Abstract In this work, we apply an iterative energy method à la de Giorgi in order to establish $L^\infty$ bounds for numerical solutions of noncoercive convection-diffusion equations with mixed Dirichlet-Neumann boundary conditions.

Key words: finite volume schemes, uniform bounds, noncoercive elliptic equations

MSC (2010): 65M08, 35B40.

1 Introduction

The continuous problem. Let $\Omega$ be an open bounded polygonal domain of $\mathbb{R}^p$ with $p = 2$ or 3. We denote by $m(\cdot)$ both the Lebesgue and $p-1$ dimensional Hausdorff measure. We assume that $\partial \Omega = \Gamma_D \cup \Gamma_N$ with $\Gamma_D \cap \Gamma_N = \emptyset$ and $m(\Gamma_D) > 0$ and we denote by $n$ the exterior normal to $\partial \Omega$. Let $U \in C(\bar{\Omega})^2$ be a velocity field, $b \in L^\infty(\Omega)$ assumed to be nonnegative, $f \in L^\infty(\Omega)$ a source term and $v^D \in L^\infty(\Gamma_D)$ a boundary condition.

We consider the following convection-diffusion equation with mixed boundary conditions:

\[
\begin{align*}
\text{div}(-\nabla v + Uv) + bv &= f & \text{in } \Omega, \quad (1a) \\
(-\nabla v + Uv) \cdot n &= 0 & \text{on } \Gamma_N, \quad (1b) \\
v &= v^D & \text{on } \Gamma_D. \quad (1c)
\end{align*}
\]
This noncoercive elliptic linear problem has been widely studied by Droniou and coauthors, even with less regularity on the data, see for instance [2, 4, 3, 5]. Nevertheless, up to our knowledge, the derivation of explicit $L^\infty$ bounds on numerical solutions has not been done in the literature.

The numerical scheme. The mesh of the domain $\Omega$ is denoted by $\mathcal{M} = (\mathcal{T}, \mathcal{E}, \mathcal{P})$ and classically given by: $\mathcal{T}$, a set of open polygonal or polyhedral control volumes; $\mathcal{E}$, a set of edges or faces; $\mathcal{P} = (x_K)_{K \in \mathcal{T}}$ a set of points. In the following, we also use the denomination “edge” for a face in dimension 3. As we deal with a Two-Point Flux Approximation (TPFA) of convection-diffusion equations, we assume that the mesh is admissible in the sense of [6] (Definition 9.1).

We distinguish in $\mathcal{E}$ the interior edges, $\sigma = KL$, from the exterior edges: $\mathcal{E} = \mathcal{E}_{\text{int}} \cup \mathcal{E}_{\text{ext}}$. Among the exterior edges, we distinguish the edges included in $\Gamma^N$: $\mathcal{E}_{\text{ext}}^N = \mathcal{E}_{\text{ext}}^D \cup \mathcal{E}_{\text{ext}}^N$. For each edge $\sigma \in \mathcal{E}$, we pick one cell in the non empty set $\{K : \sigma \in \mathcal{E}_K\}$ and denote it by $K_\sigma$. In the case of an interior edge $\sigma = KL$, $K_\sigma$ is either $K$ or $L$.

Let $d(\cdot, \cdot)$ denote the Euclidean distance. For all edges $\sigma \in \mathcal{E}$, we set $d_{\sigma} = d(x_K, x_L)$ if $\sigma = KL \in \mathcal{E}_{\text{int}}$ and $d_{\sigma} = d(x_K, \sigma)$ if $\sigma \in \mathcal{E}_{\text{ext}}$ with $\sigma \in \mathcal{E}_K$ and the transmissibility coefficient is defined by $\tau_{\sigma} = m(\sigma)/d_{\sigma}$, for all $\sigma \in \mathcal{E}$. We also denote by $n_{K,\sigma}$ the normal to $\sigma \in \mathcal{E}_K$ outward $K$. We assume that the mesh satisfies the regularity constraint:

$$\exists \xi > 0 \text{ such that } d(x_K, \sigma) \geq \xi d_{\sigma}, \quad \forall K \in \mathcal{T}, \forall \sigma \in \mathcal{E}_K. \quad (2)$$

As a consequence, we obtain that

$$\sum_{\sigma \in \mathcal{E}_K} m(\sigma)d_{\sigma} \leq \frac{p}{\xi}m(K) \quad \forall K \in \mathcal{T}. \quad (3)$$

The size of the mesh is defined by $h = \max \{\text{diam}(K) : K \in \mathcal{T}\}$.

Let us define

$$f_K = \frac{1}{m(K)} \int_K f, \quad b_K = \frac{1}{m(K)} \int_K b \quad \forall K \in \mathcal{T},$$

$$U_{K,\sigma} = \frac{1}{m(\sigma)} \int_{\sigma} U \cdot n_{K,\sigma}, \quad \forall K \in \mathcal{T}, \forall \sigma \in \mathcal{E}_K,$$

$$v_{\sigma}^D = \frac{1}{m(\sigma)} \int_{\sigma} v^D, \quad \forall \sigma \in \mathcal{E}_D.$$

Given a Lipschitz-continuous function on $\mathbb{R}$ which satisfies

$$B(0) = 1, \quad B(s) > 0 \quad \text{and} \quad B(s) - B(-s) = -s \quad \forall s \in \mathbb{R}, \quad (4)$$

we consider the B-scheme defined by
In this section, we consider the particular case where the source $f$ is non-negative and the boundary condition $v^D$ is non-negative and bounded by 1.
Let us start with some notations. Given $m \geq 1$, we denote the $m$-th truncation threshold by
\[
C_m = 2(1 - 2^{-m}),
\]
Then, we introduce the $m$-th energy
\[
E_m(v) = \sum_{\sigma \in \mathcal{E}_\text{int} \cup \partial \Omega} \tau_{\sigma} \left[ \log(1 + (v_{K,\sigma} - C_m)^+) - \log(1 + (v_K - C_m)^+) \right]^2. \tag{9}
\]
When there is no ambiguity we write $E_m = E_m(v)$. The first proposition is a fundamental estimate of the energy.

**Proposition 1.** Assume that $f_K \geq 0$ for all $K \in \mathcal{I}$ and $v^D_{\sigma} \in [0,1]$ for all $\sigma \in \mathcal{E}^D$, so that the solution $v$ to (5)-(8) satisfies $v_K \geq 0$ for all $K \in \mathcal{I}$. Then one has for all $m \geq 1$
\[
E_m \leq \frac{4p}{p_U^2} (\|U\|_{L^\infty}^2 + \|f\|_{L^\infty}) \sum_{K \in \mathcal{I}} m(K), \tag{10}
\]
where $\beta_U := \inf_{x \in [-\|U\|_{L^\infty}, \|U\|_{L^\infty}]} B(\text{diam}(\Omega) x)$ (because of (4), $\beta_U \in (0,1)$).

**Proof.** In order to shorten some expressions hereafter, let us introduce $w^m_K = v_K - C_m$ for all $K \in \mathcal{I}$ and $w^{m,D}_{\sigma} = v^D_{\sigma} - C_m$ for all $\sigma \in \mathcal{E}^D$. Let us note that we identify $w^m = (w^m_K)_{K \in \mathcal{I}}$ and the associate piecewise constant function. Therefore, we can write
\[
m(\{w^m > 0\}) = \sum_{w^m_K > 0} m(K).
\]
First, observe that $E_m$ is the discrete counterpart of
\[
\int_{\Omega} |\nabla \log(1+w^m)|^2 1_{\{w^m > 0\}} = \int_{\Omega} \nabla w^m \cdot \nabla w^m (1+w^m)^{-2} 1_{\{w^m > 0\}}, \text{ with } w^m = v - C_m,
\]
where $1_A$ is the indicator function of $A$. Let us define $\varphi : s \mapsto s/(1 + s) 1_{\{s \geq 0\}}$, which satisfies $\varphi'(s) = 1/(1 + s)^2 1_{\{s \geq 0\}}$ and let us introduce $F_m$ another discrete counterpart of the preceding quantity
\[
F_m = \sum_{\sigma \in \mathcal{E}_\text{int} \cup \partial \Omega} \tau_{\sigma} \left( (w^m_{K,\sigma})^+ - (w^m_K)^+ \right) (\varphi(w^m_{K,\sigma}) - \varphi(w^m_K)).
\]
It is clear that $E_m \leq F_m$ for all $m \geq 1$, as for all $x, y \in \mathbb{R}$ we have
\[
(\log(1 + x^+) - \log(1 + y^+))^2 \leq (x^+ - y^+) (\varphi(x) - \varphi(y)).
\]
Let us now multiply the scheme (5) by $\varphi(w^m_K)$ and sum over $K \in \mathcal{I}$. Due to the non-negativity of $b$ and $v$, we obtain, after a discrete integration by parts,
\[
\sum_{\sigma \in \mathcal{E}_\text{int} \cup \partial \Omega} \mathcal{F}_{K,\sigma} (\varphi(w^m_K) - \varphi(w^m_{K,\sigma})) \leq \sum_{K \in \mathcal{I}} m(K) f_K \varphi(w^m_K).
\]
Using that $\varphi$ is bounded by 1 and vanishes on $\mathbb{R}_-$, we deduce that
\[
\sum_{\sigma \in \partial_{int} \cup \partial^D} \mathcal{T}_{K, \sigma}(\varphi(w^m_K) - \varphi(w^m_{K, \sigma})) \leq \|f\|_{L^\infty} \mathcal{m}\{|w^m > 0\}). \tag{11}
\]

We focus now on the left-hand-side of (11). Due to (7) and the definition of $w^m_{K, \sigma}$, we can rewrite $\mathcal{T}_{K, \sigma}$ as
\[
\mathcal{T}_{K, \sigma} = \tau_\sigma B((U_{K, \sigma}|d\sigma)(w^m_K - w^m_{K, \sigma}) + m(\sigma) \left( U^+_{K, \sigma}(w^m_K + C_m) - U^-_{K, \sigma}(w^m_{K, \sigma} + C_m) \right).
\]

Observe that since $\varphi$ is a non-decreasing function, one has
\[
(x - y)(\varphi(x) - \varphi(y)) \geq (x^+ - y^+)(\varphi(x) - \varphi(y)), \quad \forall x, y \in \mathbb{R}.
\]

Therefore, using the definition of $\beta_U$ we obtain that
\[
\sum_{\sigma \in \partial_{int} \cup \partial^D} \mathcal{T}_{K, \sigma}(\varphi(w^m_K) - \varphi(w^m_{K, \sigma})) \geq \beta_U F_m - G_m, \tag{12}
\]

with
\[
G_m = -\sum_{\sigma \in \partial_{int} \cup \partial^D} m(\sigma) \left( U^+_{K, \sigma}(w^m_K + C_m) - U^-_{K, \sigma}(w^m_{K, \sigma} + C_m) \right) (\varphi(w^m_K) - \varphi(w^m_{K, \sigma})).
\]

For an interior edge, $w^m_K$ and $w^m_{K, \sigma}$ play a symmetric role in the preceding sum. As $w^m_{\sigma} \leq 0$ for all $\sigma \in \partial^D$ and $\varphi$ vanishes on $\mathbb{R}_-$, we can always assume that $w^m_K \geq w^m_{K, \sigma}$ and an edge has a contribution in the sum if at least $w^m_K > 0$. Then, under these assumptions one has
\[
-m(\sigma) \left( U^+_{K, \sigma}(w^m_K + C_m) - U^-_{K, \sigma}(w^m_{K, \sigma} + C_m) \right) (\varphi(w^m_K) - \varphi(w^m_{K, \sigma})) \leq \|U\|_{L^\infty} m(\sigma)(w^m_K + C_m)(\varphi(w^m_K) - \varphi(w^m_{K, \sigma})).
\]

But, $w^m_{K, \sigma} + C_m \leq 2(1 + (w^m_{K, \sigma})^+)$ and applying the definition of $\varphi$, we get
\[
(w^m_{K, \sigma} + C_m)(\varphi(w^m_K) - \varphi(w^m_{K, \sigma})) \leq 2 \frac{(w^m_K)^+ - (w^m_{K, \sigma})^+}{1 + (w^m_K)^+} \leq 2 \frac{(w^m_K)^+ - (w^m_{K, \sigma})^+}{\sqrt{1 + (w^m_K)^+} \sqrt{1 + (w^m_{K, \sigma})^+}}.
\]

Therefore,
\[
G_m \leq 2\|U\|_{L^\infty} \sum_{\sigma \in \partial_{int} \cup \partial^D} m(\sigma) \frac{|(w^m_K)^+ - (w^m_{K, \sigma})^+|}{\sqrt{1 + (w^m_K)^+} \sqrt{1 + (w^m_{K, \sigma})^+}}.
\]
We apply now Cauchy-Schwarz inequality in order to get

\[ G_m \leq 2\|U\|_{L^\infty(\mathcal{F}_m)}^{1/2} \left( \sum_{\sigma \in \mathcal{E}^{ip}} m(\sigma) d_\sigma \right)^{1/2}, \]  

(13)

where \( \mathcal{E}^{ip} \) is the set of interior and Dirichlet boundary edges on which \((w^m_K) - (w^m_{K,\sigma}) \neq 0 \). It appears that, due to (3),

\[ \sum_{\sigma \in \mathcal{E}^{sp}_m} m(\sigma) d_\sigma \leq \sum_{K \in \mathcal{T} \cap \{w^m_K > 0\}} \sum_{\sigma \in \mathcal{E}^{sp}_K} m(\sigma) d_\sigma \leq \frac{P}{\xi} m(\{w^m > 0\}). \]  

(14)

We deduce from (11), (12), (13) and (14) that

\[ \beta_U F_m \leq 2\|U\|_{L^\infty(\mathcal{F}_m)}^{1/2} \left( \frac{P}{\xi} m(\{w^m > 0\}) \right)^{1/2} + \|f\|_{L^\infty(\{w^m > 0\})}, \]

which yields (10) using Young’s inequality and the bounds \( E_m \leq F_m \) and \( \beta_U \leq 1 \).

Before stating the main result of the section, we need a technical lemma.

**Lemma 1.** Let \((u_n)_{n \in \mathbb{N}}\) be a sequence of non-negative real numbers and let \(K, \rho > 0\) and \(\alpha > 1\). Then if for all \(n \in \mathbb{N}\)

\[ u_{n+1} \leq K \rho^n u_n^\alpha, \]

one has

\[ 0 \leq u_n \leq \left( u_0 \rho^{\frac{1}{\alpha-1}} K^{\frac{1}{\alpha}} \right)^{\alpha^n} \rho^{-\frac{\alpha(n-1)+1}{(\alpha-1)^2}} K^{-\frac{1}{\alpha}} \]

for all \(n \in \mathbb{N}\) and the bound is optimal. In particular, if \(u_0 \leq \rho^{\frac{1}{\alpha-1}} K^{\frac{1}{\alpha}}\), then

\[ \lim_{n \to \infty} u_n = 0. \]

**Proof.** Just observe that the sequence \(v_n = u_n \rho^{\frac{\alpha(n-1)+1}{(\alpha-1)^2}} K^{\frac{1}{\alpha}}\) satisfies \(0 \leq v_{n+1} \leq v_n^\alpha\)

for all \(n \geq 0\) which directly yields the result.

**Proposition 2.** Assume that \(f_K \geq 0\) for all \(K \in \mathcal{T}\) and \(v^{D}_\sigma \in [0, 1]\) for all \(\sigma \in \mathcal{E}^{D}\), so that \(v_K \geq 0\) for all \(K \in \mathcal{T}\). Then, there exists \(\eta > 0\) depending only on \(\Omega, p\) and \(\xi\) such that one has the implication

\[ E_1 \leq \eta \frac{\beta_U^4}{\left( \|U\|_{L^\infty(\mathcal{F}_m)} + \|f\|_{L^\infty(\mathcal{F}_m)} \right)^2} \Rightarrow (v_K \leq 2, \forall K \in \mathcal{T}). \]

(15)

**Proof.** The proof consists in establishing an induction property on \(E_m\) which guarantees that if \(E_1\) is small enough then \(\lim E_m = 0\). Then, as \(\lim C_m = 2\) and thanks to the discrete Poincaré inequality, we deduce that

\[ \sum_{K \in \mathcal{T}} m(K) \left( \log(1 + (v_K - 2)^+) \right)^2 = 0, \]
which implies \( v_K \leq 2 \) for all \( K \in \mathcal{F} \).

For establishing the induction, first observe that as \( C_m = C_{m-1} + 2^{-m+1} \), for any \( q > 0 \) we have:

\[
1_{\{w^m > 0\}} \leq \left( \frac{1}{(\log(1 + 2^{-m+1}))^q} \right)^q 1_{\{w^{m-1} > 0\}}.
\]

and thus

\[
m(\{w^m > 0\}) \leq \frac{1}{(\log(1 + 2^{-m+1}))^q} \sum_{K \in \mathcal{F}} m(K) \left( \log(1 + (w^{m-1}_K)^+) \right)^q.
\]

We may choose for instance \( q = 3 \) and apply a discrete Poincaré-Sobolev inequality (whose constant \( C_{\Omega, p} \) depends only on \( \Omega \) and \( p \)), which leads to

\[
m(\{w^m > 0\}) \leq \frac{1}{(\log(1 + 2^{-m+1}))^3} \left( \frac{C_{\Omega}}{\xi^{3/2}} \right)^{3/2} E_{m-1}^{3/2}.
\]

Noticing that for \( x \in [0, 1] \), \((\log(1 + x))^3 \geq (2)^3 x^3\), we deduce from (10) and (17) that

\[
E_m \leq \frac{4}{\beta^3} \left( \| U \|_{L^\infty}^2 + \| f \|_{L^\infty} \right) \left( \frac{C_{\Omega}}{\xi^{3/2}} \right)^{3/2} \frac{\beta^{3/2}}{8} E_{m-1}^{3/2}.
\]

Thus the sequence \((E_m)_{m \geq 0}\) satisfies the hypothesis of Lemma \([1]\) with \( \alpha = 3/2 \) and \( K \) proportional to \((\| U \|_{L^\infty}^2 + \| f \|_{L^\infty})/\beta^3\). We deduce the upper bound for \( E_1 \) under which \( \lim E_m = 0 \).

Remark: The arguments developed in this section still hold, up to minor adaptation, for \( f \in L^r(\Omega) \) with \( r > p/2 \).

3 Proof of Theorem \([1]\)

First observe that if one replaces the data \( f \) and \( v^D \) by either \( f^+ \) and \((v^D)^+ \), or \( f^- \) and \((v^D)^- \), in the scheme (5)-(6), then the corresponding solutions, say respectively \( P = (p_K)_{K \in \mathcal{F}} \) and \( N = (N_K)_{K \in \mathcal{F}} \), are non-negative and such that \( v = P - N \) is the solution to (5)-(6) in the original framework.

From there let us show that there is \( \overline{M} > V^D := \max(||(v^D)^+||_{L^\infty}, 1) \) such that for all \( K \in \mathcal{F} \) one has \( 0 \leq p_K \leq \overline{M} \). The bound for \( N \), which is denoted by \( \underline{M} \), can be obtained in the same way.

Let \( M > V^D \). First observe that \( P^M := P/M \) satisfies the scheme (5)-(6) where the source term and boundary data have been replaced by \( f^+/M \) and \((v^D)^+/M \) respectively. Moreover, one can apply Proposition \([1]\) which yields

\[
E_1(P^M) \leq \frac{4p}{\beta^3} \left( \| U \|_{L^\infty}^2 + \frac{\| f^+ \|_{L^\infty}}{M} \right) m(\{P^M > 1\}).
\]
Now observe that $P = M P^M = V^D _+ P^D$. Therefore,

$$E_1(P^M) \leq \frac{4p}{k^2} \left( \|U\|_{L^\infty} \sum_{K \in \mathcal{T}} m(K) \frac{\log(1 + (P^D_K - 1)^2)}{\log(M/V^D_+)^2} + \frac{\|f^+\|_{L^\infty}}{M} m(\Omega) \right)$$

$$\leq \frac{4p}{k^2} \left( \|U\|_{L^\infty} \sum_{K \in \mathcal{T}} m(K) \frac{\log(1 + (P^D_K - 1)^2)}{\log(M/V^D_+)^2} + \frac{4p m(\Omega)}{\beta^2_U} \frac{\|f^+\|_{L^\infty}}{M} \right),$$

where we used an argument similar to (16) in the second inequality and a discrete Poincaré inequality in the third one. Then, by using (19) again we get

$$E_1(P^D) \leq \frac{4p m(\Omega)}{\beta^2_U} \left( \|U\|_{L^\infty}^2 + \frac{\|f^+\|_{L^\infty}}{V^D_+} \right)$$

Therefore, the smallness condition of Proposition 2 is satisfied by $E_1(P^M)$ if

$$\left[ \|U\|_{L^\infty}^2 + \frac{\|f^+\|_{L^\infty}}{V^D_+} \right] \left( \|U\|_{L^\infty} + \frac{\|f^+\|_{L^\infty}}{M} \log \left( \frac{M}{V^D_+} \right)^2 \right) \leq C_{\Omega, \xi, p} \beta^2_U \log \left( \frac{M}{V^D_+} \right)^2 . \quad (19)$$

It is clear that $E_1(P^D)$ is satisfied for $M$ large enough, which permits to define $\overline{M}$. Observe that if $\nu^D = 0$ ($\nu^D = 1$) and $U = 0$, $\overline{M} = C_{\Omega, \xi, p} \|f^+\|_{L^\infty}$ works as expected.

Acknowledgements. The authors thank the Labex CEMPI (ANR-11-LABX-0007-01) and the ANR MOHYCON (ANR-17-CE40-0027-01) for their support. They also want to thank Alexis F. Vasseur for fruitful exchanges on the subject.

References

1. Chainais-Hillairet, C., Herda, M.: Large-time behaviour of a family of finite volume schemes for boundary-driven convection-diffusion equations. IMA J. Numer. Anal. (2019). DOI 10.1093/imanum/drz037
2. Droniou, J.: Non-coercive linear elliptic problems. Potential Anal. 17(2), 181–203 (2002)
3. Droniou, J.: Error estimates for the convergence of a finite volume discretization of convection-diffusion equations. J. Numer. Math. 11(1), 1–32 (2003). DOI 10.1163/156939503322004873
4. Droniou, J., Gallouët, T.: Finite volume methods for convection-diffusion equations with right-hand side in $H^{-1}$. M2AN Math. Model. Numer. Anal. 36(4), 705–724 (2002)
5. Droniou, J., Gallouët, T., Herbin, R.: A finite volume scheme for a noncoercive elliptic equation with measure data. SIAM J. Numer. Anal. 41(6), 1997–2031 (2003)
6. Eymard, R., Gallouët, T., Herbin, R.: Finite volume methods. In: Handbook of numerical analysis, vol. VII, pp. 713–1020. North-Holland, Amsterdam (2000)
7. Vasseur, A.F.: The De Giorgi method for elliptic and parabolic equations and some applications. In: Lectures on the analysis of nonlinear partial differential equations., vol. 4 (2016)