

On the time growth of the error of the DG method for advective problems

VÁCLAV KUČERA*

*Faculty of Mathematics and Physics, Charles University, Sokolovská 83,
Praha 8, 18675, Czech Republic*

*Corresponding author: vaclav.kucera@email.cz

AND

CHI-WANG SHU

*Division of Applied Mathematics, Brown University, Providence, RI 02912, USA
shu@dam.brown.edu*

[Received on 16 March 2017; revised on 15 January 2018]

In this paper we derive *a priori* $L^\infty(L^2)$ and $L^2(L^2)$ error estimates for a linear advection–reaction equation with inlet and outlet boundary conditions. The goal is to derive error estimates for the discontinuous Galerkin method that do not blow up exponentially with respect to time, unlike the usual case when Gronwall’s inequality is used. While this is possible in special cases, such as divergence-free advection fields, we take a more general approach using exponential scaling of the exact and discrete solutions. Here we use a special scaling function, which corresponds to time taken along individual pathlines of the flow. For advection fields, where the time that massless particles carried by the flow spend inside the spatial domain is uniformly bounded from above by some \hat{T} , we derive $\mathcal{O}(h^{p+1/2})$ error estimates where the constant factor depends only on \hat{T} , but not on the final time T . This can be interpreted as applying Gronwall’s inequality in the error analysis along individual pathlines (Lagrangian setting), instead of physical time (Eulerian setting).

Keywords: discontinuous Galerkin method; advection–reaction problem; *a priori* error estimates; time growth of error; exponential scaling.

1. Introduction

The discontinuous Galerkin (DG) finite element method introduced in the study by [Reed & Hill \(1973\)](#) is an increasingly popular method for the numerical solution of partial differential equations. The DG method was first formulated for a neutron transport equation and such problems remain the major focus of the DG community. It is for problems of an advective or convective nature that the DG method is suited best and shows its strengths compared to other numerical methods for such problems.

In this paper, we shall consider a scalar time-dependent linear advection–reaction equation of the form

$$\frac{\partial u}{\partial t} + a \cdot \nabla u + cu = 0. \quad (1.1)$$

We will discretize the problem in space using the DG method with the upwind numerical flux on unstructured simplicial meshes in \mathbb{R}^d which may contain hanging nodes. The first analysis of the DG

method for a stationary advection–reaction problem with constant coefficients was made in the study by Lesaint & Raviart (1974), later improved in the study by Johnson & Pitkäranta (1986).

In this paper we derive *a priori* estimates of the error $e_h = u - u_h$, where u_h is the DG solution. The goal is to derive estimates of e_h in the $L^\infty(L^2)$ - and $L^2(L^2)$ -norms of the order $Ch^{p+1/2}$, where the constant C does not depend exponentially on the final time $T \rightarrow +\infty$. Such results already exist in the literature; however, they are derived under the ellipticity condition

$$c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0 \quad (1.2)$$

for some constant γ_0 . In the paper by Feistauer & Švadlenka (2004), $\gamma_0 = 0$ is also admissible, which corresponds to the interesting case of a divergence-free advection field a . Without assuming (1.2), when one proceeds straightforwardly, at some point Gronwall’s inequality must be used in the proofs. This, however, results in exponential growth of the constant factor C in the error estimate with respect to time.

Here, we will circumvent condition (1.2) by considering an exponential scaling transformation

$$u(x, t) = e^{\mu(x, t)} \tilde{u}(x, t), \quad u_h(x, t) = e^{\mu(x, t)} \tilde{u}_h(x, t) \quad (1.3)$$

of the exact and discrete solutions. Substituting (1.3) into (1.1) and its discretization results in a new equation for \tilde{u} with a modified reaction term $\tilde{c} = \frac{\partial \mu}{\partial t} + a \cdot \nabla \mu + c$. The function μ can then be chosen in various ways in order to satisfy the ellipticity condition (1.2) for the new equation. After performing the error analysis, the resulting error estimates for $\tilde{e}_h = \tilde{u} - \tilde{u}_h$ can be transformed back via (1.3) to obtain estimates for e_h .

The use of transformations similar to (1.3) is not new. The trick known as *exponential scaling*, corresponding to taking $\mu(x, t) = \alpha t$ for some sufficiently large constant α , is very well known from the partial differential equation and numerical analysis communities. However, the application of such a scaling transformation inherently leads to exponential dependence of the error on t after transforming from \tilde{e}_h back to e_h . Other choices of μ are possible, e.g. $\mu(x, t) = \mu_0 \cdot x$ for some constant vector μ_0 . This choice was used in the stationary case in the studies by Nävert (1982) and Johnson & Pitkäranta (1986). Taking $\mu(x)$ independent of t corresponds to the analysis of the stationary case from the study by Ayuso & Marini (2009). In the nonstationary case the use of this stationary transform would lead to too restrictive assumptions on the advection field.

In this paper we construct the function μ from (1.3) using characteristics of the advection field. Namely, μ will be proportional to time along individual pathlines of the flow. If pathlines exist only for a finite time bounded by some \hat{T} for each particle entering and leaving the spatial domain Ω , we obtain error estimates exponential in \hat{T} and not the final physical time T . This is a result we would expect if we applied Gronwall’s lemma in the Lagrangian framework along individual pathlines (which exist only for a finite time uniformly bounded by \hat{T}) instead of the usual application of Gronwall with respect to physical time in the Eulerian framework. The analysis can be carried out under mild assumptions that $a(\cdot, \cdot)$ satisfies the assumptions of the Picard–Lindelöf theorem and that there are no characteristic boundary points on the inlet (i.e. $a \cdot \mathbf{n}$ is uniformly bounded away from zero on the inlet).

The paper is organized as follows. After introducing the continuous problem in Section 2, we discuss the exponential scaling transform, its variants and application in the weak formulation in Section 3. In Section 4 we construct the scaling function μ and prove results on its regularity and other properties needed in the analysis. In Section 5 we introduce the DG formulation and its basic properties. In Sections 6 and 7 we analyse the DG advection and reaction forms and estimate the error of the method.

In order to focus on the main ideas of the analysis, we postpone some of the technical estimates of the DG forms to the Appendix.

We use (\cdot, \cdot) to denote the $L^2(\Omega)$ scalar product and $\|\cdot\|$ for the $L^2(\Omega)$ -norm. To simplify the notation, we shall drop the argument Ω in Sobolev norms, e.g. $\|\cdot\|_{H^{p+1}}$ denotes the $H^{p+1}(\Omega)$ -norm. We will also denote the Bochner norms over the whole considered interval $(0, T)$ in concise form, e.g. $\|u\|_{L^2(H^{p+1})}$ denotes the $L^2(0, T; H^{p+1}(\Omega))$ -norm of u .

2. Continuous problem

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$ be a bounded polygonal (polyhedral) domain with Lipschitz boundary $\partial\Omega$. Let $0 < T \leq +\infty$ and let $Q_T = \Omega \times (0, T)$ be the space–time domain. We note that we admit the time interval to be infinite in the special case of $T = +\infty$.

We seek $u : Q_T \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \frac{\partial u}{\partial t} + a \cdot \nabla u + cu &= 0 & \text{in } Q_T, \\ u &= u_D & \text{on } \partial\Omega^- \times (0, T), \\ u(x, 0) &= u^0(x), & x \in \Omega. \end{aligned} \tag{2.1}$$

Here $a : \overline{Q_T} \rightarrow \mathbb{R}^d$ and $c : \overline{Q_T} \rightarrow \mathbb{R}$ are the given advective field and reaction coefficient, respectively. By $\partial\Omega^-$ we denote the *inflow* part of the boundary, i.e.

$$\partial\Omega^- = \{x \in \partial\Omega; a(x, t) \cdot \mathbf{n}(x) < 0 \quad \forall t \in (0, T)\},$$

where $\mathbf{n}(x)$ is the unit outer normal to $\partial\Omega$ at x . For simplicity we assume that the inflow boundary is independent of t , in other words we assume that $a(x, t) \cdot \mathbf{n}(x) < 0$ for all t and all $x \in \partial\Omega^-$, and $a(x, t) \cdot \mathbf{n}(x)$ is non-negative on $\partial\Omega \setminus \partial\Omega^-$ for all t . In general $\partial\Omega^-$ can depend on t due to the nonstationary nature of a . However, we do not consider this case since it would only make the notation more complicated without changing the key points of the analysis itself.

We assume that the reaction coefficient satisfies $c \in C([0, T]; L^\infty(\Omega)) \cap L^\infty(Q_T)$. Furthermore, let $a \in C([0, T]; W^{1,\infty}(\Omega))$ with $a, \nabla a$ uniformly bounded a.e. in Q_T . In other words, $a(\cdot, \cdot)$ satisfies the assumptions of the Picard–Lindelöf theorem: continuity with respect to x, t and uniform Lipschitz continuity with respect to x . The spatial Lipschitz constant of a will be denoted by L_a .

The DG method for the stationary version of problem (2.1) was analysed in the studies by Nävert (1982), Johnson & Pitkäranta (1986) and Ayuso & Marini (2009). In this case the well-posedness of the problem is guaranteed by the assumption that the flow field possesses neither closed curves nor stationary points, cf. the study by Devinatz *et al.* (1974). For the nonstationary problem (2.1), the situation is simpler. Here, the problem can be solved using the method of characteristics, similarly to the method we use in Section 4, cf. Evans (2010). Along each characteristic curve (pathline), problem (2.1) reduces to a simple, ordinary differential equation, which can be solved in closed form. The result is exponential growth or decay of u along the pathline with the rate given by the reaction coefficient c . As the flow field a satisfies the assumptions of the Picard–Lindelöf theorem, we have existence and uniqueness of the pathlines. Since the Dirichlet boundary condition is prescribed only at the inflow boundary, this gives the well-posedness of problem (2.1). We note that also a ‘parabolic’ approach to

(2.1) is possible to obtain ‘energy’ estimates of the time growth of the $L^2(\Omega)$ -norm of the solution. This is done by Gronwall’s lemma after testing the weak form of (2.1) with the solution itself and applying Green’s theorem in the advection terms; cf. e.g. Mizohata (1973) and Larsson & Thomée (2003). The result is an upper bound on the time growth of the solution that is exponential in the quantity $c - \frac{1}{2}\operatorname{div} a$ if this quantity is non-negative; compare with the ellipticity condition (3.1). We note that from the theoretical and practical points of view, the case of *systems* of equations of the considered type is much more interesting; cf. e.g. Benzoni-Gavage & Serre (2007).

3. Exponential scaling

The standard assumption on the coefficients a, c found throughout the numerical literature is

$$c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0 \quad \text{on } Q_T \quad (3.1)$$

for some constant γ_0 . This assumption comes from the requirement that the weak formulations of the advection and reaction terms give an elliptic bilinear form on the corresponding function space. In the study by Feistauer & Švadlenka (2004), this assumption is avoided by using *exponential scaling* in time, i.e. the transformation $u(x, t) = e^{\alpha t} w(x, t)$ for $\alpha \in \mathbb{R}$. Substitution into (2.1) and division of the whole equation by the common positive factor $e^{\alpha t}$ leads to a modified reaction coefficient $\tilde{c} = c + \alpha$ in the new equation for w . By choosing the constant α sufficiently large, (3.1) will be satisfied for the new equation. In the study by Feistauer & Švadlenka (2004), error estimates that grow linearly in time are then derived for the DG scheme. However, for $\alpha > 0$, if one transforms the resulting estimates back to the original problem using the exponential scaling transformation, the result is an estimate that depends exponentially on T .

Another possibility to avoid assumption (3.1) is a transformation similar to exponential scaling, but with respect to the spatial variable; cf. the studies by Nävert (1982), Johnson & Pitkäranta (1986) and Roos *et al.* (2008): let $\mu_0 \in \mathbb{R}^d$, then we write

$$u(x, t) = e^{\mu_0 \cdot x} \tilde{u}(x, t). \quad (3.2)$$

If we assume that u is sufficiently regular, as we shall in this paper, by substituting (3.2) into (2.1) and dividing the whole equation by the strictly positive function $e^{\mu_0 \cdot x}$, we get the new problem

$$\frac{\partial \tilde{u}}{\partial t} + a \cdot \nabla \tilde{u} + (a \cdot \mu_0 + c) \tilde{u} = 0. \quad (3.3)$$

The condition corresponding to (3.1) now reads as follows: there exists $\mu_0 \in \mathbb{R}^d$ such that

$$a \cdot \mu_0 + c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0 \quad \text{on } Q_T. \quad (3.4)$$

A possible generalization of the exponential scaling transformation (3.2) is taking a sufficiently smooth function $\mu : \Omega \rightarrow \mathbb{R}$ and setting

$$u(x, t) = e^{\mu(x)} \tilde{u}(x, t). \quad (3.5)$$

Again, substituting into (2.1) and dividing by $e^{\mu(x)}$, we get the new problem

$$\frac{\partial \tilde{u}}{\partial t} + a \cdot \nabla \tilde{u} + (a \cdot \nabla \mu + c) \tilde{u} = 0.$$

The condition corresponding to (3.1) and (3.4) now reads as follows: there exists $\mu : \Omega \rightarrow \mathbb{R}$ such that

$$a \cdot \nabla \mu + c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0 \quad \text{on } Q_T. \quad (3.6)$$

This is essentially the approach used in the study by Ayuso & Marini (2009) for a stationary advection–diffusion–reaction problem. As shown in the study by Devinatz *et al.* (1974), the existence of a function $\mu : \Omega \rightarrow \mathbb{R}$ such that $a \cdot \nabla \mu \geq \gamma_0$ is equivalent to the property that the advective field a possesses neither closed curves nor stationary points. The uniformly positive term $a \cdot \nabla \mu$ can then be used to dominate the possibly negative term $c - \frac{1}{2} \operatorname{div} a$ in order to satisfy condition (3.6). If we used the choice (3.5) in our analysis, we would need to assume the nonexistence of closed curves or stationary points of the flow field for all t . This assumption is too restrictive. Furthermore, in the study by Ayuso & Marini (2009), rather high regularity of μ is needed in the analysis, namely $\mu \in W^{p+1,\infty}(\Omega)$, where p is the polynomial degree of the DG scheme.

In this paper, we shall generalize the transformation (3.5) using a function $\mu : Q_T \rightarrow \mathbb{R}$ and set

$$u(x, t) = e^{\mu(x,t)} \tilde{u}(x, t). \quad (3.7)$$

In our analysis we will need only minimum regularity of the scaling function μ , in contrast to the study by Ayuso & Marini (2009). Namely, we assume that μ is Lipschitz continuous both in space and time. This implies that the derivatives μ_t and $\nabla \mu$ exist for almost all x and t . Hence, the following steps are valid for almost all x and t and particularly, they will hold in the integral (i.e. weak) sense. We note that the Lipschitz continuity in space and time will hold for the specific construction of μ from Section 4, as shown in Section 4.1.

As in the previous cases, substituting (3.7) into (2.1) and dividing by $e^{\mu(x,t)}$ gives the new problem

$$\frac{\partial \tilde{u}}{\partial t} + a \cdot \nabla \tilde{u} + \left(\frac{\partial \mu}{\partial t} + a \cdot \nabla \mu + c \right) \tilde{u} = 0. \quad (3.8)$$

The condition corresponding to (3.1), (3.4) and (3.6) now reads as follows: there exists $\mu : Q_T \rightarrow \mathbb{R}$ such that

$$\frac{\partial \mu}{\partial t} + a \cdot \nabla \mu + c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0 \quad \text{a.e. in } Q_T. \quad (3.9)$$

This is the condition we will assume throughout the paper. Using the transformation (3.7) instead of (3.5) gives sub-exponential growth or even uniform boundedness of the constant factor in the error estimate with respect to time.

If the original problem (2.1) does not satisfy (3.1), one is tempted to numerically solve the transformed equation (3.8) instead, using DG or any other method. This is done e.g. in Roos *et al.* (2008) for the case of linear μ , i.e. (3.2). However, then we are numerically solving a different problem and obtain different results, as the DG solutions of (2.1) and (3.8) are not related by the simple relation (3.7), unlike the exact solutions. However, as we will show in this paper, one can analyse the DG method

for the original problem (2.1), while taking advantage of the weaker ellipticity condition (3.9) for the transformed problem (3.8).

Since the DG scheme is based on a suitable weak formulation, the first step is to reformulate the transformation (3.7) within the weak, rather than strong formulation of (2.1). The key step in deriving (3.8) from (2.1) is dividing the whole equation by the common factor $e^{\mu(x,t)}$. However, if we substitute (3.7) into the weak formulation

$$\int_{\Omega} \frac{\partial u}{\partial t} v + a \cdot \nabla uv + cuv \, dx = 0, \quad (3.10)$$

it is not easy to divide the equation by $e^{\mu(x,t)}$, since this is inside the integral. The solution is to test (3.10) by the new test function $\hat{v}(x, t) = e^{-\mu(x,t)} v(x, t)$. Due to the opposite signs in the exponents, the exponential factors cancel each other and we obtain the weak formulation of (3.8):

$$\int_{\Omega} \frac{\partial \tilde{u}}{\partial t} v + a \cdot \nabla \tilde{u} v + \left(\frac{\partial \mu}{\partial t} + a \cdot \nabla \mu + c \right) \tilde{u} v \, dx = 0.$$

We note that the transformations $u \mapsto \tilde{u}$ and $v \mapsto \hat{v}$ are bijections. Furthermore, since we assume that μ is Lipschitz continuous in space, then the factor $e^{\mu(\cdot, t)}$ lies in $W^{1,\infty}(\Omega)$ and the transformation (3.7) is a bijection from the function space V into itself, where

$$V = \{u \in H^1(\Omega); u|_{\partial\Omega^-} = 0\}$$

is the space for weak solutions and test functions for (2.1) we consider. This is the case $u_D = 0$, the general case can be treated by the standard Dirichlet lifting procedure. We note that in the general case, the appropriate function spaces for the solution of (2.1) would be e.g. $L^1(\Omega)$ or $BV(\Omega)$ with respect to space. However, since we assume smooth solutions and the linear case, it is possible to consider the Hilbert setting of $H^1(\Omega)$ which makes the analysis easier in the finite element or DG framework.

In the DG method, the above procedure using transform (3.7) cannot be directly applied, since if $v \in S_h$ —the discrete space of discontinuous piecewise polynomial functions—then $\hat{v} = e^{-\mu} v$ will no longer lie in S_h and therefore cannot be used as a test function in the formulation of the method. The solution is to test with a suitable projection of \hat{v} onto S_h and estimate the difference, as we shall do in Section 6.

4. Construction of the function μ

In this section we show a construction of the function μ satisfying (3.9). Many different constructions of μ are possible depending on the assumptions on the vector field a . For example, we can always take $\mu(x, t) = \alpha t$ for some $\alpha \geq 0$. This corresponds to standard exponential scaling and, as we have seen, this choice leads to exponential growth in time of the error estimate. Another possibility was the approach of Ayuso & Marini (2009) mentioned in Section 2, where a suitable function $\mu(x)$ exists if a , which is stationary, possesses no closed curves or stationary points. Here we show another possibility with an interesting interpretation.

If $c - \frac{1}{2} \operatorname{div} a$ is negative or changes sign frequently, we can use the expression $\mu_t + a \cdot \nabla \mu$ to dominate this term everywhere. If we choose μ_1 such that

$$\frac{\partial \mu_1}{\partial t} + a \cdot \nabla \mu_1 = 1 \quad \text{on } Q_T, \quad (4.1)$$

then by multiplying μ_1 by a sufficiently large constant, we can satisfy the ellipticity condition (3.9) for a chosen $\gamma_0 > 0$.

Equation (4.1) can be explicitly solved using characteristics. We define *pathlines* of the flow, i.e. the family of curves $S(t; x_0, t_0)$ by

$$S(t_0; x_0, t_0) = x_0 \in \overline{\Omega}, \quad \frac{dS(t; x_0, t_0)}{dt} = a(S(t; x_0, t_0), t). \quad (4.2)$$

This means that $S(\cdot; t_0, x_0)$ is the trajectory of a massless particle in the nonstationary flow field a passing through point x_0 at time t_0 . It is convenient to choose the parameter t_0 minimal for each pathline—then the pair (x_0, t_0) is the ‘origin’ of the pathline. In other words, for each $x_0 \in \overline{\Omega}$, there is a pathline $S(\cdot; x_0, 0)$ corresponding to trajectories of particles present in Ω at the initial time $t_0 = 0$. Then there are trajectories of particles entering through the inlet part of $\partial\Omega$: for each $x_0 \in \partial\Omega^-$ and all t_0 there exists a pathline $S(\cdot; x_0, t_0)$ originating at $(x_0, t_0) \in \partial\Omega^- \times [0, T)$.

Equation (4.1) can then be rewritten along the pathlines as

$$\frac{d\mu_1(S(t; x_0, t_0), t)}{dt} = \left(\frac{\partial \mu_1}{\partial t} + a \cdot \nabla \mu_1 \right) (S(t; x_0, t_0), t) = 1,$$

therefore

$$\mu_1(S(t; x_0, t_0), t) = t - t_0. \quad (4.3)$$

Here any constant can be chosen instead of t_0 ; however, this choice is the most convenient. With this choice, we have at the origin of a pathline

$$\mu_1(S(t_0; x_0, t_0), t_0) = 0$$

and the value of μ_1 along this pathline is simply the time elapsed since t_0 . In other words, $\mu_1(x, t)$ is the time a particle carried by the flow passing through $x \in \Omega$ at time t has spent in Ω up to the time t . In this paper we assume this quantity to be uniformly bounded in order for μ_1 and μ to be uniformly bounded; cf. assumption (4.7). It is important to note that t_0 depends on the specific trajectory considered and that t_0 will in general be different for each (x, t) . Therefore, even though t and t_0 may be arbitrarily large if we consider an infinite time interval, the difference $t - t_0 = \mu_1$, which measures the time elapsed since the entry of the corresponding particle into Ω , will remain bounded under this assumption.

Now that we have constructed a function satisfying (4.1), we can choose γ_0 and define e.g.

$$\mu(x, t) = \mu_1(x, t) \left(\left| \inf_{Q_T} \left(c - \frac{1}{2} \operatorname{div} a \right) \right| + \gamma_0 \right), \quad (4.4)$$

where $f^- = \min\{0, f\}$ is the negative part of f . Then

$$\frac{\partial \mu}{\partial t} + a \cdot \nabla \mu + c - \frac{1}{2} \operatorname{div} a = \left| \inf_{Q_T} \left(c - \frac{1}{2} \operatorname{div} a \right)^- \right| + \gamma_0 + c - \frac{1}{2} \operatorname{div} a \geq \gamma_0.$$

In the main result, of this paper, Theorem 7.1 and Corollary 7.2, this choice of μ leads to estimates of the following form for the DG error e_h (cf. (7.1) and (7.10)):

$$\|e_h\|_{L^\infty(L^2)} + \sqrt{\gamma_0} \|e_h\|_{L^2(L^2)} \leq C e^{\widehat{T}(|\inf_{Q_T} (c - \frac{1}{2} \operatorname{div} a)^-| + \gamma_0)} h^{p+1/2}, \quad (4.5)$$

where $\widehat{T} = \sup_{Q_T} \mu_1$, i.e. \widehat{T} is the maximal particle ‘lifetime’, i.e. the maximal time any particle carried by the flow field a spends in Ω , since by (4.3), $\mu_1(x, t)$ is the time the particle at (x, t) carried by the flow has spent in Ω . If we compare this to estimates obtained by a straightforward analysis using Gronwall’s lemma without any ellipticity assumption, we would expect

$$\|e_h\|_{L^\infty(L^2)} \leq C e^{T(\sup_{Q_T} |c - \frac{1}{2} \operatorname{div} a|)} h^{p+1/2}. \quad (4.6)$$

Comparing (4.6) to (4.5), we see that the estimates are essentially of similar form, only the exponential dependence on global physical time T has been replaced by dependence on time \widehat{T} along pathlines, which is bounded. Effectively, our analysis replaces the application of Gronwall’s lemma in the Eulerian framework with its application in the Lagrangian framework—along individual pathlines which have bounded length.

Throughout the paper, we assume the particle lifetime \widehat{T} to be bounded. In general, we could consider dependencies of the form

$$\widehat{T}(t) = \sup_{(x, \vartheta) \in \Omega \times (0, t)} \mu_1(x, \vartheta),$$

i.e. $\widehat{T}(t)$ is the maximal time any particle carried by the flow a spends in Ω up to time t . From (4.5), we can expect exponential dependence of the $L^\infty(0, t; L^2)$ and $L^2(0, t; L^2)$ norms on $\widehat{T}(t)$. The result of the standard analysis (4.6) corresponds to the ‘worst case’ $\widehat{T}(t) = t$, i.e. that there is a particle that stays inside Ω for all $t \in [0, T]$. However, considering more general dependencies on time is possible, e.g. $\widehat{T} = \sqrt{t}$, leading to growth of the error that is exponential in \sqrt{t} .

4.1 Regularity of the function μ

In our analysis of the DG scheme we will assume that μ satisfies

$$\begin{aligned} 0 &\leq \mu(x, t) \leq \mu_{\max}, \\ |\mu(x, t) - \mu(y, t)| &\leq L_\mu |x - y|, \end{aligned} \quad (4.7)$$

for all $x, y \in \Omega$ and $t \in (0, T)$. In other words, μ is non-negative, uniformly bounded and Lipschitz continuous in space, where the Lipschitz constant is uniformly bounded for all t . Since Ω is a Lipschitz domain, hence quasiconvex, this means that $\mu(t) \in W^{1, \infty}(\Omega)$ for all t , with its $W^{1, \infty}(\Omega)$ seminorm uniformly bounded by L_μ for all t .

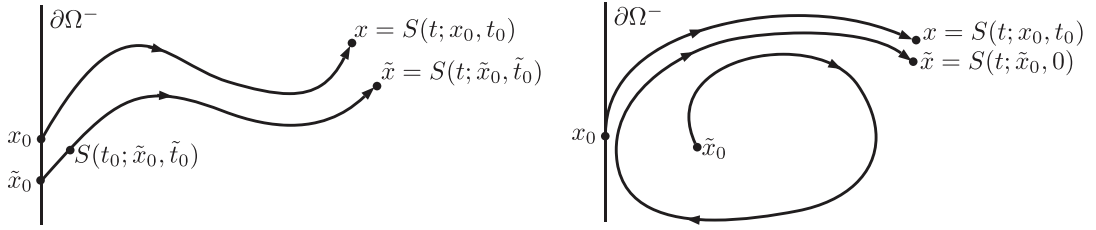


FIG. 1. Left: proof of Lemma 4.1. Right: vortex touching $\partial\Omega^-$ with $t_0 \gg \tilde{t}_0 = 0$, i.e. $t_0 - \tilde{t}_0$ is large, hence μ_1 is not Lipschitz continuous at x .

Now we show when μ_1 defined by (4.3) satisfies conditions (4.7), especially Lipschitz continuity in space which is necessary in the analysis. We note that since μ is obtained from μ_1 by multiplication by a constant, properties derived for μ_1 imply those for μ .

An obvious example of when μ_1 defined by (4.3) will not be Lipschitz continuous is when a vortex ‘touches’ $\partial\Omega^-$ as in Fig. 1. Then if (x, t) lies on a pathline originating at (x_0, t_0) such that $a(x_0, t_0)$ is tangent to $\partial\Omega$, we can find \tilde{x} arbitrarily close to x such that the corresponding pathline is much longer, perhaps winding several times around the vortex. Then $\mu_1(x, t) - \mu_1(\tilde{x}, t) = \tilde{t}_0 - t_0$ can be very large, while $\|x - \tilde{x}\|$ is arbitrarily small, hence μ_1 is not Lipschitz continuous. In fact μ_1 is discontinuous at (x, t) . In the following lemma, we show that inlet points where a is tangent to $\partial\Omega$ are the only troublemakers.

We note that since Ω is a bounded Lipschitz domain (hence quasiconvex), when proving Lipschitz continuity of μ_1 in space it is sufficient to prove local Lipschitz continuity of μ_1 in some neighborhood of each $x \in \overline{\Omega}$ with a Lipschitz constant independent of x .

LEMMA 4.1 Let $a \in L^\infty(Q_T)$ be continuous with respect to time and Lipschitz continuous with respect to space. Let there exist a constant $a_{\min} > 0$ such that

$$-a(x, t) \cdot \mathbf{n} \geq a_{\min} \quad (4.8)$$

for all $x \in \partial\Omega^-, t \in [0, T)$. Let μ_1 be defined by (4.3) on $\overline{\Omega} \times [0, T)$. Let the time any particle carried by the flow field $a(\cdot, \cdot)$ spends in Ω be uniformly bounded by \hat{T} . Then μ_1 satisfies assumption (4.7).

Proof. By definition, we have $\mu_1(x, t) \geq 0$ for all x, t . By the above considerations, μ_1 is bounded by the maximal time particles spend in Ω , which is uniformly bounded. This implies $\mu_1(x, t) \leq \hat{T} = \mu_{\max}$ for some μ_{\max} . Now we will prove Lipschitz continuity.

Let $t \in (0, T)$ be fixed and let $x, \tilde{x} \in \overline{\Omega}$ such that $|x - \tilde{x}| \leq \varepsilon$, where ε will be chosen sufficiently small in the following. Due to the assumptions on a , the pathlines passing through x and \tilde{x} are uniquely determined and originate at some (x_0, t_0) and $(\tilde{x}_0, \tilde{t}_0)$, respectively. In other words,

$$x = S(t; x_0, t_0), \quad \tilde{x} = S(t; \tilde{x}_0, \tilde{t}_0).$$

Without loss of generality, let $t_0 \geq \tilde{t}_0 > 0$, hence $x_0, \tilde{x}_0 \in \partial\Omega^-$. The case when $\tilde{t}_0 = 0$ can be treated similarly. Furthermore, we assume that x_0 is not a vertex of $\partial\Omega^-$ —this case will be treated at the end of the proof.

Since a is uniformly bounded and Lipschitz continuous in space, there exists δ , such that if $t \in [0, T]$ and $\text{dist}\{x, \partial\Omega^-\} \leq \delta$ then

$$-a(x, t) \cdot \mathbf{n} \geq a_{\min}/2, \quad (4.9)$$

due to (4.8). If ε is sufficiently small, then for the distance of the two considered pathlines at time t_0 we have $|x_0 - S(t_0; \tilde{x}_0, \tilde{t}_0)| = |S(t_0; x_0, t_0) - S(t_0; \tilde{x}_0, \tilde{t}_0)| \leq \delta$. Since $x_0 \in \partial\Omega^-$ this means that $S(t_0; \tilde{x}_0, \tilde{t}_0)$ is in the δ -neighborhood of $\partial\Omega^-$ and by (4.9), $\text{dist}\{S(\vartheta; \tilde{x}_0, \tilde{t}_0), \partial\Omega^-\}$ decreases as ϑ goes from t_0 to \tilde{t}_0 with a rate of at least $a_{\min}/2$ due to the uniformity of the bound (4.9). Therefore, $S(\vartheta; \tilde{x}_0, \tilde{t}_0)$ stays in the δ -neighborhood of $\partial\Omega^-$ for all $\vartheta \in [\tilde{t}_0, t_0]$ and

$$-a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) \cdot \mathbf{n} \geq a_{\min}/2$$

for all $\vartheta \in [\tilde{t}_0, t_0]$. Moreover, since x_0 lies in the interior of an edge of $\partial\Omega^-$, by choosing ε sufficiently small, we can ensure that \tilde{x}_0 also lies on this edge (face).

Now we estimate $|\mu_1(x, t) - \mu_1(\tilde{x}, t)| = |t_0 - \tilde{t}_0|$. We have

$$\begin{aligned} x - x_0 &= \int_{t_0}^t \frac{dS}{dt}(\vartheta; x_0, t_0) d\vartheta = \int_{t_0}^t a(S(\vartheta; x_0, t_0), \vartheta) d\vartheta, \\ \tilde{x} - \tilde{x}_0 &= \int_{\tilde{t}_0}^t \frac{dS}{dt}(\vartheta; \tilde{x}_0, \tilde{t}_0) d\vartheta = \int_{\tilde{t}_0}^t a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) d\vartheta. \end{aligned}$$

Subtracting these two identities gives us

$$x_0 - \tilde{x}_0 = x - \tilde{x} + \int_{\tilde{t}_0}^{t_0} a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) d\vartheta + \int_{t_0}^t a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) - a(S(\vartheta; x_0, t_0), \vartheta) d\vartheta. \quad (4.10)$$

If we consider \mathbf{n} , the normal to $\partial\Omega^-$ at x_0 , we see that $(x_0 - \tilde{x}_0) \cdot \mathbf{n} = 0$, as both x_0 and \tilde{x}_0 lie on the same edge on $\partial\Omega^-$. Therefore, if we multiply (4.10) by $-\mathbf{n}$, we get

$$\begin{aligned} 0 &= (x - \tilde{x}) \cdot (-\mathbf{n}) + \int_{\tilde{t}_0}^{t_0} a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) \cdot (-\mathbf{n}) d\vartheta \\ &\quad + \int_{t_0}^t (a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) - a(S(\vartheta; x_0, t_0), \vartheta)) \cdot (-\mathbf{n}) d\vartheta. \end{aligned} \quad (4.11)$$

Due to (4.9), we can estimate the first integral as

$$\int_{\tilde{t}_0}^{t_0} a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) \cdot (-\mathbf{n}) d\vartheta \geq \frac{a_{\min}}{2} (t_0 - \tilde{t}_0).$$

As for the second integral in (4.11), we have

$$\begin{aligned} & \left| \int_{t_0}^t (a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) - a(S(\vartheta; x_0, t_0), \vartheta)) \cdot (-\mathbf{n}) \, d\vartheta \right| \\ & \leq \widehat{TL}_a \sup_{\vartheta \in (t_0, t)} |S(\vartheta; \tilde{x}_0, \tilde{t}_0) - S(\vartheta; x_0, t_0)| \leq \widehat{TL}_a (e^{\widehat{TL}_a} - 1) \|x - \tilde{x}\|, \end{aligned} \quad (4.12)$$

where L_a is the Lipschitz constant of a with respect to x . The last inequality in (4.12) follows from standard results on ordinary differential equations, namely continuous dependence of the solution on the initial condition—here we consider the ordinary differential equations defining $S(\cdot; \cdot, \cdot)$ backward in time on the interval t_0, t with ‘initial’ conditions x and \tilde{x} at time t .

Finally, $|(x - \tilde{x}) \cdot (-\mathbf{n})| \leq \|x - \tilde{x}\|$. Therefore, we get from (4.11),

$$\frac{a_{\min}}{2} |\mu_1(x, t) - \mu_1(\tilde{x}, t)| = \frac{a_{\min}}{2} |t_0 - \tilde{t}_0| = \frac{a_{\min}}{2} (t_0 - \tilde{t}_0) \leq \left(1 + \widehat{TL}_a (e^{\widehat{TL}_a} - 1)\right) \|x - \tilde{x}\|. \quad (4.13)$$

Dividing by $a_{\min}/2 > 0$ gives Lipschitz continuity of $\mu_1(\cdot, t)$.

Now we return to the case when x_0 is a vertex of $\partial\Omega^-$. Reasoning as above, by choosing ε sufficiently small, we can ensure that $\tilde{x}_0 \in \partial\Omega^-$ is sufficiently close to x_0 , i.e. \tilde{x}_0 lies on one of the edges adjoining x_0 . Then we can again multiply (4.11) by $-\mathbf{n}$, the normal to $\partial\Omega^-$ at \tilde{x}_0 . Hence, $(\tilde{x} - \tilde{x}_0) \cdot (\mathbf{n}) = 0$ will also be satisfied and we can proceed as in the previous case. \square

REMARK 4.2 We note that the Lipschitz constant L_{μ_1} of μ_1 can be estimated from (4.13) by

$$L_{\mu_1} \leq \frac{2}{a_{\min}} \left(1 + \widehat{TL}_a (e^{\widehat{TL}_a} - 1)\right).$$

The Lipschitz constant L_μ of μ can then be obtained by multiplying by the constant factor from (4.4).

Having established uniform Lipschitz continuity of μ_1 in space, we can prove its Lipschitz continuity with respect to time. This implies the differentiability of μ_1 with respect to x and t a.e. in Q_T ; therefore the left-hand side of (4.1) is well defined a.e. in Q_T and this expression is equal to the derivative of μ_1 along pathlines. This is important, as all the following considerations are thus justified.

LEMMA 4.3 Let a satisfy the assumptions of Lemma 4.1. Then μ_1 is uniformly Lipschitz continuous with respect to time t : there exists $L_t \geq 0$ such that

$$|\mu(x, t) - \mu(x, \tilde{t})| \leq L_t |t - \tilde{t}|$$

for all $x \in \Omega$ and $t, \tilde{t} \in [0, T)$.

Proof. Let e.g. $\tilde{t} > t$ and let (x_0, t_0) and $(\tilde{x}_0, \tilde{t}_0)$ denote the origin of the pathlines passing through (x, t) and (x, \tilde{t}) , respectively. Therefore $x = S(t; x_0, t_0) = S(\tilde{t}; \tilde{x}_0, \tilde{t}_0)$. Denote $\tilde{x} = S(t; \tilde{x}_0, \tilde{t}_0)$. Then

$$|x - \tilde{x}| = \left| \int_t^{\tilde{t}} a(S(\vartheta; \tilde{x}_0, \tilde{t}_0), \vartheta) \, d\vartheta \right| \leq \|a\|_{L^\infty(Q_T)} |t - \tilde{t}|. \quad (4.14)$$

Therefore, by (4.3),

$$\begin{aligned} |\mu(x, t) - \mu(x, \tilde{t})| &\leq |t - \tilde{t}| + |\tilde{t}_0 - t_0| = |t - \tilde{t}| + |\mu(x, t) - \mu(\tilde{x}, t)| \leq |t - \tilde{t}| + L_\mu |x - \tilde{x}| \\ &\leq (1 + L_\mu \|a\|_{L^\infty(Q_T)}) |t - \tilde{t}| \end{aligned}$$

due to (4.14) and Lemma 4.1. This completes the proof. \square

EXAMPLE 4.4 In simple cases, the function μ_1 can be explicitly written down. As a trivial example, we take the one-dimensional stationary flow field $a(x, t) = x + 1$ on $\Omega = (0, 1)$ and the time interval $(0, +\infty)$. Then (4.2) can be easily solved to obtain

$$S(t; x_0, t_0) = (x_0 + 1)e^{t-t_0} - 1.$$

The (x, t) -plane is then separated into two regions separated by the pathline $S(t; 0, 0)$, which is the curve $x = e^t - 1$, i.e. $t = \ln(x + 1)$. For points (x, t) beneath this curve, i.e. $t \leq \ln(x + 1)$, we have $t_0 = 0$, hence $\mu_1(x, t) = t - t_0 = t$. For points (x, t) above the separation curve, we have $x_0 = 0$, hence

$$x = S(t; 0, t_0) = e^{t-t_0} - 1 \implies \mu_1(x, t) = t - t_0 = \ln(x + 1).$$

Altogether, we have

$$\mu_1(x, t) = \begin{cases} t & \text{if } t \leq \ln(x + 1), \\ \ln(x + 1) & \text{otherwise.} \end{cases}$$

This function is globally bounded and globally Lipschitz continuous. We note that the standard exponential scaling trick which gives exponential growth of the error corresponds to taking $\mu_1(x, t) = t$ for all x, t , which is an unbounded function.

5. DG method

Having introduced the necessary tools for the analysis using the general exponential scaling argument, we can finally proceed to the DG method itself.

Let \mathcal{T}_h be a triangulation of Ω , i.e. a partition of $\overline{\Omega}$ into a finite number of closed simplices with mutually disjoint interiors. As usual with DG, \mathcal{T}_h need not be conforming, i.e. hanging nodes are allowed. For $K \in \mathcal{T}_h$ we set $h_K = \text{diam}(K)$, $h = \max_{K \in \mathcal{T}_h} h_K$.

For each $K \in \mathcal{T}_h$ we define its *inflow* and *outflow* boundaries by

$$\begin{aligned} \partial K^-(t) &= \{x \in \partial K; a(x, t) \cdot \mathbf{n}(x) < 0\}, \\ \partial K^+(t) &= \{x \in \partial K; a(x, t) \cdot \mathbf{n}(x) \geq 0\}, \end{aligned}$$

where $\mathbf{n}(x)$ is the unit outer normal to ∂K . To simplify the notation, we shall usually omit the argument t in the following and write simply ∂K^\pm . Here we remark that \mathcal{T}_h need not conform to $\partial\Omega^-$. In other words, the intersection of $\partial\Omega^-$ and ∂K need not be an entire face (edge) of K .

The triangulation \mathcal{T}_h defines the *broken Sobolev space*

$$H^1(\mathcal{T}_h) = \left\{ v \in L^2(\Omega); v|_K \in H^1(K) \quad \forall K \in \mathcal{T}_h \right\}.$$

The approximate solution will be sought in the space of discontinuous piecewise polynomial functions

$$S_h = \{v_h; v_h|_K \in P^p(K) \quad \forall K \in \mathcal{T}_h\},$$

where $P^p(K)$ is the space of all polynomials on K of degree at most $p \in \mathbb{N}$.

Given an element $K \in \mathcal{T}_h$, for $v_h \in S_h$ we define v_h^- as the trace of v_h on ∂K from the side of the element adjacent to K . Furthermore, on $\partial K \setminus \partial\Omega$ we define the *jump* of v_h as $[v_h] = v_h - v_h^-$, where v_h is the trace from inside K .

The DG formulation of (2.1) then reads as follows: we seek $u_h \in C^1([0, T]; S_h)$ such that $u_h(0) = u_h^0$, an S_h -approximation of the initial condition u^0 , and for all $t \in (0, T)$,

$$\left(\frac{\partial u_h}{\partial t}, v_h \right) + b_h(u_h, v_h) + c_h(u_h, v_h) = l_h(v_h) \quad \forall v_h \in S_h. \quad (5.1)$$

Here b_h , c_h and l_h are bilinear and linear forms, respectively, defined for $u, v \in H^1(\mathcal{T}_h)$ as follows. The bilinear *advection form* b_h is

$$b_h(u, v) = \sum_{K \in \mathcal{T}_h} \int_K (a \cdot \nabla u) v \, dx - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial\Omega} (a \cdot \mathbf{n}) [u] v \, dS - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial\Omega} (a \cdot \mathbf{n}) u v \, dS,$$

the *reaction form* is defined by

$$c_h(u, v) = \int_{\Omega} c u v \, dx$$

and l_h is the *right-hand side form*

$$l_h(v) = - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial\Omega} (a \cdot \mathbf{n}) u_D v \, dx.$$

The definition of b_h corresponds to the concept of upwinding which has already been applied in the context of DG in the original paper by [Reed & Hill \(1973\)](#). The right-hand side form l_h is based on the weak enforcement of the Dirichlet boundary condition on $\partial\Omega^-$ by penalization. For a detailed derivation of the specific forms of b_h and l_h as used in the paper, we refer e.g. to the study by [Feistauer & Švadlenka \(2004\)](#).

6. Analysis of the advection and reaction terms

In this section, we prove the estimates of the bilinear forms b_h and c_h necessary for the error analysis. First, we need some standard results from approximation theory.

6.1 Auxiliary results

In the following analysis we assume that the exact solution u is sufficiently regular, namely

$$u, u_t := \frac{\partial u}{\partial t} \in L^2(H^{p+1}).$$

We consider a system $\{\mathcal{T}_h\}_{h \in (0, h_0)}$, $h_0 > 0$ of nonconforming triangulations of Ω that are *shape regular* and satisfy the *inverse assumption*, cf. Ciarlet (1980). Under these assumptions we have the following standard results:

LEMMA 6.1 (Inverse inequality). There exists a constant $C_I > 0$ independent of h , K such that for all $K \in \mathcal{T}_h$, and all $v \in P^p(K)$,

$$|v|_{H^1(K)} \leq C_I h_K^{-1} \|v\|_{L^2(K)}.$$

For $v \in L^2(\Omega)$ we denote by $\Pi_h v \in S_h$ the $L^2(\Omega)$ -projection of v onto S_h :

$$(\Pi_h v - v, \varphi_h) = 0 \quad \forall \varphi_h \in S_h. \quad (6.1)$$

Let $\eta_h(t) = u(t) - \Pi_h u(t)$ and $\xi_h(t) = \Pi_h u(t) - u_h(t) \in S_h$. Then we can write the error of the method as $e_h(t) := u(t) - u_h(t) = \eta_h(t) + \xi_h(t)$. For simplicity, we shall usually drop the subscript h . We have the following standard approximation result; cf. Ciarlet (1980).

LEMMA 6.2 There exists a constant $C_\Pi > 0$ independent of h , K such that for all $h \in (0, h_0)$,

$$\begin{aligned} \|\eta(t)\|_{L^2(K)} &\leq C_\Pi h^{p+1} |u(t)|_{H^{p+1}(K)}, \\ |\eta(t)|_{H^1(K)} &\leq C_\Pi h^p |u(t)|_{H^{p+1}(K)}, \\ \|\eta(t)\|_{L^2(\partial K)} &\leq C_\Pi h^{p+1/2} |u(t)|_{H^{p+1}(K)}, \\ \left\| \frac{\partial \eta(t)}{\partial t} \right\|_{L^2(K)} &\leq C_\Pi h^{p+1} |u_t(t)|_{H^{p+1}(K)}. \end{aligned}$$

Throughout the paper, C will be a generic constant independent of h , t , T . In order to track the dependence of the estimates on the function μ , we will also assume that the generic constant C is independent of μ , particularly μ_{\max} and L_μ from (4.7), and we will explicitly track these dependencies. The dependence of the resulting estimates of Theorem 7.1 on μ will then be clarified in Section 7 for the specific construction of μ from Section 4.

6.2 Estimates of the advection and reaction terms

In our analysis we will assume there exists a constant γ_0 and a function $\mu : Q_T \rightarrow \mathbb{R}$ such that the ellipticity assumption (3.9) holds. Furthermore, we assume that μ satisfies (4.7), i.e. μ is non-negative, bounded and Lipschitz continuous in space. Such a function μ was constructed in Section 4.

Similarly to Section 2, we wish to write $\xi(x, t) = e^{\mu(x, t)} \tilde{\xi}(x, t)$ and test the error equation with $\phi(x, t) = e^{-\mu(x, t)} \tilde{\xi}(x, t) = e^{-2\mu(x, t)} \xi(x, t)$ to obtain estimates for $\tilde{\xi}$. However, since $\phi(t) \notin S_h$ this is not possible. One possibility is to test by $\Pi_h \phi(t) \in S_h$ and estimate the resulting difference $\Pi_h \phi(t) - \phi(t)$.

This is done in the stationary case in the study by [Ayuso & Marini \(2009\)](#) and the analysis is carried out under the assumption $\mu \in W^{p+1,\infty}(\Omega)$. Such high regularity can be achieved by mollification of μ . However, this would be somewhat technical in the evolutionary case, as space-time smoothing would be required in which case the dependence of all constants on T must be carefully considered. Also Q_T is potentially an unbounded domain (for $T = +\infty$) which leads to technical difficulties. Here we carry out the analysis under the weaker assumption (4.7), i.e. $\mu(t) \in W^{1,\infty}(\Omega)$.

LEMMA 6.3 Let μ satisfy assumptions (4.7). Let $\phi(x, t) = e^{-\mu(x,t)} \tilde{\xi}(x, t) = e^{-2\mu(x,t)} \xi(x, t)$, where $\xi(t) \in S_h$. Then there exists C independent of $h, t, \xi, \tilde{\xi}$ and μ such that

$$\begin{aligned} \|\Pi_h \phi(t) - \phi(t)\|_{L^2(K)} &\leq CL_\mu e^{h_K L_\mu} h_K \max_{x \in K} e^{-\mu(x,t)} \|\tilde{\xi}(t)\|_{L^2(K)}, \\ \|\Pi_h \phi(t) - \phi(t)\|_{L^2(\partial K)} &\leq CL_\mu e^{h_K L_\mu} h_K^{1/2} \max_{x \in K} e^{-\mu(x,t)} \|\tilde{\xi}(t)\|_{L^2(K)}. \end{aligned} \quad (6.2)$$

Proof. Let x_K be the centroid of K . On element K we introduce the constant $\mu_K(t) = \mu(x_K, t)$; then the function $e^{-2\mu_K(t)} \xi(\cdot, t)$ lies in $P^p(K)$, hence is fixed by the projection Π_h . Therefore,

$$\begin{aligned} \Pi_h \phi(t) - \phi(t) &= \Pi_h \left(e^{-2\mu(t)} \xi(t) - e^{-2\mu_K(t)} \xi(t) \right) - \left(e^{-2\mu(t)} \xi(t) - e^{-2\mu_K(t)} \xi(t) \right) \\ &= \Pi_h w(t) - w(t), \end{aligned}$$

where $w(x, t) = (e^{-2\mu(x,t)} - e^{-2\mu(x_K,t)}) \xi(x, t)$. Standard estimates of the interpolation error of Π_h give

$$\|\Pi_h \phi(t) - \phi(t)\|_{L^2(K)}^2 = \|\Pi_h w(t) - w(t)\|_{L^2(K)}^2 \leq Ch_K^2 |w(t)|_{H^1(K)}^2. \quad (6.3)$$

For the right-hand side seminorm we have

$$\begin{aligned} |w(t)|_{H^1(K)}^2 &= \int_K \left| \nabla \left((e^{-2\mu(x,t)} - e^{-2\mu(x_K,t)}) \xi(x, t) \right) \right|^2 dx \\ &\leq 2 \int_K \left| \nabla e^{-2\mu(x,t)} \xi(x, t) \right|^2 dx + 2 \int_K \left| (e^{-2\mu(x,t)} - e^{-2\mu(x_K,t)}) \nabla \xi(x, t) \right|^2 dx. \end{aligned}$$

If $\mu(t) \in C^1(\overline{\Omega})$, by the mean value theorem

$$\left| e^{-2\mu(x_K,t)} - e^{-2\mu(x,t)} \right| \leq h_K \left| \nabla e^{-2\mu(\zeta,t)} \right| = h_K e^{-2\mu(\zeta,t)} 2 |\nabla \mu(\zeta, t)|$$

for some point ζ on the line between x and x_K . Therefore, by the inverse inequality,

$$\begin{aligned} |w(t)|_{H^1(K)}^2 &\leq 2 \int_K 4e^{-4\mu(x,t)} |\nabla \mu(x,t)|^2 |\xi(x,t)|^2 dx + 2 \int_K h_K^2 e^{-4\mu(\zeta,t)} 4 |\nabla \mu(\zeta,t)|^2 |\nabla \xi(x,t)|^2 dx \\ &\leq 8 |\mu(t)|_{W^{1,\infty}}^2 \int_K e^{-4\mu(x,t)} e^{2\mu(x,t)} |\tilde{\xi}(x,t)|^2 dx + 8 C_I^2 \max_{x \in K} e^{-4\mu(x,t)} |\mu(t)|_{W^{1,\infty}}^2 \|\xi(t)\|^2 \\ &\leq 8 L_\mu^2 \max_{x \in K} e^{-2\mu(x,t)} e^{2h_K L_\mu} \|\tilde{\xi}(t)\|^2 + 8 C_I^2 \max_{x \in K} e^{-2\mu(x,t)} e^{2h_K L_\mu} L_\mu^2 \|\tilde{\xi}(t)\|^2. \end{aligned}$$

Substituting into (6.3) gives us the first inequality in (6.2) for $\mu(t) \in C^1(\overline{\Omega})$. The case $\mu(t) \in W^{1,\infty}(\Omega)$ follows by a density argument. The second inequality in (6.2) can be obtained similarly, only intermediately applying the trace inequality $\|\xi(t)\|_{L^2(\partial K)} \leq C h_K^{-1/2} \|\xi(t)\|_{L^2(K)}$. \square

REMARK 6.4 The estimates of Lemma 6.3 are effectively $\mathcal{O}(h_K) \|\tilde{\xi}\|$ and $\mathcal{O}(h_K^{1/2}) \|\tilde{\xi}\|$ estimates, respectively. We note that the dependence of the estimates on L_μ is rather mild, effectively linear, due to the small factor h_K in the exponent.

Now we shall estimate individual terms in the DG formulation. Due to the consistency of the DG scheme, the exact solution u also satisfies (5.1). We subtract the formulations for u and u_h to obtain the error equation

$$\left(\frac{\partial \xi}{\partial t}, v_h \right) + \left(\frac{\partial \eta}{\partial t}, v_h \right) + b_h(\xi, v_h) + b_h(\eta, v_h) + c_h(\xi, v_h) + c_h(\eta, v_h) = 0 \quad (6.4)$$

for all $v_h \in S_h$.

As stated earlier we want to test (6.4) by $\phi(x, t) = e^{-\mu(x,t)} \tilde{\xi}(x, t)$; however, $\phi(t) \notin S_h$. We therefore set $v_h = \Pi_h \phi(t)$ and estimate the difference using Lemma 6.3. We write (6.4) as

$$\begin{aligned} &\left(\frac{\partial \xi}{\partial t}, \Pi_h \phi \right) + b_h(\xi, \phi) + b_h(\xi, \Pi_h \phi - \phi) + b_h(\eta, \phi) + b_h(\eta, \Pi_h \phi - \phi) \\ &+ c_h(\xi, \phi) + c_h(\xi, \Pi_h \phi - \phi) + c_h(\eta, \Pi_h \phi) + \left(\frac{\partial \eta}{\partial t}, \Pi_h \phi \right) = 0. \end{aligned} \quad (6.5)$$

We will estimate the individual terms of (6.5) in a series of lemmas. For this purpose, we introduce the following norm on a subset ω of ∂K or $\partial \Omega$:

$$\|f\|_{a,\omega} = \left\| \sqrt{a \cdot \mathbf{n}} f \right\|_{L^2(\omega)},$$

where \mathbf{n} is the outer normal to ∂K or $\partial \Omega$. We will usually omit the argument t to simplify the notation.

In the following lemma we show that selected terms from (6.5) indeed give us the desired ellipticity similarly to (3.8) and (3.9) plus additional elliptic terms due to the use of the upwind flux.

LEMMA 6.5 Let $\xi = e^\mu \tilde{\xi}$, $\phi = e^{-\mu} \tilde{\xi}$ as above and let μ satisfy assumptions (3.9) and (4.7). Then

$$\left(\frac{\partial \xi}{\partial t}, \Pi_h \phi \right) + b_h(\xi, \phi) + c_h(\xi, \phi) \geq \frac{1}{2} \frac{d}{dt} \|\tilde{\xi}\|^2 + \gamma_0 \|\tilde{\xi}\|^2 + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \left(\|\tilde{\xi}\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}\|_{a, \partial K \cap \partial \Omega}^2 \right). \quad (6.6)$$

Proof. Since $\partial \xi / \partial t \in S_h$ for each t , by definition (6.1) of Π_h we have

$$\left(\frac{\partial \xi}{\partial t}, \Pi_h \phi \right) = \left(\frac{\partial \xi}{\partial t}, \phi \right) = \left(e^\mu \frac{\partial \tilde{\xi}}{\partial t} + e^\mu \frac{\partial \mu}{\partial t} \tilde{\xi}, e^{-\mu} \tilde{\xi} \right) = \frac{1}{2} \frac{d}{dt} \|\tilde{\xi}\|^2 + \left(\frac{\partial \mu}{\partial t} \tilde{\xi}, \tilde{\xi} \right). \quad (6.7)$$

The reactive term satisfies

$$c_h(\xi, \phi) = \int_{\Omega} c \xi \phi \, dx = \int_{\Omega} c e^\mu \tilde{\xi} e^{-\mu} \tilde{\xi} \, dx = \int_{\Omega} c \tilde{\xi}^2 \, dx. \quad (6.8)$$

From the definition of b_h , we get

$$\begin{aligned} b_h(\xi, \phi) &= \sum_{K \in \mathcal{T}_h} \int_K a \cdot (\nabla \mu \tilde{\xi} + \nabla \tilde{\xi}) e^\mu e^{-\mu} \tilde{\xi} \, dx - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [e^\mu \tilde{\xi}] e^{-\mu} \tilde{\xi} \, dS \\ &\quad - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) e^\mu \tilde{\xi} e^{-\mu} \tilde{\xi} \, dS \\ &= \sum_{K \in \mathcal{T}_h} \int_K a \cdot (\nabla \mu \tilde{\xi} + \nabla \tilde{\xi}) \tilde{\xi} \, dx - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [\tilde{\xi}] \tilde{\xi} \, dS - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS. \end{aligned} \quad (6.9)$$

By Green's theorem,

$$\int_K a \cdot \nabla \tilde{\xi} \tilde{\xi} \, dx = -\frac{1}{2} \int_K \operatorname{div} a \tilde{\xi}^2 \, dx + \frac{1}{2} \int_{\partial K} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS. \quad (6.10)$$

Splitting the last integral over the separate parts of ∂K , i.e. $\partial K^- \setminus \partial \Omega$, $\partial K^- \cap \partial \Omega$, $\partial K^+ \setminus \partial \Omega$ and $\partial K^+ \cap \partial \Omega$, and by substituting (6.10) into (6.9), we get

$$\begin{aligned} b_h(\xi, \phi) &= \sum_{K \in \mathcal{T}_h} \int_K \left(a \cdot \nabla \mu - \frac{1}{2} \operatorname{div} a \right) \tilde{\xi}^2 \, dx \\ &\quad + \sum_{K \in \mathcal{T}_h} \left(-\frac{1}{2} \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) (\tilde{\xi}^2 - 2\tilde{\xi} \tilde{\xi}^-) \, dS - \frac{1}{2} \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS \right. \\ &\quad \left. + \frac{1}{2} \int_{\partial K^+ \setminus \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS + \frac{1}{2} \int_{\partial K^+ \cap \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS \right). \end{aligned} \quad (6.11)$$

We note that

$$\sum_{K \in \mathcal{T}_h} \int_{\partial K^+ \setminus \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi}^2 \, dS = - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) (\tilde{\xi}^-)^2 \, dS. \quad (6.12)$$

Using the facts that $\tilde{\xi}^2 - 2\tilde{\xi}\tilde{\xi}^- + (\tilde{\xi}^-)^2 = [\tilde{\xi}]^2$ and $-a \cdot \mathbf{n} = |a \cdot \mathbf{n}|$ on ∂K^- and $a \cdot \mathbf{n} = |a \cdot \mathbf{n}|$ on ∂K^+ , we get (6.6) by substituting (6.12) into (6.11) and applying assumption (3.9) in the resulting interior terms along with (6.7), (6.8). \square

The following series of lemmas deals with estimating the remaining terms in (6.5). These estimates more or less follow standard procedures with several important differences pertaining to the specific choice of test functions ϕ and $\Pi_h \phi - \phi$. These include the inability to apply the inverse inequality to $\tilde{\xi} \notin S_h$ but only to $\xi \in S_h$ and Lemma 6.3. In order to keep the main body of the text less cluttered, we leave the proofs of the following lemmas to the Appendix.

LEMMA 6.6 The advection terms satisfy

$$|b_h(\xi, \Pi_h \phi - \phi)| \leq C(1 + L_\mu)^2 e^{4hL_\mu} h \|\tilde{\xi}\|^2 + \frac{1}{8} \sum_{K \in \mathcal{T}_h} \left(\|[\tilde{\xi}]\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}\|_{a, \partial K \cap \partial \Omega}^2 \right), \quad (6.13)$$

$$|b_h(\eta, \phi)| \leq C(1 + L_\mu)^2 h \|\tilde{\xi}\|^2 + Ch^{2p+1} |u(t)|_{H^{p+1}}^2 + \frac{1}{8} \sum_{K \in \mathcal{T}_h} \left(\|[\tilde{\xi}]\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}\|_{a, \partial K \cap \partial \Omega}^2 \right), \quad (6.14)$$

$$|b_h(\eta, \Pi_h \phi - \phi)| \leq CL_\mu^2 e^{2hL_\mu} h \|\tilde{\xi}\| + Ch^{2p+1} |u(t)|_{H^{p+1}}^2. \quad (6.15)$$

LEMMA 6.7 The reaction terms satisfy

$$\begin{aligned} |c_h(\xi, \Pi_h \phi - \phi)| &\leq CL_\mu e^{2hL_\mu} h \|\tilde{\xi}\|^2, \\ |c_h(\eta, \Pi_h \phi)| &\leq C(1 + L_\mu e^{hL_\mu})^2 h \|\tilde{\xi}\|^2 + Ch^{2p+1} |u(t)|_{H^{p+1}}^2. \end{aligned}$$

LEMMA 6.8 The evolutionary term satisfies

$$\left| \left(\frac{\partial \eta}{\partial t}, \Pi_h \phi \right) \right| \leq C(1 + L_\mu e^{hL_\mu})^2 h \|\tilde{\xi}\|^2 + Ch^{2p+1} |u_t(t)|_{H^{p+1}}^2.$$

7. Error analysis

Finally, we come to the error analysis. The starting point is the error identity (6.5) to which we apply the derived estimates of its individual terms.

THEOREM 7.1 Let there exist a function $\mu : Q_T \rightarrow [0, \mu_{\max}]$ for some constant μ_{\max} , such that $\mu(t) \in W^{1,\infty}(\Omega)$ and let there exist a constant $\gamma_0 > 0$ such that the coefficients of (2.1) satisfy $\mu_t + a \cdot \nabla \mu + c - \frac{1}{2} \operatorname{div} a \geq \gamma_0 > 0$ a.e. in Q_T . Let the initial condition u_h^0 satisfy $\|\Pi_h u^0 - u_h^0\| \leq Ch^{p+1/2} |u^0|_{H^{p+1}}$.

Then there exists a constant C independent of μ , h and T such that for h sufficiently small the error of the DG scheme (5.1) satisfies

$$\begin{aligned} & \max_{t \in [0, T]} \|e_h(t)\| + \sqrt{\gamma_0} \|e_h\|_{L^2(Q_T)} + \left(\frac{1}{2} \int_0^T \sum_{K \in \mathcal{T}_h} \left(\|[e_h(\vartheta)]\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|e_h(\vartheta)\|_{a, \partial K \cap \partial \Omega}^2 \right) d\vartheta \right)^{1/2} \\ & \leq C e^{\mu_{\max}} h^{p+1/2} \left(|u^0|_{H^{p+1}} + |u|_{L^2(H^{p+1})} + |u_t|_{L^2(H^{p+1})} + h^{1/2} |u|_{L^\infty(H^{p+1})} \right). \end{aligned} \quad (7.1)$$

Proof. Applying Lemmas 6.5–6.8 to (6.5), multiplying by 2 for convenience and collecting similar terms, we get for all t ,

$$\begin{aligned} & \frac{d}{dt} \|\tilde{\xi}(t)\|^2 + 2\gamma_0 \|\tilde{\xi}(t)\|^2 + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \left(\|\tilde{\xi}(t)\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}(t)\|_{a, \partial K \cap \partial \Omega}^2 \right) \\ & \leq C_1 (1 + L_\mu)^2 e^{4hL_\mu} h \|\tilde{\xi}(t)\|^2 + C_2 h^{2p+1} \left(|u(t)|_{H^{p+1}}^2 + |u_t(t)|_{H^{p+1}}^2 \right), \end{aligned} \quad (7.2)$$

where we have used $C_1(1 + L_\mu)^2 e^{4hL_\mu}$ as a common upper bound for the constants from Lemmas 6.6–6.8 at the $\mathcal{O}(h) \|\tilde{\xi}(t)\|^2$ terms. Here C_1 and C_2 are independent of h , t , T , μ and $\tilde{\xi}$.

Now if h is sufficiently small so that

$$C_1 (1 + L_\mu)^2 e^{4hL_\mu} h \leq \gamma_0, \quad (7.3)$$

then the first right-hand side term can be absorbed by the left-hand side term $2\gamma_0 \|\tilde{\xi}(t)\|^2$:

$$\begin{aligned} & \frac{d}{dt} \|\tilde{\xi}(t)\|^2 + \gamma_0 \|\tilde{\xi}(t)\|^2 + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \left(\|\tilde{\xi}(t)\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}(t)\|_{a, \partial K \cap \partial \Omega}^2 \right) \\ & \leq C_2 h^{2p+1} \left(|u(t)|_{H^{p+1}}^2 + |u_t(t)|_{H^{p+1}}^2 \right). \end{aligned}$$

Substituting $t = \vartheta$ and integrating over $(0, t)$ give us

$$\begin{aligned} & \|\tilde{\xi}(t)\|^2 + \gamma_0 \int_0^t \|\tilde{\xi}(\vartheta)\|^2 d\vartheta + \frac{1}{2} \int_0^t \sum_{K \in \mathcal{T}_h} \left(\|\tilde{\xi}(\vartheta)\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}(\vartheta)\|_{a, \partial K \cap \partial \Omega}^2 \right) d\vartheta \\ & \leq \|\tilde{\xi}(0)\|^2 + C_2 h^{2p+1} \left(|u|_{L^2(0, t; H^{p+1})}^2 + |u_t|_{L^2(0, t; H^{p+1})}^2 \right) \\ & \leq C h^{2p+1} \left(|u^0|_{H^{p+1}}^2 + |u|_{L^2(0, t; H^{p+1})}^2 + |u_t|_{L^2(0, t; H^{p+1})}^2 \right), \end{aligned} \quad (7.4)$$

since the assumptions give

$$\|\tilde{\xi}(0)\|^2 \leq \|\xi(0)\|^2 = \|\Pi_h u^0 - u_h^0\|^2 \leq C h^{2p+1} |u^0|_{H^{p+1}}^2.$$

Now we reformulate estimate (7.4) as an estimate of ξ instead of $\tilde{\xi}$. Because $\tilde{\xi} = e^{-\mu}\xi$, we can estimate

$$\|\tilde{\xi}(t)\|^2 \geq \min_{Q_T} e^{-2\mu(x,\vartheta)} \|\xi(t)\|^2 = e^{-2\max_{Q_T} \mu(x,\vartheta)} \|\xi(t)\|^2 = e^{-2\mu_{\max}} \|\xi(t)\|^2$$

and similarly for the remaining left-hand side norms in (7.4). Substituting into (7.4) and multiplying by $e^{2\mu_{\max}}$ gives us

$$\begin{aligned} \|\xi(t)\|^2 + \gamma_0 \|\xi\|_{L^2(0,t;L^2)}^2 + \frac{1}{2} \int_0^t \sum_{K \in \mathcal{T}_h} \left(\|\xi(\vartheta)\|_{a,\partial K^- \setminus \partial\Omega}^2 + \|\xi(\vartheta)\|_{a,\partial K \cap \partial\Omega}^2 \right) d\vartheta \\ \leq Ch^{2p+1} e^{2\mu_{\max}} \left(|u^0|_{H^{p+1}}^2 + |u|_{L^2(0,t;H^{p+1})}^2 + |u_t|_{L^2(0,t;H^{p+1})}^2 \right). \end{aligned}$$

By taking the square root of the last inequality and taking the maximum over all $t \in [0, T]$, we get

$$\begin{aligned} \max_{t \in [0, T]} \|\xi(t)\| + \sqrt{\gamma_0} \|\xi\|_{L^2(L^2)} + \left(\frac{1}{2} \int_0^T \sum_{K \in \mathcal{T}_h} \left(\|\xi(\vartheta)\|_{a,\partial K^- \setminus \partial\Omega}^2 + \|\xi(\vartheta)\|_{a,\partial K \cap \partial\Omega}^2 \right) d\vartheta \right)^{1/2} \\ \leq Ch^{p+1/2} e^{\mu_{\max}} \left(|u^0|_{H^{p+1}} + |u|_{L^2(H^{p+1})} + |u_t|_{L^2(H^{p+1})} \right) \end{aligned} \quad (7.5)$$

which is estimate (7.1) for the discrete part ξ of the error. Lemma 6.2 gives a similar inequality for η :

$$\begin{aligned} \max_{t \in [0, T]} \|\eta(t)\| + \sqrt{\gamma_0} \|\eta\|_{L^2(L^2)} + \left(\frac{1}{2} \int_0^T \sum_{K \in \mathcal{T}_h} \left(\|\eta(\vartheta)\|_{a,\partial K^- \setminus \partial\Omega}^2 + \|\eta(\vartheta)\|_{a,\partial K \cap \partial\Omega}^2 \right) d\vartheta \right)^{1/2} \\ \leq Ch^{p+1/2} \left(h^{1/2} |u|_{L^\infty(H^{p+1})} + (h^{1/2} + 1) |u|_{L^2(H^{p+1})} \right). \end{aligned} \quad (7.6)$$

Since $e_h = \xi + \eta$, the triangle inequality along with (7.5) and (7.6) gives us the statement of the theorem. \square

We now interpret the results of Theorem 7.1 given the specific construction of the scaling function μ from Section 4.

If the assumptions of Lemma 4.1 are satisfied, the function μ_1 exists and has suitable regularity. If we choose γ_0 (this can be chosen e.g. as $\gamma_0 = 1$ for simplicity), then μ is constructed by the scaling (4.4). As μ_1 is bounded by \widehat{T} , the exponential factor in (7.1) becomes

$$e^{\mu_{\max}} = e^{\widehat{T}(|\inf_{Q_T} (c - \frac{1}{2} \operatorname{div} a)^-| + \gamma_0)}.$$

Now we turn to the requirement ‘ h is sufficiently small’ in Theorem 7.1. Specifically, (7.3) must be satisfied. If we denote $A := C_1 (1 + L_\mu)^2 / \gamma_0$ and $B := 4L_\mu$, then condition (7.3) reads

$$Ae^{Bh} h \leq 1. \quad (7.7)$$

It can be easily seen that if $h \leq \frac{1}{2} \min \{A^{-1}, B^{-1}\}$, then

$$Ae^{Bh}h \leq \frac{1}{2}e^{1/2} \leq 1,$$

i.e. (7.7) holds. Taking the definitions of A and B , we get the restriction on h ,

$$h \leq \frac{1}{2} \min \left\{ \frac{\gamma_0}{C_1 (1 + L_\mu)^2}, \frac{1}{4L_\mu} \right\}. \quad (7.8)$$

If we take e.g. $\gamma_0 = 1$ as mentioned and assume $C_1 \geq 1$ (if not, C_1 can be increased to 1 in (7.2)), condition (7.8) reduces to

$$h \leq \frac{1}{2C_1 (1 + L_\mu)^2}, \quad (7.9)$$

where L_μ is the Lipschitz constant of μ from Lemma 4.1, which can be estimated as in Remark 4.2 using data from the continuous problem (2.1) such as L_a , the spatial Lipschitz constant of a . Therefore the condition (7.8) on the mesh size and the final error estimate can be formulated without μ , only in terms of \hat{T} and the data of the problem. Thus we can state the following corollary derived from Theorem 7.1 using the construction of μ from Section 4.

COROLLARY 7.2 Let $a \in L^\infty(Q_T)$ be continuous with respect to time and Lipschitz continuous with respect to space. Let there exist a constant $a_{\min} > 0$ such that $-a(x, t) \cdot \mathbf{n} \geq a_{\min}$ for all $x \in \partial\Omega^-$, $t \in [0, T]$. Let the time any particle carried by the flow field $a(\cdot, \cdot)$ spends in Ω be uniformly bounded by \hat{T} . Let the initial condition u_h^0 satisfy $\|\Pi_h u^0 - u_h^0\| \leq Ch^{p+1/2} |u^0|_{H^{p+1}}$. Then there exist constants $c > 0$ and C independent of h and T such that for $h \leq c$, the error of the DG scheme (5.1) satisfies

$$\begin{aligned} & \max_{t \in [0, T]} \|e_h(t)\| + \|e_h\|_{L^2(Q_T)} + \left(\frac{1}{2} \int_0^T \sum_{K \in \mathcal{T}_h} \left(\| [e_h(\vartheta)] \|_{a, \partial K^- \setminus \partial\Omega}^2 + \| e_h(\vartheta) \|_{a, \partial K \cap \partial\Omega}^2 \right) d\vartheta \right)^{1/2} \\ & \leq C e^{\hat{T}(|\inf_{Q_T} (c - \frac{1}{2} \operatorname{div} a)^-| + 1)} h^{p+1/2} \left(|u^0|_{H^{p+1}} + |u|_{L^2(H^{p+1})} + |u_t|_{L^2(H^{p+1})} + h^{1/2} |u|_{L^\infty(H^{p+1})} \right). \end{aligned} \quad (7.10)$$

REMARK 7.3 As mentioned in Section 4, estimate (7.10) is exponential in the maximal ‘lifetime’ of particles \hat{T} , instead of the final time T , which would be expected when applying Gronwall’s lemma directly. Since we assume the maximal particle lifetime to be finite, the exponential factor in (7.10) remains uniformly bounded independent of time T , which can be infinite.

8 Conclusion and future work

In this paper we derived *a priori* error estimates for a linear advection–reaction problem with inlet and outlet boundary conditions in the $L^\infty(L^2)$ - and $L^2(L^2)$ -norms. Unlike previous works, the analysis was performed without the usual ellipticity assumption $c - \frac{1}{2} \operatorname{div} a \geq 0$. This is achieved by applying a general exponential scaling transformation in space and time to the exact and discrete solutions of the problem. We considered the case when the time spent by particles carried by the flow field inside the spatial

domain Ω is uniformly bounded by some \hat{T} . The resulting error estimates are of the order $Ch^{p+1/2}$, where C depends exponentially on \hat{T} (which is a constant) and not on the final time $T \leq +\infty$, as would be expected from the use of Gronwall's inequality. Effectively, due to the exponential scaling, we apply Gronwall's lemma in the Lagrangian setting along pathlines, which exist only for time at most \hat{T} , and not in the usual Eulerian sense.

As for future work, we plan to extend the analysis to fully discrete DG schemes with discretization in time. Furthermore, we wish to extend the analysis to nonlinear convective problems, following the arguments of Zhang & Shu (2004) and Kučera (2014) to obtain error estimates without the exponential dependence on time of the error estimates.

Funding

The work of V. Kučera was supported by the J. William Fulbright Commission in the Czech Republic and research project No. 17-01747S of the Czech Science Foundation. The work of C.-W. Shu was supported by DOE grant DE-FG02-08ER25863 and NSF grants DMS-1418750 and DMS-1719410.

REFERENCES

- AYUSO, B. & MARINI, L. D. (2009) Discontinuous Galerkin methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.*, **47**, 1391–1420.
- BENZONI-GAVAGE, S. & SERRE, D. (2007) *Multi-Dimensional Hyperbolic Partial Differential Equations: First-Order Systems and Applications*. Oxford Mathematical Monographs. Oxford: Oxford University Press.
- CIARLET, P. G. (1980) *The finite element method for elliptic problems*. Studies in Mathematics and its Applications. Amsterdam, New York: North-Holland.
- DEVINATZ, A., ELLIS, R. & FRIEDMAN, A. (1974) The asymptotic behavior of the first real eigenvalue of second order elliptic operators with a small parameter in the highest derivatives, II. *Indiana Univ. Math. J.*, **23**, 991–1011.
- EVANS, L. C. (2010) *Partial Differential Equations*. Providence, RI: American Mathematical Society.
- FEISTAUER, M. & ŠVADLENKA, K. (2004) Discontinuous Galerkin method of lines for solving nonstationary singularly perturbed linear problems. *J. Numer. Math.*, **12**, 97–117.
- JOHNSON, C. & PITKÄRANTA, J. (1986) An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, **46**, 1–26.
- KUČERA, V. (2014) On diffusion-uniform error estimates for the DG method applied to singularly perturbed problems. *IMA J. Numer. Anal.*, **34**, 820–861.
- LARSSON, S. & THOMÉE, V. (2003) *Partial Differential Equations with Numerical Methods*. Texts in Applied Mathematics. Berlin: Springer.
- LESAINT, P. & RAVIART, P. A. (1974) On a finite element method for solving the neutron transport equation. *Mathematical Aspects of Finite Elements in Partial Differential Equations* (C. de Boor ed). New York: Academic Press, pp. 89–145.
- MIZOHATA, S. (1973) *The Theory of Partial Differential Equations*. Cambridge: Cambridge University Press.
- NÄVERT, U. (1982) A finite element method for convection-diffusion problems. *Ph.D. Thesis*, Chalmers University of Technology.
- REED, W. H. & HILL, T. (1973) Triangular mesh methods for the neutron transport equation. *Los Alamos Report LA-UR-73-479*.
- ROOS, H.-G., STYNES, M. & TOBISKA, L. (2008) *Robust Numerical Methods for Singularly Perturbed Differential Equations: Convection-Diffusion-Reaction and Flow Problems*. Springer Series in Computational Mathematics, vol. 24. Berlin: Springer.
- ZHANG, Q. & SHU, C.-W. (2004) Error estimates to smooth solutions of Runge-Kutta discontinuous Galerkin methods for scalar conservation laws. *SIAM J. Numer. Anal.*, **42**, 641–666.

Appendix

A.1 Proof of Lemma 6.6

A.1.1 *Proof of estimate (6.13).* We estimate the terms of $b_h(\xi, \Pi_h \phi - \phi)$ over the interiors and boundaries of elements separately. Let Π_h^1 be the $L^2(\Omega)$ -projection onto the space of discontinuous piecewise linear polynomials on \mathcal{T}_h . Since on each $K \in \mathcal{T}_h$ it holds that $\nabla \tilde{\xi}|_K \in P^{p-1}(K)$, then $\Pi_h^1 a \cdot \nabla \tilde{\xi} \in S_h$. Hence, due to (6.1), we have

$$\sum_K \int_K \Pi_h^1 a \cdot \nabla \tilde{\xi} (\Pi_h \phi - \phi) \, dx = 0.$$

Due to standard approximation results, we have $\|a - \Pi_h^1 a\|_{L^\infty(K)} \leq Ch_K |a|_{W^{1,\infty}(K)}$, thus we can estimate the interior terms of $b_h(\xi, \Pi_h \phi - \phi)$ as

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} \int_K a \cdot \nabla \tilde{\xi} (\Pi_h \phi - \phi) \, dx &= \sum_{K \in \mathcal{T}_h} \int_K (a - \Pi_h^1 a) \cdot \nabla \tilde{\xi} (\Pi_h \phi - \phi) \, dx \\ &\leq \sum_{K \in \mathcal{T}_h} Ch_K |a|_{W^{1,\infty}} C_I h_K^{-1} \|\tilde{\xi}\|_{L^2(K)} \|\Pi_h \phi - \phi\|_{L^2(K)} \\ &\leq C \sum_{K \in \mathcal{T}_h} \max_{x \in K} e^{\mu(x,t)} \|\tilde{\xi}\|_{L^2(K)} CL_\mu e^{hL_\mu} h \max_{x \in K} e^{-\mu(x,t)} \|\tilde{\xi}\|_{L^2(K)} \\ &\leq CL_\mu e^{2hL_\mu} h \|\tilde{\xi}\|^2, \end{aligned} \tag{A.1}$$

due to the inverse inequality, Lemma 6.3 and the estimate $\max_{x \in K} e^{\mu(x,t)} \max_{x \in K} e^{-\mu(x,t)} \leq e^{L_\mu h_K}$. For the boundary terms of $b_h(\xi, \Pi_h \phi - \phi)$ we get

$$\begin{aligned} & - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [\tilde{\xi}] (\Pi_h \phi - \phi) \, dS - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \tilde{\xi} (\Pi_h \phi - \phi) \, dS \\ & \leq \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} |a \cdot \mathbf{n}| \max_{x \in K} (e^{\mu(x,t)}) |\tilde{\xi}| |\Pi_h \phi - \phi| \, dS + \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial \Omega} |a \cdot \mathbf{n}| \max_{x \in K} (e^{\mu(x,t)}) |\tilde{\xi}| |\Pi_h \phi - \phi| \, dS \\ & \leq \frac{1}{8} \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \partial \Omega} |a \cdot \mathbf{n}| \|\tilde{\xi}\|^2 \, dS + \frac{1}{8} \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \partial \Omega} |a \cdot \mathbf{n}| \|\tilde{\xi}\|^2 \, dS + 2 \sum_{K \in \mathcal{T}_h} \max_{x \in K} e^{2\mu(x,t)} \int_{\partial K} |a \cdot \mathbf{n}| |\Pi_h \phi - \phi|^2 \, dS \\ & \leq \frac{1}{8} \sum_{K \in \mathcal{T}_h} \left(\|\tilde{\xi}\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}\|_{a, \partial K^- \cap \partial \Omega}^2 \right) + CL_\mu^2 e^{2hL_\mu} \sum_K \max_{x \in K} e^{2\mu(x,t)} h_K \max_{x \in K} e^{-2\mu(x,t)} \|\tilde{\xi}\|_{L^2(K)}^2 \end{aligned}$$

by Young's inequality and Lemma 6.3. Again we estimate $\max_{x \in K} e^{2\mu(x,t)} \max_{x \in K} e^{-2\mu(x,t)} \leq e^{2L_\mu h_K}$ in the last inequality, which completes the proof after combining with (A.1). \square

8.1.2 *Proof of estimate (6.14).* We have by Green's theorem

$$b_h(\eta, \phi) = \sum_{K \in \mathcal{T}_h} \left(\int_{\partial K} a \cdot \mathbf{n} \eta \phi \, dS - \int_K (\operatorname{div} a) \eta \phi \, dx - \int_K a \cdot \nabla \phi \eta \, dx \right. \\ \left. - \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [\eta] \phi \, dS - \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \eta \phi \, dS \right). \quad (\text{A.2})$$

The first integral over K can be estimated as

$$- \sum_{K \in \mathcal{T}_h} \int_K (\operatorname{div} a) \eta \phi \, dx \leq Ch^{p+1} |u(t)|_{H^{p+1}} \max_{x \in \Omega} e^{-\mu(x,t)} \|\tilde{\xi}\| \leq Ch^{p+1} |u(t)|_{H^{p+1}} \|\tilde{\xi}\|,$$

because $\mu \geq 0$. Since $\phi = e^{-2\mu} \xi$, we get for the second integral over K in (A.2),

$$- \sum_{K \in \mathcal{T}_h} \int_K a \cdot \nabla \phi \, \eta \, dx = \sum_{K \in \mathcal{T}_h} \int_K 2a \cdot \nabla \mu e^{-2\mu} \xi \eta \, dx - \sum_{K \in \mathcal{T}_h} \int_K e^{-2\mu} a \cdot \nabla \xi \, \eta \, dx. \quad (\text{A.3})$$

The first right-hand side term in (A.3) can be estimated by

$$\sum_{K \in \mathcal{T}_h} \int_K 2a \cdot \nabla \mu e^{-2\mu} \xi \eta \, dx = \sum_{K \in \mathcal{T}_h} \int_K 2a \cdot \nabla \mu e^{-\mu} \tilde{\xi} \eta \, dx \leq \sum_{K \in \mathcal{T}_h} CL_\mu \max_{x \in K} e^{-\mu(x,t)} h^{p+1} |u(t)|_{H^{p+1}(K)} \|\tilde{\xi}\|_{L^2(K)} \\ \leq CL_\mu^2 h \|\tilde{\xi}\|^2 + Ch^{2p+1} |u(t)|_{H^{p+1}}^2.$$

The second right-hand side term in (A.3) can be estimated similarly to (A.1), by the definition of η :

$$- \sum_{K \in \mathcal{T}_h} \int_K e^{-2\mu} a \cdot \nabla \xi \, \eta \, dx = \sum_{K \in \mathcal{T}_h} \int_K \left(\Pi_h^1(e^{-2\mu} a) - e^{-2\mu} a \right) \cdot \nabla \xi \, \eta \, dx \\ \leq \sum_{K \in \mathcal{T}_h} Ch_K |e^{-2\mu} a|_{W^{1,\infty}} C_I h_K^{-1} \|\xi\|_{L^2(K)} h^{p+1} |u(t)|_{H^{p+1}(K)} \\ \leq C(1 + L_\mu) \max_{x \in \Omega} e^{-\mu(x,t)} h^{p+1} |u(t)|_{H^{p+1}} \|\tilde{\xi}\| \\ \leq C(1 + L_\mu)^2 h \|\tilde{\xi}\|^2 + Ch^{2p+1} |u(t)|_{H^{p+1}}^2,$$

since $|e^{-2\mu} a|_{W^{1,\infty}} = \|e^{-2\mu} \operatorname{div} a - 2\nabla \mu e^{-2\mu} a\|_{L^\infty} \leq C(1 + L_\mu)$.

As for the boundary terms in (A.2), we can split the integral over ∂K into integrals over the separate parts $\partial K^- \setminus \partial \Omega$, $\partial K^- \cap \partial \Omega$, $\partial K^+ \setminus \partial \Omega$ and $\partial K^+ \cap \partial \Omega$, similarly to the proof of Lemma 6.5. Thus,

several terms are canceled out:

$$\begin{aligned}
& \sum_{K \in \mathcal{T}_h} \left(\int_K (a \cdot \mathbf{n}) \eta \phi \, dS - \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [\eta] \phi \, dS - \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \eta \phi \, dS \right) \\
&= \sum_{K \in \mathcal{T}_h} \left(\int_{\partial K^+ \setminus \partial \Omega} (a \cdot \mathbf{n}) \eta \phi \, dS + \int_{\partial K^+ \cap \partial \Omega} (a \cdot \mathbf{n}) \eta \phi \, dS + \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) \eta^- \phi \, dS \right) \\
&= \sum_{K \in \mathcal{T}_h} \left(\int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) \eta^- [\phi] \, dS + \int_{\partial K^+ \cap \partial \Omega} (a \cdot \mathbf{n}) \eta \phi \, dS \right) \tag{A.4}
\end{aligned}$$

using a similar identity to (6.12). Finally, we can use Young's inequality to estimate (A.4) further as

$$\begin{aligned}
& \dots \leq \sum_{K \in \mathcal{T}_h} \left(\int_{\partial K^- \setminus \partial \Omega} |a \cdot \mathbf{n}| |\eta^-| e^{-\mu} |[\tilde{\xi}]| \, dS + \int_{\partial K^+ \cap \partial \Omega} |a \cdot \mathbf{n}| |\eta| e^{-\mu} |\tilde{\xi}| \, dS \right) \\
& \leq \frac{1}{8} \sum_{K \in \mathcal{T}_h} \left(\int_{\partial K^- \setminus \partial \Omega} |a \cdot \mathbf{n}| [\tilde{\xi}]^2 \, dS + \int_{\partial K^+ \cap \partial \Omega} |a \cdot \mathbf{n}| \tilde{\xi}^2 \, dS \right) + 2 \sum_K \int_K |a \cdot \mathbf{n}| \eta^2 e^{-2\mu} \, dS \\
& \leq \frac{1}{8} \sum_{K \in \mathcal{T}_h} \left(\|[\tilde{\xi}]\|_{a, \partial K^- \setminus \partial \Omega}^2 + \|\tilde{\xi}\|_{a, \partial K^+ \cap \partial \Omega}^2 \right) + Ch^{2p+1} |u(t)|_{H^{p+1}}^2.
\end{aligned}$$

The proof is completed by gathering all the above estimates of the individual terms of $b_h(\eta, \phi)$. \square

8.1.3 *Proof of estimate (6.15).* We use Lemmas 6.2 and 6.3 to estimate

$$\begin{aligned}
& b_h(\eta, \Pi_h \phi - \phi) \\
&= \sum_{K \in \mathcal{T}_h} \left(\int_K (a \cdot \nabla \eta) (\Pi_h \phi - \phi) \, dx - \int_{\partial K^- \setminus \partial \Omega} (a \cdot \mathbf{n}) [\eta] (\Pi_h \phi - \phi) \, dS - \int_{\partial K^- \cap \partial \Omega} (a \cdot \mathbf{n}) \eta (\Pi_h \phi - \phi) \, dS \right) \\
&\leq Ch^p |u(t)|_{H^{p+1}} CL_\mu e^{hL_\mu} h \max_{x \in \Omega} e^{-\mu(x,t)} \|\tilde{\xi}\| + Ch^{p+1/2} |u(t)|_{H^{p+1}} CL_\mu e^{hL_\mu} h^{1/2} \max_{x \in \Omega} e^{-\mu(x,t)} \|\tilde{\xi}\|.
\end{aligned}$$

The application of Young's inequality completes the proof. \square

8.2 Proof of Lemma 6.7

Lemma 6.3 gives us

$$\begin{aligned}
\sum_{K \in \mathcal{T}_h} \int_K c \xi (\Pi_h \phi - \phi) \, dx &\leq C \sum_{K \in \mathcal{T}_h} \max_{x \in K} e^{\mu(x,t)} \|\tilde{\xi}\|_{L^2(K)} L_\mu e^{hL_\mu} h_K \max_{x \in K} e^{-\mu(x,t)} \|\tilde{\xi}\|_{L^2(K)} \\
&\leq CL_\mu e^{2hL_\mu} h \|\tilde{\xi}\|^2.
\end{aligned}$$

As for the second estimate, we write $c_h(\eta, \Pi_h \phi) = c_h(\eta, \phi) + c_h(\eta, \Pi_h \phi - \phi)$ and estimate by Lemmas 6.2 and 6.3:

$$\begin{aligned} c_h(\eta, \phi) &= \int_{\Omega} c\eta\phi \, dx = \int_{\Omega} c\eta e^{-\mu} \tilde{\xi} \, dx \leq Ch^{p+1} |u(t)|_{H^{p+1}} \|\tilde{\xi}\|, \\ c_h(\eta, \Pi_h \phi - \phi) &= \int_{\Omega} c\eta(\Pi_h \phi - \phi) \, dx \leq Ch^{p+1} |u(t)|_{H^{p+1}} CL_{\mu} e^{hL_{\mu}} h \|\tilde{\xi}\|. \end{aligned}$$

Combining these two estimates with Young's inequality gives the desired result. \square

8.3 Proof of Lemma 6.8

We use Lemmas 6.2, 6.3 and the definition of ϕ :

$$\left(\frac{\partial \eta}{\partial t}, \Pi_h \phi \right) = \left(\frac{\partial \eta}{\partial t}, \phi \right) + \left(\frac{\partial \eta}{\partial t}, \Pi_h \phi - \phi \right) \leq Ch^{p+1} |u_t(t)|_{H^{p+1}} \|\tilde{\xi}\| + Ch^{p+1} |u_t(t)|_{H^{p+1}} CL_{\mu} e^{hL_{\mu}} h \|\tilde{\xi}\|.$$

Again, Young's inequality yields the final estimate. \square