Hypocoercivity-compatible finite element methods for the long-time computation of Kolmogorov’s equation

Emmanuil H. Georgoulis

October 12, 2018

Abstract

This work is concerned with the development of a family of Galerkin finite element methods for the classical Kolmogorov’s equation. Kolmogorov’s equation serves as a sufficiently rich, for our purposes, model problem for kinetic-type equations and is characterised by diffusion in one of the two (or three) spatial directions only. Nonetheless, its solution admits typically decay properties to some long time equilibrium, depending on closure by suitable boundary/decay-at-infinity conditions. A key attribute of the proposed family of methods is that they also admit similar decay properties at the (semi)discrete level for very general families of triangulations. The method construction uses ideas by the general theory of hypocoercivity developed by Villani [23], along with judicious choice of numerical flux functions. These developments turn out to be sufficient to imply that the proposed finite element methods admit a priori error bounds with constants independent of the final time, despite Kolmogorov equation’s degenerate diffusion nature. Thus, the new methods provably allow for robust error analysis for final times tending to infinity. The extension to three spatial dimensions is also briefly discussed.

1 Introduction

Degenerate parabolic problems in kinetic modelling are often characterised by the explicit presence of diffusion/dissipation in some of the spatial directions only, yet may still admit decay properties to some long time equilibrium. The development of Galerkin finite element methods for such problems satisfying provably also such decay properties is, remarkably, an unexplored area. This is despite the potential attraction that general, possibly highly non-uniform or adaptive, grid approximations can bring to kinetic simulations, especially those concerning expansive long-time simulation phenomena. Aiming to contribute in this direction, we shall develop and analyse a new class of Galerkin finite element methods for a simple kinetic equation, equipped with suitable closures by known boundary/periodicity conditions from the literature. To this end, we shall be predominantly concerned with the classical Kolmogorov’s partial differential equation (PDE)

\[ \mathcal{L}u \equiv u_t - u_{xx} + xu_y = f, \quad \text{in } (0, \infty) \times \Omega \subset \mathbb{R}^2, \]  

(1)
for suitably smooth forcing $f \in L_2(\mathbb{R}_+; H^1(\Omega))$, and Cauchy initial data $u(0, \cdot) = u_0 \in L_2(\Omega)$. The notational conventions used above are somewhat non-standard in kinetic theory, whereby $\mathcal{L}$ may be written as $\mathcal{L}f := f_t - f_{vv} + v f_x$ with $v$ denoting the particle velocity variable, $x$ the displacement/position and $f$ the respective probability density function. In Section 6 we shall also briefly discuss the extension of the developments in this work to the three-dimensional version of (1), namely

$$\mathcal{L}_3u \equiv u_t - u_{xx} + xu_y + yu_z = f, \quad \text{in } (0, \infty) \times \Omega \subset \mathbb{R}^3.$$  

(2)

Kolmogorov showed already in 1934 that (1) admits a smooth fundamental solution outside the pole and, hence, it is hypoelliptic, i.e., it admits a smooth solution in $(0, \infty) \times \mathbb{R}^2$ for any smooth initial and forcing data $u_0, f$, respectively [10]. This may be a surprising assertion at first sight, since there is no explicit dissipation built into the PDE in the $y$-variable. The advection field $(0, x)^T$, however, is “appropriately non-constant” so that it suffices to propagate the built-in dissipation onto the entire spatial domain. In 1967, Hörmander, in the celebrated work [14], gave a sufficient condition for hypoellipticity for a wide class of 2nd order operators with non-negative characteristic form, which are nowadays often referred to as Hörmander sum-of-squares operators; interestingly, a motivating example for the developments in [14] has, indeed, been Kolmogorov’s equation (1). Hence, although (1) may appear to be a rather special equation at first sight, its significance as a model problem is paramount, for it encompasses a number of pertinent structural properties of large classes of kinetic models; we refer, e.g., to [23] for very instructive expositions.

The proof of trend to equilibrium for the directly related inhomogeneous Fokker-Planck equation has been given by Hérau & Nier in [13] upon realising that certain entropies admitting mixed derivatives ($u_{xy}$ or, respectively, $f_{xy}$) give rise to full gradients in certain weighted Sobolev spaces, from which Poincaré-Friedrichs inequalities (also known as spectral gap properties in this context) are sufficient to prove decay to an equilibrium distribution. Related ideas have also been used by Eckmann & Hairer in [7] to study the spectral properties of certain hypoelliptic operators involving degenerate diffusions and by Mouhot & Neumann [18] in the study of kinetic models with integral-type collision operators, among others. The idea of using entropies involving mixed derivatives was elevated to a general framework in proving decay to equilibrium for kinetic equations by Villani [23] via the introduction of the concept of hypocoercivity. Roughly speaking, hypocoercivity is the property of certain degenerate elliptic or parabolic differential operators to yield dissipation of the solution also in the directions where no diffusion is explicitly present. Astonishingly, sufficient conditions for hypocoercivity are given by the rank spanned by Hörmander vector fields and their commutators with respect to the vector field related to 1st order term [23]. Thus, sufficient conditions for hypocoercivity are closely related, but not identical, to the classical Hörmander’s rank condition for hypoellipticity.

To fix ideas, denoting by $(\cdot, \cdot) \equiv (\cdot, \cdot)_{L_2(\Omega)}$ the inner product of $L_2(\Omega)$, and respective norm $\|\cdot\|^2 := \sqrt{(\cdot, \cdot)}$, and assuming decay of $u$ as $|(x, y)| \to \infty$, the classical energy-type analysis for (1) equation yields

$$\frac{1}{2} \frac{d}{dt} \|u(t)\|^2 + \|u_x\|^2_{L_2(\Omega)} = (f, u),$$  

(3)

for almost all $t \in (0, t_f]$, for some final time $t_f \in \mathbb{R}_+$, upon observing that the skew-symmetric term $(xu_y, u)$ vanishes. A combination of the Cauchy-Schwarz’ and Young’s inequalities on the last term on the right-hand side [3], therefore, yields

$$\int_0^{t_f} (f, u) \, dt \leq \frac{1}{2\delta} \|f\|^2_{L_2(0, t_f; L_2(\Omega))} + \frac{\delta}{2} \|u\|^2_{L_2(0, t_f; L_2(\Omega))},$$  

(4)
for any $\delta > 0$. Therefore, Grönwall’s Lemma implies
\[ \|u(t_f)\|^2 \leq e^{\delta t_f} \left( \delta^{-1} \|f\|_{L^2(0,t_f;L^2(\Omega))}^2 + \|u_0\|^2 \right). \]

Since the right-hand side of (5) is always greater than or equal to $\|u_0\|^2$ for every $f \in L^2(0,\infty;L^2(\Omega))$, we conclude that the basic energy estimate may be severely pessimistic as $t_f \to \infty$ for any fixed $\delta > 0$: for any finite final time $t_f > 0$, the above implies that the right-hand side of (5) will be positive away from zero for any non-trivial initial condition. Notice that when $f = 0$, we can take $\delta \to 0$, yielding the stability estimate $\|u(t_f)\| \leq \|u_0\|$. This is at odds with the fast decay to 0 observed for this problem by the solution of Kolmogorov’s equation for the respective homogeneous problem, at least when $\Omega = \mathbb{R}^2$. Evidently the key culprit in the above analysis is the absence of control in a time-integrated norm on the left-hand side of (5).

The situation does not appear to improve if different versions of Grönwall-type inequalities are applied instead.

On the other hand, for the non-degenerate parabolic equation
\[ \tilde{L}u \equiv u_t - u_{xx} - u_{yy} + xu_y = f, \quad \text{in } (0,\infty) \times \Omega \subset \mathbb{R}^2, \]
with the same forcing and initial conditions, we can easily deduce the energy identity
\[ \frac{1}{2} \frac{d}{dt} \|u(t)\|^2 + \|\nabla u\|_{L^2(\Omega)}^2 = (f,u). \]

Assuming now the validity of a Poincaré-Friedrichs inequality of the form $\|u\|^2 \leq C_{PF}\|\nabla u\|^2$, and working as before, we can arrive at the estimate
\[ \frac{1}{2} \frac{d}{dt} \|u(t)\|^2 + C_{PF}^{-1}\|u\|_{L^2(\Omega)}^2 \leq (f,u), \]
from which, Grönwall’s Lemma implies
\[ \|u(t_f)\|^2 \leq e^{-C_{PF}^{-1}t_f}\|u_0\|^2 + C_{PF} \int_0^{t_f} e^{-C_{PF}^{-1}(t_f-t)}\|f(t)\|_{L^2(\Omega)}^2 \, dt, \]
whose right-hand side clearly decays exponentially as $t_f \to \infty$ when $f = 0$, for instance. This is in sharp contrast with (5).

The above is obviously not a satisfactory state of affairs not only at the PDE level, but also under the prism of the numerical analysis for these problems. Indeed, almost all provably convergent, general numerical methods which can be potentially applied to this class of PDEs use such standard energy-type arguments for their error analysis, resulting to possibly severely pessimistic a priori and/or a posteriori error bounds for large final times $t_f$. Indeed, this is the state of affairs discussed in various works on Galerkin-type numerical methods for (degenerate) second order evolution problems of diffusion type; we non-exhaustively refer to [1, 15, 8, 3] as representatives of different approaches.

To the best of our knowledge, there exist only few, yet quite inspiring, exceptions which employ hypocoercivity ideas in the context of numerical methods. In particular, the work of Porretta & Zuazua [20] proves the hypocoercivity of classical finite difference operators discretising Kolmogorov’s equation for $f = 0$ on uniform spatial grids over $[0,\infty) \times \mathbb{R}^2$. The recent manuscript by Dujardin, Herau, & Lafitte [6] takes the approach of [20] further by establishing discrete versions of weighted Poincaré inequalities for difference operators, thereby showing decay to equilibrium for finite difference approximations of the inhomogeneous two-dimensional Fokker-Planck equation, which is, of course, closely related to Kolmogorov’s equation. Both
these approaches apply directly the abstract semigroup result of Villani [23, Theorem 18] to the finite difference operators used by proving the necessary commutator properties required for its validity. With regard to practical approaches, Foster, Lohéac & Tran [10] discuss the development a Lagrangian-type splitting method based on a carefully constructed similarity transformation and linear finite elements over quasiuniform meshes, although no proof of decay to equilibrium is given for the numerical method itself. Also, Bessemoulin-Chatard & Filbet [2] present design principles for the construction of equilibrium-preserving finite volume methods for nonlinear degenerate parabolic problems.

For practical computations, one is typically forced to confine the spatial computational domain into an open and bounded one, say $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. If the objective is to compute equilibrium states for the whole of $\mathbb{R}^d$, the truncation of the computational domain and the imposition of boundary conditions should be performed carefully, for it may alter the rate of decay to equilibrium compared to the non-confined problem [10]. If one is interested instead in computations on bounded domains as such, then the modelling often dictates the imposition of additional boundary conditions/constraints, such as periodicity or far field decay. As we shall see below, the imposition of boundary conditions in a fashion allowing concepts of hypocoercivity to be ported to bounded spatial domains is a challenge both in the PDE and the numerical analysis contexts. 2

This work is concerned with the development of a Galerkin-type finite element method which is compatible with hypocoercivity concepts and, therefore, allows for the proof of error bounds whose error constants do not grow as the final time $t_f \to \infty$, such as the ones in (5) do. Thus, as we shall show below, there will be no dependence of the error constant on the final time $t_f$. This, in turn, supports the potential advantages of the numerical methods presented below for long time computations. The construction of the finite element method is based on a variational interpretation of the modified entropies involving mixed derivatives of Hérau & Nier [13] and Villani [23], which are able to reveal the implicit diffusion properties in the $y$-variable also. The derived modified variational problem is based on the PDE and higher order laws stemming from the PDE and involving up to 4th order partial derivatives, when viewed formally in strong form. The finite element method is constructed via a Galerkin projection onto a suitable $C^0$-finite element space subordinate to a non-uniform, in general, triangulation. The proposed method is, nevertheless, non-conforming due to the 4th order symbol of the variational problem. To recover, therefore, the consistency of the finite element method without compromising on the hypocoercivity properties, the variational form has to be enriched by carefully constructed consistent numerical fluxes over the skeleton of the triangulation. The proposed finite element method is proven to decay to equilibrium when $f = 0$, when we equip the PDE with additional natural boundary conditions; this is, to the best of our knowledge, the first manifestation of hypocoercivity for a Galerkin scheme in the literature. Moreover, we prove a priori error bounds in the natural norm determined by the hypocoercivity result.

We view the developments presented below as proof of concept that Galerkin finite element methods over general non-uniform grids can be constructed in a compatible fashion to retain hypocoercivity properties at discrete level. This, in turn, leads to a priori error bounds which are robust with respect to their dependence on the final time $t_f$. Kolmogorov’s equation has the role of a sufficiently rich model problem to highlight a new finite element methodology for hypocoercive PDEs on bounded computational domains with view to retaining the long-time behaviour of the exact initial/boundary value problems. Apart from the mathematical challenge, the development of numerical methods for kinetic equations on general, possibly non-quasi-uniform meshes has important practical ramifications. Indeed, modern applications of kinetic equations involve the movement of large particles for which Lagrangian schemes appear to be preferential by practitioners which greatly benefits from the ability to discretise over
unstructured, possibly moving, meshes.

The remainder of this work is structured as follows. In Section 2, we introduce a variational formulation motivated by design concepts of hypocoercivity \cite{23}, which is shown to be coercive with respect to an $H^1$-equivalent norm in Section 3. A discrete bilinear form using continuous, yet non-conforming, elements with appropriately constructed numerical fluxes ensuring the coercivity in the discrete setting also is proposed in Section 4 and is shown to admit completely analogous decay to equilibrium properties as the respective continuous problem in the non-external forcing scenario. The error analysis of the proposed finite element method is given in Section 5 along with a discussion on its properties. To highlight the potential of the proposed framework for the class of kinetic equations with degenerate diffusion, we briefly discuss the extension of the developments in the first five sections to the case of the three-dimensional (in space) version of Kolmogorov’s equation; this is the content of Section 6.

2 A special weak formulation

The problem of closing a degenerate parabolic PDE with suitable boundary conditions is well understood via the, now classical, theory of linear second order equations with non-negative characteristic form \cite{9, 19}. In particular, with $n(\cdot) := (n_1(\cdot), n_2(\cdot))^T$ denoting the unit outward normal vector at each point of $\partial \Omega$, we define the elliptic boundary

$$\partial_0 \Omega := \{(x, y) \in \partial \Omega : n_1(x, y) \neq 0\},$$

along with the inflow and outflow parts of the non-elliptic boundary:

$$\partial_- \Omega := \{(x, y) \in \partial \Omega \setminus \partial_0 \Omega : xn_2(x, y) < 0\}, \quad \partial_+ \Omega := \{(x, y) \in \partial \Omega \setminus \partial_0 \Omega : xn_2(x, y) > 0\}.$$

Introducing, now, the notation $Lu \equiv -u_{xx} + xu_y$, we consider the initial/boundary-value problem:

$$u_t + Lu = u_t - u_{xx} + xu_y = f, \quad \text{in } (0, t_f] \times \Omega,$$

$$u = u_0, \quad \text{on } \{0\} \times \Omega,$$

$$u = 0, \quad \text{on } (0, t_f] \times \partial_- \Omega,$$

with $\partial_- \Omega := \partial_- \Omega \cup \partial_0 \Omega$, for $t_f > 0$ and for $f \in H^1(\Omega)$; note the non-standard regularity assumption for $f$. The well-posedness of the above problem is assured upon assuming that $\partial_- \Omega$ has positive one-dimensional Hausdorff measure \cite{19}. It is also possible to impose Neumann-type boundary conditions on parts of $\partial_0 \Omega$; this is not done here in the interest of simplicity of the presentation only. For the three-dimensional counterpart of the above model problem we refer to Section 6.

Let $A := \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ a symmetric and non-negative definite matrix with $\alpha, \beta, \gamma$ non-negative parameters, and set

$$\delta := 1 - \frac{\beta}{\sqrt{\alpha \gamma}};$$

the fact that $A$ is non-negative definite implies that $0 \leq \delta \leq 1$. Assuming sufficient regularity for the exact solution $u$ for the moment, so that the following calculations are well defined, \cite{11} implies $\nabla u_t + \nabla Lu = \nabla f$, which, tested against $A \nabla v$ for any $v \in H^1(\Omega)$, results to

$$(\nabla u_t, A \nabla v) + (\nabla Lu, A \nabla v) = (\nabla f, A \nabla v);$$
here and below, we denote $(\cdot, \cdot) \equiv (\cdot, \cdot)_{\mathcal{L}_2(\Omega)}$ and $\| \cdot \| \equiv \| \cdot \|_{\mathcal{L}_2(\Omega)}$ for brevity. Resorting to (11) once more, after an integration by parts, we deduce the variational form

$$(u_t, v) + (\nabla u_t, A \nabla v) + (u_x, v_x) + (x u_y, v) + (\nabla L u, A \nabla v) = (f, v) + (\nabla f, A \nabla v),$$

(12)

for all $v \in H^1(\Omega) \equiv \{v \in H^1(\Omega) : v|_{\partial_\nu \partial \Omega} = 0\}$. The latter will be the basis of constructing a finite element method with the sought-after properties. To highlight the idea behind this construction in the context of hypocoercivity, we shall now discuss the properties of the spatial part of the operator on the left-hand side of (12), viz.,

$$a(u, v) \equiv (u_x, v_x) + (x u_y, v) + (\nabla L u, A \nabla v).$$

(13)

3 The hypocoercivity of $a(\cdot, \cdot)$

We begin by somewhat modifying the approach presented in Villani [23] arriving, nevertheless, to an effectively similar result. For accessibility, we shall first revert to (1), rather than (10), i.e., we have $\Omega = \mathbb{R}^2$, assuming that $|u(x, y)| \rightarrow 0$ as $|(x, y)| \rightarrow \infty$; the case $\Omega \subset \mathbb{R}^2$ will be discussed below. When the domain is clear by the context, we shall drop the subscript in the $L_2(\Omega)$-norm, viz., $\| \cdot \|_{L_2(\Omega)} \equiv \| \cdot \|$, for brevity. Setting $v = u$ into (13) gives

$$a(u, u) = \|u_x\|^2 + (x u_y, u) + (-u_{xxx} + (x u_y)_x, \alpha u_x + \beta u_y) + (-u_{xxy} + x u_{yy}, \beta u_x + \gamma u_y),$$

which, upon integration by parts, already yields, respectively,

$$a(u, u) = \|u_x\|^2 + (u_{xx}, \alpha u_{xx} + \beta u_{yy}) + ((x u_y)_x, \alpha u_x + \beta u_y)
\quad + (u_{xy}, \beta u_{xx} + \gamma u_{xx}) - (x u_y, \beta u_{xy} + \gamma u_{yy})
\quad = \|u_x\|^2 + \alpha \|u_{xx}\|^2 + \beta \|u_y\|^2 + \gamma \|u_{xy}\|^2 + \alpha (u_y, u_x) + 2\beta (u_{xy}, u_{xx}),$$

noting that $(x u_{xy}, u_x) = 0 = (x u_y, u_{yy})$ from divergence theorem and the decay properties of $u$ as $|(x, y)| \rightarrow \infty$. Young’s and Cauchy-Schwarz’ inequalities now imply

$$a(u, u) \geq (1 - \epsilon) \|u_x\|^2 + (\alpha - \sqrt{\beta}) \|u_{xx}\|^2 + (\beta - \frac{\alpha^2}{4\epsilon}) \|u_y\|^2 + (\gamma - \beta \frac{\alpha}{2\epsilon}) \|u_{xy}\|^2.$$

Selecting $\alpha > 0$, $\beta = 4\alpha^2/9$, $\gamma = \alpha^3/3$, and $\epsilon = 3/4$, for instance, we deduce

$$a(u, u) \geq \frac{1}{4} \|u_x\|^2 + \frac{\alpha}{3} \|u_{xx}\|^2 + \frac{\alpha^2}{9} \|u_y\|^2 + \frac{\alpha^3}{27} \|u_{xy}\|^2.$$  

(14)

Another instructive choice is $\alpha > 0$, $\beta = \alpha^2$, $\gamma = \alpha^3$ and $\epsilon = 1/2$, giving

$$2a(u, u) \geq \|u_x\|^2 + \alpha^2 \|u_y\|^2.$$  

(15)

Thus, remarkably, using the non-standard test function $v = -\nabla \cdot (A \nabla v)$, results into $L$ providing coercivity with respect to $u_y$ also! This is a manifestation of the concept of hypocoercivity. As briefly discussed above, the mechanics of this idea can be traced back to the celebrated work of Hörmander [14] for sufficient conditions for hypoellipticity for second order linear operators: the first order term is “appropriately non-constant” with respect to the independent variables $t, x, y$, allowing for control also in the commutator vector field $[\partial_x, x \partial_y] = \partial_y$. Informally speaking, the first order vector field in conjunction with the direction of $\partial_x$ is able to “propagate” the $x$-variable dissipation of $L$ to the whole of $\mathbb{R}^2$, resulting to the property of hypocoercivity [23]. Notice that the choice of $\alpha, \beta, \gamma$ yielding (14) corresponds to $v = -\nabla \cdot (A \nabla v)$ being an elliptic
differential operator, while the respective choice for (15) corresponds to the parabolic operator \( v - \nabla \cdot (A \nabla v) \).

A comment on the different scalings present on each term on the right-hand side of (14) is also in order. It is possible to introduce dependence of the tensor \( A \) on the time variable \( t \) for the study of the singularities near \( t = 0 \); we refer to [12, 20] for details. Since our interest lies in the long time provable computation robustness for long times, we shall not pursue this direction further here.

On the other hand, kinetic PDE models posed on bounded domains \( \Omega \subset \mathbb{R}^d \), are often complemented with additional boundary /compatibility conditions to convey the physical properties in the vicinity of boundaries, e.g., reflection, diffusion, periodicity, etc. In the case of the classical Kolmogorov’s equation considered here, the \( x \)-variable is typically associated with particle velocity, while the \( y \)-variable models position/displacement. Performing the above calculations on a bounded domain \( \Omega \) instead, integrations by parts yield

\[
a(u, u) = \|u_x\|^2 + \frac{1}{2}\|\sqrt{xn_2u}\|_{\partial_+ \Omega}^2 + \|\sqrt{A} \nabla u_x\|^2 + \beta \|u_y\|^2 + \alpha(u_y, u_x)
\]

(16)

A sufficient condition, therefore, to arrive again at (14) or (15) is to assert that

\[
\frac{1}{2}xn_2 \nabla u - n_1 \nabla u_x, A \nabla u)_{\partial_\Omega} \geq 0,
\]

(17)
or, at least, that part of the left-hand side of (17) can be controlled by the non-negative terms on the right-hand side of (16) and the remaining to satisfy a respective positivity condition.

For example, assuming that \( \Omega := (x^-, x^+) \times (y^-, y^+) \), for \( x^- \geq 0 \), we have

\[
\partial_x \Omega = \{(x, y) : x^- < x \leq x^+\}, \quad \partial_0 \Omega = \{(x^-, y), (x^+, y) : y^- \leq y < y^+\}.
\]

Therefore, assuming \( H^2(\Omega) \)-smoothness of the solution, imposing additional periodic boundary conditions on \( \partial_0 \Omega \), we deduce

\[
\frac{1}{2}xn_2 \nabla u - n_1 \nabla u_x, A \nabla u)_{\partial_\Omega} = 0,
\]

(18)
since \( xn_2 \) on \( \partial_0 \Omega \). Supplementing also with a no-flux boundary conditions on the inflow boundary \( \partial_- \Omega \), viz.,

\[
n_2 u_y = 0 \quad \text{on} \quad \partial_- \Omega,
\]

(19)
we conclude

\[
\frac{1}{2}xn_2 \nabla u - n_1 \nabla u_x, A \nabla u)_{\partial_0 \cup \partial_+ \Omega} = \frac{1}{2}(xn_2 \nabla u, A \nabla u)_{\partial_0 \cup \partial_+ \Omega} = 0,
\]
since \( \nabla u = (u_x, u_y)^T = 0 \) and \( n_1 = 0 \) on \( \partial_\pm \Omega \). The periodicity in the \( x \)-variable and (19), therefore, implies

\[
\frac{1}{2}xn_2 \nabla u - n_1 \nabla u_x, A \nabla u)_{\partial_\Omega} = (xn_2 \nabla u, A \nabla u)_{\partial_+ \Omega} = \|\sqrt{xn_2 A} \frac{\nabla u}{2}\|_{\partial_+ \Omega}^2 \geq 0.
\]

We refer, e.g., to [21, 6] for further discussion on additional boundary conditions. Different combinations of assumptions are also possible to yield (17), such as vanishing second derivative
conditions across the elliptic boundary, etc. For the remaining of this work, we shall assume the validity of the additional boundary conditions:

\[
\left(\frac{1}{2}x_n \nabla u - n_1 \nabla u_x\right)|_{\partial \Omega \setminus \partial \omega, \Omega} = 0 \quad \text{and} \quad n_1 \nabla u_x|_{\partial \omega, \Omega} = 0,
\]

which trivially implies (17). We do so, instead of prescribing particular additional boundary conditions; we shall clearly state if the validity of (20) is assumed or not for the proof of various results below. We stress that, under minor modifications only, the theory presented below can still be valid for other suitable assumptions also by careful inspection of the proofs. This is not pursued here in the interest of a unified and clearer presentation.

4 A finite element method

Motivated by the above discussion, we return to the variational representation (12) of problem (10) on a bounded domain \( \Omega \subset \mathbb{R}^2 \). For simplicity of the presentation we, henceforth, assume that \( \Omega \) is polygonal also; extension to curved domains is possible via isoparametric versions of the finite element spaces discussed below. As is standard in discretisation by a Galerkin type finite element method, we introduce a family of triangulation of \( \Omega \), say \( \mathcal{T} \), consisting of mutually disjoint open triangular elements \( T \in \mathcal{T} \), whose closures cover \( \bar{\Omega} \) exactly. Let also \( h : \cup_{T \in \mathcal{T}} T \to \mathbb{R}^+ \) be the local meshsize function defined elementwise by \( h_T := h_T := \text{diam}(T) \).

For simplicity, are further assume that \( \mathcal{T} \) to be shape-regular, in the sense that the radius \( \rho_T \) of the largest inscribed circle of each \( T \in \mathcal{T} \) is bounded with respect to each element’s diameter \( h_T \), uniformly as \( \|h\|_{L^\infty(\Omega)} \to 0 \) under mesh refinement. Also, we assume that \( \mathcal{T} \) is locally quasi-uniform in the sense that the diameters of adjacent elements are uniformly bounded from above and below. Finally, let \( \Gamma := \cup_{T \in \mathcal{T}} \partial T \) denote the mesh skeleton, and \( \Gamma_{\text{int}} := \Gamma \setminus \partial \Omega \).

We define the finite element space subordinate to \( \mathcal{T} \) by

\[
V_h \equiv V_h^p := \{ V \in H_{-\text{div}}^1(\Omega) : V|_{T} \in \mathbb{P}_p(T), \ T \in \mathcal{T} \},
\]

with \( \mathbb{P}_p(\omega) \), \( \omega \subset \mathbb{R}^2 \), denoting the space of polynomials of total degree at most \( p \) over \( \omega \), \( p = 2, 3, \ldots \). Further, we define the *jump* \( \{w\} \) of \( w \) across the common element interface \( e := \partial T^+ \cap \partial T^- \) of two adjacent elements \( T^+, T^- \), by \( \{w\}_e := w|_{T^+} n_{T^+} + w|_{T^-} n_{T^-} \) with \( n_T \) denoting the unit outward normal vector to each point of \( \partial T \); correspondingly, we also define the *average* \( \{w\}_e := (w|_{T^+} + w|_{T^-})/2 \) of \( w \) across \( e \). Finally, for boundary faces \( e \subset \partial T^+ \cap \partial \Omega \), we set \( \{w\}_e := w|_{T} n_{T} \) and \( \{w\}_e := w|_{T} \), respectively. The above jump and average definitions are trivially extended to vector-valued functions by component-wise application. We also define the, so-called, broken Sobolev spaces \( H^r(\Omega, \mathcal{T}) \), \( r > 0 \), subordinate to the mesh \( \mathcal{T} \) by \( H^r(\Omega, \mathcal{T}) := \{ v|_{T} \in H^r(T), T \in \mathcal{T} \} \), with respective norm \( \|v\|_{H^r(\Omega, \mathcal{T})} := \sum_{T \in \mathcal{T}} \|v\|_{H^r(T)} + \|\{v\}\|_{\Gamma_{\text{int}}} \). Finally, by \( \nabla_T \), we shall denote the element-wise broken gradient.

Starting from (12), we consider the spatially discrete finite element method: for every \( t \in (0, t_f] \), find \( U \equiv U(t) \in V_h \), such that

\[
(U_t, V) + (U_x, V_x) + (x U_y, V) \\
+(\nabla U_t, A \nabla V) + (A \nabla U_x, \nabla V_x) + (\nabla_T (x U_y), A \nabla V) + s_h(U, V) \\
= (f, V) + (\nabla f, A \nabla V),
\]

(21)
for all $V \in V_h$, whereby
\[
sh(U, V) := -\int_{\Gamma_{\text{int}} \cup \partial-\Omega} (0, x)^T \cdot \left( \alpha [U_x] [V_x] + \beta [U_y] [V_y] + \gamma [U_y] [V_y] \right) ds \\
+ \int_{\Gamma_{\text{int}}} \frac{|x n_2|}{2} \left( \kappa [U_x] \cdot [V_x] + \lambda [U_y] \cdot [V_y] \right) ds \\
- \int_{\Gamma_{\text{int}} \cup \partial-\Omega} \left( \{ A \nabla U_x \} \cdot [\nabla V]_1 + \{ A \nabla V_x \} \cdot [\nabla U]_1 - \tau [\nabla U]_1 \cdot A [\nabla V]_1 \right) ds,
\]
with $[v]_e := (v|_e + n_1^e) + (v|_e - n_1^e)$, for each internal face $e$ and $n_1^e$ denoting the first component of the unit normal vector to $e$ and $[v]_e := (v|_e + n_1^e)$, for $e \in \partial\Omega$, for some $\tau : \Gamma \rightarrow \mathbb{R}$ non-negative function to be determined precisely below and user-defined parameters $\kappa, \lambda \geq 0$; we also set $U(0) := \Pi u_0 \in V_h$, with $\Pi : L_2(\Omega) \rightarrow V_h$ denoting the orthogonal $L_2$ projection onto the finite element space. For brevity, we shall use the notation:
\[
s_{1,h}(w, v) := \int_{\Gamma_{\text{int}}} \frac{|x n_2|}{2} \left( \kappa [w_x] \cdot [v_x] + \lambda [w_y] \cdot [v_y] \right) ds.
\]
We also observe that the first term of $s_h$ vanishes for $(x, y)^T \in \partial\Omega \setminus (\partial-\Omega \cap \partial+\Omega)$ since $x_n = 0$ there. As we shall see below the stabilisation $s_{1,h}(\cdot, \cdot)$ is constructed so that in the method \[21\] is consistent with respect to \[12\] upon assuming the additional boundary conditions \[20\].

**Lemma 4.1.** Let $\tau := \Gamma \rightarrow \mathbb{R}$ with $\tau = C_\tau p^2 / \|h\|$, for $C_\tau > 0$ large enough but independent of $h$ and of $p$. Then, there exists a positive constant $c_0$, independent of the approximate solution $U \in V_h$ and of $t_f$, such that for any $0 < \epsilon, \zeta < 1$, we have
\[
\|U(t_f)\|^2 + \|\sqrt{A} \nabla U(t_f)\|^2 + (1 - \zeta) \int_0^{t_f} \left( c_0 \|U\|^2 + \|\sqrt{A} \nabla U\|^2 \right) dt \\
+ \int_0^{t_f} \left( \|\sqrt{x n_2 U}\|_{\partial+\Omega}^2 + \|\sqrt{x n_2 A \nabla U}\|_{\partial+\Omega}^2 \\
+ \|\sqrt{A} \nabla T U_x\|^2 + s_{1,h}(U, U) + \|\sqrt{\tau A} [\nabla U]_1\|_{\Gamma_{\text{int}} \cup \partial-\Omega}^2 \right) dt \leq \frac{4}{\zeta} \int_0^{t_f} \left( c_0^{-1} \|f\|^2 + \|\sqrt{A} \nabla f\|^2 \right) dt + \|\Pi u_0\|^2 + \|\sqrt{A} \nabla \Pi u_0\|^2;
\]
when $f \equiv 0$, we can trivially select $\zeta = 0$.

**Proof.** Setting $V = U$ in \[21\], standard arguments give
\[
\frac{1}{2} \frac{d}{dt} \left( \|U\|^2 + \|\sqrt{A} \nabla U\|^2 \right) + \|U_x\|^2 + \frac{1}{2} \|\sqrt{x n_2 U}\|_{\partial+\Omega}^2 \\
+ \|\sqrt{A} \nabla T U_x\|^2 + (\nabla T(x U y), A \nabla U) + s_h(U, U) \\
= (f, U) + (\nabla f, A \nabla U),
\]
upon observing that $(x U y, U) = \|\sqrt{x n_2 U}\|_{\partial+\Omega}^2 / 2$ as $U = 0$ on $\partial-\Omega$. We calculate
\[
(\nabla (x U y), A \nabla U)_T = \alpha (U_y, U_x)_T + \beta |U_y|^2 + \alpha (U_x, x U y)_T + \beta (U_y, x U y)_T \\
+ \beta (x U y, U_x)_T + \gamma (x U y, U y)_T.
\]
Observing now the identities
\[
\alpha (U_x, x U y)_T = \frac{\alpha}{2} \int_{\partial T} x n_2 U_x^2 ds, \quad \gamma (U_y, x U y)_T = \frac{\gamma}{2} \int_{\partial T} x n_2 U_y^2 ds,
\]

\[9\]
\[
\beta(xU_{yy}, U_x)_T = -\beta(xU_y, U_{xy})_T + \beta \int_{\partial T} x n_2 U_x U_y \, ds,
\]

Then, summing over all \( T \in \mathcal{T} \), we deduce

\[
(\nabla_T(xU_y), A\nabla U) = \alpha(U_y, U_x) + \beta\|U_y\|^2 + \frac{1}{2} \int_{\Gamma} (0, x)^T \cdot [\alpha U_x^2 + 2\beta U_x U_y + \gamma U_y^2] \, ds.
\] (24)

Using (24) into (25), noting that \( \|WV\| = \{W\} \|V\| + \{V\} \|W\| \) on \( \Gamma_{\text{int}} \) for \( V, W \in V_h \), and integrating with respect to \( t \in [0, t_f] \), therefore, gives

\[
\|U(t_f)\|^2 + \|\sqrt{A} \nabla U(t_f)\|^2 + 2 \int_0^{t_f} \|\sqrt{A} \nabla_T U_x\|^2 \, dt + \int_0^{t_f} \|\nabla_T n_2 U\|^2_{\partial_\omega, \Omega} \, dt
\]

\[
+ \int_0^{t_f} \int_{\partial_\omega, \Omega} x n_2 (\alpha U_x^2 + 2\beta U_x U_y + \gamma U_y^2) \, ds \, dt + 2 \int_0^{t_f} \left( \|U_x\|^2 + \alpha(U_y, U_x) + \beta\|U_y\|^2 \right) \, dt
\]

\[
- 4 \int_0^{t_f} \int_{\Gamma_{\text{int}}} \|A \nabla U_x\| [\nabla U]_1 \, ds \, dt + 2 \int_0^{t_f} \left( \|\sqrt{\tau A}[\nabla U]_1\|_{\Gamma_{\text{int}}} + s_{1,h}(U, U) \right) \, dt,
\]

\[
= 2 \int_0^{t_f} \left( (f, U) + (\nabla f, A \nabla U) \right) \, dt + \|\Pi u_0\|^2 + \|\sqrt{A} \nabla \Pi u_0\|^2.
\] (25)

Focusing on the fifth and sixth terms on the left-hand side of (25), recalling (11), the Cauchy-Schwarz' and Young’s inequalities along with elementary manipulations imply

\[
\int_{\partial_\omega, \Omega} x n_2 (\alpha U_x^2 + 2\beta U_x U_y + \gamma U_y^2) \, ds = \int_{\partial_\omega, \Omega} x n_2 |\sqrt{A} \nabla U|^2 \, ds \geq 0,
\] (26)

and

\[
2(\|U_x\|^2 + \alpha(U_y, U_x) + \beta\|U_y\|^2) \geq 2(1-\epsilon)\|U_x\|^2 + (2\beta - \frac{\alpha^2}{2\epsilon})\|U_y\|^2 = \|\sqrt{B} \nabla U\|^2,
\] (27)

respectively, upon defining the diagonal matrix \( B = \text{diag}(2(1-\epsilon), 2\beta - \alpha^2/(2\epsilon)) \), for any \( 0 < \epsilon < 1 \). We, then, have

\[
\|\sqrt{B} \nabla U\|^2 = \|\sqrt{B - A} \nabla U\|^2 + \|\sqrt{A} \nabla U\|^2 \geq \lambda_{\text{min}}^{B-A}\|\nabla U\|^2 + \|\sqrt{A} \nabla U\|^2,
\]

where \( \lambda_{\text{min}}^{B-A} \) denotes the smallest eigenvalue of \( B - A \). Since \( U = 0 \) on \( \partial_\Omega \) which was assumed to be of positive one-dimensional Hausdorff measure, the validity of a Poincaré-Friedrichs/spectral gap inequality \( \|U\| \leq C_{PF}\|\nabla U\| \) with \( C_{PF} \equiv C_{PF}(\Omega) \) positive constant, independent of \( U \), is assured, thereby allowing us to conclude the lower bound

\[
\|\sqrt{B} \nabla U\|^2 \geq C_{PF}^{-1}\lambda_{\text{min}}^{B-A}\|U\|^2 + \|\sqrt{A} \nabla U\|^2.
\] (28)

Also, the symmetry of \( A \) along with the Cauchy-Schwarz inequality and a standard inverse estimate yield, respectively,

\[
\int_{\Gamma_{\text{int}} \cup \partial_\Omega} \{A \nabla U_x\} \cdot [\nabla U]_1 \, ds \leq \|\varphi^{-1}\{\sqrt{A} \nabla U_x\}\|_{\Gamma_{\text{int}} \cup \partial_\Omega} \|\varphi \sqrt{A}[\nabla U]_1\|_{\Gamma_{\text{int}} \cup \partial_\Omega}
\]

\[
\leq \|\sqrt{C_{\text{inv}}^{PF}/h_{\text{T}}} \sqrt{A} \nabla_T U_x\| \|\varphi \sqrt{A}[\nabla U]_1\|_{\Gamma_{\text{int}} \cup \partial_\Omega}
\]

\[
\leq \|\sqrt{A} \nabla_T U_x\|^2 + \|\sqrt{\tau A}[\nabla U]_1\|^2_{\Gamma_{\text{int}} \cup \partial_\Omega}
\] (29)

using the standard inverse inequality \( \|v\|^2_{\partial T} \leq C_{\text{inv}}^{PF}/h_{\text{T}}\|v\|^2 \) on (29), and selecting \( \varphi^2 = \tau \geq 2C_{\text{inv}}^{PF}/\|h\| \) and resorting to the local quasuniformity of the mesh. Inserting the last two bounds into (25) and, using the Cauchy-Schwarz inequality, setting \( c_0 := C_{PF}^{-1}\lambda_{\text{min}}^{B-A} \), we deduce the result for any \( 0 < \zeta < 1 \); when \( f \equiv 0 \) we can trivially select \( \zeta = 0 \). \( \square \)
Remark 4.2. Setting $\beta = \alpha^2 \gamma = \alpha^3$, and $\epsilon = 1/2$, we have $\det(B - A) = \alpha^2(1 - 2\alpha)$. A numerical investigation reveals that $\max_{0 \leq \alpha \leq 1/2} \lambda_{\min}^{\text{B-A}} \approx 0.054429$ attained for a value $\alpha \approx 0.35060$, noting that $B - A$ has positive determinant for $0 < \alpha < 1/2$. It is possible to optimise further the constant $c_0$ by working as follows: starting from $\|B \nabla U\|^2 = \|B - \nu A \nabla U\|^2 + \nu \|\nabla U\|^2 \geq \lambda_{\min}^{\text{B-\nu A}} \|\nabla U\|^2 + \nu \|\nabla U\|^2$ and selecting $c_0 = \lambda_{\min}^{\text{B-\nu A}} C_{PF}^{-1}$ for suitable $\nu > 0$. We will revisit this idea in Section 6.

The last proof highlights that the validity of the Poincaré-Friedrichs inequality is of central importance here, and in general in the study of trend to equilibrium for kinetic equations [21, 13, 23]. In the present context of bounded domain spatial $\Omega$, we have taken the viewpoint of using standard (non-weighted) integral norms, whose measure is present in the constant $c_0$. The extension of the above framework to general Fokker-Planck equations on function spaces weighted by equilibrium distributions is an important question and will be discussed elsewhere. The availability of such a Poincaré-Friedrichs inequality (also known as spectral gap property in kinetic theory) facilitates the crucial feature of the proposed method: the absence of an exponential term of the form $\exp(\epsilon t f)$ multiplying the data terms in the above stability estimate. As a consequence, (21) also immediately implies the decay of numerical solutions for the respective homogeneous problem.

Theorem 4.3 (Decay via hypocoercivity). The finite element method (21) for the homogeneous problem (10) with $f \equiv 0$ satisfies

$$\|U(t_f)\| + \|\sqrt{A} \nabla U(t_f)\| \leq e^{-\min\{1, \epsilon\} t_f} \left(\|\Pi u_0\| + \|\sqrt{A} \nabla \Pi u_0\|\right),$$

for all $t_f > 0$.

Proof. From (22) we deduce

$$\|U(t_f)\|^2 + \|\sqrt{A} \nabla U(t_f)\|^2 \leq \|\Pi u_0\|^2 + \|\sqrt{A} \nabla \Pi u_0\|^2,$$

from which an application of Grönwall’s Lemma already implies the result. \qed

5 Error analysis

We continue by proving a priori error bounds for (21). For brevity, we shall denote by $B(\cdot, \cdot) : (H^{5/2+\varepsilon}(\Omega, \mathcal{T}) + V_h) \times (H^{5/2+\varepsilon}(\Omega, \mathcal{T}) + V_h) \to \mathbb{R}$, $\varepsilon > 0$, the bilinear form given by

$$B(w, v) := \langle w_x, v_x \rangle + \langle \nabla_x w_y, v \rangle + \langle \nabla \tau \wedge w, \nabla v \rangle + \langle \nabla \tau (xw_y), A \nabla v \rangle + s_h(w, v).$$

Implicitly in the proof of Lemma 4.1 we proved also the following (hypo)coercivity result.

Lemma 5.1 ((Hypo)coercivity). Let

$$|||w||| := \left(\|\sqrt{B} \nabla w\|^2 + \|\sqrt{\alpha} \nabla w\|_{\partial_+ \Omega}^2 + \|\sqrt{A} \nabla w\|_{\partial_+ \Omega}^2 + \|\sqrt{\alpha} \nabla w\|_{\partial_\Omega}^2 + \|\nabla \tau \nabla w\|_{\Gamma_{\text{int}} \cup \partial_0 \Omega}^2 + s_{\lambda, h}(w, w)\right)^{1/2},$$

with $B$ as in the proof of Lemma 4.1. Then, $|||\cdot|||$ is a norm on $H^{1, \varepsilon}_0(\Omega) \cap H^2(\Omega, \mathcal{T})$, and we have

$$B(w, w) \geq \frac{1}{2}|||w|||^2,$$

for all $w \in V_h$. \qed
For the remaining of this work, unless explicitly stated, we assume that $\varepsilon$ is away from $1$ and that $\beta \sim \alpha^2$, so that $B = c_B \text{diag}(1, \alpha^2)$ for a given constant $0 < c_B \leq 1$. Thus, we can take $C_0 = (2c_B)^{-1}$. Note that this also covers the case of $A$ being singular, i.e., the operator $\nabla \cdot A\nabla(\cdot)$ being parabolic. Next, we establish the consistency of the bilinear form $B$.

**Lemma 5.2.** Assume that for the solution $u$ of (10) we have $u(t, \cdot) \in H^1_{-0}(\Omega) \cap H^3(\Omega)$, for almost all $t \in (0, t_f]$, and that (20) also holds. Then, for almost all $t \in (0, t_f]$ and for all $V \in V_h$, we have

$$ (u_t, V) + (\nabla u_t, A\nabla V) + B(u, V) = (f, V) + (\nabla f, A\nabla V). \quad (30) $$

**Proof.** The proof follows by integration by parts and by the smoothness of $\pi_0$ precisely below, and set $A \nabla \cdot \beta$ with $s$ almost all. Setting, now Uniqueness (and, therefore, existence due to linearity) follows from the coercivity of $B$. We shall now construct a $\pi$, so that $B(\rho, \vartheta) = 0$.

**Lemma 5.3** (hypo)elliptic projection. Assume that $B = c_B \text{diag}(1, \alpha^2)$ for a given constant $0 < c_B \leq 1$. For every $v \in H^1_{-0}(\Omega) \cap H^{5/2+\varepsilon}(\Omega, T)$, $\varepsilon > 0$, there exists a unique $\pi v \in V_h$ defined by

$$ B(\pi v, V) = B(v, V), \quad (32) $$

for all $V \in V_h$. Moreover, assuming further that $v \in H^{k_T}(T)$ for $T \in T$, $k_T \geq 3$, we have the approximation estimate:

$$ \|v - \pi v\|^2 \leq C(A, \kappa, \lambda) \sum_{T \in T} h_T^{2s_T} x_T |u|_{H^{k_T}(T)}^2, \quad (33) $$

with $s_T := \min\{p + 1, k_T\} - 2$ and $x_T := \max_{x \in T}\{1, |x|^2\}$, $T \in T$.

**Proof.** Uniqueness (and, therefore, existence due to linearity) follows from the coercivity of $B$ shown in Lemma 5.1. Setting, now $\eta := v - \mathcal{P}v$ and $\xi := \mathcal{P}v - \pi v$, with $\mathcal{P}$ denoting an optimal projection operator from $H^1_{-0}(\Omega)$ onto $V_h$, whose particular properties apart from approximation are irrelevant at this point, we have

$$ \frac{1}{2} \|\xi\|^2 \leq B(\xi, \xi) = -B(\eta, \xi) = - (\eta_x, \xi_x) - (\eta_y, x\xi) - (\nabla \mathcal{A} \nabla T \eta_x, \sqrt{A} \nabla T \xi_x) - (\nabla T (x\eta_y), A\nabla \xi) - s_h(\eta, \xi). \quad (34) $$

Now, integration by parts yields

$$ - (\eta_y, x\xi) = (x\eta, \xi_y) - \int_{\partial_\Omega} \eta y \xi ds, \quad (35) $$
and
\[
\begin{align*}
(\nabla_T(x\eta_y), AV\xi) \\
= \sum_{T \in T} \left( (\eta_y + x\eta_{xy} + \alpha\xi_x + \beta\xi_y + \gamma\xi_T) + (x\eta_{yy} + \beta\xi_x + \gamma\xi_y) \right) \\
= \alpha(\eta_y, \xi_x) + \beta(\eta_y, \xi_y) - \sum_{T \in T} \left( (\eta_x, x(\alpha\xi_x + \beta\xi_y) + \eta_y, x(\beta\xi_x + \gamma\xi_y)) \right) \\
+ \sum_{T \in T} \int_{\Omega} x\nu_2(\alpha\eta_x\xi_x + \beta\eta_x\xi_y + \beta\eta_y\xi_x + \gamma\eta_y\xi_y) \, ds.
\end{align*}
\] (36)

Observe that the last term on the right-hand side of (36) is equal to

\[
\begin{align*}
\int_{\Gamma_{\text{int}}} (0, x)^T \cdot (\alpha\eta_x, \xi_x + \beta\eta_y, \xi_y + \gamma\eta_y) \, ds \\
+ \int_{\Gamma_{\text{int}}} (0, x)^T \cdot (\alpha\xi_x, \xi_x + \beta\xi_y, \xi_y + \gamma\xi_y) \, ds \\
+ \int_{\partial\Omega} x\nu_2 A\nabla\eta \cdot \nabla\xi \, ds.
\end{align*}
\] (37)

and we note that the first term of (37) cancels with the first term of \( s_h(\eta, \xi) \). Using now (35), (36) and (37) into (34), along with standard Cauchy-Schwarz’ and Young’s inequality arguments, and working as in (29) gives

\[
C^{-1}\|\xi\|^2 \leq \|\eta_x\|^2 + \beta^{-1}\|\eta_y\|^2 + \|\sqrt{x\nu_2}\|_{\partial\Omega}^2 + \|\sqrt{\sqrt{x\nu_2}A\nabla\eta}\|_{\partial\Omega}^2 + \|\sqrt{A}\nabla\eta_x\|_\Gamma^2 \\
+ (\alpha^2 + \beta)\|\eta_y\|^2 + \|\sqrt{T\nabla\eta}\|_{\Gamma_{\text{int}}} + \kappa\|\sqrt{x\nu_2}\|_{\Gamma_{\text{int}}}^2 + \lambda\|\sqrt{x\nu_2}\|_{\Gamma_{\text{int}}}^2 \\
+ \sum_{T \in T} \left( \|\mu^{-1}\eta_x\|_T (\alpha\|\mu(\xi_x)\|_T + \beta\|\mu(\xi_y)\|_T) + \|\mu^{-1}\eta_y\|_T (\beta\|\mu(\xi_x)\|_T + \gamma\|\mu(\xi_y)\|_T) \right) \\
+ \|\sqrt{A/T}\|_{\Gamma_{\text{int}}}^2 + (\alpha^2\kappa^{-1} + \beta^2\lambda^{-1})\|\sqrt{x\nu_2}\|_{\Gamma_{\text{int}}}^2 + (\beta^2\kappa^{-1} + \gamma^2\lambda^{-1})\|\sqrt{x\nu_2}\|_{\Gamma_{\text{int}}}^2,
\]

for any \( \mu, \kappa, \lambda > 0 \), for some \( C > 1 \), independent of the, relevant to the argument, parameters. Alternatively, when \( \kappa, \lambda = 0 \), we work as follows: using the inverse estimate \( \|\xi_x\|_{\Omega}^2 \leq CP^2/h_T\|\xi_x\|_T^2 \) and, similarly for \( \xi_y \), we further estimate the second term on the right-hand side of (37) to arrive instead at

\[
C^{-1}\|\xi\|^2 \leq \|\eta_x\|^2 + \beta^{-1}\|\eta_y\|^2 + \|\sqrt{x\nu_2}\|_{\partial\Omega}^2 + \|\sqrt{\sqrt{x\nu_2}A\nabla\eta}\|_{\partial\Omega}^2 + (\alpha^2 + \beta)\|\eta_y\|^2 + \|\sqrt{T\nabla\eta}\|_{T}^2 \\
+ \sum_{T \in T} \left( \|\mu^{-1}\eta_x\|_T (\alpha\|\mu(\xi_x)\|_T + \beta\|\mu(\xi_y)\|_T) + \|\mu^{-1}\eta_y\|_T (\beta\|\mu(\xi_x)\|_T + \gamma\|\mu(\xi_y)\|_T) \right) \\
+ C|AB|^{-1/2}2\left( \|x\nu_2\|_{\Gamma_{\text{int}}}^2 + \|x\nu_2\|_{\Gamma_{\text{int}}}^2 \right) + \frac{1}{4}\|\sqrt{B}\nabla\xi\|_T^2,
\]

with \( |\cdot|_2 \) denoting the matrix-2-norm. Applying the inverse estimate \( \|\nabla v\|_T^2 \leq CP^4/h_T^2\|v\|_T \), for \( v \in P_p(T) \), on the terms containing \( \mu\xi \) in the last two alternative estimates, selecting \( \mu = h/p^2 \).
and performing standard manipulations, we arrive at the combined estimate

\[
C^{-1}\|\xi\|^2 \leq \|\eta_\kappa\|^2 + \beta^{-1}\|\eta\|^2 + \|\sqrt{x_{n2}}\eta\|^2_{\Omega} + \|\sqrt{A}\nabla \tau \eta\|_T^2 + \|\sqrt{\tau A}(\nabla \eta)\|_T^2
\]
\[
+ (\alpha^2 + \beta)(\|\mu^{-1}x_{n\kappa}\|^2 + \|\eta_\kappa\|^2) + (\beta^2 + \gamma^2)^2\|\mu^{-1}x\eta\|^2
\]
\[
+ \min \left\{ \left(\alpha^2\kappa^{-1} + \beta^2\lambda^{-1}\right)\|\sqrt{x_{n2}}\eta\|_T^2 + \left(\beta^2\kappa^{-1} + \gamma^2\lambda^{-1}\right)\|\sqrt{x_{n2}}\eta\|_T^2 \right\}
\]
\[
C|A|B^{-1/2}2(\|\sqrt{x_{n2}}h^{-1/2}\|_{\Omega} + \|\sqrt{x_{n2}}\eta\|_T^2)
\]
\[
+ k\|\sqrt{|x_{n2}}|\|\eta\|_T^2 + 2\|\sqrt{|x_{n2}}|\|\eta\|_T^2
\]

which holds for any \(\kappa, \lambda \geq 0\). Choosing \(P\) to be an \(hp\)-optimal projection operator onto the finite element space \(V_h\), we can approximate estimates of the form

\[
\|\nabla^m \eta\|_T \leq C h_T^{\min(p+1,k)-m} p^{m-k} \|u\|_{H^k(T)}
\]

for \(k > m\), with \(u \in H^k(T)\), \(T \in \mathcal{T}\) (see, e.g., [22, 3] for examples of such operators). These, together with the trace estimate \(\|w\|^2_{\partial T} \leq C\|w\|_T\|w\|_{H^1(T)}\) and the triangle inequality \(\|\rho\| \leq \|\eta\| + \|\xi\|\), already imply (33).

We now show the (potentially super-)approximation of for the left-hand side of (31). Such results are typical to finite element method for parabolic problems. We shall show that this is the case also for (21) discretising the, degenerate, Kolmogorov’s equation.

**Lemma 5.4.** Assume that (20) are satisfied. Then, under the assumptions of Lemma 4.1, we have the following bound

\[
\|\vartheta(t)\|^2 + \|\sqrt{A}\nabla \vartheta(t)\|^2 + \frac{1}{2} \int_0^t \|\vartheta\|^2 \, ds \leq \|\vartheta(0)\|^2 + \|\sqrt{A}\nabla \vartheta(0)\|^2
\]
\[
+ 4 \max\{1, c_0^{-2}\} \int_0^t \|\sqrt{B}\nabla \rho(t)\|^2 \, ds,
\]

for any \(t \in (0, t_f]\), for some positive constant \(C\), depending only on the shape-regularity and the local quasi-uniformity of the mesh.

**Proof.** Integrating in time between 0 and \(t \in (0, t_f]\) multiplying by 2 and using standard arguments, gives

\[
\|\vartheta(t)\|^2 + \|\sqrt{A}\nabla \vartheta(t)\|^2 + \int_0^t \|\vartheta\|^2 \, ds \leq \|\vartheta(0)\|^2 + \|\sqrt{A}\nabla \vartheta(0)\|^2
\]
\[
+ 2 \int_0^t (c_0^{-1}\|\rho(t)\|^2 + \|\sqrt{A}\nabla \rho(t)\|^2) \, ds
\]
\[
+ \frac{1}{2} \int_0^t (c_0\|\vartheta\|^2 + \|\sqrt{A}\nabla \vartheta\|^2) \, ds,
\]

recalling the notation \(c_0 = C_{PF}^{-1}\min_{B-A}^\lambda\). Using now (28), we deduce the result. \(\square\)

The above developments yield an a priori error bound in the norm appearing on the left-hand side of (38).
Theorem 5.5. Assume that $u_0, u, u_t \in H_{-\sigma}^1(\Omega) \cap H_T^k(T)$ for $k_T \geq 3$, $T \in T$, the latter two for almost all $t \in (0, t_f]$. Assume also that (20) are satisfied. Then, under the assumptions of Lemmata 4.1 and 5.3, the error $e := u - U$ of the finite element method (21) satisfies the bound:

$$
\|e(t)\|^2 + \|\sqrt{A}\nabla e(t)\|^2 + \int_0^t \|e\|^2 \, ds \leq C(A, \kappa, \lambda) \sum_{T \in T} \mathcal{E}_T(u_0, t, u),
$$

where

$$
\mathcal{E}_T(u_0, t, u) := \frac{h_T^{2s_T} x_T}{p^{2(k_T-3)}} \left( \max\{1, c_0^{-1}\} |u_0|^2_{H^s_T(T)} + |u|^2_{L_2(0,t; H^s_T(T))} + |u|^2_{L_2(0,t; H^s_T(T))} \right),
$$

with $s_T := \min\{p + 1, k_T\} - 2$, and the constant $C(A, \kappa, \lambda) > 0$ is bounded away from $+\infty$ for $0 < \alpha \leq 1/2$ and all $\kappa, \lambda \geq 0$ and independent of the mesh parameters and of $u$. Moreover, $C(A, \kappa, \lambda)$ is independent of the final time $t_f$.

Proof. Working as for (28), we have

$$
\|\hat{\vartheta}(0)\|^2 + \|\sqrt{A}\nabla \hat{\vartheta}(0)\|^2 \leq \max\{1, c_0^{-1}\} \|\sqrt{B}\nabla \hat{\vartheta}(0)\|^2 \\
\leq 2 \max\{1, c_0^{-1}\} \left( \|\sqrt{B}\nabla \rho(0)\|^2 + \|\sqrt{B}\nabla (u_0 - \Pi u_0)\|^2 \right).
$$

Similarly, we also have

$$
\|\rho(0)\|^2 + \|\sqrt{A}\nabla \rho(0)\|^2 \leq \max\{1, c_0^{-1}\} \|\sqrt{B}\nabla \rho(0)\|^2.
$$

The last two estimates can be further bounded from above by (33) and $hp$-approximation bounds for the orthogonal $L_2$-projection [17, 4], respectively. Also, from (33), we have

$$
\int_0^t \left( \|\sqrt{B}\nabla \rho_t\|^2 + \|\rho\|^2 \right) \, ds \leq C \sum_{T \in T} \frac{h_T^{2s_T} x_T}{p^{2(k_T-3)}} \left( |u_t|^2_{L_2(0,t; H^s_T(T))} + |u|^2_{L_2(0,t; H^s_T(T))} \right).
$$

Combining the above bounds with the triangle inequality recalling that $u - u_h = \rho + \hat{\vartheta}$, the result already follows. \(\square\)

The crucial property of the above error estimate, i.e., the independence of the respective constant $C(A, \kappa, \lambda)$ from the final time $t_f$, is a direct consequence of the hypocoercivity-compatible discretisation (21).

Assumption $u_0 \in H_{-\sigma}^1(\Omega)$ is a natural compatibility condition in this context. Since a typical setting in kinetic simulation concerns initial profiles with compact support within a computational domain $\Omega$, $u_0 \in H_{-\sigma}^1(\Omega)$ is of significant practical relevance also. Nonetheless, we shall now discuss the dependence of the error committed by the numerical method as $t_f \to \infty$ and we shall see that the effect of the initial condition error $u_0 - \Pi u_0$ diminishes exponentially as $t_f \to \infty$. Thus, possibly incompatible initial conditions $u_0 \in H^1(\Omega) \setminus H_{-\sigma}^1(\Omega)$ will have an exponentially diminishing effect in the accuracy of the method with respect to $t_f$.

Theorem 5.6. Assume that $u, u_t \in H_{-\sigma}^1(\Omega) \cap H_T^k(T)$ for $k_T \geq 3$, $T \in T$ for almost all $t \in (0, t_f]$. Assume that (20) are satisfied. Then, under the assumptions of Lemmata 4.1 and 5.3, the error $e := u - U$ of the finite element method (21) satisfies the bound:

$$
\|e(t_f)\|^2 + \|\sqrt{A}\nabla e(t_f)\|^2 \leq e^{-\min\{1, c_0\} t_f} \left( \|\Pi u_0 - \Pi u_0\|^2 + \|\sqrt{A}\nabla (\Pi u_0 - \Pi u_0)\|^2 \right) + C(A, \kappa, \lambda) \min\{1, c_0\} \sum_{T \in T} \frac{h_T^{2s_T} x_T}{p^{2(k_T-3)}} \mathcal{H}_T(t, u).
$$

(42)
where
\[ \mathcal{H}_T(t, u) := |u(t_f)|^2_{H^s_T(T)} + \int_0^{t_f} e^{-\min\{1, c_0\}(t_f-t)}|u_t|^2_{H^s_T(T)} \, dt, \]
for \( s_T := \min\{p+1, k_T\} - 2 \), with \( C(A, \kappa, \lambda) \) independent of \( t_f \), of \( u \), and of the mesh parameters (cf. Theorem 5.5).

**Proof.** Starting from (31), we use Lemmata 5.1 and 5.3 and, subsequently, apply (28) to arrive at
\[ \frac{d}{dt}(\|\vartheta(t)\|^2 + \|\sqrt{A}\nabla \vartheta(t)\|^2) + \min\{1, c_0\}(\|\vartheta\|^2 + \|\sqrt{A}\nabla \vartheta\|^2) \leq c_0^{-1}\|\vartheta_t\|^2 + \|\sqrt{A}\nabla \vartheta_t\|^2, \]
upon multiplication by 2. Grönwall’s Lemma, thus, implies
\[ \|\vartheta(t_f)\|^2 + \|\sqrt{A}\nabla \vartheta(t_f)\|^2 \leq C(A, \kappa, \lambda) \min\{1, c_0\} \sum_{t \in T} \frac{h_T^{2s_T}}{h_T^{2(k_T-3)}} |u_t|^2_{H^{s_T}(T)}. \]
As before, we also have
\[ \|\vartheta_t\|^2 + \|\sqrt{A}\nabla \vartheta_t\|^2 \leq \max\{1, c_0^{-1}\} \|\sqrt{B}\nabla \vartheta_t\|^2 \leq \frac{C(A, \kappa, \lambda) \min\{1, c_0\}}{C(A, \kappa, \lambda) \min\{1, c_0\}} \sum_{t \in T} \frac{h_T^{2s_T}}{h_T^{2(k_T-3)}} |u_t|^2_{H^{s_T}(T)}. \]

The last bound combined with (43) and the completely analogous estimate for
\[ \|\vartheta(t_f)\|^2 + \|\sqrt{A}\nabla \vartheta(t_f)\|^2 \leq \max\{1, c_0^{-1}\} \|\sqrt{B}\nabla \vartheta(t_f)\|^2, \]
already imply the result.

Therefore, (21) admits completely analogous long time properties compared to the PDE problem (10), (20).

**Remark 5.7.** The first term on the right-hand side of (43) vanishes upon altering the finite element method (21) at \( t = 0 \) from \( U(0) = \Pi u_0 \) to \( U(0) = \pi u_0 \). In practical terms, this results to a computational overhead of a stiffness matrix solve as opposed to a mass matrix one.

The a priori error estimate derived in Theorem 5.5 is optimal with respect to mesh-size \( h \), upon observing that the bilinear form \( B \) is the weak form of a formally 4th order differential operator. At the same time, the highest order terms involving \( A \) of \( B(u, v) \) play the role of “stabilisation” in this non-standard Galerkin context, resulting to increase of the spatial operator order from second to fourth. The estimate (40) is slightly suboptimal with respect to the polynomial degree \( p \), as is typical in error analyses of \( hp \)-version interior penalty procedures, such as (21), involving inverse estimates [11]. Correspondingly, the a priori error bound in Theorem 5.6 is formally suboptimal with respect to the mesh size \( h \) by one order; the errors are measured in weaker norms that the ones appearing on the right-hand side of (42). We view this as a reasonable price to pay for the long time robustness of method when \( t_f \to \infty \).

Indeed, in many practical scenarios, such as long time computations of decay to equilibrium distributions, the favourable exponential dependence with respect to \( t_f \) may be preferable to a slight suboptimality in the rate of convergence with respect to the mesh size \( h \). Moreover, given the typically high smoothness of the exact solution \( u \), the use of high order finite element spaces may diminish further the practical significance of this slight \( h \)-suboptimality.

One may be tempted, therefore, to scale the stabilisation terms involving \( A \) by appropriate powers of the mesh-size \( h \) to reduce the stiffness matrix scaling to one of a second order operator;

16
this is a standard practice in classical Galerkin contexts, e.g., for streamline upwind Petrov-
Galerkin methods for convection-dominated problems. In the present context, however, such
scaling would introduce new challenges, most important of which is that the spectral gap
constant $c_0$ would be proportional to a negative power of $h$. Although such dependence may
not turn out to be catastrophic in certain scenarios, e.g., when $u_t$ decays very fast (cf., (42) in
Theorem 5.6), this is a challenging question that requires further research.

Another possibility to retrieve optimality in the rate of convergence for the method (21)
in weaker norms is the incorporation of Aubin-Nitsche type duality techniques allowing for
$h$-optimal error control of the (hypo)elliptic projection error $u - \pi u$ with respect to $L_2$- and/or
$H^1$-equivalent norms. Although this is classical in parabolic problems, it appears to be a
challenge in the present context of hypocoercive operators. This is due to the non-closedness
of the hypocoercivity property with respect to the adjoint operation in the present variational
context [12]. This constitutes an interesting direction for future research.

Finally, we remark on the practical performance of the proposed method. Preliminary
computations show agreement with the theoretical developments presented above. Since this
is, however, an interesting topic in its own right, it will be discussed in detail elsewhere [5].

6 On extension to three spatial dimensions

To highlight the potential generality of the approach, we briefly consider the three-dimensional
version of Kolmogorov’s equation: for $\Omega \subset \mathbb{R}^3$, find $u : (0, t_f] \times \Omega \to \mathbb{R}$, such that

\[ u_t + L_3 u = u_t - u_{xx} + x u_y + y u_z = f, \quad \text{in } (0, t_f] \times \Omega, \]
\[ u = u_0, \quad \text{on } \{0\} \times \Omega, \]
\[ u = 0, \quad \text{on } (0, t_f] \times \partial_+ \Omega, \]

for $t_f > 0$ and $f \in H^1(\Omega)$, with

\[ \partial_0 \Omega := \{x \in \partial \Omega : n_1(x) \neq 0\}, \]
\[ \partial_\Omega := \{x \in \partial \Omega \setminus \partial_0 \Omega : b(x) \cdot n(x) < 0\}, \]
\[ \partial_\Omega := \{x \in \partial \Omega \setminus \partial_0 \Omega : b(x) \cdot n(x) > 0\}, \]

with $n(x) := (n_1(x), n_2(x), n_3(x))^T$ denoting the unit outward normal at $x := (x, y, z)^T$; for
brevity, we used the notation $b(x) := (0, x, y)^T$ and, thus, $L_3 u = b \cdot \nabla u - u_{xx}$. To ensure decay
to equilibrium for the homogeneous problem (i.e., $f = 0$), we assume further the additional boundary conditions

\[ \left(\frac{b}{2} \cdot \nabla u - n_1 \nabla u_x\right)|_{\partial_1 \partial_+ \Omega} = 0 \quad \text{and} \quad n_1 \nabla u_x|_{\partial_+ \Omega} = 0. \]

Remarkably, (44) is smoothing, despite possessing explicit diffusion in one spatial dimension only. Indeed, we have, respectively,

\[ [\partial_x, b \cdot \nabla] = \partial_y, \quad [\partial_y, b \cdot \nabla] = \partial_z, \]

thereby, Hörmander’s rank condition is satisfied [14], implying that (45) is, in fact, hypoelliptic!
Moreover, since full rank is achieved via commutators involving the skew-symmetric 1st order part of the PDE $b \cdot \nabla u$, (44) is also hypoelliptic [23, Theorem 24].

In a modest deviation from [23], we consider the matrix $A_3 \in \mathbb{R}^{3 \times 3}$ with

\[ A_3 := \begin{pmatrix} \alpha & \beta_1 & 0 \\ \beta_1 & \gamma_1 & \beta_2 \\ 0 & \beta_2 & \gamma_2 \end{pmatrix}. \]
which we use to define the weak form
\[(u_t, v) + (\nabla u_t, A_3 \nabla v) + a_3(u, v) = (f, v) + (\nabla f, A_3 \nabla v),\] (46)
whereby
\[a_3(u, v) := (u_x, v_x) + (b \cdot \nabla u, v) + (\nabla u_x, A_3 \nabla v_x) + (\nabla (b \cdot \nabla u), A_3 \nabla v).\] (47)
We now discuss the (hypo)coercivity of $a_3$. To this end, we compute
\[a_3(u, u) := \|u_x\|^2 + \frac{1}{2}\|\sqrt{b \cdot n} u\|^2_{\partial_\Omega} + \|\sqrt{A_3 \nabla u}\|^2 + (\nabla (b \cdot \nabla u), A_3 \nabla v),\]
since $\nabla \cdot b = 0$. We observe the identity
\[(\nabla (b \cdot \nabla u), A_3 \nabla u) = ((u_y, u_z, 0)^T, A_3 \nabla u) + (x \nabla u_y + y \nabla u_z, A_3 \nabla u).\]
Now, we have
\[(x \nabla u_y + y \nabla u_z, A_3 \nabla u) = \frac{1}{2}(b \cdot n \nabla u, A_3 \nabla u)_{\partial\Omega} = \frac{1}{2}\|\sqrt{b \cdot n} A_3 \nabla u\|^2_{\partial_\Omega},\]
using (45). Also, upon observing the identity
\[(u_y, u_z, 0)^T = \nabla u^T S, \quad \text{with} \quad S := \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},\]
denoting the left-shift operator, we have
\[
\|u_x\|^2 + ((u_y, u_z, 0)^T, A_3 \nabla u) = \int_{\Omega} \nabla u^T \tilde{B}_3 \nabla u \, dx = \|\sqrt{\tilde{B}_3} \nabla u\|^2,
\]
where for $e_1 := (1, 0, 0)^T$, we have set
\[\tilde{B}_3 := e_1 e_1^T + S A_3 = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta_1 & 0 \\ \beta_1 & \gamma_1 & \beta_2 \end{pmatrix},\]
and, thus,
\[\|\sqrt{\tilde{B}_3} \nabla u\|^2 = \|u_x\|^2 + \beta_1 \|u_y\|^2 + \beta_2 \|u_z\|^2 + \alpha (u_y, u_z) + \beta_1 (u_z, u_x) + \gamma_1 (u_z, u_y).\]
Selecting now $\beta_1 = \alpha^2$, $\beta_2 = \alpha^3$, $\gamma_1 = 2 \alpha^3$ and $\gamma_2 = \alpha^3$, for instance, $A_3$ is non-negative definite, and standard arguments yield the lower bound
\[
\|\sqrt{\tilde{B}_3} \nabla u\|^2 \geq \frac{1}{4} \left( \|u_x\|^2 + \alpha^2 \|u_y\|^2 + 4(1 - 17\alpha)\alpha^3 \|u_z\|^2 \right),
\]
for $\alpha \in (0, \frac{1}{17})$. We remark that $v - \nabla \cdot A_3 \nabla v$ is a parabolic operator for the above choice of $A_3$, the latter possessing one zero and two positive eigenvalues. Next, we consider the matrix
\[B_3 = \frac{1}{4} \text{diag}(1, \alpha^2, 4(1 - 17\alpha)\alpha^3),\]
and we have for $w \in H^1_{-\partial}(\Omega)$,
\[
\|\sqrt{B_3} \nabla w\|^2 = \|\sqrt{B_3 - \nu^2 A_3} \nabla w\|^2 + \nu \|\sqrt{A_3} \nabla w\|^2 \geq \lambda^{\min}_{B_3-\nu^2 A_3} \|\nabla w\|^2 + \nu \|\sqrt{A_3} \nabla w\|^2.
\]
Selecting now $\nu > 0$ small enough, we can ensure that $\lambda_{\min}^{B_3 - \nu A_3} > 0$. Thus, we can conclude

$$\|\sqrt{B_3} \nabla w\|^2 \geq c_3(\alpha, \nu)(\|\nabla w\|^2 + \|\sqrt{A_3} \nabla w\|^2),$$

(48)

for some positive constant $c_3(\alpha, \nu)$, yielding the coercivity of the modified bilinear form (47).

Starting now from (46), we consider the spatially discrete finite element method: for every $t \in (0, t_f]$, find $U \equiv U(t) \in V_h$, (with $V_h$ the $d$-dimensional version of the finite element space,) such that (21) holds for all $V \in V_h$, with $s_A(U, V)$ replaced by:

$$s_{h,3}(U, V) := -\int_{\Gamma_{int} \cup \partial_0 \Omega} b \cdot \left[\begin{array}{c} \nabla V \\ A_3 \nabla U \end{array}\right] ds + s_{1,h,3}(U, V)$$

$$-\int_{\Gamma_{int} \cup \partial_0 \Omega} \left( \left[ A_3 \nabla U_x \right] \cdot [\nabla V]_1 + \left[ A_3 \nabla V_x \right] \cdot [\nabla U]_1 - \tau [\nabla U]_1 \cdot A_3 [\nabla V]_1 \right) ds,$$

for $\tau$ as before, with

$$s_{1,h,3}(U, V) := \int_{\Gamma_{int}} \frac{|b \cdot n|}{2} \left( \kappa [U_x] \cdot [V_x] + \lambda_1 [U_y] \cdot [V_y] + \lambda_2 [U_z] \cdot [V_z] \right) ds.$$

Following an analogous approach we can show a corresponding results to Lemma 4.1 and Theorem 4.3 with $A$ replaced by $A_3$, a different $c_0 > 0$ arising from $c_3(\alpha, \nu)$ and $s_{1,h}$ replaced by $s_{1,h,3}$. Similarly corresponding results to Lemmata 5.1 and 5.2 follow to finally arrive at error bounds completely analogous to Theorems 5.5 and 5.6.

Acknowledgements

I thank Zhaonan (Peter) Dong (University of Leicester) for helpful comments on an earlier version of this work. The financial support of The Leverhulme Trust (research project grant no. RPG-2015-306) is also gratefully acknowledged.

References

[1] G. Akrivis, M. Crouzeix, and V. Thomée, Numerical methods for ultraparabolic equations, Calcolo, 31 (1994), pp. 179–190 (1996).
[2] M. Bessemoulin-Chatard and F. Filbet, A finite volume scheme for nonlinear degenerate parabolic equations, SIAM J. Sci. Comput., 34 (2012), pp. B559–B583.
[3] A. Cangiani, Z. Dong, E. H. Georgoulis, and P. Houston, hp-version discontinuous Galerkin methods on polygonal and polyhedral meshes, SpringerBriefs in Mathematics, Springer, Cham, 2017.
[4] A. Chernov, Optimal convergence estimates for the trace of the polynomial $L^2$-projection operator on a simplex, Math. Comp., 81 (2012), pp. 765–787.
[5] Z. Dong and E. H. Georgoulis, In preparation.
[6] G. Dujardin, F. Hérau, and P. Lapitte, Coercivity, hypocoercivity, exponential time decay and simulations for discrete Fokker-Planck equations, arXiv:1802.02173v1, (2018).
[7] J.-P. Eckmann and M. Hairer, *Spectral properties of hypoelliptic operators*, Comm. Math. Phys., 235 (2003), pp. 233–253.

[8] A. Ern and J.-L. Guermond, *Discontinuous Galerkin methods for Friedrichs’ systems. II. Second-order elliptic PDEs*, SIAM J. Numer. Anal., 44 (2006), pp. 2363–2388.

[9] G. Fichera, *Sulle equazioni differenziali lineari ellittico-paraboliche del secondo ordine*, Atti Accad. Naz. Lincei. Mem. Cl. Sci. Fis. Mat. Nat. Sez. I. (8), 5 (1956), pp. 1–30.

[10] E. L. Foster, J. Lohéac, and M.-B. Tran, *A structure preserving scheme for the Kolmogorov-Fokker-Planck equation*, J. Comput. Phys., 330 (2017), pp. 319–339.

[11] E. H. Georgoulis, E. Hall, and J. M. Melenk, *On the suboptimality of the p-version interior penalty discontinuous Galerkin method*, J. Sci. Comput., 42 (2010), pp. 54–67.

[12] F. Hérau, *Short and long time behavior of the Fokker-Planck equation in a confining potential and applications*, J. Funct. Anal., 244 (2007), pp. 95–118.

[13] F. Hérau and F. Nier, *Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential*, Arch. Ration. Mech. Anal., 171 (2004), pp. 151–218.

[14] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171.

[15] 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.

[16] A. Kolmogoroff, *Zufällige Bewegungen (zur Theorie der Brownschen Bewegung)*, Ann. of Math. (2), 35 (1934), pp. 116–117.

[17] H. Li and J. Shen, *Optimal error estimates in Jacobi-weighted Sobolev spaces for polynomial approximations on the triangle*, Math. Comp., 79 (2010), pp. 1621–1646.

[18] C. Mouhot and L. Neumann, *Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus*, Nonlinearity, 19 (2006), pp. 969–998.

[19] O. A. Oleinik and E. V. Radkevič, *Second order equations with nonnegative characteristic form*, Plenum Press, New York-London, 1973. Translated from the Russian by Paul C. Fife.

[20] A. Porretta and E. Zuazua, *Numerical hypocoercivity for the Kolmogorov equation*, Math. Comp., 86 (2017), pp. 97–119.

[21] H. Risken, *The Fokker-Planck equation*, vol. 18 of Springer Series in Synergetics, Springer-Verlag, Berlin, second ed., 1989. Methods of solution and applications.

[22] C. Schwab, *p- and hp-finite element methods*, Numerical Mathematics and Scientific Computation, The Clarendon Press, Oxford University Press, New York, 1998. Theory and applications in solid and fluid mechanics.

[23] C. Villani, *Hypocoercivity*, Mem. Amer. Math. Soc., 202 (2009), pp. iv+141.