Mathematical Modelling and Numerical Analysis
Modélisation Mathématique et Analyse Numérique

ESAIM: M2AN
M2AN, Vol. 36, N° 4, 2002, pp. 705–724
DOI: 10.1051/m2an:2002031

FINITE VOLUME METHODS FOR CONVECTION-DIFFUSION EQUATIONS WITH RIGHT-HAND SIDE IN H^{-1}

Jérôme Droniou[1] and Thierry Gallouët[2]

Abstract. We prove the convergence of a finite volume method for a noncoercive linear elliptic problem, with right-hand side in the dual space of the natural energy space of the problem.

Mathematics Subject Classification. 65N12, 65N30.

Received: November 15, 2001. Revised: April 4, 2002.

1. Introduction

We take Ω a polygonal open subset of $\mathbb{R}^d$ ($d = 2$ or 3), and we study the problem

$$\begin{cases} -\Delta u + \operatorname{div}(\mathbf{v}u) + bu = L & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1.1}$$

with the following hypotheses on the data:

$$\begin{gathered} \exists p > d \text{ such that } \mathbf{v} \in (L^p(\Omega))^d\,, \\ b \in L^r(\Omega) \text{ with } r > 1 \text{ if } d = 2 \text{ and } r = \tfrac{3}{2} \text{ if } d = 3\,,\ b \geq 0 \text{ a.e. on } \Omega\,, \\ L \in H^{-1}(\Omega). \end{gathered} \tag{1.2}$$

Of course, solutions to (1.1) are taken in a weak sense, that is to say

$$\begin{cases} u \in H_0^1(\Omega)\,, \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \varphi - \int_\Omega u\mathbf{v} \cdot \nabla \varphi + \int_\Omega bu\varphi = \langle L, \varphi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)}\,,\ \forall \varphi \in H_0^1(\Omega). \end{cases} \tag{1.3}$$

Existence and uniqueness of a solution to (1.3) have already been proved in [3] (see also [4] for nonlinear problems).

Our purpose is to prove the convergence of a finite volume discretization of (1.1). Finite volume methods have been widely used to approximate solutions to convection-diffusion equations, either using structured or

Keywords and phrases. Finite volumes, convection-diffusion equations, noncoercivity, non-regular data.

[1] UMPA, ENS Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France. e-mail: jdroniou@umpa.ens-lyon.fr

[2] Université de Provence, CMI, Technopôle de Château Gombert, 39 rue F. Joliot Curie, 13453 Marseille Cedex 13, France. e-mail: gallouet@cmi.univ-mrs.fr

unstructured grids (see for example [2, 5, 6, 8, 9]). The grids we consider here are the same as in [5], that is to say grids made of convex polygonal control volumes with some geometrical properties (see the next section).

There are two main originalities in the work we present here. First, we consider elliptic problems which are not necessarily coercive, because it is not supposed that $\frac{1}{2}\mathrm{div}(\mathbf{v}) + b$ is nonnegative. Moreover, the regularity we have taken on the velocity $\mathbf{v}$ is minimal (that is, just enough for (1.3) to make sense — in previous papers on the finite volume discretization of convection-diffusion equations, the convection velocity is in general C^1-continuous, see *e.g.* [5] or [9]); considering a non-regular convection velocity is a first step toward the treatment of coupled systems, in which $\mathbf{v}$ comes from the resolution of another partial differential equation.

The second originality concerns the right-hand side: here too, we consider a datum with minimal regularity (that is, in the dual space of the energy space associated to the equation — previous papers take in general a right-hand side in $L^2(\Omega)$); in fact, $H^{-1}(\Omega)$ is a natural space for right-hand sides of convection-diffusion equations.

In the next section, we define the finite volume scheme used to discretize (1.1), and we state the main convergence result of this paper; since we consider data $\mathbf{v}$ and L which lack of regularity (with respect to previous works), we present a new way to discretize them, using what we call "half-diamonds". We also give, in this section, technical results useful to the rest of the paper. In Section 3, we prove *a priori* estimates on the solutions to our finite volume discretization of (1.1); the problem being noncoercive, obtaining estimates on these solutions is not straightforward: we must adapt the techniques of [3] to the discrete setting. Along with the compactness results of [5], these *a priori* estimates allow us, in Section 4, to prove our main result, that is to say existence and uniqueness of the approximate solutions and their convergence toward the solution of (1.3); to prove the convergence result with our irregular data, we approximate them by regular data and adapt then known techniques (see [5], for example). In the last section, we present a modified scheme which consists in discretizing the data $\mathbf{v}$ and L using another method (based on the "full-diamonds"); comparing this scheme to the one of Section 2, we easily obtain the convergence of the associated approximate solutions.

2. Definition of the scheme and main result

Definition 2.1. An admissible mesh $\mathcal{T}$ of Ω is a finite family of polygonal open convex subsets of Ω (the "control volumes"), together with a finite family $\mathcal{E}$ of disjoint subsets of $\overline{\Omega}$ contained in affine hyperplanes (the "edges") and a family $\mathcal{P} = (x_K)_{K\in\mathcal{T}}$ of points in Ω such that:

(i) $\overline{\Omega} = \bigcup_{K\in\mathcal{T}} \overline{K}$;
(ii) each $\sigma \in \mathcal{E}$ is a non-empty open subset of ∂K for some $K \in \mathcal{T}$;
(iii) by denoting $\mathcal{E}_K = \{\sigma \in \mathcal{E} \mid \sigma \subset \partial K\}$, $\partial K = \cup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$ for all $K \in \mathcal{T}$;
(iv) for all $K \neq L$ in $\mathcal{T}$, either the $(d-1)$-dimensional measure of $\overline{K} \cap \overline{L}$ is null, or $\overline{K} \cap \overline{L} = \overline{\sigma}$ for some $\sigma \in \mathcal{E}$, that we denote then $\sigma = K|L$;
(v) for all $K \in \mathcal{T}$, $x_K \in K$;
(vi) for all $\sigma = K|L \in \mathcal{E}$, the line (x_K, x_L) intersects and is orthogonal to σ;
(vii) for all $\sigma \in \mathcal{E}$, $\sigma \subset \partial\Omega \cap \partial K$, the line which is orthogonal to σ and going through x_K intersects σ.

The size of the mesh is then defined by $\mathrm{size}(\mathcal{T}) = \sup_{K\in\mathcal{T}} \mathrm{diam}(K)$ (where $\mathrm{diam}(K)$ is the diameter of K). We denote by $\mathrm{meas}(K)$ the Lebesgue measure of $K \in \mathcal{T}$. The unit normal to $\sigma \in \mathcal{E}_K$ outward to K is denoted by $\mathbf{n}_{K,\sigma}$.

We define $\mathcal{E}_{\mathrm{int}} = \{\sigma \in \mathcal{E} \mid \sigma \not\subset \partial\Omega\}$ and $\mathcal{E}_{\mathrm{ext}} = \mathcal{E}\backslash\mathcal{E}_{\mathrm{int}}$. If $\sigma \in \mathcal{E}$, $m(\sigma)$ is the $(d-1)$-dimensional measure of σ; if $\sigma = K|L \in \mathcal{E}_{\mathrm{int}}$, d_σ is the Euclidean distance between the points (x_K, x_L) and $d_{K,\sigma}$ denotes the distance between x_K and σ; if $\sigma \in \mathcal{E}_{\mathrm{ext}} \cap \mathcal{E}_K$, $d_\sigma = d_{K,\sigma}$ is the distance between x_K and σ. The transmissivity through an edge σ is $\tau_\sigma = \frac{m(\sigma)}{d_\sigma}$. We denote by γ the $(d-1)$-dimensional measure on the edges of the mesh.

If $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, the "half-diamond" $\triangle_{K,\sigma}$ is defined by $\triangle_{K,\sigma} = \{tx_K + (1-t)x\,,\ t \in [0,1]\,,\ x \in \sigma\}$. It will be useful to notice that $\mathrm{meas}(\triangle_{K,\sigma}) = \frac{m(\sigma)d_{K,\sigma}}{d}$.

The following quantity measures the "regularity" of the mesh:

$$\mathrm{reg}(\mathcal{T}) = \inf_{K\in\mathcal{T}} \left(\inf_{\sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}}{d_\sigma} \right).$$

If $\mathcal{T}$ is an admissible mesh, and under Hypothesis (1.2), we can define the finite volume discretization of (1.1).

We first write

$$L = f + \mathrm{div}(G), \quad \text{with } f \in L^2(\Omega) \text{ and } G \in (L^2(\Omega))^d.$$

It is well-known that any element of $H^{-1}(\Omega)$ can be written this way; in fact, in models of physical problems, the right-hand side naturally appears in this form, see *e.g.* [7], and there is thus no trouble to define the following scheme (this is also why we have kept f, which can be taken, from a theoretical point of view, null).

The finite volume discretization consists in integrating the equation $-\Delta u + \mathrm{div}(\mathbf{v}u) + bu = f + \mathrm{div}(G)$ on a control volume K: with some integrates by parts, we formally obtain

$$\sum_{\sigma\in\mathcal{E}_K} -\int_\sigma \nabla u\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma + \sum_{\sigma\in\mathcal{E}_K}\int_\sigma u\mathbf{v}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma + \int_K bu = \int_K f + \sum_{\sigma\in\mathcal{E}_K}\int_\sigma G\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma.$$

By letting u_K be an approximate value of u on the control volume K, we must then discretize each term of this relation. To this aim, we denote, for $K\in\mathcal{T}$ and $\sigma\in\mathcal{E}_K$,

$$\begin{aligned} &v_{K,\sigma} = \left(\frac{1}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}}\mathbf{v}\right)\cdot\mathbf{n}_{K,\sigma}\,, \quad b_K = \frac{1}{\mathrm{meas}(K)}\int_K b\,,\\ &f_K = \frac{1}{\mathrm{meas}(K)}\int_K f \quad \text{and} \quad G_{K,\sigma} = \left(\frac{1}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} G\right)\cdot\mathbf{n}_{K,\sigma} \end{aligned} \tag{2.1}$$

(these are, respectively, approximate values of $\mathbf{v}\cdot\mathbf{n}_{K,\sigma}$ on σ, of b on K, of f on K and of $G\cdot\mathbf{n}_{K,\sigma}$ on σ), and the finite volume scheme is written

$$\forall K\in\mathcal{T}\,,\ \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} + \sum_{\sigma\in\mathcal{E}_K} m(\sigma)v_{K,\sigma}u_{K,\sigma,+} + \mathrm{meas}(K)b_Ku_K = \mathrm{meas}(K)f_K + \sum_{\sigma\in\mathcal{E}_K} m(\sigma)G_{K,\sigma}\,, \tag{2.2}$$

$$\forall K\in\mathcal{T}\,,\forall\sigma\in\mathcal{E}_K\,,\ F_{K,\sigma} = -\tfrac{m(\sigma)}{d_{K,\sigma}}(u_\sigma - u_K)\,, \tag{2.3}$$

$$\begin{aligned} &\forall\sigma = K|L\in\mathcal{E}_{\mathrm{int}}\,,\ F_{K,\sigma} + m(\sigma)v_{K,\sigma}u_{K,\sigma,+} - m(\sigma)G_{K,\sigma} = -(F_{L,\sigma} + m(\sigma)v_{L,\sigma}u_{L,\sigma,+} - m(\sigma)G_{L,\sigma})\,,\\ &\forall\sigma\in\mathcal{E}_{\mathrm{ext}}\,, \qquad u_\sigma = 0\,, \end{aligned} \tag{2.4}$$

$$\begin{aligned} &\forall\sigma = K|L\in\mathcal{E}_{\mathrm{int}}\,,\ u_{K,\sigma,+} = u_K \ \text{ if } \ v_{K,\sigma}\geq 0\,,\ u_{K,\sigma,+} = u_L \ \text{ otherwise,}\\ &\forall\sigma\in\mathcal{E}_{\mathrm{ext}}\cap\mathcal{E}_K\,, \quad u_{K,\sigma,+} = u_K \ \text{ if } \ v_{K,\sigma}\geq 0\,,\ u_{K,\sigma,+} = 0 \ \text{ otherwise.} \end{aligned} \tag{2.5}$$

Equations (2.2–2.5) are a linear system in $(u_K)_{K\in\mathcal{T}}$ and $(u_\sigma)_{\sigma\in\mathcal{E}}$, but thanks to (2.4) (which describes the conservativity of the fluxes), we can eliminate the unknowns $(u_\sigma)_{\sigma\in\mathcal{E}}$, so that (2.2–2.5) can be considered as a linear system of size $\mathrm{Card}(\mathcal{T})$, with unknowns $(u_K)_{K\in\mathcal{T}}$.

We naturally identify the set $\mathbb{R}^{\mathrm{Card}(\mathcal{T})}$ to the set $X(\mathcal{T})$ of functions defined a.e. on Ω and constant on each control volume $K\in\mathcal{T}$.

Our main result is the following.

Theorem 2.1. *If $\mathcal{T}$ is an admissible mesh, then there exists a unique solution to (2.2–2.5). Moreover, let $\alpha > 0$; denoting by $u_\mathcal{T} \in X(\mathcal{T})$ the solution to (2.2–2.5), $u_\mathcal{T}$ converges in $L^q(\Omega)$, for all $q < \frac{2d}{d-2}$, to the unique solution of (1.3), as* $\mathrm{size}(\mathcal{T}) \to 0$ *with* $\mathrm{reg}(\mathcal{T}) \geq \alpha$.

Remark 2.1. We will not use, to prove this theorem, the existence of a solution to (1.3). The finite volume method allows, as usual, to prove the existence of a solution to the continuous problem.

Remark 2.2. In dimension $d = 2$, the regularity we suppose on $\mathbf{v}$ is minimal in order for all the terms in (1.3) to make sense (see the Sobolev imbeddings in [1]). But, if $d = 3$, the minimal regularity on the convection velocity would be: $\mathbf{v} \in (L^3(\Omega))^3$; in fact, cutting $\mathbf{v}$ in two parts (one small in $(L^3(\Omega))^3$, the other in $(L^\infty(\Omega))^3$ — see [3] for the reasoning in the continuous case), we could also prove Theorem 2.1 under this minimal hypothesis on $\mathbf{v}$. However, for the legibility of the following proofs, we prefer to suppose Hypothesis (1.2).

2.1. Technical results

To prove this existence, uniqueness and convergence result, we first search for *a priori* estimates on the solutions to (2.2–2.5). These estimates are obtained *via* the following discrete H_0^1 norm.

Definition 2.2. If $\mathcal{T}$ is an admissible mesh and $v_\mathcal{T} = (v_K)_{K\in\mathcal{T}} \in X(\mathcal{T})$, we define

$$||v_\mathcal{T}||_{1,\mathcal{T}} = \left(\sum_{\sigma\in\mathcal{E}} \tau_\sigma (D_\sigma v_\mathcal{T})^2 \right)^{1/2},$$

where $D_\sigma v_\mathcal{T} = |v_K - v_L|$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ and $D_\sigma v_\mathcal{T} = |v_K|$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$.

Notice that this norm takes into account a boundary condition "$v_\mathcal{T} = 0$ on $\partial\Omega$", since we have defined $D_\sigma v_\mathcal{T} = |v_K|$ if $\sigma \subset \partial\Omega$ (this comes down to consider that functions of $X(\mathcal{T})$ are defined on $\mathbb{R}^N$ and are null outside Ω).

The following proposition sums up a few useful properties of the norm $||\cdot||_{1,\mathcal{T}}$.

Proposition 2.1.

(i) *(Discrete Poincaré inequality) If $\mathcal{T}$ is an admissible mesh and $v_\mathcal{T} \in X(\mathcal{T})$, then $||v_\mathcal{T}||_{L^2(\Omega)} \leq \mathrm{diam}(\Omega)||v_\mathcal{T}||_{1,\mathcal{T}}$ (where* $\mathrm{diam}(\Omega)$ *is the diameter of Ω).*

(ii) *(Discrete Sobolev inequality) If $\mathcal{T}$ is an admissible mesh and $0 < \zeta \leq \mathrm{reg}(\mathcal{T})$, then there exists C only depending on (Ω, ζ) such that, for all $q \in [1, \frac{2d}{d-2}[$, for all $v_\mathcal{T} \in X(\mathcal{T})$, $||v_\mathcal{T}||_{L^q(\Omega)} \leq Cq||v_\mathcal{T}||_{1,\mathcal{T}}$.*

(iii) *(Discrete Rellich Theorem) If $(\mathcal{T}_n)_{n\geq 1}$ is a sequence of admissible meshes such that* $\mathrm{size}(\mathcal{T}_n) \to 0$ *and if $v_n \in X(\mathcal{T}_n)$ is such that $(||v_n||_{1,\mathcal{T}_n})_{n\geq 1}$ is bounded, then $(v_n)_{n\geq 1}$ is relatively compact in $L^2(\Omega)$ and any adherence value in $L^2(\Omega)$ of $(v_n)_{n\geq 1}$ belongs to $H_0^1(\Omega)$.*

For a proof of these properties, see [5].

The following discrete integrate by parts formula will be quite useful in the sequel.

Lemma 2.1. *Let $\mathcal{T}$ be an admissible mesh and $u_\mathcal{T} = (u_K)_{K\in\mathcal{T}}$ satisfy (2.2–2.5). Then, for all $\varphi_\mathcal{T} = (\varphi_K)_{K\in\mathcal{T}} \in X(\mathcal{T})$, we have*

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}} \tau_\sigma (u_K - u_L)(\varphi_K - \varphi_L) + \sum_{K\in\mathcal{T}} \mathrm{meas}(K) b_K u_K \varphi_K = \sum_{K\in\mathcal{T}} \mathrm{meas}(K) f_K \varphi_K \\
&+ \sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (\varphi_K - \varphi_L) + \sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} G_{L,\sigma} \right) (\varphi_K - \varphi_L),
\end{aligned} \tag{2.6}$$

where we have denoted $\sigma = K|L$ if $\sigma \in \mathcal{E}_{\rm int}$ and $u_L = u_{L,\sigma,+} = v_{L,\sigma} = d_{L,\sigma} = G_{L,\sigma} = \varphi_L = 0$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$.

Proof of Lemma 2.1. We notice that, thanks to (2.4), the quantity $a_{K,\sigma} = F_{K,\sigma} + m(\sigma)v_{K,\sigma}u_{K,\sigma,+} - m(\sigma)G_{K,\sigma}$ is conservative, that is to say, if $\sigma = K|L \in \mathcal{E}_{\rm int}$, then $a_{K,\sigma} = -a_{L,\sigma}$.

Multiplying (2.2) by φ_K and summing on the control volumes $K \in \mathcal{T}$, we have

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} a_{K,\sigma}\varphi_K + \sum_{K\in\mathcal{T}} {\rm meas}(K) b_K u_K \varphi_K = \sum_{K\in\mathcal{T}} {\rm meas}(K) f_K \varphi_K.$$

Using the conservativity of $a_{K,\sigma}$ and gathering by edges, we deduce

$$\sum_{\sigma\in\mathcal{E}} a_{K,\sigma}(\varphi_K - \varphi_L) + \sum_{K\in\mathcal{T}} {\rm meas}(K) b_K u_K \varphi_K = \sum_{K\in\mathcal{T}} {\rm meas}(K) f_K \varphi_K \tag{2.7}$$

where $\sigma = K|L$ if $\sigma \in \mathcal{E}_{\rm int}$ and $\varphi_L = 0$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$.

Let us now compute the $(a_{K,\sigma})_{K\in\mathcal{T},\,\sigma\in\mathcal{E}_K}$. If $\sigma = K|L \in \mathcal{E}_{\rm int}$, then (2.3) and (2.4) give u_σ; indeed, dividing (2.4) by $m(\sigma)$, we have

$$-\frac{u_\sigma}{d_{K,\sigma}} + \frac{u_K}{d_{K,\sigma}} + v_{K,\sigma}u_{K,\sigma,+} - G_{K,\sigma} = \frac{u_\sigma}{d_{L,\sigma}} - \frac{u_L}{d_{L,\sigma}} - v_{L,\sigma}u_{L,\sigma,+} + G_{L,\sigma},$$

that is, noticing that $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$,

$$\frac{d_\sigma}{d_{K,\sigma}d_{L,\sigma}}u_\sigma = \frac{u_K}{d_{K,\sigma}} + \frac{u_L}{d_{L,\sigma}} + v_{K,\sigma}u_{K,\sigma,+} + v_{L,\sigma}u_{L,\sigma,+} - G_{K,\sigma} - G_{L,\sigma},$$

which gives

$$u_\sigma = \frac{d_{L,\sigma}}{d_\sigma}u_K + \frac{d_{K,\sigma}}{d_\sigma}u_L + \frac{d_{K,\sigma}d_{L,\sigma}}{d_\sigma}\left(v_{K,\sigma}u_{K,\sigma,+} + v_{L,\sigma}u_{L,\sigma,+} - G_{K,\sigma} - G_{L,\sigma}\right).$$

With this value of u_σ, we obtain

$$\begin{aligned} a_{K,\sigma} &= -\frac{m(\sigma)}{d_{K,\sigma}}\left(\frac{d_{K,\sigma}}{d_\sigma}u_L - \frac{d_{K,\sigma}}{d_\sigma}u_K\right) - \frac{m(\sigma)d_{L,\sigma}}{d_\sigma}\left(v_{K,\sigma}u_{K,\sigma,+} + v_{L,\sigma}u_{L,\sigma,+} - G_{K,\sigma} - G_{L,\sigma}\right) \\ &\quad + m(\sigma)v_{K,\sigma}u_{K,\sigma,+} - m(\sigma)G_{K,\sigma} \\ &= \tau_\sigma(u_K - u_L) + m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+} - \frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}\right) \\ &\quad - m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right). \end{aligned}$$

Note that this equality is also valid if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$, providing that we define $u_L = u_{L,\sigma,+} = v_{L,\sigma} = G_{L,\sigma} = \varphi_L = 0$ in this case.

Using this expression in (2.7), we obtain the desired formula. □

3. *A PRIORI* ESTIMATES

We prove here some *a priori* estimates on the solution to (2.2–2.5). As already said, we adapt the methods of [3] to the discrete setting; however, the estimation of the convection term (the noncoercive part of the equation) requires new ideas, to take advantage of the upwind choice in (2.5).

3.1. Estimate on $\ln(1+|u_{\mathcal{T}}|)$

Proposition 3.1. *Let $\mathcal{T}$ be an admissible mesh. If $(u_K)_{K\in\mathcal{T}}$ is a solution to (2.2–2.5), then*

$$||\ln(1+|u_{\mathcal{T}}|)||^2_{1,\mathcal{T}} \leq 2||f||_{L^1(\Omega)} + 2d\left(||\,|G|\,||_{L^2(\Omega)} + ||\,|\mathbf{v}|\,||_{L^2(\Omega)}\right)^2,$$

where $|X|$ denotes the Euclidean norm of a vector $X\in\mathbb{R}^d$.

Proof of Proposition 3.1.

Step 1: A preliminary estimate.

Let $\varphi(s)=\int_0^s \frac{\mathrm{d}t}{(1+|t|)^2}$. Applying Formula (2.6) to $(\varphi_K)_{K\in\mathcal{T}}=(\varphi(u_K))_{K\in\mathcal{T}}$, and since φ is bounded by 1 and $b_K u_K\varphi(u_K)\geq 0$ for all $K\in\mathcal{T}$, we have

$$\begin{aligned}\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L)) \leq{}& \sum_{K\in\mathcal{T}}\mathrm{meas}(K)|f_K| \\ &+\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)) \\ &+\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(u_K)-\varphi(u_L)) \qquad (3.1)\end{aligned}$$

(with the notation $\sigma=K|L$ if $\sigma\in\mathcal{E}_{\rm int}$ and $u_L=u_{L,\sigma,+}=v_{L,\sigma}=d_{L,\sigma}=G_{L,\sigma}=\varphi(u_L)=0$ if $\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}_K$).

We have

$$\sum_{K\in\mathcal{T}}\mathrm{meas}(K)|f_K| \leq \sum_{K\in\mathcal{T}}\int_K |f| = ||f||_{L^1(\Omega)}. \qquad (3.2)$$

Using the Cauchy-Schwarz inequality, we can write

$$\begin{aligned}&\left|\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(u_K)-\varphi(u_L))\right| \\ &\qquad\leq\left(\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)^2\right)^{1/2}\left(\sum_{\sigma\in\mathcal{E}}\tau_\sigma(\varphi(u_K)-\varphi(u_L))^2\right)^{1/2}. \qquad (3.3)\end{aligned}$$

Since $\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)^2 \leq 2\frac{d_{K,\sigma}}{d_\sigma}G^2_{K,\sigma}+2\frac{d_{L,\sigma}}{d_\sigma}G^2_{L,\sigma}$ (we have used the fact that $\frac{d_{K,\sigma}}{d_\sigma}$ and $\frac{d_{L,\sigma}}{d_\sigma}$ are bounded by 1), gathering by control volumes, we have

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)^2 \leq 2\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}G^2_{K,\sigma}.$$

But, for all $K\in\mathcal{T}$ and all $\sigma\in\mathcal{E}_K$, by Jensen's inequality and since $\mathrm{meas}(\triangle_{K,\sigma})=\frac{m(\sigma)d_{K\sigma}}{d}$, we have $m(\sigma)d_{K,\sigma}G^2_{K,\sigma}\leq d\int_{\triangle_{K,\sigma}}|G|^2$. Using the fact that $\{\triangle_{K,\sigma},\ K\in\mathcal{T},\ \sigma\in\mathcal{E}_K\}$ is (up to a set of null Lebesgue measure) a partition of Ω, we deduce

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)^2 \leq 2d\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\int_{\triangle_{K,\sigma}}|G|^2 = 2d||\,|G|\,||^2_{L^2(\Omega)}. \qquad (3.4)$$

φ being nondecreasing and Lipschitz-continuous with Lipschitz constant 1, we have $(\varphi(u_K)-\varphi(u_L))^2 \le (u_K - u_L)(\varphi(u_K)-\varphi(u_L))$; (3.3) and (3.4) give then

$$\left|\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(u_K)-\varphi(u_L))\right| \le \sqrt{2d}\,||\,|G|\,||_{L^2(\Omega)}\left(\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L))\right)^{1/2}. \quad (3.5)$$

Now, we need to estimate the terms of (3.1) coming from the discretization of the convection part of (1.1). We first notice that, if $\sigma \in \mathcal{E}_{\rm ext}\cap\mathcal{E}_K$,

$$\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)) = -\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\varphi(u_K) \le 0.$$

Indeed, if $v_{K,\sigma}\ge 0$, this last term is $-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_K\varphi(u_K)$, which is nonpositive since $s\varphi(s)\ge 0$ for all $s\in\mathbb{R}$; if $v_{K,\sigma}<0$, this last term is null (because $u_{K,\sigma,+}=0$ in this case). Thus,

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)) \le \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)). \quad (3.6)$$

Let $\sigma = K|L \in \mathcal{E}_{\rm int}$ and denote

$$\Lambda = \left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)), \quad A = \left|\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}\right| \text{ and } B = \left|\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}\right|.$$

We separate the cases:

- If $v_{K,\sigma}$ and $v_{L,\sigma}$ are nonnegative, then $\Lambda = (Au_L - Bu_K)(\varphi(u_K)-\varphi(u_L))$ is, by item (i) of Lemma 3.1 below, bounded from above by 0 if $u_Ku_L\le 0$ and by $|A-B|\inf(|u_L|,|u_K|)|\varphi(u_K)-\varphi(u_L)|$ otherwise.
- If $v_{K,\sigma}$ and $v_{L,\sigma}$ are negative, then $\Lambda = (-Au_K+Bu_L)(\varphi(u_K)-\varphi(u_L)) = (Bu_L-Au_K)(\varphi(u_K)-\varphi(u_L))$ is once again bounded from above by 0 if $u_Ku_L\le 0$ and by $|A-B|\inf(|u_L|,|u_K|)|\varphi(u_K)-\varphi(u_L)|$ otherwise.
- If $v_{K,\sigma}\ge 0$ and $v_{L,\sigma}<0$, then $\Lambda = -(A+B)u_K(\varphi(u_K)-\varphi(u_L))$ is, by item (ii) of Lemma 3.1 below, bounded from above by 0 if $u_Ku_L\le 0$ and by $(A+B)\inf(|u_L|,|u_K|)|\varphi(u_K)-\varphi(u_L)|$ otherwise.
- If $v_{K,\sigma}<0$ and $v_{L,\sigma}\ge 0$, then $\Lambda = (A+B)u_L(\varphi(u_K)-\varphi(u_L)) = -(A+B)u_L(\varphi(u_L)-\varphi(u_K))$ is, as before, bounded from above by 0 if $u_Ku_L\le 0$ and by $(A+B)\inf(|u_K|,|u_L|)|\varphi(u_L)-\varphi(u_K)|$ otherwise.

In either case, we notice that $\Lambda\le 0$ if $u_Ku_L\le 0$ and that $\Lambda\le (A+B)\inf(|u_K|,|u_L|)|\varphi(u_K)-\varphi(u_L)|$ otherwise; thus, by denoting $\mathcal{A}=\{\sigma=K|L\in\mathcal{E}_{\rm int}\mid u_Ku_L>0\}$, (3.6) gives

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L)) \le \sum_{\sigma\in\mathcal{A}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}|v_{L,\sigma}| + \frac{d_{K,\sigma}}{d_\sigma}|v_{K,\sigma}|\right)\inf(|u_K|,|u_L|)|\varphi(u_K)-\varphi(u_L)|.$$

Since $\frac{d_{K,\sigma}}{d_\sigma}$ and $\frac{d_{L,\sigma}}{d_\sigma}$ are bounded by 1, we obtain

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L))$$
$$\leq \left(2\sum_{\sigma\in\mathcal{A}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}^2+\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}^2\right)\right)^{1/2}\left(\sum_{\sigma\in\mathcal{A}}\tau_\sigma \inf(|u_K|,|u_L|)^2(\varphi(u_K)-\varphi(u_L))^2\right)^{1/2}. \quad (3.7)$$

Gathering by control volumes, and using Jensen's inequality, we can write

$$\sum_{\sigma\in\mathcal{A}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}^2+\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}^2\right)\leq \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)d_{K,\sigma}}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}}|\mathbf{v}|^2.$$

Since $\frac{m(\sigma)d_{K,\sigma}}{\mathrm{meas}(\triangle_{K,\sigma})}=d$ and $\{\triangle_{K,\sigma}\,,\ K\in\mathcal{T}\,,\ \sigma\in\mathcal{E}_K\}$ is (up to a set of null Lebesgue measure) a partition of Ω, we deduce

$$\sum_{\sigma\in\mathcal{A}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}^2+\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}^2\right)\leq d\,||\,|\mathbf{v}|\,||_{L^2(\Omega)}^2. \quad (3.8)$$

For all $\sigma=K|L\in\mathcal{A}$, since $u_Ku_L>0$, item (iii) of Lemma 3.1 gives

$$\inf(|u_K|,|u_L|)^2(\varphi(u_K)-\varphi(u_L))^2\leq(u_K-u_L)(\varphi(u_K)-\varphi(u_L)).$$

Using this and (3.8) in (3.7), we finally obtain

$$\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(u_K)-\varphi(u_L))$$
$$\leq\sqrt{2d}\,||\,|\mathbf{v}|\,||_{L^2(\Omega)}\left(\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L))\right)^{1/2}. \quad (3.9)$$

Gathering (3.2, 3.5) and (3.9) in (3.1), we get

$$\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L))$$
$$\leq||f||_{L^1(\Omega)}+\sqrt{2d}\left(||\,|G|\,||_{L^2(\Omega)}+||\,|\mathbf{v}|\,||_{L^2(\Omega)}\right)\left(\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L))\right)^{1/2},$$

which gives, thanks to Young's inequality,

$$\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(u_K)-\varphi(u_L))\leq 2||f||_{L^1(\Omega)}+2d\left(||\,|G|\,||_{L^2(\Omega)}+||\,|\mathbf{v}|\,||_{L^2(\Omega)}\right)^2. \quad (3.10)$$

Step 2: Estimate on $\ln(1+|u_{\mathcal{T}}|)$.

We notice that, for all $s \in \mathbb{R}$, $\ln(1+|s|) = \int_0^s \frac{\operatorname{sgn}(t)\,\mathrm{d}t}{1+|t|}$. Thus, for all $(x,y) \in \mathbb{R}^2$, by the Cauchy-Schwarz inequality and since φ is nondecreasing,

$$\begin{aligned}(\ln(1+|x|) - \ln(1+|y|))^2 &= \left(\int_y^x \frac{\operatorname{sgn}(t)\,\mathrm{d}t}{1+|t|}\right)^2 \\ &\le |x-y| \left|\int_y^x \frac{\mathrm{d}t}{(1+|t|)^2}\right| = |x-y||\varphi(x)-\varphi(y)| = (x-y)(\varphi(x)-\varphi(y)).\end{aligned}$$

Using this bound in (3.10), we deduce the desired estimate on $\ln(1+|u_{\mathcal{T}}|)$. □

It remains to state and prove the following technical result, which has been used in the course of the preceding proof. This lemma shows the usefulness of the upwind choice in (2.5): thanks to the first two items of the lemma, the upwind choice allows to reduce the estimate on the discrete convection term to the cases $u_K u_L > 0$; these cases are then, thanks to item (iii), bounded by the discrete diffusion term.

Lemma 3.1. *Let $\varphi(s) = \int_0^s \frac{\mathrm{d}t}{(1+|t|)^2}$.*

(i) *Let A and B be nonnegative real numbers and $(x,y) \in \mathbb{R}^2$. If $xy \le 0$, then*

$$(Ax - By)(\varphi(y) - \varphi(x)) \le 0$$

and, if $xy > 0$, then

$$(Ax - By)(\varphi(y) - \varphi(x)) \le |A-B| \inf(|x|,|y|)|\varphi(y) - \varphi(x)|.$$

(ii) *Let $(x,y) \in \mathbb{R}^2$. If $xy \le 0$, then*

$$-y(\varphi(y) - \varphi(x)) \le 0$$

and, if $xy > 0$, then

$$-y(\varphi(y) - \varphi(x)) \le \inf(|x|,|y|)|\varphi(y) - \varphi(x)|.$$

(iii) *Let $(x,y) \in \mathbb{R}^2$. If $xy > 0$, then*

$$\inf(|x|,|y|)^2(\varphi(y) - \varphi(x))^2 \le (y-x)(\varphi(y) - \varphi(x)).$$

Proof of Lemma 3.1. The first two items are only consequences of the nondecreasingness of φ and of the fact that $s\varphi(s) \ge 0$ for all $s \in \mathbb{R}$.

Consider (i). Suppose first that $xy \le 0$. Up to a permutation of x and y, there is no loss of generality if we assume that $x \le 0$. If $x = 0$, then $(Ax - By)(\varphi(y) - \varphi(x)) = -By\varphi(y) \le 0$. If $x < 0$, then $y \ge 0 > x$ and, A and B being nonnegative, we have $By \ge 0 \ge Ax$, thus $Ax - By \le 0$; φ being nondecreasing, we deduce that $(Ax - By)(\varphi(y) - \varphi(x)) \le 0$.

Suppose now that $xy > 0$. Up to a permutation of x and y, we can suppose that $|x| \le |y|$. We have then

$$(Ax - By)(\varphi(y) - \varphi(x)) = (A-B)x(\varphi(y) - \varphi(x)) + B(x-y)(\varphi(y) - \varphi(x)).$$

φ being nondecreasing, the second term of the right-hand side of this equality is nonpositive, and we obtain thus $(Ax - By)(\varphi(y) - \varphi(x)) \le (A-B)x(\varphi(y) - \varphi(x)) \le |A-B|\,|x|\,|\varphi(y) - \varphi(x)|$ as desired.

Let us now study the second item. If $xy \le 0$, then either $x = 0$, or $y = 0$, or $x < 0 < y$ or $y < 0 < x$. In the first case, $-y(\varphi(y) - \varphi(x)) = -y\varphi(y) \le 0$; in the second case, $-y(\varphi(y) - \varphi(x)) = 0$; in the third case, $-y \le 0$

and $\varphi(y)-\varphi(x)\geq 0$ so that the result holds; in the fourth case, $-y\geq 0$ but $\varphi(y)-\varphi(x)\leq 0$ and the result still holds. Assume now that $xy>0$; the result is obvious if $|y|=\inf(|x|,|y|)$, so that we can take $|y|\geq|x|$; then either $0<x\leq y$ or $y\leq x<0$. In both cases, the nondecreasingness of φ easily gives $-y(\varphi(y)-\varphi(x))\leq 0$, and the desired inequality is thus satisfied.

To prove the third item, we notice that, since φ is C^1-continuous on $\mathbb{R}$, there exists $\theta\in[x,y]$ such that $\varphi(y)-\varphi(x)=\varphi'(\theta)(y-x)$. Using the fact that φ is nondecreasing, we obtain

$$\begin{aligned}\inf(|x|,|y|)^2(\varphi(y)-\varphi(x))^2 &\leq \frac{\inf(|x|,|y|)^2}{(1+|\theta|)^2}|y-x|\,|\varphi(y)-\varphi(x)|\\ &\leq \frac{\inf(|x|,|y|)^2}{(1+|\theta|)^2}(y-x)(\varphi(y)-\varphi(x)).\end{aligned}$$

But, since x and y have the same sign and $\theta\in[x,y]$, we have $\inf(|x|,|y|)\leq|\theta|$, and the result is thus a consequence of the previous inequality. □

3.2. Estimate on $||u_{\mathcal{T}}||_{1,\mathcal{M}}$

Theorem 3.1. *Let $\mathcal{T}$ be an admissible mesh, $0<\zeta\leq\mathrm{reg}(\mathcal{T})$ and M be an upper bound of $||\,|\mathbf{v}|\,||_{L^p(\Omega)}$. There exists $C>0$ only depending on (Ω,p,M,ζ) such that, if $u_{\mathcal{T}}$ is a solution to (2.2–2.5), then*

$$||u_{\mathcal{T}}||_{1,\mathcal{T}}\leq C(||f||_{L^2(\Omega)}+||\,|G|\,||_{L^2(\Omega)}).$$

Proof of Theorem 3.1. (2.2–2.5) being a linear system, proving a bound on $u_{\mathcal{T}}$ whenever

$$||f||_{L^2(\Omega)}+||\,|G|\,||_{L^2(\Omega)}\leq 1 \tag{3.11}$$

is enough to prove the theorem in the general case.

We denote, for $k>0$, $T_k(s)=\max(-k,\min(s,k))$ and $S_k(s)=s-T_k(s)$.

Step 1: estimate on $S_k(u_{\mathcal{T}})$.

Let $k>0$. We use (2.6) with $\varphi_K=S_k(u_K)$; since $(S_k(u_K)-S_k(u_L))^2\leq(u_K-u_L)(S_k(u_K)-S_k(u_L))$ (S_k is nondecreasing and Lipschitz-continuous with 1 as Lipschitz constant) and $b_Ku_KS_k(u_K)\geq 0$ ($S_k(s)$ has the same sign as s), we get

$$\begin{aligned}\sum_{\sigma\in\mathcal{E}}\tau_\sigma(S_k(u_K)-S_k(u_L))^2 \leq{}& \sum_{K\in\mathcal{T}}\mathrm{meas}(K)f_KS_k(u_K)\\ &+\sum_{\sigma\in\mathcal{E}}m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(S_k(u_K)-S_k(u_L))\\ &+\sum_{\sigma\in\mathcal{E}}m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(S_k(u_K)-S_k(u_L)).\end{aligned} \tag{3.12}$$

By means of the Cauchy-Schwarz inequality, the discrete Poincaré inequality and (3.11), we have

$$\begin{aligned}\left|\sum_{K\in\mathcal{T}}\mathrm{meas}(K)f_KS_k(u_K)\right| &\leq \left(\sum_{K\in\mathcal{T}}\mathrm{meas}(K)f_K^2\right)^{1/2}\left(\sum_{K\in\mathcal{T}}\mathrm{meas}(K)(S_k(u_K))^2\right)^{1/2}\\ &\leq ||f||_{L^2(\Omega)}||S_k(u_{\mathcal{T}})||_{L^2(\Omega)}\\ &\leq \mathrm{diam}(\Omega)||S_k(u_{\mathcal{T}})||_{1,\mathcal{T}}.\end{aligned} \tag{3.13}$$

The Cauchy-Schwarz inequality, associated to (3.4) and (3.11), gives

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(S_k(u_K)-S_k(u_L))\\
&\le\left(\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)^2\right)^{1/2}\left(\sum_{\sigma\in\mathcal{E}}\tau(\sigma)(S_k(u_K)-S_k(u_L))^2\right)^{1/2}\\
&\le\sqrt{2d}||S_k(u_\mathcal{T})||_{1,\mathcal{T}}. \qquad (3.14)
\end{aligned}$$

We bound now the convection term, beginning with

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(S_k(u_K)-S_k(u_L))\\
&\qquad\le\left(\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)^2\right)^{1/2}||S_k(u_\mathcal{T})||_{1,\mathcal{T}}. \qquad (3.15)
\end{aligned}$$

Since $d_{K,\sigma}/d_\sigma\le 1$ and $d_{L,\sigma}/d_\sigma\le 1$, gathering by control volumes and using Hölder's inequality (with $p/2>1$ and $p/(p-2)$), we find

$$\begin{aligned}
\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)^2 &\le 2\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}^2u_{L,\sigma,+}^2+\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}^2u_{K,\sigma,+}^2\right)\\
&\le 2\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}v_{K,\sigma}^2u_{K,\sigma,+}^2\\
&\le 2\left(\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|v_{K,\sigma}|^p\right)^{\frac{2}{p}}\\
&\quad\times\left(\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|u_{K,\sigma,+}|^{\frac{2p}{p-2}}\right)^{\frac{p-2}{p}}. \qquad (3.16)
\end{aligned}$$

But, by Jensen's inequality,

$$m(\sigma)d_{K,\sigma}|v_{K,\sigma}|^p\le\frac{m(\sigma)d_{K,\sigma}}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}}|\mathbf{v}|^p=d\int_{\triangle_{K,\sigma}}|\mathbf{v}|^p$$

so that, since $\{\triangle_{K,\sigma}\,,\ K\in\mathcal{T}\,,\ \sigma\in\mathcal{E}_K\}$ is (up to a set of null Lebesgue measure) a partition of Ω,

$$\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|v_{K,\sigma}|^p\le d||\,|\mathbf{v}|\,||_{L^p(\Omega)}^p\le dM^p. \qquad (3.17)$$

On the other hand,

$$\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|u_{K,\sigma,+}|^{\frac{2p}{p-2}}=\sum_{K\in\mathcal{T}}|u_K|^{\frac{2p}{p-2}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d(K,\sigma)$$

where

- $d(K,\sigma) = d_{K,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ satisfies $v_{K,\sigma} \geq 0$ and $v_{L,\sigma} \geq 0$;
- $d(K,\sigma) = d_{K,\sigma} + d_{L,\sigma} = d_\sigma$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ satisfies $v_{K,\sigma} \geq 0$ and $v_{L,\sigma} < 0$;
- $d(K,\sigma) = d_{L,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ satisfies $v_{K,\sigma} < 0$ and $v_{L,\sigma} < 0$;
- $d(K,\sigma) = 0$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ satisfies $v_{K,\sigma} < 0$ and $v_{L,\sigma} \geq 0$;
- $d(K,\sigma) = d_{K,\sigma}$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$ satisfies $v_{K,\sigma} \geq 0$;
- $d(K,\sigma) = 0$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$ satisfies $v_{K,\sigma} < 0$.

In either case, we have $d(K,\sigma) \leq d_\sigma \leq \frac{d_{K,\sigma}}{\zeta}$, so that

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} |u_{K,\sigma,+}|^{\frac{2p}{p-2}} \leq \frac{1}{\zeta} \sum_{K\in\mathcal{T}} |u_K|^{\frac{2p}{p-2}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} \leq \frac{d}{\zeta} ||u_\mathcal{T}||^{\frac{2p}{p-2}}_{L^{\frac{2p}{p-2}}(\Omega)} \tag{3.18}$$

(we have used $\sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} = d\,\mathrm{meas}(K)$).

(3.16, 3.17) and (3.18) together give

$$\left(\sum_{\sigma\in\mathcal{E}} m(\sigma) d_\sigma \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right)^2 \right)^{1/2} \leq \frac{\sqrt{2} d^{\frac{1}{p}+\frac{p-2}{2p}}}{\zeta^{\frac{p-2}{2p}}} M ||u_\mathcal{T}||_{L^{\frac{2p}{p-2}}(\Omega)}. \tag{3.19}$$

Since $|u_\mathcal{T}| \leq k + |S_k(u_\mathcal{T})|$, (3.15) entails

$$\sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (S_k(u_K) - S_k(u_L)) \\ \leq C_1 k ||S_k(u_\mathcal{T})||_{1,\mathcal{T}} + C_1 ||S_k(u_\mathcal{T})||_{L^{\frac{2p}{p-2}}(\Omega)} ||S_k(u_\mathcal{T})||_{1,\mathcal{T}} \tag{3.20}$$

where C_1 only depends on (Ω, p, M, ζ) (a dependence on Ω takes into account a dependence on d).

But $p > d$, so that $\frac{2p}{p-2} < \frac{2d}{d-2}$. Let $q \in]\frac{2p}{p-2}, \frac{2d}{d-2}[$ (the choice of such a q only depends on (d,p)). Since $S_k(u_\mathcal{T}) = 0$ outside $E_k = \{x \in \Omega \mid |u_\mathcal{T}(x)| \geq k\}$, the Hölder inequality and the discrete Sobolev inequality give

$$||S_k(u_\mathcal{T})||_{L^{\frac{2p}{p-2}}(\Omega)} \leq \mathrm{meas}(E_k)^{\frac{p-2}{2p}-\frac{1}{q}} ||S_k(u_\mathcal{T})||_{L^q(\Omega)} \leq C_2 \mathrm{meas}(E_k)^{\frac{p-2}{2p}-\frac{1}{q}} ||S_k(u_\mathcal{T})||_{1,\mathcal{T}}$$

where C_2 only depends on (Ω, q, ζ) (*i.e.* on (Ω, p, ζ)). (3.20) leads then to

$$\sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (S_k(u_K) - S_k(u_L)) \\ \leq C_3 k ||S_k(u_\mathcal{T})||_{1,\mathcal{T}} + C_3 \mathrm{meas}(E_k)^{\frac{p-2}{2p}-\frac{1}{q}} ||S_k(u_\mathcal{T})||^2_{1,\mathcal{T}} \tag{3.21}$$

where C_3 only depends on (Ω, p, M, ζ).

Gathering (3.13, 3.14) and (3.21) in (3.12), we obtain

$$||S_k(u_\mathcal{T})||^2_{1,\mathcal{T}} \leq (\mathrm{diam}(\Omega) + \sqrt{2d} + C_3 k) ||S_k(u_\mathcal{T})||_{1,\mathcal{T}} + C_3 \mathrm{meas}(E_k)^{\frac{p-2}{2p}-\frac{1}{q}} ||S_k(u_\mathcal{T})||^2_{1,\mathcal{T}}. \tag{3.22}$$

But, by Tchebycheff's inequality, the discrete Poincaré inequality and Proposition 3.1, we have

$$\mathrm{meas}(E_k) = \mathrm{meas}(\{x \in \Omega \mid \ln(1+|u_\mathcal{T}(x)|) \geq \ln(1+k)\}) \leq \frac{||\ln(1+|u_\mathcal{T}|)||^2_{L^2(\Omega)}}{(\ln(1+k))^2} \leq \frac{C_4}{(\ln(1+k))^2},$$

where C_4 only depends on (Ω, p, M). Thus, since $\frac{p-2}{2p} - \frac{1}{q} > 0$, we can find k_0 only depending on (Ω, p, M, ζ) such that $C_3 \mathrm{meas}(E_{k_0})^{\frac{p-2}{2p}-\frac{1}{q}} < \frac{1}{2}$ and (3.22) allows to write

$$||S_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}} \leq 2(\mathrm{diam}(\Omega) + \sqrt{2d} + C_3 k_0) = C_5 \tag{3.23}$$

with C_5 only depending on (Ω, p, M, ζ).

Step 2: estimate on $T_{k_0}(u_\mathcal{T})$ and conclusion.

With the k_0 obtained in the previous step, using $\varphi_K = T_{k_0}(u_K)$ in (2.6), the fact that $(T_{k_0}(u_K) - T_{k_0}(u_L))^2 \leq (u_K - u_L)(T_{k_0}(u_K) - T_{k_0}(u_L))$, that $b_K u_K T_{k_0}(u_K) \geq 0$ and that $|T_{k_0}(u_K)| \leq k_0$, we find

$$\begin{aligned} ||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}}^2 \leq k_0 ||f||_{L^1(\Omega)} &+ \sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (T_{k_0}(u_K) - T_{k_0}(u_L)) \\ &+ \sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} G_{L,\sigma} \right) (T_{k_0}(u_K) - T_{k_0}(u_L)). \end{aligned} \tag{3.24}$$

The Cauchy-Schwarz inequality, (3.4) and (3.11) lead to

$$\sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} G_{L,\sigma} \right) (T_{k_0}(u_K) - T_{k_0}(u_L)) \leq \sqrt{2d} ||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}}. \tag{3.25}$$

Thanks to the Cauchy-Schwarz inequality and to (3.19), we also have

$$\begin{aligned} \sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (T_{k_0}(u_K) - T_{k_0}(u_L)) &\leq C_6 ||u_\mathcal{T}||_{L^{\frac{2p}{p-2}}(\Omega)} ||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}} \\ &\leq \left(C_7 + C_7 ||S_{k_0}(u_\mathcal{T})||_{L^{\frac{2p}{p-2}}(\Omega)} \right) ||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}} \end{aligned}$$

where C_6 and C_7 only depend on (Ω, p, M, ζ) (we have used $|u_\mathcal{T}| \leq k_0 + |S_{k_0}(u_\mathcal{T})|$). Thanks to the discrete Sobolev inequality (recall that $\frac{2p}{p-2} < \frac{2d}{d-2}$) and to (3.23), we deduce that there exists C_8 only depending on (Ω, p, M, ζ) such that

$$\sum_{\sigma\in\mathcal{E}} m(\sigma) \left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+} \right) (T_{k_0}(u_K) - T_{k_0}(u_L)) \leq C_8 ||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}}.$$

This inequality, injected in (3.24) together with (3.25), gives $||T_{k_0}(u_\mathcal{T})||_{1,\mathcal{T}} \leq C_9$ with C_9 only depending on (Ω, p, M, ζ).

Since $u_\mathcal{T} = T_{k_0}(u_\mathcal{T}) + S_{k_0}(u_\mathcal{T})$, we deduce that $||u_\mathcal{T}||_{1,\mathcal{T}} \leq C_5 + C_9$, which concludes this proof. □

4. Proof of the existence, uniqueness and convergence result

Proof of Theorem 2.1. The existence of a unique solution to (2.2–2.5) is an immediate consequence of the estimate of Theorem 3.1: indeed, if $f = G = 0$, then this theorem shows that any solution to (2.2–2.5) is null, that is to say that the square matrix defining this linear system is invertible.

Let us now prove the convergence result.

Since the solution to (1.3) is unique (see [3]), it is sufficient to prove that, for any sequence of admissible meshes $(\mathcal{T}_n)_{n\geq 1}$ such that $\mathrm{size}(\mathcal{T}_n)\to 0$ and $\mathrm{reg}(\mathcal{T}_n)\geq\alpha$, we can extract a subsequence (still denoted $(\mathcal{T}_n)_{n\geq 1}$) such that the solution $u_{\mathcal{T}_n}$ to (2.2–2.5) (with $\mathcal{T}_n$ instead of $\mathcal{T}$) converges to the solution of (1.3).

Take such a sequence $(\mathcal{T}_n)_{n\geq 1}$. Thanks to Theorem 3.1 and to item (iii) of Proposition 2.1, we see that, up to a subsequence, we can suppose that $u_{\mathcal{T}_n}\to u$ in $L^2(\Omega)$, for some $u\in H_0^1(\Omega)$; by the discrete Sobolev inequality, $(u_{\mathcal{T}_n})_{n\geq 1}$ is also bounded in $L^q(\Omega)$ for all $q<\frac{2d}{d-2}$, so that Vitali's Theorem gives the convergence of $(u_{\mathcal{T}_n})_{n\geq 1}$ to u in $L^q(\Omega)$ for all $q<\frac{2d}{d-2}$.

We are now going to prove that u is a solution to (1.3), which is enough, as noticed above, to conclude the proof of the theorem.

To simplify the notation, we forget the index n.

Of course, it is sufficient to prove that u satisfies the equation of (1.3) for all $\varphi\in C_c^\infty(\Omega)$. Take such a φ. Using (2.6) with $\varphi_K=\varphi(x_K)$, we have

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(x_K)-\varphi(x_L))+\sum_{K\in\mathcal{T}}\mathrm{meas}(K)b_Ku_K\varphi(x_K)\\
&=\sum_{K\in\mathcal{T}}\mathrm{meas}(K)f_K\varphi(x_K)+\sum_{\sigma\in\mathcal{E}}m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\\
&\quad+\sum_{\sigma\in\mathcal{E}}m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(x_K)-\varphi(x_L))
\end{aligned}\tag{4.1}$$

(with $\varphi(x_L)=0$ whenever $\sigma\in\mathcal{E}_{\mathrm{ext}}\cap\mathcal{E}_K$).

Step 1: convergence of the diffusion and the lower order terms.

The convergence proof in [5] immediately gives

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}}\tau_\sigma(u_K-u_L)(\varphi(x_K)-\varphi(x_L))\to\int_\Omega\nabla u\cdot\nabla\varphi\,,\quad\sum_{K\in\mathcal{T}}\mathrm{meas}(K)b_Ku_K\varphi(x_K)\to\int_\Omega bu\varphi\\
&\text{and}\quad\sum_{K\in\mathcal{T}}\mathrm{meas}(K)f_K\varphi(x_K)\to\int_\Omega f\varphi
\end{aligned}\tag{4.2}$$

as $\mathrm{size}(\mathcal{T})\to 0$ (in fact, to prove the convergence of $\sum_{K\in\mathcal{T}}\mathrm{meas}(K)b_Ku_K\varphi(x_K)$, we must slightly adapt the method of [5], since b is constant in this reference).

Step 2: convergence of the term involving G.

Let us study the convergence of $\sum_{\sigma\in\mathcal{E}}m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(x_K)-\varphi(x_L))$. We first notice that, for $\mathrm{size}(\mathcal{T})$ small enough, since φ has a compact support in Ω, this sum is reduced to $\mathcal{E}_{\mathrm{int}}$; we take, from now on, $\mathrm{size}(\mathcal{T})$ satisfying this property.

Fix $\varepsilon>0$ and take $H\in(C^1(\overline{\Omega}))^d$ such that $\||\,|G-H|\,\|_{L^1(\Omega)}\leq\varepsilon$; let, for $K\in\mathcal{T}$ and $\sigma\in\mathcal{E}_K$, $H_{K,\sigma}=\left(\frac{1}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}}H\right)\cdot\mathbf{n}_{K,\sigma}$.

By regularity of φ and gathering by control volumes, we write

$$\begin{aligned}
&\left| \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} G_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} G_{L,\sigma} \right) (\varphi(x_K) - \varphi(x_L)) \right. \\
&\qquad \left. - \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} H_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} H_{L,\sigma} \right) (\varphi(x_K) - \varphi(x_L)) \right| \\
&\le C_1 \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma) d_\sigma \left(\frac{d_{K,\sigma}}{d_\sigma} |G_{K,\sigma} - H_{K,\sigma}| + \frac{d_{L,\sigma}}{d_\sigma} |G_{L,\sigma} - H_{L,\sigma}| \right) \\
&\le C_1 \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma) d_{K,\sigma} |G_{K,\sigma} - H_{K,\sigma}| \\
&\le C_1 d \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \int_{\triangle_{K,\sigma}} |G - H| \le C_1 d\varepsilon
\end{aligned} \tag{4.3}$$

where C_1 only depends on φ.

By regularity of H and φ, we have

$$\begin{aligned}
&\left| \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma) \left(\frac{d_{K,\sigma}}{d_\sigma} H_{K,\sigma} - \frac{d_{L,\sigma}}{d_\sigma} H_{L,\sigma} \right) (\varphi(x_K) - \varphi(x_L)) \right. \\
&\qquad \left. - \sum_{\sigma\in\mathcal{E}_{\rm int}} \left(\frac{d_{K,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma - \frac{d_{L,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma \right) (\varphi(x_K) - \varphi(x_L)) \right| \\
&\qquad\qquad \le C_2 \sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma) d_\sigma \left(\frac{d_{K,\sigma}}{d_\sigma} \mathrm{size}(\mathcal{T}) + \frac{d_{L,\sigma}}{d_\sigma} \mathrm{size}(\mathcal{T}) \right) \le C_2\, d\, \mathrm{meas}(\Omega) \mathrm{size}(\mathcal{T})
\end{aligned} \tag{4.4}$$

where C_2 only depends on (H,φ).

Gathering by control volumes and noticing that $\mathbf{n}_{K,\sigma} = -\mathbf{n}_{L,\sigma}$ whenever $\sigma = K|L \in \mathcal{E}_{\rm int}$, we can moreover write

$$\begin{aligned}
\sum_{\sigma\in\mathcal{E}_{\rm int}} \left(\frac{d_{K,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma - \frac{d_{L,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma \right) (\varphi(x_K) - \varphi(x_L)) &= \sum_{K\in\mathcal{T}} \varphi(x_K) \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm int}} \int_\sigma H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma \\
&= \sum_{K\in\mathcal{T}} \varphi(x_K) \int_{\partial K\setminus\partial\Omega} H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma.
\end{aligned}$$

Since $\varphi = 0$ on the control volumes $K\in\mathcal{T}$ such that $\partial K\cap\partial\Omega \neq \emptyset$, we have in fact

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}_{\rm int}} \left(\frac{d_{K,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma - \frac{d_{L,\sigma}}{d_\sigma} \int_\sigma H\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma \right) (\varphi(x_K) - \varphi(x_L)) \\
&\quad = \sum_{K\in\mathcal{T}} \varphi(x_K) \int_{\partial K} H\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma \\
&\quad = \sum_{K\in\mathcal{T}} \varphi(x_K) \int_K \mathrm{div}(H) \overset{\mathrm{size}(\mathcal{T})\to 0}{\longrightarrow} \int_\Omega \varphi\,\mathrm{div}(H) = -\int_\Omega H\cdot\nabla\varphi,
\end{aligned} \tag{4.5}$$

the convergence being a consequence of the regularity of φ and H.

We also remark that

$$\left|\int_\Omega H\cdot\nabla\varphi-\int_\Omega G\cdot\nabla\varphi\right|\le C_3\varepsilon, \tag{4.6}$$

where C_3 only depends on φ.

Gathering (4.3–4.5) and (4.6), we deduce that

$$\limsup_{\mathrm{size}(\mathcal{T})\to 0}\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(x_K)-\varphi(x_L))-\left(-\int_\Omega G\cdot\nabla\varphi\right)\right|\le (C_1 d+C_3)\varepsilon$$

for all $\varepsilon>0$ and, since C_1 and C_3 only depend on φ, this gives

$$\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{K,\sigma}}{d_\sigma}G_{K,\sigma}-\frac{d_{L,\sigma}}{d_\sigma}G_{L,\sigma}\right)(\varphi(x_K)-\varphi(x_L))\to-\int_\Omega G\cdot\nabla\varphi \tag{4.7}$$

as $\mathrm{size}(\mathcal{T})\to 0$.

Step 3: convergence of the convective term.

It remains to study the convergence of the term in (4.1) coming from the convection, that is to say $\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))$ (the sum is reduced to $\mathcal{E}_{\rm int}$ because $\mathrm{size}(\mathcal{T})$ has been chosen small enough).

Take $\varepsilon>0$ and $\mathbf{w}\in(C^1(\overline{\Omega}))^d$ such that $||\,|\mathbf{v}-\mathbf{w}|\,||_{L^2(\Omega)}\le\varepsilon$. Let, if $K\in\mathcal{T}$ and $\sigma\in\mathcal{E}_K$, $\mathbf{w}_{K,\sigma}=\left(\frac{1}{\mathrm{meas}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}}\mathbf{w}\right)\cdot\mathbf{n}_{K,\sigma}$. We have, by regularity of φ,

$$\begin{aligned}
&\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right.\\
&\quad\left.-\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}w_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}w_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right|\\
&\le C_1\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}|v_{L,\sigma}-w_{L,\sigma}|\,|u_{L,\sigma,+}|+\frac{d_{K,\sigma}}{d_\sigma}|v_{K,\sigma}-w_{K,\sigma}|\,|u_{K,\sigma,+}|\right)\\
&\le C_1\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|v_{K,\sigma}-w_{K,\sigma}|\,|u_{K,\sigma,+}|\\
&\le C_1\left(\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}(v_{K,\sigma}-w_{K,\sigma})^2\right)^{1/2}\left(\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}(u_{K,\sigma,+})^2\right)^{1/2}
\end{aligned}$$

(C_1, which only depends on φ, is the same constant as before). The same way we have obtained (3.18), we can prove that

$$\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}(u_{K,\sigma,+})^2\le\frac{d}{\zeta}||u_\mathcal{T}||^2_{L^2(\Omega)}\le C_4$$

where C_4 only depends on $(\Omega, p, ||\,|\mathbf{v}|\,||_{L^p(\Omega)}, \zeta)$ (we use here Th. 3.1 and the discrete Poincaré inequality to obtain a bound on $u_\mathcal{T}$ in $L^2(\Omega)$). Moreover, by Jensen's inequality,

$$\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}(v_{K,\sigma}-w_{K,\sigma})^2 \le \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\frac{m(\sigma)d_{K,\sigma}}{\triangle_{K,\sigma}}\int_{\triangle_{K,\sigma}}|\mathbf{v}-\mathbf{w}|^2 = d||\,|\mathbf{v}-\mathbf{w}|\,||^2_{L^2(\Omega)} \le d\varepsilon^2.$$

Thus, we have

$$\begin{aligned}&\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}v_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}v_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right.\\&\quad\left.-\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}w_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}w_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right|\\&\le \varepsilon C_1\sqrt{C_4 d}.\end{aligned}\tag{4.8}$$

By regularity of $\mathbf{w}$ and φ, and gathering by control volumes, we find C_5 only depending on $\mathbf{w}$ and φ such that

$$\begin{aligned}&\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}w_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}w_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right.\\&\quad\left.-\sum_{\sigma\in\mathcal{E}_{\rm int}}\left(\frac{d_{L,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma\,u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma\,u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right|\\&\le C_5\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)d_\sigma\left(\frac{d_{L,\sigma}}{d_\sigma}\mathrm{size}(\mathcal{T})|u_{L,\sigma,+}|+\frac{d_{K,\sigma}}{d_\sigma}\mathrm{size}(\mathcal{T})|u_{K,\sigma,+}|\right)\\&\le C_5\mathrm{size}(\mathcal{T})\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|u_{K,\sigma,+}|.\end{aligned}$$

Once again we can prove, the same way we have obtained (3.18), that $\sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K} m(\sigma)d_{K,\sigma}|u_{K,\sigma,+}| \le \frac{d}{\zeta}||u_\mathcal{T}||_{L^1(\Omega)}$, which is bounded by C_6 only depending on $(\Omega, p, ||\,|\mathbf{v}|\,||_{L^p(\Omega)}, \zeta)$. Thus we get

$$\begin{aligned}&\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma}w_{L,\sigma}u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}w_{K,\sigma}u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right.\\&\quad\left.-\sum_{\sigma\in\mathcal{E}_{\rm int}}\left(\frac{d_{L,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma\,u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma\,u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L))\right|\\&\le C_5C_6\mathrm{size}(\mathcal{T}).\end{aligned}\tag{4.9}$$

Denoting by $\overline{w}_{K,\sigma}=\int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$, $\overline{u}_\sigma=\frac{d_{L,\sigma}u_{L,\sigma,+}+d_{K,\sigma}u_{K,\sigma,+}}{d_\sigma}$ and noticing that $\overline{w}_{K,\sigma}=-\overline{w}_{L,\sigma}$ whenever $\sigma=K|L\in\mathcal{E}_{\rm int}$, we can write

$$\begin{aligned}&\sum_{\sigma\in\mathcal{E}_{\rm int}}\left(\frac{d_{L,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma\,u_{L,\sigma,+}-\frac{d_{K,\sigma}}{d_\sigma}\int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma\,u_{K,\sigma,+}\right)(\varphi(x_K)-\varphi(x_L)) =\\&\qquad -\sum_{\sigma\in\mathcal{E}_{\rm int}}\overline{w}_{K,\sigma}\overline{u}_\sigma(\varphi(x_K)-\varphi(x_L)).\end{aligned}$$

Gathering by control volumes (and since $\overline{w}_{K,\sigma} = -\overline{w}_{L,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$), this gives

$$\sum_{\sigma\in\mathcal{E}_{\rm int}} \left(\frac{d_{L,\sigma}}{d_\sigma} \int_\sigma \mathbf{w}\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}\gamma\; u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma\; u_{K,\sigma,+} \right) (\varphi(x_K) - \varphi(x_L)) =$$
$$- \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm int}} \overline{w}_{K,\sigma}\overline{u}_\sigma\varphi(x_K) = - \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}\overline{u}_\sigma\varphi(x_K) \quad (4.10)$$

(recall that size($\mathcal{T}$) is small enough so that $\varphi(x_K) = 0$ whenever $\mathcal{E}_{\rm ext}\cap\mathcal{E}_K \neq \emptyset$).

The technique is then the same as in [5]: we decompose

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}\overline{u}_\sigma\varphi(x_K) = \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}(\overline{u}_\sigma - u_K)\varphi(x_K) + \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}u_K\varphi(x_K). \quad (4.11)$$

Since $\sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma} = \int_K \mathrm{div}(\mathbf{w})$, by convergence of $u_\mathcal{T}$ to u in $L^2(\Omega)$ and by regularity of φ, we have

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}u_K\varphi(x_K) = \sum_{K\in\mathcal{T}} u_K\varphi(x_K) \int_K \mathrm{div}(\mathbf{w}) \longrightarrow \int_\Omega u\varphi\,\mathrm{div}(\mathbf{w}) \quad (4.12)$$

as size($\mathcal{T}$) $\to 0$. We also have

$$\left| \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}(\overline{u}_\sigma - u_K)\varphi(x_K) - \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} (\overline{u}_\sigma - u_K) \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi\,\mathrm{d}\gamma \right|$$
$$\leq C_1\mathrm{size}(\mathcal{T})||\,|\mathbf{w}|\,||_{C(\overline{\Omega})} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)|\overline{u}_\sigma - u_K|.$$

But $\overline{u}_\sigma$ is a convex combinaison of (u_K, u_L) if $\sigma = K|L \in \mathcal{E}_{\rm int}$, and $\overline{u}_\sigma \in \{0, u_K\}$ if $\sigma \in \mathcal{E}_{\rm ext}\cap\mathcal{E}_K$, so that, in either case, $|\overline{u}_\sigma - u_K| \leq D_\sigma u_\mathcal{T}$ and

$$\left| \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}(\overline{u}_\sigma - u_K)\varphi(x_K) - \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} (\overline{u}_\sigma - u_K) \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi\,\mathrm{d}\gamma \right|$$
$$\leq C_1\mathrm{size}(\mathcal{T})||\,|\mathbf{w}|\,||_{C(\overline{\Omega})} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} m(\sigma)D_\sigma u_\mathcal{T}$$
$$\leq 2C_1\mathrm{size}(\mathcal{T})||\,|\mathbf{w}|\,||_{C(\overline{\Omega})} \sum_{\sigma\in\mathcal{E}} m(\sigma)D_\sigma u_\mathcal{T}$$
$$\leq 2C_1\mathrm{size}(\mathcal{T})||\,|\mathbf{w}|\,||_{C(\overline{\Omega})} \left(\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma \right)^{1/2} ||u_\mathcal{T}||_{1,\mathcal{T}}.$$

$||u_\mathcal{T}||_{1,\mathcal{T}}$ being bounded as size($\mathcal{T}$) $\to 0$ and $\sum_{\sigma\in\mathcal{E}} m(\sigma)d_\sigma$ being constant (it is $d\,\mathrm{meas}(\Omega)$), we deduce that

$$\lim_{\mathrm{size}(\mathcal{T})\to 0} \left(\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}(\overline{u}_\sigma - u_K)\varphi(x_K) - \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} (\overline{u}_\sigma - u_K) \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi\,\mathrm{d}\gamma \right) = 0. \quad (4.13)$$

We have, gathering by edges and since $\varphi = 0$ on σ whenever $\sigma \in \mathcal{E}_{\rm ext}$,

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{u}_\sigma \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi\,\mathrm{d}\gamma = \sum_{\sigma\in\mathcal{E}_{\rm int}} \overline{u}_\sigma \left(\int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi\,\mathrm{d}\gamma + \int_\sigma \mathbf{w}\cdot\mathbf{n}_{L,\sigma}\varphi\,\mathrm{d}\gamma \right) = 0$$

since $\mathbf{n}_{K,\sigma} = -\mathbf{n}_{L,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$. Moreover,

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} u_K \int_\sigma \mathbf{w}\cdot\mathbf{n}_{K,\sigma}\varphi \,\mathrm{d}\gamma = \sum_{K\in\mathcal{T}} u_K \int_K \mathrm{div}(\varphi\mathbf{w}) \longrightarrow \int_\Omega u\,\mathrm{div}(\varphi\mathbf{w})$$

as size$(\mathcal{T}) \to 0$. Thus, (4.13) implies

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \overline{w}_{K,\sigma}(\overline{u}_\sigma - u_K)\varphi(x_K) \longrightarrow -\int_\Omega u\,\mathrm{div}(\varphi\mathbf{w})$$

as size$(\mathcal{T}) \to 0$. Together with (4.9–4.11) and (4.12), this gives

$$\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma} w_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} w_{K,\sigma} u_{K,\sigma,+}\right)(\varphi(x_K) - \varphi(x_L)) \longrightarrow -\int_\Omega u\varphi\,\mathrm{div}(\mathbf{w}) + \int_\Omega u\,\mathrm{div}(\varphi\mathbf{w}) = \int_\Omega u\mathbf{w}\cdot\nabla\varphi \quad (4.14)$$

as size$(\mathcal{T}) \to 0$.

By noticing that

$$\left|\int_\Omega u\mathbf{w}\cdot\nabla\varphi - \int_\Omega u\mathbf{v}\cdot\nabla\varphi\right| \le ||u||_{L^2(\Omega)}||\,|\mathbf{v}-\mathbf{w}|\,||_{L^2(\Omega)}||\,|\nabla\varphi|\,||_{L^\infty(\Omega)} \le C_7\varepsilon$$

where C_7 only depends on u and φ, (4.8) and (4.14) allow to write

$$\limsup_{\mathrm{size}(\mathcal{T})\to 0}\left|\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+}\right)(\varphi(x_K) - \varphi(x_L)) - \int_\Omega u\mathbf{v}\cdot\nabla\varphi\right| \le (C_1\sqrt{C_4 d} + C_7)\varepsilon$$

for all $\varepsilon > 0$ and with C_1, C_4 and C_7 not depending on ε, that is to say

$$\sum_{\sigma\in\mathcal{E}_{\rm int}} m(\sigma)\left(\frac{d_{L,\sigma}}{d_\sigma} v_{L,\sigma} u_{L,\sigma,+} - \frac{d_{K,\sigma}}{d_\sigma} v_{K,\sigma} u_{K,\sigma,+}\right)(\varphi(x_K) - \varphi(x_L)) \longrightarrow \int_\Omega u\mathbf{v}\cdot\nabla\varphi \quad (4.15)$$

as size$(\mathcal{T}) \to 0$.

Gathering (4.2, 4.7) and (4.15) in (4.1), we see that u satisfies the equation of (1.3). □

5. Another scheme

The scheme of Section 2 is based on a discretization of (1.1) that brings in approximate values of $\int_\sigma \mathbf{v}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$ and $\int_\sigma G\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$ based on the values of $\mathbf{v}$ and G on a subset of K (the "half-diamond"). The choice of such approximate values seems to be quite adapted when there is a link between the mesh and $\mathbf{v}$ or G: for example, if $\mathbf{v}$ or G is constant on each side of an hyperplane and if we take meshes such that each control volume is on one side of this hyperplane.

But when there is no relation between $\mathbf{v}$ or G and the mesh, the reasons for using the values of $\mathbf{v}$ or G only on K to approximate $\int_\sigma \mathbf{v}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$ or $\int_\sigma G\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$ are not so clear: we could approximate $\mathbf{v}$ or G on σ by some quantity $\mathbf{v}_\sigma$ or G_σ, and then consider $m(\sigma)\mathbf{v}_\sigma\cdot\mathbf{n}_{K,\sigma}$ or $m(\sigma)G_\sigma\cdot\mathbf{n}_{K,\sigma}$ as a coherent approximate value of $\int_\sigma \mathbf{v}\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$ or $\int_\sigma G\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma$. This is what the following scheme does.

Let $\mathcal{T}$ be an admissible mesh. If $\sigma = K|L \in \mathcal{E}_{\rm int}$, we define the "full-diamond" around σ by $\diamondsuit_\sigma = \diamondsuit_{K,\sigma} \cup \diamondsuit_{L,\sigma}$; if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$, the "full-diamond" around σ is simply $\diamondsuit_\sigma = \diamondsuit_{K,\sigma}$. We let then, for $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}$,

$$\mathbf{v}_\sigma = \frac{1}{{\rm meas}(\diamondsuit_\sigma)} \int_{\diamondsuit_\sigma} \mathbf{v}\,, \quad b_K = \frac{1}{{\rm meas}(K)} \int_K b\,,$$
$$f_K = \frac{1}{{\rm meas}(K)} \int_K f \quad \text{and} \quad G_\sigma = \frac{1}{{\rm meas}(\diamondsuit_\sigma)} \int_{\diamondsuit_\sigma} G.$$

The new scheme for (1.1) is

$$\forall K \in \mathcal{T}\,,\ \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} + \sum_{\sigma\in\mathcal{E}_K} m(\sigma)\mathbf{v}_\sigma\cdot\mathbf{n}_{K,\sigma}u_{\sigma,+} + {\rm meas}(K)b_K u_K = {\rm meas}(K)f_K + \sum_{\sigma\in\mathcal{E}_K} m(\sigma)G_\sigma\cdot\mathbf{n}_{K,\sigma}\,, \quad (5.1)$$

$$\begin{aligned} &\forall K \in \mathcal{T}\,, \forall \sigma = K|L \in \mathcal{E}_K \cap \mathcal{E}_{\rm int}\,,\ F_{K,\sigma} = \tfrac{m(\sigma)}{d_\sigma}(u_K - u_L)\,,\\ &\forall K \in \mathcal{T}\,, \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext}\,, \qquad F_{K,\sigma} = \tfrac{m(\sigma)}{d_\sigma}u_K\,, \end{aligned} \quad (5.2)$$

$$\begin{aligned} &\forall \sigma = K|L \in \mathcal{E}_{\rm int}\,,\ u_{\sigma,+} = u_K \ \text{ if } \ \mathbf{v}_\sigma\cdot\mathbf{n}_{K,\sigma} \geq 0\,,\ u_{\sigma,+} = u_L \ \text{ otherwise,}\\ &\forall \sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K\,, \quad u_{\sigma,+} = u_K \ \text{ if } \ \mathbf{v}_\sigma\cdot\mathbf{n}_{K,\sigma} \geq 0\,,\ u_{\sigma,+} = 0 \ \text{ otherwise.} \end{aligned} \quad (5.3)$$

In fact, we can remark that (5.1–5.3) is exactly (2.2–2.5), provided that we define $v_{K,\sigma} = \mathbf{v}_\sigma\cdot\mathbf{n}_{K,\sigma}$, $G_{K,\sigma} = G_\sigma\cdot\mathbf{n}_{K,\sigma}$ and let $u_{K,\sigma,+} = u_{\sigma,+}$ (for all $K \in \mathcal{T}$ and all $\sigma \in \mathcal{E}_K$). Indeed, in this case, if $\sigma = K|L \in \mathcal{E}_{\rm int}$, we have $v_{K,\sigma} = -v_{L,\sigma}$, so that (5.3) is equivalent to (2.5) (with the notation $u_{K,\sigma,+} = u_{\sigma,+}$), and $G_{K,\sigma} = -G_{L,\sigma}$, so that (2.4) comes down to $F_{K,\sigma} = -F_{L,\sigma}$ (or $u_\sigma = 0$ if $\sigma \in \mathcal{E}_{\rm ext}$) which, associated to (2.3), is equivalent to (5.2).

Thus, we easily see that the preceding techniques to obtain *a priori* estimates on the solutions to (2.2–2.5) give us estimates on the solutions to (5.1–5.3), which proves the existence and uniqueness of the solution to this problem. The convergence proof also works as before, and we deduce that, if $\alpha > 0$ is fixed and $u_\mathcal{T}$ denotes the solution to (5.1–5.3), then $u_\mathcal{T}$ converges in $L^q(\Omega)$, for all $q < \frac{2d}{2-d}$, to the unique solution of (1.3), as ${\rm size}(\mathcal{T}) \to 0$ with ${\rm reg}(\mathcal{T}) \geq \alpha$.

References

[1] R.A. Adams, *Sobolev Spaces*. Academic Press, New York (1975).

[2] Y. Coudière, J.P. Vila and P. Villedieu, Convergence rate of a finite volume scheme for a two dimensional convection diffusion problem. *ESAIM: M2AN* **33** (1999) 493–516.

[3] J. Droniou, Non-coercive linear elliptic problems. *Potential Anal.* **17** (2002) 181–203.

[4] J. Droniou, Ph.D. thesis, CMI, Université de Provence.

[5] R. Eymard, T. Gallouët and R. Herbin, Finite Volume Methods, in *Handbook of Numerical Analysis*, Vol. VII, P.G. Ciarlet and J.L. Lions Eds., North-Holland, Amsterdam (1991) 713–1020.

[6] R. Eymard, T. Gallouët and R. Herbin, Convergence of finite volume approximations to the solutions of semilinear convection diffusion reaction equations. *Numer. Math.* **82** (1999) 91–116.

[7] J.M. Fiard and R. Herbin, Comparison between finite volume finite element methods for the numerical simulation of an elliptic problem arising in electrochemical engineering. *Comput. Methods Appl. Mech. Engrg.* **115** (1994) 315–338.

[8] P.A. Forsyth and P.H. Sammon, Quadratic convergence for cell-centered grids. *Appl. Numer. Math.* **4** (1988) 377–394.

[9] T. Gallouët, R. Herbin and M.H. Vignal, Error estimate for the approximate finite volume solutions of convection diffusion equations with Dirichlet, Neumann or Fourier boundary conditions. *SIAM J. Numer. Anal.* (2000).