# VIRTUAL ELEMENT METHOD FOR GENERAL SECOND ORDER ELLIPTIC PROBLEMS ON POLYGONAL MESHES

L. BEIRÃO da VEIGA

*Dipartimento di Matematica, Università di Milano, Via Saldini 50, 20133 Milano (Italy),*
*and IMATI del CNR, Via Ferrata 1, 27100 Pavia, (Italy)*
*lourenco.beirao@unimi.it*


F. BREZZI

*IUSS, Piazza della Vittoria 15, 27100 Pavia (Italy),*
*and IMATI del CNR, Via Ferrata 1, 27100 Pavia, (Italy)*
*brezzi@imati.cnr.it*


L.D. MARINI

*Dipartimento di Matematica, Università di Pavia,*
*and IMATI del CNR, Via Ferrata 1, 27100 Pavia, (Italy)*
*marini@imati.cnr.it*


A. RUSSO

*Dipartimento di Matematica e Applicazioni, Università di Milano-Bicocca, via Cozzi 57, 20125*
*Milano (Italy),*
*and IMATI del CNR, Via Ferrata 1, 27100 Pavia, (Italy)*
*alessandro.russo@unimib.it*

We consider the discretization of a boundary value problem for a general linear second-order elliptic operator with smooth coefficients using the Virtual Element approach. As in [34] the problem is supposed to have a unique solution, but the associated bilinear form is not supposed to be coercive. Contrary to what was previously done for Virtual Element Methods (as for instance in [5]), we use here, in a systematic way, the $L^2$-projection operators as designed in [1]. In particular, the present method *does not* reduce to the original Virtual Element Method of [5] for simpler problems as the classical Laplace operator (apart from the lowest order cases). Numerical experiments show the accuracy and the robustness of the method, and they show as well that a simple-minded extension of the method in [5] to the case of variable coefficients produces, in general, sub-optimal results.

AMS Subject Classification: 65N30

## 1. Introduction

The aim of this paper is to design and analyze the use of Virtual Element Methods (in short, VEM) for the approximate solution of general linear second order elliptic problems in two dimensions. In particular we shall deal with diffusion-convection-reaction problems with variable coefficients.

For the simpler case of Laplace operator in two dimensions the Virtual Element Method in the primal form (see [5]) could be seen essentially as a re-formulation (in a simpler, more elegant and easier to analyze manner) of the Mimetic Finite Difference method as presented in [14] for the lowest order case, and extended to arbitrary order in [9].

Actually, in more recent times both Mimetic Finite Differences and Virtual Element Methods have been growing very fast, allowing a much wider type of discretizations (arbitrary degree, arbitrary continuity, nonconforming or discontinuous variants) as well as different types of applications. See in particular, for Mimetic Finite Differences [15], and mostly [10,27] (and the references therein), and [1,2,11,6,7,16,17,26,28,30] for Virtual Elements.

We point out, on the other hand, that the use of polygonal and polyhedral meshes for the approximate solution of Partial Differential Equations, but also for several other branches of Scientific Computing, is surely not reduced to Mimetic Finite Differences or Virtual Element Methods. Indeed, polygonal (and then polyhedral) decompositions have already a long story, and often are based on approaches that are substantially different from MFDs or VEMs. We recall for instance [3,4,12,13,19,21,24,25,29,31,32,33,35].

Most of these methods use trial and test functions of a rather complicate nature, that often could be computed (and integrated) only in some approximate way. The same is (even more) true for Virtual Element Methods where trial and test functions are solutions of PDE problems inside each element. However, these local problems are not solved, not even in an approximate way, and the general idea is (roughly speaking) to try to compute *exactly* the values of the local (stiffness) bilinear form when one of the two entries is of polynomial type, and then stabilize the rest, in a rather brutal way. Keeping this in mind, it is clear that for Virtual Element Methods the extension from the constant coefficients to the variable ones is less trivial than for other methods, and in particular, simple minded approaches to variable coefficients can lead to a loss of optimality, especially for higher order methods, as we show with numerical evidence at the end of this paper.

In more recent times several other methods for polygonal decompositions have been introduced in which the trial and test functions are *pairs of polynomial* (instead of a single non-polynomial function). See [18,19,20,22,36]. Though different, these methods surely have many points in common with each other, and with Virtual Element Methods. The main difference is that in the Virtual Element Methods we have indeed, on each element, both boundary and internal degrees of freedom, but they refer to *the same function* (as it is normal for traditional Finite Element Meth-

ods), that however is not a polynomial, while in these other methods we have *two different functions* that are both polynomials.

However we could consider that the internal degrees of freedom refer to a different (polynomial) function, that has the same moments as the VEM one (as it is done for instance in Mimetic Finite Differences, where the degrees of freedom are treated more as *co-chains* rather than values attached to a specific function). In this respect, the relationships among all these methods definitely deserves a deeper analysis.

The most recent Virtual Element approach (already hinted in [1] for dealing with Laplace operator in three dimensions and later extended to mixed formulations in [7]) consists in a tricky way to make the $L^2$-projection operator computable in an exact way starting from the degrees of freedom, with the idea to use, as often as possible, the $L^2$-projection of test and trial functions in place of the functions themselves.

A question that often arises when presenting Virtual Element approximations is: "Since the approximate solution is not explicitly known inside the elements, how can it be represented? And/or how can we compute its value at points of interest that are internal to elements?" What we suggest is simply to use the $L^2-$projection of the VEM-solution onto piecewise polynomials of degree $k$. In Section 6 we provide numerical results showing the general behavior of the error, and also the error in some internal point following this path.

An outline of the paper is as follows. After stating the problem and its formal adjoint in Section 2, we recall in Section 3 the variational formulation. Then, in Section 4 we introduce the Virtual Element approximation. Section 5 is devoted to prove optimal error estimates in $H^1$ and in $L^2$, given in Theorem 5.1 and Theorem 5.2, respectively. Finally, numerical results are presented in Section 6.

Throughout the paper we will use the standard notation $(\cdot, \cdot)$ or $(\cdot, \cdot)_0$ to indicate the $L^2$ scalar product. Whenever confusion may arise, we will underline the domain explicitly; for instance $(\cdot, \cdot)_{0,E}$ will denote the $L^2(E)$ scalar product on a generic polygon $E$. For every geometrical object $\mathcal{O}$ and for every integer $k \geq -1$ we denote by $\mathbb{P}_k(\mathcal{O})$ the set of polynomials of degree $\leq k$ on $\mathcal{O}$, with $\mathbb{P}_{-1}(\mathcal{O}) \equiv \{0\}$, as usual. Whenever no confusion may arise, we will simply use $\mathbb{P}_k$, without declaring explicitly the domain.

## 2. The problem and the adjoint problem

Let $\Omega \subset \mathbb{R}^2$ be a bounded convex polygonal domain with boundary $\Gamma$, let $\kappa$ and $\gamma$ be smooth functions $\Omega \to \mathbb{R}$ with $\kappa(\mathbf{x}) \geq \kappa_0 > 0$ for all $\mathbf{x} \in \Omega$, and let $\mathbf{b}$ be a smooth vector valued function $\Omega \to \mathbb{R}^2$. In the sequel $\kappa_{\max}, \gamma_{\max}$ and $b_{\max}$ will denote the ($L^\infty-$like) norm of the coefficients $\kappa, \gamma, \mathbf{b}$, respectively.

Assume that the problem

$$\begin{cases} \mathfrak{L} p := \operatorname{div}(-\kappa(\mathbf{x})\nabla p + \mathbf{b}(\mathbf{x})p) + \gamma(\mathbf{x})\, p = f(\mathbf{x}) & \text{in } \Omega \\ \qquad\qquad\qquad\qquad\qquad\qquad\quad p = 0 & \text{on } \Gamma \end{cases} \tag{2.1}$$

4   *Beirão da Veiga, Brezzi, Marini, Russo*

is solvable for any $f \in H^{-1}(\Omega)$, and that the estimates

$$\|p\|_{1,\Omega} \le C\|f\|_{-1,\Omega} \tag{2.2}$$

and

$$\|p\|_{2,\Omega} \le C\|f\|_{0,\Omega} \tag{2.3}$$

hold with a constant $C$ independent of $f$. We point out that these assumptions imply, among other things, that existence and uniqueness hold, as well, for the (formal) adjoint operator $\mathfrak{L}^*$ given by

$$\mathfrak{L}^*p := \operatorname{div}(-\kappa(\mathbf{x})\nabla p) - \mathbf{b}(\mathbf{x}) \cdot \nabla p + \gamma(\mathbf{x})\,p. \tag{2.4}$$

Moreover, for every $g \in L^2(\Omega)$ there exists a unique $\varphi \in H^2(\Omega) \cap H_0^1(\Omega)$ such that $\mathfrak{L}^*\varphi = g$, and

$$\|\varphi\|_{2,\Omega} \le C^*\|g\|_{0,\Omega} \tag{2.5}$$

for a constant $C^*$ independent of $g$. As we shall see, the 2-regularity (2.3) and (2.5) is not strictly necessary in order to get the results of the present work, and an $s$-regularity with $s > 1$ would be sufficient. Here however we are not interested in minimizing the regularity assumptions.

We also point out that the choice of having a scalar diffusion coefficient was done just for simplicity. Having a full diffusion tensor would not change the analysis in a substantial way. Actually, in the numerical results presented in Section 6 a full tensor is used.

## 3.   Variational formulation

Set:

$$a(p,q) := \int_\Omega \kappa\nabla p \cdot \nabla q \,\mathrm{d}x, \quad b(p,q) := -\int_\Omega p(\mathbf{b}\cdot\nabla q)\,\mathrm{d}x, \quad c(p,q) := \int_\Omega \gamma p\,q\,\mathrm{d}x \tag{3.1}$$

and define

$$B(p,q) := a(p,q) + b(p,q) + c(p,q). \tag{3.2}$$

The variational formulation of problem (2.1) is

$$\begin{cases} \text{Find } p \in H_0^1(\Omega) \text{ such that} \\ B(p,q) = (f,q) \quad \forall q \in H_0^1(\Omega). \end{cases} \tag{3.3}$$

**Remark 3.1.** It is immediate to check that our assumptions on the coefficients imply that the bilinear form $B(\cdot,\cdot)$ verifies

$$B(p,q) \le M\|p\|_1\|q\|_1, \quad p,q \in H^1(\Omega) \tag{3.4}$$

and hence

$$\|\mathfrak{L}p\|_{-1} = \sup_{q\in H_0^1} \frac{<\mathfrak{L}p,q>}{\|q\|_1} = \sup_{q\in H_0^1} \frac{B(p,q)}{\|q\|_1} \le M\|p\|_1.$$

It is also easy to check that this, together with (2.2), implies that

$$\sup_{q \in H_0^1} \frac{B(p,q)}{\|q\|_1} \geq C_B \|p\|_1 \quad \forall p \in H_0^1(\Omega), \tag{3.5}$$

for some constant $C_B > 0$ independent of $p$. On the other hand it is also well known that (3.4) and (3.5) imply existence and uniqueness of the solution of problem (3.3).

## 4. VEM approximation

In the present section we introduce the virtual element discretization of (3.3).

### 4.1. *The Virtual Element space*

Let $\mathcal{T}_h$ be a decomposition of $\Omega$ into star-shaped polygons $E$, and let $\mathcal{E}_h$ be the set of edges $e$ of $\mathcal{T}_h$.

**Remark 4.1.** To be precise, we assume that (i) every element $E$ is star-shaped with respect to every point of a disk $D_\rho$ of radius $\rho^E h_E$ (where $h_E$ is the diameter of $E$), and (ii) that every edge $e$ of $E$ has lenght $|e| \geq \rho^E h_E$. The first assumption could be relaxed in order to allow unions of star-shaped elements and the second one could be essentially avoided; since such technical generalizations are beyond the scope of the present work, we prefer to keep the simpler conditions stated above. When considering a *sequence* of decompositions $\{\mathcal{T}_h\}_h$ we will obviously assume $\rho^E \geq \rho_0 > 0$ for some $\rho_0$ independent of $E$ and of the decomposition. As usual, $h$ will denote the maximum diameter of the elements of $\mathcal{T}_h$.

Following [5,1], for every integer $k \geq 1$ and for every element $E$ we start by defining a *preliminary* local space:

$$\widetilde{\mathcal{Q}}_h^k(E) := \{q \in H^1(E): \ q_{|e} \in \mathbb{P}_k(e) \ \forall e \in \partial E, \ \Delta q \in \mathbb{P}_k(E)\}. \tag{4.1}$$

On $\widetilde{\mathcal{Q}}_h^k(E)$ the following set of linear operators are well defined. For all $q \in \widetilde{\mathcal{Q}}_h^k(E)$:

($D_1$)  the values $q(V_i)$ at the vertices $V_i$ of $E$,

and for $k \geq 2$

($D_2$)  the edge moments $\int_e q\, p_{k-2}\, \mathrm{d}s$, $p_{k-2} \in \mathbb{P}_{k-2}(e)$, on each edge $e$ of $E$,
($D_3$)  the internal moments $\int_E q\, p_{k-2}\, \mathrm{d}x$, $p_{k-2} \in \mathbb{P}_{k-2}(E)$.

We point out that for each element $E$ and for all $k$ the operators $D_1$–$D_3$ satisfy the following property:

$$\{q \in \mathbb{P}_k(E)\} \text{ and } \{D_i(q) = 0, i = 1, 2, 3\} \text{ imply } \{q = 0\}. \tag{4.2}$$

Property (4.2) implies that on each element $E$ we can easily construct a projection operator from $\widetilde{\mathcal{Q}}_h^k$ to $\mathbb{P}_k$ that depends only on $D_1$–$D_3$ and is explicitly computable starting from them. Let us see how. Let $n_V$ be the number of vertices of $E$, and let

$$n_D := n_V k + k(k-1)/2 \tag{4.3}$$

be the "cardinality" of $D_1$–$D_3$ (with obvious meaning). Consider the mapping from $\widetilde{\mathcal{Q}}_h^k(E)$ to $\mathbb{R}^{n_D}$ defined by $Dq := (D_1$–$D_3)(q)$, and choose a bilinear symmetric positive form $\mathcal{G}$ on $\mathbb{R}^{n_D} \times \mathbb{R}^{n_D}$ (for instance, the Euclidean scalar product on $\mathbb{R}^{n_D}$). For every $q \in \widetilde{\mathcal{Q}}_h^k(E)$ we define $\Pi_k^{\mathcal{G}} q \in \mathbb{P}_k$ as the unique solution of

$$\mathcal{G}(Dq - D\Pi_k^{\mathcal{G}}q, Dz) = 0 \qquad \forall\, z \in \mathbb{P}_k. \tag{4.4}$$

It is obvious that $\Pi_k^{\mathcal{G}} q_k \equiv q_k$ for every $q_k \in \mathbb{P}_k$, and also that $\Pi_k^{\mathcal{G}} q$ depends only on the values of $D_1 q$, $D_2 q$, and $D_3 q$. It can be rather easily proved that *every projection operator $\widetilde{\mathcal{Q}}_h^k(E) \to \mathbb{R}^{n_D}$ depending only on the values of $D_1$–$D_3$ can be obtained by* (4.4) *for a suitable choice of the bilinear form $\mathcal{G}$.* It is also obvious that collecting all the local projection operators we can construct every global projection operator from $\widetilde{\mathcal{Q}}_h^k$ to the space of piecewise $\mathbb{P}_k$ functions.

Here however (both for *historical* reasons and for convenience of computation) we will focus our attention on a particular choice of projection operator. For this we recall from [5,1] the definition of the operator $\Pi_k^\nabla$: for any $q \in H_0^1(\Omega)$, the function $\Pi_k^\nabla q$ on each element $E$ is a polynomial in $\mathbb{P}_k(E)$, defined by

$$(\nabla(\Pi_k^\nabla q - q), \nabla p_k)_{0,E} = 0 \quad \text{and} \quad \int_{\partial E}(\Pi_k^\nabla q - q)\mathrm{d}s = 0 \qquad \forall p_k \in \mathbb{P}_k. \tag{4.5}$$

This operator is well defined on $\widetilde{\mathcal{Q}}_h^k(E)$ and, most important, for all $q \in \widetilde{\mathcal{Q}}_h^k(E)$ the polynomial $\Pi_k^\nabla q$ can be computed using only the values of the operators (D) calculated on $q$. This follows easily with an integration by parts, see for instance [5].

We are now ready to introduce our local Virtual space

$$\mathcal{Q}_h^k(E) := \{q \in \widetilde{\mathcal{Q}}_h^k(E) \,:\, \int_E q\, p_k \,\mathrm{d}x = \int_E (\Pi_k^\nabla q) p_k \,\mathrm{d}x \ \ \forall p_k \in (\mathbb{P}_k/\mathbb{P}_{k-2}(E))\}, \tag{4.6}$$

where the space $\big(\mathbb{P}_k/\mathbb{P}_{k-2}(E)\big)$ denotes the polynomials in $\mathbb{P}_k(E)$ that are $L^2(E)$ orthogonal to $\mathbb{P}_{k-2}(E)$. The corresponding global space is:

$$\mathcal{Q}_h^k := \{q \in H_0^1(\Omega) :\ q_{|E} \in \mathcal{Q}_h^k(E) \ \forall E \in \mathcal{T}_h\}. \tag{4.7}$$

Let now $\Pi_k^0$ denote the $L^2-$ projection onto $\mathbb{P}_k$, defined locally, as usual, by

$$(q - \Pi_k^0 q, p_k)_{0,E} = 0 \qquad \forall p_k \in \mathbb{P}_k. \tag{4.8}$$

For simplicity of notation, in the following we will denote by the same symbol also the $L^2-$ projection of vector valued functions onto the polynomial space $[\mathbb{P}_k]^2$.

We note that it can be proved, see again [5,1] that the set of linear operators (D) are a set of degrees of freedom for the virtual space $\mathcal{Q}_h^k(E)$.

Clearly the degrees of freedom (D) define an *interpolation operator* that associates to each smooth enough function $\varphi$ its *interpolant* $\varphi_I \in \mathcal{Q}_h^k(E)$ that shares with $\varphi$ the values of the degrees of freedom. Moreover, the virtual space $\mathcal{Q}_h^k(E)$ satisfies the following four properties:

- $\mathbb{P}_k(E) \subseteq \mathcal{Q}_h^k(E)$ (trivial to check);

- for all $q \in \mathcal{Q}_h^k(E)$, the function $\Pi_k^\nabla q$ can be explicitly computed from the degrees of freedom $(D)$ of $q$ (see [5,1]);
- for all $q \in \mathcal{Q}_h^k(E)$, the function $\Pi_k^0 q$ can be explicitly computed from the degrees of freedom $(D)$ of $q$ (see [1]);
- for all $q \in \mathcal{Q}_h^k(E)$, the vector function $\Pi_{k-1}^0 \nabla q$ can be explicitly computed from the degrees of freedom $(D)$ of $q$ (see below).

While the second and third properties above can be found in the literature, and thus are not detailed here, we need to spend some words on the last one. In order to compute $\Pi_{k-1}^0 \nabla q$, for all $E \in \mathcal{T}_h$ we must be able to calculate

$$\int_E \nabla q \cdot \mathbf{p}_{k-1}\, \mathrm{d}x \quad \forall \mathbf{p}_{k-1} \in [\mathbb{P}_{k-1}(E)]^2.$$

An integration by parts, denoting by $\mathbf{n}$ the outward unit normal to the element boundary $\partial E$, gives

$$\int_E \nabla q \cdot \mathbf{p}_{k-1}\, \mathrm{d}x = -\int_E q\, \mathrm{div}(\mathbf{p}_{k-1})\, \mathrm{d}x + \int_{\partial E} q\, (\mathbf{p}_{k-1} \cdot \mathbf{n})\, \mathrm{d}s.$$

The first term in the right hand side above clearly depends only on the moments of $q$ appearing in $(D_3)$. The second term can also be computed since $q$ is a polynomial of degree $k$ on each edge and therefore $q_{|\partial E}$ is uniquely determined by the values of $(D_1)$ and $(D_2)$. Needless to say, all the above properties extend in an obvious way to the global space (4.7). In particular, we point out that, for a smooth function $\varphi \in H_0^1(\Omega)$, its global interpolant $\varphi_I$ is in $Q_h^k$.

We end this section by showing some simple bounds on the operator $\Pi_k^\nabla$. Applying (4.5) for $p_k = \Pi_k^\nabla q$ we have

$$\|\nabla \Pi_k^\nabla q\|_{0,E}^2 = (\nabla q, \nabla \Pi_k^\nabla q)_{0,E} \leq \|\nabla q\|_{0,E} \|\nabla \Pi_k^\nabla q\|_{0,E}$$

giving immediately

$$\|\nabla \Pi_k^\nabla q\|_{0,E} \leq \|\nabla q\|_{0,E}. \tag{4.9}$$

Moreover, always from the definition (4.5),

$$(\nabla(q - \Pi_k^\nabla q), \nabla(q - \Pi_k^\nabla q))_{0,E} = (\nabla(q - \Pi_k^\nabla q), \nabla q)_{0,E} \leq |q - \Pi_k^\nabla q|_{1,E} |\nabla q|_{0,E}$$

that immediately gives

$$|q - \Pi_k^\nabla q|_{1,E} \leq |q|_{1,E}. \tag{4.10}$$

Finally, using again the definition (4.5) we have

$$\|\nabla(q - \Pi_k^\nabla q)\|_{0,E}^2 = (\nabla(q - \Pi_k^\nabla q), \nabla(q - \Pi_k^0 q))_{0,E} \leq \|\nabla(q - \Pi_k^\nabla q)\|_{0,E} \|\nabla(q - \Pi_k^0 q)\|_{0,E},$$

giving

$$\|\nabla(q - \Pi_k^\nabla q)\|_{0,E} \leq \|\nabla(q - \Pi_k^0 q)\|_{0,E}. \tag{4.11}$$

8   *Beirão da Veiga, Brezzi, Marini, Russo*

### 4.2.  *The discrete problem*

We now introduce the discrete bilinear forms that will be used in the method. Since we will mostly work on a generic element $E$, we will denote by $a^E(\cdot,\cdot), b^E(\cdot,\cdot), c^E(\cdot,\cdot)$, and $B^E(\cdot,\cdot)$ the restriction to $E$ of the corresponding bilinear forms defined in (3.1)-(3.2). Let $S^E(p,q)$ be a symmetric bilinear form on $\mathcal{Q}_h^k(E) \times \mathcal{Q}_h^k(E)$ that scales like $a^E(\cdot,\cdot)$ on the kernel of $\Pi_k^\nabla$. More precisely, we assume that $\exists\, \alpha_*, \alpha^*$ independent of $h$ with $0 < \alpha_* \le \alpha^*$ such that

$$\alpha_* a^E(q_h, q_h) \le S^E(q_h, q_h) \le \alpha^* a^E(q_h, q_h) \quad \forall q_h \in \mathcal{Q}_h^k(E) \text{ with } \Pi_k^\nabla q_h = 0. \quad (4.12)$$

Examples on how to build the bilinear form above can be found in [5,8]. Note that, due to the symmetry of $S^E$, this implies, for all $p_h, q_h \in \mathcal{Q}_h^k(E)$ with $\Pi_k^\nabla p_h = \Pi_k^\nabla q_h = 0$,

$$S^E(p_h, q_h) \le (S^E(p_h, p_h))^{1/2}(S^E(q_h, q_h))^{1/2} \le \alpha^*(a^E(p_h, p_h))^{1/2}(a^E(q_h, q_h))^{1/2}. \quad (4.13)$$

We can now define, on each element $E \in \mathcal{T}_h$ and for every $p$, $q$ in $\mathcal{Q}_h^k(E)$, the local forms and loading term:

$$a_h^E(p,q) := \int_E \kappa [\Pi_{k-1}^0 \nabla p] \cdot [\Pi_{k-1}^0 \nabla q]\, \mathrm{d}x + S^E((I - \Pi_k^\nabla)p, (I - \Pi_k^\nabla)q)$$

$$b_h^E(p,q) := -\int_E [\Pi_{k-1}^0 p]\, [\mathbf{b} \cdot \Pi_{k-1}^0 \nabla q]\, \mathrm{d}x, \qquad\qquad (4.14)$$

$$c_h^E(p,q) := \int_E \gamma [\Pi_{k-1}^0 p]\, [\Pi_{k-1}^0 q]\, \mathrm{d}x, \quad (f_h, q)_E := \int_E f\, \Pi_{k-1}^0 q\, \mathrm{d}x,$$

$$B_h^E(p,q) := a_h^E(p,q) + b_h^E(p,q) + c_h^E(p,q).$$

We just recall that, since $\Pi_k^\nabla$ is a projection, then

$$S^E((I - \Pi_k^\nabla)p_k, (I - \Pi_k^\nabla)q) = 0 \quad \forall p_k \in \mathbb{P}_k,\ \forall q \in \mathcal{Q}_h^k(E), \qquad (4.15)$$

and thus, since $S^E$ is symmetric, the $S^E$ term will vanish whenever one of the two entries of $a_h^E(\cdot,\cdot)$ is a polynomial in $\mathbb{P}_k$.
Then we set for all $p, q \in \mathcal{Q}_h^k$

$$a_h(p,q) := \sum_E a_h^E(p,q), \quad b_h(p,q) := \sum_E b_h^E(p,q),$$

$$c_h(p,q) := \sum_E c_h^E(p,q), \quad (f_h, q) := \sum_E (f_h, q)_E,$$

and

$$B_h(p,q) := a_h(p,q) + b_h(p,q) + c_h(p,q) = \sum_E B_h^E(p,q). \qquad (4.16)$$

The approximate problem is:

$$\begin{cases} \text{Find } p_h \in \mathcal{Q}_h^k \text{ such that} \\ B_h(p_h, q) = (f_h, q) \quad \forall q \in \mathcal{Q}_h^k. \end{cases} \qquad (4.17)$$

**Remark 4.2.** The bilinear forms $b_h^E$ and $c_h^E$ in (4.14) are well defined for all $p, q \in H^1(E)$, as well as the global forms $b_h$ and $c_h$, which are well defined on the whole $H_0^1(\Omega)$. This does not hold for $a_h^E$, due to the presence of the stabilizing term $S^E$ that is defined only on $\mathcal{Q}_h^k(E)$.

**Remark 4.3.** We recall that the choice indicated in [5] would have suggested to define

$$a_h^E(p, q) := \int_E \kappa [\nabla \Pi_k^\nabla p] \cdot [\nabla \Pi_k^\nabla q] \, \mathrm{d}x + S^E((I - \Pi_k^\nabla)p, (I - \Pi_k^\nabla)q). \qquad (4.18)$$

Actually, it can be easily seen that for $k = 1$ this coincides with our choice (4.14). This is not the case for $k \geq 2$. In particular, a deeper analysis shows heavy losses in the order of convergence for $k \geq 3$. In Section 6 we provide an example for $k = 4$. On the other hand, it can be shown that if $\kappa \nabla p$ happens to be a gradient the choice (4.18) does work.

## 5. Error estimates

In the present section we derive error estimates for the proposed method.

### 5.1. *Preliminary results*

We now present some preliminary results useful in the sequel. We start by the following approximation lemma, that mainly comes from the mesh regularity assumptions in Remark 4.1 and standard approximation results on polygonal domains (see for instance [5,30]).

Here and in the sequel $C$ will denote a generic positive constant independent of $h$, with different meaning in different occurrencies, and generally depending on the coefficients of the operator $\mathfrak{L}$. Whenever needed to better follow the steps of the proofs, for a smooth scalar or vector-valued function $\aleph$, we shall use $C_\aleph$ to denote a constant depending on $\aleph$ and possibly on its derivatives up to the needed order.

**Lemma 5.1.** *There exists a positive constant* $C = C(\rho_0, k)$ *such that, for all $E$ in* $\mathcal{T}_h$ *and all smooth enough functions* $\varphi$ *defined on $E$, it holds*

$$\|\varphi - \Pi_k^0 \varphi\|_{m,E} \leq C h_E^{s-m} |\varphi|_{s,E} \qquad m, s \in \mathbb{N}, \ m \leq s \leq k+1,$$

$$\|\varphi - \Pi_k^\nabla \varphi\|_{m,E} \leq C h_E^{s-m} |\varphi|_{s,E}, \qquad m, s \in \mathbb{N}, \ m \leq s \leq k+1, \ s \geq 1,$$

$$\|\varphi - \varphi_I\|_{m,E} \leq C h_E^{s-m} |\varphi|_{s,E}, \qquad m, s \in \mathbb{N}, \ m \leq s \leq k+1, \ s \geq 2.$$

We also have the following continuity lemma.

**Lemma 5.2.** *The bilinear form* $B_h(\cdot, \cdot)$ *is continuous in* $\mathcal{Q}_h^k \times \mathcal{Q}_h^k$, *that is,*

$$B_h(p, q) \leq C_{\kappa, \mathbf{b}, \gamma} \|p\|_1 \|q\|_1 \quad p, q \in \mathcal{Q}_h^k, \qquad (5.1)$$

*with* $C_{\kappa, \mathbf{b}, \gamma}$ *a positive constant depending on* $\kappa, \mathbf{b}, \gamma$ *but independent of $h$.*

**Proof.** The continuity of $b_h$ and $c_h$ is obvious, and actually holds on the whole $H_0^1(\Omega)$ space. We have

$$b_h(p,q) \leq b_{\max}\|p\|_0 |q|_1, \quad c_h(p,q) \leq \gamma_{\max}\|p\|_0 \|q\|_0, \quad p,q \in H_0^1(\Omega). \tag{5.2}$$

The continuity of $a_h$ is proved upon observing that, thanks to (4.13) and (4.10),

$$S^E((I - \Pi_k^\nabla)p, (I - \Pi_k^\nabla)q)) \leq \alpha^* \kappa_{\max} |p - \Pi_k^\nabla p|_{1,E} |q - \Pi_k^\nabla q|_{1,E}$$
$$\leq \alpha^* \kappa_{\max} |p|_{1,E} |q|_{1,E}. \tag{5.3}$$

Thus:

$$a_h(p,q) \leq (1 + \alpha^*)\kappa_{\max} |p|_1 |q|_1 \quad p,q \in \mathcal{Q}_h^k, \tag{5.4}$$

and the result follows. $\qquad\square$

In many occasions we will need to estimate the difference between continuous and discrete bilinear forms. This is done once and for all in the following preliminary Lemma.

**Lemma 5.3.** *Let $E \in \mathcal{T}_h$, let $\mu$ be a smooth function on $E$, and let $p, q$ denote smooth scalar or vector-valued functions on $E$. For a generic $\varphi \in L^2(E)$ (or in $(L^2(E))^2$) we define*

$$\mathcal{E}_E^k(\varphi) := \|\varphi - \Pi_k^0 \varphi\|_{0,E}. \tag{5.5}$$

*Then we have the estimate:*

$$(\mu p, q)_{0,E} - (\mu \Pi_k^0 p, \Pi_k^0 q)_{0,E} \leq \mathcal{E}_E^k(\mu p)\mathcal{E}_E^k(q) + \mathcal{E}_E^k(\mu q)\mathcal{E}_E^k(p) + C_\mu \mathcal{E}_E^k(p)\mathcal{E}_E^k(q), \tag{5.6}$$

*where $C_\mu$ is a constant depending on $\mu$.*

**Proof.** For simplifying the notation we will set $\overline{p} := \Pi_k^0 p$, $\overline{q} := \Pi_k^0 q$. By adding and subtracting terms, and by the definition of projection we have

$$(\mu p, q)_{0,E} - (\mu \overline{p}, \overline{q})_{0,E} = (\mu p, q - \overline{q})_{0,E} + (p - \overline{p}, \mu \overline{q})_{0,E}$$
$$= (\mu p - \overline{\mu p}, q - \overline{q})_{0,E} + (p - \overline{p}, \mu \overline{q} - \overline{\mu q})_{0,E}$$
$$= (\mu p - \overline{\mu p}, q - \overline{q})_{0,E} + (p - \overline{p}, \mu \overline{q} - \overline{\mu q} + \mu q - \mu q)_{0,E}$$
$$= (\mu p - \overline{\mu p}, q - \overline{q})_{0,E} + (p - \overline{p}, \mu q - \overline{\mu q})_{0,E} - (p - \overline{p}, \mu(q - \overline{q}))_{0,E}, \tag{5.7}$$

and the result follows by Cauchy-Schwarz inequality with $C_\mu = \|\mu\|_\infty$. $\qquad\square$

The following result follows immediately by a direct application of Lemma 5.3.

**Lemma 5.4.** *For all $E \in \mathcal{T}_h$ it holds*

$$a_h^E(p,q) - a^E(p,q) \leq \mathcal{E}_E^{k-1}(\kappa \nabla p)\mathcal{E}_E^{k-1}(\nabla q) + \mathcal{E}_E^{k-1}(\kappa \nabla q)\mathcal{E}_E^{k-1}(\nabla p)$$
$$+ C_\kappa \mathcal{E}_E^{k-1}(\nabla p)\mathcal{E}_E^{k-1}(\nabla q)$$
$$+ S^E((I - \Pi_k^\nabla)p, (I - \Pi_k^\nabla)q)) \quad \forall p, q \in \mathcal{Q}_h^k(E), \tag{5.8}$$

$$b_h^E(p,q) - b^E(p,q) \leq \mathcal{E}_E^{k-1}(\mathbf{b} \cdot \nabla q)\mathcal{E}_E^{k-1}(p) + \mathcal{E}_E^{k-1}(\nabla q)\mathcal{E}_E^{k-1}(\mathbf{b}p)$$
$$+ C_{\mathbf{b}}\mathcal{E}_E^{k-1}(\nabla q)\mathcal{E}_E^{k-1}(p) \quad \forall p,q \in H^1(E), \tag{5.9}$$

$$c_h^E(p,q) - c^E(p,q) \leq \mathcal{E}_E^{k-1}(\gamma p)\mathcal{E}_E^{k-1}(q) + \mathcal{E}_E^{k-1}(\gamma q)\mathcal{E}_E^{k-1}(p)$$
$$+ C_{\gamma}\mathcal{E}_E^{k-1}(p)\mathcal{E}_E^{k-1}(q) \quad \forall p,q \in H^1(E). \tag{5.10}$$

In the next Lemma we evaluate the consistency error.

**Lemma 5.5 (consistency).** *For all $p$ sufficiently regular and for all $q_h \in \mathcal{Q}_h^k$ it holds*

$$B^E(\Pi_k^0 p, q_h) - B_h^E(\Pi_k^0 p, q_h) \leq C_{\kappa,\mathbf{b},\gamma} h_E^k \|p\|_{k+1,E} \|q_h\|_{1,E} \quad \forall E \in \mathcal{T}_h. \tag{5.11}$$

**Proof.** From the definition of $B^E$ and $B_h^E$ we have

$$B^E(\Pi_k^0 p, q_h) - B_h^E(\Pi_k^0 p, q_h) = a^E(\Pi_k^0 p, q_h) - a_h^E(\Pi_k^0 p, q_h)$$
$$+ b^E(\Pi_k^0 p, q_h) - b_h^E(\Pi_k^0 p, q_h) + c^E(\Pi_k^0 p, q_h) - c_h^E(\Pi_k^0 p, q_h). \tag{5.12}$$

We first observe that when $p \in \mathbb{P}_k(E)$, then obviously we have $\Pi_k^0 p \equiv p$, $\Pi_{k-1}^0 \nabla p \equiv \nabla p$, and then by (4.15) the term containing $S^E$ vanishes. Therefore, a direct application of (5.8) implies

$$a_h^E(\Pi_k^0 p, q_h) - a^E(\Pi_k^0 p, q_h) \leq \mathcal{E}_E^{k-1}(\kappa \nabla \Pi_k^0 p)\,\mathcal{E}_E^{k-1}(\nabla q_h), \tag{5.13}$$

for all $q_h \in \mathcal{Q}_h^k(E)$. The first factor in the right-hand side of (5.13) can be easily bounded by

$$\mathcal{E}_E^{k-1}(\kappa \nabla \Pi_k^0 p) = \|\kappa \nabla \Pi_k^0 p - \Pi_{k-1}^0(\kappa \nabla \Pi_k^0 p)\|_{0,E} \leq \|\kappa \nabla \Pi_k^0 p - \Pi_{k-1}^0(\kappa \nabla p)\|_{0,E}$$
$$\leq \|\kappa \nabla \Pi_k^0 p - \kappa \nabla p\|_{0,E} + \|\kappa \nabla p - \Pi_{k-1}^0(\kappa \nabla p)\|_{0,E}$$
$$\leq C\,h_E^k(\kappa_{\max}|p|_{k+1,E} + |\kappa \nabla p|_{k,E}) \leq C_{\kappa} h_E^k \|p\|_{k+1,E}, \tag{5.14}$$

and the second factor can by simply bounded by $\|q_h\|_{1,E}$. Thus,

$$a_h^E(\Pi_k^0 p, q_h) - a^E(\Pi_k^0 p, q_h) \leq C_{\kappa} h_E^k \|p\|_{k+1,E} \|q_h\|_{1,E}. \tag{5.15}$$

With similar arguments we have, for instance,

$$\mathcal{E}_E^{k-1}(\mathbf{b}\Pi_k^0 p) \leq C(h_E^{k+1}\mathbf{b}_{\max}|p|_{k+1,E} + h_E^k|\mathbf{b}p|_{k,E}) \leq C_{\mathbf{b}} h_E^k \|p\|_{k+1,E},$$
$$\mathcal{E}_E^{k-1}(\gamma\Pi_k^0 p) \leq C(h_E^{k+1}\gamma_{\max}|p|_{k+1,E} + h_E^k|\gamma p|_{k,E}) \leq C_{\gamma} h_E^k \|p\|_{k+1,E}. \tag{5.16}$$

Consequently,

$$b_h^E(\Pi_k^0 p, q_h) - b^E(\Pi_k^0 p, q_h) \leq C_{\mathbf{b}} h_E^k \|p\|_{k+1,E} \|q_h\|_{1,E},$$
$$c_h^E(\Pi_k^0 p, q_h) - c^E(\Pi_k^0 p, q_h) \leq C_{\gamma} h_E^k \|p\|_{k+1,E} \|q_h\|_{1,E}. \tag{5.17}$$

The proof follows by inserting (5.15) and (5.17) in (5.12). $\qquad\square$

**Remark 5.1.** We point out that (5.11) holds for a generic $q_h \in \mathcal{Q}_h^k$, for which only $H^1$ regularity can be used. If for instance $q_h = q_I$, that is, $q_h$ is the interpolate of

a more regular function, (5.11) can be improved. Indeed, looking e.g. at (5.13) we would have

$$
\begin{aligned}
\mathcal{E}_E^{k-1}(\nabla q_I) = \|\nabla q_I - \Pi_{k-1}^0 \nabla q_I\|_{0,E} &\leq \|\nabla q_I - \Pi_{k-1}^0 \nabla q\|_{0,E} \\
&\leq \|\nabla(q_I - q)\|_{0,E} + \|\nabla q - \Pi_{k-1}^0 \nabla q\|_{0,E} \leq C\,h\|q\|_{2,E},
\end{aligned}
\tag{5.18}
$$

and in (5.11) we would gain an extra power of $h$:

$$
B^E(\Pi_k^0 p, q_I) - B_h^E(\Pi_k^0 p, q_I) \leq C_{\kappa,\mathbf{b},\gamma} h_E^{k+1}\|p\|_{k+1,E}\|q\|_{2,E}.
\tag{5.19}
$$

Before going to study the error estimates for our problem, we have to prove a final technical Lemma.

**Lemma 5.6.** *For every $q^* \in H_0^1(\Omega)$ there exists a $q_h^* \in \mathcal{Q}_h^k$ such that*

$$
a_h(q_h^*, q_h) = a(q^*, q_h) \quad \forall\, q_h \in \mathcal{Q}_h^k.
\tag{5.20}
$$

*Moreover, there exists a constant $C$, independent of $h$, such that*

$$
h\|q^* - q_h^*\|_{1,\Omega} + \|q^* - q_h^*\|_{0,\Omega} \leq C\,h\,\|q^*\|_{1,\Omega}.
\tag{5.21}
$$

**Proof.** We first remark that, by definition of projection, we have

$$
\|\nabla q - \Pi_{k-1}^0 \nabla q\|_{0,E} \leq \|\nabla q - \nabla\Pi_k^\nabla q\|_{0,E},
\tag{5.22}
$$

since $\nabla\Pi_k^\nabla q$ is a (vector) polynomial of degree $\leq k-1$. Hence, for $q \in \mathcal{Q}_h^k$ and for every integer $k \geq 1$:

$$
a_h(q,q) \geq C \sum_E \left( \|\Pi_{k-1}^0 \nabla q\|_{0,E}^2 + \|(I - \Pi_{k-1}^0)\nabla q\|_{0,E}^2 \right) \geq C|q|_1^2,
\tag{5.23}
$$

and this immediately implies that (5.20) has a unique solution, and that, using (5.4), we also have $\|q_h^*\|_1 \leq C\,\|q^*\|_1$. In order to show the second part of (5.21) we shall use duality arguments. Let $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ be the solution of

$$
a(q,\psi) = (q^* - q_h^*, q)_{0,\Omega} \quad \forall q \in H_0^1(\Omega),
\tag{5.24}
$$

and let $\psi_I \in \mathcal{Q}_h^k$ be its interpolant, for which it holds

$$
\|\psi - \psi_I\|_1 \leq Ch|\psi|_2 \leq C\,h\,\|q^* - q_h^*\|_0.
\tag{5.25}
$$

We have easily that, for every $k \geq 0$ (and obvious notation for $\mathcal{E}^k$)

$$
\mathcal{E}^k(\nabla\psi_I) \leq \mathcal{E}^0(\nabla\psi_I) \leq \mathcal{E}^0(\nabla(\psi_I - \psi)) + \mathcal{E}^0(\nabla\psi) \leq C\,h\|\psi\|_2 \leq C\,h\|q^* - q_h^*\|_0,
$$

and similarly

$$
\mathcal{E}^k(\kappa\nabla\psi_I) \leq \mathcal{E}^0(\kappa\nabla\psi_I) \leq C_\kappa\,h\|q^* - q_h^*\|_0.
$$

By recalling (4.13) and the definition of the projectors, then using standard approximation estimates, we easily get

$$
\begin{aligned}
S^E((I - \Pi_k^\nabla)q_h^*, (I - \Pi_k^\nabla)\psi_I)) &\leq \alpha^*\kappa_{\max}\|\nabla q_h^* - \nabla\Pi_k^\nabla q_h^*\|_{0,E}\|\nabla\psi_I - \nabla\Pi_k^\nabla\psi_I\|_{0,E} \\
&\leq \alpha^*\kappa_{\max}|q_h^*|_{1,E}\|\nabla\psi_I - \nabla\Pi_k^\nabla\psi\|_{0,E} \\
&\leq \alpha^*\kappa_{\max}|q_h^*|_{1,E}(\|\nabla(\psi_I - \psi)\|_{0,E} + \|\nabla(\psi - \Pi_k^\nabla\psi)\|_{0,E}) \\
&\leq C\,h_E\,|q^*|_{1,E}|\psi|_{2,E}.
\end{aligned}
\tag{5.26}
$$

Summation on the elements and (5.25) give

$$\sum_{E \in \mathcal{T}_h} S^E((I - \Pi_k^\nabla)q_h^*, (I - \Pi_k^\nabla)\psi_I) \leq C\,h\,|q^*|_1 \|q^* - q_h^*\|_0. \qquad (5.27)$$

On the other hand, both $\mathcal{E}^k(\nabla q_h^*)$ and $\mathcal{E}^k(\kappa \nabla q_h^*)$ are just bounded by, say, $C_\kappa \|q^*\|_1$. Then, using (5.24), (5.20), and (5.8) (with (5.27) and (5.18)) we obtain

$$\begin{aligned}
\|q^* - q_h^*\|_0^2 &= a(q^* - q_h^*, \psi) = a(q^* - q_h^*, \psi - \psi_I) + a(q^* - q_h^*, \psi_I) \\
&= a(q^* - q_h^*, \psi - \psi_I) + a_h(q_h^*, \psi_I) - a(q_h^*, \psi_I) \\
&\leq C_\kappa \|q^* - q_h^*\|_1 \|\psi - \psi_I\|_1 + C_\kappa \|q^*\|_1\, h\, \|q^* - q_h^*\|_0,
\end{aligned} \qquad (5.28)$$

and the result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

### 5.2. $H^1$ *Estimate*

We have the following discrete stability lemma.

**Lemma 5.7.** *The bilinear form $B_h(\cdot, \cdot)$ satisfies the following condition (discrete counterpart of (3.5)): there exists an $h_0 > 0$ and a constant $\overline{C}_B$ such that, for all $h < h_0$:*

$$\sup_{q_h \in \mathcal{Q}_h^k} \frac{B_h(p_h, q_h)}{\|q_h\|_1} \geq \overline{C}_B \|p_h\|_1 \quad \forall\, p_h \in \mathcal{Q}_h^k. \qquad (5.29)$$

**Proof.** In order to prove (5.29) we follow Schatz [34]. For $p_h \in \mathcal{Q}_h^k$, from (3.5) we have

$$\exists q^* \in H_0^1(\Omega) \text{ such that } \frac{B(p_h, q^*)}{\|q^*\|_1} \geq C_B \|p_h\|_1. \qquad (5.30)$$

Thanks to Lemma 5.6, the problem

$$\text{Find } q_h^* \in \mathcal{Q}_h^k \text{ such that } a_h(q_h^*, v_h) = a(q^*, v_h) \quad \forall v_h \in \mathcal{Q}_h^k \qquad (5.31)$$

has a unique solution, that satisfies

$$\|q_h^*\|_1 \leq C \|q^*\|_1, \quad \text{and} \quad \|q^* - q_h^*\|_{0,\Omega} \leq C\,h \|q^*\|_1. \qquad (5.32)$$

Then,

$$\begin{aligned}
B_h(p_h, q_h^*) &= a_h(p_h, q_h^*) + b_h(p_h, q_h^*) + c_h(p_h, q_h^*) \\
&= a(p_h, q^*) + b_h(p_h, q_h^*) - b(p_h, q^*) + c_h(p_h, q_h^*) - c(p_h, q^*) \\
&\quad + b(p_h, q^*) + c(p_h, q^*) \\
&= B(p_h, q^*) + b_h(p_h, q_h^*) - b(p_h, q^*) + c_h(p_h, q_h^*) - c(p_h, q^*) \\
&= B(p_h, q^*) + b(p_h, q_h^* - q^*) + b_h(p_h, q_h^*) - b(p_h, q_h^*) \\
&\quad + c_h(p_h, q_h^* - q^*) + c_h(p_h, q^*) - c(p_h, q^*).
\end{aligned} \qquad (5.33)$$

From (5.32) and (5.2) we have

$$c_h(p_h, q_h^* - q^*) \leq \gamma_{\max} h \|p_h\|_0 \|q^*\|_1, \qquad (5.34)$$

14   *Beirão da Veiga, Brezzi, Marini, Russo*

while an integration by parts and again (5.32) yield

$$b(p_h, q_h^* - q^*) = -\int_\Omega p_h \mathbf{b} \cdot \nabla(q_h^* - q^*) \, \mathrm{d}x = \int_\Omega \mathrm{div}(\mathbf{b}p_h) \, (q_h^* - q^*) \, \mathrm{d}x$$
$$\leq \|\mathrm{div}(\mathbf{b}p_h)\|_0 \, h \|q^*\|_1 \leq C_{\mathbf{b}}^\dagger \|p_h\|_1 \, h \|q^*\|_1. \tag{5.35}$$

Moreover from (5.9) with $p = p_h$, $q = q_h^*$

$$b_h(p_h, q_h^*) - b(p_h, q_h^*) \leq \mathcal{E}^{k-1}(\mathbf{b} \cdot \nabla q_h^*) \, \mathcal{E}^{k-1}(p_h) + \mathcal{E}^{k-1}(\nabla q_h^*) \, \mathcal{E}^{k-1}(\mathbf{b}p_h)$$
$$+ C_{\mathbf{b}} \mathcal{E}^{k-1}(\nabla q_h^*) \, \mathcal{E}^{k-1}(p_h)$$
$$= \|\mathbf{b} \cdot \nabla q_h^* - \Pi_{k-1}^0 (\mathbf{b} \cdot \nabla q_h^*)\|_0 \|p_h - \Pi_{k-1}^0 p_h\|_0$$
$$+ \|\nabla q_h^* - \Pi_{k-1}^0 \nabla q_h^*\|_0 \|\mathbf{b}p_h - \Pi_{k-1}^0 (\mathbf{b}p_h)\|_0$$
$$+ C_{\mathbf{b}} \|\nabla q_h^* - \Pi_{k-1}^0 \nabla q_h^*\|_0 \|p_h - \Pi_{k-1}^0 p_h\|_0$$
$$\leq \|\mathbf{b} \cdot \nabla q_h^*\|_0 \, C \, h \, |p_h|_1 + |q_h^*|_1 \, C \, h \, |\mathbf{b}p_h|_1 + C_{\mathbf{b}} |q_h^*|_1 \, C \, h \, |p_h|_1$$
$$\leq C_{\mathbf{b}}^* \, h \|p_h\|_1 \|q^*\|_1. \tag{5.36}$$

Similarly, from (5.10) we deduce

$$c_h(p_h, q^*) - c(p_h, q^*) \leq C_\gamma^* \, h \|p_h\|_0 \|q^*\|_1 \leq C_\gamma^* \, h \|p_h\|_1 \|q^*\|_1. \tag{5.37}$$

Choosing then $h_0 := \frac{C_B}{2(C_{\mathbf{b}}^* + C_\gamma^* + C_{\mathbf{b}}^\dagger + \gamma_{\max})}$ we obviously have for $h \leq h_0$,

$$(C_{\mathbf{b}}^* + C_\gamma^* + C_{\mathbf{b}}^\dagger + \gamma_{\max}) \, h \leq \frac{C_B}{2}. \tag{5.38}$$

Hence, for $h \leq h_0$,

$$B_h(p_h, q_h^*) \geq \frac{C_B}{2} \|p_h\|_1 \|q_h^*\|_1, \tag{5.39}$$

and the proof is concluded.                                                  $\square$

**Remark 5.2.** Clearly, if $\mathbf{b} = 0$, and $\gamma = 0$, (5.38) holds for any $h$ (and, indeed, we are back at the situation of Lemma 5.6).

We are now ready to prove the following Theorem.

**Theorem 5.1.** *For h sufficiently small, problem* (4.17) *has a unique solution* $p_h \in \mathcal{Q}_h^k$, *and the following error estimate holds:*

$$\|p - p_h\|_1 \leq C h^k \, (\|p\|_{k+1} + |f|_k), \tag{5.40}$$

*with C a constant depending on* $\kappa, \boldsymbol{\beta}$, *and* $\gamma$ *but independent of h.*

**Proof.** The existence and uniqueness of the solution of problem (4.17), for $h$ small, is a consequence of Lemma 5.7. To prove the estimate (5.40), using (5.29) we have that for $h \leq h_0$ there exists a $q_h^* \in \mathcal{Q}_h^k$ verifying

$$\frac{B(p_h - p_I, q_h^*)}{\|q_h^*\|_1} \geq \overline{C}_B \|p_h - p_I\|_1. \tag{5.41}$$

Recalling that $B_h(p_h, q_h^*) = (f_h, q_h^*)$, and $B(p, q_h^*) = (f, q_h^*)$, adding and subtracting $\Pi_k^0 p$ some simple algebra yields:

$$
\begin{aligned}
\overline{C}_B \|p_h - p_I\|_1 \|q_h^*\|_1 &\leq B_h(p_h - p_I, q_h^*) = B_h(p_h, q_h^*) - B_h(p_I, q_h^*) \\
&= (f_h, q_h^*) + B_h(\Pi_k^0 p - p_I, q_h^*) - B_h(\Pi_k^0 p, q_h^*) + B(\Pi_k^0 p, q_h^*) \\
&\quad + B(p - \Pi_k^0 p, q_h^*) - B(p, q_h^*) \\
&= (f_h - f, q_h^*) + B_h(\Pi_k^0 p - p_I, q_h^*) + \big(B(\Pi_k^0 p, q_h^*) - B_h(\Pi_k^0 p, q_h^*)\big) \\
&\quad + B(p - \Pi_k^0 p, q_h^*).
\end{aligned}
\tag{5.42}
$$

The first term in the right hand side of (5.42) is bounded by the Cauchy-Schwarz inequality and standard approximation estimates on the load $f$. The second and fourth term are bounded similarly using the continuity of $B_h$ and $B$ combined with approximation estimates for $p$. Finally, the third term is bounded using Lemma 5.5 on each element $E$. We get

$$
\overline{C}_B \|p_h - p_I\|_1 \|q_h^*\|_1 \leq C\, h^k \Big(C_{\kappa, \mathbf{b}, \gamma}\, \|p\|_{k+1} + |f|_k\Big) \|q_h^*\|_1,
$$

and the proof is concluded. $\qquad\square$

**Remark 5.3.** It is immediate to check that, by the same proof, also the following refined result holds:

$$
\|p - p_h\|_1 \leq C \Big( \sum_{E \in \mathcal{T}_h} h_E^{2k} \left(\|p\|_{k+1,E}^2 + |f|_{k,E}^2\right) \Big)^{1/2}.
$$

### 5.3.  $L^2$ *estimate*

We have the following result.

**Theorem 5.2.** *For $h$ sufficiently small, the following error estimate holds:*

$$
\|p - p_h\|_0 \leq Ch^{k+1} \left(\|p\|_{k+1} + |f|_k\right),
\tag{5.43}
$$

*where $C$ is a constant depending on $\kappa, \boldsymbol{\beta}$, and $\gamma$ but independent of $h$.*

**Proof.**  Once more, we shall use duality arguments. Let $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ be the solution of the adjoint problem (see (2.4))

$$
\mathfrak{L}^* \psi = p - p_h,
\tag{5.44}
$$

and let $\psi_I \in \mathcal{Q}_h^k$ be its interpolant, for which it holds

$$
\|\psi - \psi_I\|_1 \leq Ch|\psi|_2 \leq Ch\|p - p_h\|_0.
\tag{5.45}
$$

16   *Beirão da Veiga, Brezzi, Marini, Russo*

Then:

$$
\begin{aligned}
\|p - p_h\|_0^2 &= B(p - p_h, \psi) = B(p, \psi - \psi_I) + B(p, \psi_I) - B(p_h, \psi) \\
&= B(p, \psi - \psi_I) + (f, \psi_I) + B_h(p_h, \psi_I) - (f_h, \psi_I) - B(p_h, \psi) \\
&= B(p - p_h, \psi - \psi_I) + (f - f_h, \psi_I) + B_h(p_h, \psi_I) - B(p_h, \psi_I) \\
&= B(p - p_h, \psi - \psi_I) + (f - f_h, \psi_I - \Pi_{k-1}^0 \psi_I) \\
&\quad + B_h(p_h - \Pi_k^0 p, \psi_I) - B(p_h - \Pi_k^0 p, \psi_I) \\
&\quad + B_h(\Pi_k^0 p, \psi_I) - B(\Pi_k^0 p, \psi_I).
\end{aligned}
\tag{5.46}
$$

Next:

$$
\begin{aligned}
B(p - p_h, \psi - \psi_I) &\leq C h^{k+1} \|p\|_{k+1} \|p - p_h\|_0, \\
(f - f_h, \psi_I - \Pi_{k-1}^0 \psi_I) &\leq C h^{k+1} |f|_k \|p - p_h\|_0.
\end{aligned}
\tag{5.47}
$$

From (5.19) with $q_I = \psi_I$, and (5.45)

$$
B_h(\Pi_k^0 p, \psi_I) - B(\Pi_k^0 p, \psi_I) \leq C_{\kappa, \mathbf{b}, \gamma} h^{k+1} \|p\|_{k+1} \|p - p_h\|_0,
\tag{5.48}
$$

and from (5.8)–(5.10) with $p = p_h - \Pi_k^0 p$, $q = \psi_I$, adding and subtracting $p$,

$$
B_h(p_h - \Pi_k^0 p, \psi_I) - B(p_h - \Pi_k^0 p, \psi_I) \leq C_{\kappa, \mathbf{b}, \gamma} h^{k+1} \|p\|_{k+1} \|p - p_h\|_0.
\tag{5.49}
$$

Hence,

$$
\|p - p_h\|_0 \leq C_{\kappa, \mathbf{b}, \gamma} h^{k+1} (\|p\|_{k+1} + |f|_k).
\tag{5.50}
$$

**Remark 5.4.** As it can be easily seen from our proofs, the extension to the three-dimensional case would not present major difficulties. We chose to skip it here in order to avoid the use of a heavier notation and a certain amount of technicalities.

## 6. Numerical Experiments

We will consider problem (2.1) on the unit square with

$$
\kappa(x, y) = \begin{pmatrix} y^2 + 1 & -xy \\ -xy & x^2 + 1 \end{pmatrix}, \quad \mathbf{b} = (x, y), \quad \gamma = x^2 + y^3,
\tag{6.1}
$$

and with right hand side and Dirichlet boundary conditions defined in such a way that the exact solution is

$$
p_{\mathrm{ex}}(x, y) := x^2 y + \sin(2\pi x) \sin(2\pi y) + 2.
\tag{6.2}
$$

We will show, in a loglog scale, the convergence curves of the error in $L^2$ and $H^1$ between $p_{\mathrm{ex}}$ and the solution $p_h$ given by the Virtual Element Method (4.17). As the VEM solution $p_h$ is not explicitly known inside the elements, we compare $p_{\mathrm{ex}}$ with the $L^2-$projection of $p_h$ onto $\mathbb{P}_k$, that is, with $\Pi_k^0 p_h$. We will also show the behaviour of $|p_{\mathrm{ex}} - \Pi_k^0 p_h|$ at the maximum point of $p_{\mathrm{ex}}$ which is approximately at $(x_{\max}, y_{\max}) = (0.781, 0.766)$.
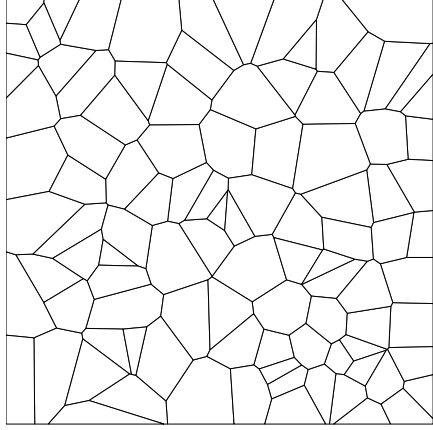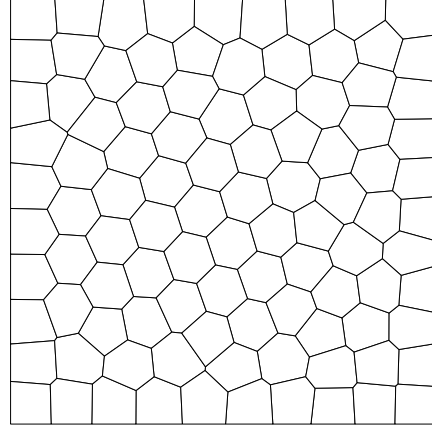
Fig. 1. `Lloyd-0` mesh                    Fig. 2. `Lloyd-100` mesh

### 6.1. *Meshes*

For the convergence test we consider four sequences of meshes.

The first sequence of meshes (labelled `Lloyd-0`) is a random Voronoi polygonal tessellation of the unit square in 25, 100, 400 and 1600 polygons. The second sequence (labelled `Lloyd-100`) is obtained starting from the previous one and performing 100 Lloyd iterations leading to a Centroidal Voronoi Tessellation (CVT) (see e.g. [23]). The 100-polygon mesh of each family is shown in Fig. 1 (`Lloyd-0`) and in Fig. 2 (`Lloyd-100`) respectively.

The third sequence of meshes (labelled `square`) is simply a decomposition of the domain in 25, 100, 400 and 1600 equal squares, while the fourth sequence (labelled `concave`) is obtained from the previous one by subdividing each small square into two non-convex (quite nasty) polygons. As before, the second meshes of the two sequences are shown in Fig. 3 and in Fig. 4 respectively.

### 6.2. *Case $k = 1$*

We start to show the convergence results for $k = 1$. In Figs. 5 and 6 we report the relative error in $L^2$ and $H^1$, respectively, for the four mesh sequences. In Fig. 7 we report the relative error at the maximum point $(x_{\max}, y_{\max})$. Finally, Fig. 8 shows the relative error in $L^2$ obtained with the method (4.18) (that is, the simple-minded extension of [5]). As observed in Remark 4.3, $\Pi_k^0 \nabla \equiv \nabla \Pi_k^\nabla$ for $k = 1$, hence the graphs of Fig. 5 and of Fig. 8 are identical.

### 6.3. *Case $k = 4$*

We show the convergence results for $k = 4$; we proceed as done in the case $k = 1$. In Figs. 9 and 10 we report the relative error in $L^2$ and in $H^1$, respectively, on
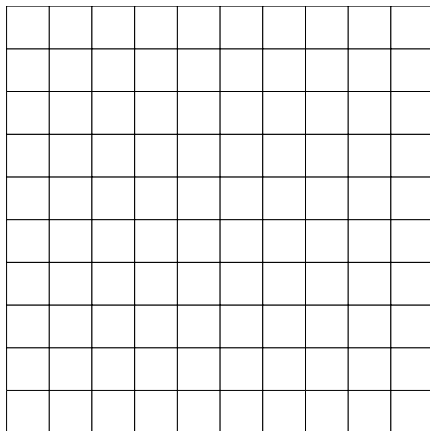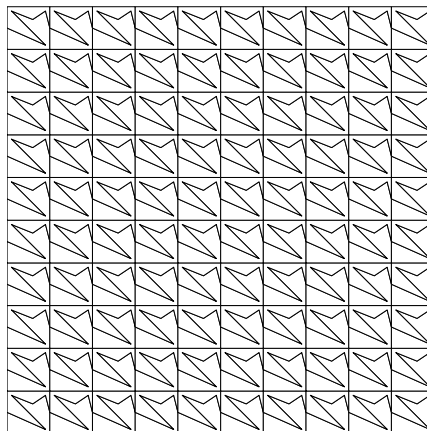
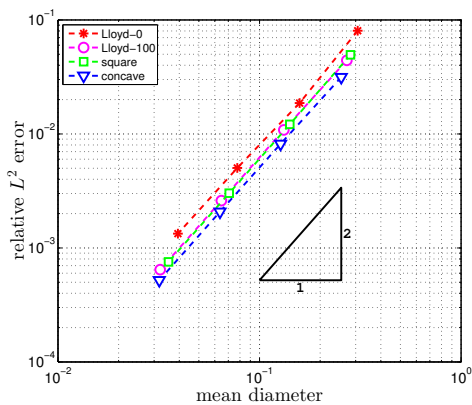Fig. 3. `square` mesh



Fig. 4. `concave` mesh
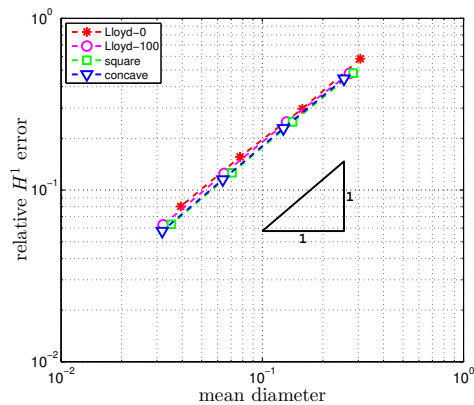


Fig. 5. $k = 1$, relative $L^2$ error



Fig. 6. $k = 1$, relative $H^1$ error

the four mesh sequences. In Fig. 11 we report the relative error at the maximum point $(x_{\max}, y_{\max})$. The last figure (Fig. 12) shows the relative error in $L^2$ obtained with the method (4.18). As announced, a heavy loss in the order of convergence is produced.

We conclude that the Virtual Element Method behaves as expected and shows a remarkable stability with respect to the shape of the mesh polygons.

## References

1. B. Ahmad, A. Alsaedi, F. Brezzi, L. D. Marini, and A. Russo, *Equivalent projectors for virtual element methods*, Comput. Math. Appl. **66** (2013), no. 3, 376–391.
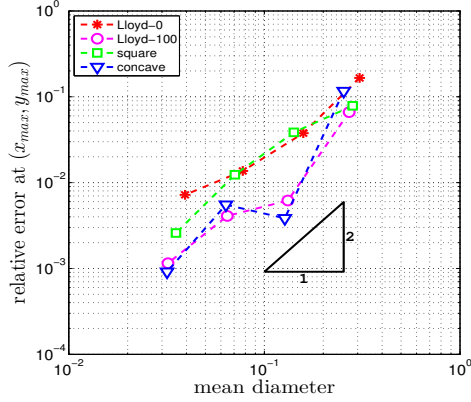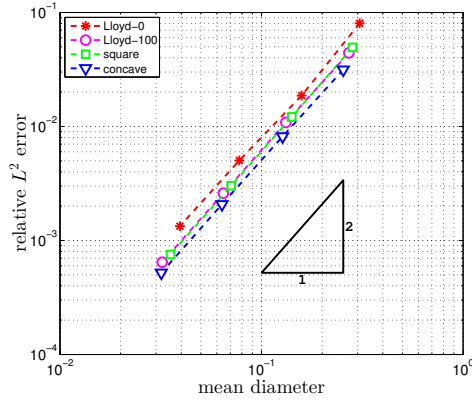
Fig. 7. $k = 1$, relative error at $(x_{\max}, y_{\max})$



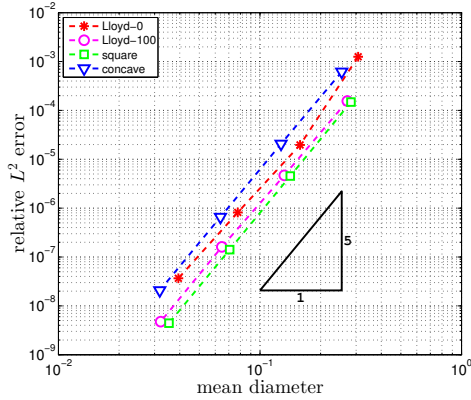Fig. 8. $k = 1$, relative $L^2$ error for method (4.18)



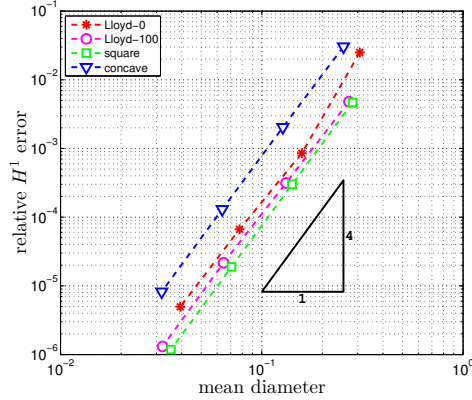Fig. 9. $k = 4$, relative $L^2$ error



Fig. 10. $k = 4$, relative $H^1$ error

2. P. F. Antonietti, L. Beirão da Veiga, D. Mora, and M. Verani, *A stream virtual element formulation of the Stokes problem on polygonal meshes*, SIAM J. Numer. Anal. **52** (2014), no. 1, 386–404.

3. M. Arroyo and M. Ortiz, *Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods*, Internat. J. Numer. Methods Engrg. **65** (2006), no. 13, 2167–2202.

4. I. Babuska, U. Banerjee, and J. E. Osborn, *Survey of meshless and generalized finite element methods: a unified approach*, Acta Numer. **12** (2003), 1–125.

5. L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo, *Basic principles of virtual element methods*, Math. Models Methods Appl. Sci. **23** (2013), no. 1, 199–214.

6. L. Beirão da Veiga, F. Brezzi, and L. D. Marini, *Virtual elements for linear elasticity problems*, SIAM J. Numer. Anal. **51** (2013), no. 2, 794–812.

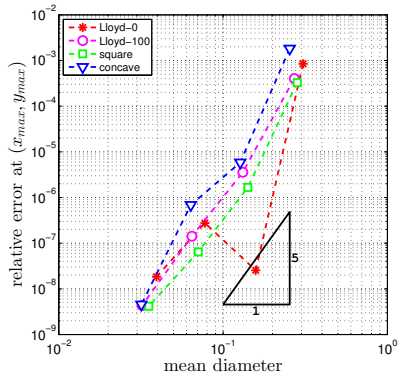7. L. Beirão da Veiga, F. Brezzi, L. D. Marini, and A. Russo, *H(div) and H(curl)-*

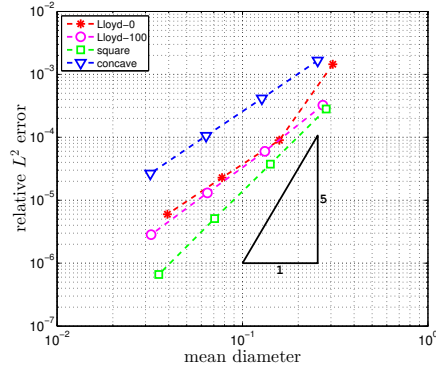Fig. 11. $k = 4$, relative error at $(x_{\max}, y_{\max})$     Fig. 12. $k = 4$, relative $L^2$ error for (4.18)

*conforming VEM*, Numer. Math., online July 17, 2015.

8. ———, *The hitchhiker's guide to the virtual element method*, Math. Models Methods Appl. Sci. **24** (2014), no. 8, 1541–1573.

9. L. Beirão da Veiga, K. Lipnikov, and G. Manzini, *Arbitrary-order nodal mimetic discretizations of elliptic problems on polygonal meshes*, SIAM J. Numer. Anal. **49** (2011), no. 5, 1737–1760.

10. ———, *The mimetic finite difference method for elliptic problems*, MS&A. Modeling, Simulation and Applications, vol. 11, Springer-Verlag, 2014.

11. M. F. Benedetto, S. Berrone, S. Pieraccini, and S. Scialò, *The virtual element method for discrete fracture network simulations*, Comput. Methods Appl. Mech. Engrg. **280** (2014), 135–156.

12. J. E. Bishop, *A displacement-based finite element formulation for general polyhedra using harmonic shape functions*, Internat. J. Numer. Methods Engrg. **97** (2014), no. 1, 1–31.

13. P. B. Bochev and J. M. Hyman, *Principles of mimetic discretizations of differential operators*, Compatible spatial discretizations, IMA Vol. Math. Appl., vol. 142, Springer, New York, 2006, pp. 89–119.

14. F. Brezzi, A. Buffa, and K. Lipnikov, *Mimetic finite differences for elliptic problems*, M2AN Math. Model. Numer. Anal. **43** (2009), no. 2, 277–295.

15. F. Brezzi, K. Lipnikov, and M. Shashkov, *Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes*, SIAM J. Numer. Anal. **43** (2005), no. 5, 1872–1896.

16. F. Brezzi and L. D. Marini, *Virtual element methods for plate bending problems*, Comput. Methods Appl. Mech. Engrg. **253** (2013), 455–462.

17. A. Cangiani, G. Manzini, A. Russo, and N. Sukumar, *Hourglass stabilization and the virtual element method*, Internat. J. Numer. Methods Engrg. **102** (2015), no. 3-4, 404–436.

18. B. Cockburn, *The hybridizable discontinuous Galerkin methods*, Proceedings of the International Congress of Mathematicians. Volume IV, Hindustan Book Agency, New Delhi, 2010, pp. 2749–2775.

19. B. Cockburn, J. Gopalakrishnan, and R. Lazarov, *Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems*, SIAM J. Numer. Anal. **47** (2009), no. 2, 1319–1365.

20. D. Di Pietro and A. Alexandre Ern, *A hybrid high-order locking-free method for linear elasticity on general meshes*, Comput. Methods Appl. Mech. Engrg. **283** (2015), no. 0, 1–21.

21. J. Droniou, R. Eymard, T. Gallouët, and R. Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, Math. Models Methods Appl. Sci. **20** (2010), no. 2, 265–295.

22. _____, *Gradient schemes: a generic framework for the discretisation of linear, non-linear and nonlocal elliptic and parabolic equations*, Math. Models Methods Appl. Sci. **23** (2013), no. 13, 2395–2432.

23. Q. Du, V. Faber, and M. Gunzburger, *Centroidal Voronoi tessellations: applications and algorithms*, SIAM Rev. **41** (1999), no. 4, 637–676.

24. M. Floater, A. Gillette, and N. Sukumar, *Gradient bounds for Wachspress coordinates on polytopes*, SIAM J. Numer. Anal. **52** (2014), no. 1, 515–532.

25. T.-P. Fries and T. Belytschko, *The extended/generalized finite element method: an overview of the method and its applications*, Internat. J. Numer. Methods Engrg. **84** (2010), no. 3, 253–304.

26. A. L. Gain, C. Talischi, and G. H. Paulino, *On the Virtual Element Method for three-dimensional linear elasticity problems on arbitrary polyhedral meshes*, Comput. Methods Appl. Mech. Engrg. **282** (2014), 132–160.

27. K. Lipnikov, G. Manzini, and M. Shashkov, *Mimetic finite difference method*, J. Comput. Phys. **257** (2014), no. part B, 1163–1227.

28. G. Manzini, A. Russo, and N. Sukumar, *New perspectives on polygonal and polyhedral finite element methods*, Math. Models Methods Appl. Sci. **24** (2014), no. 8, 1665–1699.

29. S. Mohammadi, *Extended finite element method*, Blackwell Publishing Ltd, 2008.

30. D. Mora, G. Rivera, and R. Rodríguez, *A virtual element method for the Steklov eigenvalue problem*, Math. Models Methods Appl. Sci. **25** (2015), no. 8, 1421–1445.

31. T. Rabczuk, S. Bordas, and G. Zi, *On three-dimensional modelling of crack growth using partition of unity methods*, Computers & Structures **88** (2010), no. 2324, 1391 – 1411, Special Issue: Association of Computational Mechanics  United Kingdom.

32. A. Rand, A. Gillette, and C. Bajaj, *Interpolation error estimates for mean value coordinates over convex polygons*, Advances in Computational Mathematics **39** (2013), no. 2, 327–347.

33. S. Rjasanow and S. Weisser, *Fem with trefftz trial functions on polyhedral elements*, J. of Comp. and Appl. Math. **263** (2014), 202–217.

34. A. H. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Math. Comp. **28** (1974), 959–962.

35. E. Wachspress, *A rational finite element basis*, Academic Press, Inc., New York-London, 1975, Mathematics in Science and Engineering, Vol. 114.

36. J. Wang and X. Ye, *A weak Galerkin mixed finite element method for second order elliptic problems*, Math. Comp. **83** (2014), no. 289, 2101–2126.