

STABLE AND EFFICIENT PETROV-GALERKIN METHODS FOR A KINETIC FOKKER-PLANCK EQUATION

JULIA BRUNKEN AND KATHRIN SMETANA

ABSTRACT. We propose a stable Petrov-Galerkin discretization of a kinetic Fokker-Planck equation constructed in such a way that uniform inf-sup stability can be inferred directly from the variational formulation. Inspired by well-posedness results for parabolic equations, we derive a lower bound for the dual inf-sup constant of the Fokker-Planck bilinear form by means of stable pairs of trial and test functions. The trial function of such a pair is constructed by applying the kinetic transport operator and the inverse velocity Laplace-Beltrami operator to a given test function. For the Petrov-Galerkin projection we choose an arbitrary discrete test space and then define the discrete trial space using the same application of transport and inverse Laplace-Beltrami operator. As a result, the spaces replicate the stable pairs of the continuous level and we obtain a well-posed numerical method with a discrete inf-sup constant identical to the inf-sup constant of the continuous problem independently of the mesh size. We show how the specific basis functions can be efficiently computed by low-dimensional elliptic problems, and confirm the practicability and performance of the method with numerical experiments.

1. INTRODUCTION

In this manuscript we develop a stable and efficient Petrov-Galerkin approximation scheme for certain kinetic Fokker-Planck equations, including the equation

$$(1) \quad \partial_t u((t, x), v) + v \cdot \nabla_x u((t, x), v) = \Delta_v \left(\frac{u((t, x), v)}{q(x, v)} \right) \quad \text{in } \Omega = I_t \times \Omega_x \times \Omega_v$$

with suitable inflow boundary conditions. Equation (1) describes a particle density u dependent on time $t \in I_t$, position $x \in \Omega_x \subset \mathbb{R}^d$, $d \in \{2, 3\}$, and direction $v \in \Omega_v = S^{d-1}$, where S^{d-1} is the $(d-1)$ -dimensional unit sphere and $q \in L^\infty(\Omega_x \times \Omega_v)$ with $q > 0$ a.e. and $q(x, \cdot) \in C^1(\Omega_v)$ for a.e. $x \in \Omega_x$.

Formulations for particle densities governed by kinetic equations arise in various contexts. Beyond the classical applications of radiative transfer and kinetic gas theory (see e.g. [17, 20]), kinetic equations are, for instance, also used to describe densities of tumor cells in multiscale descriptions of tumor spreading [25, 34]. In this manuscript, we are mainly interested in the latter application. More precisely, we focus on a discretization of a prototype of a glioma tumor equation described in [34], where the velocity is driven by a Brownian motion resulting in the specific

Date: April 13, 2021.

2010 Mathematics Subject Classification. 65N30, 65M12, 65J10.

Key words and phrases. Kinetic Fokker-Planck equation, Petrov-Galerkin method, well-posedness, inf-sup stability.

The work of Julia Brunken was supported by the German Federal Ministry of Education and Research under grant BMBF 05M2016 - GlioMaTh and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044 –390685587, Mathematics Münster: Dynamics–Geometry–Structure.

Laplace-Beltrami term of (1). However, other variants including, e.g., ∇_v terms, are also included in the more general setting considered in the course of this work.

We aim for a finite element discretization with guaranteed stability. Therefore, we focus on a Petrov-Galerkin discretization based on a stable variational formulation of (1), since in such a framework the well-posedness of the discrete scheme can often be inferred from respective results on the continuous level, see e.g. [15, 19, 41, 44].

First, we establish a full-dimensional variational formulation for (1) based on Bochner-type spaces, mapping the combined space-time domain $\Omega_{t,x} = I_t \times \Omega_x$ to a Sobolev space defined on the velocity domain Ω_v similar to spaces defined in [1, 11]. Taking the viewpoint that the Fokker-Planck equation can be interpreted as a “generalization” of a parabolic equation with a $(d+1)$ -dimensional kinetic transport operator $\partial_t + v \cdot \nabla_x$ instead of a one-dimensional time derivative ∂_t , we analyze the well-posedness of the variational formulation for (1) by combining respective approaches developed for parabolic equations [26, 41, 44] and for transport equations [10, 15, 19]. We show existence of a weak solution by verifying the dual inf-sup condition. To that end, similarly to [26, 41], specific function pairs in the trial and test spaces are constructed. We associate a test space function p to a trial space function roughly defined as $w_p = p - (\Delta_v)^{-1}(\partial_t p + v \cdot \nabla_x p)$. Then the bilinear form evaluated in w_p and p can be bounded from below by the respective norms of w_p and p , which leads to a lower bound for the dual inf-sup constant. This approach is a generalization of proofs for parabolic equations using a variant of w_p containing only the time derivative instead of the kinetic transport operator [26, 41] and of proofs for transport equations, where a “stable function pair” consists roughly of $-(\partial_t p + v \cdot \nabla_x p)$ and p , when choosing the kinetic transport operator in the linear transport equation, see [10, 15, 19]. Under an additional assumption on the global traces of certain functions, we also show uniqueness of the solution similar to proofs for parabolic equations [26] and transport equations [4], and have a stability estimate dependent on the inf-sup constant, which is similar to the respective estimates for parabolic equations.

To design the Petrov-Galerkin discretization, we use problem-specific trial spaces ensuring stability: We first choose an arbitrary discrete test space \mathcal{Y}_δ and then define the discrete trial space roughly as $\mathcal{X}_\delta = \mathcal{Y}_\delta + (\Delta_v)^{-1}(\partial_t + v \cdot \nabla_x)\mathcal{Y}_\delta$. The spaces thus consist of pairs w_p^δ, p^δ that are the discrete counterparts of the pairs w_p, p used in the proof for the lower bound of the dual inf-sup constant. This approach automatically yields a well-posed discrete problem with the same stability constant as for the continuous problem independently of the choice of the test space and thus of the mesh size. The strategy to use an application of the transport operator for defining a stable trial space was already used for linear first-order transport equations [10] and for the wave equation [31] as an alternative to computing stable test spaces by approximately inverting the transport operator [15, 19]. Our choice ensures that the spaces can be efficiently computed in the course of the numerical scheme, where we apply the high-dimensional transport operator and only solve low-dimensional elliptic problems in the velocity domain due to the inverse Laplace-Beltrami operator. As a result, we can guarantee the stability of the method with low-dimensional computations that are not dominant in the computational costs of the full solution process.

Weak solutions and variational formulations for different types of kinetic Fokker-Planck equations have been defined and analyzed in various works, see e.g. [1, 11, 18, 35, 43]. However, these approaches focus on the properties of the weak solution without an orientation towards a subsequent discretization. On the other hand, discretizations of kinetic Fokker-Planck equations are often not based on the direct connection to a weak solution or do not specifically consider stability estimates. In [37], a finite element discretization of a kinetic Fokker-Planck equation is described, where the well-posedness of the discrete problem is however not analyzed. Applying the framework of [22], a mixed variational formulation with a subsequent discretization for a generalized Fokker-Planck equation is proposed in [29]. In the context of neuronal networks, a Fokker-Planck equation is discretized with finite differences in [14]. Another well-established approach to discretize kinetic equations is the method of moments, applied to Fokker-Planck equations, for instance, in [27, 40], while a related approach in the context of hierarchical model reduction is proposed in [9]. For the related Vlasov-Fokker-Planck system there are, for instance, works based on finite differences [39, 46] and streamline-diffusion discontinuous Galerkin approximations [2, 3]. For the more general class of equations with nonnegative characteristic form, discontinuous Galerkin methods [32, 33] and also sparse tensor approximations [42] have been developed.

This paper is structured as follows. After a more detailed description of the considered Fokker-Planck equation in section 2, we introduce the suitable Bochner-type function spaces and establish density and trace properties in section 3. We then derive the variational formulation and prove the existence and uniqueness results in section 4. In section 5, we introduce the discrete scheme, show well-posedness and describe an efficient computation. These properties of the proposed method are finally confirmed for a numerical example in section 6.

2. THE KINETIC FOKKER-PLANCK EQUATION

In this paper we consider a simplified version of the kinetic Fokker-Planck equation developed in [34, sect. 2.4.2] that gives a mesoscopic description of the density of glioma tumor cells. Let $\Omega_x \subset \mathbb{R}^d$, $d \in \{2, 3\}$ be the spatial domain¹ with piecewise C^1 boundary that is globally Lipschitz and let $I_t := (0, T)$ be the time interval. Moreover, let the velocity domain be the $(d - 1)$ -dimensional unit sphere $\Omega_v := S^{d-1}$, which corresponds to the assumption of particles with constant speed but varying direction. As we will often treat space and time variables simultaneously, we denote by $\Omega_{t,x} := I_t \times \Omega_x$ the space-time domain. The full domain is defined as $\Omega := \Omega_{t,x} \times \Omega_v$.

To prescribe suitable inflow boundary conditions, we first define relevant boundaries. First, we denote by

$$\hat{\Gamma} := \{0, T\} \times \bar{\Omega}_x \times \Omega_v \cup [0, T] \times \partial\Omega_x \times \Omega_v$$

the essential boundary of Ω . Then, we define the spatial out- and inflow domains $\Gamma_\pm^x(v) := \{x \in \partial\Omega_x : n(x) \cdot v \gtrless 0\} \subset \partial\Omega_x$, where $n(x)$ is the unit outer normal to $\partial\Omega_x$ at x . The full out- and inflow domains Γ_+ and Γ_- are then defined as

$$\Gamma_\pm := \{((t, x), v) \in \partial\Omega_{t,x} \times \Omega_v : (\frac{1}{v}) \cdot n(t, x) \gtrless 0\} \subset \hat{\Gamma},$$

¹One can also define a Fokker-Planck equation on a one-dimensional spatial domain, where the velocity has to be defined as a one-dimensional projection variable, see, e.g., [40]. We leave out this special case for ease of presentation.

where $n(t, x)$ is the unit outer normal to $\partial\Omega_{t,x}$ at (t, x) . The sets Γ_\pm thus contain both the temporal and the spatial boundaries, i.e., Γ_- contains the “initial boundary” and the (v -dependent) spatial inflow boundary whereas Γ_+ contains the final time boundary and the spatial outflow boundary.

The strong form of the Fokker-Planck equation then reads

$$(2) \quad \begin{aligned} \partial_t u((t, x), v) + v \cdot \nabla_x u((t, x), v) &= \Delta_v \left(\frac{u((t, x), v)}{q(x, v)} \right) && \text{in } \Omega, \\ u((t, x), v) &= g((t, x), v) && \text{on } \Gamma_-, \end{aligned}$$

where Δ_v is the Laplace-Beltrami operator on the unit sphere $\Omega_v = S^{d-1}$, $q \in L^\infty(\Omega_x \times \Omega_v)$ is the so-called “tissue fiber orientation distribution” satisfying $q(x, \cdot) \in C^1(\Omega_v)$ for a.e. $x \in \Omega_x$ and $q \geq \alpha_q > 0$ a.e. in $\Omega_x \times \Omega_v$ and $g : \Gamma_- \rightarrow \mathbb{R}$ is the inflow boundary condition that contains the initial condition $g|_{\{t=0\}}$ as well as the spatial inflow boundary condition $g|_{\Gamma_-^x(v)}$, $v \in \Omega_v$. Since q is assumed to be sufficiently regular, we can bring the respective differential operator in (2) in divergence form.¹

In section 4, we develop a variational formulation for this equation, where we allow for a more general differential operator on Ω_v and give specific conditions on q and g leading to well-posedness.

3. FUNCTION SPACES

To develop a variational formulation for (2) we first introduce the necessary function spaces. Since we aim for a full space-time-velocity formulation, we use Bochner spaces mapping the space-time domain $\Omega_{t,x}$ to a space of functions on Ω_v .

We start with the function space for the velocity variable: Since the equation contains a Laplace-Beltrami operator on the velocity domain $\Omega_v = S^{d-1}$, we define $V := H^1(\Omega_v) \subset L^2(\Omega_v)$ as the Sobolev space of weakly differentiable functions on the surface $\Omega_v = S^{d-1}$ with squared norm $\|\phi\|_V^2 = \|\phi\|_{L^2(\Omega_v)}^2 + \|\nabla_v \phi\|_{L^2(\Omega_v)}^2$. For details on the definition of Sobolev spaces on manifolds, see [21, 30]. We denote the dual space of V by $V' := H^{-1}(\Omega_v)$. The space V is a dense subspace of $L^2(\Omega_v)$ and we will make use of the Gelfand triple $V \hookrightarrow L^2(\Omega_v) \hookrightarrow V'$, where we denote the dual pairing by $\langle \cdot, \cdot \rangle_{V', V}$.

As a function space for the full domain, we will use the space $L^2(\Omega_{t,x}; V)$ with squared norm

$$(3) \quad \|w\|_{L^2(\Omega_{t,x}; V)}^2 = \int_{\Omega_{t,x}} \|w(t, x)\|_V^2 d(t, x).$$

From now on, we will denote the kinetic advection field $(\frac{1}{v})$ by $k \in C^1(\bar{\Omega}, \mathbb{R}^{d+1})$, $k((t, x), v) := (\frac{1}{v})$, so that the kinetic space-time transport operator is given as $k \cdot \nabla_{t,x} p = \partial_t p + v \cdot \nabla_x p$. We then define

$$(4) \quad H_{\text{FP}}^1(\Omega) := \{p \in L^2(\Omega_{t,x}; V) : k \cdot \nabla_{t,x} p \in L^2(\Omega_{t,x}; V')\},$$

with squared norm

$$(5) \quad \|p\|_{H_{\text{FP}}^1(\Omega)}^2 := \|p\|_{L^2(\Omega_{t,x}; V)}^2 + \|k \cdot \nabla_{t,x} p\|_{L^2(\Omega_{t,x}; V')}^2.$$

This definition is similar to the spaces used for other variants of the kinetic Fokker-Planck equation, e.g., in [1, 5, 11]. We use ideas from [1] to show the following:

Proposition 3.1. *The set $C^\infty(\bar{\Omega}_{t,x} \times \Omega_v)$ is dense in $H_{\text{FP}}^1(\Omega)$.*

Proof. For the proof one constructs approximations of a function $f \in H_{\text{FP}}^1(\Omega)$ by a mollification in $\Omega_{t,x}$ analogously to [1, Prop. 7.1] and a suitable basis expansion in Ω_v . For more details see the supplementary material. \square

To discuss the boundary behavior of functions in $H_{\text{FP}}^1(\Omega)$, we introduce weighted L^2 -spaces, as usually used for transport and kinetic equations (e.g. [6, 12], [16, XXI, §2]) and for different versions of the kinetic Fokker-Planck equation [1, 11]. For any $\Gamma \subseteq \hat{\Gamma}$ we introduce $L^2(\Gamma, |k \cdot n|)$ with squared norm

$$(6) \quad \|w\|_{L^2(\Gamma, |k \cdot n|)}^2 := \int_{\Gamma} w^2 |k \cdot n| \, ds.$$

Then, we can show that functions in $H_{\text{FP}}^1(\Omega)$ admit local traces on $\Gamma_+ \cup \Gamma_-$:

Proposition 3.2. *For every compact set $K \subset \Gamma_+$ (resp. $K \subset \Gamma_-$), the trace operator $w \mapsto w|_K$ from $C^\infty(\bar{\Omega})$ to $L^2(K, |k \cdot n|)$ extends to a continuous linear operator on $H_{\text{FP}}^1(\Omega)$.*

For the proof we need to estimate the product of $H_{\text{FP}}^1(\Omega)$ functions with different test functions in the following way, where the proof can be found in Appendix A.

Lemma 3.3. *Let $\phi \in C^1(\bar{\Omega})$. Then, the mapping $f \mapsto \phi f$ is continuous in $H_{\text{FP}}^1(\Omega)$ with the estimate*

$$\|\phi f\|_{H_{\text{FP}}^1(\Omega)} \leq C \|\phi\|_{C^1(\Omega)} \|f\|_{H_{\text{FP}}^1(\Omega)}.$$

Proof of Proposition 3.2. We use ideas of the proof of a similar result for transport equations, e.g., in [16, Chap. XXI, Thm. 1, p. 220]. Analogous results for spaces similar to $H_{\text{FP}}^1(\Omega)$ are also given in [1, Proofs of Lemmas 4.3, 7.6].

Given a compact set $K \subset \Gamma_+$, let $\eta_K \in C^1(\bar{\Omega})$ with $\eta_K = 1$ on K and $\text{supp } \eta_K \cap \Gamma_- = \emptyset$. We then obtain by integrating by parts for $w \in C^\infty(\bar{\Omega})$

$$\begin{aligned} \int_K w^2 |k \cdot n| \, ds &= \int_K (\eta_K w)^2 |k \cdot n| \, ds \leq \int_{\hat{\Gamma}} (\eta_K w)^2 |k \cdot n| \, ds \\ &\stackrel{(*)}{=} \int_{\hat{\Gamma}} (\eta_K w)^2 k \cdot n \, ds = 2 \int_{\Omega} \eta_K w k \cdot \nabla_{t,x}(\eta_K w) \, d((t, x), v) \\ &\leq 2 \|\eta_K w\|_{L^2(\Omega_{t,x}, V)} \|k \cdot \nabla_{t,x}(\eta_K w)\|_{L^2(\Omega_{t,x}, V')} \\ &\leq 2 \|\eta_K w\|_{H_{\text{FP}}^1(\Omega)}^2 \stackrel{\text{Lemma 3.3}}{\leq} C \|\eta_K\|_{C^1(\Omega)}^2 \|w\|_{H_{\text{FP}}^1(\Omega)}^2. \end{aligned}$$

We thus have continuity of the mapping $w \mapsto w|_K$ for all $w \in C^\infty(\bar{\Omega})$, and by density (Proposition 3.1) the mapping extends to a continuous operator $H_{\text{FP}}^1(\Omega) \rightarrow L^2(K, |k \cdot n|)$. For $K \subset \Gamma_-$ the claim can be shown analogously using $|k \cdot n| = -k \cdot n$ on $\text{supp } \eta_K$ in (*). \square

This result ensures that $H_{\text{FP}}^1(\Omega)$ functions have a trace on the non-characteristic boundary² $\Gamma_+ \cup \Gamma_-$. However, from the local existence of traces we cannot directly deduce that these generally lie in global trace spaces as e.g. $L^2(\partial\Omega, |k \cdot n|)$.

We now define

$$(7) \quad H_{\text{FP}, \Gamma_{\pm}}^1(\Omega) := \text{clos}_{\|\cdot\|_{H_{\text{FP}}^1(\Omega)}} \{f \in C^\infty(\bar{\Omega}) : f \equiv 0 \text{ on } \Gamma_{\pm}\}.$$

To avoid boundary integrals on the outflow domain in the variational formulation, we will use $H_{\text{FP}, \Gamma_+}^1(\Omega)$ as the test space for our variational formulation. With the

²The non-characteristic boundary is the part of the boundary where $|k \cdot n| \neq 0$.

restriction of functions in $H_{\text{FP},\Gamma_+}^1(\Omega)$ on the outflow boundary and the definition through the closure, we can show that these functions have a trace in $L^2(\Gamma_-, |k \cdot n|)$:

Proposition 3.4. *There exists a linear continuous mapping $\gamma_- : H_{\text{FP},\Gamma_+}^1(\Omega) \rightarrow L^2(\Gamma_-, |k \cdot n|)$ such that*

$$\|\gamma_-(w)\|_{L^2(\Gamma_-, |k \cdot n|)} \leq C \|w\|_{H_{\text{FP}}^1(\Omega)} \quad \forall w \in H_{\text{FP},\Gamma_+}^1(\Omega).$$

Furthermore, the integration by parts formula

$$\int_{\Omega_{t,x}} \langle k \cdot \nabla_{t,x} w, w \rangle_{V,V} d(t,x) = \frac{1}{2} \int_{\Gamma_-} w^2 k \cdot n ds$$

holds for all $w \in H_{\text{FP},\Gamma_+}^1(\Omega)$.

Proof. The proof is similar to the respective result for transport equations e.g. in [10, Prop. 2.4], see also [1, sect. 4]. Let $w \in C^\infty(\bar{\Omega})$ with $w \equiv 0$ on Γ_+ . Performing integration by parts we obtain

$$\int_{\Omega} wk \cdot \nabla_{t,x} w d((t,x), v) = - \int_{\Omega} \nabla_{t,x} w \cdot kw d((t,x), v) + \int_{\Gamma_-} w^2 \underbrace{k \cdot n}_{<0} ds,$$

and thus

$$\begin{aligned} \|w\|_{L^2(\Gamma_-, |k \cdot n|)}^2 &= \int_{\Gamma_-} w^2 |k \cdot n| ds = 2 \int_{\Omega} (-k \cdot \nabla_{t,x} w) w d((t,x), v) \\ &\leq 2 \| -k \cdot \nabla_{t,x} w \|_{L^2(\Omega_{t,x}; V')} \|w\|_{L^2(\Omega_{t,x}; V)} \leq 2 \|w\|_{H_{\text{FP}}^1(\Omega)}^2. \end{aligned}$$

By density (due to the definition of $H_{\text{FP},\Gamma_+}^1(\Omega)$), the integration by parts formula and the bound for $\|w\|_{L^2(\Gamma_-, |k \cdot n|)}$ hold for all $w \in H_{\text{FP},\Gamma_+}^1(\Omega)$. \square

Remark 3.5. *Similarly, it can be shown that the space $H_{\text{FP},\Gamma_-}^1(\Omega)$ admits a continuous trace operator $\gamma_+ : H_{\text{FP},\Gamma_-}^1(\Omega) \rightarrow L^2(\Gamma_+, |k \cdot n|)$.*

To later show the uniqueness of the weak solution in section 4, we also need to verify the existence of a global trace and the integration by parts formula for certain functions in $H_{\text{FP}}^1(\Omega)$ with vanishing trace on Γ_- , but not necessarily in $H_{\text{FP},\Gamma_-}^1(\Omega)$. This is established for spaces where the advective or kinetic terms lie in $L^2(\Omega)$ (see, e.g., [6, Thm. 2.2, Prop. 2.5]), [16, Chap. XXI, Remark 3]). Similar or even stronger results for respective functions in $H_{\text{FP}}^1(\Omega)$ are claimed to be proven in [1, 5, 11], however, we believe the arguments to be incomplete, for more details see the supplementary material.

Since we were not able to prove the existence of a global trace for $H_{\text{FP}}^1(\Omega)$ functions with vanishing trace on the inflow or the outflow boundary, we will formulate the exact result needed for uniqueness of the weak solution as an assumption in section 4.

4. VARIATIONAL FORMULATION

In this section, we develop a variational formulation for (2) and show its well-posedness.

Let $a : \Omega_{t,x} \times V \times V \rightarrow \mathbb{R}$ be a potentially (x,t) -dependent bilinear form defined on the velocity space V . Moreover, let a satisfy the following assumptions:

- (8) the map $(t,x) \mapsto a((t,x); \phi, \psi)$ is measurable on $\Omega_{t,x}$ for all $\phi, \psi \in V$,
- (9) $a((t,x); \cdot, \cdot)$ is bilinear for a.e. $(t,x) \in \Omega_{t,x}$,
- (10) $a((t,x); \phi, \psi) \leq c_a \|\phi\|_V \|\psi\|_V$ with $c_a < \infty$ for all $\phi, \psi \in V$, a.e. $(x,t) \in \Omega_{t,x}$,
- (11) $a((t,x); \phi, \phi) + \lambda_a \|\phi\|_{L^2(\Omega_v)}^2 \geq \alpha_a \|\phi\|_V^2$ with $\lambda_a \in \mathbb{R}, \alpha_a > 0$
for all $\phi \in V$, a.e. $(x,t) \in \Omega_{t,x}$.

Note that c_a, λ_a , and α_a are assumed to be independent of (x,t) .

Example 4.1. For the strong form of the Fokker-Planck equation (2), a is given for all $\phi, \psi \in V$, a.e. $x \in \Omega_x$ by

$$\begin{aligned} a(x; \phi, \psi) &= (\nabla_v (q(x,v)^{-1} \phi(v)), \nabla_v \psi(v))_{L^2(\Omega_v)} \\ &= (q(x,v)^{-1} \nabla_v \phi(v), \nabla_v \psi(v))_{L^2(\Omega_v)} + (\nabla_v q(x,v)^{-1} \phi(v), \nabla_v \psi(v))_{L^2(\Omega_v)}, \end{aligned}$$

where ∇_v is the tangential gradient on Ω_v , see, e.g., [21] for a formal definition. If $q^{-1} \in L^\infty(\Omega_x \times \Omega_v)$ with $\nabla_v q^{-1} \in L^\infty(\Omega_x \times \Omega_v)$ and $q^{-1}(x,v) \geq l_q > 0$ for a.e. (x,v) , then a fulfills the conditions (8)–(11), for instance, with $c_a = \|q^{-1}\|_{L^\infty} + \|\nabla_v q^{-1}\|_{L^\infty}$, $\alpha_a = \frac{1}{2} l_q$, and $\lambda_a = \|\nabla_v q^{-1}\|_{L^\infty}^2 / (2l_q) + \frac{1}{2} l_q$. Depending on q , other estimates might be better, e.g., for $q = q(x)$ and thus $\nabla_v q = 0$ we can get $\alpha_a = \lambda_a = l_q$.

Recalling the function spaces introduced in (3) and (7), we define the space-time-velocity trial and test spaces as

$$(12) \quad \mathcal{X} := L^2(\Omega_{t,x}, V), \quad \mathcal{Y} := H_{\text{FP}, \Gamma_+}^1(\Omega).$$

with squared norms (cf. (3), (5))

$$(13) \quad \|w\|_{\mathcal{X}}^2 = \int_{\Omega_{t,x}} \|w(t,x)\|_V^2 \, d(t,x),$$

$$(14) \quad \|p\|_{\mathcal{Y}}^2 = \|p\|_{\mathcal{X}}^2 + \|k \cdot \nabla_{t,x} p\|_{\mathcal{X}'}^2.$$

We then define the full bilinear form $b : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ for $w \in \mathcal{X}, p \in \mathcal{Y}$ by

$$(15) \quad b(w, p) := \int_{\Omega_{t,x}} \langle w(t,x), -k(t,x) \cdot \nabla_{t,x} p(t,x) \rangle_{V,V'} + a((t,x); w(t,x), p(t,x)) \, d(t,x).$$

The functional $f : \mathcal{Y} \rightarrow \mathbb{R}$ containing the boundary condition $g \in L^2(\Gamma_-, |k \cdot n|)$ is given as

$$f(p) := \int_{\Gamma_-} g p |k \cdot n| \, d((t,x), v) \quad \forall p \in \mathcal{Y},$$

which is well-defined due to Proposition 3.4, and we thus have $f \in \mathcal{Y}'$.

We call $u \in \mathcal{X}$ a weak solution of (2), if

$$(16) \quad b(u, p) = f(p) \quad \forall p \in \mathcal{Y}.$$

In the following, we examine the well-posedness of the variational formulation, using the Banach-Nečas-Babuška (or inf-sup) Theorem (see e.g. [26, Thm. 2.6]). We first prove existence of a weak solution in subsection 4.1. Then, in subsection 4.2 we

also show uniqueness of the weak solution under an additional assumption on the trace of certain $H_{\text{FP}}^1(\Omega)$ -functions.

4.1. Existence of a weak solution. We show the existence of a weak solution u to (16) by verifying a dual inf-sup condition. To that end, we construct stable pairs of trial and test space functions such that the application of the bilinear form to the function pairs can be estimated from below by the respective norms of the functions. In these pairs, the trial space functions are derived from the test space functions by the application of the kinetic transport operator and the inverse elliptic velocity operator. We thus generalize similar proofs for parabolic equations [26, 41], where a time derivative was used instead of the kinetic transport operator, and for transport equations, where only an application of the transport operator was used [10, 15, 19].

Theorem 4.2. *The bilinear form b satisfies the dual inf-sup condition*

$$\inf_{\substack{p \in \mathcal{Y} \\ p \neq 0}} \sup_{\substack{w \in \mathcal{X} \\ w \neq 0}} \frac{b(w, p)}{\|w\|_{\mathcal{X}} \|p\|_{\mathcal{Y}}} \geq \beta$$

with an inf-sup constant

$$(17) \quad \beta \geq \frac{\alpha_a}{\sqrt{2} \max\{1, c_a\}}, \quad \text{if } \lambda_a \leq 0,$$

$$(18) \quad \beta \geq \frac{\alpha_a}{\sqrt{2} \max\{1, c_a + \lambda_a\}} \frac{e^{-\lambda_a T}}{\sqrt{\max\{1 + 2\lambda_a^2, 2\}}}, \quad \text{if } \lambda_a > 0.$$

Consequently, the variational formulation (16) has at least one weak solution $u \in \mathcal{X}$.

Remark 4.3. *The estimates for β are not worse than estimates for space-time variational formulations for parabolic equations from [41]. In fact, for $\lambda_a \leq 0$ and assuming $\alpha_a \leq 1$ and $c_a \geq 1$, the estimate in [41, (A.6)] roughly translates³ to $\beta_{\text{parab}} \geq \alpha_a^2 / (\sqrt{2} c_a^2)$, while we have $\beta \geq \alpha_a / \sqrt{2} c_a$. The exponential dependence on the final time T for the non-coercive case is the same for both types of equations.*

Proof of Theorem 4.2. We start with the case of a being coercive, i.e., $\lambda_a \leq 0$; the non-coercive case will be treated afterwards via a temporal transformation.

To show the inf-sup condition we combine ideas from well-posedness results for parabolic equations as e.g. in [26, 41] and for transport equations as, e.g., in [10]. To that end, we take $0 \neq p \in \mathcal{Y}$ arbitrary, but fixed. We want to construct a suitable $w_p \in \mathcal{X}$ and show $b(w_p, p) \geq \beta \|w_p\|_{\mathcal{X}} \|p\|_{\mathcal{Y}}$ for a constant β independent of p , which makes β a lower bound for the inf-sup constant.

Since $p \in \mathcal{Y}$, we have $r_p := -k \cdot \nabla_{t,x} p \in L^2(\Omega_{t,x}; V') = \mathcal{X}'$. Similar to [38, pp. 235], we define the bilinear form $m : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ by

$$m(w_1, w_2) := \int_{\Omega_{t,x}} a((t, x); w_1(t, x), w_2(t, x)) \, d(t, x), \quad \forall w_1, w_2 \in \mathcal{X}.$$

Since the function $(t, x) \mapsto a((t, x); \phi, \psi)$ is assumed to be measurable for all $\phi, \psi \in V$ (see (8)) and $a((t, x), \cdot, \cdot)$ is continuous and coercive with constants c_a, α_a independent of (t, x) ((10) and (11) with $\lambda_a \leq 0$), m is well-defined, continuous,

³More precisely, using the notation of this paper, the complete estimate in [41, (A.6)] reads $\beta_{\text{parab}} \geq \min(\alpha_a / c_a^2, \alpha_a) / (2 \max(\alpha_a^{-2}, 1) + M_e^2)^{1/2}$, where M_e is an additional positive constant that appears due to a different boundary treatment and that we can leave out here.

and coercive over $\mathcal{X} \times \mathcal{X}$ with constants c_a and α_a . Therefore, by the Lax-Milgram theorem it exists a unique $z_p \in \mathcal{X}$ with

$$(19) \quad m(z_p, w) = \langle r_p, w \rangle_{\mathcal{X}', \mathcal{X}} \quad \forall w \in \mathcal{X}.$$

Due to the definitions of z_p , r_p , and m , there holds⁴

$$(20) \quad \int_{\Omega_{t,x}} a(z_p, w) d(t, x) = \int_{\Omega_{t,x}} \langle -k \cdot \nabla_{t,x} p, w \rangle_{V', V} d(t, x) \quad \forall w \in \mathcal{X}.$$

We now define $w_p := p + z_p \in \mathcal{X}$. To bound $b(w_p, p)$ from below we use (20) for $w = w_p$, and the integration by parts formula from Proposition 3.4:

$$\begin{aligned} b(w_p, p) &= \int_{\Omega_{t,x}} \langle p + z_p, -k \cdot \nabla_{t,x} p \rangle_{V, V'} + a(p + z_p, p) d(t, x) \\ &= \int_{\Omega_{t,x}} \langle p, -k \cdot \nabla_{t,x} p \rangle_{V, V'} + a(z_p, z_p) + a(p, p) + \langle -k \cdot \nabla_{t,x} p, p \rangle_{V', V} d(t, x) \\ (21) \quad &\geq \alpha_a (\|p\|_{\mathcal{X}}^2 + \|z_p\|_{\mathcal{X}}^2) + 2 \int_{\Omega_{t,x}} \langle -k \cdot \nabla_{t,x} p, p \rangle_{V', V} d(t, x) \\ &= \alpha_a (\|p\|_{\mathcal{X}}^2 + \|z_p\|_{\mathcal{X}}^2) + \int_{\Gamma_-} p^2 |k \cdot n| ds \geq \alpha_a (\|p\|_{\mathcal{X}}^2 + \|z_p\|_{\mathcal{X}}^2). \end{aligned}$$

Since we have $\langle r_p, w \rangle_{\mathcal{X}', \mathcal{X}} = m(z_p, w) \leq c_a \|z_p\|_{\mathcal{X}} \|w\|_{\mathcal{X}}$ for all $w \in \mathcal{X}$, it holds

$$(22) \quad \|r_p\|_{\mathcal{X}'} \leq c_a \|z_p\|_{\mathcal{X}}.$$

Using the definition of w_p , r_p , and the norm of \mathcal{Y} as defined in (5), we can then estimate

$$\begin{aligned} \|w_p\|_{\mathcal{X}} \|p\|_{\mathcal{Y}} &= \|p + z_p\|_{\mathcal{X}} (\|p\|_{\mathcal{X}}^2 + \|r_p\|_{\mathcal{X}'}^2)^{1/2} \\ (23) \quad &\stackrel{(22)}{\leq} [\|p + z_p\|_{\mathcal{X}}^2 (\|p\|_{\mathcal{X}}^2 + c_a^2 \|z_p\|_{\mathcal{X}}^2)]^{1/2} \\ &\leq [2 (\|p\|_{\mathcal{X}}^2 + \|z_p\|_{\mathcal{X}}^2) (\|p\|_{\mathcal{X}}^2 + c_a^2 \|z_p\|_{\mathcal{X}}^2)]^{1/2} \\ &\leq \sqrt{2} \max\{1, c_a\} (\|p\|_{\mathcal{X}}^2 + \|z_p\|_{\mathcal{X}}^2) \stackrel{(21)}{\leq} \frac{\sqrt{2} \max\{1, c_a\}}{\alpha_a} b(w_p, p). \end{aligned}$$

Since $p \in \mathcal{Y}$ was chosen arbitrarily, we thus have

$$(24) \quad \inf_{p \in \mathcal{Y}} \sup_{w \in \mathcal{X}} \frac{b(w, p)}{\|w\|_{\mathcal{X}} \|p\|_{\mathcal{Y}}} \geq \beta := \frac{\alpha_a}{\sqrt{2} \max\{1, c_a\}},$$

i.e., the claim for coercive a .

To address the case that a fulfills the Gårding inequality (11) with $\lambda_a > 0$, we use a standard temporal transformation of the full problem as proposed e.g. in [41, 44]. We set $\hat{w} := e^{-\lambda_a t} w$ for $w \in \mathcal{X}$, $\hat{p} = e^{\lambda_a t} p$ for $p \in \mathcal{Y}$, and define the bilinear form $\hat{b} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ by

$$(25) \quad \hat{b}(\hat{w}, \hat{p}) := \int_{\Omega_{t,x}} \langle \hat{w}, -k \cdot \nabla_{t,x} \hat{p} \rangle_{V, V'} + a((t, x); \hat{w}, \hat{p}) + \lambda_a (\hat{w}, \hat{p})_{L^2(\Omega_v)} d(t, x).$$

Then it holds $b(w, p) = \hat{b}(\hat{w}, \hat{p})$ for all $w \in \mathcal{X}, p \in \mathcal{Y}$. The transformed bilinear form \hat{b} is the same as b , with a transformed velocity bilinear form $\hat{a} : V \times V \rightarrow \mathbb{R}$ defined by $\hat{a}((t, x); \phi, \psi) = a((t, x); \phi, \psi) + \lambda_a (\phi, \psi)_{L^2(\Omega_v)}$ for $\phi, \psi \in V$. Due to the Gårding

⁴In the following, we omit the (t, x) dependence in the integrals.

inequality (11) and continuity (10) of a , \hat{a} is coercive with constant $\hat{\alpha}_a = \alpha_a$ and continuous with constant $\hat{c}_a = c_a + \lambda_a$. As in [41], we can estimate the norms of $\hat{w} \in \mathcal{X}$ and $\hat{p} \in \mathcal{Y}$ by

$$\|\hat{w}\|_{\mathcal{X}} \geq e^{-\lambda_a T} \|w\|_{\mathcal{X}}, \quad \|\hat{p}\|_{\mathcal{Y}} \geq (\max\{1 + 2\lambda_a^2, 2\})^{-\frac{1}{2}} \|p\|_{\mathcal{Y}},$$

where we use $\|\psi\|_{V'} \leq \|\psi\|_{L^2(\Omega_v)} \leq \|\psi\|_V$ for the estimation of the \mathcal{Y} -norm.

Then, the dual inf-sup constant of b can be bounded from below as follows

$$\begin{aligned} \inf_{p \in \mathcal{Y}} \sup_{w \in \mathcal{X}} \frac{b(w, p)}{\|w\|_{\mathcal{X}} \|p\|_{\mathcal{Y}}} &= \inf_{\hat{p} \in \mathcal{Y}} \sup_{\hat{w} \in \mathcal{X}} \frac{\hat{b}(\hat{w}, \hat{p})}{\|\hat{w}\|_{\mathcal{X}} \|\hat{p}\|_{\mathcal{Y}}} \frac{\|\hat{w}\|_{\mathcal{X}}}{\|w\|_{\mathcal{X}}} \frac{\|\hat{p}\|_{\mathcal{Y}}}{\|p\|_{\mathcal{Y}}} \\ &\geq \frac{\alpha_a}{\sqrt{2} \max\{1, c_a + \lambda_a\}} \frac{e^{-\lambda_a T}}{\sqrt{\max\{1 + 2\lambda_a^2, 2\}}}. \end{aligned}$$

Since the dual inf-sup condition implies surjectivity of the operator $B : \mathcal{X} \rightarrow \mathcal{Y}'$ defined by $\langle B \cdot, \cdot \rangle_{\mathcal{Y}', \mathcal{Y}} = b(\cdot, \cdot)$ and thus existence of a weak solution to (16) (see for instance [26, Lemma A.40, Remark A.41]), this concludes the proof. \square

4.2. Uniqueness of the weak solution. As already mentioned in section 3, we were not able to prove all necessary trace results in our specific function space. To show uniqueness of the weak solution, we therefore assume the following:

Assumption 4.4. *Let $w \in H_{\text{FP}}^1(\Omega)$ such that $w = 0$ a.e. on Γ_- and $b(w, p) = 0$ for all $p \in \mathcal{Y}$. Then, we assume this implies $w \in L^2(\partial\Omega, |k \cdot n|)$ and the integration by parts formula*

$$(26) \quad \int_{\Omega_{t,x}} \langle k \cdot \nabla_{t,x} w, w \rangle_{V', V} \, d(t, x) = \frac{1}{2} \int_{\partial\Omega} w^2 k \cdot n \, ds$$

holds.

As discussed in more detail in the supplementary material, we do not know how to prove Assumption 4.4, since, for instance, ideas from existing approaches for the related space $H_{\text{NT}}^1(\Omega) = \{w \in L^2(\Omega) : k \cdot \nabla_{t,x} w \in L^2(\Omega)\}$ cannot be readily transferred to the $H_{\text{FP}}^1(\Omega)$ case. We therefore leave it as an open problem. We emphasize that the respective trace and integration by parts result holds for all $H_{\text{NT}}^1(\Omega)$ -functions with zero inflow or outflow trace (cf. [6, 12, 13], [16, Chap. XXI]), and also for all $H_{\text{FP}}^1(\Omega)$ -functions that can be approximated by smooth functions vanishing on the inflow or outflow boundary (Proposition 3.4). Additionally, Assumption 4.4 only refers to $H_{\text{FP}}^1(\Omega)$ -functions with vanishing trace on Γ_- and satisfying a weak form of the differential equation with zero boundary condition. This additional condition on the considered functions might make it possible to show and exploit a higher regularity of the considered functions to prove existence of suitable traces and (26).

We now show uniqueness of the weak solution in the form of surjectivity of the dual operator. To that end, we follow the general structure of respective proofs for parabolic equations [26, Thm 6.6, p. 283] and transport equations [4, Thm. 16]. We take a function $w \in \mathcal{X}$ solving (16) with zero right-hand side and prove that $w = 0$ by showing that w possesses space- and time derivatives, that w has trace zero on the outflow boundary, and finally that w must therefore vanish on the whole domain.

Theorem 4.5. *If Assumption 4.4 holds, then for all $0 \neq w \in \mathcal{X}$ we have*

$$\sup_{p \in \mathcal{Y}} b(w, p) > 0.$$

Proof. Let $w \in \mathcal{X}$ such that

$$(27) \quad b(w, p) = 0 \quad \forall p \in \mathcal{Y}.$$

To prove the claim, we need to show that $w = 0$. First, we show that w has a weak derivative $-k \cdot \nabla_{t,x} w \in \mathcal{X}' = L^2(\Omega_{t,x}; V')$. To that end, let $\psi \in C_0^\infty(\Omega_{t,x})$ and $\phi \in V$ be arbitrary. Then $\psi\phi = 0$ on $\hat{\Gamma}$, and by approximating ϕ in $C^\infty(\Omega_v)$ we see that $\psi\phi \in \mathcal{Y}$. Using the definition of the weak (t, x) -derivative and testing (27) with $p = \psi\phi$ we obtain

$$\begin{aligned} & \int_{\Omega_{t,x}} \langle k(t, x) \cdot \nabla_{t,x} w(t, x), \phi \rangle_{V', V} \psi(t, x) d(t, x) \\ &= - \int_{\Omega_{t,x}} \langle w(t, x), k(t, x) \cdot \nabla_{t,x} \psi(t, x) \phi \rangle_{V', V} d(t, x) \\ &= - \int_{\Omega_{t,x}} a((t, x); w(t, x), \psi(t, x)\phi) d(t, x) \\ &= - \int_{\Omega_{t,x}} \langle A(t, x)w(t, x), \phi \rangle_{V', V} \psi(t, x) d(t, x), \end{aligned}$$

where the operator $A_v(t, x) \in \mathcal{L}(V, V')$ is defined as $\langle A_v(t, x)\phi, \rho \rangle_{V', V} = a((t, x); \phi, \rho)$ for all $\phi, \rho \in V$, a.e. $(t, x) \in \Omega_{t,x}$. Due to the density of $C_0^\infty(\Omega_{t,x})$ in $L^2(\Omega_{t,x})$ have

$$(28) \quad -k \cdot \nabla_{t,x} w = A_v w \in \mathcal{X}',$$

which especially means that $w \in H_{\text{FP}}^1(\Omega)$.

Next, let $K \subset\subset \Gamma_-$ be an arbitrary but fixed compactly embedded subset of Γ_- . Moreover, let $z \in C^\infty(\bar{\Omega})$ with $z = 0$ on $\hat{\Gamma} \setminus K$. We show $wz \in \mathcal{Y}$: Since $w \in H_{\text{FP}}^1(\Omega)$, due to Proposition 3.1 there is a sequence $(w_n)_{n \in \mathbb{N}} \subset C^\infty(\bar{\Omega})$ with $\|w_n - w\|_{H_{\text{FP}}^1(\Omega)} \xrightarrow{n \rightarrow \infty} 0$. Therefore, we have $w_n z \in C^\infty(\bar{\Omega})$ with $wz = 0$ on Γ_+ . Due to Lemma 3.3, it holds

$$\|wz - w_n z\|_{H_{\text{FP}}^1(\Omega)} \leq C \|z\|_{C^1(\Omega)} \|w - w_n\|_{H_{\text{FP}}^1(\Omega)}$$

and thus $w_n z \rightarrow wz$ in $H_{\text{FP}}^1(\Omega)$ as $n \rightarrow \infty$. Invoking the definition of \mathcal{Y} in (12),(7) we obtain $wz \in \mathcal{Y}$.

Since $K \subset \Gamma_-$ is compact, we may apply Proposition 3.2 to infer that w has a trace on K and $w|_K \in L^2(K, |k \cdot n|)$. Thanks to $z|_{\hat{\Gamma}} \in L^\infty(\hat{\Gamma})$ and $\text{supp } z|_{\hat{\Gamma}} \subset K$, we have

$$\left| \int_{\hat{\Gamma}} w^2 z |k \cdot n| ds \right| = \left| \int_K w^2 z |k \cdot n| ds \right| \leq \|z\|_{L^\infty(K)} \|w\|_{L^2(K, |k \cdot n|)}^2 < \infty.$$

As a consequence we can apply the linear functional in (28) to $wz \in \mathcal{Y} \subset \mathcal{X}$, perform integration by parts, since the boundary integral exists, and use (27):

$$\begin{aligned} 0 &= \int_{\Omega_{t,x}} \langle k \cdot \nabla_{t,x} w + A_v w, wz \rangle_{V',V} d(t,x) \\ &= \int_{\Omega_{t,x}} \langle w, -k \cdot \nabla_{t,x}(wz) \rangle_{V,V'} + a(w, wz) d(t,x) + \int_{\tilde{\Gamma}} w^2 z k \cdot n \, ds \\ &= \underbrace{b(w, wz)}_{=0} - \int_K w^2 z |k \cdot n| \, ds = - \int_K w^2 z |k \cdot n| \, ds. \end{aligned}$$

Since $z|_K \in C_0^\infty(K)$ can be chosen arbitrarily and $|k \cdot n| > 0$ on K , the fundamental lemma of calculus of variations yields $w = 0$ a.e. on K . As also $K \subset \Gamma_-$ was chosen arbitrarily, we have $w = 0$ a.e. on Γ_- .

Thanks to Assumption 4.4, it therefore holds $w \in L^2(\partial\Omega, |k \cdot n|)$. We can thus use integration by parts for (28) applied to w . Assuming first that a is coercive, i.e., $\lambda_a \leq 0$, we obtain

$$\begin{aligned} 0 &= \int_{\Omega_{t,x}} \langle k \cdot \nabla_{t,x} w + A_v w, w \rangle_{V',V} d(t,x) \\ &= \int_{\Omega_{t,x}} \langle k \cdot \nabla_{t,x} w, w \rangle_{V',V} d(t,x) + \int_{\Omega_{t,x}} a(w, w) d(t,x) \\ &\geq \frac{1}{2} \int_{\Gamma_+} w^2 \underbrace{k \cdot n}_{>0} \, ds + \alpha_a \|w\|_{\mathcal{X}}^2, \end{aligned}$$

which implies $w = 0$.

If a is not coercive, we use the temporal transformation described in the proof of Theorem 4.2. Setting $\hat{w} = e^{-\lambda_a t} w$ and using the definition of \hat{b} in (25), we see that (27) is equivalent to $\hat{b}(\hat{w}, \hat{p}) = 0$ for all $\hat{p} \in \mathcal{Y}$. Since \hat{a} is coercive, we have proven that $\hat{w} = 0$ and thus also $w = 0$. \square

We summarize our findings in the following theorem.

Theorem 4.6 (Well-posedness). *There exists a solution $u \in \mathcal{X}$ to the variational problem (16). If Assumption 4.4 holds, the solution is unique and satisfies the stability estimate*

$$\|u\|_{\mathcal{X}} \leq \frac{1}{\beta} \|f\|_{\mathcal{Y}'}$$

for β as defined in Theorem 4.2.

Proof. Standard inf-sup theory ensures the existence of a solution due to the continuity of b and the dual inf-sup condition stated in Theorem 4.2. Under Assumption 4.4, Theorem 4.5 yields the dual surjectivity, which implies uniqueness and the stability estimate. \square

5. DISCRETIZATION

We now design a stable and efficient discretization scheme for (16). To that end, we use a Petrov-Galerkin projection onto problem-dependent discrete spaces realizing the stable function pairs with test functions $p \in \mathcal{Y}$ and trial functions $w_p \in \mathcal{X}$ developed in the proof of Theorem 4.2. As a result, the discrete inf-sup stability and thus the well-posedness of the discrete problem follow analogously to

the continuous results with the same stability constant. We then illustrate for a class of data functions how the trial space functions w_p^δ can be efficiently computed by solving low-dimensional elliptic problems in the velocity domain.

5.1. Stable Petrov-Galerkin schemes. To define an approximation of the solution $u \in \mathcal{X}$ of (16), we use a Petrov-Galerkin projection onto suitable discrete spaces: Given discrete trial and test spaces $\mathcal{X}_\delta \subset \mathcal{X}$ and $\mathcal{Y}_\delta \subset \mathcal{Y}$, the Petrov-Galerkin approximation $u^\delta \in \mathcal{X}_\delta$ is defined by

$$(29) \quad b(u^\delta, v^\delta) = f(v^\delta) \quad \forall v^\delta \in \mathcal{Y}_\delta.$$

Well-posedness then depends on the inf-sup stability of the discrete problem. To find a pair of spaces leading to a stable scheme, we transfer ideas from [10] to our setting. In [10], a stable discretization with a discrete inf-sup constant equal to one was built for a transport equation by fixing a discrete test space and defining a problem dependent trial space with optimal stability properties. In this manuscript, we will use the same strategy: We start with a discrete test space and define the corresponding trial space based on the trial space functions used in the proof of Theorem 4.2.

To that end, we first define a discrete space $V_h \subset V$ for the discretization in the velocity direction. Since the \mathcal{Y} -norm contains a term in the $\mathcal{X}' = L^2(\Omega_{t,x}, V')$ -norm (see (14)) which is not computable, we consider the norm

$$(30) \quad \|w\|_{L^2(\Omega_{t,x}, V'_h)}^2 := \int_{\Omega_{t,x}} \|w(t, x)\|_{V'_h}^2 d(t, x), \quad \|\psi\|_{V'_h} := \sup_{\phi^h \in V_h} \frac{\langle \psi, \phi^h \rangle_{V', V}}{\|\phi^h\|_V}$$

instead of $\|\cdot\|_{L^2(\Omega_{t,x}, V')}$ where necessary.

Let $\mathcal{Y}_\delta \subset \mathcal{Y}$ be a discrete space for which we assume $w^\delta(t, x) \in V_h$ for all $w^\delta \in \mathcal{Y}_\delta$ and a.e. $(t, x) \in \Omega_{t,x}$. The space \mathcal{Y}_δ will be used as the test space for the Petrov-Galerkin approximation. We define the discrete version of the \mathcal{Y} -norm by

$$(31) \quad \|w\|_{\mathcal{Y}_\delta}^2 := \|w\|_{L^2(\Omega_{t,x}, V)}^2 + \|k \cdot \nabla_{t,x} w\|_{L^2(\Omega_{t,x}, V'_h)}^2.$$

Since we will make use of the function pairs developed in the proof of Theorem 4.2, we assume for the discretization that the velocity bilinear form a is coercive, i.e., $\lambda_a \leq 0$. For problems, where a only satisfies the Gårding inequality (11) with $\lambda_a > 0$, a temporal transformation of the problem as described in section 4 can be performed. Then, the transformed problem with a coercive bilinear form \hat{a} can be discretized.

We now define a problem-dependent discrete trial space. For each $p^\delta \in \mathcal{Y}_\delta$, we denote $f_p^\delta := -k \cdot \nabla_{t,x} p^\delta(t, x) \in \mathcal{X}'$. We then define the function $z_p^\delta \in \mathcal{X}$ as the solution of

$$(32) \quad a(z_p^\delta(t, x), \phi^h) = \langle r_p^\delta(t, x), \phi^h \rangle_{V', V}, \quad \forall \phi^h \in V_h, \text{ a.e. } (t, x) \in \Omega_{t,x}.$$

The function z_p^δ is the discrete counterpart of z_p defined in (19), here it is defined pointwise in $\Omega_{t,x}$ due to the discrete setting. Then, the discrete trial space $\mathcal{X}_\delta \subset \mathcal{X}$ is defined as

$$(33) \quad \mathcal{X}_\delta := \{p^\delta + z_p^\delta : p^\delta \in \mathcal{Y}_\delta\}.$$

Proposition 5.1. *If $\lambda_a \leq 0$ in (11) and thus a is coercive, and if the discrete trial and test spaces \mathcal{X}_δ and \mathcal{Y}_δ are chosen according to (33), then there exists a unique*

solution $u^\delta \in \mathcal{X}_\delta$ to (29). Moreover, we have discrete inf-sup estimate

$$(34) \quad \inf_{\substack{p^\delta \in \mathcal{Y}_\delta \\ p^\delta \neq 0}} \sup_{\substack{w^\delta \in \mathcal{X}_\delta \\ w^\delta \neq 0}} \frac{b(w^\delta, p^\delta)}{\|w^\delta\|_{\mathcal{X}} \|p^\delta\|_{\mathcal{Y}_\delta}} \geq \beta_\delta \geq \alpha_a (\sqrt{2} \max\{1, c_a\})^{-1}.$$

Remark 5.2. For $\lambda_a > 0$ the respective result holds for the discretization of the transformed problem according to (25) with \hat{a} being coercive.

Proof. We can reuse all essential parts of the proof of the inf-sup constant for the continuous problem to also prove discrete inf-sup stability of (29).

Let $0 \neq w^\delta \in \mathcal{X}_\delta$ be fixed. Then, by definition of \mathcal{X}_δ there is $p^\delta \in \mathcal{Y}_\delta$ such that $w^\delta = p^\delta + z_p^\delta$ with z_p^δ defined as in (32). By using (32) and the same arguments as in (21) we obtain

$$(35) \quad b(w^\delta, p^\delta) = b(p^\delta + z_p^\delta, p^\delta) \geq \alpha_a (\|p^\delta\|_{\mathcal{X}}^2 + \|z_p^\delta\|_{\mathcal{X}}^2).$$

As we have

$$\langle r_p^\delta(t, x), \phi^h \rangle_{V', V} = a(z_p^\delta(t, x), \phi^h) \leq c_a \|z_p^\delta(t, x)\|_V \|\phi^h\|_V \quad \forall \phi^h \in V_h, \text{ a.e. } (t, x) \in \Omega_{t,x}$$

we can inflect that

$$(36) \quad \|r_p^\delta\|_{L^2(\Omega_{t,x}, V'_h)} \leq c_a \|z_p^\delta\|_{\mathcal{X}}.$$

Therefore, we obtain analogously to (23), but using the discrete \mathcal{Y}_δ -norm,

$$\begin{aligned} \|w_p^\delta\|_{\mathcal{X}} \|p^\delta\|_{\mathcal{Y}_\delta} &= \|p^\delta + z_p^\delta\|_{\mathcal{X}} \left(\|p^\delta\|_{\mathcal{X}}^2 + \|r_p^\delta\|_{L^2(\Omega_{t,x}, V'_h)}^2 \right)^{1/2} \\ &\stackrel{(36)}{\leq} \left[\|p^\delta + z_p^\delta\|_{\mathcal{X}}^2 (\|p^\delta\|_{\mathcal{X}}^2 + c_a^2 \|z_p^\delta\|_{\mathcal{X}}^2) \right]^{1/2} \\ &\leq [2 (\|p^\delta\|_{\mathcal{X}}^2 + \|z_p^\delta\|_{\mathcal{X}}^2) (\|p^\delta\|_{\mathcal{X}}^2 + c_a^2 \|z_p^\delta\|_{\mathcal{X}}^2)]^{1/2} \\ &= \sqrt{2} \max\{1, c_a\} (\|p^\delta\|_{\mathcal{X}}^2 + \|z_p^\delta\|_{\mathcal{X}}^2) \stackrel{(35)}{\leq} \frac{\sqrt{2} \max\{1, c_a\}}{\alpha_a} b(w_p^\delta, p^\delta). \end{aligned}$$

This means that b is inf-sup stable on the spaces $(\mathcal{X}_\delta, \|\cdot\|_{\mathcal{X}}), (\mathcal{Y}_\delta, \|\cdot\|_{\mathcal{Y}_\delta})$ with constant $\beta_\delta \geq \alpha_a (\sqrt{2} \max\{1, c_a\})^{-1}$. Since for all $0 \neq p^\delta$ it holds $b(w_p^\delta, p^\delta) > 0$ and thus $w_p^\delta \neq 0$, we have $\dim(\mathcal{X}_\delta) = \dim(\mathcal{Y}_\delta)$. Therefore, inf-sup stability already guarantees well-posedness of the discrete problem (29). \square

Remark 5.3. Due to the finite-dimensional spaces, the Petrov-Galerkin approximation $u^\delta \in \mathcal{X}_\delta$ is unique even if Assumption 4.4 does not hold.

Remark 5.4 (Choice of λ_a in the case $\lambda_a > 0$). For possibly non-coercive problems, there is usually some flexibility in the choice of α_a and λ_a such that the Gårding inequality (11) is fulfilled: On the one hand, if (11) holds for a specific λ_a , all $\tilde{\lambda}_a > \lambda_a$ are also possible. On the other hand, often (11) holds for all $\lambda_a > 0$ with different respective $\alpha_a > 0$; think, for instance, of $a(\psi, \theta) = (\nabla_v \psi, \nabla_v \theta)_{L^2(\Omega_v)}$, where (11) holds for any $\lambda_a > 0$ with $\alpha_a = \min(1, \lambda_a)$. When using a temporal transformation before the discretization, the constant λ_a should not be too large: Since $e^{\lambda_a t}$ appears in the temporal transformation, a large λ_a leads to error amplification and a very small effective inf-sup constant of the “non-transformed” discrete problem (cf. (18)). Therefore, a suitable balancing of λ_a and α_a with possibly small λ_a and large α_a should be sought to obtain a stable discretization when using the temporal transformation.

5.2. Efficient numerical scheme. Regarding the computational realization of the Petrov-Galerkin approximation, we have to take into account the specific choice of the discrete spaces according to (33). To assemble the linear system and to represent the discrete solution, the functions z_p^δ defined by (32), have to be computed for all basis functions of \mathcal{Y}_δ . We illustrate how this can be done very efficiently for the case where a is coercive and has the separable form

$$(38) \quad a((t, x), \phi, \psi) = d(t, x)\tilde{a}(\phi, \psi),$$

where $d \in L^\infty(\Omega_{t,x})$ satisfies $d(t, x) \geq \alpha^d > 0$ for a.e. $(t, x) \in \Omega_{t,x}$ and $\tilde{a} : V \times V \rightarrow \mathbb{R}$ is a coercive bilinear form.

To build the discrete test space, let first $\bar{\mathcal{Y}}_\delta^{t,x} \subset H^1(\Omega_{t,x})$ be a discrete space in the space-time domain with basis $(p_i^{t,x,\delta}(t, x))_{i=1}^{n_{t,x}}$ and let $V_h \subset V$ be the already defined velocity discrete space with basis $(\psi_j^h(v))_{j=1}^{n_v}$. Denoting the tensor product of these spaces by $\bar{\mathcal{Y}}_\delta := \bar{\mathcal{Y}}_\delta^{t,x} \otimes V_h$, we then set

$$\mathcal{Y}_\delta := \text{span}\{p_{i,j}^\delta = p_i^{t,x,\delta} \psi_j^h : p_{i,j}^\delta|_{\Gamma_+} = 0\} \subset \bar{\mathcal{Y}}_\delta \cap \mathcal{Y}.$$

We may then use this tensor product structure to efficiently solve (32): Fixing a basis function $p_{i,j}^\delta = p_i^{t,x,\delta} \psi_j^h$ of \mathcal{Y}_δ , the right-hand side of (32) reads

$$\begin{aligned} \langle -k \cdot \nabla_{t,x} p_{i,j}^\delta(t, x), \phi^h \rangle_{V', V} &= -\partial_t p_i^{t,x,\delta}(t, x) \int_{\Omega_v} \psi_j^h(v) \phi^h(v) \, dv \\ &\quad - \sum_{k=1}^d \partial_{x_k} p_i^{t,x,\delta}(t, x) \int_{\Omega_v} v_k \psi_j^h(v) \phi^h(v) \, dv \end{aligned}$$

for all $\phi^h \in V_h$, a.e. $(t, x) \in \Omega_{t,x}$. Using the separable form of a (38), we can rewrite (32) as follows: Find $z_{i,j}^\delta := z_{p_{i,j}^\delta}^\delta \in \mathcal{X}$, such that

$$\begin{aligned} d(t, x)\tilde{a}(z_{i,j}^\delta(t, x), \phi^h) &= -\partial_t p_i^{t,x,\delta}(t, x) \int_{\Omega_v} \psi_j^h(v) \phi^h(v) \, dv \\ &\quad - \sum_{k=1}^d \partial_{x_k} p_i^{t,x,\delta}(t, x) \int_{\Omega_v} v_k \psi_j^h(v) \phi^h(v) \, dv \\ &\quad \forall \phi^h \in V_h, \text{ a.e. } (t, x) \in \Omega_{t,x}. \end{aligned}$$

Hence, the computation of all $z_{i,j}^\delta$ can be separated in the following way: We first compute the solutions $\rho_j^1, \rho_j^{v_1}, \dots, \rho_j^{v_d} \in V_h$ to the problems

$$(39) \quad \begin{aligned} \tilde{a}(\rho_j^1, \phi^h) &= \int_{\Omega_v} \psi_j^h(v) \phi^h(v) \, dv, \quad \forall \phi^h \in V_h, \\ \tilde{a}(\rho_j^{v_k}, \phi^h) &= \int_{\Omega_v} v_k \psi_j^h(v) \phi^h(v) \, dv, \quad \forall \phi^h \in V_h, k = 1, \dots, d, \end{aligned}$$

for all basis functions $\psi_j^h \in V_h, j = 1, \dots, n_v$. Then, the $z_{i,j}^\delta$ are given by

$$(40) \quad z_{i,j}^\delta(t, x, v) = -d(t, x)^{-1} \left(\partial_t p_i^{t,x,\delta}(t, x) \rho_j^1(v) + \sum_{k=1}^d \partial_{x_k} p_i^{t,x,\delta}(t, x) \rho_j^{v_k}(v) \right).$$

The full solution process thus consists of the following steps:

- (1) Precompute $\rho_j^1, \rho_j^{v_k}$, i.e., solve $(d+1) \times n_v$ problems of size n_v , which can be done in parallel.

- (2) Assemble the stiffness matrix $[b(p_{i,j}^\delta + z_{i,j}^\delta, p_{k,l}^\delta)]_{(k,l),(i,j)}$, using (40), and assemble the load vector $[f(p_{k,l}^\delta)]_{(k,l)}$.
- (3) Solve the linear system of equations to obtain the coefficient vector $[u_{i,j}]_{(i,j)}$.
- (4) Compose the solution $u^\delta = \sum_{i,j} u_{i,j} (p_{i,j}^\delta + z_{i,j}^\delta) \in \mathcal{X}_\delta$ by again using (40) for $z_{i,j}^\delta$.

Compared to using finite element spaces without any stabilization, the additional costs thus only lie in the n_v -sized problems (step 1) and possibly more nonzero elements in the stiffness matrix. These effects only depend on the dimension n_v of V_h . Therefore, the proposed discretization strategy is especially well-suited for using specific spaces V_h of low dimension, which can be achieved for example by using polynomial bases or a hierarchical model reduction approach as proposed in [9].

In order to efficiently compute the problem-dependent basis functions, we heavily rely on the separable form of the bilinear form a given in (38), which is unfortunately often not fulfilled for realistic data. For general bilinear forms, (32) remains a variational problem in all dimensions that is not directly decomposable in single low-dimensional problems. However, as the velocity operator is elliptic, for realistic data functions we usually expect the problem to be well-suited for model reduction strategies. Therefore, it might be possible to use low-rank approximations as done in a related setting in [7] to find sufficiently accurate approximate solutions to (32) in a computationally efficient manner.

More generally, due to the high-dimensionality of the problem, it is especially desirable to combine the approach proposed in this manuscript with further approximations as the already mentioned hierarchical model reduction [9] or tensor-based methods that have already been used in similar Petrov-Galerkin settings [7, 31] and to discretize kinetic equations like the radiative transfer equation [28, 45] or the Vlasov equation [23, 24, 36].

6. NUMERICAL EXPERIMENTS

We investigate the properties of the method developed in section 5 by implementing the discretization for the Fokker-Planck equation (1) on a two-dimensional spatial domain as well as for a modified stationary equation. We are especially interested in the convergence of the discretization error, analyzing how sharp the lower bound for the inf-sup constant is and examining the efficiency in light of the nonstandard discrete spaces. The source code to reproduce all results is provided in [8].

6.1. Test Cases. Let $\Omega_x = (0, 1)^2 \subset \mathbb{R}^2$ be the spatial domain and $I_t = (0, 0.75)$ be the time interval. We parametrize $\Omega_v = S^1$ by the angle $\varphi \in [0, 2\pi)$, leading to $v = (\begin{smallmatrix} \cos \varphi \\ \sin \varphi \end{smallmatrix})$ and $\Delta_v u = \frac{\partial^2}{\partial \varphi^2} u$.

We consider the Fokker-Planck equation (2) for a constant $q \in \mathbb{R}_+$. Then, the equation reads

$$(41) \quad \begin{aligned} \partial_t u((t, x), \varphi) + \left(\begin{smallmatrix} \cos \varphi \\ \sin \varphi \end{smallmatrix} \right) \cdot \nabla_x u((t, x), \varphi) &= q^{-1} \frac{\partial^2}{\partial \varphi^2} u((t, x), \varphi) \quad \text{in } \Omega, \\ u((t, x), \varphi) &= g((t, x), \varphi) \quad \text{on } \Gamma_-, \end{aligned}$$

where we choose the initial condition

$$g((0, x), \varphi) := \begin{cases} \frac{1}{2\pi}(128r(x)^3 - 48r(x)^2 + 1), & r(x) < \frac{1}{4}, \\ 0, & r(x) \geq \frac{1}{4}, \end{cases}$$

with $r(x_1, x_2) := \sqrt{(0.5 - x_1)^2 + (0.5 - x_2)^2}$ and zero spatial inflow boundary conditions $g|_{\Gamma_-^*(\varphi)} \equiv 0$ for all $\varphi \in [0, 2\pi]$.

The corresponding velocity bilinear form

$$a(\psi, \rho) := q^{-1} \int_0^{2\pi} \psi'(\varphi) \rho'(\varphi) d\varphi \quad \forall \psi, \rho \in V = H^1(\Omega_v)$$

fulfills the Gårding inequality (11) for any $\lambda_a > 0$ with $\alpha_a = \min(q^{-1}, \lambda_a)$. As mentioned in Remark 5.4, a choice with possibly large α_a and possibly small λ_a is desirable to obtain good results when using a temporal transformation according to (25). We only consider cases where $0.1 \leq q^{-1} \leq 1$, therefore we select $\lambda_a = q^{-1}$, $\alpha_a = q^{-1}$. Then, we discretize the transformed problem, where the transformed velocity bilinear form coincides with the scaled V -scalar product, i.e., $\hat{a} = q^{-1}(\cdot, \cdot)_V$.

For the discretization we choose $V_h \subset V$ as the continuous linear FE space on $[0, 2\pi)$ with periodic boundary condition and uniform mesh with size $h_v = 2\pi/n_v$. The space $\bar{\mathcal{Y}}_\delta^{t,x} \subset H^1(\Omega_{t,x})$ is chosen as the continuous \mathbb{Q}_2 FE space on a 3D rectangular mesh with uniform 1D mesh sizes $h_t = 0.75/n_t$ and $h_{x_1} = h_{x_2} = 1/n_x$. The trial space \mathcal{X}_δ is computed as described in subsection 5.2 by first solving $3n_v$ problems of dimension n_v . From the definition we see that $\mathcal{X}_\delta \subset \bar{\mathcal{X}}_\delta^{t,x} \otimes V_h$, with $\mathcal{X}_\delta^{t,x} \subset L^2(\Omega_{t,x})$ being the respective discontinuous \mathbb{Q}_2 FE space. After computing the transformed solution $\hat{u}^\delta \in \mathcal{X}_\delta$, we obtain the discrete solution to (41) by setting $u^\delta := e^t \hat{u}^\delta$.

To investigate the convergence rate of the newly proposed scheme, we additionally consider a stationary (and thus lower-dimensional) problem with a manufactured solution $u(x_1, x_2, \varphi) = \sin^2(\pi x_1) \sin^2(\pi x_2) \sin^2(\varphi)$ and corresponding right-hand side f_0 ; therefore slightly deviating from the original problem. More precisely, we consider

$$(42) \quad \left(\begin{array}{c} \cos \varphi \\ \sin \varphi \end{array} \right) \cdot \nabla_x u(x, \varphi) + c u(x, \varphi) = d \frac{\partial^2}{\partial \varphi^2} u(x, \varphi) + f_0(x, \varphi) \quad \text{in } \Omega_x \times \Omega_v$$

with reaction and velocity diffusion constants $c, d \in \mathbb{R}$, $c, d > 0$ and zero inflow boundary conditions on $\Gamma_- \subset \partial\Omega_x \times \Omega_v$. Note that we require $c > 0$ here in order to obtain a coercive bilinear form $a : V \times V \rightarrow \mathbb{R}$

$$a(\psi, \rho) = \int_0^{2\pi} d \psi'(\varphi) \rho'(\varphi) + c \psi(\varphi) \rho(\varphi) d\varphi \quad \forall \psi, \rho \in V.$$

Then, the bilinear form a is coercive with constant $\alpha_a = \min(c, d) > 0$ and continuous with constant $\gamma_v = \max(c, d)$. The variational formulation for the stationary equation (42) is based on $\mathcal{X}_{\text{st}} := L^2(\Omega_x; H^1(\Omega_v))$, and $\mathcal{Y}_{\text{st}} = \text{clos}_{\|\cdot\|_{\mathcal{Y}_{\text{st}}}} \{w \in C^1(\Omega_x \times \Omega_v) : w = 0 \text{ on } \Gamma_{\text{st},+}\}$, where

$$\begin{aligned} \Gamma_{\text{st},+} &= \{(x, v) \in \partial\Omega_x \times \Omega_v : \left(\begin{array}{c} \cos \varphi \\ \sin \varphi \end{array} \right) \cdot n_x > 0\}, \\ \|w\|_{\mathcal{Y}_{\text{st}}}^2 &= \|w\|_{\mathcal{X}_{\text{st}}}^2 + \left\| \left(\begin{array}{c} \cos \varphi \\ \sin \varphi \end{array} \right) \cdot \nabla_x w \right\|_{\mathcal{X}'_{\text{st}}}^2. \end{aligned}$$

The space-velocity bilinear form is

$$b_{\text{st}}(w, p) := \int_{\Omega_x} \langle w(x), -\left(\begin{array}{c} \cos \\ \sin \end{array} \right) \cdot \nabla_x p(x) \rangle_{V, V'} + a(w(x), p(x)) dx, \quad \forall w \in \mathcal{X}_{\text{st}}, p \in \mathcal{Y}_{\text{st}}$$

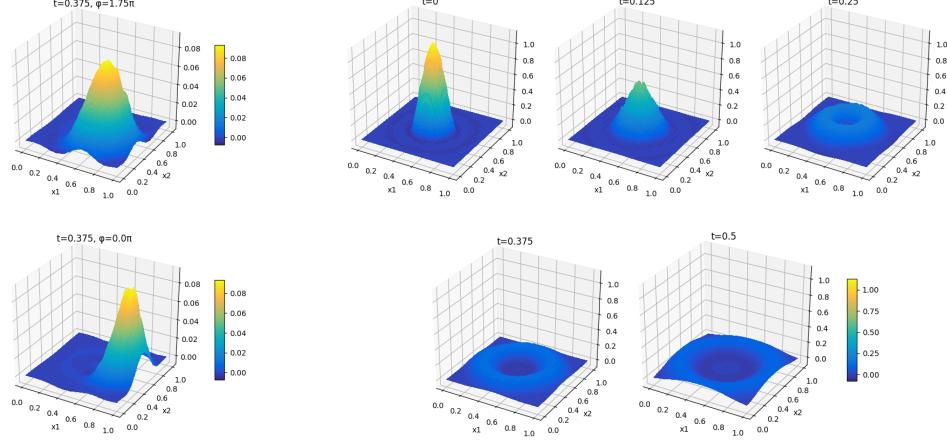


FIGURE 1. Plots of the solution u^δ of (41) with $q^{-1} = 0.8$ for $h_v = h_{x_1} = h_{x_2} = h_t = 1/16$. Left: Solution for fixed $t = 0.375$ and $\varphi = 1.75\pi$ (upper) and $\varphi = 0$ (lower). Right: Spatial density, i.e., moment $\int_0^{2\pi} u(\cdot, \cdot, \varphi) d\varphi$ for different t .

TABLE 1. Discretization of (41): Computed discrete inf-sup constants β_δ of the transformed problem in relation to the respective lower bound β_{lb} for varying mesh sizes with $n = 1/h_{x_1} = 1/h_{x_2} = 1/h_t = 2\pi/h_v$.

n	$q^{-1} = 0.8$		$q^{-1} = 0.4$		$q^{-1} = 0.1$	
	β_δ	β_δ/β_{lb}	β_δ	β_δ/β_{lb}	β_δ	β_δ/β_{lb}
4	0.8878	1.569	0.6418	2.269	0.45005	6.365
8	0.81141	1.434	0.44126	1.56	0.18668	2.64
16	0.80072	1.415	0.40317	1.425	0.11112	1.573

and the functional describing the source term is defined as

$$f_{st}(p) := \int_{\Omega_x} \int_0^{2\pi} f_0(x, \varphi) p(x, \varphi) d\varphi dx \quad \forall p \in \mathcal{Y}_{st}.$$

Well-posedness of the weak formulation of (42) follows completely analogously to the time-dependent case, as a is coercive and $f_{st} \in \mathcal{Y}'_{st}$. As in the time-dependent case, we choose V_h as linear FE space and $\mathcal{Y}_\delta^x \subset H^1(\Omega_x)$ as continuous \mathbb{Q}_2 FE space on a 2D uniform rectangular mesh.

6.2. Numerical results. We first compute the discrete solution to (41) for $q^{-1} = 0.8$ and $h_v = h_{x_1} = h_{x_2} = h_t = 1/16$. The assembly of the system matrices which includes the computation of the \mathcal{X}_δ basis functions as described in subsection 5.2 takes up about 11% of the computational time in our experiments. Hence, the additional low-dimensional problems in V_h are not dominant in the computational costs. In Fig. 1, plots of the solution are shown, where we see that the dynamics of the solution are captured well and that no instabilities or oscillations occur.

To investigate whether the estimate for the discrete inf-sup constant from section 5 is sharp, we compute the constants for the transformed problem with $\hat{a}(\cdot, \cdot) = q^{-1}(\cdot, \cdot)_V$ for different q^{-1} and different mesh sizes. In Table 1, we show the evaluated constants in relation to the lower bound (34), which is given for this test case as $q^{-1}/(\sqrt{2} \max(1, q^{-1}))$. We see that the estimate is sharp up to a factor of about $\sqrt{2}$.

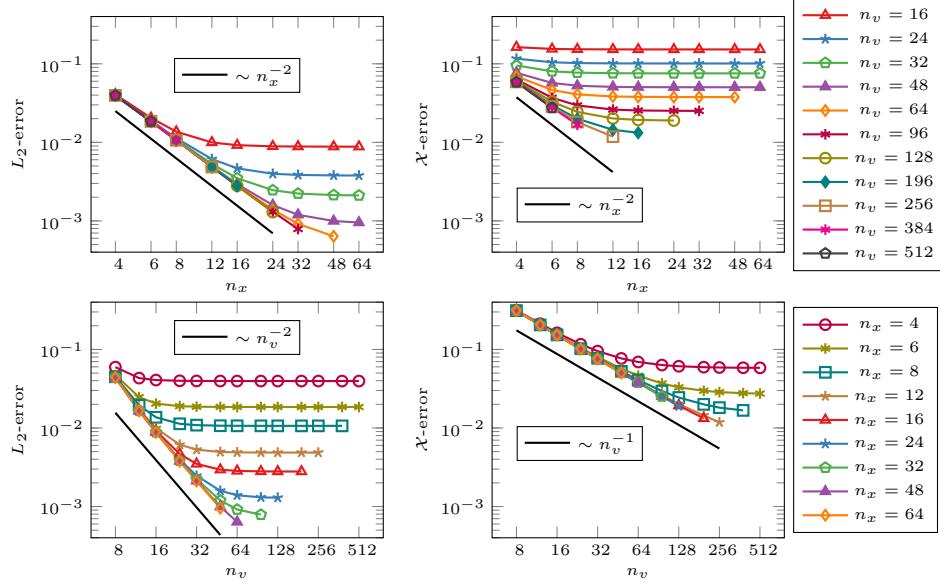


FIGURE 2. Discretization of (42) with $d = 0.1$, $c = 0.1$. Left: L^2 -errors $\|u - u_\delta\|_{L^2(\Omega_x \times \Omega_v)}$, right: \mathcal{X}_{st} -errors $\|u - u_\delta\|_{L^2(\Omega_x, V)}$. Upper plots: convergence in $n_x = 1/h_{x_1} = 1/h_{x_2}$ for different fixed $n_v = 2\pi/h_v$. Lower plots: Convergence in n_v for different fixed n_x .

TABLE 2. Discretization of (42): Computed discrete inf-sup constants β_δ in relation to the lower bound β_{lb} for varying mesh sizes with $n = 1/h_{x_1} = 1/h_{x_2} = 2\pi/h_v$ and varying values for the constants d and c .

n	$d = 0.4, c = 1$		$d = 0.1, c = 1$		$d = 0.1, c = 0.1$	
	β_δ	$\beta_\delta/\beta_{\text{lb}}$	β_δ	$\beta_\delta/\beta_{\text{lb}}$	β_δ	$\beta_\delta/\beta_{\text{lb}}$
4	0.61855	2.187	0.41087	5.811	0.30579	4.324
8	0.44891	1.587	0.18628	2.634	0.14924	2.111
16	0.40915	1.447	0.11688	1.653	0.10585	1.497
32	0.40202	1.421	0.1033	1.461	0.10041	1.42
48	0.40088	1.417	0.10137	1.434	0.10008	1.415

TABLE 3. Ratio of nonzero elements in the stiffness matrix for varying mesh sizes

n	$\frac{n_{\text{nz}}}{n_{\text{entries}}}$	$\frac{n_{\text{nz}}}{(n_{x_1} n_{x_2} n_v^2)}$
4	20.05%	39.3
8	5.52%	53.05
16	1.463%	58.98
32	0.378%	61.61
48	0.17%	62.44
64	0.096%	62.84

To examine the convergence behavior of our scheme, we compute discrete solutions to (42), where the exact solution is known. We compare the discretization errors for different mesh sizes in the $L^2(\Omega_x \times \Omega_v)$ norm as well as in the \mathcal{X}_{st} norm in Fig. 2. We see that the L^2 -error converges with second order in both $h_{x_1} = h_{x_2} = 1/n_x$ and $h_v = 2\pi/n_v$. The \mathcal{X}_{st} -error, which includes the L^2 -norm of the v -derivative, converges with second order in h_{x_i} and first order in h_v . For a further investigation of the estimate for the discrete inf-sup constant we compute the constants for the discretization of (42) for different mesh sizes and reaction and diffusion constants c and d ; see Table 2. The estimate (34) is given here as $\beta_\delta \geq \min\{c, d\}/(\sqrt{2} \max\{1, c, d\})$, which is $\min\{c, d\}/\sqrt{2}$ for all considered data values in Table 2. As can be seen in the table, the estimate is here again sharp up to a factor of about $\sqrt{2}$.

Since the basis functions of the discrete trial space \mathcal{X}_δ are not chosen as standard nodal basis functions but have larger support, one can ask if the choice of spaces still leads to an efficient numerical scheme. Therefore, in Table 3 we list the ratio

of nonzero elements in the stiffness matrix, which decreases significantly with larger problem sizes. However, as \mathcal{X}_δ includes solutions of problems in Ω_v , the nonzero elements increase linearly in the dimension of the x -discretization and quadratically in the dimension of V_h .

7. CONCLUSIONS

In this paper, we present a stable Petrov-Galerkin discretization of a kinetic Fokker-Planck equation. Based on an estimate for the dual inf-sup constant of the bilinear form, where ‘stable pairs’ of trial and test functions are introduced, we propose a discretization where these pairs are directly built into the spaces: By defining the discrete trial space dependent on the chosen discrete test space through the application of the kinetic transport and the inverse velocity Laplace-Beltrami operator, we obtain a well-posed numerical scheme with the same lower bound of the discrete inf-sup constant as for the continuous problem independently of the mesh size. We show that under suitable conditions on the data functions these spaces can be computed efficiently. Numerical experiments show favorable convergence orders of the discretization error for a manufactured solution of the stationary equation (order 2 in x both in the L^2 -norm and the \mathcal{X} -norm, order 2 and 1 in v for the respective norms). For both the examined time-dependent and stationary test cases, the estimate of the discrete inf-sup constant is sharp up to a factor of $\sqrt{2}$.

The new method is especially beneficial for spaces with few degrees of freedom in the velocity domain. Therefore, a promising application might be a combination with a hierarchical model order reduction scheme such as [9], which realizes small spaces in the velocity domain and has stability problems that might be resolved using the new method.

APPENDIX A. PROOFS OF FUNCTION SPACE RESULTS

Proof of Lemma 3.3. We estimate $\|\phi f\|_{H_{\text{FP}}^1(\Omega)}$. Using the definition of the V -norm and the product rule we obtain for the first term⁵

$$\begin{aligned} \|\phi f\|_{\mathcal{X}}^2 &= \|\phi f\|_{L^2(\Omega)}^2 + \|(\nabla_v \phi) f + \phi \nabla_v f\|_{L^2(\Omega)}^2 \\ &\leq \|\phi^2\|_{L^\infty(\Omega)} \|f\|_{L^2(\Omega)}^2 + 2\|\nabla_v \phi\|_{L^\infty(\Omega)}^2 \|f\|_{L^2(\Omega)}^2 + 2\|\phi^2\|_{L^\infty(\Omega)} \|\nabla_v f\|_{L^2(\Omega)}^2 \\ (43) \quad &\leq 2 \left(\|\phi\|_{L^\infty(\Omega)}^2 + \|\nabla_v \phi\|_{L^\infty(\Omega)}^2 \right) \|f\|_{\mathcal{X}}^2. \end{aligned}$$

By using the product rule, the identification $\langle \cdot, \cdot \rangle_{\mathcal{X}', \mathcal{X}} = (\cdot, \cdot)_{L^2(\Omega)}$, and the density of $C^\infty(\Omega)$ in $H_{\text{FP}}^1(\Omega)$ we see that for arbitrary $\psi \in \mathcal{X}$ it holds

$$\begin{aligned} \langle k \cdot \nabla_{t,x}(\phi f), \psi \rangle_{\mathcal{X}', \mathcal{X}} &= \langle k \cdot \nabla_{t,x} f, \phi \psi \rangle_{\mathcal{X}', \mathcal{X}} + \langle f(k \cdot \nabla_{t,x} \phi), \psi \rangle_{L^2(\Omega)} \\ &\leq \|k \cdot \nabla_{t,x} f\|_{\mathcal{X}'} \|\phi \psi\|_{\mathcal{X}} + \|f(k \cdot \nabla_{t,x} \phi)\|_{L^2(\Omega)} \|\psi\|_{L^2(\Omega)}. \\ (43) \quad &\stackrel{(43)}{\leq} \sqrt{2} \left(\|\phi\|_{L^\infty(\Omega)}^2 + \|\nabla_v \phi\|_{L^\infty(\Omega)}^2 \right)^{\frac{1}{2}} \|k \cdot \nabla_{t,x} f\|_{\mathcal{X}'} \|\psi\|_{\mathcal{X}} \\ &\quad + \|k \cdot \nabla_{t,x} \phi\|_{L^\infty(\Omega)} \|f\|_{L^2(\Omega)} \|\psi\|_{L^2(\Omega)} \\ &\leq \sqrt{2} \left(\|\phi\|_{L^\infty(\Omega)} + \|\nabla_v \phi\|_{L^\infty(\Omega)} + \|k \cdot \nabla_{t,x} \phi\|_{L^\infty(\Omega)} \right) \|f\|_{H_{\text{FP}}^1(\Omega)} \|\psi\|_{\mathcal{X}}. \end{aligned}$$

⁵As introduced in section 4, we write $\mathcal{X} = L^2(\Omega_{t,x}, V)$.

We thus have

$$(44) \quad \|k \cdot \nabla_{t,x}(\phi f)\|_{\mathcal{X}'} \leq 2\sqrt{2} (\|\phi\|_{L^\infty(\Omega)} + \|\nabla_v \phi\|_{L^\infty(\Omega)} \\ + \|k \cdot \nabla_{t,x} \phi\|_{L^\infty(\Omega)}) \|f\|_{H_{\text{FP}}^1(\Omega)}.$$

Combining (43) and (44) and using that $|k|$ is bounded in Ω , we thus have

$$\|\phi f\|_{H_{\text{FP}}^1(\Omega)} \leq C \|\phi\|_{C^1(\Omega)} \|f\|_{H_{\text{FP}}^1(\Omega)}.$$

□

ACKNOWLEDGMENTS

We would like to thank Dr. M. Schlottbom (University of Twente) and Prof. M. Ohlberger (University of Münster) for fruitful discussions.

REFERENCES

- [1] S. ARMSTRONG AND J.-C. MOURRAT, *Variational methods for the kinetic Fokker-Planck equation*, Feb. 2019, <https://arxiv.org/abs/1902.04037v1>.
- [2] M. ASADZADEH AND P. KOWALCZYK, *Convergence analysis of the streamline diffusion and discontinuous Galerkin methods for the Vlasov-Fokker-Planck system*, Numer. Methods Partial Differential Equations, 21 (2005), pp. 472–495, <https://doi.org/10.1002/num.20044>.
- [3] M. ASADZADEH AND A. SOPASAKIS, *Convergence of a hp-streamline diffusion scheme for Vlasov-Fokker-Planck system*, Math. Models Methods Appl. Sci., 17 (2007), pp. 1159–1182, <https://doi.org/10.1142/S0218202507002236>.
- [4] P. AZÉRAD, *Analyse des équations de Navier-Stokes en bassin peu profond et de l'équation de transport*, PhD thesis, Université de Neuchâtel, 1996.
- [5] G. BAL AND B. PALACIOS, *Pencil-beam approximation of stationary Fokker-Planck*, SIAM J. Math. Anal., 52 (2020), pp. 3487–3519, <https://doi.org/10.1137/19M1295775>.
- [6] C. BARDOS, *Problèmes aux limites pour les équations aux dérivées partielles du premier ordre à coefficients réels; théorèmes d'approximation; application à l'équation de transport*, Ann. Sci. École Norm. Sup. (4), 3 (1970), pp. 185–233, <https://doi.org/10.24033/asens.1190>.
- [7] M. BILLAUD-FRIESS, A. NOUY, AND O. ZAHM, *A tensor approximation method based on ideal minimal residual formulations for the solution of high-dimensional problems*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 1777–1806, <https://doi.org/10.1051/m2an/2014019>.
- [8] J. BRUNKEN, *Source code to “Stable and efficient Petrov-Galerkin methods for a kinetic Fokker-Planck equation”*, 2021, <https://doi.org/10.5281/zenodo.4106756>.
- [9] J. BRUNKEN, T. LEIBNER, M. OHLBURGER, AND K. SMETANA, *Problem adapted hierachical model reduction for the Fokker-Planck equation.*, in Proceedings of ALGORITMY 2016, the 20th Conference on Scientific Computing (Vysoké Tatry, Podbanske, Slovakia, 2016), A. Handlovičova and D. Sevcovič, eds., Publishing House of Slovak University of Technology in Bratislava, 2016, pp. 13–22.
- [10] J. BRUNKEN, K. SMETANA, AND K. URBAN, *(Parametrized) first order transport equations: Realization of optimally stable Petrov-Galerkin methods*, SIAM Journal on Scientific Computing, 41 (2019), pp. A592–A621, <https://doi.org/10.1137/18M1176269>.
- [11] J. A. CARRILLO, *Global weak solutions for the initial-boundary-value problems to the Vlasov-Poisson-Fokker-Planck system*, Math. Methods Appl. Sci., 21 (1998), pp. 907–938, [https://doi.org/10.1002/\(SICI\)1099-1476\(19980710\)21:10<907::AID-MMA977>3.3.CO;2-N](https://doi.org/10.1002/(SICI)1099-1476(19980710)21:10<907::AID-MMA977>3.3.CO;2-N).
- [12] M. CESSENAT, *Théorèmes de trace L^p pour des espaces de fonctions de la neutronique*, C. R. Acad. Sci. Paris Sér. I Math., 299 (1984), pp. 831–834.
- [13] M. CESSENAT, *Théorèmes de trace pour des espaces de fonctions de la neutronique*, C. R. Acad. Sci. Paris Sér. I Math., 300 (1985), pp. 89–92.
- [14] M. J. CÁCERES, J. A. CARRILLO, AND L. TAO, *A numerical solver for a nonlinear Fokker-Planck equation representation of neuronal network dynamics*, Journal of Computational Physics, 230 (2011), pp. 1084 – 1099, <https://doi.org/10.1016/j.jcp.2010.10.027>.
- [15] W. DAHMEN, C. HUANG, C. SCHWAB, AND G. WELPER, *Adaptive Petrov-Galerkin methods for first order transport equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445, <https://doi.org/10.1137/110823158>.

- [16] R. DAUTRAY AND J.-L. LIONS, *Mathematical analysis and numerical methods for science and technology. Vol. 6*, Springer-Verlag, Berlin, 1993, <https://doi.org/10.1007/978-3-642-58004-8>. Evolution problems. II.
- [17] B. DAVISON AND J. B. SYKES, *Neutron transport theory*, Oxford, at the Clarendon Press, 1957.
- [18] P. DEGOND AND S. MAS-GALLIC, *Existence of solutions and diffusion approximation for a model Fokker-Planck equation*, in Proceedings of the conference on mathematical methods applied to kinetic equations (Paris, 1985), vol. 16, 1987, pp. 589–636, <https://doi.org/10.1080/00411458708204307>.
- [19] L. DEMKOWICZ AND J. GOPALAKRISHNAN, *A class of discontinuous Petrov-Galerkin methods. II. Optimal test functions*, Numer. Methods Partial Differential Equations, 27 (2011), pp. 70–105, <https://doi.org/10.1002/num.20640>.
- [20] J. J. DUDESTADT AND W. R. MARTIN, *Transport theory*, John Wiley & Sons, New York-Chichester-Brisbane, 1979.
- [21] G. DZIUK AND C. M. ELLIOTT, *Finite element methods for surface PDEs*, Acta Numerica, 22 (2013), pp. 289–396, <https://doi.org/10.1017/S0962492913000056>.
- [22] H. EGGER AND M. SCHLOTTEBOM, *A mixed variational framework for the radiative transfer equation*, Math. Models Methods Appl. Sci., 22 (2012), pp. 1150014, 30, <https://doi.org/10.1142/S021820251150014X>.
- [23] V. EHRLACHER AND D. LOMBARDI, *A dynamical adaptive tensor method for the Vlasov-Poisson system*, J. Comput. Phys., 339 (2017), pp. 285–306, <https://doi.org/10.1016/j.jcp.2017.03.015>.
- [24] L. EINKEMMER AND C. LUBICH, *A low-rank projector-splitting integrator for the Vlasov-Poisson equation*, SIAM J. Sci. Comput., 40 (2018), pp. B1330–B1360, <https://doi.org/10.1137/18M116383X>.
- [25] C. ENGWER, T. HILLEN, M. KNAPPITSCH, AND C. SURULESCU, *Glioma follow white matter tracts: a multiscale dti-based model*, Journal of Mathematical Biology, 71 (2015), pp. 551–582, <https://doi.org/10.1007/s00285-014-0822-7>.
- [26] A. ERN AND J.-L. GUERMOND, *Theory and Practice of Finite Elements*, Applied Mathematical Sciences, Springer New York, 2004, <https://doi.org/10.1007/978-1-4757-4355-5>.
- [27] M. FRANK, H. HENSEL, AND A. KLAR, *A fast and accurate moment method for the Fokker-Planck equation and applications to electron radiotherapy*, SIAM J. Appl. Math., 67 (2006/07), pp. 582–603, <https://doi.org/10.1137/06065547X>.
- [28] K. GRELLA AND C. SCHWAB, *Sparse tensor spherical harmonics approximation in radiative transfer*, J. Comput. Phys., 230 (2011), pp. 8452–8473, <https://doi.org/10.1016/j.jcp.2011.07.028>.
- [29] W. HAN, Y. LI, Q. SHENG, AND J. TANG, *A numerical method for generalized Fokker-Planck equations*, in Recent advances in scientific computing and applications, vol. 586 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2013, pp. 171–179, <https://doi.org/10.1090/conm/586/11649>.
- [30] E. HEBEY, *Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities*, Courant lecture notes in mathematics, Courant Institute of Mathematical Sciences, 2000.
- [31] J. HENNING, D. PALITTA, V. SIMONCINI, AND K. URBAN, *Matrix oriented reduction of space-time Petrov-Galerkin variational problems*, in ENUMATH 2019 Proceedings, 2019. to appear.
- [32] P. HOUSTON, C. SCHWAB, AND E. SÜLI, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163, <https://doi.org/10.1137/S0036142900374111>.
- [33] P. HOUSTON AND E. SÜLI, *Stabilised hp-finite element approximation of partial differential equations with nonnegative characteristic form*, Computing, 66 (2001), pp. 99–119, <https://doi.org/10.1007/s006070170030>.
- [34] A. HUNT, *DTI-Based Multiscale Models for Glioma Invasion*, PhD thesis, TU Kaiserslautern, 2017, <https://nbn-resolving.org/urn:nbn:de:hbz:5:kluedo-53575>.
- [35] H. J. HWANG, J. JANG, AND J. JUNG, *The Fokker-Planck equation with absorbing boundary conditions in bounded domains*, SIAM J. Math. Anal., 50 (2018), pp. 2194–2232, <https://doi.org/10.1137/16M1109928>.
- [36] K. KORMANN, *A semi-Lagrangian Vlasov solver in tensor train format*, SIAM J. Sci. Comput., 37 (2015), pp. B613–B632, <https://doi.org/10.1137/140971270>.

- [37] O. LEHTIKANGAS, T. TARVAINEN, V. KOLEHMAINEN, A. PULKKINEN, S. ARRIDGE, AND J. KAIPIO, *Finite element approximation of the fokker-planck equation for diffuse optical tomography*, Journal of Quantitative Spectroscopy and Radiative Transfer, 111 (2010), pp. 1406 – 1417, <https://doi.org/10.1016/j.jqsrt.2010.03.003>.
- [38] J.-L. LIONS AND E. MAGENES, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
- [39] J. SCHAEFFER, *Convergence of a difference scheme for the Vlasov-Poisson-Fokker-Planck system in one dimension*, SIAM J. Numer. Anal., 35 (1998), pp. 1149–1175, <https://doi.org/10.1137/S0036142996302554>.
- [40] F. SCHNEIDER, G. ALLDREDGE, M. FRANK, AND A. KLAR, *Higher order mixed-moment approximations for the Fokker-Planck equation in one space dimension*, SIAM J. Appl. Math., 74 (2014), pp. 1087–1114, <https://doi.org/10.1137/130934210>.
- [41] C. SCHWAB AND R. STEVENSON, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp., 78 (2009), pp. 1293–1318, <https://doi.org/10.1090/S0025-5718-08-02205-9>.
- [42] C. SCHWAB, E. SÜLI, AND R. A. TODOR, *Sparse finite element approximation of high-dimensional transport-dominated diffusion problems*, ESAIM: M2AN, 42 (2008), pp. 777–819, <https://doi.org/10.1051/m2an:2008027>.
- [43] Q. SHENG AND W. HAN, *Well-posedness of the Fokker-Planck equation in a scattering process*, J. Math. Anal. Appl., 406 (2013), pp. 531–536, <https://doi.org/10.1016/j.jmaa.2013.04.063>.
- [44] K. URBAN AND A. PATERA, *An improved error bound for reduced basis approximation of linear parabolic problems*, Math. Comp., 83 (2014), pp. 1599–1615, <https://doi.org/10.1090/S0025-5718-2013-02782-2>.
- [45] G. WIDMER, R. HIPTMAIR, AND C. SCHWAB, *Sparse adaptive finite elements for radiative transfer*, J. Comput. Phys., 227 (2008), pp. 6071–6105, <https://doi.org/10.1016/j.jcp.2008.02.025>.
- [46] S. WOLLMAN AND E. OZIZMIR, *Numerical approximation of the Vlasov-Poisson-Fokker-Planck system in two dimensions*, Journal of Computational Physics, 228 (2009), pp. 6629 – 6669, <https://doi.org/10.1016/j.jcp.2009.05.027>.

UNIVERSITY OF MÜNSTER, APPLIED MATHEMATICS, EINSTEINSTR. 62, 48149 MÜNSTER, GERMANY, JULIA.BRUNKEN@UNI-MUENSTER.DE

UNIVERSITY OF TWENTE, FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS & COMPUTER SCIENCE, ZILVERLING, P.O. BOX 217, 7500 AE ENSCHEDE, THE NETHERLANDS. CURRENT ADDRESS: DEPARTMENT OF MATHEMATICAL SCIENCES, STEVENS INSTITUTE OF TECHNOLOGY, 1 CASTLE POINT TERRACE, HOBOKEN, NJ 07030, UNITED STATES OF AMERICA, KSMETANA@STEVENS.EDU.