Journal of Computational and Applied Mathematics 310 (2017) 19–31

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Arbitrary dimension convection–diffusion schemes for space–time discretizations☆

Randolph E. Bank [a], Panayot S. Vassilevski [b,*], Ludmil T. Zikatanov [c,d]

[a] Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, United States

[b] Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, Mail Stop L-561, Livermore, CA 94551, United States

[c] Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, United States

[d] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 20 April 2016

Keywords:
Space–time formulation
Convection–diffusion problems
Finite-element method
Exponential fitting
Streamline-diffusion

ABSTRACT

This note proposes embedding a time dependent PDE into a convection–diffusion type PDE (in one space dimension higher) with singularity, for which two discretization schemes, the classical streamline-diffusion and the EAFE (edge average finite element) one, are investigated in terms of stability and error analysis. The EAFE scheme, in particular, is extended to be arbitrary order which is of interest on its own. Numerical results, in combined space–time domain demonstrate the feasibility of the proposed approach.

1. Introduction

The embedding of time-dependent problems into a one space dimension higher stationary problem is not a new idea. It has many appealing properties, such as: using already existing tools developed for stationary problems; using adaptive methods with reliable and efficient error control; the ability to use existing efficient solver libraries developed for stationary problems. There is, however, a drawback: typically, the memory needed to run a simulation using the combined space–time discretization approach is increased by an order of magnitude. One way to keep the memory required by such methods under control is to use time intervals with fixed length. Another, more general, remedy to the extensive use of computer memory in space–time simulations is to employ accurate dimension reduction algorithms, both in space and in time, which can lead to coarser problems with fewer degrees of freedom, also known as upscaled discretizations. Indeed, an accurate coarser problem can replace the expensive, in terms of memory, fine-grid one and still provide a reliable discretization tool. For a general dimension reduction approach by coarsening (in three space dimensions), we refer to [1]. The extension of the technique proposed in [1] to 4D space–time elements is a work in progress. Another feasible approach for dimension reduction in space–time discretizations is to exploit sparse grids, as proposed in [2]. More recently, discrete space–time schemes using B-splines and Non-Uniform Rational Basis Splines (NURBS) have been employed (see [3]) to yield stable isogeometric analysis methods for the numerical solution of parabolic PDEs in fixed and moving spatial domains.

☆ This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The work of the first author was supported in part by the National Science Foundation under contract DMS-1318480 and the work of the third author was supported in part by the National Science Foundation under contracts DMS-1418843 and DMS-1522615.

* Corresponding author.
E-mail addresses: rbank@ucsd.edu (R.E. Bank), panayot@llnl.gov (P.S. Vassilevski), ludmil@psu.edu (L.T. Zikatanov).

http://dx.doi.org/10.1016/j.cam.2016.04.029
0377-0427/

We point out that in the present note we do not consider dimension reduction techniques. Rather, as a first step, we study the accuracy and stability of the proposed embedding. More specifically, for the discretization of the space–time formulation of a parabolic problem we exploit two well-known techniques for convection–diffusion equations: the streamline diffusion method [4] (see also [5,6]) and the EAFE—Edge Average Finite Element scheme [7] (see also [8,9]). Let us add that the high order EAFE method developed here provides a novel, high order, exponentially fitted discretization for convection–diffusion problems with suitable stability and approximation properties.

The structure of the remainder of this note is as follows. In Section 2, we introduce the space–time formulation of parabolic problems. Then, in Section 3, we present the streamline diffusion method in our space–time setting. Section 4, contains the derivation of the high order EAFE scheme on simplicial finite element grids in arbitrary spatial dimension. The application details for the lowest order EAFE discretization to parabolic problems are given in Section 5. Finally, in Section 6, we present numerical tests showing the optimality and efficiency of both schemes for space–time formulation of parabolic problems. We conclude this paragraph with remark on the terminology: as the EAFE scheme may be viewed as a multidimensional Scharfetter–Gummel discretization [10], in what follows, we use the terms "EAFE discretization" and "Scharfetter–Gummel discretization" interchangeably.

2. Space–time formulation of parabolic problems

We consider the following parabolic problem:

$$\begin{aligned} &u_t - \operatorname{div}(K(x)\nabla u - \boldsymbol{\beta}\cdot u) + \gamma u = f, \quad x \in \Omega_s, \\ &u = 0, \quad x \in \Gamma = \partial\Omega_s; \qquad u(x,0) = u_0(x), \quad x \in \Omega_s. \end{aligned} \tag{1}$$

Here, $\boldsymbol{\beta}$ is a vector field (a velocity) and $K(x)$ is, in general, a scalar (or $d \times d$ tensor valued) function. Let $\Omega_t = (0, t_{\max})$ be the time interval of interest. The space–time domain is $\Omega = \Omega_s \times \Omega_t$. For convenience we have assumed homogeneous Dirichlet boundary conditions $u = 0$ on $\partial\Omega_s \times \Omega_t$. In treating time as a space-like variable, the initial condition at $t = 0$ becomes a Dirichlet boundary condition for the $(d+1)$ dimensional problem.

In a space–time formulation, introducing a new variable $y = (x, t)$ then gives the following convection–diffusion equation: Find $u = u(y)$ such that

$$\begin{aligned} &-\operatorname{div}_y(D\nabla_y u + \mathbf{b}\cdot\nabla_y u) + \gamma u = f \quad \text{in } \Omega\ \mathbf{b} = (\boldsymbol{\beta}^t, 1)^t : \Omega \mapsto \mathbb{R}^{d+1}, \\ &u = 0 \quad \text{on } \Gamma = \partial\Omega \times \Omega_t; \qquad u = u_0 \quad \text{on } \Gamma_0 = \Omega_s \times \{t = 0\}. \end{aligned}$$

Without loss of generality we may assume that $u_0 = 0$ and we define $\mathcal{H}^1_E(\Omega)$ as the subspace of $\mathcal{H}^1(\Omega)$ satisfying these homogeneous Dirichlet boundary conditions.

In the following we consider two schemes for discretization of convection–diffusion problems and apply them to space–time formulations of (1). These are the Streamline Diffusion and the Scharfetter–Gummel (EAFE) discretizations. For the latter we need a non-singular D, while above $D = \begin{bmatrix} K & 0 \\ 0 & 0 \end{bmatrix}$ is actually degenerate. To remedy this, we perturb it to make it invertible, i.e., we let $D = \begin{bmatrix} K & 0 \\ 0 & \epsilon \end{bmatrix}$ for a small parameter $\varepsilon > 0$.

3. Streamline diffusion

We first consider a simple case when $K = \alpha I, \alpha > 0, \gamma \geq 0$, and $\boldsymbol{\beta}$ is constant. Then Eq. (1) has the form:

$$Lu \equiv u_t - \alpha\Delta u + \boldsymbol{\beta}\cdot\nabla u + \gamma u = f. \tag{2}$$

The results below generalize to the variable coefficient case in a straightforward and well-studied fashion. Here we consider the constant coefficients case only in an attempt to keep the focus on the important aspect of time discretization. In allowing for different sizes of α, $\boldsymbol{\beta}$ and γ, our analysis covers several scenarios of interest. For simplicity we assume the initial condition $u_0 = 0$.

The weak form of (2) is given by: find $u \in \mathcal{H}^1_E$ such that

$$(u_t, v) + \alpha(\nabla u, \nabla v) + (\boldsymbol{\beta}\cdot\nabla u, v) + \gamma(u, v) = f(v)$$

for all $v \in \mathcal{H}^1_E$. The space–time bilinear for $B(u, v)$ is given by

$$B(u, v) = \int_0^T (u_t, v) + \alpha(\nabla u, \nabla v) + (\boldsymbol{\beta}\cdot\nabla u, v) + \gamma(u, v)\, dt$$

where

$$(u, v) = \int_{\Omega_s} uv\, dx$$

is the usual $\mathcal{L}_2$ inner product on Ω_s. The right hand side is given by the linear functional

$$F(v) = \int_0^T f(v)\, dt.$$

We assume that the space–time domain Ω is covered by a shape regular quasiuniform tessellation $\mathcal{T}_h$ of elements of size h. The energy norm for this problem is given by

$$|||u|||^2 = \|u(T)\|^2 + \int_0^T \alpha \|\nabla u\|^2 + h^p \nu \|\boldsymbol{\beta} \cdot \nabla u + u_t\|^2 + \gamma \|u\|^2\, dt$$

where

$$\nu = \frac{1}{\sqrt{|\boldsymbol{\beta}|^2 + 1}}.$$

For technical reasons made clear below, we set $p = 1$ for the important case of continuous piecewise linear approximation, or the special case $\alpha = 0$; otherwise we choose $p = 2$.

We make a standard Petrov–Galerkin streamline diffusion discretization for this $d + 1$ dimensional problem. Let $V_h \subset \mathcal{H}_E^1$ denote a C^0 conforming piecewise polynomial finite element space. The space V_h itself is the trial space. In our Petrov–Galerkin formulation, the test functions are given by $v + \theta h^p \nu(\boldsymbol{\beta} \cdot \nabla v + v_t)$ for $v \in V_h$, where θ is a parameter to be characterized below. The discrete problem is: find $u_h \in V_h$ such that

$$B_h(u_h, v) \equiv B(u_h, v) + \int_0^T (Lu_h, \theta h^p \nu(\boldsymbol{\beta} \cdot \nabla v + v_t))\, dt = F(v + \theta h^p \nu(\boldsymbol{\beta} \cdot \nabla v + v_t))$$

for all $v \in V_h$. Because V_h is only C^0, the term $(Lu_h, \theta h \nu(\boldsymbol{\beta} \cdot \nabla v + v_t))$ is formally interpreted elementwise due to possible discontinuities on inter-element boundaries.

We begin with a basic stability result.

Lemma 1. *Let V_h be the space of continuous piecewise linear polynomials or $\alpha = 0$. For $v \in V_h$, and θ sufficiently small, there exists $C > 0$, independent of h, such that*

$$B_h(v, v) \geq C|||v|||^2. \tag{3}$$

Proof. We first note the term $(\alpha \Delta v, \theta h \nu(v_t + \boldsymbol{\beta} \cdot \nabla v)) = 0$.

The term

$$\int_0^T (v_t + \boldsymbol{\beta} \cdot \nabla v, v)\, dt = \frac{\|v(T)\|^2}{2}$$

and

$$\alpha(\nabla v, \nabla v) = \alpha \|\nabla v\|^2$$
$$(v_t + \boldsymbol{\beta} \cdot \nabla v, \theta h \nu(v_t + \boldsymbol{\beta} \cdot \nabla v)) = \theta h \nu \|v_t + \boldsymbol{\beta} \cdot \nabla v\|^2.$$

Finally

$$\begin{aligned}(\gamma v, v + \theta h \nu(v_t + \boldsymbol{\beta} \cdot \nabla v)) &\geq \gamma \|v\|^2 - \gamma \theta h \nu \|v\|\, \|v_t + \boldsymbol{\beta} \cdot \nabla v\| \\ &\geq \gamma \|v\|^2 \left(1 - \frac{\gamma \theta h \nu}{2}\right) - \frac{\theta h \nu}{2} \|v_t + \boldsymbol{\beta} \cdot \nabla v\|^2.\end{aligned}$$

Combining all these estimates, and taking θ sufficiently small proves (3). □

The orthogonality-like relation for the error $e = u - u_h$ in our approximation is given by

$$B_h(e, v) = 0 \tag{4}$$

for all $v \in V_h$.

For $\chi \in V_h$, let

$$\phi = u_h - \chi$$
$$\eta = u - \chi.$$

Our error relation can be expressed in terms of ϕ and η as

$$B_h(\phi, v) = B_h(\eta, v)$$

for all $v \in V_h$. We take $v = \phi \in V_h$ and use Lemma 1. Then we have

$$|||\phi|||^2 \leq C B_h(\phi, \phi) \leq C B_h(\eta, \phi). \tag{5}$$

Let δ be a sufficiently small parameter to be characterized below. We now estimate all the terms on the right hand side of (5). First,

$$\alpha(\nabla\eta, \nabla\phi) \le C\alpha\|\nabla\eta\|^2 + \delta\alpha\|\nabla\phi\|^2$$
$$(\eta_t + \boldsymbol{\beta}\cdot\nabla\eta, \theta h\nu(\phi_t + \boldsymbol{\beta}\cdot\nabla\phi)) \le C\theta h\nu\|\eta_t + \boldsymbol{\beta}\cdot\nabla\eta\|^2 + \delta\theta h\nu\|\phi_t + \boldsymbol{\beta}\cdot\nabla\phi\|^2$$
$$(\gamma\eta, \phi + \theta h\nu(\phi_t + \boldsymbol{\beta}\cdot\nabla\phi)) \le C\gamma\|\eta\|^2 + \delta(\gamma\|\phi\|^2 + \theta h\nu\|\phi_t + \boldsymbol{\beta}\cdot\nabla\phi\|^2).$$

The fourth term is a bit more involved.

$$\int_0^T (\eta_t + \boldsymbol{\beta}\cdot\nabla\eta, \phi)\,dt = (\eta(T), \phi(T)) - \int_0^T (\eta, \phi_t + \boldsymbol{\beta}\cdot\nabla\phi)\,dt$$
$$\le C\left(\|\eta(T)\|^2 + (h\nu)^{-1}\int_0^T \|\eta\|^2\,dt\right) + \delta\left(\|\phi(T)\|^2 + h\nu\int_0^T \|\phi_t + \boldsymbol{\beta}\cdot\nabla\phi\|^2\,dt\right).$$

Combining these estimates, and making δ sufficiently small, we have

$$|||\phi|||^2 \le C\left(|||\eta|||^2 + \int_0^T (h\nu)^{-1}\|\eta\|^2\,dt\right). \tag{6}$$

Using (6) and the triangle inequality, we obtain

Theorem 2. *Let V_h be the space of continuous piecewise linear polynomials or $\alpha = 0$. Then the error $e = u - u_h$ satisfies*

$$|||u - u_h|||^2 \le C \inf_{\chi\in V_h}\left(|||u - \chi|||^2 + \int_0^T (h\nu)^{-1}\|u - \chi\|^2\,dt\right). \tag{7}$$

Suppose $\alpha = O(1)$ and V_h contains piecewise linear polynomials. Then if $u \in \mathcal{H}^2(\Omega)$, (7) yields an $O(h^{1/2})$ rate of convergence of the space–time gradient in the streamline direction $\mathbf{b} = (\boldsymbol{\beta}^t, 1)^t$, and an optimal $O(h)$ convergence rate for $(\int_0^T \|\nabla(u - u_h)\|^2)^{1/2}$. If $\gamma = O(1)$ we have $O(h)$ convergence for the space–time $\mathcal{L}_2$ norm. While not optimal in every norm considered, overall this is in alignment with well-known behavior for the classical streamline diffusion method. If $\alpha = 0$ and V_h contains piecewise polynomials of degree r, we lose control of the gradient $\|\nabla(u - u_h)\|$ but gain improved $O(h^r)$ convergence for the space–time gradient in the streamline direction, and if $\gamma = O(1)$ we have improved $O(h^{r+1/2})$ convergence for the $\mathcal{L}_2$ norm. These again correspond with classical results for the streamline diffusion method.

If $\alpha \ne 0$ and V_h contains piecewise polynomials of degree $r > 1$, terms similar to $\alpha(\Delta v, v_t)$ become problematic since Δv contains no time derivatives and at present forces us to choose $p = 2$. (One might alternatively consider replacing the diffusion term Δu with $\Delta u + \epsilon u_{tt}$, and then analyzing as in the standard streamline diffusion scenario, but this dilutes the advantage one obtains through the use of higher order approximation since we lose consistency with the original PDE). Here is the analog of Lemma 1.

Lemma 3. *Let V_h be the space of continuous piecewise linear polynomials of degree $r > 1$ and $\alpha \ne 0$. For $v \in V_h$, and θ sufficiently small, there exists $C > 0$, independent of h, such that*

$$B_h(v, v) \ge C|||v|||^2. \tag{8}$$

Proof. Generally the proof follows the same pattern as Lemma 1. The new term is $(\Delta v, \theta h^2\nu(v_t + \boldsymbol{\beta}\cdot\nabla v))$. On a single element $\tau \in \mathcal{T}_h$ we can use a local inverse assumption

$$|(-\Delta v, v_t + \boldsymbol{\beta}\cdot\nabla v)_\tau \le Ch^{-1}\|\nabla v\|_\tau\|v_t + \boldsymbol{\beta}\cdot\nabla v\|_\tau|.$$

Using this estimate, we have

$$\int_0^T \alpha(\nabla v, \nabla v) - (\alpha\Delta v, \theta h^2\nu(v_t + \boldsymbol{\beta}\cdot\nabla v))\,dt \ge \int_0^T \alpha(1 - C\alpha\nu\theta)\|\nabla v\|^2\,dt - \frac{\theta h^2\nu}{4}\|v_t + \boldsymbol{\beta}\cdot\nabla v\|^2.$$

The remaining estimates in the proof of Lemma 1 are the same with h replaced by h^2. □

We analyze the error similar to the proof of Theorem 2. The new term is

$$\alpha(\Delta\eta, \theta h^2\nu(\phi_t + \boldsymbol{\beta}\cdot\nabla\phi)) \le C\alpha^2\nu h^2\|\Delta\eta\|^2 + \delta h^2\nu\|\phi_t + \boldsymbol{\beta}\cdot\nabla\phi\|^2.$$

The remaining terms are estimated as in Theorem 2 with h replaced by h^2, leading to

Theorem 4. *Let V_h be the space of continuous piecewise linear polynomials of degree $r > 1$ and $\alpha \neq 0$. Then the error $e = u - u_h$ satisfies*

$$|||u - u_h|||^2 \leq C \inf_{\chi \in V_h} \left(|||u - \chi|||^2 + \int_0^T (h^2 \nu)^{-1} \|u - \chi\|^2 + \alpha h^2 \|\Delta(u - \chi)\|^2 \, dt \right). \tag{9}$$

If $\alpha = O(1)$, V_h is the space of continuous polynomials of degree $r > 1$, and u is sufficiently smooth, we have optimal $O(h^r)$ convergence for $(\int_0^T \|\nabla(u - u_h)\|^2)^{1/2}$, but only $O(h^{r-1})$ convergence for the space–time gradient in the streamline direction. If $\gamma = O(1)$ we also obtain $O(h^r)$ convergence in the space–time $\mathcal{L}_2$ norm.

A practical remark

Suppose that the space domain Ω_s has a generic length scale L. Since the time units for $\Omega_t = [0, T]$ could be completely unrelated to the space units, the space–time domain $\Omega = \Omega_s \times \Omega_t$ could be quite anisotropic. It could be very long if $T \gg L$ or very short if $T \ll L$. Filling such potentially thin domains with a small number of shape regular elements could be problematic from the practical point of view. Therefore it could be useful to rescale the time variable such that it has a similar scale to the space variables. For example, one could change variables as in

$$\tilde{t} = \frac{Lt}{T} \equiv \kappa t$$

for $0 \leq \tilde{t} \leq L$. The modified space time-domain $\Omega_s \times [0, L]$ is more isotropic, and likely could be tessellated with far fewer shape regular elements. In terms of the partial differential equation,

$$\frac{\partial u}{\partial t} = \kappa \frac{\partial u}{\partial \tilde{t}}$$

making the convection in the time direction larger or smaller depending on the value of κ. In terms of our analysis, we could replace

$$\alpha \to \frac{\alpha}{\kappa} \equiv \tilde{\alpha}$$
$$\beta \to \frac{\beta}{\kappa} \equiv \tilde{\beta}$$
$$\gamma \to \frac{\gamma}{\kappa} \equiv \tilde{\gamma}$$

and directly apply the analysis of the previous section to this modified constant coefficient equation.

4. High order Scharfetter–Gummel discretization

In this section we derive a high order Scharfetter–Gummel scheme on simplicial finite element grids in dimension $d \geq 1$. The original Scharfetter–Gummel difference scheme [10] is a method used in simulating 1-dimensional semiconductor equations. After its discovery, it has been generalized and used for the numerical solution of convection–diffusion equations of the form:

$$-\operatorname{div} J(u) = f, \quad x \in \Omega \subset \mathbb{R}^d \tag{10}$$
$$J(u) = (D(x)\nabla_x u - \mathbf{b}u), \tag{11}$$
$$u(x) = 0, \quad x \in \Gamma_D, \qquad J(u) \cdot \mathbf{n} = 0, \quad x \in \Gamma_N \tag{12}$$
$$Du \cdot \mathbf{n} = 0, \quad x \in \Gamma_R. \tag{13}$$

Here, $J(u)$ is the flux variable which plays an important role in approximating the weak form of the equation. We note that the natural boundary condition is the one given on Γ_N and the boundary condition (13) is of a Robin type for this problem. The weak form of the equation above is: Find $u \in V$ such that

$$a(u, v) + m_R(u, v) = f(v), \tag{14}$$
$$a(u, v) = \int_\Omega J(u) \cdot \nabla v, \qquad f(v) = \int_\Omega f v \tag{15}$$
$$m_R(u, v) = \int_{\Gamma_R} (\mathbf{b} \cdot \mathbf{n}) uv. \tag{16}$$

The variational form is obtained after integration by parts and using the fact that on Γ_R, $Du \cdot \mathbf{n} = J(u) \cdot \mathbf{n} - \mathbf{b} \cdot \mathbf{n}u$.

The Scharfetter–Gummel scheme was extended to more than 1 spatial dimension as the Edge Average Finite Element (EAFE) Scheme. A priori error estimates in any dimension were shown in [7]. This work only considered scalar valued

diffusion coefficients (although in any spatial dimension); a discretization for matrix valued diffusion coefficients was proposed and analyzed in [9]. Related work on exponential fitting in discretizing convection–diffusion equations via mixed finite element methods is [11]. More recently, the techniques from [7] have been utilized to yield a second order gauge invariant discretizations for Pauli and Schrödinger equations (see [12]).

Here, we provide a novel approach which gives a Scharfetter–Gummel discretization for finite element spaces of order $r \geq 1$. Our approach follows the ideas in [7,9]. The extension to $r \geq 1$, however is not at all straightforward and requires results from the recently developed Finite Element Exterior Calculus. The rationale of constructing the high order Scharfetter–Gummel scheme is:

(i) approximate the flux $J(u)$ via the Nédélec elements (discrete differential 1-forms with polynomial coefficients);
(ii) eliminate the flux variable and write the resulting discrete problem in terms of the scalar valued finite element approximation of the solution of (1) u (a 0-form).

To set up the finite element approximation, let us itemize some of the ingredients and the main assumptions needed for the discretization.

- We assume that Ω is covered by a conforming, simplicial, shape-regular mesh $\mathcal{T}_h$. We have $\Omega = \cup\{T \mid T \in \mathcal{T}_h\}$.
- The space V_h is the space of conforming Lagrange finite elements of degree r and for the derivation of the scheme, we also need the 1st-kind-Nédélec polynomial spaces on a fixed element (cf. e.g. [13–16]). The details are described below in Section 4.1.
- We assume that the fluxes J and u are smooth enough so that all the norms of functions below make sense. In particular $J \in W^{1,p}(T)$, for all $T \in \mathcal{T}_h$ and for some $p > d$. The solution u is at least continuous, so that its Lagrange interpolant is well defined.
- We assume that the coefficients D, $\mathbf{b}$ are piece-wise constants with discontinuities aligned with $\mathcal{T}_h$.

Remark 1. The assumption $J \in W^{1,p}(\Omega)$ needs some comments. One important feature of the Scharfetter–Gummel scheme is that the estimates on $\|u_I - u_h\|_{1,\Omega}$ are in terms of norms of the flux $J(u)$. We thus approximate more accurately the interpolant $u_I \in V_h$ of the solution if the flux is smooth, while both the solution and the coefficients can be rough functions. For example, if we look at the 1D problem on $(0, 1)$:

$$-(u' - \beta u)' = 0, \qquad u(0) = 0, \qquad u(1) = 1,$$

we observe that $J(u) = (u' - \beta u)$ is a constant, i.e. smooth, while the solution u may exhibit a sharp boundary layer, depending on β. In fact, in this idealized situation in 1D, the estimate in Theorem 8 implies that $u_I = u_h$, i.e. we have captured the exact solution at the vertices.

Next, we show that (i) and (ii) in the rationale given earlier are computationally feasible steps.

4.1. Notation and Nédélec spaces

Consider the Nédélec space $P^{\mathcal{N}}$, which restricted to any element T is the following polynomial space

$$\mathcal{P}^{\mathcal{N}} = (P_{r-1})^d \oplus \mathbf{S}_r, \qquad P^d_{r-1} \subsetneq P^{\mathcal{N}} \subsetneq (P_r)^d, \tag{17}$$

where P_j, $j = (r-1)$, r is the space of polynomials of degree $\leq j$ on T, and $\mathbf{S}_r$ is a subspace of the space $\mathbf{H}_r$ of vector valued homogeneous polynomials of degree r defined as

$$\mathbf{S}_r = \left\{\mathbf{s} \in \mathbf{H}_r \mid \mathbf{s} \cdot \mathbf{x} = 0\right\}.$$

By definition, the inclusion relations given in Eq. (17) hold on any element $T \in \mathcal{T}_h$. From now on we fix this element. We refer to [13,14,17–19], for the classical and the modern description of these spaces and studies of their properties. In what follows we use some of the tools from [17,18]. In our notation, the lowest order of such polynomials corresponds to $r = 1$.

Further, let $M = \dim P^{\mathcal{N}}$ be the dimension of the Nédélec polynomial space on T. The elements of the basis in the dual space of $P^{\mathcal{N}}$ are known as *degrees of freedom* and we denote them by $\{\eta_j\}_{j=1}^M$. Next, the basis in $P^{\mathcal{N}}$, dual to the degrees of freedom we denote by $\{\boldsymbol{\varphi}_j\}_{j=1}^M$. For general simplex in $\mathbb{R}^d$, the explicit form of the degrees of freedom and their dual basis is found in [17]. For our purposes it is sufficient to note that the functionals η_j can be thought as integrals of traces of functions over sub-simplicies. For the lowest order case, we have

$$\langle \eta_e, \mathbf{v} \rangle = \int_e \mathbf{v} \cdot \boldsymbol{\tau}_e, \qquad \boldsymbol{\varphi}_e = \lambda_i \nabla \lambda_j - \lambda_j \nabla \lambda_i$$

for every edge $e = (i, j)$ of T (there are $\frac{d(d+1)}{2}$ edges). Here, $\boldsymbol{\tau}_e$ is the tangent for edge e, and $\{\lambda_i\}$ are the usual barycentric coordinates for element T (cf., e.g., [20]). Using this notation, we have that any function $\mathbf{v} \in P^{\mathcal{N}}$ can be written as

$$\mathbf{v} = \sum_{j=1}^{M} \langle \eta_j, \mathbf{v} \rangle \boldsymbol{\varphi}_j(\mathbf{x}). \tag{18}$$

We stress that this representation is unique and provides a canonical interpolation operator, which for sufficiently smooth vector valued $\mathbf{v}$ is defined as

$$\Pi^{\mathcal{N}}\mathbf{v} = \sum_{j=1}^{M} \langle \eta_j, \mathbf{v}\rangle \boldsymbol{\varphi}_j(\mathbf{x}). \tag{19}$$

The smoothness of $\mathbf{v}$ must be such that the linear forms $\langle \eta_j, \cdot\rangle$ are bounded.

Consider now the space V_h of Lagrange finite elements of order r. The standard set of the degrees of freedom in such case are point evaluations (see [20, Theorem 2.2.1]) and we denote them by $\{\mu_j\}$. Further, the polynomial basis, dual to these degrees of freedom, we denote by $\{\xi_j\}$. We then have a canonical interpolation operator, well defined for any continuous v. The image of $v \in C^0(\overline{\Omega})$ under this interpolation is denoted by v_I and we have

$$v_I = \sum_{j=1}^{N_h} \langle \mu_j, v\rangle \xi_j(\mathbf{x}). \tag{20}$$

There is no need to distinguish the global interpolation operator (on Ω) and the local one (on $T \in \mathcal{T}_h$) for our considerations and we use the same notation for both. Let us note, however, that when working on fixed $T \in \mathcal{T}_h$ we will use $N = \dim P_r = \binom{r+d}{d}$, instead of $N_h = \dim V_h$.

As is well known [17], we have commutative diagrams linking the Nédélec elements and the Lagrange elements of matching orders (order r here), and on every element T we have

$$\Pi^{\mathcal{N}} \nabla v = \nabla v_I.$$

This relation is in fact a relation between degrees of freedom, namely

$$\langle \eta_j, \nabla v\rangle = \langle \eta_j, \nabla v_I\rangle. \tag{21}$$

This is obvious by using the definition of $\Pi^{\mathcal{N}}$, the fact that $\nabla v_I \in P^{\mathcal{N}}$, and the uniqueness of the representation in (18).

4.2. Derivation of a high order Scharfetter–Gummel scheme

Let us fix $T \in \mathcal{T}_h$ and we start with the definition of J and use that D and $\mathbf{b}$ are constants on T.

$$J(u) = D\nabla u - \mathbf{b}u = \exp(\mathbf{q}\cdot\mathbf{x})D\nabla(\exp(-\mathbf{q}\cdot\mathbf{x})u), \quad \mathbf{q} = D^{-1}\mathbf{b}.$$

Hence, we have

$$\exp(-\mathbf{q}\cdot\mathbf{x})D^{-1}J(u) = \nabla(\exp(-\mathbf{q}\cdot\mathbf{x})u). \tag{22}$$

If we apply now $\langle \eta_j, \cdot\rangle, j = 1 : M$ on both sides, and then use (21) we get

$$\begin{aligned}\langle \eta_j, e^{(-\mathbf{q}\cdot\mathbf{x})}D^{-1}J(u)\rangle &= \langle \eta_j, \nabla(e^{(-\mathbf{q}\cdot\mathbf{x})}u)\rangle \\ &= \left\langle \eta_j, \nabla\left(e^{(-\mathbf{q}\cdot\mathbf{x})}u\right)_I\right\rangle, \\ &= \left\langle \eta_j, \nabla\left(e^{(-\mathbf{q}\cdot\mathbf{x})}u_I\right)_I\right\rangle.\end{aligned}$$

The latter identity on the right hand side above, uses the fact that the "I"-interpolant is based on the functionals μ_j that are based on nodal evaluation. As expected, the right hand side is a gradient of a function in V_h and in summary we have

$$\langle \eta_j, e^{(-\mathbf{q}\cdot\mathbf{x})}D^{-1}J(u)\rangle = \left\langle \eta_j, \nabla\left(e^{(-\mathbf{q}\cdot\mathbf{x})}u_I\right)_I\right\rangle, \quad j = 1, \ldots, M. \tag{23}$$

Introducing now $\boldsymbol{G}(J(u)) \in \mathbb{R}^M$ and $\mathbf{d}(u) \in \mathbb{R}^M$ by

$$[\boldsymbol{G}(J(u))]_j = \langle \eta_j, e^{(-\mathbf{q}\cdot\mathbf{x})}D^{-1}J(u)\rangle, \quad j = 1, \ldots, m \tag{24}$$

$$[\mathbf{d}(u)]_j = \left\langle \eta_j, \nabla\left(e^{(-\mathbf{q}\cdot\mathbf{x})}u_I\right)_I\right\rangle, \quad j = 1, \ldots, M \tag{25}$$

and we can write (23) as

$$\boldsymbol{G}(J(u)) = \mathbf{d}(u). \tag{26}$$

Note that both $\boldsymbol{G}$ and $\mathbf{d}$ are linear operators, mapping vector fields and functions to $\mathbb{R}^M$. We remark that the relation (26) is used later in the definition of the approximate bilinear form, and, in the proof of the error estimates, and, we further stress on the fact that $\mathbf{d}(u) = \mathbf{d}(u_I)$, by definition. The advantage of the exponential weighting in (23) is that it provides a way to accurately approximate the flux by polynomials. Note that the right side of (23) are the values of η_j evaluated on a polynomial, while the left side contains the "true" flux $J(u)$. As we shall see later, it is advantageous to use discretizations schemes based on (23) when the flux is piece-wise smooth. This not only includes the case when u and the PDE coefficients are piece-wise smooth, but also includes many other cases (see Remark 1 for a simple 1D example on this).

The main idea of the Scharfetter–Gummel and EAFE schemes is to approximate $J(u)$,

$$J(u) \approx J_T(u) \in P^{\mathcal{N}},$$

or equivalently, we seek

$$J_T(u) = \sum c_j \boldsymbol{\varphi}_j,$$

for some coefficient vector $\mathbf{c} = (c_j)$. The coefficient $\mathbf{c}$ is chosen so that the relation (23) still holds for the approximation. An important question is whether this is possible. If $r = 1$, and we use the lowest order Nédélec elements, this is definitely the case as shown in the earlier works [7,9].

A construction of exponentially-fitted discretizations with higher order polynomial spaces is a bit more intricate. In general, we would like to find $J_T(u) \in P^{\mathcal{N}}$. A key observation is that in order to derive our scheme, we use the weak form of Eq. (10) and we will aim to approximate the weak form as follows:

$$\int_T J(u_h) \cdot \nabla v_h \approx \int_T J_T(u_h) \cdot \nabla v_h,$$

for functions $v_h \in V_h$ and $J_T(u_h) \approx J(u_h)$. As on T, $\nabla v_h \in (P_{r-1})^d$, it is sufficient to look for approximations $J_T(u) \in (P_{r-1})^d \subset P^{\mathcal{N}}$.

We now explore this observation and look at how it affects the identity (26). Let P be the matrix representation of the embedding $(P_{r-1})^d \subset P^{\mathcal{N}}$. To define this matrix, let $\{\boldsymbol{\psi}_m\}_{m=1}^{M_0}$ be a basis in $(P_{r-1})^d$, with $M_0 = \dim(P_{r-1})^d$ and let $\{\boldsymbol{\varphi}_j\}_{j=1}^{M}$ be the basis in $P^{\mathcal{N}}$, dual to the degrees of freedom $\{\eta_m\}_{m=1}^{M}$. Then the entries of P are the coefficients in the expansion of $\boldsymbol{\psi}_m$ in terms of $\{\boldsymbol{\varphi}_j\}_{j=1}^{M}$, and we have,

$$\boldsymbol{\psi}_k = \sum_{j=1}^{M} p_{jk} \boldsymbol{\varphi}_j, \quad \text{with } p_{mk} = \langle \eta_m, \boldsymbol{\psi}_k \rangle. \tag{27}$$

Note that $\nabla\big(\exp(-\mathbf{q}\cdot\mathbf{x})u_I\big)_I$ is an element $(P_{r-1})^d$, and, as such, it can be written as a linear combination via $\{\boldsymbol{\psi}_k\}_{k=1}^{M_0}$. Recalling the definition of $\mathbf{d}(u)$ in (25) then leads to the following useful relations:

$$\nabla\big(\exp(-\mathbf{q}\cdot\mathbf{x})u_I\big)_I = \sum_{k=1}^{M_0} \widetilde{d}_k \boldsymbol{\psi}_k, \qquad [\mathbf{d}(u)]_j = \sum_{k=1}^{M_0} \langle \eta_j, \boldsymbol{\psi}_k \rangle \widetilde{d}_k,$$

$$\mathbf{d}(u) = P\widetilde{\mathbf{d}}.$$

As a consequence, to define the approximation J_T, we need to find a solution of the following problem

$$P^* Z P \widetilde{\mathbf{c}} = P^* P \widetilde{\mathbf{d}}, \tag{28}$$

where we have set $\mathbf{c} = P\widetilde{\mathbf{c}}$, and, as we have shown, $\mathbf{d} = \widetilde{\mathbf{d}}$. Above the matrix $Z \in \mathbb{R}^{M\times M}$ has entries

$$Z_{jk} = \langle \eta_j, e^{-\mathbf{b}\cdot D^{-1}\mathbf{x}} D^{-1} \boldsymbol{\varphi}_k \rangle.$$

The following remark is in order. In general, we may have tried to solve the following system of equations for the coefficients $\mathbf{c}$:

$$Z\mathbf{c} = \mathbf{d}. \tag{29}$$

Clearly, if this is a well posed problem, then we can find the approximation $J_T(u)$. However, as it described above, we only need the solution in subspace, i.e., to solve problem (28). To show that the subspace problem (28) is solvable it is sufficient to show that ZP is injective, and since P is injective, it is sufficient to show that Z is injective on the range of P.

We let

$$Z^\dagger = P\left(P^* Z P\right)^{-1} P^*. \tag{30}$$

In the following, we will simply denote $\mathbf{c} = P\widetilde{\mathbf{c}}$ by $\mathbf{c} = Z^\dagger \mathbf{d}$, or, by (26), by $Z^\dagger G(J(u))$.

The following lemma follows by the construction of the approximation $J_T(u)$.

Lemma 5. *If $J(u)$ is polynomial of degree $r-1$, then its approximation $J_T(u)$ defined by $Z^\dagger G(J(u))$ coincides with $J(u)$.*

4.2.1. A unisolvence result

Note that, when $\mathbf{b} = 0$, the solvability of such system follows from the fact that the Nédélec degrees of freedom form a unisolvent set of functionals on $P^{\mathcal{N}} \supset (P_{r-1})^d$. Multiplying by the exponent changes the game, and, we need to prove some

of the basic results on unisolvence of Nédélec degrees of freedom for quasi-polynomials which we state in the following lemma.

Lemma 6. *The matrix ZP is injective, or, equivalently, if* $\mathbf{p} \in (P_{r-1})^d$ *and* $\langle \eta_j, e^{(-\mathbf{b}\cdot D^{-1}\mathbf{x})}\mathbf{p}\rangle = 0$, *for all* $j = 1 : M$, *then* $\mathbf{p} = 0$.

Proof. This proof follows exactly the lines of the proofs of [18, Lemma 4.5, Lemma 4.6]. The only modifications needed are that we multiply by an exponential function, which is positive everywhere. The rest of the arguments carry over without any change. A different proof of this lemma, in terms of vector proxies, which parallels the proofs of [18, Lemma 4.5, Lemma 4.6] for the specific case considered here, is found in [21, Appendix A]. □

4.2.2. Derivation of the discrete problem

Since now the approximation $J_T(u_h)$, for $u_h \in V_h$ is well defined, due Lemma 6, we have a natural approximating bilinear form. For $u_h \in V_h$ and $v_h \in V_h$ we set

$$\begin{aligned} a_h(u_h, v_h) &= \sum_T \int_T J_T(u_h) \cdot \nabla v_h \\ &= \sum_T \sum_{j=1}^{M} \int_T [Z_T^{\dagger} d_T(u_h)]_j \int_T \boldsymbol{\varphi}_j \cdot \nabla v_h. \end{aligned} \tag{31}$$

The coefficients in $J_T(u_h)$ are determined by $Z_T^{\dagger}\mathbf{d}_T(u_h)$, for all $T \in \mathcal{T}_h$, which in turn indeed makes the right side of (31) to depend only on the degrees of freedom of u_h.

We then define the following discrete problem: Find $u_h \in V_h$ such that

$$a_h(u_h, v) = f(v), \quad \text{for all } v \in V_h. \tag{32}$$

We note another useful relation which follows from the derivation above and is used in the error estimates below. It is an analogue of [7, Equation (3.16)], and [9, Equation (3.8)] and it plays a crucial role in the a priori error estimates. In particular it is useful to estimate the deviation of the derived discrete scheme from the standard Galerkin one (with bilinear form $a(u, v) = \int_\Omega (D\nabla u - \mathbf{b}u) \cdot \nabla v$).

Lemma 7. *For any continuous u, and sufficiently smooth J, such that* $\Pi^{\mathcal{N}} J(u)$ *is well defined we have:*

$$a_h(u_I, v_h) = \sum_T \sum_{j=1}^{M} [Z_T^{\dagger} G_T(J(u))]_j \boldsymbol{\varphi}_j \cdot \nabla v_h. \tag{33}$$

Proof. Recalling that $d_T(u_I) = d_T(u)$, and substituting (26) in (31) gives the desired result. □

Next, we show that, under certain conditions, this is a well posed problem, and we also prove an *a priori* error estimate.

4.3. Stability and error analysis

Error estimates and other properties of such discretization schemes are found in [7,9]. Here we give an estimate for higher order Scharfetter–Gummel discretization and assume for simplicity, and without loss of any generality that we have Dirichlet boundary conditions. We have the following theorem:

Theorem 8. *Assume that* $a(\cdot, \cdot)$ *is invertible on* V_h. *Then, for sufficiently small h, the discrete variational problem* (32) *is well posed and the following error estimate holds:*

$$|u_I - u_h|_{1,\Omega} \le ch^r |J(u)|_{r,p,\Omega}. \tag{34}$$

Proof. From the definition of $a_h(\cdot, \cdot)$, for all $v \in V_h$ we have

$$\begin{aligned} |a(u, v) - a_h(u_I, v)| &= \left| \sum_T a_T(u, v) - a_{h,T}(u_I, v) \right| \\ &\le \sum_T \left| \int_T J(u) \cdot \nabla v - \sum_{j=1}^{M} [Z_T^{\dagger} G_T(J(u))]_j \int_T \boldsymbol{\varphi}_j \cdot \nabla v \right|. \end{aligned}$$

Note that from Lemmas 7 and 5, the right side vanishes for all $J(u)$ that are polynomials of degree $(r - 1)$, and, standard scaling argument shows the estimate

$$|a(u, v) - a_h(u_I, v)| \le ch^r |J(u)|_{r,p,\Omega} |v|_{1,q,\Omega}, \qquad p^{-1} + q^{-1} = 1. \tag{35}$$

The solvability of the discrete problem then follows from the fact that by assumption $a(\cdot,\cdot)$ provides a solvable problem, and hence it satisfies an inf–sup condition on V_h. According to (35) $a(\cdot,\cdot)$ and $a_h(\cdot,\cdot)$ are close when $h \to 0$, and, hence, $a(\cdot,\cdot)$ also satisfies an inf–sup condition for sufficiently small h. This in turn implies that the discrete problem (32) is well posed. The error estimate (34) then follows from the inequality (35). □

Remark 2. For the case $r = 1$ our proof here is analogous to the one given in [7, Lemma 6.2 and Theorem 6.3].

5. Application to parabolic problems

In this section, we recall the parabolic equation (1):

$$\begin{aligned} &u_t - \operatorname{div}(K(x)\nabla_x u - \boldsymbol{\beta} u) = f, \\ &u(x,0) = u_0(x), \quad \text{for } t = 0; \\ &u(x,t) = 0, \quad x \in \Gamma = \partial\Omega \times \{[0, t_{\max})\}. \end{aligned} \tag{36}$$

This equation and the equation discretized by the streamline diffusion method match, if $\operatorname{div}_x \boldsymbol{\beta} = 0$, which we assume to hold. In general, the divergence form comes from a material law and many mathematical models of physical phenomena (if not all) are in divergence form.

The space–time formulation, (written in terms of a flux J_0, and with $y = (x,t)$) then is:

$$\begin{aligned} &-\operatorname{div}_y J_0(u) = f, \qquad \widetilde{J_0}(u) = D_0(x)\nabla_x u - \mathbf{b}u \\ &u(y) = 0, \quad x \in \Gamma = \partial\Omega \times \{(0, t_{\max}]\}, \\ &u(y) = u_0(x), \quad x \in \Gamma_0 = \overline{\Omega} \times \{t = 0\}. \end{aligned} \tag{37}$$

Here we have introduced the semidefinite, tensor valued function $D_0(x)$, and more generally, we denote, $D_\varepsilon(x) : \Omega \mapsto \mathbb{R}^{(d+1)\times(d+1)}$:

$$D_\varepsilon = \begin{pmatrix} K(x) & 0 \\ 0 & \varepsilon \end{pmatrix}, \qquad J_\varepsilon = D_\varepsilon \nabla_y u - \mathbf{b}u. \tag{38}$$

The well known heat equation, $u_t - \Delta u = f$, corresponds to $K(x) = I$ and $\boldsymbol{\beta} = 0$ and $\varepsilon = 0$.

The technique described in the previous section does not work in a straightforward fashion in the case of space–time formulation, because D_0 is a singular matrix. In fact, there is no obvious construction that works in the case of singular D_0. We consider then a formulation using perturbation of the diffusion tensor D_ε and the flux J_ε. Thus, for the parabolic problem we set

$$J_\varepsilon(u) = D_\varepsilon \nabla_y u - \mathbf{b}u, \quad \mathbf{b} = (\boldsymbol{\beta}^T, 1)^T.$$

5.1. Lowest order discretization for parabolic equations

In this section we discuss the Scharfetter–Gummel discretization when applied to space–time formulation of a parabolic equation, in the lowest order case. As a simple, but important example, we consider the simple case of heat equation, i.e. $\boldsymbol{\beta} = 0$, which implies that $\mathbf{b} = \mathbf{e}_{d+1}$, and $\mathbf{e}_{d+1} = (\underbrace{0, \ldots, 0}_{d}, 1)^T$.

We next compute the action of the local stiffness matrix corresponding to a parabolic problem on a vector of degrees of freedom u representing a function in V_h. We fix an element ($(d+1)$ dimensional simplex) $T \in \mathcal{T}_h$ and we denote its barycentric coordinates by $\{\lambda_i\}_{i=1}^{d+2}$ and the space–time coordinates of its vertices are $\{\mathbf{y}_i\}_{i=1}^{d+2} = \{(\mathbf{x}_i, t_i)\}_{i=1}^{d+2}$. The degrees of freedom of a linear polynomial $u \in V_h$ restricted to T are $\{u_i\}_{i=1}^{d+2} = \{u(\mathbf{y}_i)\}_{i=1}^{d+2}$ and we have $u(\mathbf{y}) = \sum_{i=1}^{d+2} u_i \lambda_i(\mathbf{y})$. For an edge $E \in T$, $E = (\mathbf{y}_i, \mathbf{y}_j)$, $i = 1, \ldots (d+2)$, $j = 1, \ldots (d+2)$, we denote

$$\boldsymbol{\tau}_{ij} = \boldsymbol{\tau}_E = \frac{(\mathbf{y}_i - \mathbf{y}_j)}{|\mathbf{y}_i - \mathbf{y}_j|}, \qquad |\mathbf{r}| = \sqrt{\sum_{l=1}^{d+1} r_l^2}, \quad \text{for all } \mathbf{r} \in \mathbb{R}^{d+1}.$$

We note that $\boldsymbol{\tau}_{ij} = -\boldsymbol{\tau}_{ji}$, but as we shall see, this is of no consequence for the final form of the local stiffness matrix. To avoid complications in the presentation coming from unnecessary subscripts we will write D (resp. J) instead of D_ε and (resp. J_ε).

For any $u \in V_h$, as D and $J(u)$ are constants on T, we have the following obvious identities from the definition of J:

$$\begin{aligned} &D^{-1}J \cdot \boldsymbol{\tau}_E = e^{t/\varepsilon}\nabla_y \left(e^{-t/\varepsilon}u\right), \\ &\int_E e^{-t/\varepsilon} D^{-1}J \cdot \tau_E dE = \int_E \nabla_y \left(e^{-t/\varepsilon}u\right) \cdot \tau_E dE, \\ &(D^{-1}J \cdot \tau_E) \int_E e^{-t/\varepsilon} dE = [e^{-t_i/\varepsilon}u(\mathbf{x}_i, t_i) - e^{-t_j/\varepsilon}u(\mathbf{x}_j, t_j)]. \end{aligned}$$

Computing the integral on the left side gives

$$\begin{aligned}\int_E e^{-\frac{t}{\varepsilon}}\,dE &= |E|\int_0^1 \exp\left(-\frac{t_j+s(t_i-t_j)}{\varepsilon}\right)ds\\ &= \frac{|E|\varepsilon}{t_j-t_i}\int_{-t_j/\varepsilon}^{-t_i/\varepsilon} e^{\xi}\,d\xi = |E|\varepsilon\frac{e^{-t_i/\varepsilon}-e^{-t_j/\varepsilon}}{t_j-t_i}\\ &= \frac{|E|e^{-t_i/\varepsilon}}{B\left(\frac{t_i-t_j}{\varepsilon}\right)} = \frac{|E|e^{-t_j/\varepsilon}}{B\left(\frac{t_j-t_i}{\varepsilon}\right)},\end{aligned}$$

where $B(s)=\frac{s}{e^s-1}$ is the Bernoulli function ($B(0)=1$). Note that, $B(s)=e^{-s}B(-s)$ and $B(s)>0$ for all $s\in\mathbb{R}$. We then conclude that on every edge E in T we have:

$$|E|(D^{-1}J\cdot\tau_E) = B\left(\frac{t_i-t_j}{\varepsilon}\right)u(\mathbf{y}_i) - B\left(\frac{t_j-t_i}{\varepsilon}\right)u(\mathbf{y}_j). \tag{39}$$

In the derivation for general order of polynomials we needed the Nédélec basis and spaces. In the lowest order case, we can take a route that does not use these spaces explicitly. In the evaluation of the stiffness matrix entries, we need to compute integrals of the form

$$\int_T (J\cdot\nabla_y\lambda_j) = |T|(J\cdot\nabla_y\lambda_j).$$

We note that since $D^{-1}J$ is a constant on T, we can write it as a gradient of a linear function, namely

$$J = D(D^{-1}J) = D\nabla_y(D^{-1}J\cdot\mathbf{y}) = \sum_{i=1}^{d+2}(D^{-1}J\cdot\mathbf{y}_i)D\nabla_y\lambda_i. \tag{40}$$

Since $\sum_{i=1}^{d+2}\nabla_y\lambda_i\equiv 0$ on T, we have that

$$0 = (D^{-1}J\cdot\mathbf{y}_j)\left(D\sum_{i=1}^{d+2}\nabla_y\lambda_i\cdot\nabla_y\lambda_j\right) = \sum_{i=1}^{d+2}(D^{-1}J\cdot\mathbf{y}_j)(D\nabla_y\lambda_i\cdot\nabla_y\lambda_j).$$

Hence,

$$\begin{aligned}|T|(J\cdot\nabla_y\lambda_j) &= \sum_{i=1}^{d+2}(D^{-1}J\cdot\mathbf{y}_i)(D\nabla_y\lambda_i\cdot\nabla_y\lambda_j)\\ &= \sum_{i\neq j}(D^{-1}J\cdot(\mathbf{y}_i-\mathbf{y}_j))(D\nabla_y\lambda_i\cdot\nabla_y\lambda_j)\\ &= \sum_{i\neq j}|E|(D^{-1}J\cdot\boldsymbol{\tau}_{ij})(D\nabla_y\lambda_i\cdot\nabla_y\lambda_j).\end{aligned}$$

We have computed earlier (see (39)) the quantity $|E|(D^{-1}J\cdot\boldsymbol{\tau}_{ij})$ for all $E\subset\partial T$. Therefore,

$$|T|(J\cdot\nabla_y\lambda_j) = \sum_{i=1;\ i\neq j}^{d+2} d_{ji}^T\left[B\left(\frac{t_i-t_j}{\varepsilon}\right)u_i - B\left(\frac{t_j-t_i}{\varepsilon}\right)u_j\right]. \tag{41}$$

Here $d_{ji}^T=\int_T D\nabla_y\lambda_i\cdot\nabla_y\lambda_j$ are the entries of the local stiffness matrix corresponding to the discretization of $(-\operatorname{div}D\nabla)$ with linear elements on T. Therefore on T we get

$$[A_T]_{jj} = -\sum_{i=1;i\neq j}^{d+2} d_{ji}^T B\left(\frac{t_j-t_i}{\varepsilon}\right), \qquad [A_T]_{ji} = d_{ji}^T B\left(\frac{t_i-t_j}{\varepsilon}\right). \tag{42}$$

The global stiffness matrix is assembled from A_T. It is invertible for sufficiently small mesh size, invertible whenever the assembly of d_{ij}^T gives an M-matrix.

For more detailed discussions about sufficient conditions which lead to a stiffness matrix which is an M-matrix, as well as relations to finite volume methods we refer to [8]. More importantly, the work [8] provides techniques for consistent modification of the local stiffness matrices, leading to solvable linear systems for wide range of meshes. In 2 dimensions, a sufficient condition for the stiffness matrix to be an M-matrix is that the triangulation is a Delaunay triangulation which is easily achieved by any standard mesh generator. For spatial dimensions greater than 2, meshes satisfying the condition given in [7, Lemma 2.1] yield discretization with M-stiffness matrix. If this condition is violated by the mesh, then the techniques proposed in [8] can be used to modify consistently the local stiffness matrices so that the resulting global stiffness matrix is an M-matrix.

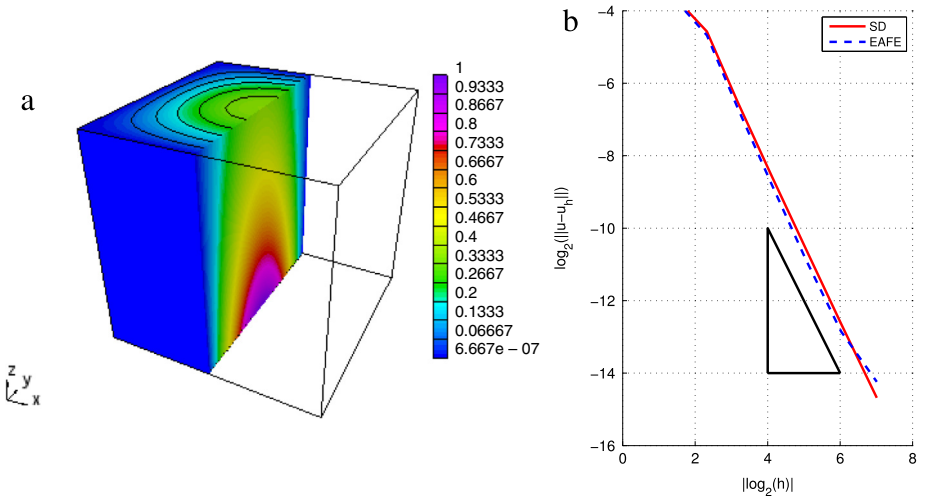

Fig. 1. (a) Trace of the solution on the plane $x = \frac{1}{2}$. (b) Error reduction in L^2-norm. Quadratic convergence is clearly observed.

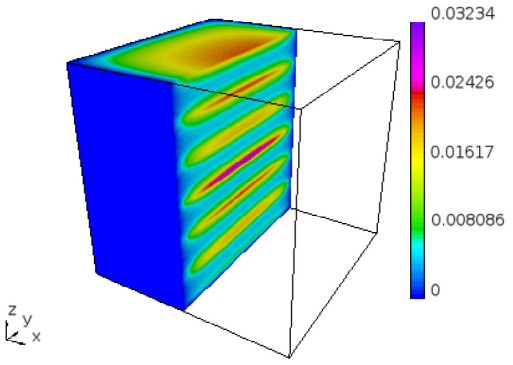

Fig. 2. Trace of the numerical solution of Eq. (43) on the plane $x = \frac{1}{2}$. The effect of the time dependent convection is clearly seen in the plot.

6. Numerical tests

We consider the 2D heat equation with Dirichlet boundary conditions on the unit square $(0, 1) \times (0, 1)$. We test both schemes: StreamLineDIffusion (SLDI) and EAFE on a uniform triangulation of the unit square. The exact solution is

$$U(x, t) = e^{-t} \sin \pi x \sin \pi y.$$

The domain is the unit square in 2D, and the space–time problem is solved as fully coupled 3D convection–diffusion problem.

We have tested the lowest order streamline diffusion scheme which has the same number of degrees of freedom as the EAFE scheme. The convergence behavior of both discretizations is shown in Fig. 1(b).

We have tested the convergence on a family of successively refined triangulations. The coarsest one has a mesh size $h_0 \approx \frac{1}{2}$ and the finest 2^{-8} in 3D. The parameter ε in the diffusion tensor D_ε for the EAFE scheme was 10^{-5} on all grids. The parameter θ in the streamline diffusion method was set to 10^{-2} on all grids. Such pool of tests corresponds to mesh with 27 vertices on the coarsest grid, and, $\approx 2.1 \times 10^6$ vertices on the finest grid. In Fig. 1(a) we have plotted the trace of the approximate solution on the plane $x = 0$. The approximate solution obtained via the EAFE scheme looks exactly the same, as is also the exact solution.

We next show a plot of a solution to an equation with convection depending on time. The equation is

$$u_t - \operatorname{div}(K(x)\nabla u - \mathbf{b}u) = 1, \quad x \in \Omega_s, \ \mathbf{b} = \begin{pmatrix} 100 \sin(6\pi t) \\ 0 \end{pmatrix} \tag{43}$$

and the boundary and initial conditions are homogeneous, the domain is the unit square and the time interval is $(0, 1)$. The solution via the Scharfetter–Gummel (EAFE) scheme is shown in Fig. 2. Note that with such convection term, the convection is 0 for $t = k/6$ and k integer; it is, however, convection dominated for other values of t.

We have mentioned already the software used in performing the tests. In summary, we have used the C++ library and examples from the `mfem` package [22] (discretization); The solutions of the resulting linear systems are done using the

Algebraic Multigraph Multilevel ILU algorithm by Bank and Smith [23,24] found at http://ccom.ucsd.edu/~reb/software.html. The visualization was done using the `glvis` tool [25].

7. Concluding remarks

We introduced a class of numerical methods for convection–diffusion equations in arbitrary spatial dimensions. In principle, these schemes can be applied to wide range of problems, such as linearization of the Nernst–Planck equations for transport of species in a charged media and the space–time discretizations of such equations. We have derived novel exponentially fitted (higher order Scharfetter–Gummel and streamline diffusion) discretizations for convection–diffusion equations. Distinctive features of the proposed Scharfetter–Gummel discretization are: (1) the monotonicity in the lowest order case; (2) its applicability in any spatial dimension; and (3) the a priori estimates are in terms of the flux only. For order higher than 1, the derivation and the analysis of this scheme is new, and its implementation for space–time formulation of parabolic problems is a subject of a current research.

References

[1] I.V. Lashuk, P.S. Vassilevski, The Construction of coarse de Rham complexes with improved approximation properties, Comput. Methods Appl. Math. 14 (2) (2014) 257–303. URL: http://www.degruyter.com/view/j/cmam.2014.14.issue-2/cmam-2014-0004/cmam-2014-0004.xml.
[2] M. Griebel, D. Oeltz, P. Vassilevski, Space–time approximation with sparse grids, SIAM J. Sci. Comput. 28 (2006) 701–727.
[3] U. Langer, S. Moore, M. Neumüller, Space–time isogeometric analysis of parabolic evolution equations, RICAM Report 19, Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, 2015, Also available at: http://arxiv.org/abs/1509.02008.
[4] A.N. Brooks, T.J.R. Hughes, Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Engrg. (ISSN: 0045-7825) 32 (1–3) (1982) 199–259. URL: http://dx.doi.org/10.1016/0045-7825(82)90071-8, fENOMECH'81, Part I (Stuttgart, 1981).
[5] C. Johnson, Numerical Solution of Partial Differential Equations by the Finite Element Method, Cambridge University Press, Cambridge, ISBN: 0-521-34514-6, 1987, 0-521-34758-0.
[6] O. Axelsson, V. Eijkhout, B. Polman, P. Vassilevski, Incomplete block-matrix factorization iterative methods for convection–diffusion problems, BIT (ISSN: 0006-3835) 29 (4) (1989) 867–889. URL: http://dx.doi.org/10.1007/BF01932750.
[7] J. Xu, L. Zikatanov, A monotone finite element scheme for convection–diffusion equations, Math. Comp. 68 (228) (1999) 1429–1446.
[8] R.E. Bank, W. Coughran Jr., L.C. Cowsar, The finite volume Scharfetter-Gummel method for steady convection diffusion equations, Comput. Vis. Sci. (ISSN: 1432-9360) 1 (3) (1998) 123–136. URL: http://dx.doi.org/10.1007/s007910050012.
[9] R.D. Lazarov, L.T. Zikatanov, An exponential fitting scheme for general convection–diffusion equations on tetrahedral meshes, Comput. Appl. Math. 92 (1) (2005) 60–69. (Obchysljuval'na ta prykladna matematyka, Kiev). URL: http://isc.tamu.edu/publications-reports/#2004 also available as Tech. Report 04-15, Institute for Scientific Computations, Texas A&M University, 2004.
[10] D. Scharfetter, H. Gummel, Large-signal analysis of a silicon read diod oscilator, IEEE Trans. Electron Devices ED-16 (205) (1969) 959–962.
[11] F. Brezzi, L.D. Marini, P. Pietra, Two-dimensional exponential fitting and applications to drift-diffusion models, SIAM J. Numer. Anal. (ISSN: 0036-1429) 26 (6) (1989) 1342–1355. URL: http://dx.doi.org/10.1137/0726078.
[12] S. Harald Christiansen, T.G. Halvorsen, Second order gauge invariant discretizations to the Schrödinger and Pauli equations, ArXiv e-prints.
[13] J.-C. Nédélec, Mixed finite elements in $\mathbf{R}^3$, Numer. Math. (ISSN: 0029-599X) 35 (3) (1980) 315–341. URL: http://dx.doi.org/10.1007/BF01396415.
[14] J.-C. Nédélec, A new family of mixed finite elements in $\mathbf{R}^3$, Numer. Math. (ISSN: 0029-599X) 50 (1) (1986) 57–81. URL: http://dx.doi.org/10.1007/BF01389668.
[15] A. Bossavit, A rationale for edge elements in 3D field computations, IEEE Trans. Magn. 24 (1988) 74–79.
[16] P. Monk, Finite Element Methods for Maxwell's Equations, in: Numerical Mathematics and Scientific Computation, Oxford University Press, New York, ISBN: 0-19-850888-3, 2003, URL: http://dx.doi.org/10.1093/acprof:oso/9780198508885.001.0001.
[17] R. Hiptmair, Canonical construction of finite elements, Math. Comp. (ISSN: 0025-5718) 68 (228) (1999) 1325–1346. URL: http://dx.doi.org/10.1090/S0025-5718-99-01166-7.
[18] D.N. Arnold, R.S. Falk, R. Winther, Finite element exterior calculus, homological techniques, and applications, Acta Numer. (ISSN: 0962-4929) 15 (2006) 1–155. URL: http://dx.doi.org/10.1017/S0962492906210018.
[19] D.N. Arnold, R.S. Falk, R. Winther, Finite element exterior calculus: from Hodge theory to numerical stability, Bull. Amer. Math. Soc. (NS) (ISSN: 0273-0979) 47 (2) (2010) 281–354. URL: http://dx.doi.org/10.1090/S0273-0979-10-01278-4.
[20] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, in: Classics in Applied Mathematics, vol. 40, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, ISBN: 0-89871-514-8, 2002, URL: http://dx.doi.org/10.1137/1.9780898719208 reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].
[21] R.E. Bank, P.S. Vassilevski, L.T. Zikatanov, Arbitrary dimension convection–diffusion schemes for space–time discretizations, ArXiv e-prints. URL: http://adsabs.harvard.edu/abs/2016arXiv160106351B arxiv report arXiv:1601.06351 (math.NA).
[22] Software, MFEM: Modular finite element methods, 2015. http://mfem.org.
[23] R.E. Bank, R.K. Smith, An algebraic multilevel multigraph algorithm, SIAM J. Sci. Comput. (ISSN: 1064-8275) 23 (5) (2002) 1572–1592 (electronic). URL: http://dx.doi.org/10.1137/S1064827500381045.
[24] R.E. Bank, Multigraph 1.0: An algebraic multilevel multigraph algorithm, 2015. http://ccom.ucsd.edu/~reb/software.html.
[25] Software, GLVis: Accurate finite element visualization, 2015. http://glvis.org.