SIAM J. NUMER. ANAL.
Vol. 44, No. 6, pp. 2245–2269

CFL CONDITION AND BOUNDARY CONDITIONS FOR DGM APPROXIMATION OF CONVECTION-DIFFUSION*

JEAN-BAPTISTE APOUNG KAMGA† AND BRUNO DESPRÉS‡

Abstract. We propose a general method for the design of discontinuous Galerkin methods (DGMs) for nonstationary linear equations. The method is based on a particular splitting of the bilinear forms that appear in the weak DGM. We prove that an appropriate time splitting gives a stable linear explicit scheme whatever the order of the polynomial approximation. Numerical results are presented.

Key words. discontinuous Galerkin method, advection diffusion, stability, CFL condition

AMS subject classifications. 65M12, 65M60

DOI. 10.1137/050633159

1. Introduction.

The convection-diffusion equation is widely used in real-life problems such as contaminant transport in porous media [1, 8, 26]. Due to the geological structure of the problem, the equation is convection-dominant in random distributed parts of the media. This makes its numerical resolution difficult. While difference schemes suffer from the complex geometry of the domain, ordinary finite element methods suffer from their lack of local conservativity [28], and finite volume methods suffer from their low order of accuracy (due to low order polynomial approximation). The discontinuous Galerkin method (DGM or DG), introduced in 1973 by Reed and Hill [32], in its development [24] found here a good field of application. In a computational aspect, the DGM can be used efficiently to handle the advection part in an operator splitting technique scheme [27]. But this strategy may break apart at boundary conditions of mixed type, where it is difficult to determine whether the boundary condition is more in the advection step or in the diffusion step. For real-life problems [8], the Dirichlet part of the boundary can also be split into inflow and outflow parts. This boundary condition can astutely be distributed in between the advection terms and diffusion terms [5, 6]. In a mathematical aspect, it is more convenient to have a unique bilinear form even if the splitting technique is used [20, 37, 33]. This leads to an ordinary differential equation, a different approach is [39]. Assuming for example that the DGM is used only in space to exploit the block diagonal mass matrix obtained, most time discretizations are explicit and therefore require a CFL condition.

In the one-dimensional case, using Von Neumann analysis, Chavent & Cockburn [11] proved that explicit linear Euler time integration of the DGM is unconditionally unstable if the ratio $\frac{\Delta t}{\Delta x}$ is held constant. To overcome this striking difficulty and still keep high order accuracy, Cockburn and Shu [19, 20, 21, 22] introduced the RKDG (Runge–Kutta discontinuous Galerkin method). It uses at each time step an

*Received by the editors June 6, 2005; accepted for publication (in revised form) April 25, 2006; published electronically November 24, 2006.
http://www.siam.org/journals/sinum/44-6/63315.html

†University Paris VI, Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, 175 rue du Chevaleret, 75013 Paris, France (apoung@ann.jussieu.fr). This author's research was supported by BQR of the University of Paris VI and GDR MOMAS under the guidance of O. Pironneau.

‡CEA 31/33 rue de la Fédération, 75752 Paris Cedex 15, France, and University Paris VI, Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, 175 rue du Chevaleret, 75013 Paris, France (despres@ann.jussieu.fr).

explicit Euler scheme, stabilized by a particular slope limiter, which makes the scheme nonlinear. Due to this nonlinearity, proof of convergence of the fully discrete explicit DGM is not possible except perhaps in very rare and special cases. We refer the reader to Cockburn [18] for a presentation of the convergence theory for the DGM. Despite this lack of theory, numerical experiments show the convergence. For example, in the one-dimensional case for advection, the convergence is observed if the CFL condition is of the form $\frac{1}{2k+1}$ for polynomials of order k [18]. To the best of our knowledge, the analysis of the fully discrete explicit DGM scheme remains an open problem.

In this work we propose a way to solve this problem. We propose an abstract functional formalism. Within this formalism, it is easy to design explicit (only local-in-the-cell) computations, which are linear and stable under CFL DGMs. Then we apply this method to our model problem, which is advection diffusion in two dimensions,

$$\partial_t c + \mathbf{u}.\nabla c - \nabla.(K\nabla c) = 0, \quad x \in \mathbf{R}^2,\ t > 0. \tag{1.1}$$

The diffusion coefficient is nonnegative $K \geq 0$, and the velocity is divergence-free $\nabla.\mathbf{u} = 0$. Boundary conditions are general and are specified in the core of the paper. Due to the stability (under the CFL condition) and linearity of our explicit DGM scheme, we are able to prove the convergence by a standard method. For example, we obtain the estimate of convergence in two dimensions for the advection case ($K = 0$),

$$\|c(n\Delta t) - c_h^n\|_{L^2} \leq C_1 \Delta t^2 + C_2 h^p + E.$$

E is an error term due the discretization of the initial condition and can be taken as small as desired. This estimate is true for the second order in time discretization. The order in space is p, which is the degree of the polynomial basis. Since the optimal order in space is $p + 1/2$, we think this loss of $1/2$ is an artifact of the analysis, which could be corrected with a more sophisticated technique [13, 14, 15]. To our knowledge, such an estimate is new and was not possible to get for previous fully discrete explicit DGM schemes.

At the theoretical level the key idea is to reformulate (1.1) as a weak problem

$$\left(\frac{\partial}{\partial t}U, V\right) + \mathcal{A}_0(U,V) + \mathcal{A}_1(U,V) - \mathcal{A}_2(U,V) = 0 \ \ \forall V \in \mathcal{V}, \tag{1.2}$$

where U is the solution, V is a test function, $(.,.)$ is the standard L^2 scalar product, and $\mathcal{A}_{0,1,2}$ are some bilinear forms defined later in this paper. The space is $\mathcal{V} \subset \sum_k L^2(\Omega_k)$, where (Ω_k) is a partition of the plane, i.e., is the mesh. Among other properties, the local bilinear forms $\mathcal{A}_0(U,V)$, $\mathcal{A}_1(U,V)$, and $\mathcal{A}_2(U,V)$ satisfy

$$\mathcal{A}_0(U,U) + \mathcal{A}_1(U,U) - \mathcal{A}_2(U,U) \geq 0. \tag{1.3}$$

The first order time discretization of (1.3) is as follows: Find $U_h^n, U_h^{n+1} \in \mathcal{V}_h$ such that for all test functions $V_h \in \mathcal{V}_h$,

$$\left(\frac{U_h^{n+1} - U_h^n}{\Delta t}, V_h\right) + \mathcal{A}_0(U_h^{n+1}, V_h) + \mathcal{A}_1(U_h^n, V_h) - \mathcal{A}_2(U_h^n, V_h) = 0. \tag{1.4}$$

When applied to (1.1), the bilinear form $\mathcal{A}_0$ is local-in-the-cell, and this is why the scheme is explicit. The main stability property that we prove is the inequality

$$||U_h^{n+1}||_{L^2(\mathbf{R}^2)} \leq ||U_h^n||_{L^2(\mathbf{R}^2)} \quad \forall n \ \in \mathbb{N}, \tag{1.5}$$

which is true under a CFL condition that is studied in detail. It guarantees stability whatever the order of the polynomial approximation. Since $\mathcal{A}_0$ is in practice a local-in-the-cell bilinear form, the scheme is explicit at the price of the resolution of a local-in-the-cell linear system. At the implementation level, it does not cost more than inverting the local mass matrix. We also study the second order discretization in time,

$$\begin{aligned}(1.6)\qquad \frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},V_h\right)+\frac{2}{3}\mathcal{A}_0(U_h^{n+1},V_h)\\ +\frac{2}{3}\mathcal{A}_1(2U_h^n-U_h^{n-1},V_h)-\frac{2}{3}\mathcal{A}_2(2U_h^n-U_h^{n-1},V_h)=0.\end{aligned}$$

The CFL condition is twice as stringent for (1.6) than for (1.4). It is possible to define all the parameters of the method in order to optimize the CFL condition. We will apply this method for our convection-diffusion problem.

The paper is organized as follows. In section 2 we consider a general setting. We present the properties which the bilinear forms should satisfy in this framework. Assuming these properties, we discuss some time schemes and derive the abstract CFL condition that guarantees their stability. In section 3 we address the convection-diffusion equation within the discontinuous Galerkin approximation and show how to cast the bilinear form to fit within the abstract formalism. We show how to introduce commonly used boundary conditions. In section 4 we analyze the abstract CFL condition in the case of a uniform grid and give values to all constants. We give the bilinear forms in particular cases of pure advection and pure diffusion. We conclude that the totally discrete schemes introduced for the convection-diffusion equation make up a continuous interpolation between the scheme for pure advection and the scheme for pure diffusion. In section 5 we analyze the convergence of the second order schemes in the case of the pure advection equation. Finally, in section 6 we present numerical results for advection and diffusion and compare them with other DGMs.

2. The abstract discontinuous Galerkin formalism. We first consider an abstract formalism in a more general setting and derive some time discretization, which will be stable under an abstract CFL condition.

2.1. Abstract formalism. Let us define the spaces

$$\mathcal{V}\subset\mathcal{H}. \tag{2.1}$$

$\mathcal{H}$ is endowed with a scalar product, namely $(.,.)$. In practice $\mathcal{H}=L^2(\Omega)$.

DEFINITION 2.1. *A sequence $(U^p)_p\in\mathcal{V}$ will be said to be L^2 stable if there exists a constant $C\in\mathbb{R}$ such that $(U^p,U^p)\le C$ for all $p\in\mathbb{N}$.*

Let $\mathcal{A}_i, i=0,1,2$, be three bilinear forms on $\mathcal{V}$ satisfying the following properties:

$$(2.2)\qquad \begin{cases}\mathcal{A}_1 \text{ is symmetric nonnegative.}\\ \text{There exist a bilinear form } \mathcal{A}_3 \text{ also defined on } \mathcal{V} \text{ such that}\\ \mathcal{A}_0(U,U)\ge\frac{1}{2}\left(-\mathcal{A}_1(U,U)+\mathcal{A}_3(U,U)\right) \text{ and } \mathcal{A}_2(U,V)\le\frac{1}{2}\left(\mathcal{A}_1(U,U)+\mathcal{A}_3(V,V)\right).\end{cases}$$

A consequence of (2.2) is

$$\mathcal{A}_0(U,U)+\mathcal{A}_1(U,U)-\mathcal{A}_2(U,U)\ge 0\quad\forall U\in\mathcal{V}.$$

We now consider the problem (2.3):

$$
\begin{cases}
\text{Given} \quad U_0 \in \mathcal{V}, \\
\text{find} \quad U \in C^1(0,T;\mathcal{V}) \quad \text{such that} \quad \forall V \in \mathcal{V}, \\
\left(\frac{\partial}{\partial t} U, V\right) + \mathcal{A}_0(U,V) + \mathcal{A}_1(U,V) - \mathcal{A}_2(U,V) = 0, \\
U = U_0 \quad \text{at} \quad t = 0.
\end{cases} \tag{2.3}
$$

In what follows we will assume that it has a unique solution.

LEMMA 2.2. *Assume that the bilinear forms $\mathcal{A}_i, i = 0,1,2$, satisfy* (2.2). *Then the solution to* (2.3) *is L^2 stable.*

Proof. Choosing $V = U$ and using the property of (2.2) one gets directly that $d_t\left[\frac{1}{2}(U,U)(t)\right] \le 0$. Therefore the energy $t \mapsto (U,U)(t)$ decreases. ∎

2.2. Time and space discretizations and abstract CFL conditions. Let $\mathcal{V}_h \subset \mathcal{V}$ be a finite-dimensional vectorial subspace of $\mathcal{V}$. The unknown at time step n is $U_h^n \in \mathcal{V}_h$. The test function is denoted by $V_h^n \in \mathcal{V}_h$. Under assumptions (2.2) on bilinear forms $\mathcal{A}_i, i = 0,1,2$, we can now derive some fully discrete schemes, which are stable under abstract CFL conditions.

2.2.1. First order scheme. The first order scheme reads

$$
\left(\frac{U_h^{n+1} - U_h^n}{\Delta t}, V_h\right) + \mathcal{A}_0(U_h^{n+1}, V_h) + \mathcal{A}_1(U_h^n, V_h) - \mathcal{A}_2(U_h^n, V_h) = 0 \quad \forall V_h. \tag{2.4}
$$

We have the following result.

THEOREM 2.3. *Assuming that the bilinear forms $\mathcal{A}_i, i = 0,1,2$, satisfy the properties* (2.2), *we assume that the time step satisfies the abstract CFL requirement*

$$
\Delta t \mathcal{A}_1(U_h, U_h) \le (U_h, U_h) \quad \forall U_h \in \mathcal{V}_h. \tag{2.5}
$$

Then scheme (2.4) *is L^2 stable and*

$$
\left(U_h^{n+1}, U_h^{n+1}\right) \le \left(U_h^n, U_h^n\right). \tag{2.6}
$$

$\Delta t > 0$ exists because the dimension of $\mathcal{V}_h$ is finite.

Proof. The proof explicitly uses the inequalities of (2.2). The scalar product of (2.4) with U_h^{n+1} gives

$$
\begin{aligned}
&\left(U_h^{n+1}, U_h^{n+1}\right) \\
&= \left(U_h^n, U_h^{n+1}\right) - \Delta t \mathcal{A}_0(U_h^{n+1}, U_h^{n+1}) - \Delta t \mathcal{A}_1(U_h^n, U_h^{n+1}) + \Delta t \mathcal{A}_2(U_h^n, U_h^{n+1}) \\
&\le \left(U_h^n, U_h^{n+1}\right) - \Delta t \mathcal{A}_0(U_h^{n+1}, U_h^{n+1}) - \Delta t \mathcal{A}_1(U_h^n, U_h^{n+1}) \\
&\quad + \frac{\Delta t}{2}\left(\mathcal{A}_1(U_h^n, U_h^n) + \mathcal{A}_3(U_h^{n+1}, U_h^{n+1})\right) \\
&\le \left(U_h^n, U_h^{n+1}\right) + \frac{\Delta t}{2}\left(\mathcal{A}_1(U_h^n, U_h^n) - 2\mathcal{A}_1(U_h^n, U_h^{n+1}) + \mathcal{A}_1(U_h^{n+1}, U_h^{n+1})\right).
\end{aligned}
$$

Using the symmetry of bilinear form $\mathcal{A}_1$ and the scalar product, we rewrite the previous inequality as

$$
\begin{aligned}
\left(U_h^{n+1}, U_h^{n+1}\right) &\le \left(U_h^{n+1}, U_h^{n+1}\right) \\
&- \left(\left(U_h^{n+1} - U_h^n, U_h^{n+1} - U_h^n\right) - \Delta t \mathcal{A}_1\left(U_h^{n+1} - U_h^n, U_h^{n+1} - U_h^n\right)\right).
\end{aligned}
$$

Assuming the abstract CFL-like condition (2.5), the result is proved. ∎

2.2.2. Second order scheme. Extending to second order time discretization the abstract DGM already mentioned is not easy. After numerous attempts, we focused on the following approach, which is based on the theory of A-stable time integration for stiff equations; see [25]. First, we begin with the retrograde second order time integration,

$$\frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},V_h\right)+\frac{2}{3}\left(\mathcal{A}_0+\mathcal{A}_1-\mathcal{A}_2\right)(U_h^{n+1},V_h)=0\quad\forall V_h. \tag{2.7}$$

Its stability can be proved, by taking $V_h=U_h^{n+1}$ in (2.7). The scheme is fully implicit in the sense that it requires the inversion of a global linear system to get the new value. Let us now define a semi-implicit second order time scheme. The idea is to get rid of the cell-to-cell coupling that appears in (2.7). For this we use the relation $U((n+1)\Delta t)=2U(n\Delta t)-U((n-1)\Delta t)+O(\Delta t^2)$, which is true provided that U is smooth. Then we eliminate some occurrences of U_h^{n+1} in (2.7) using transformation $U_h^{n+1}\leftarrow 2U_h^n-U_h^{n-1}$. It gives the scheme

$$\begin{aligned}\frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},V_h\right)&+\frac{2}{3}\mathcal{A}_0(U_h^{n+1},V_h)\\&+\frac{2}{3}\mathcal{A}_1(2U_h^n-U_h^{n-1},V_h)-\frac{2}{3}\mathcal{A}_2(2U_h^n-U_h^{n-1},V_h)=0\quad\forall V_h.\end{aligned} \tag{2.8}$$

We will see that in practice, $\mathcal{A}_0$ is of local-in-the-cell bilinear form. In this case, scheme (2.8) is only locally implicit, and we need only inverse local linear systems to get the new solution. Hence scheme (2.8) is in practice an explicit one.

THEOREM 2.4. *Assume the bilinear forms $\mathcal{A}_i, i=0,1,2$, satisfy the properties* (2.2), *and assume the time step satisfies the abstract CFL requirement*

$$2\Delta t\mathcal{A}_1(U_h,U_h)\le(U_h,U_h)\quad\forall U_h\in\mathcal{V}_h. \tag{2.9}$$

Then scheme (2.8) *is L^2 stable and*

$$\begin{aligned}&\left(U_h^{n+1},U_h^{n+1}\right)+\left(2U_h^{n+1}-U_h^n,2U_h^{n+1}-U_h^n\right)\\&\le(U_h^n,U_h^n)+\left(2U_h^n-U_h^{n-1},2U_h^n-U_h^{n-1}\right).\end{aligned} \tag{2.10}$$

Proof. Let us take $V_h=U_h^{n+1}$ in (2.8). We get

$$\begin{aligned}\frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},U_h^{n+1}\right)&+\frac{2}{3}\mathcal{A}_0(U_h^{n+1},U_h^{n+1})\\&+\frac{2}{3}\mathcal{A}_1(2U_h^n-U_h^{n-1},U_h^{n+1})-\frac{2}{3}\mathcal{A}_2(2U_h^n-U_h^{n-1},U_h^{n+1})=0.\end{aligned}$$

We can give a lower bound to $\mathcal{A}_0(U_h^{n+1},U_h^{n+1})$ and $-\mathcal{A}_2(2U_h^n-U_h^{n-1},U_h^{n+1})$ using (2.2). Therefore

$$\begin{aligned}&\frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},U_h^{n+1}\right)+\frac{1}{3}(\mathcal{A}_3-\mathcal{A}_1)(U_h^{n+1},U_h^{n+1})\\&+\frac{2}{3}\mathcal{A}_1(2U_h^n-U_h^{n-1},U_h^{n+1})-\frac{1}{3}\mathcal{A}_1(2U_h^n-U_h^{n-1},2U_h^n-U_h^{n-1})-\frac{1}{3}\mathcal{A}_3(U_h^{n+1},U_h^{n+1})\le 0,\end{aligned}$$

that is,

$$\frac{1}{3}\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},U_h^{n+1}\right)$$
$$-\frac{1}{3}\mathcal{A}_1(U_h^{n+1}-2U_h^n+U_h^{n-1},U_h^{n+1}-2U_h^n+U_h^{n-1})\le 0.$$

Let us define the energy

$$E(n+1)=\left(U_h^{n+1},U_h^{n+1}\right)+\left(2U_h^{n+1}-U_h^n,2U_h^{n+1}-U_h^n\right).$$

One has the equality

$$E(n+1)-E(n)=6\left(\frac{3U_h^{n+1}-4U_h^n+U_h^{n-1}}{\Delta t},U_h^{n+1}\right)$$
$$-\left(U_h^{n+1}-2U_h^n+U_h^{n-1},U_h^{n+1}-2U_h^n+U_h^{n-1}\right).$$

Plugging in the previous inequality, we obtain

$$E(n+1)\le E(n)-\left(U_h^{n+1}-2U_h^n+U_h^{n-1},U_h^{n+1}-2U_h^n+U_h^{n-1}\right)$$
$$+2\Delta t\mathcal{A}_1(U_h^{n+1}-2U_h^n+U_h^{n-1},U_h^{n+1}-2U_h^n+U_h^{n-1}).$$

Under the abstract CFL condition (2.9), the result is proved. □

2.2.3. Implicit scheme. The implicit scheme is

$$\left(\frac{U_h^{n+1}-U_h^n}{\Delta t},V_h\right)+\mathcal{A}_0(U_h^{n+1},V_h)+\mathcal{A}_1(U_h^{n+1},V_h)-\mathcal{A}_2(U_h^{n+1},V_h)=0. \tag{2.11}$$

LEMMA 2.5. *The implicit scheme* (2.11) *is L^2 stable unconditionally.*

Proof. The proof is left to the reader. □

2.3. Optimization of numerical parameters. It is well known that the DGM applied to convection-diffusion needs the definition of some arbitrary numerical parameters in order to completely define the bilinear forms at interfaces. We refer to [30, 12], where the dependence between the convergence of the DGM for stationary problems and the numerical parameters is analyzed. In what follows, we analyze the influence of the numerical parameters on the CFL condition (for nonstationary problems, of course). An open problem is to show that the parameter which is optimal with respect to the CFL condition is also optimal for convergence.

By inspection of the bilinear forms defined in the following section for convection-diffusion, it is enough to consider the abstract problem

$$\left(\frac{\partial}{\partial t}U,V\right)+\mathcal{A}_0(U,V)+\mathcal{A}_1^\alpha(U,V)-\mathcal{A}_2^\alpha(U,V)=0\ \ \forall V\in\mathcal{V}. \tag{2.12}$$

The bilinear forms $\mathcal{A}_0,\mathcal{A}_1^\alpha,\mathcal{A}_2^\alpha$ satisfy (2.2). The dependence to the arbitrary parameters is represented by α. The CFL condition takes the form

$$\left(\max_{U_h\in\mathcal{V}_h,\,U_h\neq 0}\frac{\mathcal{A}_1^\alpha(U_h,U_h)}{(U_h,U_h)}\right)\Delta t\le C, \tag{2.13}$$

where $C = 1$ for the first order scheme (2.5) and $C = \frac{1}{2}$ for the second order scheme (2.9). So the best α, denoted as α_{opt}, is the one that minimizes the constant in this inequality. We obtain the min-max problem for α_{opt},

$$\left(\max_{U_h \in \mathcal{V}_h,\, U_h \neq 0} \frac{\mathcal{A}_1^{\alpha_{\text{opt}}}(U_h, U_h)}{(U_h, U_h)} \right) \leq \left(\max_{U_h \in \mathcal{V}_h,\, U_h \neq 0} \frac{\mathcal{A}_1^{\alpha}(U_h, U_h)}{(U_h, U_h)} \right) \quad \forall \alpha.$$

We will apply this method in order to define optimized coefficients for DGM discretization for convection-diffusion in section 3.

3. Advection-diffusion with discontinuous coefficients and boundary conditions. In what follows, we describe the introduction of mixed-type boundary conditions in an advection-diffusion problem. We show that physically correct boundary conditions fit into the framework. So the stability of the scheme is guaranteed for all boundary conditions described below. Let us recall the model equation

$$\partial_t c + \mathbf{u}.\nabla c - \nabla.(K \nabla c) = 0, \quad x \in \Omega \subset \mathbf{R}^2,\ t > 0. \tag{3.1}$$

Ω is a bounded smooth open set of $\mathbf{R}^2$.

3.1. Abstract discontinuous Galerkin formalism of problem (3.1). We are now going to show how to cast the discontinuous Galerkin formulation of problem (3.1) so that the bilinear forms fit with properties (2.2).

3.1.1. Notation. We begin with some notation. Let (Ω_k) be a mesh of the plane. The cells Ω_k do not overlap. They cover the plane. The boundary of cell Ω_k is $\partial \Omega_k$. The intersection of the boundary of cell Ω_j and cell Ω_k is referred to as $\Sigma_{jk} = \Sigma_{kj}$. The outgoing normal from Ω_k is $\mathbf{n}_k$.

The velocity field $\mathbf{u}$ is not necessarily constant but is divergence-free. Therefore the degrees of freedom of $\mathbf{u}$ are naturally described in terms of its fluxes $(\mathbf{u}_{kj}, \mathbf{n}_k)$ across Σ_{jk}. The diffusion coefficient is assumed to be positive and lower bounded, but not necessarily constant. Let K_k denote the value of the diffusion coefficient in cell Ω_k. For simplicity, K_k is considered constant in the cell, but there is no real issue if it is not, except at the implementation level. We will describe the boundary conditions later on. If necessary we will assumed that the outgoing unit normal is split into two parts

$$\begin{cases} \text{if } (\mathbf{u}, \mathbf{n}_k) \geq 0, & \text{then } \mathbf{n}_k^+ = \mathbf{n}_k \quad \text{and } \mathbf{n}_k^- = 0, \\ \text{if } (\mathbf{u}, \mathbf{n}_k) < 0, & \text{then } \mathbf{n}_k^+ = 0 \quad\ \ \text{and } \mathbf{n}_k^- = \mathbf{n}_k. \end{cases} \tag{3.2}$$

Let us define the spaces

$$\mathcal{V} = \oplus_k H^2(\Omega_k) \subset \mathcal{H} = \oplus_k L^2(\Omega_k). \tag{3.3}$$

$\mathcal{H}$ is endowed with a scalar product $(U, V) = \sum_k \int_{\Omega_k} u_k(x) v_k(x) dx$.

3.1.2. Construction of the bilinear forms. Next we assume that c is smooth. Let us define $U = (u_k)$ with $u_k = c_{|\Omega_k}$. The test function is $V = (v_k)$. Let us define the local bilinear form

$$\mathcal{A}_0(U, V) = \sum_k \int_{\Omega_k} \left(-u_k(t, x) \mathbf{u}.\nabla v_k(x) + u_k \nabla.(K_k \nabla v_k) + 2K_k \nabla u_k.\nabla v_k \right) dx. \tag{3.4}$$

We also need to define $\mathcal{A}_1$ and $\mathcal{A}_2$. So let us compute

$$(\partial_t U, V) + \mathcal{A}_0(U,V) = \sum_k \int_{\Omega_k} \left(-\mathbf{u}.\nabla u_k + \nabla.(K_k \nabla u_k)\right) v_k$$
$$+ \sum_k \int_{\Omega_k} \left(-u_k(t,x)\mathbf{u}.\nabla v_k(x) + u_k \nabla.(K_k \nabla v_k) + 2K_k \nabla u_k.\nabla v_k\right) dx$$
$$= \sum_k \int_{\partial\Omega_k} \left(-u_k v_k (\mathbf{u}_{kj}, \mathbf{n}_k) + u_k K_k \frac{\partial}{\partial n_k} v_k + v_k K_k \frac{\partial}{\partial n_k} u_k .\right) d\sigma = \text{R.H.S.}$$

Next we need to transform the right-hand side (R.H.S.) in order to be able to define $\mathcal{A}_1$ and $\mathcal{A}_2$. For this task we define

$$\begin{cases} w_k^+ = K_k \frac{\partial}{\partial n_k} u_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k + \alpha_{jk} u_k, \\ w_k^- = -K_k \frac{\partial}{\partial n_k} u_k + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k + \alpha_{jk} u_k \end{cases}$$

and

$$\begin{cases} z_k^+ = K_k \frac{\partial}{\partial n_k} v_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k + \alpha_{jk} v_k, \\ z_k^- = -K_k \frac{\partial}{\partial n_k} v_k + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k + \alpha_{jk} v_k. \end{cases}$$

The value of the positive parameter $\alpha_{jk} = \alpha_{kj}$ will be specified later on. Then the R.H.S. is also

$$\text{R.H.S.} = \sum_k \int_{\partial\Omega_k} \left[u_k \left(K_k \frac{\partial}{\partial n_k} v_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k \right) + v_k \left(K_k \frac{\partial}{\partial n_k} u_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k \right) \right] d\sigma,$$
$$\text{R.H.S.} = \sum_k \int_{\partial\Omega_k} \frac{1}{2\alpha_{jk}} (w_k^+ z_k^+ - w_k^- z_k^-) d\sigma.$$

The nonnegative symmetric bilinear form is given by the $w^- z^-$ part of the integral. Therefore we define

$$\mathcal{A}_1(U,V) = \sum_k \int_{\partial\Omega_k} \frac{1}{2\alpha_{jk}} \left(-K_k \frac{\partial}{\partial n_k} u_k + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k + \alpha_{jk} u_k \right) \times \left(-K_k \frac{\partial}{\partial n_k} v_k + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k + \alpha_{jk} v_k \right) d\sigma \tag{3.5}$$

so that we now have the relation

$$(\partial_t U, V) + \mathcal{A}_0(U,V) + \mathcal{A}_1(U,V) - \sum_k \int_{\partial\Omega_k} \frac{1}{2\alpha_{jk}} w_k^+ z_k^+ d\sigma = 0. \tag{3.6}$$

It is the place into which boundary conditions must be plugged. Let us start with some notation. The boundary between two cells Ω_k and Ω_j is still referred to as Σ_{jk}. The exterior boundary of cell Ω_k is Γ_k,

$$\Gamma_k = \partial\Omega_k \cap \partial\Omega, \quad \partial\Omega_k = (\cup_j \Sigma_{jk}) \cup \Gamma_k. \tag{3.7}$$

To transform the residual in (3.6) we use the continuity equation

$$w_k^+ = w_j^- \text{ on } \Sigma_{jk} \tag{3.8}$$
$$\Longleftrightarrow K_k \frac{\partial}{\partial n_k} u_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k + \alpha_{jk} u_k = -K_k \frac{\partial}{\partial n_j} u_j + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_j) u_j + \alpha_{jk} u_j.$$

For mathematical convenience we consider that all boundary conditions may be rewritten as

$$w_k^+ = R_k^\alpha w_k^- \text{ on } \Gamma_k, \tag{3.9}$$

where $R_k^\alpha \in \mathbf{R}$ characterizes the boundary condition. This coefficient R_k^α is very similar to a reflexion coefficient in time-harmonic wave equations. It will be more obvious later on that physically correct boundary conditions are such that $|R_k^\alpha| \le 1$. α_{kk} stands for the value of the artificial parameter on Γ_k, and $(\mathbf{u}_{kj}, \mathbf{n}_k)$ stands for the value of the velocity flux on the boundary. We now define

$$\begin{aligned}\mathcal{A}_2(U,V) = \sum_{kj} \int_{\Sigma_{kj}} \frac{1}{2\alpha_{jk}} &\left(-K_j \frac{\partial}{\partial n_j} u_j + \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_j) u_j + \alpha_{jk} u_j\right) \\ &\times \left(K_k \frac{\partial}{\partial n_k} v_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k + \alpha_{jk} v_k\right) d\sigma \\ + \sum_k \int_{\Gamma_k} \frac{R_k^\alpha}{2\alpha_{kk}} &\left(-K_k \frac{\partial}{\partial n_k} u_k + \frac{1}{2}(\mathbf{u}_{kk}, \mathbf{n}_k) u_k + \alpha_{kk} u_k\right) \\ &\times \left(K_k \frac{\partial}{\partial n_k} v_k - \frac{1}{2}(\mathbf{u}_{kk}, \mathbf{n}_k) v_k + \alpha_{kk} v_k\right) d\sigma.\end{aligned} \tag{3.10}$$

The bilinear form $\mathcal{A}_3$ is

$$\begin{aligned}\mathcal{A}_3(U,V) = \sum_k \int_{\partial\Omega_k} \frac{1}{2\alpha_{jk}} &\left(K_k \frac{\partial}{\partial n_k} u_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) u_k + \alpha_{jk} u_k\right) \\ &\times \left(K_k \frac{\partial}{\partial n_k} v_k - \frac{1}{2}(\mathbf{u}_{kj}, \mathbf{n}_k) v_k + \alpha_{jk} v_k\right) d\sigma.\end{aligned} \tag{3.11}$$

Now that we have defined all the bilinear forms, let us show that they satisfy the required properties.

LEMMA 3.1. *Consider the bilinear forms* (3.4), (3.5), (3.10), (3.11). *Assume that* $|R_k^\alpha| \le 1$. *Then properties* (2.2) *are satisfied.*

Proof. One has

$$\begin{aligned}\mathcal{A}_0(U,U) &= \sum_k \int_{\Omega_k} \left(-u_k(t,x)\mathbf{u}.\nabla u_k(x) + u_k \nabla.(K_k \nabla u_k) + 2K_k \nabla u_k.\nabla u_k\right) dx \\ &\ge \sum_k \int_{\partial\Omega_k} \left(-\frac{1}{2}(\mathbf{u}, \mathbf{n}_k) u_k^2 + u_k K \frac{\partial}{\partial n_k} u_k\right) d\sigma = \frac{1}{2}\left(-\mathcal{A}_1(U,U) + \mathcal{A}_3(U,U)\right),\end{aligned}$$

which proves the first part of (2.2). Then using the Cauchy–Schwarz inequality and property $|R_k^\alpha| \le 1$, one gets $\mathcal{A}_2(U,V) \le \frac{1}{2}(\mathcal{A}_1(U,U) + \mathcal{A}_3(V,V))$, which is the second part of (2.2). $\mathcal{A}_1$ is obviously symmetric nonnegative. ∎

3.1.3. Boundary conditions. One major particularity of this formalism is the way boundary conditions are introduced. They are all defined by giving different values to parameter R_k^α. Equation (3.9) shows how to introduce homogeneous boundary conditions. The expressions of R_k^α for commonly used boundary conditions are given in Table 3.1. For the Robin-type boundary condition, we need to restrict the admissible boundary conditions to $\frac{1}{2}(\mathbf{u}, \mathbf{n}) + \sigma \ge 0$ so that $|R_k^\alpha| \le 1$.

LEMMA 3.2. *All* R_k^α *given in Table* 3.1 *satisfy the inequality* $|R_k^\alpha| \le 1$.

Proof. The proof is obtained by straightforward computation. ∎

TABLE 3.1
Values of R_k^α for commonly used boundary conditions in the convection-diffusion equation.

Outgoing	$K_k = 0$, $(\mathbf{u}, \mathbf{n}) > 0$	$R_k^\alpha = \frac{-(\mathbf{u},\mathbf{n})+\alpha}{(\mathbf{u},\mathbf{n})+\alpha}$
Ingoing Dirichlet	$K_k = 0$, $(\mathbf{u}, \mathbf{n}) < 0$	$R_k^\alpha = 0$
Dirichlet	$K_k > 0$, $(\mathbf{u}, \mathbf{n}) = 0$	$R_k^\alpha = -1$
Neumann	$K_k > 0$, $(\mathbf{u}, \mathbf{n}) = 0$	$R_k^\alpha = 1$
Mixed or Robin	$K_k \frac{\partial}{\partial n} c + \sigma c = 0$	$R_k^\alpha = \frac{\alpha - \frac{1}{2}(\mathbf{u},\mathbf{n}) - \sigma}{\alpha + \frac{1}{2}(\mathbf{u},\mathbf{n}) + \sigma}$

3.2. Fully discrete DGM. Now we need to choose the space $\mathcal{V}_h$. The standard choice for DGMs is $\mathcal{V}_h = \mathcal{V}_p \subset \mathcal{V}$ with

$$\mathcal{V}_p = \oplus_k P_p(\Omega_k), \tag{3.12}$$

where $P_p(\Omega_k)$ is the space of all polynomial functions of degree $p \in \mathbf{N}$ or less on cell Ω_k. Applying the time discretization defined in section 2.2, we obtain the fully discrete DGM. By construction, this DGM is L^2 stable for all p and without the need of any limiter. Therefore this method is different from the standard RKDG approach. The bilinear form $\mathcal{A}_0$ is local-to-one-cell so that both the first (2.4) and the second (2.7) order schemes are semi-implicit. In fact, one needs only inverse local linear systems to get the new solutions. Let us now analyze the abstract CFL condition in the case of uniform meshes.

4. CFL analysis. In this section we show that the abstract CFL condition (2.5) is equivalent to standard CFL requirements for the convection-diffusion equation, which is a kind of interpolation between pure convection and pure diffusion.

LEMMA 4.1. *Consider (for simplicity) a sequence of triangular and conformal meshes. Assume the sequence of meshes is uniformly regular. Denote by h a characteristic length of the mesh. Consider the first order scheme* (2.4) *with bilinear forms* (3.4), (3.5), (3.10).

For all $p \in \mathbf{N}$, there exists two constants $C_p^1 > 0, C_p^2 > 0$ such that if

$$\frac{3}{2}\Delta t \max_k \left(\frac{\alpha_{kj}}{C_p^1 h} + \frac{|\mathbf{u}|^2}{4\alpha_{kj} C_p^1 h} + \frac{K_k^2}{\alpha_{kj} C_p^2 h^3} \right) \le 1, \tag{4.1}$$

then the abstract CFL condition (2.5) *holds, and* (2.4) *is L^2 stable. Assuming that K is constant for simplicity, the optimal value of α corresponding to the least stringent CFL constraint is*

$$\alpha_{opt} = \sqrt{\frac{|\mathbf{u}|^2}{4} + \frac{K^2 C_p^1}{C_p^2 h^2}}. \tag{4.2}$$

Proof. First, the abstract CFL condition (2.5) is

$$\Delta t \max_k \left(\max_{\text{degree}(u_k) \le p} \frac{1}{2\alpha_{kj}} \frac{\int_{\partial\Omega_k} (\alpha_{kj} u_k + \frac{1}{2}(\mathbf{u}, \mathbf{n}_k) u_k - K \frac{\partial}{\partial n_k} u_k)^2}{\int_{\Omega_k} u_k^2} \right) \le 1.$$

This is true once the following inequality is satisfied:

$$\Delta t \max_k (T_1^k + T_2^k + T_3^k) \le 1,$$

where T_1^k, T_2^k, T_3^k are given by

$$T_1^k = 3 \max_{\text{degree}(u_k)\le p} \frac{1}{2\alpha_{kj}} \frac{\int_{\partial\Omega_k} (\alpha_{kj} u_k)^2}{\int_{\Omega_k} u_k^2},$$

$$T_2^k = 3 \max_{\text{degree}(u_k)\le p} \frac{1}{2\alpha_{kj}} \frac{\int_{\partial\Omega_k} (\frac{1}{2}(\mathbf{u}, \mathbf{n}_k) u_k)^2}{\int_{\Omega_k} u_k^2},$$

$$T_3^k = 3 \max_{\text{degree}(u_k)\le p} \frac{1}{2\alpha_{kj}} \frac{\int_{\partial\Omega_k} (K \frac{\partial}{\partial n_k} u_k)^2}{\int_{\Omega_k} u_k^2}.$$

Let us introduce the linear transformation F_k that maps the triangular cell Ω_k onto the reference cell $\hat{T}$. Using the regularity of the mesh,

$$T_1^k \le 3 \frac{\alpha_{kj}}{2hc_k} \left(\max_{\text{degree}(\hat{u}_k)\le p} \frac{\int_{\partial\hat{T}} \hat{u}_k^2}{\int_{\hat{T}} \hat{u}_k^2} \right),$$

where c_k depends on transformation F_k. Since the mesh is assumed to be uniformly regular, c_k is uniformly bounded from below. Let us define

$$c^p = \max_{\text{degree}(\hat{u}_k)\le p} \frac{\int_{\partial\hat{T}} \hat{u}_k^2}{\int_{\hat{T}} \hat{u}_k^2} \quad \text{and} \quad C_p^1 = \frac{\min_k c^k}{c^p}. \quad \text{Then} \quad T_1^k \le \frac{3}{2} \frac{\alpha_{kj}}{h} \frac{1}{C_p^1}.$$

Also, one has

$$T_2^k \le \frac{3}{2} \frac{|\mathbf{u}|^2}{4h} \frac{1}{\alpha_{kj} C_p^1}.$$

Using again the regularity of the mesh, we have

$$T_3^k \le \frac{3}{2} \frac{K_k^2}{\alpha_{kj}} \frac{d_k}{h^3} \left(\max_{\text{degree}(u_k)\le p} \frac{\int_{\partial\hat{T}} (\frac{\partial}{\partial \hat{n}_k} u_k)^2}{\int_{\hat{T}} \hat{u}_k^2} \right),$$

where d_k depends on F_k. Since the mesh is assumed to be uniformly regular, d_k is uniformly upper bounded. Let us define

$$e_p = \max_{\text{degree}(\hat{u}_k)\le p} \frac{\int_{\partial\hat{T}} \frac{\partial}{\partial \hat{n}} u_k^2}{\int_{\hat{T}} \hat{u}_k^2} \quad \text{and} \quad C_p^2 = \frac{1}{e_p \max\limits_k d_k}. \quad \text{Then} \quad T_3^k \le \frac{3}{2} \frac{K_k^2}{\alpha_{kj}} \frac{d_k}{h^3} \frac{1}{C_p^2}.$$

Putting this all together, we have

$$\Delta t \max_k (T_1^k + T_2^k + T_3^k) \le \frac{3}{2} \Delta t \max_k \left(\frac{\alpha_{kj}}{C_p^1 h} + \frac{|\mathbf{u}|^2}{4\alpha_{kj} C_p^1 h} + \frac{K_k^2}{\alpha_{kj} C_p^3 h^3} \right).$$

The abstract CFL condition is thus satisfied once we have

$$\frac{3}{2} \Delta t \max_k \left(\frac{\alpha_{kj}}{C_p^1 h} + \frac{|\mathbf{u}|^2}{4\alpha_{kj} C_p^1 h} + \frac{K_k^2}{\alpha_{kj} C_p^3 h^3} \right) \le 1.$$

Assuming K is constant, the optimal value of parameter α is the one that minimizes the multiplicative constant in front of Δt. Since the constant is $a\alpha + \frac{1}{\alpha b}$, where $a > 0$ and $b > 0$ are constants, then the optimal value is the solution of the equation $\frac{d}{d\alpha}\left(a\alpha + \frac{b}{\alpha}\right) = 0$, that is, $\alpha = \sqrt{\frac{b}{a}}$. Expanding with the definition of a and b, it gives (4.2). ∎

4.1. Particular cases. This section discusses the particular cases of the pure advection equation (i.e., $K \equiv 0, \mathbf{u}$ constant) and pure diffusion equation (i.e., $\mathbf{u} \equiv 0$ but $K > 0$).

4.1.1. Pure advection. In this particular case we have ($K \equiv 0, \mathbf{u}$ constant). Notation is still the same as in section 3.1.1. A consequence of Lemma 4.1 is the following.

LEMMA 4.2. *Consider a sequence of triangular and conformal meshes. Assume the sequence of meshes is uniformly regular. Denote by h a characteristic length of the mesh. For all $p \in \mathbf{N}$, there exists a $C_p^1 > 0$ such that if*

$$|\mathbf{u}|\Delta t \leq C_p^1 h, \tag{4.3}$$

then the abstract CFL condition is true.

4.1.2. Pure diffusion. In this particular case we have ($K > 0$ but $\mathbf{u} \equiv 0$). The equation is

$$\partial_t c - \nabla.(K\nabla c) = 0. \tag{4.4}$$

We consider $\alpha_{kj} \equiv \alpha > 0$ for simplicity of notation. Asfwe s in Lemma 4.1 we have the following.

LEMMA 4.3. *Consider a sequence of triangular and conformal meshes. Assume the sequence of meshes is uniformly regular. Denote by h a characteristic length of the mesh. For all $p \in \mathbf{N}$, there exists a $C_p^2 > 0$ such that if*

$$\Delta t \leq \frac{1}{\frac{\alpha}{C_p^1 h} + \frac{K^2}{\alpha C_p^2 h^3}}, \tag{4.5}$$

then the abstract CFL condition is true. Both constants C_p^1, C_p^2 depend only on the mesh and the degree of the polynomials, and not on the parameters of the equations or on α.

For an optimal value for parameter α, we also have the following.

LEMMA 4.4. *Consider the CFL inequality* (4.5), *with parameter α set to*

$$\alpha = \frac{K}{h}. \tag{4.6}$$

Then inequality (4.5) *is equivalent to the more standard CFL inequality*

$$K\Delta t \leq C_p^3 h^2, \quad \frac{1}{C_p^3} = \frac{1}{C_p^1} + \frac{1}{C_p^2}. \tag{4.7}$$

The proof is left to the reader.

The value (4.6) is optimal, since we recover the classical time step CFL constraint for explicit discretization of diffusion.

Remark. Formula (4.2) is a kind of continuous interpolation between (4.3) and (4.7). More importantly, if $K \equiv 0$, then $\alpha = \frac{|\mathbf{u}|}{2}$ and the scheme defined by (3.4), (3.5), (3.10) is *equal* to the standard DGM for the pure advection case. On the other hand, if $\mathbf{u} \equiv 0$, then the method is *equal* to the DGM defined above for the pure diffusion case. Therefore (4.2) ensures that the scheme for advection-diffusion is a continuous interpolation between the scheme for pure advection and the scheme for pure diffusion.

5. Convergence analysis for the advection case. Let us now state the convergence result. We restrict the analysis to the DGM for advection and leave convergence analysis of diffusion for future studies. Let us define an L^2 projection $\pi_h : \mathcal{H} \to \mathcal{V}_p$,

$$\pi_h(u) = (u_k) \Longleftrightarrow \int_{\Omega_k} u_k(x)v_k(x)dx = \int_{\Omega_k} u(x)v_k(x)dx \quad \forall v_k,\ \forall k. \tag{5.1}$$

The scheme that we analyze in this section is defined by

$$\begin{cases} U_h^0 = \pi_h(u_0), \text{ where } u_0 \text{ is the initial condition,} \\ U_h^1 \text{ is the solution of the first order time scheme (2.4),} \\ U_h^{n+1} \text{ is the solution of the second order time scheme (2.8),} \\ \text{the bilinear forms are } \mathcal{A}_0,\ \mathcal{A}_1,\ \mathcal{A}_2,\ \mathcal{A}_3, \\ \qquad\qquad \text{as defined in section 3.1.2 in the case } K \equiv 0. \end{cases} \tag{5.2}$$

We will use the following approximation property of the projection π_h.

Lemma 5.1. *Let E be an element (a triangle or a tetrahedron) in $\mathbb{R}^n(n = 2, 3)$ of diameter h_E. Then for any $u \in H^{k+1}(E)$,*

$$\|u - \pi_h u\|_{H^r(E)} \le C h_E^{k+1-r} \|u\|_{H^{k+1}(E)} \qquad r = 0,\ 1,$$

where C is independent of h_E. See [2].

Lemma 5.2 (trace inequality). *Let E be an element in $\mathbb{R}^n(n = 2, 3)$ of diameter h_E. Let e_k be an edge or a face of E. Then for any f in $H^s(E)$ and for $s \ge 2$,*

$$\|f\|_{L^2(e_k)} \le \hat{C} |e_k|^{\frac{1}{2}} |E|^{-\frac{1}{2}} (\|f\|_{L^2(E)} + h_E \|\nabla f\|_{L^2(E)}).$$

If f is a polynomial of degree $p > 0$ on E,

$$\|f\|_{L^2(e_k)} \le \hat{C} p^2 |e_k|^{\frac{1}{2}} |E|^{-\frac{1}{2}} (\|f\|_{L^2(E)}).$$

Here $\hat{C}$ is a constant independent of h_E and p. See [33].

Lemma 5.3. *Let $c \in \mathcal{V}$ be the solution of the advection equation and $U_h^n \in \mathcal{V}_p$ be the solution of* (5.2). *Then*

$$\theta_{l+1}^2 - \theta_l^2 \le 6\Delta t r^{l+1}, \tag{5.3}$$

where

$$\begin{aligned} \theta_l^2 &= (\xi^l, \xi^l) + (2\xi^l - \xi^{l-1}, 2\xi^l - \xi^{l-1}) \quad \forall\ l \ge 1, \\ \xi^l &= \pi_h u(l\Delta t) - U_h^l, \\ 6\chi^l &= \pi_h u(l\Delta t) - u(l\Delta t) \text{ and} \\ r^{l+1} &= \frac{1}{3}\left(\frac{3\chi^{l+1} - 4\chi^l + \chi^{l-1}}{\Delta t}, \xi^{l+1}\right) + \frac{2}{3}\mathcal{A}_0(\chi^{l+1}, \xi^{l+1}) \\ &\quad + \frac{2}{3}\mathcal{A}_1(2\chi^l - \chi^{l-1}, \xi^{l+1}) - \frac{2}{3}\mathcal{A}_2(2\chi^l - \chi^{l-1}, \xi^{l+1}) \\ &\quad + \frac{1}{3}\left(\frac{3u^{l+1} - 4u^l + u^{l-1}}{\Delta t} - 2\partial_t u((l+1)\Delta t), \xi^{l+1}\right) \\ &\qquad\qquad + \frac{2}{3}\mathcal{A}_1(2u^l - u^{l-1} - u^{l+1}, \xi^{l+1})r \\ &\qquad\qquad - \frac{2}{3}\mathcal{A}_2(2u^l - u^{l-1} - u^{l+1}, \xi^{l+1}). \end{aligned}$$

Proof. Taking $V_h = \xi^{l+1}$ in (2.8) with U_h^l replaced by $\pi_h u(l\Delta t)$, and subtracting the resulting equation in which $V_h = \xi^{l+1}$, from (2.8), gives

$$\frac{1}{3}\left(\frac{3\xi^{l+1} - 4\xi^l + \xi^{l-1}}{\Delta t}, \xi^{l+1}\right) + \frac{2}{3}\mathcal{A}_0(\xi^{l+1}, \xi^{l+1}) + \frac{2}{3}\mathcal{A}_1(2\xi^l - \xi^{l-1}, \xi^{l+1}) - \frac{2}{3}\mathcal{A}_2(2\xi^l - \xi^{l-1}, \xi^{l+1}) = r^{l+1}.$$

Using the lower bounds of $\mathcal{A}_0$ and $\mathcal{A}_2$ given by (2.2) and the symmetry of the bilinear form $\mathcal{A}_1$, we have

$$\frac{1}{3}\left(\frac{3\xi^{l+1} - 4\xi^l + \xi^{l-1}}{\Delta t}, \xi^{l+1}\right) + \frac{1}{3}\mathcal{A}_1(\xi^{l+1} - 2\xi^l + \xi^{l-1}, \xi^{l+1} - 2\xi^l + \xi^{l-1}) \leq r^{l+1}.$$

Now applying the abstract CFL condition (2.9), we further obtain

$$\frac{1}{3}\left(\frac{3\xi^{l+1} - 4\xi^l + \xi^{l-1}}{\Delta t}, \xi^{l+1}\right) - \frac{1}{6\Delta t}(\xi^{l+1} - 2\xi^l + \xi^{l-1}, \xi^{l+1} - 2\xi^l + \xi^{l-1}) \leq r^{l+1}$$

which, from the equality

$$(\theta_{l+1}^2 - \theta_l^2)/(6\Delta t) = \frac{1}{3}\left(\frac{3\xi^{l+1} - 4\xi^l + \xi^{l-1}}{\Delta t}, \xi^{l+1}\right) - \frac{1}{6\Delta t}(\xi^{l+1} - 2\xi^l + \xi^{l-1}, \xi^{l+1} - 2\xi^l + \xi^{l-1}),$$

reduces to $\theta_{l+1}^2 - \theta_l^2 \leq 6\Delta t\, r^{l+1}$. This ends the proof. ∎

LEMMA 5.4. *Notation is the same as in Lemma* 5.3. *Let us assume that the solution c is sufficiently smooth. Then there exist two constants, C_1 and C_2 not depending on l, Δt, and h such that*

$$|r^{l+1}| \leq (C_1(\Delta t)^2 + C_2 h^{\mu-1})\theta_{l+1}. \tag{5.4}$$

Here $\mu = \min(p+1, s)$ and s is the order of regularity of the solution in Sobolev's spaces.[1]

Proof. The velocity $\mathbf{u}$ is constant. In this proof we denote its module by $c_{vel} = |\mathbf{u}|$. The method consists of estimating all the terms in the right hand side in the definition of r^{l+1} in lemma 5.3. By the definition of the projection π_h, we have

$$\frac{1}{3}\left(\frac{3\chi^{l+1} - 4\chi^l + \chi^{l-1}}{\Delta t}, \xi^{l+1}\right) = 0.$$

Since $\mathbf{u}.\nabla\xi_k^{l+1} \in \mathcal{V}_p$, we have $\int_{\Omega_k} \chi_k^{l+1}\mathbf{u}.\nabla\xi_k^{l+1}dx = 0$. Therefore $\mathcal{A}_0(\chi^{l+1}, \xi^{l+1}) = 0$. Let us estimate $|\mathcal{A}_1(2\chi^l - \chi^{l-1}, \xi^{l+1})|$.

$$\begin{aligned}
|\mathcal{A}_1(2\chi^l - \chi^{l-1}, \xi^{l+1})| &\leq \sum_k \int_{e\in\partial\Omega_k} c_{vel}|(2\chi_k^l - \chi_k^{l-1})||\xi_k^{l+1}| \\
&\leq \sum_k c_{vel}h^{-1}\left(\|(2\chi_k^l - \chi_k^{l-1})\|_{L^2(\Omega_k)}\right. \\
&\qquad \left. + h\|\nabla(2\chi_k^l - \chi_k^{l-1})\|_{L^2(\Omega_k)}\right)\|\xi_k^{l+1}\|_{L^2(\Omega_k)} \\
&\leq \sum_k c_{vel}c_1 h^{\mu-1}\|\xi_k^{l+1}\|_{L^2(\Omega_k)} \\
&\leq Ch^{\mu-1}(\xi^{l+1}, \xi^{l+1})^{\frac{1}{2}}.
\end{aligned} \tag{5.5}$$

[1] A requirement of which is that $u \in C^1([0,T]; H^s(\Omega))$, $u_{tt} \in L^\infty([0,T]; L^\infty(\Omega))$, and $u_{ttt} \in L^\infty([0,T]; L^2(\Omega))$.

Similarly,

$$
\begin{aligned}
|\mathcal{A}_2(2\chi^l-\chi^{l-1},\xi^{l+1})| &\le \sum_{k,j}\int_{e\in\partial\Omega_k\cap\partial\Omega_j} c_{vel}|(2\chi_j^l-\chi_j^{l-1})||\xi_k^{l+1}| \\
&\le \sum_{k,j} c_{vel}h^{-1}\left(\|(2\chi_j^l-\chi_j^{l-1})\|_{L^2(\Omega_j)}\right. \\
&\qquad\left. + h\|\nabla(2\chi_j^l-\chi_j^{l-1})\|_{L^2(\Omega_j)}\right)\|\xi_k^{l+1}\|_{L^2(\Omega_k)} \\
&\le Ch^{\mu-1}(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}}.
\end{aligned}
\tag{5.6}
$$

The two other terms are

$$
\begin{aligned}
\left|\left(\frac{3u^{l+1}-4u^l+u^{l-1}}{\Delta t}-2(\partial_t u)^{l+1},\xi_k^{l+1}\right)\right| &\le (\Delta t)^2\sum_k\int_{\Omega_k} c_{vel}|\partial_{ttt}u(t^*,x)||\xi_k^{l+1}(x)| \\
&\le C(\Delta t)^2\|\partial_{ttt}u\|_{L^\infty(0,T;L^2(\Omega))}(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}} \\
&\le C(\Delta t)^2(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
|\mathcal{A}_1(2u^l-u^{l-1}-u^{l+1},\xi^{l+1})| &\le (\Delta t)^2\sum_k\int_{e\in\partial\Omega_k} c_{vel}|\partial_{tt}u(t^*,x)||\xi_k^{l+1}(x)| \\
&\le (\Delta t)^2\sum_k c_{vel}\|\partial_{tt}u(t^*)\|_{L^\infty(\Omega_k)}\int_{e\in\partial\Omega_k}|\xi_k^{l+1}(x)| \\
&\le (\Delta t)^2\sum_k c_{vel}\|\partial_{tt}u(t^*)\|_{L^\infty(\Omega_k)}h^{\frac{1}{2}}h^{-\frac{1}{2}}\|\xi_k^{l+1}\|_{L^2(\Omega_k)} \\
&\le C(\Delta t)^2\|\partial_{tt}u\|_{L^\infty(0,T;L^\infty(\Omega))}(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}} \\
&\le C(\Delta t)^2(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}}.
\end{aligned}
$$

Proceeding as above, we have

$$|\mathcal{A}_2(2u^l-u^{l-1}-u^{l+1},\xi^{l+1})| \le C(\Delta t)^2(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}}.$$

Now observing that $(\xi^{l+1},\xi^{l+1})^{\frac{1}{2}} \le \theta_{l+1}$, we obtain the result by summing all the above inequalities. □

THEOREM 5.5 (L^2 error estimate for pure advection). *Let $c\in\mathcal{V}$ be the solution of* (1.1) *in the advection case ($K\equiv 0$) with initial condition $c_0\in H^s(s\ge 2)$ and $U_h\in\mathcal{V}_p$ the solution of* (2.8)*, with the initial condition given by* (5.2)*. Assume the CFL condition* (2.9)*. Then there exist two constants C_1 and C_2 depending only on T and u such that*

$$\|(u-U_h)(T)\|_{L^2} \le 3\|\pi_h u(\Delta t)-U_h^1\|_{L^2}+C_1(\Delta t)^2+C_2h^{\mu-1},$$

where $\mu=\min(p+1,s)$.

Proof. Using the triangular inequality, we have

$$\|(u-U_h)(T)\|_{L^2} \le \|(u-\pi_h u)(T)\|_{L^2}+\|(\pi_h u-U_h)(T)\|_{L^2}.$$

The first term on the R.H.S. is bounded using the classical approximation theory [16] $\|(u-\pi_h u)(T)\|_{L^2}\le c(u)h^\mu$. Observe that by

$$\|(\pi_h u-U_h)(T)\|_{L^2}^2=(\xi^N,\xi^N),$$

it is possible to give an upper bound where N is defined by $T = N\Delta t$. Therefore according to Lemma 5.3, we have

$$\left(\theta_{n+1}^2 - \theta_n^2\right)/6\Delta t \le r^{n+1}.$$

From Lemma 5.4 there exist two constants C_1 and C_2 such that

$$\theta_{n+1}^2 - \theta_n^2 \le 6\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1})\theta_{n+1}.$$

We then have $\theta_{n+1}^2 - 6\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1})\theta_{n+1} \le \theta_n^2$, which can be rewritten as

$$\left(\theta_{n+1} - 3\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1})\right)^2 \le \theta_n^2 + \left(3\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1})\right)^2.$$

Therefore $\theta_{n+1} - \theta_n \le 6\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1})$. Summing this inequality over all n from 1 to $N-1$ produces

$$\theta_N \le \theta_0 + \sum_{n=1}^{n=N-1} 6\Delta t(C_1(\Delta t)^2 + C_2 h^{\mu-1}).$$

Since

$$\begin{aligned}\theta_0^2 &= (\xi^1, \xi^1) + (2\xi^1 - \xi^0, 2\xi^1 - \xi^0)\\ &\le ((\xi^1, \xi^1)^{\frac{1}{2}} + (2\xi^1 - \xi^0, 2\xi^1 - \xi^0)^{\frac{1}{2}})^2\\ &\le (3(\xi^1, \xi^1)^{\frac{1}{2}} + (\xi^0, \xi^0)^{\frac{1}{2}})^2,\end{aligned}$$

we have $\theta_0 \le 3(\xi^1, \xi^1)^{\frac{1}{2}} + (\xi^0, \xi^0)^{\frac{1}{2}}$. By definition of the scheme, initials values are such that

$$\xi^1 = \pi_h u(\Delta t) - U_h^1 \text{ and } \xi^0 = 0.$$

Also one has $N\Delta t = T$ so that $\sum_1^{N-1}(6\Delta t) \le 6T$. Therefore taking $C_i = C_i 6T, i = 1, 2$, ends the proof. □

Remark.

- The above theorem shows the convergence of the second order time discretization. Note that since it is second order in time, two initial conditions are needed: U_h^0, U_h^1. We have taken U_h^1 as the solution of a particular iteration of the first order scheme. So $\pi_h u(\Delta t) - U_h^1$ can be kept as small as we need.
- One can observe that in the demonstration above, except in Lemma 5.4, we have used only the property of the bilinear forms $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$. So by just giving an analogous lemma for pure diffusion and for mixed convection-diffusion equations, one obtains the convergence result for those equations. It is possible to guess that, in general, one has

  $$|r^{l+1}| \le (C_1(\Delta t)^\nu + C_2 h^\mu)(\xi^{l+1}, \xi^{l+1})^{\frac{1}{2}},$$

  where $\nu = 1, 2$ is the order of time discretization and μ is the order of the approximation error seen by the bilinear forms $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$. Note that μ can be kept optimal by replacing the L^2-projection with a well-chosen projection $\mathbf{R}_h$ related to the Gauss quadrature formula; see [17].

6. Numerical results. This section is devoted to the study of the order of convergence of our method by means of numerical tests and comparison with other methods. The algorithm presented in this work is denoted by the words "new formalism."

6.1. Pure advection. In this example, we consider (1.1) in the case when $K \equiv 0$. The computational domain is $(\Omega = (-0.5, 0.5)^2)$. The initial condition and the inflow boundary condition are taken from the exact solution, which is chosen here to be

$$c(t, x, y) = \exp\left(-\frac{(\hat{x} - x_c)^2 + (\hat{y} - y_c)^2}{2\sigma^2}\right).$$

The velocity field is $u = (-1, 1)^T$ and $\hat{x} = x + t, \hat{y} = y - t$. The parameters are $x_c = 0.25, y_c = -0.25, 2\sigma^2 = 0.004$. The time interval for the simulation is $(0, 0.5)$, which is the required time to shift the cone from its initial position to the symmetric position with respect to the center $(0, 0)$. The domain is subdivided into an initial mesh consisting of $8\times8\times2 = 138$ uniform regular triangles. We then successively refine the mesh and compute L^2 and L^∞ errors e_h on the mesh of size h and the numerical convergence rates by the ratio $\ln(e_h/e_{h/2})/\ln(2)$. The use of uniform meshes leads to the following values for the parameters in the CFL analysis. In formula (4.3) the value of C_p^1 is

$$C_p^1 = \begin{cases} \frac{1}{4+4\sqrt{2}} & \text{for } p = 1, \\ \frac{1}{6+6\sqrt{2}} & \text{for } p = 2. \end{cases}$$

For a second order in time discretization the value of C_p^1 is divided by 2. In our computations we divide it by 10, just to stay away from the optimal value. Table 6.1 shows the behavior of our formalism with respect to the order of the polynomial basis and time discretization. In Table 6.2 we compare the new formalism with RKDG (without flux limiting), RKDG (with the Cockburn–Shu flux limiting) that we call TVBMRKDG (total variation bounded modified slope limiter; see [23]), and with a Crank–Nicholson scheme applied to the stabilized DGM formulation of convection-equation introduced by Brezzi, Marini, and Süli [10]. The last one is introduced to compare our results to schemes in which the global matrix is inverted at every time step. We have done an element renumbering in that Crank–Nicholson scheme in order to have a thin band global matrix. We factor the global matrix before entering into loops, which leads to a gain in time compared to a sparse direct resolution of the global algebraic equation at every time step. The time required to do this operation is denoted by R in Table 6.2.

Observations. From Table 6.1, the error at the time T is of the form $C_1(\Delta t)^\alpha + C_2 h^\beta$, where α is the order of the time discretization and β is a real whose optimal value is $\beta = p + 1$ (where p is the degree of the polynomials). Even if constants C_1, C_2 influence the computed convergence rate, one can still observe that when using polynomials of order p with second order time discretization, the L^2 error is at least of order p in space. By comparison with other theoretical results [10] it is possible to conjecture a behavior of the form $O(\Delta t^2) + O(h^{p+\frac{1}{2}})$. But for this test problem the error in time is clearly dominant over the error in space. Therefore it is difficult to clearly identify the asymptotic order of convergence when using the second order in time discretization. At a more general level, it shows the interest of the second order in time discretization. This is seen in Table 6.2, where we observe the same convergence rate with RKDG without flux limiting, which is of order 2 for polynomials of order 1. The same convergence rate is observed for the Crank–Nicholson scheme applied to the formulation of [10]. These three second order formulations produce the same convergence rate for first order polynomials.

Table 6.1

Numerical L^2 errors, L^∞ errors, and convergence rate at time $t = 0.5s$, for first and second order in time with first and second order basis polynomials, in the new formalism ((2.4), (2.8)) scheme applied to the pure advection equation.

	First order in time				Second order in time			
h	L^2 error	Rate	L^∞ error	Rate	L^2 error	Rate	L^∞ error	Rate
P_1 basis polynomials								
1/8	$5.47E-02$	–	$7.28E-01$	–	$5.15E-02$	–	$6.42E-01$	–
1/16	$4.08E-02$	0.49	$6.15E-01$	0.31	$3.43E-02$	0.59	$5.07E-01$	0.34
1/32	$2.11E-02$	1.02	$3.54E-01$	0.94	$1.31E-02$	1.39	$2.16E-01$	1.23
1/64	$9.72E-03$	1.16	$1.63E-01$	1.17	$3.08E-03$	2.09	$5.65E-02$	1.93
1/128	$4.78E-03$	1.02	$7.55E-02$	1.11	$5.87E-04$	2.40	$1.13E-02$	2.32
P_2 basis polynomials								
1/8	$4.23E-02$	–	$6.39E-01$	–	$3.14E-02$	–	$4.83E-01$	–
1/16	$2.05E-02$	1.05	$3.21E-01$	0.99	$6.99E-03$	2.17	$1.10E-01$	1.80
1/32	$1.09E-02$	0.91	$1.59E-01$	1.01	$5.44E-04$	3.68	$1.17E-02$	3.23
1/64	$5.90E-03$	0.89	$8.49E-02$	0.91	$4.37E-05$	3.64	$1.87E-03$	2.65
1/128	$3.10E-03$	0.93	$4.49E-02$	0.92	$6.66E-06$	2.73	$2.53E-04$	2.88

Table 6.2

Comparison of numerical errors and convergence rates at time $t = 0.5s$, for second order in time with first order basis polynomials. R is the time spent renumbering the elements and factoring the global matrix. Computational times are for the finest mesh, using a Pentium III/1.266 GHZ.

	New formalism		RKDG		TVBMRKDG		Crank–Nicholson	
h	Error	Rate	Error	Rate	Error	Rate	Error	Rate
L^2 errors								
1/8	$5.15E-02$	–	$5.18E-02$	–	$5.23E-02$	–	$5.15E-02$	–
1/16	$3.43E-02$	0.59	$3.44E-02$	0.59	$3.83E-02$	0.45	$3.43E-02$	0.59
1/32	$1.31E-02$	1.39	$1.31E-02$	1.39	$2.96E-02$	0.37	$1.31E-02$	1.39
1/64	$3.08E-03$	2.09	$3.08E-03$	2.09	$1.39E-02$	1.09	$3.07E-03$	2.09
L^∞ errors								
1/8	$6.42E-01$	–	$6.48E-01$	–	$6.57E-01$	–	$6.43E-01$	–
1/16	$5.07E-01$	0.34	$5.08E-01$	0.35	$5.58E-01$	0.24	$5.05E-01$	0.34
1/32	$2.16E-01$	1.23	$2.16E-01$	1.23	$4.63E-01$	0.27	$2.15E-01$	1.23
1/64	$5.65E-02$	1.93	$5.62E-02$	1.96	$2.79E-01$	0.73	$5.62E-02$	1.93
CPU time								
1/64	81.38		90.34		32400		$553.93+R$	

6.2. Pure diffusion. In this example we consider the Dirichlet equation (1.1) with $(K \equiv 1, u \equiv 0)$. The computational domain is $\Omega = (0,1)^2$. The boundary condition is homogeneous so that the exact solution is

$$c(t,x,y) = \sin(\pi x)\sin(\pi y)\exp(-2\pi^2 t).$$

The initial condition is taken from this exact solution. The time interval is $(0, 1.510^{-2})$. This is the required time to reduce the maximum of the exact solution by about 25%. The domain is meshed into 16 uniform regular triangles. We successively refine this mesh uniformly. For each mesh of size h we compute the L^2 and L^∞ errors e_h and the numerical convergence rates given by the ratio $\ln(e_h/e_{h/2})/\ln(2)$. The use of uniform meshes leads to the following values of C_p^2 in formula (4.5): $C_p^2 = \frac{1}{12+6\sqrt{2}}$ for $p=1$ and $C_p^2 = \frac{1}{120+66\sqrt{2}}$ for $p=2$.

In order to enforce a better interelement continuity for small p, one can choose the parameter α to be of the form $\alpha = \beta\frac{K}{h}$, where $\beta \geq 1$ is a user-defined constant.

The optimal value of β is $\beta = \sqrt{\frac{C_p^1}{C_p^2}}$. Therefore our optimal value for C_p^3 in formula (4.7) is in this case $C_p^3 = \sqrt{C_p^1 C_p^2}$. In Table 6.4 we compare the new formalism for first order in time and second order polynomials with computed solutions obtained by Nonsymmetric Interior Penalty Galerkin (NIPG) and Symmetric Interior Penalty Galerkin (SIPG) GDMs [7, 35, 36]. For this first order in time, we have used an implicit scheme to discretize the SIPG and NIPG methods. We intended to do the same comparison for the second order in time. We tried a θ-scheme (see [34]) to discretize time in both SIPG and NIPG (note that implicit scheme corresponds to a θ-scheme with $\theta = 1$, as in [29], while the Crank–Nicholson scheme corresponds to $\theta = 0$ as described in [34]). But we noticed that using the same time step for the new formalism and for SIPG and NIPG Galerkin methods with the Crank–Nicholson scheme leads to instabilities in SIPG and NIPG. So for that time step, θ must stay in the interval $]0, 1]$, and therefore the θ-scheme is no longer of second order. This is a significant advantage of our formalism over the two others. We have taken the stabilization parameter $\sigma = 1$ for NIPG and $\sigma = 10$ for SIPG; see [35, 36]. The time step has also been multiplied by 10 in SIPG and NIPG, which are implicit methods ($\theta = 1$).

Observations. Here, as in the pure advection case, the error is of the form $C_1(\Delta t)^\alpha + C_2 h^\beta$. Since we have used the optimal CFL condition while refining the mesh, $\Delta t \approx Ch^2$, the convergence rate obtained numerically should be close to

$$\gamma = \min(2\alpha, \beta).$$

Let us discuss the values of α, β, and γ observed in Tables 6.3 and 6.5. For first order time discretization, $\alpha = 1$. Therefore $\gamma = \min(2, \beta)$. It shows that for first order or second order polynomials in conjunction with first order time discretization, we obtain a convergence rate of order 2. This is what we get in Table 6.3. Second order time discretization with first order polynomials gives also a convergence rate of $\gamma = 2$. Hence $\beta = 2$ for first order polynomials, and the convergence in space is optimal in this case. It is also seen in Table 6.3 that when we use second order polynomials with second order time discretization, the convergence rate starts from almost 3 and tends asymptotically to $\gamma = 2$. This shows that $\alpha = 2$ and $\beta = 2$ for second order time discretization with second order polynomials. Hence the convergence in space is suboptimal in this case. However, this is only a matter of worst behavior for even order polynomials. To view that, let us try third order polynomials with a sufficiently small CFL condition so that the error in time is absolutely negligible and we get an accurate value for β. Table 6.5 shows a convergence rate of order $\gamma = \min(2\times 2, 4) = 4$. By inspection of all these results we deduce that the new formalism presented in this paper keeps (on pure diffusion) the optimal space convergence rate for polynomials of odd order. This behavior is similar to other nonsymmetric discretizations like NIPG.

In order to analyze the advantage of our method over NIPG and SIPG for this kind of test problem, let us analyze the ratio accuracy/CPU time of the computation (see the last line of Table 6.4). We see that the error is slightly smaller for our method. But more important is the CPU time required to perform the computation. Due to well-known stability issues, NIPG and SIPG are implicit, which means a certain CPU time is needed to factorize and invert the matrix. This CPU time is denoted as R in the table. It is well known that R can be quite large. In our computations, R is about the same order as the CPU time needed to perform the whole computation. But here the matrix is factorized only once because the coefficients of the problem are constant

Table 6.3

Numerical L^2 errors, L^∞ errors, and convergence rates for first and second order in time with first and second order basis polynomials in the new formalism ((2.4), (2.8)) scheme applied to the pure diffusion equation.

	First order in time				Second order in time			
h	L^2 error	Rate	L^∞ error	Rate	L^2 error	Rate	L^∞ error	Rate
P_1 basis polynomials								
1/8	$1.00E-02$	–	$3.10E-02$	–	$8.87E-03$	–	$3.17E-02$	–
1/16	$2.50E-03$	2.00	$7.47E-03$	2.05	$2.16E-03$	2.04	$7.50E-03$	2.08
1/32	$6.20E-04$	2.01	$1.83E-03$	2.03	$5.37E-04$	2.00	$1.84E-03$	2.02
1/64	$1.55E-04$	2.00	$4.56E-04$	2.00	$1.34E-04$	2.00	$4.57E-04$	2.00
1/128	$3.87E-05$	2.00	$1.14E-04$	2.00	$3.35E-05$	2.00	$1.14E-04$	2.00
P_2 basis polynomials								
1/8	$8.95E-04$	–	$2.63E-03$	–	$7.53E-04$	–	$3.02E-03$	–
1/16	$2.10E-04$	2.09	$4.63E-04$	2.50	$1.75E-04$	2.11	$4.92E-04$	2.61
1/32	$5.16E-05$	2.02	$1.15E-04$	2.00	$4.27E-05$	2.03	$9.64E-05$	2.35
1/64	$1.28E-05$	2.01	$2.86E-05$	2.00	$1.06E-05$	2.01	$2.20E-05$	2.13
1/128	$3.20E-06$	2.00	$7.14E-06$	2.00	$2.65E-06$	2.00	$5.34E-06$	2.04

Table 6.4

Numerical comparison of L^2 errors, L^∞ errors, CPU time, and convergence rate, for first order in time with second order basis polynomials in the new formalism, and implicit scheme for SIPG and NIPG DGM. R is the time spent renumbering the elements and factoring the global matrix. Computational times were evaluated on a Pentium III/1.266 *GH processor.*

	New formalism			NIPG			SIPG		
h	Error	Rate	CPU	Error	Rate	CPU	Error	Rate	CPU
L^2 error									
1/8	$8.95E-04$	–	0.94	$1.94E-02$	–	$0.87+R$	$1.89E-02$	–	$0.86+R$
1/16	$2.10E-04$	2.09	9.29	$4.62E-03$	2.07	$5.91+R$	$4.47E-03$	2.08	$6.18+R$
1/32	$5.16E-05$	2.02	119.8	$1.14E-03$	2.02	$71.25+R$	$1.10E-03$	2.02	$71.21+R$
1/64	$1.28E-05$	2.02	1855	$2.84E-04$	2.00	$1519+R$	$2.75E-04$	2.00	$1334+R$
L^∞ error									
1/8	$2.63E-03$	–	0.94	$3.84E-02$	–	$0.87+R$	$3.75E-02$	–	$0.86+R$
1/16	$4.63E-04$	2.50	9.29	$9.22E-03$	2.06	$5.91+R$	$8.93E-03$	2.07	$6.18+R$
1/32	$1.15E-04$	2.00	119.8	$2.28E-03$	2.02	$71.25+R$	$2.21E-03$	2.01	$71.21+R$
1/64	$2.86E-05$	2.00	1855	$5.69E-04$	2.00	$1519+R$	$5.50E-04$	2.00	$1334+R$

Table 6.5

Numerical L^2 errors, L^∞ errors, and convergence rates for second order time discretization with third order basis polynomials in the new formalism scheme (2.8) *applied to pure diffusion equation. Computations are done with a very small CFL condition so as to reduce the time discretization error.*

Second order time scheme with P_3 basis polynomials				
h	L^2 error	Rate	L^∞ error	Rate
1/2	$3.885E-03$	–	$4.517E-02$	–
1/4	$3.663E-04$	3.41	$4.212E-03$	3.42
1/8	$2.636E-05$	3.80	$2.668E-04$	3.98
1/16	$1.757E-06$	3.91	$1.769E-05$	3.91
1/32	$1.139E-07$	3.95	$1.127E-06$	3.97
1/64	$7.246E-09$	3.97	$7.191E-08$	3.97

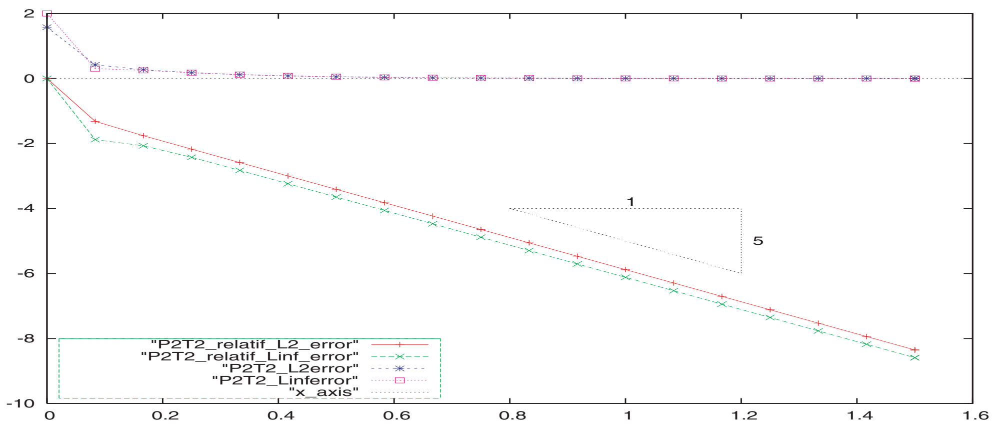

FIG. 6.1. *L^2 and L^∞ convergence errors at different times steps for the pure diffusion equation with nonhomogeneous boundary conditions. The computation is done using the new formalism with polynomials of order 2 in space and second order time discretization. The notation $P2T2$ stands for polynomials of order 2 in space ($P2$) with second order ($T2$) time discretization.*

in time. So if ever one desires to apply NIPG and SIPG to problems with variable coefficients, then R is to be multiplied by the number of iterations. Note that in our calculations, we have adapted the time step for NIPG and SIPG so that the number of time steps is already 10 times smaller for NIPG and SIPG. An even much greater time step is possible for NIPG and SIPG but at the price of a loss of accuracy of the discretization in time. In this case the new method, which is explicit, is much better than NIPG and SIPG.

6.3. An example with a nonhomogeneous Dirichlet boundary condition. Here is an example with a nonhomogeneous boundary Dirichlet condition. Instead of simply writing $\omega_k^+ = R_k^\alpha \omega_k^-$ (see Table 3.1), one uses

$$\omega_k^+ = R_k^\alpha \omega_k^- + \alpha_k (1 - R_k^\alpha) c_d \quad \text{for Dirichlet boundary condition } c = c_d,$$
$$\omega_k^+ = R_k^\alpha \omega_k^- + (1 + R_k^\alpha) g_N \quad \text{for Neumann boundary condition } K \frac{\partial}{\partial n} c = g_N.$$

We now take the same test case as above ($K \equiv 1, u \equiv 0$), with R.H.S. $f(t,x,y) = -4$, and a nonhomogeneous Dirichlet boundary condition $g_D(x,y) = x^2 + y^2$. We know that the limit of the exact solution as time tends to infinity is the solution of the stationary problem. That limit solution is in fact the function we have chosen as the Dirichlet boundary condition. In order to show that the new formalism handles nonhomogeneous boundary conditions, we have computed the solution with the initial condition taken to be $c(t = 0, x, y) = 0$ which is not related to the exact solution. The computational domain is $\Omega = (-1,1)^2$, meshed with nonuniform triangles (with 21 vertices per side) to show that the behavior of the formalism is well suited to the nonuniform mesh. Different steps of the solution are shown in Figure 6.2. Figure 6.1 shows the convergence to the exact solution as L^2 and L^∞ errors (measured by $||u(\infty) - u(t_n)||$) and relative L^2 and L^∞ errors (measured by $\log(||u(\infty) - u(t_n)||/||u(\infty)||)$) at every time step. Here $u(\infty)$ denotes the limit solution.

Observations. In Figure 6.2 the initial solution is zero, and as the time passes the convergence to the exact solution is achieved. It shows that boundary conditions of Dirichlet type are correctly discretized by this method.

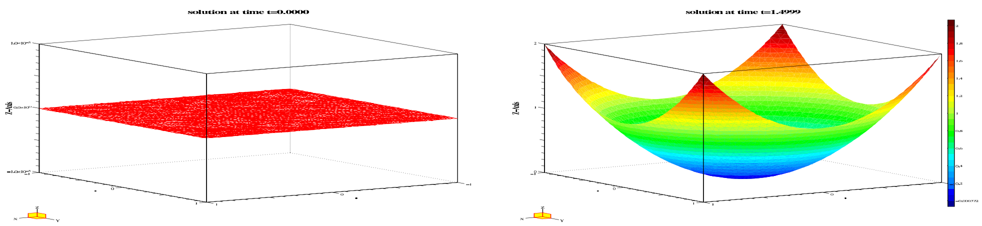

Fig. 6.2. *Asymptotic solution of the pure diffusion equation with nonhomogeneous boundary conditions, on a nonuniform mesh. On the left is the initial solution; on the right is the solution at $t = 1.5s$. The computations are done using the new formalism with second order polynomials in space and second order time discretization.*

6.4. A convection-diffusion example. In this section we consider the rotating pulse problem. The spatial domain is $\Omega = (-0.5, 0.5) \times (-0.5, 0.5)$, and the rotating field is imposed as $\mathbf{u} = (-4y, 4x)$. The initial condition and Dirichlet boundary condition are taken from the exact solution

$$c(t, x, y) = \frac{2\sigma^2}{2\sigma^2 + 4Kt} \exp(-\frac{(\bar{x} - x_c)^2 + (\bar{y} - y_c)^2}{2\sigma^2 + 4Kt},$$

where $\bar{x} = x\cos(4t) + y\sin(4t)$ and $\bar{y} = -x\sin(4t) + y\cos(4t)$. Here K is the constant diffusion coefficient. The R.H.S. is $f = 0$. This example was considered in [38], where only maxima and minima of many methods were listed. It is also used as a model equation in [4] to compare the L^2 error of a higher order DGM with various other methods on uniform rectangular meshes. Here we consider the same model problem on uniform triangular meshes, and we evaluate the L^2 and L^∞ errors and the convergence rate for the first and second order schemes presented in this paper. We take the same parameters as in [38, 4]: $K = 10^{-4}, x_c = 0.25, y_c = 0$, and $2\sigma^2 = 0.004$. The time interval for the simulation is $[0, T] = [0, \pi/4]$, which is the time for a half rotation. We begin with a uniform mesh of the domain made up of $8 \times 8 \times 2 = 138$ uniform triangles. We then successively refine the mesh and compute the L^2 and L^∞ errors e_h on the mesh of size h and the numerical convergence rates by the ratio $\ln(e_h/e_{h/2})/\ln(2)$. The time step is chosen so that the ratio $\Delta t/h$ is kept constant. The constant value is $1/82$ for first order time discretization and $1/164$ for the second order time scheme. The results obtained are recorded in Table 6.6.

Observations. This numerical test [38] is advection dominant in most parts of the domain and is diffusion dominant in the center of the domain. We solve it with the formalism presented in this paper with a constant ratio $\Delta t/h$. This constant ratio is obtained when we use the optimal parameter α (4.2) to determine the CFL condition (4.1). The second order in time scheme gives good results with higher order polynomials. Table 6.6 shows that using the constant ratio $\Delta t/h$, the convergence rate is greater than 2. Hence in second order time discretization, the time discretization error is small compared to the space discretization error for this test problem. This is a good feature when dealing with a coarse mesh. The second order time discretization is well suited for this kind of problem, where fine meshes are prohibitive due to memory management.

6.5. Conclusion driven from numerical experiments. The theoretical analysis is confirmed by numerical experiments. In particular we have L^2 stability and correct treatment of boundary conditions whatever the order of the polynomials is.

Table 6.6

Numerical L^2 errors, L^∞ errors, and convergence rates for first and second order time discretization schemes (2.4), (2.8) *applied to constant diffusion but variable velocity convection-diffusion equation. The convergence rates are obtained by computing the ratio* $\ln(e_h/e_{h/2})/\ln(2)$ *as the mesh is been refined. The polynomial space is of order* $0, 1, 2,$ *and* 3, *and the ratio* $\Delta t/h$ *is kept constant during the mesh refinement. The experimental order is* 1 *for first order in time integration and greater than* 2 *for second order in time integration.*

	First order in time				Second order in time			
h	L^2 error	Rate	L^∞ error	Rate	L^2 error	Rate	L^∞ error	Rate
P_0 basis polynomials								
1/8	$7.28E-02$	–	$3.93E-01$	–	$7.29E-02$	–	$3.93E-01$	–
1/16	$6.77E-02$	0.11	$6.92E-01$	−0.82	$6.78E-02$	0.10	$6.93E-01$	−0.82
1/32	$6.06E-02$	0.16	$7.50E-01$	−0.11	$6.09E-02$	0.16	$7.52E-01$	−0.12
1/64	$5.02E-02$	0.27	$6.77E-01$	0.15	$5.06E-02$	0.27	$6.81E-01$	0.14
1/128	$3.71E-02$	0.44	$5.36E-01$	0.34	$3.76E-02$	0.43	$5.41E-01$	0.33
P_1 basis polynomials								
1/8	$4.94E-02$	–	$5.89E-01$	–	$4.89E-02$	–	$5.76E-01$	–
1/16	$3.28E-02$	0.59	$4.49E-01$	0.39	$3.14E-02$	0.64	$4.30E-01$	0.42
1/32	$1.27E-02$	1.37	$1.86E-01$	1.27	$1.06E-02$	1.56	$1.57E-01$	1.46
1/64	$3.89E-03$	1.71	$5.64E-02$	1.72	$2.27E-03$	2.23	$3.26E-02$	2.27
1/128	$1.31E-03$	1.57	$1.89E-02$	1.58	$4.61E-04$	2.30	$6.09E-03$	2.42
P_2 basis polynomials								
1/8	$3.39E-02$	–	$4.77E-01$	–	$3.03E-02$	–	$4.29E-01$	–
1/16	$1.12E-02$	1.60	$1.55E-01$	1.62	$5.83E-03$	2.38	$7.43E-02$	2.53
1/32	$4.49E-03$	1.32	$6.55E-02$	1.24	$4.91E-04$	3.57	$1.30E-02$	2.51
1/64	$2.17E-03$	1.05	$3.11E-02$	1.07	$5.21E-05$	3.24	$2.32E-03$	2.49
1/128	$1.05E-03$	1.05	$1.48E-02$	1.07	$7.91E-06$	2.72	$3.15E-04$	2.88
P_3 basis polynomials								
1/8	$1.86E-02$	–	$2.53E-01$	–	$1.05E-02$	–	$1.31E-01$	–
1/16	$8.02E-03$	1.21	$1.26E-01$	1.01	$6.11E-04$	4.10	$1.99E-02$	2.72
1/32	$4.15E-03$	0.95	$6.87E-02$	0.87	$2.63E-05$	4.54	$2.25E-03$	3.14
1/64	$2.08E-03$	1.00	$3.20E-02$	1.10	$3.40E-06$	2.95	$1.29E-04$	4.12
1/128	$9.81E-04$	1.08	$1.49E-02$	1.10	$5.97E-07$	2.51	$9.24e-06$	3.80

REFERENCES

[1] V. Azinger, C. Dawson, B. Cockburn, and P. Castillo, *The local discontinuous Galerkin method for contaminant transport*, Adv. Wat. Res., 24 (2001), pp. 73–87.

[2] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[3] D. N. Arnold, F. Brezzi, B. Cockburn, and L. Donatella Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39 (2002), pp. 1749–1779.

[4] P. Bastian, *Higher order discontinuous Galerkin methods for flow and transport in porous media*, in Challenges in Scientific Computing—CISC 2002, Lecture Notes Comput. Sci. Eng. 35, Springer, Berlin, 2003, pp. 1–22.

[5] P. Bastian and S. Lang, *Couplex Benchmark Computations with UG*, Computational Geosciences, 8 (2004), pp. 125–147.
Tech. Report 2002-31, IWR (SFB 359), Universität Heidelberg, Germany, 2002; Comput. Geosci., submitted.

[6] P. Bastian and B. Rivière, *Superconvergence and $H($div$)$ Projection for discontinuous Galerkin methods*, Int. J. Numer. Methods Fluids, 42 (2003), pp. 1043–1057.

[7] C. E. Bauman and J. T. Oden, *A discontinuous hp finite element finite element method for convection diffusion problems*, Comput. Methods Appl. Mech. Engrg., 175 (1999), pp. 311–341.

[8] A. Bourgeat, M. Kern, S. Schumacher, and J. Talandier, *The COUPLEX Test Cases: Nuclear Waste Disposal Simulation*, February, 2002.

[9] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo, *Discontinuous Galerkin approximations for elliptic problems*, Numer. Methods Partial Differential Equations, 16 (2000), pp. 365–378.

[10] F. Brezzi, L. D. Marini, and E. Süli, *Discontinuous Galerkin methods for first-order hyperbolic problems*, Math. Models Methods Appl. Sci., 14 (2004), pp. 1893–1903.

[11] G. Chavent and B. Cockburn, *The local projection P^0P^1 -discontinuous Galerkin finite element method for scalar conservative laws*, M2AN Math. Model. Anal. Numer., 23 (1989), pp. 565–592.

[12] P. Castillo, B. Cockburn, I. Perugia, and D. Schötzau, *An a priori error analysis of the local discontinuous Galerkin method for elliptic problems*, SIAM J. Numer. Anal., 38 (2000), pp. 1676–1706.

[13] H. Chen, *Local error estimates of mixed discontinuous Galerkin methods for elliptic problems*, J. Numer. Math., 12 (2004), pp. 1–21.

[14] H. Chen and Z. Chen, *Stability and convergence of mixed discontinuous finite element methods for second-order elliptic problems*, J. Numer. Math., 11 (2003), pp. 253–324.

[15] H. Chen, Z. Chen, and B. Li, *Numerical study of hp version of mixed discontinuous finite element methods for reaction diffusion problems: 1D case*, Numer. Methods Partial Differential Equations, 19 (2003), pp. 525–553.

[16] P. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1975.

[17] B. Cockburn, *Discontinuous Galerkin Methods for Convection Dominated Problems*, School of Mathematics, University of Minnesota, Minneapolis, MN.

[18] B. Cockburn, *An introduction to the discontinuous Galerkin method for convection-dominated problems*, in Advanced Numerical Approximation of Nonlinear Hyperbolic Equations, Lecture Notes in Math. 1697, Springer, Berlin, 1998, pp. 151–268.

[19] B. Cockburn and C. W. Shu, *The Runge-Kutta local projection P^1-discontinuous Galerkin method for scalar conservation laws*, M2AN Math. Model. Anal. Numer., 25 (1991), pp. 337–361.

[20] B. Cockburn and C.-W. Shu, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM J. Numer. Anal., 35 (1998), pp. 2440–2463.

[21] B. Cockburn and C. W. Shu, *TVB Runge Kutta local projection discontinuous Galerkin finite element method for conservative laws* II: *General frame-work*, Math. Comp., 52 (1989), pp. 411–435.

[22] B. Cockburn and C. W. Shu, *TVB Runge Kutta local projection discontinuous Galerkin finite element method for conservative laws* III: *One dimensional systems*, J. Comput. Phys., 84 (1989), pp. 90–113.

[23] B. Cockburn and C. W. Shu, *The Runge-Kutta discontinuous Galerkin method for conservative laws* V: *Multidimensional systems*, J. Comput. Phys., 141 (1998), pp. 199–224.

[24] B. Cockburn, G. E. Karniadakis, and C. W. Shu, *Discontinuous Galerkin Methods, Theory, Computation and Applications*, Lecture Notes in Comput. Sci. Engrg. 11, Springer-Verlag, Berlin, 2000.

[25] M. Crouzeix and A. L. Mignot, *Analyse numérique des équations différentielles*, Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris, 1984.

[26] S. Del Pino and O. Pironneau, *Asymptotic analysis and layer decomposition for the couplex exercise*, Comput. Geosci., 8 (2004), pp. 149–162.

[27] H. Hoteit, *Simulation d'écoulements et de transports de polluants en milieu poreux: Apllication à la modélisation de la sureté des dépôts de déchets radioactifs*, Ph.D. thesis, Université de Rennes 1, Rennes, France, 2002.

[28] T. J. R. Hughes, G. Engel, L. Mazzei, and M. G. Larson, *A comparison of discontinuous and continuous Galerkin methods based on error estimates, conservation, robustness and efficiency*, in Discontinuous Galerkin Methods, Lecture Notes in Comput. Sci. Eng. 11, Springer, Berlin, 2000, pp. 135–146.

[29] W. Hundsdorfer and J. Jaffré, *Implicit-explicit time stepping with spatial discontinuous finite elements*, Appl. Numer. Math., 445 (2003), pp. 231–254.

[30] L. D. Marini, *A survey of DG methods for elliptic problems*, in Proceedings ENUMATH 2001, F. Brezzi, A. Buffa, S. Corsaro, and A. Murli, eds., Springer, Berlin, 2003, pp. 805–814.

[31] S. Prudhomme, F. Pascal, J. T. Oden, and A. Romkes, *Review of A Priori Error Estimation for Discontinuous Galerkin Methods*, TICAM Report 00-27, University of Texas at Austin, Austin, TX, October 17, 2000.

[32] W. H. Reed and T. R. Hill, *Triangular Mesh Methods for the Neutron Transport Equation*, Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM, 1973.

[33] B. Rivière, *Discontinuous Galerkin Methods for Solving the Miscible Displacement Problem in Porous Media*, Ph.D. thesis, University of Texas at Austin, Austin, TX, 2000.

[34] B. Rivière and M. F. Wheeler, *A discontinuous Galerkin method applied to nonlinear Parabolic equations*, in Discontinuous Galerkin Methods, Lecture Notes in Comput. Sci. Eng. 11, Springer, Berlin, 2000, pp. 231–244.

[35] B. Rivière, M. F. Wheeler, and V. Girault *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems* I, Comput. Geosci., 3 (1999), pp. 337–360.

[36] B. Rivière, M. F. Wheeler, and V. Girault, *A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems*, SIAM J. Numer. Anal., 39 (2001), pp. 902–931.

[37] S. Sun, *Discontinuous Galerkin Methods for Reactive Transport in Porous Media*, Ph.D. thesis, The University of Texas at Austin, Austin, TX, 2003.

[38] H. Wang, H. K. Dahle, R. E. Ewing, M. S. Espedal, R. C. Sharpley, and S. Man, *An ELLAM scheme for advection-diffusion equations in two dimensions*, SIAM J. Sci. Comput., 20 (1999), pp. 2160–2194.

[39] L. Yin, A. Acharya, N. Sobh, R. B. Haber, and D. A. Tortorelli, *A space-time discontinuous Galerkin method for elastodynamic analysis*, in Discontinuous Galerkin Methods, Lecture Notes in Comput. Sci. Eng. 11, Springer, Berlin, 1999, pp. 459–464.