Numer. Math. (2007) 105:337–374
DOI 10.1007/s00211-006-0033-2

Numerische Mathematik

Finite elements approximation of second order linear elliptic equations in divergence form with right-hand side in L^1

J. Casado-Díaz · T. Chacón Rebollo · V. Girault · M. Gómez Mármol · F. Murat

Received: 6 December 2004 /
Published online: 23 November 2006

Abstract In this paper we consider, in dimension $d \geq 2$, the standard $\mathbb{P}_1$ finite elements approximation of the second order linear elliptic equation in divergence form with coefficients in $L^\infty(\Omega)$ which generalizes Laplace's equation. We assume that the family of triangulations is regular and that it satisfies an hypothesis close to the classical hypothesis which implies the discrete maximum principle. When the right-hand side belongs to $L^1(\Omega)$, we prove that the unique solution of the discrete problem converges in $W_0^{1,q}(\Omega)$ (for every q with $1 \leq q < \frac{d}{d-1}$) to the unique renormalized solution of the problem. We obtain a weaker result when the right-hand side is a bounded Radon measure. In the case where the dimension is $d = 2$ or $d = 3$ and where the coefficients are smooth, we give an error estimate in $W_0^{1,q}(\Omega)$ when the right-hand side belongs to $L^r(\Omega)$ for some $r > 1$.

Mathematics Subject Classification 65N12 · 65N30 · 35A35 · 35J25

J. Casado-Díaz · T. Chacón Rebollo · M. Gómez Marmol
Departamento de Ecuaciones Diferenciales y Análisis Numérico, Apdo. de correos 1160, Universidad de Sevilla, 41080 Sevilla, Spain
e-mail: jcasadod@us.es

T. Chacón Rebollo
e-mail: chacon@us.es

M. Gómez Marmol
e-mail: macarena@us.es

V. Girault (✉) · F. Murat
Laboratoire Jacques-Louis Lions, CNRS UMR 7598, Université Paris VI, Boîte courrier 187, 75252 Paris cedex 05, France
e-mail: Girault@ann.jussieu.fr

F. Murat
e-mail: murat@ann.jussieu.fr

0 Introduction

In this paper we consider the $\mathbb{P}_1$ finite elements approximation of the boundary value problem

$$\begin{cases} -\operatorname{div} A\,\nabla u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{0.1}$$

where Ω is an open bounded set of $\mathbb{R}^d$, with $d \geq 2$, A is a coercive matrix with coefficients in $L^\infty(\Omega)$ and f belongs to $L^1(\Omega)$. This type of problem often arises in applications, as for example in the modelling of heat transfer and of turbulence. Then in general f is an energy dissipated by the system. The fact that f belongs to $L^1(\Omega)$ is the outstanding feature of the present paper.

For this problem the standard $\mathbb{P}_1$ finite elements approximation, namely

$$\begin{cases} u_h \in V_h\,, \\ \forall v_h \in V_h\,, \quad \int\limits_\Omega A\,\nabla u_h\,\nabla v_h\,\mathrm{d}x = \int_\Omega f\,v_h\,\mathrm{d}x\,, \end{cases} \tag{0.2}$$

where

$$V_h = \{v_h \in \mathcal{C}^0(\overline{\Omega}) \,:\, \forall T \in \mathcal{T}_h\,,\; v_h|_T \in \mathbb{P}_1, v_h|_{\partial\Omega} = 0\}\,, \tag{0.3}$$

has a unique solution, since the right-hand side $\int_\Omega f\,v_h\,\mathrm{d}x$ is correctly defined for $f \in L^1(\Omega)$.

However one cannot hope that the solution of (0.2) converges in $H_0^1(\Omega)$ to the solution u of (0.1), since the solution of (0.1) does not belong to $H_0^1(\Omega)$ for a general right-hand side in $L^1(\Omega)$. Actually, in order to correctly define the solution of (0.1), one has to consider a specific framework, the concept of renormalized solution (or equivalently of entropy solution). The definitions of these solutions (see Sect. 1 below) have been respectively introduced by Lions and Murat [19] and by Bénilan et al. [2]. These definitions allow one to prove that in this new sense problem (0.1) is well posed in the terminology of Hadamard, namely that the solution of (0.1) exists, is unique, and depends continuously on the right-hand side f.

Using the ideas which are at the root of the definition of renormalized solution, we are able to prove in the present paper (Theorem 1.3) that the unique solution u_h of (0.2) converges to the unique renormalized solution u of (0.1) in the following sense

$$\begin{cases} u_h \to u & \text{strongly in } W_0^{1,q}(\Omega), \\ \Pi_h\,(T_k(u_h)) \to T_k(u) & \text{strongly in } H_0^1(\Omega), \end{cases} \tag{0.4}$$

for every q with $1 \leq q < \frac{d}{d-1}$ and for every $k > 0$, where Π_h is the usual Lagrange interpolation operator in V_h and where T_k is the usual truncation at height k.

To prove (0.4), we assume that the family of triangulations is regular in the sense of Ciarlet [8], and that it satisfies an assumption which is close to the assumption which is usually made to ensure that the discrete maximum principle holds true. More precisely, denoting by φ_i the basis functions of V_h, we assume that the matrix with coefficients Q_{ij} defined by

$$Q_{ij} = \int_\Omega A \nabla \varphi_i \nabla \varphi_j \, \mathrm{d}x$$

is a diagonally dominant matrix (hypothesis (1.17)). This allows us to prove (Proposition 3.1) that the solution u_h of (0.2) satisfies

$$\alpha \int_\Omega |\nabla \Pi_h(T_k(u_h))|^2 \, \mathrm{d}x \leq k \, \|f\|_{L^1(\Omega)},$$

for every h and every $k > 0$. This is the main estimate of the present paper.

The assumption that Q is a diagonally dominant matrix is unfortunately a restriction on the coercive matrices A with $L^\infty(\Omega)$ coefficients and on the triangulations $\mathcal{T}_h$ of Ω. In the case of Laplace's operator, we recall in Sect. 6 the classical result (see e.g. Ciarlet and Raviart [9]) which asserts that this condition is satisfied when every inner angle of every d-simplex of the triangulations $\mathcal{T}_h$ is acute. We also show in that section that Q is a diagonally dominant matrix for an adequate regular family of triangulations when the matrix A is of the form

$$A(x) = a(x)C + E(x),$$

where

$$\begin{aligned} &a \in L^\infty(\Omega), \quad \text{a.e. } x \in \Omega, \quad a(x) \geq \alpha > 0, \\ &C \text{ is a symmetric coercive matrix with constant coefficients,} \\ &E \in L^\infty(\Omega)^{d\times d} \text{ with } \|E\|_{L^\infty(\Omega)^{d\times d}} \text{ sufficiently small.} \end{aligned}$$

In Sect. 5 we complete the main result of the present paper, namely the convergence of u_h to u in the sense of (0.4), by the error estimate (Theorem 5.1)

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq C h^{2(1-\frac{1}{r})} \, \|f\|_{L^r(\Omega)},$$

when $d = 2$ or $d = 3$, when f belongs to $L^r(\Omega)$ with $1 < r < 2$ and when the coefficients of the matrix A are smooth.

Some other error estimates have been obtained previously in similar settings. Scott [23] derived error estimates for finite element approximations of general elliptic problems with singular data in norms of order lower than the elliptic norms. In particular, when the datum is a Dirac distribution in a plane domain, he proved an error estimate of order h in the L^2 norm. Also Clain [10] obtained by duality arguments error estimates in fractional Sobolev norms for the Laplace operator in a plane convex domain with bounded Radon measure data.

In Sect. 4 we consider the case where f is a bounded Radon measure. We prove that for a subsequence (still denoted by h) the unique solution u_h of (0.2) converges to a solution u of

$$\begin{cases} \forall\, q \quad \text{with } 1 \le q < \frac{d}{d-1}, \;\; u \in W_0^{1,q}(\Omega), \\ \forall\, k > 0, \quad T_k(u) \in H_0^1(\Omega), \\ -\mathrm{div}\, A\, \nabla\, u = f \quad \text{in } \mathcal{D}'(\Omega), \end{cases} \tag{0.5}$$

in the following sense (compare with (0.4))

$$\begin{cases} u_h \rightharpoonup u & \text{weakly in } W_0^{1,q}(\Omega), \\ \Pi_h\,(T_k(u_h)) \rightharpoonup T_k(u) & \text{weakly in } H_0^1(\Omega), \end{cases} \tag{0.6}$$

for every q with $1 \le q < \frac{d}{d-1}$ and for every $k > 0$. In general it is not known whether the solution of (0.5) is unique or not. When this solution is unique (this is the case if $\partial\Omega$ is smooth and if $d = 2$ and/or if the coefficients of the matrix A are smooth), the whole sequence converges. We therefore obtain in this context a result which is similar to the result recently obtained in dimension $d = 2$ by Gallouët and Herbin [17].

0.1 Notation

In the present paper, Ω denotes an open bounded subset of $\mathbb{R}^d$ with $d \ge 2$. A particular case is the case where Ω is an open bounded polyhedron.

We use the notation $A\,v\,w$ for the scalar product of the vector $A\,v$ by the vector w (which is often denoted by ${}^t w \cdot A\,v$).

For a measurable set $S \subset \Omega$, we denote by $|S|$ the measure of S, by S^c the complement $\Omega \backslash S$ of S, and by χ_s the characteristic function of S.

For $1 < p < +\infty$, we denote by $W^{1,p}(\Omega)$ the standard Sobolev space

$$W^{1,p}(\Omega) = \{u \in L^p(\Omega) : \nabla\, u \in L^p(\Omega)^d\},$$

equipped with the norm

$$\|u\|_{W^{1,p}(\Omega)} = (\|u\|^p_{L^p(\Omega)} + \|\nabla\, u\|^p_{L^p(\Omega)^d})^{1/p}\,,$$

and by $W_0^{1,p}(\Omega)$ the closure in $W^{1,p}(\Omega)$ of $C_c^\infty(\Omega)$, the space of those C^∞ functions whose support is compact and contained in Ω. Since Ω is bounded, $W_0^{1,p}(\Omega)$ will be equipped with the equivalent norm

$$\|u\|_{W_0^{1,p}(\Omega)} = \|\nabla\, u\|_{L^p(\Omega)^d}\,.$$

We denote by $W^{-1,p'}(\Omega)$, with $p' = \frac{p}{p-1}$, the dual of $W_0^{1,p}(\Omega)$, and when $p = 2$, we denote as usual

$$H^1(\Omega) = W^{1,2}(\Omega), \quad H_0^1(\Omega) = W_0^{1,2}(\Omega) \quad \text{and} \quad H^{-1}(\Omega) = W^{-1,2}(\Omega).$$

We denote by $\mathcal{M}_b(\Omega)$ the space of Radon measures on Ω with total bounded variation, also called the space of bounded Radon measures.

For every r with $1 < r < +\infty$, we denote by $L^{r,\infty}(\Omega)$ the Marcinkiewicz space whose norm is defined by

$$\|v\|_{L^{r,\infty}(\Omega)} = \sup_{\lambda>0} (\lambda\, |\{x \in \Omega : |v(x)| \geq \lambda\}|^{1/r}). \tag{0.7}$$

For every real number $k > 0$ we define the truncation $T_k : \mathbb{R} \to \mathbb{R}$ by

$$T_k(s) = \begin{cases} s & \text{if } |s| \leq k, \\ k\dfrac{s}{|s|} & \text{if } |s| \geq k. \end{cases}$$

1 Setting of the problem and main result

We consider a matrix A such that

$$A \in L^\infty(\Omega)^{d\times d}, \tag{1.1}$$

$$\text{a.e. } x \in \Omega, \quad \forall \xi \in \mathbb{R}^d, \quad A(x)\xi\xi \geq \alpha|\xi|^2, \tag{1.2}$$

for some $\alpha > 0$, and a right-hand side f such that

$$f \in L^1(\Omega). \tag{1.3}$$

Let us recall the definition of the renormalized solution of the problem

$$\begin{cases} -\operatorname{div} A\,\nabla u = f \text{ in } \Omega, \\ u = 0 \qquad\qquad \text{on } \partial\Omega. \end{cases} \tag{1.4}$$

Definition 1.1 *A function u is a renormalized solution of* (1.4) *if u satisfies*

$$u \in L^1(\Omega)\,, \tag{1.5}$$

$$\forall\, k > 0\,, \quad T_k(u) \in H_0^1(\Omega)\,, \tag{1.6}$$

$$\lim_{k\to\infty} \frac{1}{k} \int_\Omega |\nabla T_k(u)|^2 \,\mathrm{d}x = 0\,, \tag{1.7}$$

$$\begin{cases} \forall\, k > 0, \quad \forall\, S \in C_c^1(\mathbb{R}) \ \textit{with}\ \operatorname{supp} S \subset [-k, +k], \\ \forall\, v \in H_0^1(\Omega) \cap L^\infty(\Omega)\,, \\ \int\limits_\Omega A\, \nabla\, T_k(u)\, \nabla\, v\; S(u)\, \mathrm{d}x + \int_\Omega A\, \nabla\, T_k(u)\, \nabla\, T_k(u)\; S'(u)\, v\, \mathrm{d}x \\ \quad = \int\limits_\Omega f\, S(u)\, v\, \mathrm{d}x\,. \end{cases} \tag{1.8}$$

In (1.8) every term makes sense since $T_k(u)$ belongs to $H_0^1(\Omega)$. Equation (1.8) is the correct way to write the result which is obtained formally when using $v\, S(u)$ as test function in (1.4).

It is easy to see that when f belongs to $L^1(\Omega) \cap H^{-1}(\Omega)$, the usual weak solution of (1.4), namely

$$\begin{cases} u \in H_0^1(\Omega), \\ \forall v \in H_0^1(\Omega), \quad \int\limits_\Omega A\, \nabla\, u\, \nabla\, v\, \mathrm{d}x = \int\limits_\Omega f\, v\, \mathrm{d}x, \end{cases} \tag{1.9}$$

is also a renormalized solution of (1.4) and conversely.

The above definition of renormalized solution was introduced by Lions and Murat [19] (see also [11,21,22]). Two others definitions of solutions, the entropy solution and the solution obtained as limit of approximations, were introduced at the same time respectively by Bénilan et al. [2] and by Dall'Aglio [12]. The three definitions can be proved to be equivalent (see e.g. [11]), and they can actually be given for monotone operators acting in $W_0^{1,p}(\Omega)$. In the linear case considered in the present work, the three definitions are also equivalent to the definition of solution by transposition introduced in 1969 by Stampacchia [25] (see e.g. [11]).

The main interest of the definition of renormalized solution is the following existence, uniqueness and continuity theorem.

Theorem 1.2 *Assume that A and f satisfy* (1.1), (1.2) *and* (1.3). *Then there exists a renormalized solution of* (1.4). *This solution is unique. Moreover this unique solution belongs to $W_0^{1,q}(\Omega)$ for every q with $1 \le q < \frac{d}{d-1}$. It depends continuously on the right-hand side f in the following sense: if f^ε is a sequence which satisfies*

$$f^\varepsilon \to f \quad \textit{strongly in} \;\; L^1(\Omega),$$

when ε tends to zero, then the sequence u^ε of the renormalized solutions of (1.4) *for the right-hand sides f^ε satisfies for every $k > 0$ and for every q with $1 \le q < \frac{d}{d-1}$*

$$T_k(u^\varepsilon) \to T_k(u) \quad \textit{strongly in } H_0^1(\Omega),$$
$$u^\varepsilon \to u \quad \textit{strongly in } W_0^{1,q}(\Omega),$$

when ε tends to zero, where u is the renormalized solution of (1.4) *for the right-hand side f. Finally, if f_1 and f_2 belong to $L^1(\Omega)$, and if u_1 and u_2 are the renormalized solutions of* (1.4) *for the right-hand sides f_1 and f_2, then for every $k > 0$, the function $T_k(u_1 - u_2)$ belongs to $H_0^1(\Omega)$ and for every q with $1 \le q < \frac{d}{d-1}$ one has*

$$\alpha \| T_k(u_1 - u_2)\|^2_{H_0^1(\Omega)} \le k\, \|f_1 - f_2\|_{L^1(\Omega)},$$
$$\|u_1 - u_2\|_{W_0^{1,q}(\Omega)} \le C_1(d, |\Omega|, q)\, \frac{1}{\alpha}\, \|f_1 - f_2\|_{L^1(\Omega)}, \tag{1.10}$$

where the constant $C_1(d, |\Omega|, q)$ only depends on d, $|\Omega|$ and q.

Now we consider a family of triangulations $\mathcal{T}_h$ satisfying for each $h > 0$ the following assumption:

$$\begin{cases} \text{the triangulation } \mathcal{T}_h \text{ is made of a finite number} \\ \text{of closed } d\text{-simplices } T \text{ (namely triangles when } d = 2, \\ \text{tetrahedra when } d = 3, \text{ etc.) such that:} \\ \text{(i) } \Omega_h = \cup\{T : T \in \mathcal{T}_h\} \subset \overline{\Omega}, \\ \text{(ii) for every compact set } K \text{ with } K \subset \Omega, \text{ there exists} \\ h_0(K) > 0 \text{ such that } K \subset \Omega_h \text{ for every } h \text{ with } h < h_0(K), \\ \text{(iii) for } T_1 \text{ and } T_2 \text{ of } \mathcal{T}_h \text{ with } T_1 \neq T_2, \text{ one has } |T_1 \cap T_2| = 0, \\ \text{(iv) every face of every } T \text{ of } \mathcal{T}_h \text{ is either a subset of } \partial\Omega_h, \\ \text{or a face of another } T' \text{ of } \mathcal{T}_h. \end{cases} \tag{1.11}$$

Note that because of (iv) the triangulations are conforming. A particular case is the case where Ω is a polyhedron of $\mathbb{R}^d$, and where $\overline{\Omega}_h$ coincides with $\overline{\Omega}$ for every h.

The vertices of the d-simplices T of $\mathcal{T}_h$ are denoted by a_i. There are interior and boundary vertices, namely vertices which belong to $\mathring{\Omega}_h$ and vertices which belong to $\partial\Omega_h$. We denote by I the set of indices corresponding to interior vertices and by B the set of indices corresponding to boundary vertices.

For every $T \in \mathcal{T}_h$, we denote by h_T the diameter of T and by ρ_T the diameter of the ball inscribed in T. We set

$$h = \sup_{T \in \mathcal{T}_h} h_T, \tag{1.12}$$

and we assume that h tends to zero. We also assume that the family of triangulations $\mathcal{T}_h$ is regular in the sense of Ciarlet [8], namely that there exists a constant σ such that

$$\forall h, \quad \forall T \in \mathcal{T}_h, \quad \frac{h_T}{\rho_T} \leq \sigma. \tag{1.13}$$

On every triangulation $\mathcal{T}_h$, we define the space V_h of those continuous functions which are affine on each d-simplex of $\mathcal{T}_h$ and which vanish on $\overline{\Omega}\backslash\mathring{\Omega}_h$, namely

$$V_h = \{v_h \in \mathcal{C}^0(\overline{\Omega}) \ : \ v_h = 0 \text{ in } \overline{\Omega}\backslash\mathring{\Omega}_h, \ \forall T \in \mathcal{T}_h, \ v_h|_T \in \mathbb{P}_1\}. \tag{1.14}$$

One has

$$V_h \subset H_0^1(\Omega).$$

For every (interior or boundary) vertex a_i of $\mathcal{T}_h$, i.e. for every $i \in I \cup B$, we define the function φ_i by

$$\begin{cases} \varphi_i \in C^0(\Omega_h), \quad \varphi_{i|_T} \in \mathbb{P}_1 \text{ for every } T \in \mathcal{T}_h, \\ \varphi_i(a_i) = 1, \quad \varphi_i(a_j) = 0 \quad \text{for every vertex } a_j \text{ of } \mathcal{T}_h \text{ with } a_j \neq a_i. \end{cases}$$

One has

$$\sum_{i \in I \cup B} \varphi_i = 1 \quad \text{in } \Omega_h. \tag{1.15}$$

When a_i is an interior vertex, i.e. when $i \in I$, then the function φ_i belongs to $H_0^1(\mathring{\Omega}_h)$, and extending φ_i by zero to $\overline{\Omega}\backslash\mathring{\Omega}_h$, we obtain a function of V_h, still denoted by φ_i. The functions φ_i, $i \in I$, are a basis of the space V_h.

We define the interpolation operator Π_h by

$$\begin{cases} \forall v \in C^0(\overline{\Omega}) \quad \text{with } v = 0 \text{ in } \overline{\Omega}\backslash\mathring{\Omega}_h, \\ \Pi_h(v) \in V_h, \quad (\Pi_h(v))(a_i) = v(a_i) \quad \text{for every vertex } a_i \text{ of } \mathcal{T}_h, \end{cases}$$

or equivalently by

$$\Pi_h(v) = \sum_{i \in I} v(a_i)\, \varphi_i.$$

For all interior vertices a_i and a_j of $\mathcal{T}_h$, i.e. for every i and j of I, we define the real number

$$Q_{ij} = \int_\Omega A\, \nabla \varphi_i\, \nabla \varphi_j\, \mathrm{d}x; \tag{1.16}$$

this defines an $I \times I$ matrix Q. The main assumption of the present paper is that Q satisfies

$$\forall\, i \in I, \quad Q_{ii} - \sum_{\substack{j\in I\\ j\neq i}} |Q_{ij}| \geq 0. \tag{1.17}$$

In other words, Q is assumed to be a diagonally dominant matrix. This assumption is close to the usual assumption which ensures that the discrete maximum principle holds true (see Remark 6.2 below). We present in Sect. 6 some examples where assumption (1.17) is satisfied.

For every triangulation $\mathcal{T}_h$, we consider the solution u_h of

$$\begin{cases} u_h \in V_h\,, \\ \forall\, v_h \in V_h\,, \quad \int\limits_\Omega A\,\nabla\, u_h\, \nabla\, v_h\, \mathrm{d}x = \int\limits_\Omega f\, v_h\, \mathrm{d}x\,. \end{cases} \tag{1.18}$$

Note that the right-hand side of (1.18) makes sense since f belongs to $L^1(\Omega)$ and v_h to $V_h \subset L^\infty(\Omega)$. The solution u_h of (1.18) exists and is unique.

Our main result is the following.

Theorem 1.3 *Assume that A, f and $\mathcal{T}_h$ satisfy* (1.1), (1.2), (1.3), (1.11), (1.12), (1.13) *and* (1.17). *Then the unique solution u_h of* (1.18) *satisfies for every $k > 0$ and for every q with $1 \leq q < \frac{d}{d-1}$*

$$\Pi_h\,(T_k(u_h)) \to T_k(u) \quad \text{strongly in } \; H_0^1(\Omega),$$
$$u_h \to u \quad \text{strongly in } \; W_0^{1,q}(\Omega),$$

when h tends to zero, where u is the unique renormalized solution of (1.4).

This theorem will be proved in Sect. 3, using the tools that we will prepare in Sect. 2. In Sect. 4 we will give a variant of this result in the case where f is a bounded Radon measure, and in Sect. 5 an error estimate when $d = 2$ or $d = 3$, when f belongs to $L^r(\Omega)$ with $1 < r < 2$ and when the coefficients of the matrix A are smooth.

2 Tools

In this section we prove various results which will be used in particular in the proofs of Theorems 1.3 and 4.1.

The following result is a piecewise $\mathbb{P}_1$ variant of a result of Boccardo and Gallouët [4,5] (see also Bénilan et al. [2]).

Theorem 2.1 *Assume that $v_h \in V_h$ and that v_h satisfies*

$$\forall\, k > 0, \quad \int\limits_\Omega |\nabla\, \Pi_h(T_k(v_h))|^2\, \mathrm{d}x \leq k\, M, \tag{2.1}$$

for some $M > 0$. *Then, for every* q *with* $1 \leq q < \frac{d}{d-1}$

$$\|v_h\|_{W_0^{1,q}(\Omega)} \leq C_2(d, |\Omega|, q)\, M, \tag{2.2}$$

where the constant $C_2(d, |\Omega|, q)$ *only depends on* $d, |\Omega|$ *and* q.

Remark 2.2 When $d \geq 3$, we will actually prove a result which is stronger than (2.2), namely

$$\|v_h\|_{L^{\frac{d}{d-2},\infty}(\Omega)} \leq C(d)\, M, \tag{2.3}$$

$$\|\nabla v_h\|_{L^{\frac{d}{d-1},\infty}(\Omega)^d} \leq C(d)\, M, \tag{2.4}$$

for a constant $C(d)$ which only depends on d, where $L^{r,\infty}(\Omega)$ denotes the Marcinkiewicz space whose norm is defined by (0.7). Indeed (2.4) and the embedding inequality

$$\forall q, \quad 1 \leq q < r, \quad \|\psi\|_{L^q(\Omega)} \leq C(q, r, |\Omega|)\, \|\psi\|_{L^{r,\infty}(\Omega)} \tag{2.5}$$

immediately imply (2.2). □

The proof of Theorem 2.1 uses the following lemma.

Lemma 2.3 *Let* $v_h \in V_h$ *and let* $k > 0$. *If for some* $T \in \mathcal{T}_h$ *there exists* $y \in T$ *with* $|v_h(y)| \geq k$, *then there exists a* d*-simplex* $S \subset T$ *with* $|S| = C(d)\,|T|$ *such that*

$$\forall x \in S, \quad |\Pi_h(T_k(v_h))(x)| \geq \frac{k}{2},$$

where the strictly positive constant $C(d)$ *only depends on* d.

Proof Consider $T \in \mathcal{T}_h$. In order to simplify the notation, in this proof we denote by a_i, $i = 0, \ldots, d$, the vertices of T. Let λ_i, $i = 0, \ldots, d$, be the barycentric coordinates with respect to the a_i's. Recall that

$$\forall i, j, \quad i, j = 0, \ldots, d, \quad \lambda_i \in \mathbb{P}_1, \quad \lambda_i(a_j) = \delta_{ij},$$

$$\forall x \in \mathbb{R}^d, \quad \sum_{i=0}^{d} \lambda_i(x) = 1,$$

and that T is characterized by

$$T = \{x \in \mathbb{R}^d : 0 \leq \lambda_i(x) \leq 1, \quad i = 0, \ldots, d\}.$$

If v_h is affine in T and if $|v_h(y)| \geq k$ for some $y \in T$, there exists a vertex, say a_0, where $|v_h(a_0)| \geq k$. We define S as

$$S = \left\{ x \in T : \lambda_0(x) \geq \frac{3}{4} \right\}.$$

Then S is a d-simplex contained in T and similar to T.

Since the function $\Pi_h(T_k(v_h))$ is affine in T, it satisfies for every $x \in T$

$$\Pi_h(T_k(v_h))(x) = \sum_{i=0}^{d} \lambda_i(x)\, \Pi_h(T_k(v_h))(a_i) = \sum_{i=0}^{d} \lambda_i(x)\, T_k(v_h)(a_i),$$

and therefore one has, for every $x \in S$

$$\begin{aligned} |\Pi_h(T_k(v_h))(x)| &= \left| \sum_{i=0}^{d} \lambda_i(x)\, T_k(v_h)(a_i) \right| \\ &\geq \lambda_0(x)|T_k(v_h)(a_0)| - \sum_{i=1}^{d} \lambda_i(x)|T_k(v_h)(a_i)| \\ &\geq \lambda_0(x)\, k - \sum_{i=1}^{d} \lambda_i(x)\, k = \lambda_0(x)\, k - (1 - \lambda_0(x))\, k \geq \frac{k}{2}. \end{aligned}$$

It remains to estimate the measure of S. Let $\hat{T}$ be the reference unit d-simplex with vertices $\hat{a}_0 = 0$ and $\hat{a}_i = e_i$, $i = 1, \ldots, d$, where e_i, $i = 1, \ldots, d$, is the canonical basis of $\mathbb{R}^d$. Let F_T be the invertible affine mapping that maps $\hat{T}$ onto T. Set $\hat{S} = F_T^{-1}(S)$. It is easy to check that

$$\hat{S} = \left\{ \hat{x} \in \hat{T} : \hat{\lambda}_0(\hat{x}) \geq \frac{3}{4} \right\},$$

and that $|S| = \frac{|\hat{S}|}{|\hat{T}|}\, |T| = C(d)\, |T|$, with $C(d) = \frac{|\hat{S}|}{|\hat{T}|}$ a constant that depends only on d. This proves the result. □

Proof of Theorem 2.1 Sobolev's theorem asserts that

$$\forall\, v \in H_0^1(\Omega), \quad \|v\|_{L^{2^*}(\Omega)} \leq C_S \|\nabla v\|_{L^2(\Omega)^d},$$

where $2^* = \frac{2d}{d-2}$ if $d \geq 3$ (and then C_S only depends on d), and where 2^* is any real number with $1 \leq 2^* < +\infty$ if $d = 2$ (and then C_S depends on $|\Omega|$). From this estimate and (2.1) we deduce that

$$\int_\Omega |\Pi_h(T_k(v_h))|^{2^*}\,dx \le C_S^{2^*}\left(\int_\Omega |\nabla\,\Pi_h(T_k(v_h))|^2\,dx\right)^{\frac{2^*}{2}} \le C_S^{2^*}(k\,M)^{\frac{2^*}{2}}. \quad (2.6)$$

For $k > 0$, we define the set $B(k)$ by

$$B(k) = \bigcup \{T \in \mathcal{T}_h : \exists\, y \in T \text{ with } |v_h(y)| \ge k\}.$$

From Lemma 2.3 we know that for every $T \in \mathcal{T}_h$, with $T \subset B(k)$, there exists $S \subset T$, with $|S| = C(d)\,|T|$ and

$$\forall\, x \in S, \quad |\Pi_h(T_k(v_h))(x)| \ge \frac{k}{2}.$$

Therefore if $T \subset B(k)$

$$\int_T |\Pi_h(T_k(v_h))|^{2^*}\,dx \ge \int_S |\Pi_h(T_k(v_h))|^{2^*}\,dx \ge \left(\frac{k}{2}\right)^{2^*}|S| = \left(\frac{k}{2}\right)^{2^*}C(d)\,|T|,$$

and so

$$\begin{aligned}|B(k)| = \sum_{T\subset B(k)} |T| &\le \sum_{T\subset B(k)} \frac{1}{C(d)\left(\frac{k}{2}\right)^{2^*}} \int_T |\Pi_h(T_k(v_h))|^{2^*}\,dx \\ &\le \frac{1}{C(d)\left(\frac{k}{2}\right)^{2^*}} \int_\Omega |\Pi_h(T_k(v_h))|^{2^*}\,dx.\end{aligned}$$

From (2.6) one deduces that

$$|B(k)| \le \frac{1}{C(d)\left(\frac{k}{2}\right)^{2^*}} C_S^{2^*}(kM)^{\frac{2^*}{2}} = \frac{(2C_S)^{2^*}}{C(d)}\frac{M^{\frac{2^*}{2}}}{k^{\frac{2^*}{2}}}. \quad (2.7)$$

The inclusion $\{x \in \Omega : \; |v_h(x)| \ge k\} \subset B(k)$ and inequality (2.7) imply that

$$k^{\frac{2^*}{2}}|\{x \in \Omega : \; |v_h(x)| \ge k\}| \le k^{\frac{2^*}{2}}|B(k)| \le \frac{(2C_S)^{2^*}}{C(d)}M^{\frac{2^*}{2}},$$

which is exactly (2.3) when $d \ge 3$, since $\frac{2^*}{2} = \frac{d}{d-2}$.

For every $\lambda > 0$ and for every $k > 0$ one has

$$\begin{aligned}&\{x \in \Omega : \; |\nabla\, v_h(x)| \ge \lambda\} \\ &\quad = \{x \in \Omega : \; |\nabla\, v_h(x)| \ge \lambda \text{ and } x \in B(k)\} \cup \\ &\quad \cup \{x \in \Omega : \; |\nabla\, v_h(x)| \ge \lambda \text{ and } x \in B(k)^c\},\end{aligned}$$

and therefore

$$\begin{cases} |\{x \in \Omega : |\nabla v_h(x)| \geq \lambda\}| \\ \leq |B(k)| + |\{x \in \Omega : |\nabla v_h(x)| \geq \lambda \text{ and } x \in B(k)^c\}|. \end{cases} \tag{2.8}$$

But $B(k)^c$ coincides, up to a set of measure zero, with the union of the d-simplices $T \in \mathcal{T}_h$ which are not contained in $B(k)$. On such a T, one has $|v_h(x)| \leq k$, and therefore $\Pi_h(T_k(v_h))(x) = v_h(x)$ and $\nabla \Pi_h(T_k(v_h))(x) = \nabla v_h(x)$. Therefore

$$\begin{aligned} &|\{x \in \Omega : |\nabla v_h(x)| \geq \lambda \text{ and } x \in B(k)^c\}| \\ &\quad = |\{x \in \Omega : |\nabla \Pi_h(T_k(v_h))(x)| \geq \lambda \text{ and } x \in B(k)^c\}| \\ &\quad \leq |\{x \in \Omega : |\nabla \Pi_h(T_k(v_h))(x)| \geq \lambda\}| \leq \frac{1}{\lambda^2} \int_\Omega |\nabla \Pi_h(T_k(v_h))(x)|^2 \mathrm{d}x. \end{aligned}$$

Going back to (2.8) and using (2.7) and hypothesis (2.1), we have proved that for every $\lambda > 0$ and every $k > 0$

$$|\{x \in \Omega : |\nabla v_h(x)| \geq \lambda\}| \leq \frac{(2C_S)^{2^*}}{C(d)} \frac{M^{\frac{2^*}{2}}}{k^{\frac{2^*}{2}}} + \frac{k M}{\lambda^2}.$$

Taking $k = \lambda^{\frac{4}{2^*+2}} M^{\frac{2^*-2}{2^*+2}}$ we obtain

$$\lambda^{\frac{2 2^*}{2^*+2}} |\{x \in \Omega : |\nabla v_h(x)| \geq \lambda\}| \leq \left(\frac{(2C_S)^{2^*}}{C(d)} + 1 \right) M^{\frac{2 2^*}{2^*+2}}.$$

When $d \geq 3$, since $\frac{2 2^*}{2^*+2} = \frac{d}{d-1}$, this is exactly (2.4), which implies (2.2) (see Remark 2.2). When $d = 2$, this is an estimate for $|\nabla v_h|$ in $L^{\frac{2 2^*}{2^*+2},\infty}(\Omega)$, where 2^* is any finite number, and (2.2) follows from this estimate and from (2.5). □

The next lemmas show that when v_h satisfies (2.1), then $\Pi_h(T_k(v_h))$ and $T_k(v_h)$ are close in measure.

Lemma 2.4 *Let $v_h \in V_h$. For every s and every k with $0 < s < k$, the set $B(k,s)$ defined by*

$$B(k,s) = \cup\{T \in \mathcal{T}_h : \exists x \in T, \exists y \in T, |v_h(x)| \geq k, |v_h(y)| \leq s\} \tag{2.9}$$

satisfies

$$|B(k,s)| \leq \frac{h^2}{(k-s)^2} \int_\Omega |\nabla \Pi_h(T_k(v_h))|^2 \mathrm{d}x. \tag{2.10}$$

Proof Consider $T \in \mathcal{T}_h$ which is contained in $B(k,s)$. Then there exist two points x and y in T such that

$$|v_h(x)| \geq k \quad \text{and} \quad |v_h(y)| \leq s\,.$$

Since v_h belongs to $\mathbb{P}_1$ in T, it attains its maximum and its minimum on the vertices. Since $-s \leq v_h(y) \leq s$, there are two cases:

(i) If $v_h(x) \geq k$, then there exist two vertices of T, say a_i and a_j, such that $v_h(a_i) \geq k \quad \text{and} \quad v_h(a_j) \leq s$. Hence
$T_k(v_h(a_i)) = k$, $T_k(v_h(a_j)) \leq s \quad \text{and} \quad k - s \leq T_k(v_h(a_i)) - T_k(v_h(a_j))$.

(ii) If $v_h(x) \leq -k$, then there exist two vertices of T, say a_i and a_j, such that $v_h(a_i) \leq -k \quad \text{and} \quad v_h(a_j) \geq -s$. Hence
$T_k(v_h(a_i)) = -k$, $T_k(v_h(a_j)) \geq -s \quad \text{and} \quad k - s \leq T_k(v_h(a_j)) - T_k(v_h(a_i))$.

Since the gradient of $\Pi_h(T_k(v_h))$ is a constant in T, we have in both cases that

$$\begin{aligned} k - s &\leq |T_k(v_h(a_i)) - T_k(v_h(a_j))| = |\Pi_h(T_k(v_h(a_i))) - \Pi_h(T_k(v_h(a_j)))| \\ &\leq |\nabla\, \Pi_h(T_k(v_h))|\, |a_i - a_j| \leq |\nabla\, \Pi_h(T_k(v_h))|\, h. \end{aligned}$$

Therefore

$$\int_\Omega |\nabla\, \Pi_h(T_k(v_h))|^2\, \mathrm{d}x \geq \int_{B(k,s)} |\nabla\, \Pi_h(T_k(v_h))|^2\, \mathrm{d}x \geq |B(k,s)| \frac{(k-s)^2}{h^2}\,,$$

which proves (2.10). □

Lemma 2.5 *Let $v_h \in V_h$. For every s and every k with $0 < s < k$, one has*

$$T_s(\Pi_h(T_k(v_h))) = T_s(v_h) \quad \textit{in } B(k,s)^c, \tag{2.11}$$

and

$$\nabla\, T_s(\Pi_h(T_k(v_h))) = \nabla\, T_s(v_h) \quad \textit{almost everywhere in } B(k,s)^c. \tag{2.12}$$

Proof Assertion (2.12) immediately follows from (2.11): indeed, the functions $T_s(v_h)$ and $T_s(\Pi_h(T_k(v_h)))$ belong to $H^1(\Omega)$, the set $E = B(k,s)^c$ is measurable and one has $\nabla v = 0$ a.e. in E for every $v \in H^1(\Omega)$ and for every measurable set E when $v = 0$ a.e. in E.

To prove (2.11) we fix $x \in B(k,s)^c$. Let us consider a d-simplex T with $x \in T$. There are five possibilities.

(i) If $v_h(x) \geq k$, then for every $y \in T$ one has $|v_h(y)| > s$. But actually one has $v_h(y) > s$, since if there exists $y_0 \in T$ with $v_h(y_0) < -s$, by continuity there also exists $y_1 \in T$ with $|v_h(y_1)| < s$, a contradiction with $|v_h(y)| > s$ for every $y \in T$. Hence for every $y \in T$

$$T_s(v_h)(y) = s,\ T_k(v_h)(y) > s,\ \Pi_h(T_k(v_h))(y) > s,\ T_s(\Pi_h(T_k(v_h)))(y) = s,$$

and therefore for every $y \in T$

$$T_s(\Pi_h(T_k(v_h)))(y) = T_s(v_h)(y), \tag{2.13}$$

which in particular holds for $y = x$.

(ii) If $v_h(x) \leq -k$, the proof is similar to (i).

(iii) If $|v_h(x)| \leq s$, then for every $y \in T$ one has $|v_h(y)| < k$, and therefore

$$\begin{aligned} &T_k(v_h)(y) = v_h(y), \ \Pi_h(T_k(v_h))(y) = v_h(y), \\ &T_s(\Pi_h(T_k(v_h)))(y) = T_s(v_h)(y), \end{aligned} \tag{2.14}$$

which in particular holds for $y = x$.

(iv) If $s < v_h(x) < k$, we consider some $z \in T$. If $|v_h(z)| \geq k$, we apply (i) or (ii) and we obtain (2.13), which holds for every $y \in T$, and in particular for $y = x$. If $|v_h(z)| \leq s$, we apply (iii) and we obtain (2.14), which holds for every $y \in T$, and in particular for $y = x$.
It remains to consider the case where $s < |v_h(z)| < k$ for every $z \in T$. As in case (i), by continuity one has actually $s < v_h(z) < k$ for every $z \in T$. Then

$$T_k(v_h)(z) = v_h(z), \ \Pi_h(T_k(v_h))(z) = v_h(z),$$

and therefore for every $z \in T$

$$T_s(\Pi_h(T_k(v_h)))(z) = T_s(v_h)(z),$$

which in particular holds for $z = x$.

(v) If $-k < v_h(x) < -s$, the proof is similar to (iv). □

In view of (2.10), $|B(k,s)|$ tends to zero when h tends to zero if estimate (2.1) holds. The following result is therefore an immediate consequence of Lemmas 2.5 and 2.4.

Proposition 2.6 *Assume that $v_h \in V_h$ and that v_h satisfies (2.1). Then for every s and every k with $0 < s < k$, one has*

$$T_s(\Pi_h(T_k(v_h))) - T_s(v_h) \to 0 \quad \textit{in measure}, \tag{2.15}$$

$$\nabla T_s(\Pi_h(T_k(v_h))) - \nabla T_s(v_h) \to 0 \quad \textit{in measure}, \tag{2.16}$$

when h tends to zero.

We conclude this section with an analogue in V_h of the fact that in the continuous case, for every $v \in H_0^1(\Omega)$ and every $k > 0$, one has

$$A \nabla (v - T_k(v)) \nabla T_k(v) = 0 \quad \text{almost everywhere in } \Omega.$$

Proposition 2.7 *Under assumption* (1.17), *one has for every* $v_h \in V_h$ *and every* $k > 0$

$$\int_\Omega A\,\nabla\,(v_h - \Pi_h(T_k(v_h)))\,\nabla\,\Pi_h(T_k(v_h))\,\mathrm{d}x \geq 0. \tag{2.17}$$

Proof Since

$$v_h = \sum_{i\in I} v_h(a_i)\,\varphi_i \quad \text{and} \quad \Pi_h(T_k(v_h)) = \sum_{i\in I} T_k(v_h)(a_i)\,\varphi_i,$$

using the definition (1.16) of Q_{ij}, we have

$$\int_\Omega A\,\nabla\,(v_h - \Pi_h(T_k(v_h)))\,\nabla\,\Pi_h(T_k(v_h))\,\mathrm{d}x$$
$$= \sum_{i,j\in I} Q_{ij}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_j)) = \sum_{i\in I} S_i,$$

where

$$S_i = Q_{ii}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_i))$$
$$+\sum_{\substack{j\in I\\ j\neq i}} Q_{ij}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_j)).$$

Fix $i \in I$. If $|v_h(a_i)| \leq k$, then $v_h(a_i) - T_k(v_h(a_i)) = 0$ and $S_i = 0$. If $|v_h(a_i)| > k$, then

$$(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_i)) = |v_h(a_i) - T_k(v_h(a_i))|\,k.$$

Since $|T_k(v_h(a_j))| \leq k$ for every j, one has

$$S_i \geq Q_{ii}|v_h(a_i) - T_k(v_h(a_i))|\,k - \sum_{\substack{j\in I\\ j\neq i}} |Q_{ij}||v_h(a_i) - T_k(v_h(a_i))|\,k$$
$$= |v_h(a_i) - T_k(v_h(a_i))|\,k\left(Q_{ii} - \sum_{\substack{j\in I\\ j\neq i}} |Q_{ij}|\right) \geq 0,$$

owing to hypothesis (1.17). This proves that

$$\forall\, i \in I, \quad S_i \geq 0,$$

and therefore (2.17), as desired. □

Remark 2.8 Proposition 2.7 asserts that condition (1.17) is a sufficient condition for (2.17) to hold true for every $v_h \in V_h$. Actually (1.17) is also necessary (and therefore necessary and sufficient) for (2.17) to hold true for every $v_h \in V_h$. Indeed, as seen in the above proof,

$$\begin{aligned}
&\int_\Omega A\,\nabla\,(v_h - \Pi_h(T_k(v_h)))\,\nabla\,\Pi_h(T_k(v_h))\,\mathrm{d}x \\
&\quad = \sum_{i,j\in I} Q_{ij}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_j)) \\
&\quad = \sum_{i\in I} Q_{ii}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_i)) \\
&\qquad + \sum_{\substack{j\in I\\ j\neq i}} Q_{ij}\,(v_h(a_i) - T_k(v_h(a_i)))\;T_k(v_h(a_j)).
\end{aligned}$$

Fixing $i \in I$ and taking $v_h(a_i) = k+1$ and $v_h(a_j) = -k\,sgn(Q_{ij})$ for every $j \in I$, $j \neq i$, proves that (1.17) holds true when (2.17) holds for every $v_h \in V_h$. □

3 Proof of Theorem 1.3

In this section we prove Theorem 1.3.

We first obtain an a priori estimate on the solution u_h of (1.18).

Proposition 3.1 *Under the assumptions of Theorem 1.3, the solution u_h of (1.18) satisfies for every $h > 0$ and every $k > 0$*

$$\int_\Omega A\,\nabla\,\Pi_h(T_k(u_h))\,\nabla\,\Pi_h(T_k(u_h))\,\mathrm{d}x \le \int_\Omega f\,\Pi_h(T_k(u_h))\,\mathrm{d}x. \tag{3.1}$$

In particular, u_h satisfies

$$\alpha \int_\Omega |\nabla\,\Pi_h(T_k(u_h))|^2\,\mathrm{d}x \le k\,\|f\|_{L^1(\Omega)}. \tag{3.2}$$

Proof Since $u_h \in V_h$, the function $T_k(u_h)$ is continuous and the function $\Pi_h(T_k(u_h))$ belongs to V_h. Using this function as test function in (1.18) we have

$$\int_\Omega A\,\nabla\,u_h\,\nabla\,\Pi_h(T_k(u_h))\,\mathrm{d}x = \int_\Omega f\,\Pi_h(T_k(u_h))\,\mathrm{d}x.$$

On the other hand, Proposition 2.7 shows that

$$\int_\Omega A\,\nabla\,(u_h - \Pi_h(T_k(u_h)))\,\nabla\,\Pi_h(T_k(u_h))\,\mathrm{d}x \geq 0.$$

This immediately implies (3.1). From (3.1) and from the coercivity (1.2) of A one deduces (3.2). □

Estimate (3.2) is the main estimate of the present paper. By Theorem 2.1, it implies that u_h is bounded in $W_0^{1,q}(\Omega)$ for every q with $1 \leq q < \frac{d}{d-1}$. We now prove the strong convergence of u_h in this space.

Theorem 3.2 *Under the assumptions of Theorem 1.3, the solution u_h of (1.18) satisfies for every q with $1 \leq q < \frac{d}{d-1}$*

$$u_h \to u \quad \textit{strongly in } \; W_0^{1,q}(\Omega), \tag{3.3}$$

when h tends to zero, where u is the unique renormalized solution of (1.4).

Proof Consider a sequence f^ε of functions such that

$$f^\varepsilon \in L^2(\Omega), \quad f^\varepsilon \to f \quad \text{strongly in } \; L^1(\Omega).$$

Such a sequence is easily obtained by taking for example $f^\varepsilon = T_{\frac{1}{\varepsilon}}(f)$. Let u_h^ε be the unique solution of (1.18) for the right-hand side f^ε. Then $u_h - u_h^\varepsilon$ satisfies

$$\begin{cases} u_h - u_h^\varepsilon \in V_h, \\ \forall\, v_h \in V_h, \quad \int\limits_\Omega A\,\nabla\,(u_h - u_h^\varepsilon)\,\nabla\, v_h\,\mathrm{d}x = \int\limits_\Omega (f - f^\varepsilon)\, v_h\,\mathrm{d}x. \end{cases}$$

Applying estimate (3.2) to this problem, we obtain for every $k > 0$, every $h > 0$ and every $\varepsilon > 0$

$$\alpha \int_\Omega |\nabla\,\Pi_h(T_k(u_h - u_h^\varepsilon))|^2\,\mathrm{d}x \leq k\,\|f - f^\varepsilon\|_{L^1(\Omega)},$$

which implies by Theorem 2.1 that for every q with $1 \leq q < \frac{d}{d-1}$, every $h > 0$ and every $\varepsilon > 0$

$$\|u_h - u_h^\varepsilon\|_{W_0^{1,q}(\Omega)} \leq C_2(d, |\Omega|, q)\,\frac{1}{\alpha}\,\|f - f^\varepsilon\|_{L^1(\Omega)}. \tag{3.4}$$

On the other hand, since $f^\varepsilon \in L^2(\Omega)$ and since the family of triangulations $\mathcal{T}_h$ satisfies (1.11), (1.12) and (1.13), it is well known that for every fixed ε

$$u_h^\varepsilon \to u^\varepsilon \;\; \text{strongly in } \; H_0^1(\Omega), \tag{3.5}$$

when h tends to zero, where u^ε is the unique solution of

$$\begin{cases} u^\varepsilon \in H_0^1(\Omega), \\ -\operatorname{div} A \nabla u^\varepsilon = f^\varepsilon \quad \text{in } \mathcal{D}'(\Omega). \end{cases} \tag{3.6}$$

Finally, the function u^ε, which is the unique weak solution of (3.6), is also the unique renormalized solution in the sense of Definition 1.1 of the problem

$$\begin{cases} -\operatorname{div} A \nabla u^\varepsilon = f^\varepsilon & \text{in } \Omega, \\ u^\varepsilon = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.7}$$

By estimate (1.10) we have

$$\|u^\varepsilon - u\|_{W_0^{1,q}(\Omega)} \leq C_1(d, |\Omega|, q) \frac{1}{\alpha} \|f^\varepsilon - f\|_{L^1(\Omega)}, \tag{3.8}$$

for every q with $1 \leq q < \frac{d}{d-1}$, where u is the unique renormalized solution of (1.4).

Writing now

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq \|u_h - u_h^\varepsilon\|_{W_0^{1,q}(\Omega)} + \|u_h^\varepsilon - u^\varepsilon\|_{W_0^{1,q}(\Omega)} + \|u^\varepsilon - u\|_{W_0^{1,q}(\Omega)},$$

and using (3.4), (3.5) and (3.8), we have proved that for every $\varepsilon > 0$ and every q with $1 \leq q < \frac{d}{d-1}$

$$\limsup_{h\to 0} \|u_h - u\|_{W_0^{1,q}(\Omega)} \leq \Big(C_1(d, |\Omega|, q) + C_2(d, |\Omega|, q)\Big) \frac{1}{\alpha} \|f^\varepsilon - f\|_{L^1(\Omega)}.$$

Taking the limit when ε tends to zero proves (3.3). □

To complete the proof of Theorem 1.3, it remains to prove that $\Pi_h(T_k(u_h))$ converges strongly to $T_k(u)$ in $H_0^1(\Omega)$. This is done in the following result.

Proposition 3.3 *Under the assumptions of Theorem 1.3, the solution u_h of (1.18) satisfies for every $k > 0$*

$$\Pi_h(T_k(u_h)) \to T_k(u) \quad \textit{strongly in } H_0^1(\Omega), \tag{3.9}$$

when h tends to zero.

Proof Fix $k > 0$. In view of estimate (3.2), we can extract a subsequence (which depends on k and is still denoted by h) such that for some $w_k \in H_0^1(\Omega)$

$$\Pi_h(T_k(u_h)) \rightharpoonup w_k \quad \text{weakly in } H_0^1(\Omega), \tag{3.10}$$

when h tends to zero. By estimate (3.2) and Proposition 2.6, u_h satisfies (2.15), namely

$$T_s(\Pi_h(T_k(u_h))) - T_s(u_h) \to 0 \quad \text{in measure,}$$

when h tends to zero, for every s with $0 < s < k$. The convergence (3.10), the convergence (3.3), the Rellich-Kondrashov's compactness theorem and the continuity of the function T_s prove that

$$T_s(w_k) = T_s(u),$$

for every s with $0 < s < k$. Passing to the limit when s tends to k, we obtain $T_k(w_k) = T_k(u)$. But since $|\Pi_h(T_k(u_h))| \leq k$, the convergence (3.10) implies that $|w_k(x)| \leq k$, hence $T_k(w_k) = w_k$. This yields $w_k = T_k(u)$, and since the limit does not depend on the subsequence, we have proved that

$$\Pi_h(T_k(u_h)) \rightharpoonup T_k(u) \quad \text{weakly in } H_0^1(\Omega), \tag{3.11}$$

when h tends to zero without extracting a subsequence.

Let us now prove that this convergence is strong.

Lebesgue's dominated convergence theorem combined with

$$|f\,\Pi_h(T_k(u_h))| \leq |f|\,k \in L^1(\Omega),$$

with the weak convergence (3.11) and with Rellich–Kondrashov's compactness theorem implies that

$$\int_\Omega f\,\Pi_h(T_k(u_h))\,dx \;\to\; \int_\Omega f\,T_k(u)\,dx.$$

Therefore passing to the limit with respect to h in (3.1) yields

$$\limsup_{h\to 0} \int_\Omega A\,\nabla\,\Pi_h(T_k(u_h))\,\nabla\,\Pi_h(T_k(u_h))\,dx \leq \int_\Omega f\,T_k(u)\,dx. \tag{3.12}$$

On the other hand, since u is the renormalized solution of (1.4), it is well known that one has

$$\int_\Omega A\,\nabla\,T_k(u)\,\nabla\,T_k(u)\,dx = \int_\Omega f\,T_k(u)\,dx\,, \tag{3.13}$$

but let us give a proof of (3.13) for completeness.

Take $S = \psi_n$ in (1.8), where

$$\forall\, s \in \mathbb{R}, \quad \psi_n(s) = \psi\left(\frac{s}{n}\right),$$

with $\psi \in C_c^1(\mathbb{R})$ a fixed function such that

$$\psi(s) = 1 \text{ if } |s| \le \frac{1}{2}, \quad \psi(s) = 0 \text{ if } |s| \ge 1.$$

Since supp $\psi_n \subset [-n, +n]$, (1.8) reads as

$$\int_\Omega A\,\nabla T_n(u)\,\nabla v \;\; \psi_n(u)\,\mathrm{d}x + \int_\Omega A\,\nabla T_n(u)\,\nabla T_n(u) \;\; \psi_n'(u)\,v\,\mathrm{d}x = \int_\Omega f\,\psi_n(u)\,v\,\mathrm{d}x,$$

where we take $v = T_k(u)$, that belongs to $H_0^1(\Omega) \cap L^\infty(\Omega)$. We obtain

$$\int_\Omega A\,\nabla T_n(u)\,\nabla T_k(u) \;\; \psi_n(u)\,\mathrm{d}x + \int_\Omega A\,\nabla T_n(u)\,\nabla T_n(u) \;\; \psi_n'(u)\,T_k(u)\,\mathrm{d}x$$
$$= \int_\Omega f\,\psi_n(u)\,T_k(u)\,\mathrm{d}x.$$

Since $\nabla T_k(u) = 0$ when $|u(x)| \ge k$, we observe that

$$A\,\nabla T_n(u)\,\nabla T_k(u)\;\psi_n(u) = A\,\nabla T_k(u)\,\nabla T_k(u),$$

when $n \ge 2k$. On the other hand, since $|\psi_n'| \le \frac{\|\psi'\|_{L^\infty(\mathbb{R})}}{n}$, one has

$$\left| \int_\Omega A\,\nabla T_n(u)\,\nabla T_n(u)\;\psi_n'(u)\;T_k(u)\,\mathrm{d}x \right|$$
$$\le \|A\|_{L^\infty(\Omega)^{d\times d}}\,\frac{\|\psi'\|_{L^\infty(\mathbb{R})}}{n}\,k \int_\Omega |\nabla T_n(u)|^2\,\mathrm{d}x,$$

where the right-hand side tends to zero when n tends to infinity owing to (1.7). Finally by Lebesgue's dominated convergence theorem

$$\int_\Omega f\,\psi_n(u)\,T_k(u)\,\mathrm{d}x \;\;\to\;\; \int_\Omega f\,T_k(u)\,\mathrm{d}x,$$

when n tends to infinity. This proves (3.13).

From (3.12) and (3.13) we deduce that

$$\limsup_{h\to 0} \int_\Omega A\,\nabla\,\Pi_h(T_k(u_h))\,\nabla\,\Pi_h(T_k(u_h))\,\mathrm{d}x \le \int_\Omega A\,\nabla\,T_k(u)\,\nabla\,T_k(u)\,\mathrm{d}x,$$

which combined with the weak convergence (3.11) implies the strong convergence (3.9). □

4 The case where f is a bounded Radon measure

In this section we consider the case where f no longer belongs to $L^1(\Omega)$, but belongs to $\mathcal{M}_b(\Omega)$, the space of Radon measures with total bounded variation. We obtain results which are weaker than in the case where f belongs to $L^1(\Omega)$, but which are still satisfactory in dimension $d = 2$ and/or when the coefficients of the matrix A are smooth.

In this section we assume that

$$f \in \mathcal{M}_b(\Omega). \tag{4.1}$$

Then, since V_h is contained in $C^0(\overline{\Omega})$, u_h is still correctly defined by (1.18). Moreover, the statement and the proof of Proposition 3.1 remain valid with $f \in L^1(\Omega)$ replaced by $f \in \mathcal{M}_b(\Omega)$, the measure $f\,\mathrm{d}x$ replaced by $\mathrm{d}f$ in (3.1) and $\|f\|_{L^1(\Omega)}$ replaced by $\|f\|_{\mathcal{M}_b(\Omega)}$ in (3.2). With these modifications estimate (3.2) is satisfied, and therefore by Theorem 2.1, u_h is bounded in $W_0^{1,q}(\Omega)$ for every q with $1 \le q < \frac{d}{d-1}$. So there exist some u and some subsequence, still denoted by h, such that for every q with $1 \le q < \frac{d}{d-1}$

$$u_h \rightharpoonup u \quad \text{weakly in } W_0^{1,q}(\Omega), \tag{4.2}$$

when h tends to zero along this subsequence.

Let $v \in C_c^\infty(\Omega)$. Taking $v_h = \Pi_h(v)$ in (1.18) yields

$$\int_\Omega A\,\nabla u_h\,\nabla \Pi_h(v)\,\mathrm{d}x = \int_\Omega \Pi_h(v)\,\mathrm{d}f,$$

in which it is easy to pass to the limit when h tends to zero owing to (4.2) and to the fact that for $v \in C_c^\infty(\Omega)$

$$\Pi_h(v) \to v \quad \text{strongly in } W^{1,\infty}(\Omega).$$

Moreover the first part of the proof of Proposition 3.3 remains valid (the fact that u_h is bounded in $W_0^{1,q}(\Omega)$ is sufficient to obtain $w_k = T_k(u)$) and implies that for every $k > 0$ one has

$$\Pi_h(T_k(u_h)) \rightharpoonup T_k(u) \quad \text{weakly in } H_0^1(\Omega),$$

when h tends to zero along the subsequence for which (4.2) holds.

We have proved the following Theorem.

Theorem 4.1 *Assume that A, $\mathcal{T}_h$ and f satisfy* (1.1), (1.2), (1.11), (1.12), (1.13), (1.17) *and* (4.1). *Then there exist a subsequence, still denoted by h, and a function u such that for every $k > 0$ and for every q with $1 \leq q < \frac{d}{d-1}$ one has*

$$\Pi_h(T_k(u_h)) \rightharpoonup T_k(u) \quad \text{weakly in } H_0^1(\Omega), \tag{4.3}$$

$$u_h \rightharpoonup u \quad \text{weakly in } W_0^{1,q}(\Omega), \tag{4.4}$$

when h tends to zero along this subsequence, where u satisfies

$$\forall k > 0, \quad T_k(u) \in H_0^1(\Omega), \tag{4.5}$$

$$\forall q \quad \text{with } 1 \leq q < \frac{d}{d-1}, \quad u \in W_0^{1,q}(\Omega), \tag{4.6}$$

$$\forall v \in C_c^\infty(\Omega), \quad \int_\Omega A \nabla u \nabla v \, \mathrm{d}x = \int_\Omega v \, \mathrm{d}f. \tag{4.7}$$

In (4.7), *one can also by density take $v \in W_0^{1,p}(\Omega)$ for every p with $p > d$.*

Let us discuss the assumptions and the results of Theorem 4.1. The hypotheses of this theorem are weaker than those of Theorem 1.3, since f is assumed to belong to $\mathcal{M}_b(\Omega)$ and not to $L^1(\Omega)$. But the conclusions also are weaker, since convergences (4.3) and (4.4) are weak and not strong convergences, and since they take place only for a subsequence. Indeed, when A and/or $\partial\Omega$ are not smooth, it is not known whether the solution of (4.5), (4.6), (4.7) is unique or not. This is the main reason why renormalized solutions, entropy solutions and solutions obtained as limit of approximations were introduced when $f \in L^1(\Omega)$. In particular, a counterexample due to Serrin [24] shows that for every q with $1 \leq q < 2$, one can exhibit a coercive matrix A_q with coefficients in $L^\infty(\Omega)$ and some function $u_q \neq 0$ such that

$$\begin{cases} u_q \in W_0^{1,q}(\Omega), \\ -\operatorname{div} A_q \nabla u_q = 0 \quad \text{in } \mathcal{D}'(\Omega). \end{cases} \tag{4.8}$$

Note however that in this counterexample q is fixed and that u_q does not satisfy $T_k(u_q) \in H_0^1(\Omega)$ for every $k > 0$. Observe also that Bénilan and Bouhsiss [3] showed that for the specific matrix A_q of this counterexample, every solution of (4.8) which also satisfies $T_k(u_q) \in H_0^1(\Omega)$ for every $k > 0$ is zero (this does not prove the uniqueness of the solution of (4.5), (4.6), (4.7), but it is a first step in this direction).

However there are cases where the solution of (4.6), (4.7) is known to be unique, and in such cases the whole sequences (and not just subsequences) converge in (4.3) and (4.4) (this is clear since then the limit u is uniquely determined independently of the subsequence). On the one hand, when A has sufficiently smooth coefficients and when $\partial\Omega$ is sufficiently smooth, the operator $u \to -\mathrm{div}\, A\, \nabla\, u$ is an isomorphism from $W_0^{1,q}(\Omega)$ onto $W^{-1,q}(\Omega)$ for every q with $1 < q < +\infty$. Therefore, in this case the solution of (4.6), (4.7) is unique. On the other hand, the two dimensional case presents some special feature. Indeed, in view of Meyer's regularity theorem [20], when $\partial\Omega$ is sufficiently smooth, the operator $u \to -\mathrm{div}\, A\, \nabla\, u$ is an isomorphism from $W_0^{1,q}(\Omega)$ onto $W^{-1,q}(\Omega)$ for every q with $2 - \delta < q < 2 + \delta$, where $\delta > 0$ only depends on the dimension d, on the open set Ω, on the coercivity coefficient α of the matrix A and on $\|A\|_{L^\infty(\Omega)^{d\times d}}$. Therefore since in the two dimensional case $q < \frac{d}{d-1}$ reads as $q < 2$, the solution of (4.6), (4.7) is unique when $\partial\Omega$ is sufficiently smooth.

In the two dimensional case, for Laplace's operator with a bounded Radon measure right-hand side, the weak convergence (4.4) of the solution of (1.18) to the (unique) solution of (4.6), (4.7) has recently been established by Gallouët and Herbin [17] by a proof based on the similarity between $\mathbb{P}_1$ finite elements and finite volume schemes and on one of their previous results [16] (see also [14]). The weak convergence (4.4) could also be proved by using the $W^{1,p}$-estimates of Brenner and Scott [6] in the two following cases: the case where $d = 2$ and where the matrix A is a general coercive matrix with $L^\infty(\Omega)$ coefficients, and the case where $d = 3$ and where the matrix A has smooth coefficients ; note that these estimates are established under the assumption that the family of triangulations is quasi-uniform in the sense of [6].

Let us finally return to the result of Theorem 4.1, which is unsatisfactory for a general coercive matrix A with $L^\infty(\Omega)$ coefficients but which has the advantage that its proof is self-contained. If we appeal to the very powerful result of Aguilera and Caffarelli [1] (we chose not to do so up to now in order to keep our results self-contained), we can obtain a much more complete result, namely the fact that in Theorem 4.1, the function u is the unique solution by transposition of problem (1.4). Indeed Aguilera and Caffarelli [1] claim that for a coercive matrix A with $L^\infty(\Omega)$ coefficients (the result is only proved for Laplace's operator in [1]), when $g \in W^{-1,p}(\Omega)$ for some $p > d$ and when $\partial\Omega$ is sufficiently smooth, the solution w_h of

$$\begin{cases} w_h \in V_h\,, \\ \forall v_h \in V_h\,, \quad \int\limits_\Omega {}^t\!A\, \nabla\, w_h\, \nabla\, v_h\, \mathrm{d}x = \langle g, v_h\rangle, \end{cases} \tag{4.9}$$

satisfies

$$w_h \to w \quad \text{in } C^{0,\gamma}(\overline{\Omega}), \tag{4.10}$$

for some $\gamma > 0$ which depends only on the data of the problem, where w is the unique solution of

$$\begin{cases} w \in H_0^1(\Omega), \\ -\mathrm{div}\, {}^t\!A\, \nabla w = g \quad \text{in } \mathcal{D}'(\Omega). \end{cases} \tag{4.11}$$

This is the discrete analogue of De Giorgi's regularity theorem. In the setting of Theorem 4.1, we have, taking $v_h = w_h$ in (1.18) (with $f\,\mathrm{d}x$ replaced by $\mathrm{d}f$) and $v_h = u_h$ in (4.9)

$$\int_\Omega w_h\,\mathrm{d}f = \int_\Omega A\,\nabla u_h\,\nabla w_h\,\mathrm{d}x = \int_\Omega {}^t\!A\,\nabla w_h\,\nabla u_h\,\mathrm{d}x = \langle g, u_h\rangle,$$

in which it is now easy to pass to the limit in view of (4.10) and of (4.4). This yields

$$\langle g, u\rangle = \int_\Omega w\,\mathrm{d}f, \tag{4.12}$$

for every $g \in W^{-1,p}(\Omega)$ with $p > d$, where w is the solution of (4.11). Equation (4.12) is nothing but Stampacchia's definition of the solution by transposition of (1.4) (see [25]). Recall that this solution is unique. Using Aguilera and Caffarelli's result, we have thus proved that the function u defined in Theorem 4.1 is the unique solution by transposition of problem (1.4), which implies that the whole sequences converge in (4.3) and (4.4). This result is much stronger than Theorem 4.1, whose proof is in contrast self-contained.

5 Error estimate

When f belongs to $L^1(\Omega)$, Theorem 1.3 proves the convergence of the finite element method, but it does not provide any error estimate. In this section we prove that when Ω and the coefficients of A are sufficiently smooth, and when f belongs to $L^r(\Omega)$ (or to the Marcinkiewicz space $L^{r,\infty}(\Omega)$) with $r > 1$, then the argument used in the proof of Theorem 3.2 also provides an error estimate in dimension 2 and 3.

To simplify the presentation, we assume in this section that either $d = 2$ or $d = 3$, that Ω is a convex polyhedron, that $\Omega_h = \Omega$ for every $h > 0$, and that the coefficients of A belong to $W^{1,\infty}(\Omega)$. In this case, it is well known that for every $g \in L^2(\Omega)$ the unique solution w_h of problem (1.18) with right-hand side g satisfies

$$\|w_h - w\|_{H_0^1(\Omega)} \le C\,h\,\|g\|_{L^2(\Omega)}, \tag{5.1}$$

where w is the unique weak solution of problem (1.9) with right-hand side g, and where the constant $C > 0$ is independent of h and g (but depends on Ω, α, $\|A\|_{W^{1,\infty}(\Omega)^{d\times d}}$ and on the parameter σ which measures the regularity of the family of triangulations, see (1.13)).

We also assume in this section that f belongs to the Marcinkiewicz space $L^{r,\infty}(\Omega)$ for some r with $1 < r < 2$ (this holds in particular if f belongs to $L^r(\Omega)$). For every $\varepsilon > 0$, we set

$$f^\varepsilon = T_{\frac{1}{\varepsilon}}(f),$$

which belongs to $L^\infty(\Omega) \subset L^2(\Omega)$, and we denote by u_h^ε the solution of (1.18) with right-hand side f^ε. Defining also u^ε as the solution of (3.6), we write for every q with $1 \leq q < \frac{d}{d-1}$

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq \|u_h - u_h^\varepsilon\|_{W_0^{1,q}(\Omega)} + \|u_h^\varepsilon - u^\varepsilon\|_{W_0^{1,q}(\Omega)} + \|u^\varepsilon - u\|_{W_0^{1,q}(\Omega)}. \quad (5.2)$$

From (5.1) applied to $g = f^\varepsilon$, $w_h = u_h^\varepsilon$ and $w = u^\varepsilon$, and from the continuous imbedding of $H_0^1(\Omega)$ in $W_0^{1,q}(\Omega)$, we have for a new constant C (which depends on q, Ω, α, $\|A\|_{W^{1,\infty}(\Omega)^{d\times d}}$ and σ)

$$\|u_h^\varepsilon - u^\varepsilon\|_{W_0^{1,q}(\Omega)} \leq C\,h\,\|f^\varepsilon\|_{L^2(\Omega)}.$$

Using then (3.4) and (3.8), we deduce that for a new constant C, which is independent of ε, h and f (but depends on d, q, Ω, α, $\|A\|_{W^{1,\infty}(\Omega)^{d\times d}}$ and σ), one has

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq C\left(\|f - f^\varepsilon\|_{L^1(\Omega)} + h\,\|f^\varepsilon\|_{L^2(\Omega)}\right). \quad (5.3)$$

We now estimate the right-hand side of this inequality by using the coarea formula, namely

$$\|g\|^p_{L^p(\Omega)} = p\int_0^{+\infty} t^{p-1}\,|\{x \in \Omega :\ |g(x)| \geq t\}|\,\mathrm{d}t,$$

which gives

$$\begin{cases} \|f - f^\varepsilon\|_{L^1(\Omega)} = \displaystyle\int_0^{+\infty} |\{x \in \Omega :\ |f(x) - T_{\frac{1}{\varepsilon}}(f)(x)| \geq t\}|\,\mathrm{d}t \\ \qquad = \displaystyle\int_0^{+\infty} |\{x \in \Omega :\ (|f(x)| - \tfrac{1}{\varepsilon}) \geq t\}|\,\mathrm{d}t \\ \qquad = \displaystyle\int_{\frac{1}{\varepsilon}}^{+\infty} |\{x \in \Omega :\ |f(x)| \geq t\}|\,\mathrm{d}t, \end{cases} \quad (5.4)$$

$$\begin{cases} \|f^\varepsilon\|^2_{L^2(\Omega)} = 2\displaystyle\int_0^{+\infty} t\,|\{x \in \Omega :\ |T_{\frac{1}{\varepsilon}}(f)(x)| \geq t\}|\,\mathrm{d}t \\ \qquad = 2\displaystyle\int_0^{\frac{1}{\varepsilon}} t\,|\{x \in \Omega :\ |f(x)| \geq t\}|\,\mathrm{d}t. \end{cases} \quad (5.5)$$

By the definition (0.7) of the norm in the Marcinkiewicz space $L^{r,\infty}(\Omega)$, we have

$$|\{x \in \Omega : \ |f(x)| \geq t\}| \leq \frac{\|f\|^r_{L^{r,\infty}(\Omega)}}{t^r},$$

and thus

$$\begin{cases} \|f - f^\varepsilon\|_{L^1(\Omega)} \leq \dfrac{1}{r-1}\, \varepsilon^{r-1}\, \|f\|^r_{L^{r,\infty}(\Omega)}, \\ \|f^\varepsilon\|_{L^2(\Omega)} \leq \sqrt{\dfrac{2}{2-r}}\, \dfrac{1}{\varepsilon^{1-\frac{r}{2}}}\, \|f\|^{\frac{r}{2}}_{L^{r,\infty}(\Omega)}. \end{cases} \tag{5.6}$$

Then (5.3) gives

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq C\left(\frac{1}{r-1}\, \varepsilon^{r-1}\, \|f\|^r_{L^{r,\infty}(\Omega)} + \sqrt{\frac{2}{2-r}}\, \frac{h}{\varepsilon^{1-\frac{r}{2}}}\, \|f\|^{\frac{r}{2}}_{L^{r,\infty}(\Omega)}\right).$$

Taking in this inequality $\varepsilon = \frac{h^{\frac{2}{r}}}{\|f\|_{L^{r,\infty}(\Omega)}}$ yields, for every q with $1 \leq q < \frac{d}{d-1}$ and for every $h > 0$

$$\|u_h - u\|_{W_0^{1,q}(\Omega)} \leq C(d, q, r, \Omega, \alpha, \|A\|_{W^{1,\infty}(\Omega)^{d\times d}}, \sigma)\, h^{2(1-\frac{1}{r})}\, \|f\|_{L^{r,\infty}(\Omega)}. \tag{5.7}$$

We have proved the following result.

Theorem 5.1 *Under the assumptions of Theorem 1.3, if we further assume that either $d = 2$ or $d = 3$, that $f \in L^{r,\infty}(\Omega)$ for some r with $1 < r < 2$, that Ω is a convex polyhedron, that $\Omega_h = \Omega$ and that the coefficients of the matrix A belong to $W^{1,\infty}(\Omega)$, then we have the error estimate (5.7).*

To the best of our knowledge, this estimate is new in the case where r is close to 1, but also in the case where $L^r(\Omega) \subset H^{-1}(\Omega)$. Indeed when r is such that $L^r(\Omega) \subset H^{-1}(\Omega)$, i.e. when $r > 1$ if $d = 2$ or when $r \geq 6/5$ if $d = 3$, one can interpolate between the estimate (5.1) for $g \in L^2(\Omega)$ and the easy estimate for $g \in H^{-1}(\Omega)$

$$\|w_h - w\|_{H_0^1(\Omega)} \leq \|w_h\|_{H_0^1(\Omega)} + \|w\|_{H_0^1(\Omega)} \leq \frac{2}{\alpha}\|g\|_{H^{-1}(\Omega)}.$$

This interpolation yields

$$\|w_h - w\|_{H_0^1(\Omega)} \leq C_\delta\, h^{2(1-\frac{1}{r})-\delta}\, \|g\|_{L^r(\Omega)} \quad \text{for every } \delta > 0 \quad \text{if } d = 2,$$

$$\|w_h - w\|_{H_0^1(\Omega)} \leq C\, h^{3(\frac{5}{6}-\frac{1}{r})}\, \|g\|_{L^r(\Omega)} \quad \text{if } d = 3.$$

If one compares this interpolation estimate with (5.7), the order of convergence is higher in (5.7) but the norm under consideration is weaker since the space $W_0^{1,q}(\Omega)$ is larger than $H_0^1(\Omega)$.

To conclude this section let us recall two error estimates obtained in a setting different of (but related to) the present one. In dimension $d = 2$ for Laplace's equation and f a Dirac mass, Scott [23] proved that for a quasi-uniform family of triangulations one has

$$\|u_h - u\|_{L^2(\Omega)} \leq Ch,$$

while in the same setting, when f is a bounded Radon measure, Clain [10] proved that one has

$$\|u_h - u\|_{W_0^{1,p}(\Omega)} \leq C h^s \|\mu\|_{\mathcal{M}_b(\Omega)},$$

for every s with $0 < s < 1$ and every p with $1 < p < \frac{2}{1+s}$. These estimates are neither stronger nor weaker than (5.7).

6 Examples of triangulations and matrices

In this section, we present examples of families of triangulations and of matrices for which all the assumptions of Theorem 1.3, namely (1.1), (1.2), (1.11), (1.12), (1.13) and (1.17), are satisfied. After some general considerations (which are standard), we successively consider the case where the matrix A is the identity, the case of a coercive matrix with constant coefficients, the case where A is the product of a coercive matrix with constant coefficients by a scalar function, and finally a pertubation of the latest case.

6.1 General considerations

For every d-simplex T of $\mathcal{T}_h$ and for every vertex a_i of T, we denote in this section by F_i the face opposite to a_i and by n_i the exterior (to the d-simplex T) unit normal to the face F_i.

Our results are based on the following proposition, whose proof is a straightforward adaptation of a classical result (for Laplace's operator see e.g. Drăgănescu et al. [13] and the references therein).

Proposition 6.1 *Assume that the matrix A satisfies* (1.1) *and* (1.2). *If the triangulation $\mathcal{T}_h$ is such that for every $T \in \mathcal{T}_h$*

$$\begin{cases} \forall i \in I,\ \forall j \in I \cup B, \quad j \neq i, \\ \displaystyle\sum_{\substack{T\in\mathcal{T}_h \\ a_i,\, a_j \in T}} \frac{1}{d^2} \frac{|F_i|\,|F_j|}{|T|^2} \left(\int_T A \,\mathrm{d}x \right) n_i\, n_j \leq 0, \end{cases} \tag{6.1}$$

then (1.17) *is satisfied. In particular, if for every interior vertex a_i and every (interior or boundary) vertex a_j of T with $a_j \neq a_i$, i.e. for every $i \in I$ and $j \in I \cup B$ with $j \neq i$, one has*

$$\left(\int\limits_T A \,\mathrm{d}x\right) n_i\, n_j \leq 0, \tag{6.2}$$

then (1.17) *is satisfied.*

Proof We give it here for the reader's convenience.

For this proof we extend the definition (1.16) of Q_{ij}, which was given only for i and j in I, to the case where i and j belong to $I \cup B$ by setting

$$\forall\, i, j \in I \cup B, \quad Q_{ij} = \int\limits_{\Omega_h} A \,\nabla \varphi_i \,\nabla \varphi_j \,\mathrm{d}x.$$

(In (1.16) we did not define Q_{ij} for i and/or j in B since these values are not required in the statement of hypothesis (1.17); these new Q_{ij} coincide with the Q_{ij} defined by (1.16) when i and j belong to I.)

First step. Since $\sum\limits_{j \in I \cup B} \varphi_j(x) = 1$ in Ω_h (see (1.15)), one has

$$\sum_{j \in I \cup B} \nabla \varphi_j(x) = 0 \quad \text{in } \Omega_h.$$

For every $i \in I \cup B$ this implies that

$$\sum_{j \in I \cup B} Q_{ij} = \int\limits_{\Omega_h} A \,\nabla \varphi_i \sum_{j \in I \cup B} \nabla \varphi_j \,\mathrm{d}x = 0,$$

and therefore for every $i \in I$

$$0 = \sum_{j \in I \cup B} Q_{ij} = Q_{ii} + \sum_{\substack{j \in I \\ j \neq i}} Q_{ij} + \sum_{j \in B} Q_{ij}. \tag{6.3}$$

Observe that for $i = j \in I$, one has

$$Q_{ii} = \int\limits_{\Omega_h} A \,\nabla \varphi_i \,\nabla \varphi_i \,\mathrm{d}x \geq 0.$$

If we assume that

$$\forall\, i \in I, \quad \forall\, j \in I \cup B, \;\; j \neq i, \;\; Q_{ij} \leq 0, \tag{6.4}$$

then one has $Q_{ij} = -|Q_{ij}|$ for every $i \in I$ and every $j \in I \cup B$ with $j \neq i$. Therefore, for every $i \in I$

$$Q_{ii} - \sum_{\substack{j \in I \\ j \neq i}} |Q_{ij}| = Q_{ii} + \sum_{\substack{j \in I \\ j \neq i}} Q_{ij} = -\sum_{j \in B} Q_{ij} = \sum_{j \in B} |Q_{ij}| \geq 0,$$

which proves that the matrix Q satisfies (1.17) when (6.4) holds.

Second step. Let T be a d-simplex of $\mathcal{T}_h$. When a_i is a vertex of T, one has

$$\nabla \varphi_i = -\frac{1}{d} \frac{|F_i|}{|T|} n_i \quad \text{in } T;$$

indeed $\varphi_i = 0$ on F_i, and so $\nabla \varphi_i$ is orthogonal to F_i; since $\varphi_i(a_i) = 1$, one has $\nabla \varphi_i = -\frac{1}{h_i} n_i$, where h_i is the distance of a_i to the hyperplane which contains F_i; finally $|T| = \frac{1}{d}|F_i| h_i$. Therefore, when both a_i and a_j are vertices of T, one has

$$\int_T A \nabla \varphi_i \nabla \varphi_j \, \mathrm{d}x = \frac{1}{d^2} \frac{|F_i|\,|F_j|}{|T|^2} \left(\int_T A \, \mathrm{d}x \right) n_i \, n_j. \tag{6.5}$$

On the other hand, when a_i and/or a_j is not a vertex of T, then φ_i and/or φ_j is zero on T, and then

$$\int_T A \nabla \varphi_i \nabla \varphi_j \, \mathrm{d}x = 0.$$

This implies that for every i and j in $I \cup B$, one has

$$\begin{cases} Q_{ij} = \displaystyle\int_{\Omega_h} A \nabla \varphi_i \nabla \varphi_j \, \mathrm{d}x = \sum_{\substack{T \in \mathcal{T}_h \\ a_i,\, a_j \in T}} \int_T A \nabla \varphi_i \nabla \varphi_j \, \mathrm{d}x \\ \quad = \displaystyle\sum_{\substack{T \in \mathcal{T}_h \\ a_i,\, a_j \in T}} \frac{1}{d^2} \frac{|F_i|\,|F_j|}{|T|^2} \left(\int_T A \, \mathrm{d}x \right) n_i \, n_j. \end{cases} \tag{6.6}$$

Third step. In view of (6.6), assumption (6.1) is nothing but (6.4) and the first result of Proposition 6.1 follows from the first step above. On the other hand, hypothesis (6.2) immediately implies that (6.1) holds true, which proves the second result of Proposition 6.1. □

Remark 6.2 The first step of the above proof establishes that (6.4) implies (1.17). Actually condition (6.4), i.e. $Q_{ij} \leq 0$ for $j \neq i$, is also necessary for (1.17) to hold, at least as far as "strictly interior vertices" a_i are concerned.

Let us indeed define the strictly interior vertices as those vertices a_i for which, for every d-simplex $T \in \mathcal{T}_h$ with $a_i \in T$, all the vertices of T are interior vertices. Since $Q_{ij} = 0$ when $j \neq i$ and when a_i and a_j do not belong to a same d-simplex T, one has $Q_{ij} = 0$ for every $j \in B$ when a_i is a strictly interior vertex; then (6.3) reads as

$$0 = Q_{ii} + \sum_{\substack{j \in I \\ j \neq i}} Q_{ij}.$$

But $Q_{ij} \geq -|Q_{ij}|$ for every $j \neq i$ and therefore one has

$$Q_{ii} - \sum_{\substack{j \in I \\ j \neq i}} |Q_{ij}| \leq 0,$$

when a_i is a strictly interior vertex. If (1.17) holds true, we necessarily have for every strictly interior vertex a_i

$$Q_{ii} - \sum_{\substack{j \in I \\ j \neq i}} |Q_{ij}| = 0,$$

and therefore $Q_{ij} = -|Q_{ij}|$, i.e. $Q_{ij} \leq 0$ for every $j \neq i$ when a_i is a strictly interior vertex.

We have therefore proved that condition (6.4) is a sufficient condition for (1.17) to hold, and that this condition is necessary and sufficient when a_i is a strictly interior vertex. Let us finally note that (6.1) is equivalent to (6.4), but that (6.2) is only a sufficient condition for (6.1) to hold. □

Let us now present some examples of matrices A and of regular families of triangulations $\mathcal{T}_h$ for which assumption (1.17) is satisfied.

6.2 The case where A is the identity matrix

Consider first the case where the matrix A is the identity Id, i.e. the case where the operator is Laplace's operator $-\Delta$. Then condition (6.2), which implies (1.17), is satisfied if and only if

$$\forall i \in I, \quad \forall j \in I \cup B \quad \text{with } j \neq i, \;\; n_i n_j \leq 0. \tag{6.7}$$

In the two dimensional case, (6.7) is satisfied if every inner angle of every triangle is acute, i.e. not larger than $\pi/2$. In the three dimensional case, (6.7) is satisfied if every inner dihedral angle of every tetrahedron is acute. When $d \geq 4$, we will say that the inner angles are acute if $n_i n_j \leq 0$.

We have proved the following well-known result.

Proposition 6.3 *In the d-dimensional case, (1.17) holds for Laplace's operator if every inner angle of every d-simplex of $\mathcal{T}_h$ is acute, i.e. if (6.7) holds.*

An example of family of triangulations which enjoys all the properties required in Sect. 1 for Laplace's operator is therefore obtained by triangulating $\mathbb{R}^d$ by a regular family of triangulations with acute inner angles, and by taking for $\mathcal{T}_h$ the union of the d-simplices T which satisfy $T \subset \overline{\Omega}$.

For $d = 2$, one such family of triangulations is obtained by covering $\mathbb{R}^2$ by squares of vertices (ih, jh) with $i, j \in \mathbb{Z}$, and then by subdividing each square $\{(x_1, x_2) : i \leq x_1 \leq (i+1)h, j \leq x_2 \leq (j+1)h\}$ into 2 triangles along its first or its second diagonal. Other triangulations (e.g. by equilateral triangles) are of course possible.

For $d = 3$, one such family of triangulations is obtained by covering $\mathbb{R}^3$ by cubes of vertices (ih, jh, kh) with $i, j, k \in \mathbb{Z}$, and then by subdividing each cube $\{(x_1, x_2, x_3) : ih \leq x_1 \leq (i+1)h, jh \leq x_2 \leq (j+1)h, kh \leq x_3 \leq (k+1)h\}$ into 6 tetrahedra obtained by slicing each cube along the three planes defined in the cube $(0, h)^3$ by $x_1 = x_2$, $x_2 = x_3$ and $x_3 = x_1$. It is easy to see that condition (6.7) is satisfied for this subdivision. Other subdivisions of the cube (e.g. the subdivisions into 6 similar thetrahedra where the diagonal $x_1 = x_2 = x_3$ of the cube $(0, h)^3$ is replaced by one of the other four diagonals of the cube, but also subdivisions into 5 tetrahedra) are also possible.

In order to ensure that (1.17) holds true, one can of course use, in place of the sufficient condition (6.2), the condition (6.1), which is weaker and almost necessary (see Remark 6.2). In the two dimensional case, for two given vertices a_i and a_j with $i \in I$, $j \in I \cup B$ and $j \neq i$, there is either no triangle T with $a_i \in T$ and $a_j \in T$, or a_i and a_j belong to the same triangle; in this case the edge $[a_i\, a_j]$ is not included in $\partial\Omega_h$ and there are exactly two triangles T^+ and T^- which share the two vertices a_i and a_j. When $A = Id$, condition (6.1) is nothing but

$$\begin{cases} \forall\, i \in I, \quad \forall\, j \in I \cup B, \quad j \neq i, \\ \\ \dfrac{1}{2^2}\, \dfrac{|F_i^+|\,|F_j^+|}{|T^+|}\, n_i^+\, n_j^+ + \dfrac{1}{2^2}\, \dfrac{|F_i^-|\,|F_j^-|}{|T^-|}\, n_i^-\, n_j^- \leq 0. \end{cases} \tag{6.8}$$

Denote by θ^+ the inner angle facing the edge $[a_i\, a_j]$ in T^+ and by h_i^+ the distance of a_i to the straight line which contains F_i^+. Then

$$|T^+| = \frac{1}{2}|F_i^+|h_i^+ = \frac{1}{2}|F_i^+|\,|F_j^+| \sin\theta^+,$$

$$n_i^+\, n_j^+ = \cos(\pi - \theta^+) = -\cos\theta^+,$$

$$\frac{1}{2^2}\, \frac{|F_i^+|\,|F_j^+|}{|T^+|}\, n_i^+\, n_j^+ = -\frac{1}{2}\, \frac{\cos\theta^+}{\sin\theta^+}.$$

Therefore if θ^- denotes the inner angle facing the edge $[a_i\, a_j]$ in T^-, condition (6.8) becomes

$$-\frac{1}{2}\frac{\cos\,\theta^+}{\sin\theta^+} - \frac{1}{2}\frac{\cos\,\theta^-}{\sin\,\theta^-} = -\frac{1}{2}\frac{\sin\,(\theta^+ + \theta^-)}{\sin\,\theta^+\,\sin\theta^-} \le 0.$$

Since θ^+ and θ^- belong to $(0, \pi)$, (6.8) is equivalent to

$$\forall i \in I, \quad \forall j \in I \cup B, \quad i \neq j, \quad \theta^+ + \theta^- \le \pi.$$

In the two dimensional case, we have thus proved the following classical result (see e.g. Drăgănescu et al. [13] and the references therein).

Proposition 6.4 *In the two dimensional case,* (1.17) *holds for Laplace's operator if for every edge* $[a_i\, a_j]$ *of the triangulation which is not included in* $\partial\Omega_h$*, the sum of the two inner angles* θ^+ *and* θ^- *facing* $[a_i\, a_j]$ *is not larger than* π*.*

In the case where strictly interior vertics are concerned, the requirement of Proposition 6.4 is necessary and sufficient for (1.17) to hold (see Remark 6.2).

In the two dimensional case, a triangulation which satisfies the requirement of Proposition 6.4 is called a Delaunay triangulation, see e.g. Frey and George [15] or George and Borouchaki [18].

6.3 The case where A is a coercive matrix with constant coefficients

Consider now the case where A is a coercive matrix with constant coefficients. Then we can always reduce ourselves to the case where A is a symmetric matrix since for every u

$$\begin{aligned}-\mathrm{div}\, A\,\nabla\, u &= -\sum_{k,\ell} A_{k\ell}\frac{\partial^2 u}{\partial x_k \partial x_\ell} = -\sum_{k,\ell}\frac{A_{k\ell} + A_{\ell k}}{2}\,\frac{\partial^2 u}{\partial x_k \partial x_\ell}\\ &= -\mathrm{div}\left(\frac{A + {}^t\!A}{2}\right)\nabla\, u.\end{aligned}$$

Using an orthonormal change of basis, we write A as

$$A = {}^t\!MDDM,$$

with M an orthogonal matrix and D a diagonal coercive matrix. Then condition (6.2), which implies (1.17), is satisfied if and only if

$$\forall\, i \in I, \quad \forall\, j \in I \cup B \quad \text{with} \quad i \neq j, \;\; (DM\, n_i)(DM\, n_j) \le 0. \tag{6.9}$$

On the other hand, for a triangulation $\mathcal{T}_h$, consider the triangulation $\hat{\mathcal{T}}_h$ obtained by the change of variables $\hat{x} = D^{-1}M\,x$, namely

$$\hat{\mathcal{T}}_h = \{\hat{T} : \hat{T} = D^{-1}M\,(T) \quad \text{with } T \in \mathcal{T}_h\}.$$

When a_i is a vertex of T and φ_i the basis function associated with a_i, we define $\hat{\varphi}_i$ on $\hat{T}$ by

$$\hat{\varphi}_i(\hat{x}) = \hat{\varphi}_i(D^{-1}Mx) = \varphi_i(x) = \varphi_i({}^tMD\hat{x}).$$

Then $\hat{\varphi}_i$ is the basis function associated with $\hat{a}_i = D^{-1}Ma_i$, and for every pair of vertices a_i and a_j of T, one has, since $A = {}^tMDDM$

$$\begin{aligned}\int_T A\,\nabla\varphi_i\,\nabla\varphi_j\,\mathrm{d}x &= \int_{\hat{T}} A\,{}^tMD^{-1}\,\nabla\hat{\varphi}_i\,{}^tMD^{-1}\,\nabla\,\hat{\varphi}_j\,|\det D|\,\mathrm{d}\hat{x}\\ &= |\det D|\int_{\hat{T}} \nabla\,\hat{\varphi}_i\,\nabla\,\hat{\varphi}_j\,\mathrm{d}\hat{x}.\end{aligned}$$

Therefore, in view of (6.5), $\left(\int_T A\,\mathrm{d}x\right)n_in_j = |T|A\,n_in_j$ and $\hat{n}_i\hat{n}_j$ have the same sign.

Actually by the change of variables $\hat{x} = D^{-1}M\,x$, we have transformed the problem (1.4) into the problem

$$\begin{cases} -\Delta\hat{u} = \hat{f} \text{ in } \hat{\Omega},\\ \hat{u} = 0 \qquad \text{on } \partial\hat{\Omega},\end{cases}$$

for which we will consider an acute triangulation $\hat{\mathcal{T}}_h$ of $\hat{\Omega}$.

We have proved the following result.

Proposition 6.5 *In the d-dimensional case,* (1.17) *holds for a given symmetric coercive matrix with constant coefficients $A = {}^tMDDM$ if* (6.9) *holds, or in other words if every inner angle of every d-simplex of the triangulation $\hat{\mathcal{T}}_h$ obtained from $\mathcal{T}_h$ by the change of variables $\hat{x} = D^{-1}Mx$ is acute.*

6.4 The case where A is the product of a coercive matrix with constant coefficients by a scalar function, and a perturbation

More generally, consider the case where A is a matrix of the form

$$A(x) = a(x)C,$$

where

$$a \in L^\infty(\Omega), \quad \text{a.e. } x \in \Omega, \quad a(x) \geq \alpha,$$

for some $\alpha > 0$ and where C is a symmetric coercive matrix with constant coefficients, with $C = {}^tMDDM$ as before. Then

$$\left(\int_T A\,\mathrm{d}x\right) n_i n_j = \left(\int_T a(x) C\,\mathrm{d}x\right) n_i n_j = \left(\int_T a(x)\mathrm{d}x\right) C n_i n_j$$

shows that

$$\left(\int_T A\,\mathrm{d}x\right) n_i n_j \text{ has the same sign as } C n_i n_j = (DM\,n_i)(DM\,n_j).$$

Therefore every triangulation which satisfies (6.2) for the matrix C also satisfies (6.2) for the matrix $A = a(x)C$, and condition (6.2) is here equivalent to (6.9). This condition is satisfied if the triangulation obtained by the change of variables $\hat{x} = D^{-1}Mx$ has acute inner angles.

Consider finally a (small) perturbation of the previous case, i.e. a matrix A of the form

$$A(x) = a(x)C + a(x)E(x), \tag{6.10}$$

with

$$a \in L^\infty(\Omega), \quad \text{a.e. } x \in \Omega, \quad a(x) \geq \alpha, \quad C = {}^tMDDM,$$

for some $\alpha > 0$, where M is some orthogonal matrix and D is some coercive diagonal matrix, both with constant coefficients. Assume that the triangulation $\hat{\mathcal{T}}_h$ obtained by the change of variables $\hat{x} = D^{-1}Mx$ has strictly δ-acute inner angles for some $\delta > 0$, in the sense that, for every $i \in I$ and every $j \in I \cup B$ with $i \neq j$, one has

$$(DMn_i)(DMn_j) \leq -\delta. \tag{6.11}$$

Then if

$$\|E(x)\|_{L^\infty(\Omega)^{d\times d}} \leq \delta,$$

condition (6.2) is satisfied since

$$\begin{aligned}\left(\int_T A\,\mathrm{d}x\right) n_i n_j &= \left(\int_T a(x)\mathrm{d}x\right) C\, n_i n_j + \left(\int_T a(x)E(x)\mathrm{d}x\right) n_i n_j \\ &\leq \left(\int_T a(x)\mathrm{d}x\right) (DMn_i)(DMn_j) + \left(\int_T a(x)\mathrm{d}x\right) \|E\|_{L^\infty(\Omega)^{d\times d}} \\ &\leq \left(\int_T a(x)\mathrm{d}x\right) (-\delta+\delta) = 0.\end{aligned}$$

Note also that the matrix A is coercive when $\|E\|_{L^\infty(\Omega)^{d\times d}}$ is sufficiently small, since denoting by $\beta > 0$ the coercivity coefficient of C, one has for every $\xi \in \mathbb{R}^d$

$$\begin{aligned}A(x)\xi\xi &= a(x)C\xi\xi + a(x)E(x)\xi\xi \\ &\geq a(x)\beta|\xi|^2 - a(x)\|E\|_{L^\infty(\Omega)^{d\times d}}|\xi|^2 \\ &= a(x)\left(\beta - \|E\|_{L^\infty(\Omega)^{d\times d}}\right)|\xi|^2 \geq \alpha\,\frac{\beta}{2}\,|\xi|^2,\end{aligned}$$

when $\|E\|_{L^\infty(\Omega)^{d\times d}} \leq \beta/2$.

We have proved the following result.

Proposition 6.6 *In the d-dimensional case, hypotheses* (1.11), (1.12), (1.13) *and* (1.17) *hold for a matrix A of the form* (6.10) *and for a family of triangulations $\mathcal{T}_h$ when the family of triangulations $\hat{\mathcal{T}}_h$ obtained by the change of variables $\hat{x} = D^{-1}Mx$ is regular and has δ-acute inner angles for some $\delta > 0$ (i.e. when the family of triangulations $\mathcal{T}_h$ satisfies* (6.11)*) and when $\|E\|_{L^\infty(\Omega)^{d\times d}}$ is sufficiently small.*

An example of family of triangulations which enjoys all the properties required in Sect. 1 for a matrix A of the form (6.10) with $\|E\|_{L^\infty(\Omega)^{d\times d}}$ sufficiently small is obtained by triangulating $\mathbb{R}^d$ by a regular family of triangulations $\mathcal{T}_h$ such that the transformed family of triangulations $\hat{\mathcal{T}}_h$ has δ-acute inner angles for some $\delta > 0$, and by taking for $\mathcal{T}_h$ the union of the d-simplices T which satisfy $T \subset \overline{\Omega}$.

Unfortunately, for a general coercive matrix with coefficients in $L^\infty(\Omega)$, it is not clear for us whether one can always construct a regular family of triangulations which satisfy (6.2) or (1.17) (recall that (6.2) implies (1.17) but not conversely).

Let us finally mention that in [7] we will construct in the two dimensional case a regular family of triangulations for any coercive symmetric matrix A with $L^\infty(\Omega)$ coefficients, when the matrix A given by

$$A(x) = \begin{pmatrix} A_{11}(x) & A_{12}(x) \\ A_{12}(x) & A_{22}(x) \end{pmatrix},$$

satisfies

$$|A_{12}(x)| \leq \inf\Big(A_{11}(x),\ A_{22}(x)\Big) \quad \text{a.e. } x \in \Omega.$$

Acknowledgements The authors thank Michel Crouzeix for a very illuminating discussion. The present work was made possible by reciprocal visits of the Spanish authors to Paris and of the French authors to Seville. The authors thank the various institutions which provided the corresponding financial supports. The research of J. Casado-Díaz was partially supported by the Spanish Ministerio de Ciencia y Tecnología Grant BFM2002-00672. The research of T. Chacón Rebollo was partially supported by the Spanish Ministerio de Ciencia y Tecnología Grant BFM2003-07530-C02-01 and by a Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme. The research of M. Gómez Marmol was partially supported by the Spanish Ministerio de Ciencia y Tecnología Grant BFM2003-07530-C02-01.

References

1. Aguilera, N.E., Caffarelli, L.A.: Regularity results for discrete solutions of second order elliptic problems in the finite element method. Calcolo **23**, 327–353 (1986)
2. Bénilan, P., Boccardo, L., Gallouët, T., Gariepy, R., Pierre, M., Vázquez, J.L.: An L^1-theory of existence and uniqueness of solutions of nonlinear elliptic equations. Ann. Scuola Norm. Sup. Pisa **22**, 241–273 (1995)
3. Bénilan, P., Bouhsiss, F.: Une remarque sur l'unicité des solutions pour l'opérateur de Serrin. C. R. Acad. Sci. Paris Sér. I **325**, 611–616 (1997)
4. Boccardo, L., Gallouët, T.: Nonlinear elliptic and parabolic equations involving measure data. J. Funct. Anal. **87**, 149–169 (1989)
5. Boccardo, L., Gallouët, T.: Nonlinear elliptic equations with right-hand side measures. Comm. Partial Differ. Equ. **17**, 641–655 (1992)
6. Brenner, S.C., Scott, L.R.: The mathematical theory of finite element methods. Texts in Applied Mathematics, vol. **15**. Springer, Berlin Heidelberg New York (1994)
7. Casado-Díaz, J., Chacón Rebollo, T., Girault, V., Gómez Mármol, M., Murat, F.: A condition with ensures the discrete maximum principle for $-\text{div}\, A(x)\, \nabla$ in dimension 2 (to appear)
8. Ciarlet, P.G.: The finite element method for elliptic problems. North-Holland, Amsterdam (1978)
9. Ciarlet, P.G., Raviart, P.A.: Maximum principle and uniform convergence for the finite element method. Comput. Meth. Appl. Mech. Eng. **2**, 17–31 (1973)
10. Clain, S.: Finite element approximations for the Laplace operator with a right-hand side measure. Math. Models Methods Appl. Sci. **6**, 713–719 (1995)
11. Dal Maso, G., Murat, F., Orsina, L., Prignet, A.: Renormalized solutions of elliptic equations with general measure data. Ann. Scuola Norm. Sup. Pisa **28**, 741–808 (1999)
12. Dall'Aglio, A.: Approximated solutions of equations with L^1 data. Application to the H-convergence of quasi-linear parabolic equations. Ann. Mat. Pura Appl. **170**, 207–240 (1996)
13. Drăgănescu, A., Dupont, T.F., Scott, L.R.: Failure of the discrete maximum principle for an elliptic finite element problem. Math. Comp. **74**, 1–23 (2004)
14. Droniou, J., Gallouët, T., Herbin, R.: A finite volume scheme for a noncoercive elliptic equation with measure data. SIAM J. Numer. Anal. **41**, 1997–2031 (2003)
15. Frey, P., George, P.-L.: Maillages. Hermes, Paris (1999)
16. Gallouët, T., Herbin, R.: Finite volume approximation of elliptic problems with irregular data. In: Vilsmeier, R., Benkhaldoun, F., Hänel, D. (eds.), Finite volumes for complex applications II, Hermes, Paris (1999), pp. 155–162
17. Gallouët, T., Herbin, R.: Convergence of linear finite elements for diffusion equations with measure data. C. R. Math. Acad. Sci. Paris Sér. I **338**, 81–84 (2004)
18. George, P.-L., Borouchaki, H.: Delaunay triangulation and meshing. Application to finite elements. Hermes, Paris (1998)
19. Lions, P.-L., Murat, F.: Solutions renormalisées d'équations elliptic non linéaires. (to appear)

20. Meyers, N.G.: An L^p estimate for the gradient of solutions of second order elliptic divergence equations. Ann. Scuola Norm. Sup. Pisa **17**, 189–206 (1963)
21. Murat, F.: Soluciones renormalizadas de edp elípticas no lineales. Publication 93023 du Laboratoire d'Analyse Numérique de l'Université Paris VI (1993)
22. Murat, F.: Équations elliptiques non linéaires avec second membre L^1 ou mesure. In: Actes du 26ème Congrès national d'analyse numérique (Les Karellis, juin 1994), Université de Lyon I (1994), pp. A12–A24
23. Scott, L.R.: Finite element convergence for singular data. Numer. Math. **21**, 317–327 (1973)
24. Serrin, J.: Pathological solutions of elliptic differential equations. Ann. Scuola Norm. Sup. Pisa **18**, 385–387 (1964)
25. Stampacchia, G.: Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. Ann. Inst. Fourier (Grenoble) **15**, 189–258 (1965)