of the system for high-order polynomials (we use here monomial bases). Based on these observations, using `nlhho_2` seems to be a sound choice to optimise efficiency while ensuring a good numerical stability.

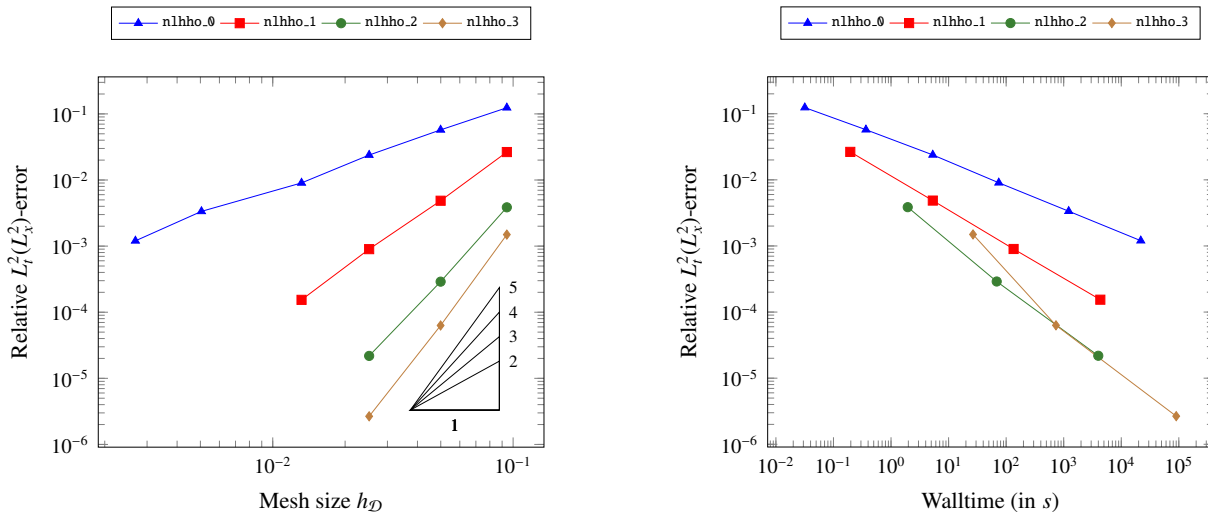


Figure 3. Accuracy on general meshes. Relative $L_t^2(L_x^2)$ -error on distorted quadrangular meshes.

4.4. Discrete long-time behaviour

We are now interested in the long-time behaviour of discrete solutions. We first use the same test-case as in Section 4.3, but this time with an anisotropic tensor: we set $l_{x_1} = 10^{-2}$. The corresponding steady-state is

$$u^\infty(x_1, x_2) = 2C_1\pi e^{x_1 - \frac{1}{2}}.$$

We compute the discrete solutions on the time interval $[0, 350]$, with $\Delta t = 10^{-1}$, on two Kershaw meshes of sizes 0.02 and 0.006. On Figure 4, we display the evolution along time of the L^1 -distance between the reconstructed discrete densities and u^∞ , computed as

$$\|u_{\mathcal{M}}^{\omega, n} - u^\infty\|_{L^1(\Omega)} \quad \text{and} \quad \|u_{\mathcal{M}}^n - u^\infty\|_{L^1(\Omega)} \quad (4.5)$$

for, respectively, the exponential fitting scheme, and the nonlinear scheme. We here focus on `exp_f_1`, and on `nlhho_k` for $k \in \{0, 1, 2\}$ (as well as on `nlhho_1_0`). For all schemes, we observe the exponential convergence towards the thermal equilibrium, until machine precision is reached. Remark that, for the test-case considered here, $\phi \in \mathbb{P}^1(\Omega)$, therefore $\phi_{\mathcal{M}} = \phi$ for all $k \geq 0$. It follows that $u_{\mathcal{M}}^\infty = u^\infty$ (recall that we always have $u_{\mathcal{M}}^{\omega, \infty} = u^\infty$). This is exactly what we observe in the numerical experiments. As previously, `nlhho_1` and `nlhho_1_0` exhibit an extremely similar behaviour. Also, for $k \geq 1$, we observe that the decay rates are similar to the exact one α , and do not seem to depend on the size of the mesh. For $k = 0$, the decay rate differs a bit from α on the coarsest mesh, but these two rates seem to coincide on a sufficiently refined mesh.

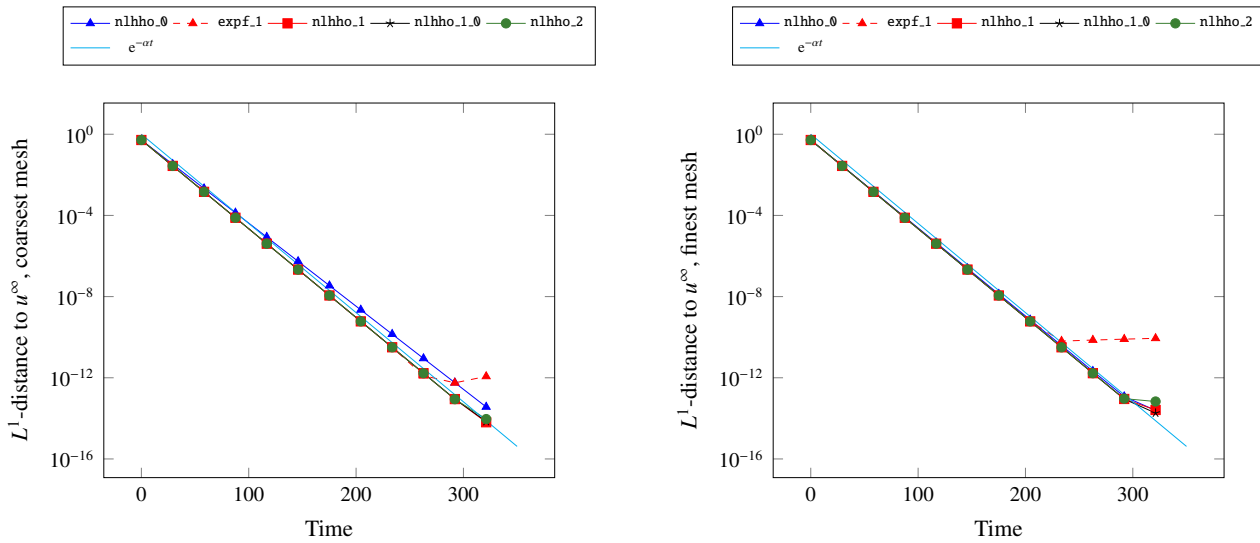


Figure 4. Long-time behaviour of discrete solutions. L^1 -distance to u^∞ on Kershaw meshes.

As a last test-case, we consider an advective potential and an anisotropy tensor set to

$$\phi(x_1, x_2) = -\frac{1}{2} \log(1 + (x_1 - x_2)^2 + 3x_2^2) \quad \text{and} \quad \Lambda = \begin{pmatrix} 10^3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Our initial datum is

$$u^0(x_1, x_2) = 1 + \frac{1}{2} \cos(2\pi x_1) \sin(2\pi x_2).$$

The corresponding thermal equilibrium therefore reads

$$u^\infty(x_1, x_2) = \frac{1}{\int_{(0,1)^2} e^{-\phi}} \sqrt{1 + (x_1 - x_2)^2 + 3x_2^2}.$$

Remark that the potential ϕ is not (piecewise) polynomial. As previously, we investigate the long-time behaviour of the schemes. We compute the discrete solutions on the time interval $[0, 5]$, with $\Delta t = 0.2$, on two distorted quadrangular meshes featuring, respectively, 64 and 1024 cells. On Figure 5, we display the evolution of the L^1 -distance to equilibrium, as defined in (4.5), for both the `expf.k` and `nlhho.k` schemes, for $k \in \{0, 1, 2, 3\}$. For all schemes, we observe the exponential convergence towards the thermal equilibrium, until some precision is reached. For the exponential fitting schemes, machine precision is attained (which is expected since $u_{\mathcal{M}}^{\omega, \infty} = u^\infty$), whereas for the nonlinear schemes (for which $u_{\mathcal{M}}^\infty$ is an approximation of u^∞), the precision increases, as expected, with the polynomial degree and as the mesh is refined. Also, all schemes with $k \geq 1$ seem to exhibit a similar, meshsize-independent decay rate. For $k = 0$, the decay rate seems slightly sensitive to the mesh size, but tends to reach the expected value on a sufficiently refined mesh.

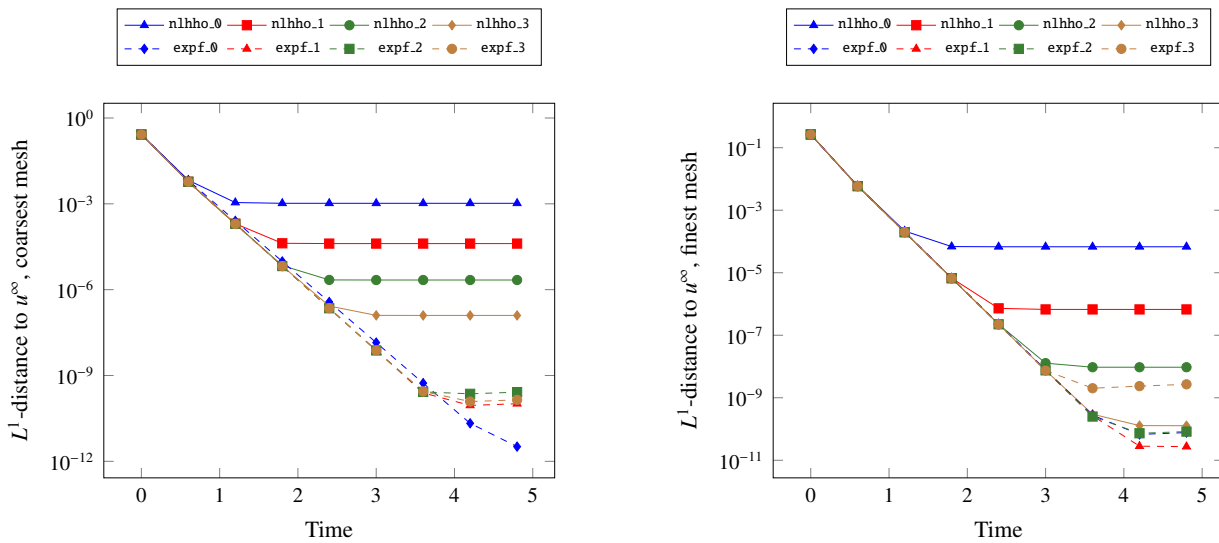


Figure 5. Long-time behaviour of discrete solutions. L^1 -distance to u^∞ on distorted quadrangular meshes.

5. Conclusions

In this paper, we have studied two arbitrary-order hybrid methods for the approximation of linear, anisotropic, potential-driven advection-diffusion equations on general polytopal meshes. The first one is a linear scheme, which is based on the exponential fitting strategy, whereas the second is a nonlinear scheme, whose building principles are adapted from the low-order constructions of [12, 13, 16]. We proved that both schemes admit solutions, possess a discrete entropy structure, and preserve the mass, the thermal equilibrium, and the long-time asymptotics. Moreover, the solutions to the nonlinear scheme are positive by construction. We have validated these theoretical results on a set of numerical test-cases. We have unraveled the positivity violation of the linear methods, which justifies the use of (more costly) nonlinear methods. In the meantime, the use of nonlinear schemes with polynomial unknowns of higher degree results in an important gain of efficiency (accuracy vs. computational cost). These results confirm the benefits of using high-order nonlinear schemes in order to get reliable approximations of dissipative problems. Future research directions include a full analysis of the nonlinear scheme, in particular of its convergence (with respect to the discretisation parameters) and time-asymptotic properties, as well as the development of similar schemes for more complex, nonlinear problems, like semiconductor models (based on [45]).

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The authors would like to thank the anonymous reviewers for their remarks and suggestions which helped improving the quality of the presentation. The authors also thank J. Droniou for his insightful

comments about this work, and for pointing out a simplification of the proof of Lemma 1. The authors finally thank C. Chainais-Hillairet and M. Herda for fruitful discussions on the topic. This research was funded in part by the Austrian Science Fund (FWF) project 10.55776/F65. The authors also acknowledge support from the LabEx CEMPI (ANR-11-LABX-0007).

Conflict of interest

The authors declare no conflicts of interest.

References

1. J. Aghili, D. A. Di Pietro, B. Ruffini, An hp -hybrid high-order method for variable diffusion on general meshes, *Comput. Methods Appl. Math.*, **17** (2017), 359–376. <https://doi.org/10.1515/cmam-2017-0009>
2. G. R. Barrenechea, E. H. Georgoulis, T. Pryer, A. Veiser, A nodally bound-preserving finite element method, *IMA J. Numer. Anal.*, 2023. <https://doi.org/10.1093/imanum/drad055>
3. G. R. Barrenechea, V. John, P. Knobloch, Finite element methods respecting the discrete maximum principle for convection-diffusion equations, *SIAM Rev.*, 2023.
4. L. Beirão da Veiga, J. Droniou, G. Manzini, A unified approach for handling convection terms in finite volumes and mimetic discretization methods for elliptic problems, *IMA J. Numer. Anal.*, **31** (2011), 1357–1401. <https://doi.org/10.1093/imanum/drq018>
5. X. Blanc, E. Labourasse, A positive scheme for diffusion problems on deformed meshes, *ZAMM-J. Appl. Math. Mech.*, **96** (2016), 660–680. <https://doi.org/10.1002/zamm.201400234>
6. T. Bodineau, J. Lebowitz, C. Mouhot, C. Villani, Lyapunov functionals for boundary-driven nonlinear drift-diffusion equations, *Nonlinearity*, **27** (2014), 2111. <https://doi.org/10.1088/0951-7715/27/9/2111>
7. F. Bonizzoni, M. Braukhoff, A. Jüngel, I. Perugia, A structure-preserving discontinuous Galerkin scheme for the Fisher-KPP equation, *Numer. Math.*, **146** (2020), 119–157. <https://doi.org/10.1007/s00211-020-01136-w>
8. M. Braukhoff, I. Perugia, P. Stocker, An entropy structure preserving space-time formulation for cross-diffusion systems: analysis and Galerkin discretization, *SIAM J. Numer. Anal.*, **60** (2022), 364–395. <https://doi.org/10.1137/20M1360086>
9. F. Brezzi, K. Lipnikov, V. Simoncini, A family of mimetic finite difference methods on polygonal and polyhedral meshes, *Math. Mod. Meth. Appl. Sci.*, **15** (2005), 1533–1551. <https://doi.org/10.1142/S0218202505000832>
10. F. Brezzi, L. D. Marini, P. Pietra, Two-dimensional exponential fitting and applications to drift-diffusion models, *SIAM J. Numer. Anal.*, **26** (1989), 1342–1355. <https://doi.org/10.1137/0726078>
11. C. Cancès, C. Chainais-Hillairet, M. Herda, S. Krell, Large time behavior of nonlinear finite volume schemes for convection-diffusion equations, *SIAM J. Numer. Anal.*, **58** (2020), 2544–2571. <https://doi.org/10.1137/19M1299311>

12. C. Cancès, C. Chainais-Hillairet, S. Krell, Numerical analysis of a nonlinear free-energy diminishing discrete duality finite volume scheme for convection diffusion equations, *Comput. Methods Appl. Math.*, **18** (2018), 407–432. <https://doi.org/10.1515/cmam-2017-0043>
13. C. Cancès, C. Guichard, Numerical analysis of a robust free-energy diminishing finite volume scheme for parabolic equations with gradient structure, *Found. Comput. Math.*, **17** (2017), 1525–1584. <https://doi.org/10.1007/s10208-016-9328-6>
14. C. Chainais-Hillairet, J. Droniou, Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions, *IMA J. Numer. Anal.*, **31** (2011), 61–85. <https://doi.org/10.1093/imanum/drp009>
15. C. Chainais-Hillairet, M. Herda, Large-time behaviour of a family of finite volume schemes for boundary-driven convection-diffusion equations, *IMA J. Numer. Anal.*, **40** (2020), 2473–2504. <https://doi.org/10.1093/imanum/drz037>
16. C. Chainais-Hillairet, M. Herda, S. Lemaire, J. Moatti, Long-time behaviour of hybrid finite volume schemes for advection-diffusion equations: linear and nonlinear approaches, *Numer. Math.*, **151** (2022), 963–1016. <https://doi.org/10.1007/s00211-022-01289-w>
17. P. G. Ciarlet, Discrete maximum principle for finite-difference operators, *Aeq. Math.*, **4** (1970), 338–352. <https://doi.org/10.1007/BF01844166>
18. P. G. Ciarlet, P. A. Raviart, Maximum principle and uniform convergence for the finite element method, *Comput. Methods Appl. Mech. Eng.*, **2** (1973), 17–31. [https://doi.org/10.1016/0045-7825\(73\)90019-4](https://doi.org/10.1016/0045-7825(73)90019-4)
19. M. Cicuttin, A. Ern, N. Pignet, *Hybrid high-order methods: a primer with applications to solid mechanics*, Cham: Springer, 2021. <https://doi.org/10.1007/978-3-030-81477-9>
20. B. Cockburn, D. A. Di Pietro, A. Ern, Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods, *ESAIM: Math. Model. Numer. Anal.*, **50** (2016), 635–650. <https://doi.org/10.1051/m2an/2015051>
21. P. L. Colin, *Numerical analysis of drift-diffusion models: convergence and asymptotic behaviors*, Ph.D. Thesis, Université de Lille 1, 2016.
22. M. Corti, F. Bonizzoni, P. F. Antonietti, Structure preserving polytopal discontinuous Galerkin methods for the numerical modeling of neurodegenerative diseases, *arXiv*, 2023. <https://doi.org/10.48550/arXiv.2308.00547>
23. D. A. Di Pietro, J. Droniou, *The hybrid high-order method for polytopal meshes: design, analysis, and applications*, Vol. 19, Cham: Springer, 2020. <https://doi.org/10.1007/978-3-030-37203-3>
24. D. A. Di Pietro, J. Droniou, A. Ern, A discontinuous-skeletal method for advection-diffusion-reaction on general meshes, *SIAM J. Numer. Anal.*, **53** (2015), 2135–2157. <https://doi.org/10.1137/140993971>
25. D. A. Di Pietro, A. Ern, A hybrid high-order locking-free method for linear elasticity on general meshes, *Comput. Methods Appl. Mech. Eng.*, **283** (2015), 1–21. <https://doi.org/10.1016/j.cma.2014.09.009>

26. D. A. Di Pietro, A. Ern, S. Lemaire, An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators, *Comput. Methods Appl. Math.*, **14** (2014), 461–472. <https://doi.org/10.1515/cmam-2014-0018>
27. K. Domelevo, P. Omnes, A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids, *ESAIM: Math. Model. Numer. Anal.*, **39** (2005), 1203–1249. <https://doi.org/10.1051/m2an:2005047>
28. J. Droniou, Finite volume schemes for diffusion equations: introduction to and review of modern methods, *Math. Mod. Meth. Appl. Sci.*, **24** (2014), 1575–1619. <https://doi.org/10.1142/S0218202514400041>
29. J. Droniou, C. Le Potier, Construction and convergence study of schemes preserving the elliptic local maximum principle, *SIAM J. Numer. Anal.*, **49** (2011), 459–490. <https://doi.org/10.1137/090770849>
30. J. Droniou, L. Yemm, Robust hybrid high-order method on polytopal meshes with small faces, *Comput. Methods Appl. Math.*, **22** (2022), 47–71. <https://doi.org/10.1515/cmam-2021-0018>
31. D. A. Dunavant, High degree efficient symmetrical Gaussian quadrature rules for the triangle, *Int. J. Numer. Methods Eng.*, **21** (1985), 1129–1148.
32. L. C. Evans, *Partial differential equations*, 2 Eds., American Mathematical Society, 2010.
33. R. Eymard, T. Gallouët, R. Herbin, Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces, *IMA J. Numer. Anal.*, **30** (2010), 1009–1043. <https://doi.org/10.1093/imanum/drn084>
34. R. Eymard, C. Guichard, R. Herbin, Small-stencil 3D schemes for diffusive flows in porous media, *ESAIM: Math. Model. Numer. Anal.*, **46** (2012), 265–290. <https://doi.org/10.1051/m2an/2011040>
35. I. Faragó, J. Karátson, S. Korotov, Discrete maximum principles for nonlinear parabolic PDE systems, *IMA J. Numer. Anal.*, **32** (2012), 1541–1573. <https://doi.org/10.1093/imanum/drr050>
36. H. Gajewski, K. Gärtner, On the discretization of van Roosbroeck’s equations with magnetic field, *ZAMM-J. Appl. Math. Mech.*, **76** (1996), 247–264. <https://doi.org/10.1002/zamm.19960760502>
37. H. Gajewski, K. Gröger, Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics, *Math. Nachr.*, **140** (1989), 7–36. <https://doi.org/10.1002/mana.19891400102>
38. F. Hermeline, A finite volume method for the approximation of diffusion operators on distorted meshes, *J. Comput. Phys.*, **160** (2000), 481–499. <https://doi.org/10.1006/jcph.2000.6466>
39. C. Lehrenfeld, *Hybrid discontinuous Galerkin methods for solving incompressible flow problems*, MS.Thesis, Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen, 2010.
40. C. Lehrenfeld, J. Schöberl, High order exactly divergence-free hybrid discontinuous Galerkin methods for unsteady incompressible flows, *Comput. Methods Appl. Mech. Eng.*, **307** (2016), 339–361. <https://doi.org/10.1016/j.cma.2016.04.025>
41. H. Liu, Z. Wang, An entropy satisfying discontinuous Galerkin method for nonlinear Fokker-Planck equations, *J. Sci. Comput.*, **68** (2016), 1217–1240. <https://doi.org/10.1007/s10915-016-0174-0>

42. H. Liu, H. Yu, Maximum-principle-satisfying third order discontinuous Galerkin schemes for Fokker-Planck equations, *SIAM J. Sci. Comput.*, **36** (2014), A2296–A2325. <https://doi.org/10.1137/130935161>
43. P. A. Markowich, A. Unterreiter, Vacuum solutions of a stationary drift-diffusion model, *Ann. Scuola Norm. Sup. Pisa-Cl. Sci.*, **20** (1993), 371–386.
44. J. Moatti, A skeletal high-order structure preserving scheme for advection-diffusion equations, In: E. Franck, J. Fuhrmann, V. Michel-Dansac, L. Navoret, *Finite volumes for complex applications X–Volume 1: elliptic and parabolic problems*, Springer Proceedings in Mathematics & Statistics, Cham: Springer, **432** (2023), 345–354. https://doi.org/10.1007/978-3-031-40864-9_29
45. J. Moatti, A structure preserving hybrid finite volume scheme for semiconductor models with magnetic field on general meshes, *ESAIM: Math. Model. Numer. Anal.*, **57** (2023), 2557–2593. <https://doi.org/10.1051/m2an/2023041>
46. E. H. Quenjel, Positive Scharfetter–Gummel finite volume method for convection-diffusion equations on polygonal meshes, *Appl. Math. Comput.*, **425** (2022), 127071. <https://doi.org/10.1016/j.amc.2022.127071>
47. D. L. Scharfetter, H. K. Gummel, Large-signal analysis of a silicon Read diode oscillator, *IEEE Trans. Electron Dev.*, **16** (1969), 64–77. <https://doi.org/10.1109/T-ED.1969.16566>
48. M. Schneider, L. Agélas, G. Enchéry, B. Flemisch, Convergence of nonlinear finite volume schemes for heterogeneous anisotropic diffusion on general meshes, *J. Comput. Phys.*, **351** (2017), 80–107. <https://doi.org/10.1016/j.jcp.2017.09.003>
49. Z. Sheng, J. Yue, G. Yuan, Monotone finite volume schemes of nonequilibrium radiation diffusion equations on distorted meshes, *SIAM J. Sci. Comput.*, **31** (2009), 2915–2934. <https://doi.org/10.1137/080721558>
50. W. Van Roosbroeck, Theory of the flow of electrons and holes in germanium and other semiconductors, *Bell Syst. Tech. J.*, **29** (1950), 560–607. <https://doi.org/10.1002/j.1538-7305.1950.tb03653.x>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>).