DISCONTINUOUS FINITE ELEMENTS FOR DIFFUSION PROBLEMS

F. Brezzi[1,2], M. Manzini[2], D. Marini[1,2], P. Pietra[2], A. Russo[2]

[1] *Dipartimento di Matematica, Università di Pavia, via Ferrata 1, 27100 Pavia, Italy*
[2] *Istituto di Analisi Numerica del CNR, via Abbiategrasso 209, 27100 Pavia, Italy*

Sunto. I metodi agli Elementi Finiti Discontinui sono attualmente molto utilizzati come una variante (o un sottoinsieme) dei metodi ai Volumi Finiti, nell'approssimazione di Equazioni Iperboliche, con particolare riferimento alle Leggi di Conservazione. Si vedano a questo proposito i lavori [9–14] e le referenze in essi contenute. D'altra parte, il loro utilizzo nell'approssimazione di Equazioni Ellittiche, e in particolare di problemi di diffusione, è allo stesso tempo molto più antico e molto più recente. Infatti, l'uso di metodi di penalizzazione per adattare elementi finiti C^0 alla discretizzazione di problemi del quarto ordine risale all'inizio degli anni settanta (si vedano per esempio [15, 16]). Per problemi del secondo ordine, l'uso di elementi finiti discontinui con penalizzazione risale a pochi anni dopo [2, 3], ma è stato in seguito praticamente abbandonato. In tempi recenti, visto il largo utilizzo di elementi finiti discontinui per i problemi iperbolici (come abbiamo già rilevato in precedenza), ci sono stati vari tentativi di applicarli anche alla discretizzazione di problemi in cui è presente un termine diffusivo piccolo ma non trascurabile, come per esempio nelle equazioni di Navier-Stokes con un elevato numero di Reynolds. L'idea è di utilizzare le loro buone proprietà per il trattamento dei termini convettivi, e di "farli funzionare" (in un modo o nell'altro) anche per i termini diffusivi. Tra tutti questi tentativi, ha suscitato un notevole interesse il metodo di Bassi e Rebay [4–7], in cui viene proposto un modo non usuale (ma molto efficace) per trattare i termini diffusivi con elementi finiti discontinui. Lo scopo di questo lavoro è di fornire a questo metodo una solida base matematica, e di indagare i suoi limiti e le sue possibilità.

In particolare, per mettere in evidenza le capacità di questo metodo di trattare i termini diffusivi, in questo lavoro consideriamo come problema modello l'operatore di Laplace in un dominio poligonale. La formulazione originale di Bassi e Rebay [4–6] viene riscritta in un modo nuovo e più elegante, maggiormente adatto ad una indagine di tipo matematico. Mostriamo che la formulazione originale può essere singolare per un problema stazionario (e questo naturalmente non incoraggia ad utilizzarla neanche per problemi dipendenti dal tempo). Inoltre, dimostriamo che la variante proposta in [7] è stabile e che l'errore di discretizzazione è ottimale, a meno di una piccola modifica nella scelta di un parametro. In realtà, allo stato attuale non è ancora chiaro se per il valore del parametro proposto in [7] (semplicemente uguale a 1) lo schema è instabile oppure no. Quello che noi dimostriamo è la stabilità (e la convergenza) del metodo se questo valore è strettamente maggiore di 3. Più precisamente, detta u la soluzione esatta e u_h la soluzione approssimata, dimostriamo che valgono le seguenti stime dell'errore:

- $|\!|\!| u - u_h |\!|\!| \leq C\, h^k |u|_{k+1,\Omega}$,

- $\|u - u_h\|_{0,\Omega} \leq C\, h^{k+1} |u|_{k+1,\Omega}$,

dove $\|\cdot\|_{0,\Omega}$ è la usuale norma L^2 e $|||\cdot|||$ è una norma che dipende dalla triangolazione e che tiene conto della norma L^2 dei gradienti ma anche dei salti tra un elemento e l'altro.

Il lavoro è strutturato nel modo seguente. Dopo l'Introduzione, nella Sezione 2 viene presentato il problema modello e la formulazione originale di [4–6]. Nella Sezione 3 mostriamo con un controesempio l'instabilità della formulazione originale e nella Sezione 4 descriviamo, dal nostro punto di vista, la stabilizzazione di [7]. La Sezione 5 infine contiene le dimostrazioni di stabilità e di convergenza.

1 Introduction

Discontinuous Finite Element Methods are nowadays rather widely used, as a variant (or subclass) of Finite Volumes, in the numerical treatment of Hyperbolic Equations, in particular related to Conservation Laws. See for instance [9–14] and the references therein. On the other hand, their use in the treatment of Elliptic Equations, in particular related to diffusion problems, is, at the same time, much older and much more recent. Indeed, the use of penalty methods to adapt C^0 elements to the discretization of fourth order problems goes back to the early Seventies (see [15, 16]). For second order problems, the use of discontinuous elements with penalty can be traced back to a few years later, in [2, 3], but was afterwards rather abandoned. In recent times, as the use of discontinuous elements became fashionable for hyperbolic problems (as we have seen), there has been an attempt to use them as well in problems where small (but not negligible) diffusive terms are present together with major convective terms (as in the Navier-Stokes equations with high Reynolds number). The idea being, clearly, to use their good features for treating the convective terms, and "make them work" (one way or another) for treating the diffusive part as well. Among these attempts, considerable interest has been aroused by the method of Bassi and Rebay [4–7], in which a rather peculiar (but quite effective) way of using discontinuous finite elements for diffusive terms has been proposed. The aim of this paper is to provide a solid mathematical background for this approach, and to trace properly its limits and its capabilities.

In particular, we consider as a model problem the Laplace operator in a polygonal domain, in order to cast out the ability of the method to deal with diffusive terms. We rewrite the original formulation of [4–6] in a new and more elegant way, better suited for a mathematical investigation. We show that the original formulation can be rank-deficient when applied to stationary problems (and this, somehow, does not encourage its use in time dependent problems either). On the other hand, the variant proposed in [7] can be proven to be stable and optimal order accurate, at least with a minor modification in the choice of a parameter. Indeed, it is not yet clear whether for the value of the parameter proposed in [7] (simply equal to 1) the scheme is unstable or not. What we prove here is that stability (together with all the other nice and desirable properties) is ensured if this value is strictly bigger than 3.

An outline of the paper is the following. In Section 2 we present the problem and the original formulation of [4–6], rearranging terms and giving it a formulation better suited for theoretical investigation. In Section 3 we show the potential instability of the original formulation, with a suitable counter-example. In Section 4 we present, in our framework, the stabilized version of [7]. Stability proofs and error estimates are then reported in Section 5.

Aknowledgments. This research has been carried out within the project MIGALE, financed by ENEL-CRT (Pisa). We are also grateful, for the fruitful discussions we had with them, to F. Bassi and S. Rebay as well as to L. Brusa and S. Arnelli (from CISE) who, moreover, discovered experimentally the instability of the original formulation in the one-dimensional case.

2 Description of the Method

Let Ω be a two-dimensional convex polygonal domain, with boundary $\Gamma = \partial\Omega$, partitioned as $\Gamma = \overline{\Gamma}_N \cup \overline{\Gamma}_D$, with $\Gamma_N \cap \Gamma_D = \emptyset$. We consider the following boundary value problem:

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = g_D & \text{on } \Gamma_D, \\ \dfrac{\partial u}{\partial n} = g_N & \text{on } \Gamma_N, \end{cases} \tag{1}$$

where f, g_D and g_N satisfy the usual regularity assumptions needed to write (1) in variational form (see e.g. [8]). By introducing the auxiliary vector variable $\boldsymbol{\theta} = \nabla u$, problem (1) can be rewritten as

$$\begin{cases} \boldsymbol{\theta} - \nabla u = 0 & \text{in } \Omega, \\ -\operatorname{div}\boldsymbol{\theta} = f & \text{in } \Omega, \\ u = g_D & \text{on } \Gamma_D, \\ \boldsymbol{\theta}\cdot\boldsymbol{n} = g_N & \text{on } \Gamma_N, \end{cases} \tag{2}$$

where $\boldsymbol{n}$ is the outward normal unit vector to $\partial\Omega$. Let $\{\mathscr{T}_h\}_h$ be a regular family of triangulations of Ω, in the sense of [8]; we shall indicate by E the triangles of $\mathscr{T}_h$, with $h = \max_{E\in\mathscr{T}_h} \operatorname{diam}(E)$. We denote by $\mathscr{E}_h$ the set of all edges of $\mathscr{T}_h$, by $\mathscr{E}_h'$ the set of the internal edges, by $\mathscr{E}_h^D$ the set of edges on Γ_D, and by $\mathscr{E}_h^N$ the set of edges on Γ_N. The length of the edge $e \in \mathscr{E}_h$ will be denoted by h_e.

The finite element spaces for the approximation of u and $\boldsymbol{\theta}$ will be denoted by V_h and $\boldsymbol{W}_h$ respectively. The precise definition of V_h and $\boldsymbol{W}_h$ will be given later; for now, we only assume that functions in V_h and $\boldsymbol{W}_h$ are (possibly) *discontinuous* along the edges, and no boundary values are enforced. We also require that $\boldsymbol{W}_h$ contain the gradients of the functions in V_h. We start by multiplying the first equation of (2) by a test vector $\boldsymbol{\tau}$ and we integrate over Ω. We have:

$$\int_\Omega \boldsymbol{\theta}\cdot\boldsymbol{\tau} - \int_\Omega \nabla u\cdot\boldsymbol{\tau} = 0. \tag{3}$$

We write the second integral as the sum of integrals on the triangles of $\mathscr{T}_h$ and we integrate by parts:

$$\int_\Omega \nabla u \cdot \boldsymbol{\tau} = \sum_{E \in \mathscr{T}_h} \int_E \nabla u \cdot \boldsymbol{\tau} = \sum_{E \in \mathscr{T}_h} \left[- \int_E u \, \mathrm{div}\, \boldsymbol{\tau} + \int_{\partial E} u \, \boldsymbol{\tau} \cdot \boldsymbol{n} \right]. \tag{4}$$

At this point, in order to get the scheme, we would like to insert into (3)–(4) the finite element approximations $u_h \in V_h$ and $\boldsymbol{\tau}_h \in \boldsymbol{W}_h$ instead of u and $\boldsymbol{\tau}$. Since we want u_h to be *discontinuous* across elements, we have to give a meaning to the "value" of u_h on the edges. We shall denote this quantity by u_h^0, defined in the following way. Let u_h^{ext} denote the value of u_h on the element facing the element under consideration across the edge e. Then we define:

- $u_h^0 = \dfrac{u_h + u_h^{\mathrm{ext}}}{2}$ on internal edges $e \in \mathscr{E}_h'$;
- $u_h^0 = g_D$ on Dirichlet boundary edges $e \in \mathscr{E}_h^D$;
- $u_h^0 = u_h$ on Neumann boundary edges $e \in \mathscr{E}_h^N$.

(Clearly, equality $u_h^0 = g_D$ has to be intended in a weak sense (see (10) below), and not pointwise.) Equation (4) then becomes:

$$\sum_{E \in \mathscr{T}_h} \int_E \nabla u_h \cdot \boldsymbol{\tau}_h = \sum_{E \in \mathscr{T}_h} \left[- \int_E u_h \, \mathrm{div}\, \boldsymbol{\tau}_h + \int_{\partial E} u_h^0 \, \boldsymbol{\tau}_h \cdot \boldsymbol{n} \right]. \tag{5}$$

By integrating again by parts $\int_E u_h \, \mathrm{div}\, \boldsymbol{\tau}_h$, we arrive at the following equation:

$$\sum_{E \in \mathscr{T}_h} \left[\int_E \boldsymbol{\theta}_h \cdot \boldsymbol{\tau}_h - \int_E \nabla u_h \cdot \boldsymbol{\tau}_h - \int_{\partial E} (u_h^0 - u_h) \, \boldsymbol{\tau}_h \cdot \boldsymbol{n} \right] = 0 \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h. \tag{6}$$

Consider now the second equation of (2). Multiplication by a test function v and elementwise integration by parts give:

$$\sum_{E \in \mathscr{T}_h} \left[\int_E \boldsymbol{\theta} \cdot \nabla v - \int_{\partial E} \boldsymbol{\theta} \cdot \boldsymbol{n} \, v \right] = \int_\Omega f v. \tag{7}$$

Before writing the discrete equation, we have to define the value of $\boldsymbol{\theta}_h$ on the interfaces. As we did for the definition of u_h^0, we denote by $\boldsymbol{\theta}_h^{\mathrm{ext}}$ the value of $\boldsymbol{\theta}_h$ on the element facing the element under consideration across the edge e. Then we define:

- $\boldsymbol{\theta}_h^0 = \dfrac{\boldsymbol{\theta}_h + \boldsymbol{\theta}_h^{\mathrm{ext}}}{2}$ on internal edges $e \in \mathscr{E}_h$;
- $\boldsymbol{\theta}_h^0 = \boldsymbol{\theta}_h$ on Dirichlet boundary edges $e \in \mathscr{E}_h^D$;
- $\boldsymbol{\theta}_h^0 \cdot \boldsymbol{n} = g_N$ on Neumann boundary edges $e \in \mathscr{E}_h^N$.

(As before, equality $\boldsymbol{\theta}_h^0 \cdot \boldsymbol{n} = g_N$ is in a weak sense (see (13) below).) We have then the following equation:

$$\sum_{E \in \mathscr{T}_h} \left[\int_E \boldsymbol{\theta}_h \cdot \nabla v_h - \int_{\partial E} \boldsymbol{\theta}_h^0 \cdot \boldsymbol{n}\, v_h \right] = \int_\Omega f v_h \quad \text{for all } v_h \in V_h. \tag{8}$$

The discrete equations (6) and (8) can be set in a more expressive form, by writing the integrals over the element boundaries as a sum on all edges of the triangulation. Given an internal edge e, we will use the notation $(\cdot)^+$ and $(\cdot)^-$ to indicate the value of the generic quantity $(\cdot)$ on the two elements sharing e. If we have a boundary edge (Dirichlet or Neumann), there is only one triangle involved, and the notation $(\cdot)$ without superscript will be employed. There is no special meaning attached to $+$ and $-$; any other symbol to distinguish between two values could be used. We start from equation (6). For each internal edge e there are two contributions from the integrals on the element boundaries:

$$-\int_e (u_h^0 - u_h^+)\, \boldsymbol{\tau}_h^+ \cdot \boldsymbol{n}^+ - \int_e (u_h^0 - u_h^-)\, \boldsymbol{\tau}_h^- \cdot \boldsymbol{n}^-. \tag{9}$$

For a boundary edge, we have to distinguish the Dirichlet case and the Neumann case. If e is a Dirichlet boundary edge, we get

$$-\int_e (g_D - u_h)\, \boldsymbol{\tau}_h \cdot \boldsymbol{n}, \tag{10}$$

while we get no contribution from Neumann boundary edges. If we sum together all contributions from (9) and (10), taking into account that $\boldsymbol{n}^+ = -\boldsymbol{n}^-$, we obtain the following expression:

$$-\sum_{E \in \mathscr{T}_h} \int_{\partial E} (u_h^0 - u_h)\, \boldsymbol{\tau}_h \cdot \boldsymbol{n} = \sum_{e \in \mathscr{E}_h'} \int_e [\![u_h]\!] \cdot \boldsymbol{\tau}_h^0 - \sum_{e \in \mathscr{E}_h^D} \int_e (g_D - u_h)\, \boldsymbol{\tau}_h \cdot \boldsymbol{n}, \tag{11}$$

where the *jump* of u_h is defined (on internal edges) as $[\![u_h]\!] = u_h^+ \boldsymbol{n}^+ + u_h^- \boldsymbol{n}^-$. Equation (6) can then be written as

$$\boxed{\sum_{E \in \mathscr{T}_h} \int_E (\boldsymbol{\theta}_h - \nabla u_h) \cdot \boldsymbol{\tau}_h + \sum_{e \in \mathscr{E}_h'} \int_e [\![u_h]\!] \cdot \boldsymbol{\tau}_h^0 - \sum_{e \in \mathscr{E}_h^D} \int_e (g_D - u_h)\, \boldsymbol{\tau}_h \cdot \boldsymbol{n} = 0.} \tag{12}$$

We then consider equation (8). Proceeding as above, we have:

$$-\sum_{E \in \mathscr{T}_h} \int_{\partial E} \boldsymbol{\theta}_h^0 \cdot \boldsymbol{n}\, v_h = -\sum_{e \in \mathscr{E}_h'} \int_e \boldsymbol{\theta}_h^0 \cdot [\![v_h]\!] - \sum_{e \in \mathscr{E}_h^D} \int_e \boldsymbol{\theta}_h \cdot \boldsymbol{n}\, v_h - \sum_{e \in \mathscr{E}_h^N} \int_e g_N v_h. \tag{13}$$

Hence, equation (8) becomes:

$$\boxed{\sum_{E \in \mathscr{T}_h} \int_E \boldsymbol{\theta}_h \cdot \nabla v_h - \sum_{e \in \mathscr{E}_h'} \int_e \boldsymbol{\theta}_h^0 \cdot [\![v_h]\!] - \sum_{e \in \mathscr{E}_h^D} \int_e \boldsymbol{\theta}_h \cdot \boldsymbol{n}\, v_h - \sum_{e \in \mathscr{E}_h^N} \int_e g_N v_h = \int_\Omega f v_h.} \tag{14}$$

Equations (12) and (14) can be written in a more compact form as a symmetric saddle-point type problem. By defining the two bilinear forms

$$a\left(\boldsymbol{\theta}_h, \boldsymbol{\tau}_h\right) = \int_\Omega \boldsymbol{\theta}_h \cdot \boldsymbol{\tau}_h, \tag{15}$$

$$b\left(u_h, \boldsymbol{\tau}_h\right) = -\sum_{E \in \mathscr{T}_h} \int_E \nabla u_h \cdot \boldsymbol{\tau}_h + \sum_{e \in \mathscr{E}_h^I} \int_e [\![u_h]\!] \cdot \boldsymbol{\tau}_h^0 + \sum_{e \in \mathscr{E}_h^D} \int_e u_h \, \boldsymbol{\tau}_h \cdot \boldsymbol{n}, \tag{16}$$

equations (12) and (14) become

$$\begin{cases} a\left(\boldsymbol{\theta}_h, \boldsymbol{\tau}_h\right) + b\left(u_h, \boldsymbol{\tau}_h\right) = \displaystyle\sum_{e \in \mathscr{E}_h^D} \int_e g_D \, \boldsymbol{\tau}_h \cdot \boldsymbol{n}, \\ -b\left(v_h, \boldsymbol{\theta}_h\right) \qquad\qquad\quad = \displaystyle\int_\Omega f v_h + \sum_{e \in \mathscr{E}_h^N} \int_e g_N v_h. \end{cases} \tag{17}$$

It is then clear that an *Inf-Sup* condition will be needed to ensure existence and uniqueness of the solution of (17). We shall discuss this point in the next section.

The next step is to solve equation (12) for $\boldsymbol{\theta}_h$. To this end, we define the space $\widetilde{V}$ as

$$\widetilde{V} = \prod_{E \in \mathscr{T}_h} H^1(E) \tag{18}$$

and the affine operator $\boldsymbol{R}_{g_D} : \widetilde{V} \to \boldsymbol{W}_h$ in the following way. Given $w \in \widetilde{V}$, we define $\boldsymbol{R}_{g_D}(w) \in \boldsymbol{W}_h$ as the solution of the variational problem

$$\int_\Omega \boldsymbol{R}_{g_D}(w) \cdot \boldsymbol{\tau}_h = -\sum_{e \in \mathscr{E}_h^I} \int_e [\![w]\!] \cdot \boldsymbol{\tau}_h^0 - \sum_{e \in \mathscr{E}_h^D} \int_e w \, \boldsymbol{\tau}_h \cdot \boldsymbol{n} + \sum_{e \in \mathscr{E}_h^D} \int_e g_D \, \boldsymbol{\tau}_h \cdot \boldsymbol{n} \tag{19}$$

for all $\boldsymbol{\tau}_h \in \boldsymbol{W}_h$. We have then from (12)

$$\boxed{\boldsymbol{R}_{g_D}(u_h) = \boldsymbol{\theta}_h - \nabla u_h \quad \text{in } \boldsymbol{W}_h.} \tag{20}$$

We also define the linear operator $\boldsymbol{R} = \boldsymbol{R}_0 : \widetilde{V} \to \boldsymbol{W}_h$ by

$$\int_\Omega \boldsymbol{R}(w) \cdot \boldsymbol{\tau}_h = -\sum_{e \in \mathscr{E}_h^I} \int_e [\![w]\!] \cdot \boldsymbol{\tau}_h^0 - \sum_{e \in \mathscr{E}_h^D} \int_e w \, \boldsymbol{\tau}_h \cdot \boldsymbol{n} \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h, \tag{21}$$

and we notice that the following relationship holds:

$$\int_\Omega \boldsymbol{R}_{g_D}(w) \cdot \boldsymbol{\tau}_h = \int_\Omega \boldsymbol{R}(w) \cdot \boldsymbol{\tau}_h + \sum_{e \in \mathscr{E}_h^D} \int_e g_D \, \boldsymbol{\tau}_h \cdot \boldsymbol{n}. \tag{22}$$

Substituting $\boldsymbol{\theta}_h = \nabla u_h + \boldsymbol{R}_{g_D}(u_h)$ in equation (14) yields

$$\sum_{E\in\mathscr{T}_h}\int_E \left[\nabla u_h + \boldsymbol{R}_{g_D}(u_h)\right]\cdot\nabla v_h \boxed{-\sum_{e\in\mathscr{E}_h'}\int_e \boldsymbol{\theta}^0\cdot[\![v_h]\!] - \sum_{e\in\mathscr{E}_h^D}\int_e \boldsymbol{\theta}_h\cdot\boldsymbol{n}\, v_h} - \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h = \int_\Omega f v_h. \quad (23)$$

On the other hand, equation (21) with $\boldsymbol{\tau}_h = \boldsymbol{\theta}_h$ and $w = v_h$ gives

$$\int_\Omega \boldsymbol{R}(v_h)\cdot\boldsymbol{\theta}_h = \boxed{-\sum_{e\in\mathscr{E}_h'}\int_e [\![v_h]\!]\cdot\boldsymbol{\theta}^0 - \sum_{e\in\mathscr{E}_h^D}\int_e v_h\,\boldsymbol{\theta}_h\cdot\boldsymbol{n}}. \quad (24)$$

The boxed quantities in the previous two equations are equal. Hence, using (24) in (23), since $\boldsymbol{\theta}_h = \nabla u_h + \boldsymbol{R}_{g_D}(u_h)$ we have:

$$\sum_{E\in\mathscr{T}_h}\left\{\int_E \left[\nabla u_h + \boldsymbol{R}_{g_D}(u_h)\right]\cdot\nabla v_h + \int_E \boldsymbol{R}(v_h)\cdot\left[\nabla u_h + \boldsymbol{R}_{g_D}(u_h)\right]\right\} = \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h + \int_\Omega f v_h. \quad (25)$$

Regrouping terms, the scheme becomes:

$$\sum_{E\in\mathscr{T}_h}\int_E \left[\nabla u_h + \boldsymbol{R}_{g_D}(u_h)\right]\cdot\left[\nabla v_h + \boldsymbol{R}(v_h)\right] = \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h + \int_\Omega f v_h, \quad (26)$$

or equivalently, using (22),

$$\sum_{E\in\mathscr{T}_h}\int_E \left[\nabla u_h + \boldsymbol{R}(u_h)\right]\cdot\left[\nabla v_h + \boldsymbol{R}(v_h)\right] = \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h + \int_\Omega f v_h - \sum_{e\in\mathscr{E}_h^D}\int_e g_D\left[\nabla v_h + \boldsymbol{R}(v_h)\right]\cdot\boldsymbol{n}. \quad (27)$$

3 The $\mathbb{P}_k$-$\mathbb{P}_k$ approximation

We are interested in the study of system (17) for discontinuous, piecewise polynomials of degree $k \geq 1$, for both $v_h \in V_h$ and $\boldsymbol{\tau}_h \in \boldsymbol{W}_h$. Hence, we define, for $k \geq 1$,

$$V_h = \left\{v_h \in L^2(\Omega) \text{ such that } v_{h|E} \in \mathbb{P}_k(E) \text{ for all } E \in \mathscr{T}_h\right\}, \quad (28)$$

$$\boldsymbol{W}_h = \left\{\boldsymbol{\tau}_h \in \left[L^2(\Omega)\right]^2 \text{ such that } \boldsymbol{\tau}_{h|E} \in [\mathbb{P}_k(E)]^2 \text{ for all } E \in \mathscr{T}_h\right\}. \quad (29)$$

As already mentioned, an *Inf-Sup* condition relating the spaces V_h and $\boldsymbol{W}_h$ is needed in order to guarantee the nonsingularity of the matrix associated with (17). Unfortunately, the *Inf-Sup* condition does not hold for this choice of spaces, as we shall show with a counter-example. In the next section we will circumvent this problem by a suitable stabilization. Assume $\Gamma_D = \partial\Omega$, $\Gamma_N = \emptyset$, $f = 0$, and $g_D = 0$, and, for simplicity, let us consider the case $k = 1$. If the matrix associated with (17) is nonsingular, the unique solution of (17) has to be $(\boldsymbol{\theta}_h,\ u_h) = (0, 0)$. Here we construct a pair $(0,\ u_h \neq 0)$ on the mesh of Fig. 1 which is a solution of (17). Such a pair verifies then

$$b\left(u_h, \boldsymbol{\tau}_h\right) = 0 \qquad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h, \tag{30}$$

that is, rewriting (16) as done in (6),

$$\sum_{E\in\mathscr{T}_h} \left[-\int_E \nabla u_h \cdot \boldsymbol{\tau}_h - \int_{\partial E} \left(u_h^0 - u_h\right) \boldsymbol{\tau}_h \cdot \boldsymbol{n} \right] = 0 \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h. \tag{31}$$

Integrating $\int_E \nabla u_h \cdot \boldsymbol{\tau}_h$ by parts, we have

$$\sum_{E\in\mathscr{T}_h} \left[(\operatorname{div} \boldsymbol{\tau}_h)_{|E} \int_E u_h - \int_{\partial E} u_h^0 \boldsymbol{\tau}_h \cdot \boldsymbol{n} \right] = 0 \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h. \tag{32}$$

It is then enough to find a nonzero $u_h \in V_h$ such that

- the mean value of u_h is zero element by element;
- $u_h^0 = 0$, that is $u_h^+ = -u_h^-$ on internal edges.

Such a u_h can be constructed as shown in Figure 1, where we have denoted by $+$, $-$ and 0 the values $+1$, -1 and 0 respectively.

4 Stabilization

The counter-example presented in the previous section shows that problem (17) does not provide any a priori control over the u_h variable. It seems then quite natural to add to the second equation of (17) a term which guarantees such a control. We shall describe here the modification proposed in [7]. Mimicking the definition of $\boldsymbol{R}_{g_D}$, for each Dirichlet boundary edge e of the triangulation we define the affine operator $\boldsymbol{r}_{e,g_D} : \widetilde{V} \to \boldsymbol{W}_h$ as the solution of the problem

$$\int_\Omega \boldsymbol{r}_{e,g_D}(w) \cdot \boldsymbol{\tau}_h = -\int_e w\, \boldsymbol{\tau}_h \cdot \boldsymbol{n} + \int_e g_D \boldsymbol{\tau}_h \cdot \boldsymbol{n} \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h. \tag{33}$$

If e is an internal edge, $\boldsymbol{r}_{e,g_D}(w)$ is defined by

$$\int_\Omega \boldsymbol{r}_{e,g_D}(w) \cdot \boldsymbol{\tau}_h = -\int_e [\![w]\!] \cdot \boldsymbol{\tau}_h^0 \quad \text{for all } \boldsymbol{\tau}_h \in \boldsymbol{W}_h \tag{34}$$

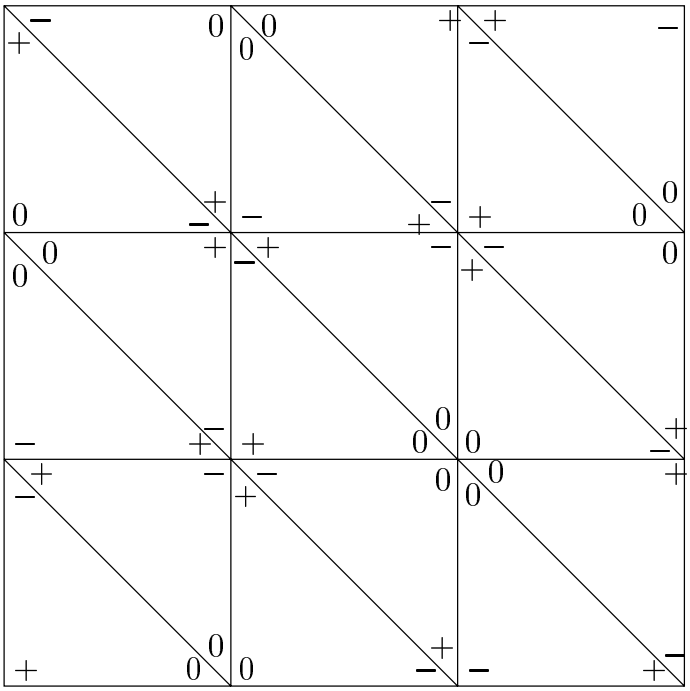


Figure 1: Singularity of the scheme

(note that in this case $\boldsymbol{r}_{e,g_D}$ does not depend on g_D), and for Neumann boundary edges $\boldsymbol{r}_{e,g_D}$ is defined to be zero. Finally, as before we set $\boldsymbol{r}_e = \boldsymbol{r}_{e,0}$. It can be easily seen that the following relationship between $\boldsymbol{R}_{g_D}$ and $\boldsymbol{r}_{e,g_D}$ holds: for any triangle $E \in \mathscr{T}_h$, we have

$$\sum_{e \subset \partial E} \boldsymbol{r}_{e,g_D} = \boldsymbol{R}_{g_D} \quad \text{on } E. \tag{35}$$

If e is an internal edge, from definition (34) it is clear that the support of $\boldsymbol{r}_{e,g_D}$ is contained in the union of the two triangles sharing the edge e. The modification proposed by Bassi and Rebay consists in *substituting* in (26), for each $E \in \mathscr{T}_h$,

$$\text{the term } \boxed{\int_E \boldsymbol{R}_{g_D}(u_h) \cdot \boldsymbol{R}(v_h)} \quad \text{with} \quad \boxed{\sum_{e \subset \partial E} \int_\Omega \boldsymbol{r}_{e,g_D}(u_h) \cdot \boldsymbol{r}_e(v_h)}. \tag{36}$$

This procedure can be interpreted in the following way. As we shall see more precisely in the next Section, the quantity $\boldsymbol{r}_e(v_h)$ allows to control the jump of v_h on e (see Lemma 2); hence, a natural stabilization of (17) consists in *adding* to $b\,(v_h, \boldsymbol{\tau}_h)$ in the second equation the term

$$\beta \sum_{e \in \mathscr{E}_h} \int_\Omega \boldsymbol{r}_{e,g_D}(u_h) \cdot \boldsymbol{r}_e(v_h) \tag{37}$$

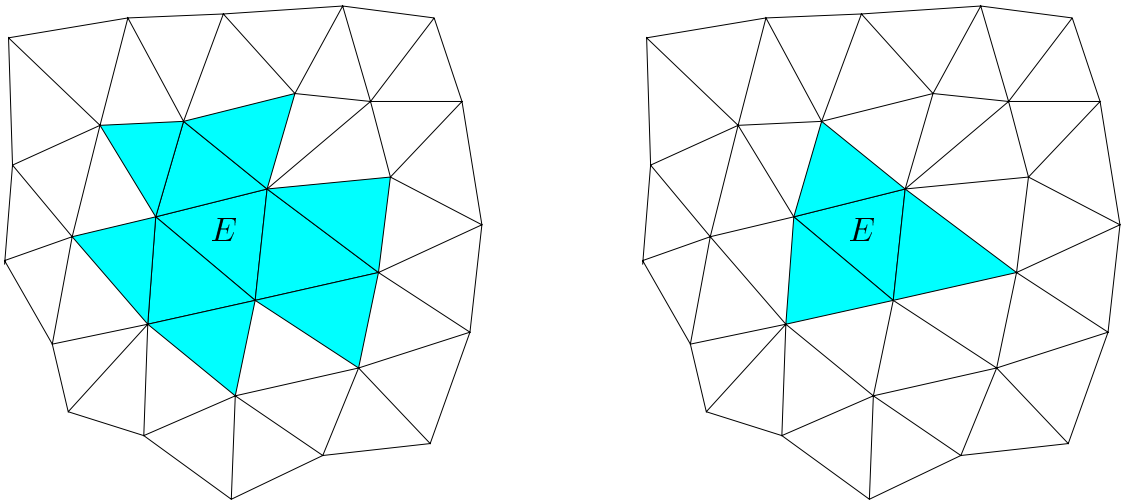
where $\beta > 0$ is a stabilization parameter, thus obtaining the new scheme

$$\sum_{E\in\mathscr{T}_h}\int_E \left[\nabla u_h + \boldsymbol{R}_{g_D}(u_h)\right]\cdot\left[\nabla v_h + \boldsymbol{R}(v_h)\right] + \beta\sum_{e\in\mathscr{E}_h}\int_\Omega \boldsymbol{r}_{e,g_D}(u_h)\cdot\boldsymbol{r}_e(v_h) = \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h + \int_\Omega f v_h. \quad (38)$$

Indeed, as we also point out at the end of the paper, it can be shown that this stabilization works for every positive β. In the next Section, we shall show that if β is large enough ($\beta > 3$), we can also get rid of the term $\int_E \boldsymbol{R}_{g_D}(u_h)\cdot\boldsymbol{R}(v_h)$ in (26), obtaining the following scheme, equivalent to the formulation of Bassi and Rebay (when $\beta = 1$):

$$\sum_{E\in\mathscr{T}_h}\int_E \left[\nabla u_h\cdot\nabla v_h + \nabla u_h\cdot\boldsymbol{R}(v_h) + \boldsymbol{R}_{g_D}(u_h)\cdot\nabla v_h\right] + \beta\sum_{e\in\mathscr{E}_h}\int_\Omega \boldsymbol{r}_{e,g_D}(u_h)\cdot\boldsymbol{r}_e(v_h) = \sum_{e\in\mathscr{E}_h^N}\int_e g_N v_h + \int_\Omega f v_h. \quad (39)$$

The advantage of this scheme with respect to (38) is the following. By a closer inspection of (39) and (38) it can be seen that the stiffness matrix of (39) is much more sparse than that of (38). Indeed, if we take a v_h with support inside one element E (far from the boundary), we see that the elements involved in (38) are always 10, while the elements involved in (39) are always 4, as depicted in Figure 2.



First stabilization (38) *Bassi-Rebay stabilization (39)*

Figure 2: Sparsity of the stiffness matrix

For the sake of simplicity, in this paper we shall limit ourselves to study in detail the case of Dirichlet homogeneous boundary condition; i.e., from now on we will assume

$\Gamma_N = \emptyset$, $\Gamma_D = \partial\Omega$, and $g_D = 0$. Hence, we define the bilinear form $a_h(\cdot,\cdot)$ after (39) as

$$a_h(u_h, v_h) = \sum_{E\in\mathscr{T}_h} \int_E \left[\nabla u_h \cdot \nabla v_h + \nabla u_h \cdot \boldsymbol{R}(v_h) + \boldsymbol{R}(u_h) \cdot \nabla v_h\right] + \beta \sum_{e\in\mathscr{E}_h} \int_\Omega \boldsymbol{r}_e(u_h) \cdot \boldsymbol{r}_e(v_h). \quad (40)$$

The discrete variational formulation then reads

$$\begin{cases} \text{find } u_h \in V_h \text{ such that} \\ a_h(u_h, v_h) = (f, v_h), \quad \text{for all } v_h \in V_h. \end{cases} \quad (41)$$

We indicate with $\|\cdot\|_{m,S}$ and $|\cdot|_{m,S}$ respectively the norm and seminorm in the Sobolev space $H^m(S)$. Error estimates will be obtained in the following mesh-dependent norm:

$$|||v|||^2 = \sum_{E\in\mathscr{T}_h} |v|^2_{1,E} + \sum_{e\in\mathscr{E}_h} \|\boldsymbol{r}_e(v)\|^2_{0,\Omega}, \quad v \in \widetilde{V}. \quad (42)$$

Notice that $|||\cdot|||$ is only a semi-norm in $\widetilde{V}$, while it is a norm in $V_h + H^1_0(\Omega)$.

Statement of the results.
In the next section we shall show coercivity and continuity on V_h of the bilinear form $a_h(\cdot,\cdot)$ with respect to the norm $|||\cdot|||$. Moreover, under the regularity assumption $u \in H^{k+1}(\Omega) \cap H^1_0(\Omega)$, with $k \geq 1$, we shall prove the following error estimates:

- $|||u - u_h||| \leq C\, h^k |u|_{k+1,\Omega}$,
- $\|u - u_h\|_{0,\Omega} \leq C\, h^{k+1} |u|_{k+1,\Omega}$,

where u is the exact solution and u_h is the solution of (41).

5 Proof of the theoretical results

We start by showing that the bilinear form $a_h(\cdot\,,\cdot)$ is coercive and continuous with respect to the norm $|||\cdot|||$.

Proposition 1 *There exists a constant $M > 0$, independent of h, such that*

(i) $a_h(u_h, v_h) \leq M\, |||u_h|||\, |||v_h|||$, *for all $u_h, v_h \in V_h$.*

Moreover, if $\beta > 3$, there exists a constant $\alpha > 0$, independent of h, such that

(ii) $a_h(v_h, v_h) \geq \alpha\, |||v_h|||^2$, *for all $v_h \in V_h$.*

Proof. We first observe that an easy consequence of (35) is

$$\|\boldsymbol{R}(v_h)\|_{0,E}^2 \leq 3 \sum_{e \subset \partial E} \|\boldsymbol{r}_e(v_h)\|_{0,E}^2. \tag{43}$$

Relation *(i)* then follows immediately from the definition of $a_h(u_h, v_h)$ and inequality (43). Next, we have

$$\begin{aligned} a_h(v_h, v_h) &= \sum_{E \in \mathscr{T}_h} \left[|v_h|_{1,E}^2 + 2 \int_E \boldsymbol{R}(v_h) \cdot \nabla v_h + \beta \sum_{e \subset \partial E} \|\boldsymbol{r}_e(v_h)\|_{0,E}^2 \right] \\ &\geq \sum_{E \in \mathscr{T}_h} \left[(1-\varepsilon)|v_h|_{1,E}^2 - \frac{1}{\varepsilon} \|\boldsymbol{R}(v_h)\|_{0,E}^2 + \beta \sum_{e \subset \partial E} \|\boldsymbol{r}_e(v_h)\|_{0,E}^2 \right] \\ &\geq \sum_{E \in \mathscr{T}_h} \left[(1-\varepsilon)|v_h|_{1,E}^2 + \left(\beta - \frac{3}{\varepsilon} \right) \sum_{e \subset \partial E} \|\boldsymbol{r}_e(v_h)\|_{0,E}^2 \right]. \end{aligned}$$

Then *(ii)* holds with $\alpha = \min\left(1-\varepsilon, \beta - \dfrac{3}{\varepsilon}\right)$, which is positive if $\beta > 3$ whenever $\varepsilon < \dfrac{3}{\beta}$.

■

For notational simplicity, we define the jump of v_h also on Dirichlet boundary edges e as

$$[\![v_h]\!] = v_h \boldsymbol{n} \quad \text{on } e. \tag{44}$$

Recall that if e is a Dirichlet boundary edge, we have defined $(\cdot)^0$ as $(\cdot)$ on e. The following properties will be used thoroughly.

Lemma 1 *Let $\varphi \in H^1(\Omega)$, with $\Delta\varphi \in L^2(\Omega)$ and let $v_h \in V_h$. Then*

$$\sum_{E \in \mathscr{T}_h} \int_{\partial E} v_h \frac{\partial \varphi}{\partial n} = \sum_{e \in \mathscr{E}_h} \int_e [\![v_h]\!] \cdot (\nabla \varphi)^0. \tag{45}$$

Proof. The regularity of φ implies that $\nabla\varphi \cdot \boldsymbol{n}$ is continuous across the interelement boundaries, that is $\boldsymbol{n}^+ \cdot \nabla\varphi^+ + \boldsymbol{n}^- \cdot \nabla\varphi^- = 0$. Thus, we have

$$\begin{aligned} \sum_{E \in \mathscr{T}_h} \int_{\partial E} v_h \frac{\partial \varphi}{\partial n} &= \sum_{e \in \mathscr{E}_h'} \int_e \left[v_h^+ \boldsymbol{n}^+ \cdot \nabla\varphi^+ + v_h^- \boldsymbol{n}^- \cdot \nabla\varphi^- \right] + \sum_{e \in \mathscr{E}_h^D} \int_e v_h \boldsymbol{n} \cdot \nabla\varphi \\ &= \sum_{e \in \mathscr{E}_h'} \int_e \left[v_h^+ \boldsymbol{n}^+ \cdot \frac{\nabla\varphi^+ + \nabla\varphi^-}{2} + v_h^- \boldsymbol{n}^- \cdot \frac{\nabla\varphi^+ + \nabla\varphi^-}{2} \right] + \sum_{e \in \mathscr{E}_h^D} \int_e v_h \boldsymbol{n} \cdot \nabla\varphi \\ &= \sum_{e \in \mathscr{E}_h} \int_e [\![v_h]\!] \cdot (\nabla\varphi)^0. \end{aligned}$$

■

Lemma 2 *There exist two positive constants C_1 and C_2, independent of h, such that*

(i) $\|[\![v_h]\!]\|_{0,e} \le C_1 h_e^{1/2} \|\boldsymbol{r}_e(v_h)\|_{0,\Omega}$

(ii) $\|\boldsymbol{r}_e(v_h)\|_{0,\Omega} \le C_2 h_e^{-1/2} \|[\![v_h]\!]\|_{0,e}$

for each $v_h \in V_h$ and for each $e \in \mathscr{E}_h$.

Proof. For any edge $e \subset \partial E$, given $\varphi_h \in \mathbb{P}_k(e)$, let us define $P_e(\varphi_h) \in \mathbb{P}_k(E)$ as the extension of φ_h on E, which is constant along the lines orthogonal to the edge e. It follows immediately that

$$\|P_e(\varphi_h)\|_{0,E} \le C h_e^{1/2} \|\varphi_h\|_{0,e}. \tag{46}$$

For a vector $\boldsymbol{\varphi}_h \in [\mathbb{P}_k(e)]^2$, $P_e(\boldsymbol{\varphi}_h)$ is defined componentwise. Taking $\boldsymbol{\tau}_h \in \boldsymbol{W}_h$ defined by $\boldsymbol{\tau}_{h|E} = P_e([\![v_h]\!])$, $\boldsymbol{\tau}_h = 0$ elsewhere, and $w = v_h$ in (34) gives

$$\|[\![v_h]\!]\|_{0,e}^2 \le \int_E |\boldsymbol{r}_e(v_h) \cdot P_e([\![v_h]\!])| \le \|\boldsymbol{r}_e(v_h)\|_{0,E} \, \|P_e([\![v_h]\!])\|_{0,E} \, . \tag{47}$$

Then, inequality *(i)* follows by using (46) in (47) and summing over all $E \in \mathscr{T}_h$. In order to prove *(ii)*, we take $\boldsymbol{\tau}_h = \boldsymbol{r}_e(v_h)$ in (34). Thus, we find

$$\|\boldsymbol{r}_e(v_h)\|_{0,\Omega}^2 \le \|[\![v_h]\!]\|_{0,e} \, \|(\boldsymbol{r}_e(v_h))^0\|_{0,e} \le C \, h_e^{-1/2} \, \|[\![v_h]\!]\|_{0,e} \, \|\boldsymbol{r}_e(v_h)\|_{0,\Omega}, \tag{48}$$

where the last inequality follows by a simple scaling argument, see [17]. ∎

Lemma 3 *Let $\varphi \in \prod_{E \in \mathscr{T}_h} H^2(E)$ and $v_h \in V_h$. Then*

$$\sum_{e \in \mathscr{E}_h} \int_e [\![v_h]\!] \cdot (\nabla \varphi)^0 \le C \sum_{E \in \mathscr{T}_h} \sum_{e \subset \partial E} \|\boldsymbol{r}_e(v_h)\|_{0,\Omega} \left(|\varphi|_{1,E} + h_e |\varphi|_{2,E} \right). \tag{49}$$

Proof. The trace inequality (see [1, 2]) gives

$$\|\nabla \varphi\|_{0,e}^2 \le C \, \left(h_e^{-1} |\varphi|_{1,E}^2 + h_e |\varphi|_{2,E}^2 \right), \tag{50}$$

where the constant C depends only on the minimal angle bound. By using (50) and Lemma 2 *(i)* we obtain

$$\int_e [\![v_h]\!] \cdot \nabla \varphi \le \, \|[\![v_h]\!]\|_{0,e} \, \|\nabla \varphi\|_{0,e} \le \, C \|\boldsymbol{r}_e(v_h)\|_{0,\Omega} \left(|\varphi|_{1,E} + h_e |\varphi|_{2,E} \right). \tag{51}$$

Thus, the thesis follows since

$$\sum_{e \in \mathscr{E}_h} \int_e [\![v_h]\!] \cdot (\nabla \varphi)^0 = \frac{1}{2} \sum_{E \in \mathscr{T}_h} \sum_{e \subset \partial E} \int_e [\![v_h]\!] \cdot \nabla \varphi. \qquad \blacksquare$$

We are now able to prove our convergence theorem.

Theorem 1 *If u is the exact solution and u_h is the solution of (41), the following estimate holds:*

$$|||u - u_h||| \leq Ch^k |u|_{k+1,\Omega}. \tag{52}$$

Proof. Let $\widetilde{V}_h$ be the usual conforming finite element space

$$\widetilde{V}_h = \{v_h \in H_0^1(\Omega) \text{ such that } v_{h|E} \in \mathbb{P}_k(E)\}, \quad \text{with } k \geq 1, \tag{53}$$

and let $u_I \in \widetilde{V}_h$ be the $\mathbb{P}_k$-interpolant of u. The definition of $|||\cdot|||$ yields

$$|||u - u_I|||^2 = \sum_{E \in \mathscr{T}_h} |u - u_I|_{1,E}^2 \leq C\, h^{2k}\, |u|_{k+1,\Omega}^2. \tag{54}$$

From Proposition 1 we have

$$\begin{aligned} \alpha |||u_I - u_h|||^2 &\leq a_h(u_I - u_h, u_I - u_h) \\ &= a_h(u_I - u_h, u_I - u) + a_h(u_I - u_h, u - u_h) \\ &\leq M |||u_I - u_h||| \, |||u_I - u||| + a_h(u_I - u_h, u - u_h). \end{aligned} \tag{55}$$

The definition of the bilinear form $a_h(\cdot\,,\cdot)$, integration by parts, application of Lemma 1 and the definition of $\boldsymbol{R}(v_h)$ (see (21)) give

$$\begin{aligned} a_h(u - u_h, u_I - u_h) &= a_h(u, u_I - u_h) - a_h(u_h, u_I - u_h) \\ &= \sum_{E \in \mathscr{T}_h} \left[\int_E \nabla u \cdot \nabla (u_I - u_h) + \int_E \nabla u \cdot \boldsymbol{R}(u_I - u_h) - \int_E f\ (u_I - u_h) \right] \\ &= \sum_{E \in \mathscr{T}_h} \left[\int_E (-\Delta u - f)(u_I - u_h) + \int_{\partial E} \frac{\partial u}{\partial n} (u_I - u_h) + \int_E \nabla u \cdot \boldsymbol{R}(u_I - u_h) \right] \\ &= \sum_{e \in \mathscr{E}_h} \int_e [\![u_I - u_h]\!] \cdot (\nabla u)^0 + \sum_{E \in \mathscr{T}_h} \int_E \nabla u \cdot \boldsymbol{R}(u_I - u_h) \\ &= \sum_{e \in \mathscr{E}_h} \int_e [\![u_I - u_h]\!] \cdot [\nabla (u - u_I)]^0 + \sum_{E \in \mathscr{T}_h} \int_E \nabla\,(u - u_I) \cdot \boldsymbol{R}(u_I - u_h). \end{aligned} \tag{56}$$

By applying Lemma 3 with $\varphi = u - u_I$ and using the standard interpolation estimates, we obtain

$$\begin{aligned} &\left| \sum_{e \in \mathscr{E}_h} \int_e [\![u_I - u_h]\!] \cdot [\nabla (u - u_I)]^0 \right| \\ &\qquad \leq C \sum_{E \in \mathscr{T}_h} \sum_{e \subset \partial E} \| \boldsymbol{r}_e (u_I - u_h) \|_{0,\Omega} \left(|u - u_I|_{1,E} + h_e\, |u - u_I|_{2,E} \right) \\ &\qquad \leq C\, h^k\, |u|_{k+1,\Omega}\, |||u_I - u_h|||. \end{aligned} \tag{57}$$

Moreover, it is immediate to see that

$$\left| \sum_{E \in \mathscr{T}_h} \int_E \nabla(u - u_I) \cdot \boldsymbol{R}(u_I - u_h) \right| \leq C\, h^k\, |u|_{k+1,\Omega}\, |\!|\!| u_I - u_h |\!|\!|. \tag{58}$$

By inserting inequalities (57)-(58) into (56) and (55), we have

$$\alpha |\!|\!| u_I - u_h |\!|\!|^2 \leq M\, |\!|\!| u_I - u_h |\!|\!|\, |\!|\!| u_I - u |\!|\!| + C\, h^k\, |u|_{k+1,\Omega}\, |\!|\!| u_I - u_h |\!|\!|. \tag{59}$$

The thesis follows from (54) and (59) by the triangle inequality. ■

A duality argument allows to prove

Theorem 2 *If u is the exact solution and u_h the solution of (41), the following estimate holds*

$$\|u - u_h\|_{0,\Omega} \leq C h^{k+1} |u|_{k+1,\Omega}. \tag{60}$$

Proof. Let us consider the following auxiliary problem

$$\begin{cases} -\Delta w = u - u_h & \text{in } \Omega \\ \quad w = 0 & \text{on } \partial\Omega \end{cases} \tag{61}$$

for which the regularity estimate

$$\|w\|_{2,\Omega} \leq C\, \|u - u_h\|_{0,\Omega} \tag{62}$$

holds. We denote by $\widetilde{w}_h \in \widetilde{V}_h$ the solution of the conforming finite element problem

$$\int_\Omega \nabla \widetilde{w}_h \cdot \nabla \widetilde{v}_h = \int_\Omega (u - u_h)\, \widetilde{v}_h \quad \text{for all } \widetilde{v}_h \in \widetilde{V}_h, \tag{63}$$

and we recall that (see e.g. [8])

$$\|w - \widetilde{w}_h\|_{1,\Omega} \leq C\, h |w|_{2,\Omega}. \tag{64}$$

From equation (61), integrating by parts and keeping the interelement terms, since u_h is discontinuous, we obtain

$$\begin{aligned} \|u - u_h\|^2_{0,\Omega} = \sum_{E \in \mathscr{T}_h} -\int_E \Delta w\, (u - u_h) = \sum_{E \in \mathscr{T}_h} \left[\int_E \nabla w \cdot \nabla(u - u_h) + \int_{\partial E} \frac{\partial w}{\partial n}\, u_h \right] = \\ \sum_{E \in \mathscr{T}_h} \left[\int_E \nabla(w - \widetilde{w}_h) \cdot \nabla(u - u_h) + \int_E \nabla \widetilde{w}_h \cdot \nabla(u - u_h) + \int_{\partial E} \frac{\partial w}{\partial n}\, u_h \right]. \end{aligned} \tag{65}$$

Since $\boldsymbol{R}(\widetilde{w}_h) = \boldsymbol{r}_e(\widetilde{w}_h) = 0$ in Ω and u_h is the solution of (41), we have

$$a_h(u_h, \widetilde{w}_h) = \sum_{E\in\mathscr{T}_h} \int_E \left[\nabla u_h + \boldsymbol{R}(u_h)\right] \cdot \nabla \widetilde{w}_h = (f, \widetilde{w}_h), \tag{66}$$

which implies

$$\sum_{E\in\mathscr{T}_h} \int_E \nabla(u - u_h) \cdot \nabla \widetilde{w}_h = \sum_{E\in\mathscr{T}_h} \int_E \boldsymbol{R}(u_h) \cdot \nabla \widetilde{w}_h. \tag{67}$$

Therefore, (65) reduces to

$$\|u - u_h\|^2_{0,\Omega} = \sum_{E\in\mathscr{T}_h} \left[\int_E \nabla(w - \widetilde{w}_h) \cdot \nabla(u - u_h) + \int_E \boldsymbol{R}(u_h) \cdot \nabla \widetilde{w}_h + \int_{\partial E} u_h \frac{\partial w}{\partial n} \right]. \tag{68}$$

By using definition (21) and Lemma 1 we find

$$\sum_{E\in\mathscr{T}_h} \left[\int_E \boldsymbol{R}(u_h) \cdot \nabla \widetilde{w}_h + \int_{\partial E} u_h \frac{\partial w}{\partial n} \right] = \sum_{e\in\mathscr{E}_h} \int_e [\![u_h]\!] \cdot [\nabla(w - \widetilde{w}_h)]^0 \leq$$
$$C \sum_{E\in\mathscr{T}_h} \sum_{e\subset\partial E} \|\boldsymbol{r}_e(u_h)\|_{0,\Omega} \left(|w - \widetilde{w}_h|_{1,E} + h_e |w - \widetilde{w}_h|_{2,E} \right), \tag{69}$$

where the last inequality follows from Lemma 3 with $\varphi = w - \widetilde{w}_h$. Thus, substituting (69) in (68), and using (64) give

$$\|u - u_h\|^2_{0,\Omega} \leq C\, h |\!|\!| u - u_h |\!|\!| \, |w|_{2,\Omega}. \tag{70}$$

Finally, applying estimate (62) and the result of Theorem 1 we conclude the proof. ■

Remark. A quick glance to the proofs of Proposition 1 and Theorems 1 and 2 immediately shows that the same results, namely error estimates (52) and (60), can be obtained for the scheme (38) for every $\beta > 0$.

References

[1] S. Agmon, *Lectures on elliptic boundary value problems*, Van Nostrand Mathematical Studies, Princeton, NJ, 1965.

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

[3] G. A. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comput., 31 (1977), pp. 45–59.

[4] F. Bassi, G. Mariotti, S. Pedinotti, and S. Rebay, *Recent developments on the solution of the full Navier-Stokes equations using a discontinuous finite element method*, in Proceedings of the Ninth International Conference on "Finite Elements in Fluids, New Trend and Applications", M. Morandi Cecchi et al., ed., vol. 1, Venice, October 15–21 1995, pp. 529–539.

[5] F. Bassi and S. Rebay, *Discontinuous finite element high order accurate numerical solution of the compressible Navier-Stokes equations*, in Proceedings of the Conference "Numerical methods for fluid dynamics V", K. W. Morton et al., ed., Oxford: Clarendon Press, April 3–6 1995, pp. 295–302.

[6] ——, *High-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations*, J. Comput. Phys., 131 (1997), pp. 267–279.

[7] F. Bassi, S. Rebay, M. Savini, G. Mariotti, and S. Pedinotti, *A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flow*, in Proceedings of the Second European Conference ASME on Turbomachinery Fluid Dynamics and Thermodynamics, 1995.

[8] P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland Publishing Company, Amsterdam, Holland, 1980.

[9] B. Cockburn, S. Hou, and C. W. Shu, *TVB Runge-Kutta local projection P^1–discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case*, Math. Comp., 54 (1990), pp. 545–581.

[10] B. Cockburn, S. Y. Lin, and C. W. Shu, *TVB Runge-Kutta local projection P^1–discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems*, J. Comput. Phys., 84 (1989), pp. 90–113.

[11] B. Cockburn and C. W. Shu, *The local discontinuous Galerkin method for time-dependent convection diffusion problems*. Submitted to SIAM J. Numer. Anal.

[12] ——, *TVB Runge-Kutta local projection P^1–discontinuous Galerkin finite element method for conservation laws II: General framework*, Math. Comp., 52 (1989), pp. 411–453.

[13] ——, *TVB Runge-Kutta local projection P^1–discontinuous Galerkin finite element method for scalar conservation laws*, M^2AN, 25 (1991), pp. 337–361.

[14] ——, *TVB Runge-Kutta local projection P^1–discontinuous Galerkin finite element method for conservation laws V: Multi-dimensional systems*, Tech. Rep. IMA Preprint Series 1492, Institute for Mathematics and its Applications, University of Minnesota, 1997.

[15] J. Douglas, Jr, *H^1–Galerkin methods for a nonlinear Dirichlet problem*, in Proceedings of the Conference "Mathematical Aspects of the Finite Element Methods", Springer Lecture Notes in Mathematics vol. 606, Rome, 1977, pp. 64–86.

[16] J. Douglas, Jr and T. Dupont, *Interior penalty procedures for elliptic and parabolic Galerkin Method*, Lectures Notes in Physics 58, Springer–Verlag, Berlin, 1976.

[17] T. Dupont and R. Scott, *Polynomial approximation of functions in Sobolev spaces*, Math. Comp., 34 (1980), pp. 441–463.