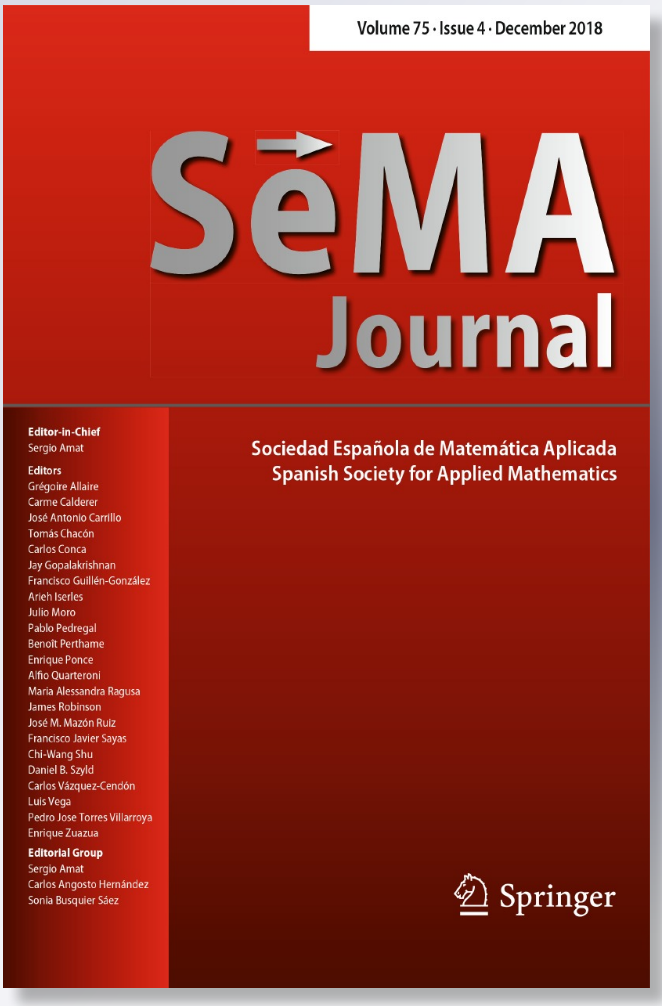
On hp convergence of stabilized finite element methods for the convection–diffusion equation

Ramon Codina

SeMA Journal
Boletin de la Sociedad Española de Matemática Aplicada

ISSN 2254-3902
Volume 75
Number 4

SeMA (2018) 75:591-606
DOI 10.1007/s40324-018-0154-4



Springer

SeMA (2018) 75:591–606
https://doi.org/10.1007/s40324-018-0154-4

On *hp* convergence of stabilized finite element methods for the convection–diffusion equation

Ramon Codina[1]

Received: 5 March 2018 / Accepted: 14 March 2018 / Published online: 22 March 2018

Abstract This work analyzes some aspects of the *hp* convergence of stabilized finite element methods for the convection-diffusion equation when diffusion is small. The methods discussed are classical-residual based stabilization techniques and also projection-based stabilization methods. The theoretical impossibility of obtaining an optimal convergence rate in terms of the polynomial order p for all possible Péclet numbers is explained. The key point turns out to be an inverse estimate that scales as p^2. The use of this estimate is not needed in a particular case of (H^1-)projection-based methods, and therefore the theoretical lack of convergence described does not exist in this case.

Keywords Convection-dominated flows · High order interpolation · Stabilized finite element methods

Mathematics Subject Classification 65N30

1 Introduction

In this work we discuss stability and convergence of classical residual-based stabilized finite element methods (FEMs), as well as some projection-based stabilized methods, both in terms of the mesh size h, and of the polynomial order p, assuming Lagrangian continuous interpolations.

As it is well known, the classical Galerkin method produces globally oscillatory solutions when convection dominates diffusion. If a is the advection velocity, $|a|$ its Euclidean norm and k the diffusion coefficient, the parameter that determines whether this dominance occurs or not is the element Péclet number, defined as $\mathrm{Pe} = |a|hk^{-1}$. Oscillations appear when Pe exceeds a critical value that depends on the polynomial order p.

✉ Ramon Codina
ramon.codina@upc.edu

1 Universitat Politècnica de Catalunya, Jordi Girona 1-3, Edifici C1, 08034 Barcelona, Spain

Residual-based stabilized FEMs add mesh-dependent terms to the discrete variational form of the problem. These terms are multiplied by a stabilization parameter τ that can be expressed as $h^2 k^{-1} \psi(\text{Pe}, p)$, where $\psi(\text{Pe}, p)$ is a function that is designed to achieve stability and optimal convergence in a suitable norm. In the simplest case, this norm is the graph norm of the convection-diffusion operator plus the L^2-norm of the convective term multiplied by $\tau^{1/2}$.

In the most popular residual-based stabilized formulation based on the so-called Variational MultiScale (VMS) concept (see [8] for the origins of the formulation and [4] for an overview), to achieve stability one needs to control second derivatives of the unknown in terms of first derivatives using an inverse estimate, which scales as $p^2 h^{-1}$ (see, for example, [11]). We will show that, because of this scaling, *the resulting a priori error estimates for the stabilized formulation cannot be optimal in p in the range $p \le \text{Pe} \le p^3$*, regardless of the design of the function $\psi(\text{Pe}, p)$. This result, however, is mainly of theoretical interest, since in our numerical tests we have observed optimal convergence for all Péclet numbers. Nevertheless, we also discuss the benefits of using the projection of the residual in the space orthogonal to the finite element space, rather than the residual itself, in the stabilization terms. In particular, optimality in the whole range of Péclet numbers can be achieved using an H^1-projection, motivated by the work developed in [1]. With appropriate projections, and in particular with the orthogonal projection proposed in [2], the resulting method can be simplified to term-by-term stabilization, similar to that encountered in Local Projection Stabilization (LPS) methods (see for example [10]).

The observation that residual-based stabilization methods are not optimal for all Péclet numbers is not new, and can be found in [9], where results presented in [7] are discussed. Here our contribution is to prove that this fact is independent of the choice of the stabilization parameter and to discuss its effect in the variants of stabilization methods described above. For this, we will consider the simplest possible setting, assuming constant coefficients in the convection-diffusion equation, using a discursive approach and avoiding technical details. Some well known methods are revisited, but we consider this necessary to highlight the places where the effect of high order interpolations appears.

The paper is organized as follows. VMS methods are described for the sake of completeness in Sect. 2. Classical residual-based stabilized FEMs are then presented in Sect. 3, emphasizing the role of the order of the polynomial interpolation and presenting their stability and convergence results. Non-residual based stabilization is the topic of Sect. 4. Section 5 presents a spectral method for Burguer's problem that serves to motivate an H^1-orthogonal projection method that happens to be optimal for all Péclet numbers. Conclusions close the paper in Sect. 6.

2 Variational multi-scale methods

Let Ω be a bounded and polyhedral domain of $\mathbb{R}^d$, with d the number of space dimensions, and consider the problem of finding a function $u : \Omega \longrightarrow \mathbb{R}$ solution of the convection-diffusion-reaction problem

$$\begin{aligned} \mathcal{L}u := -k\Delta u + a \cdot \nabla u + su = f \qquad & \text{in } \Omega \\ u = 0 \qquad & \text{on } \partial\Omega \end{aligned}$$

where $k > 0$, $a \in \mathbb{R}^d$ and $s \ge 0$ are constant (for simplicity) and f is a given function.

Let $V = H_0^1(\Omega)$. The variational form of the problem consists in finding $u \in V$ such that

$$B(u, v) = \langle f, v \rangle \quad \forall v \in V \tag{1}$$

where:

$$B(u, v) = k(\nabla u, \nabla v) + (a \cdot \nabla u, v) + s(u, v)$$

For any two functions g_1 and g_2 defined on Ω we use the notation

$$\langle g_1, g_2 \rangle := \int_\Omega g_1 g_2, \quad \text{and} \quad \langle g_1, g_2 \rangle =: (g_1, g_2) \quad \text{if} \quad g_1, g_2 \in L^2(\Omega)$$

The $L^2(\Omega)$-norm is denoted as $\| \cdot \|$, and for a subdomain $\omega \subset \Omega$ its $L^2(\omega)$-norm is denoted as $\| \cdot \|_\omega$.

To approximate problem (1), let $\mathcal{T}_h = \{K\}$ be a finite element partition of the domain Ω. To simplify the exposition, we consider $\mathcal{T}_h$ quasi-uniform, with elements of diameter h. Let V_{hp} be a conforming finite element space to approximate V constructed from $\mathcal{T}_h$ with polynomial interpolation functions of order p. The Galerkin finite element approximation consists of finding $u_h \in V_{hp}$ such that

$$B(u_h, v_h) = \langle f, v_h \rangle \quad \forall v_h \in V_{hp} \tag{2}$$

Stability for problem (2) can be obtained taking $v_h = u_h$. Since

$$B(u_h, u_h) = k\|\nabla u_h\|^2 + s\|u_h\|^2 \tag{3}$$

we see that there is no control on the convective term $\|a \cdot \nabla u_h\|$. If k is small, there will be no effective control on the derivatives of u_h and oscillations will appear. VMS is in fact a framework in which several stabilized FEMs intended to overcome these oscillations can be accommodated [8]. The starting idea is to perform the scale splitting $V = V_{hp} \oplus \tilde{V}$, where $\tilde{V}$ is any space to complete V_{hp} in V, called the space of subscales. The original continuous problem can be written as: find $u_h \in V_{hp}$ and $\tilde{u} \in \tilde{V}$ such that

$$B(u_h, v_h) + B(\tilde{u}, v_h) = \langle f, v_h \rangle \quad \forall v_h \in V_{hp}$$
$$B(u_h, \tilde{v}) + B(\tilde{u}, \tilde{v}) = \langle f, \tilde{v} \rangle \quad \forall \tilde{v} \in \tilde{V}$$

For smooth functions in $\tilde{V}$ we may write this system as

$$B(u_h, v_h) + \langle \tilde{u}, \mathcal{L}^* v_h \rangle = \langle f, v_h \rangle \quad \forall v_h \in V_{hp}$$
$$\langle \mathcal{L} u_h, \tilde{v} \rangle + \langle \mathcal{L} \tilde{u}, \tilde{v} \rangle = \langle f, \tilde{v} \rangle \quad \forall \tilde{v} \in \tilde{V}$$

which may need to be understood in a distributional sense, $\mathcal{L}^*$ being the adjoint of $\mathcal{L}$, given by $\mathcal{L}v = -k\Delta v - a \cdot \nabla v + sv$. The main objective is to model $\tilde{u}$ in terms of u_h and end up with a problem posed in terms of the latter. For these, several approximations may be required.

If $(\cdot, \cdot)_K$ is the $L^2(K)$-inner product over an element domain K, the first approximation we consider is

$$\langle \mathcal{L} v_h, \tilde{v} \rangle \approx \sum_K (\mathcal{L} v_h, \tilde{v})_K =: (\mathcal{L} v_h, \tilde{v})_h \tag{4}$$

which amounts to say that jumps of derivatives of finite element functions on ∂K are neglected. This approximation can be relaxed, as explained in [6].

The second approximation is the most important one, and consists in replacing $\mathcal{L}$ by an algebraic operator of the form

$$\langle \mathcal{L}\tilde{u}, \tilde{v}\rangle \approx \tau^{-1}(\tilde{u}, \tilde{v}), \quad \tau^{-1} = \alpha_1(p)\frac{k}{h^2} + \alpha_2(p)\frac{|a|}{h} + \alpha_3 s \tag{5}$$

where α_1 and α_2 are functions of the polynomial order p whose choice is analyzed below and $\alpha_3 < 1$. This "lumping" of $\mathcal{L}\tilde{u}$ makes the problem for $\tilde{u}$ solvable.

Both approximations (4) and (5) can be justified for example from a formal Fourier analysis requiring $\tau^{-1} \approx \|\mathcal{L}\|$ (see [4] for an overview of ways to justify these approximations). After having introduced them, the problem one has to finally solve is: find $u_h \in V_{hp}$ such that

$$B(u_h, v_h) + (\tilde{u}, \mathcal{L}^* v_h)_h = \langle f, v_h\rangle =: L(v_h) \quad \forall v_h \in V_{hp} \tag{6}$$

$$(\mathcal{L}u_h, \tilde{v})_h + \tau^{-1}(\tilde{u}, \tilde{v}) = \langle f, \tilde{v}\rangle \quad \forall \tilde{v} \in \tilde{V} \tag{7}$$

3 Classical residual-based method

Classical residual-based stabilized FEMs can be recovered by taking as space of subscales the space of element-wise finite element residuals, that is to say, of functions which are of the form $f - \mathcal{L}v_h$ within each element, for a certain $v_h \in V_{hp}$. In this case, from (7) one obtains

$$\tilde{u} = \tau(f - \mathcal{L}u_h) \tag{8}$$

in each element domain K. The resulting stabilized FEM is: find $u_h \in V_{hp}$ such that

$$\begin{aligned} B_s(u_h, v_h) :&= B(u_h, v_h) + \sum_K \tau(-\mathcal{L}^* v_h, \mathcal{L}u_h)_K \\ &= L_s(v_h) := L(v_h) + \sum_K \tau(-\mathcal{L}^*(v_h), f)_K \quad \forall v_h \in V_{hp} \\ \mathcal{L}^* v_h &= -k\Delta v_h - a\cdot\nabla v_h + s v_h \end{aligned} \tag{9}$$

Let us analyze stability of this method, for simplicity considering $s = 0$ (it will be seen in Sect.4 that the misbehavior to be described is in fact independent of s, and even of the model for the subscale). The key point in the analysis is the inverse estimate:

$$\|\nabla v_h\|_K \le C_{\text{inv}}\frac{p^2}{h}\|v_h\|_K \quad \forall v_h \in V_{hp} \tag{10}$$

where C_{inv} is a mesh-independent positive constant. This estimate is also valid if v_h is a derivative of a finite element function; in this case, C_{inv} may change, but we may take the largest inverse constant that make (10) hold for any finite element function and its derivatives. Using it we obtain:

$$\begin{aligned} B_s(v_h, v_h) &= k\|\nabla v_h\|^2 + \tau\|a\cdot\nabla v_h\|^2 - k^2\tau\sum_K \|\Delta v_h\|_K^2 \\ &\ge k\left(1 - kC_{\text{inv}}^2\frac{p^4}{h^2}\tau\right)\|\nabla v_h\|^2 + \tau\|a\cdot\nabla v_h\|^2 \end{aligned}$$

from where it follows stability in the norm

$$|||v_h|||^2 := k\|\nabla v_h\|^2 + \tau\|a\cdot\nabla v_h\|^2$$

provided

$$\alpha_1(p) > C_{\text{inv}}^2 p^4 \tag{11}$$

that is to say, $B_s(v_h, v_h) \gtrsim \|\!|v_h|\!\|^2$, where $\gtrsim$ stands for $\geq$ up to positive constants independent of the discretization and of the physical properties. Condition (11) is in fact the key point in analyzing hp-convergence of stabilized FEMs. Observe that to arrive to (11) we have not taken into account the dependence of τ on $|a|$, since we need to have stability for the whole range of $|a|$, and in particular for $|a| = 0$.

Once stability has been established, let us proceed to analyze convergence assuming that condition (11) holds. Let $u \in V$ be the continuous solution, $\tilde{u}_h \in V_{hp}$ an interpolant and $u_h \in V_{hp}$ the finite element solution. Noting the consistency property

$$B_s(u - u_h, v_h) = 0 \quad \forall v_h \in V_{hp}$$

we have that

$$\begin{aligned}
\|\!|u_h - \tilde{u}_h|\!\|^2 &\lesssim B_s(u_h - \tilde{u}_h, u_h - \tilde{u}_h)\\
&= B_s(u - \tilde{u}_h, u_h - \tilde{u}_h)\\
&\leq k^{1/2}\|\nabla u - \nabla\tilde{u}_h\| k^{1/2}\|\nabla u_h - \nabla\tilde{u}_h\|\\
&\quad + (a \cdot \nabla(u - \tilde{u}_h), u_h - \tilde{u}_h)\\
&\quad + \sum_K \tau k C_{\text{inv}} \frac{p^2}{h} \|\nabla(u_h - \tilde{u}_h)\|_K \| - k\Delta(u - \tilde{u}_h) + a \cdot \nabla(u - \tilde{u}_h)\|_K\\
&\quad + \sum_K \tau \|a \cdot \nabla(u_h - \tilde{u}_h)\|_K \| - k\Delta(u - \tilde{u}_h) + a \cdot \nabla(u - \tilde{u}_h)\|_K
\end{aligned}$$

Let $E(h)$ be the error function of the method. If it satisfies

$$E(h) \geq k^{1/2}\|\nabla u - \nabla\tilde{u}_h\| + \tau^{1/2} \sum_K \| - k\Delta(u - \tilde{u}_h) + a \cdot \nabla(u - \tilde{u}_h)\|_K$$

then we have that

$$\begin{aligned}
\|\!|u_h - \tilde{u}_h|\!\|^2 \lesssim\ & \|\!|u_h - \tilde{u}_h|\!\| E(h)\\
&+ (a \cdot \nabla(u - \tilde{u}_h), u_h - \tilde{u}_h)\\
&+ \|\!|u_h - \tilde{u}_h|\!\| E(h)
\end{aligned}$$

We have to see how to deal with the second term. To this end, we will consider two approaches. First, the strategy used in h-convergence is simple, we just use the following bounds:

$$\begin{aligned}
(a \cdot \nabla(u - \tilde{u}_h), u_h - \tilde{u}_h) &= -(a \cdot \nabla(u_h - \tilde{u}_h), u - \tilde{u}_h)\\
&\leq \tau^{1/2}\|a \cdot \nabla(u_h - \tilde{u}_h)\| \tau^{-1/2}\|u - \tilde{u}_h\|\\
&\lesssim \|\!|u_h - \tilde{u}_h|\!\| \tau^{-1/2}\|u - \tilde{u}_h\|
\end{aligned}$$

so that defining

$$\begin{aligned}
E(h) = k^{1/2}\|\nabla u - \nabla\tilde{u}_h\| &+ \tau^{1/2} \sum_K \| - k\Delta(u - \tilde{u}_h) + a \cdot \nabla(u - \tilde{u}_h)\|_K\\
&+ \tau^{-1/2}\|u - \tilde{u}_h\|
\end{aligned}$$

we get

$$|||u_h - \tilde{u}_h||| \lesssim E(h)$$

Let us analyze the behavior of $E(h)$. Consider the interpolation estimate

$$\|u - \tilde{u}_h\|_{H^s} \lesssim \frac{h^{\xi - s}}{p^{\xi - s}} \|u\|_{H^\xi}, \quad \xi = \min(p+1, r)$$

where r is the Sobolev regularity of u and $\|\cdot\|_{H^q}$ is the norm of $H^q(\Omega)$. Assuming u smooth, so that $\xi = p + 1$, we have:

$$\begin{aligned} E(h) &\lesssim k^{1/2} \frac{h^p}{p^p} \|u\|_{H^{p+1}} + \tau^{1/2} k \frac{h^{p-1}}{p^{p-1}} \|u\|_{H^{p+1}} + \tau^{1/2} |a| \frac{h^p}{p^p} \|u\|_{H^{p+1}} \\ &\quad + \tau^{-1/2} \frac{h^{p+1}}{p^{p+1}} \|u\|_{H^{p+1}} \\ &= \left(k^{1/2} + \tau^{1/2} k \frac{p}{h} + \tau^{1/2} |a| + \tau^{-1/2} \frac{h}{p} \right) \frac{h^p}{p^p} \|u\|_{H^{p+1}} \\ &=: E_0 \frac{h^p}{p^p} \|u\|_{H^{p+1}} \end{aligned}$$

The behavior of E_0 is:

$$\begin{aligned} E_0^2 &\lesssim k + \tau k^2 \frac{p^2}{h^2} + \tau |a|^2 + \tau^{-1} \frac{h^2}{p^2} \\ &\lesssim k + \frac{k^2 \frac{p^2}{h^2}}{\alpha_1(p) \frac{k}{h^2} + \alpha_2(p) \frac{|a|}{h}} + \frac{|a|^2}{\alpha_1(p) \frac{k}{h^2} + \alpha_2(p) \frac{|a|}{h}} \\ &\quad + \left(\alpha_1(p) \frac{k}{h^2} + \alpha_2(p) \frac{|a|}{h} \right) \frac{h^2}{p^2} \end{aligned}$$

Since $\alpha_1(p) \gtrsim p^2$ because of condition (11), the second term can be absorbed by the first. If furthermore, to attain optimality in the error coming from convective term we assume that $\alpha_2(p) \sim p$, i.e. $\alpha_2(p)$ behaves as p when p increases, we have that

$$E_0^2 \lesssim k + |a| \frac{h}{p} + k \frac{\alpha_1(p)}{p^2}$$

Since we had to assume that $\alpha_1(p) \sim p^4$, *the method is not optimal in* p, not even in the diffusion dominated case.

This rough strategy can be slightly improved, as done in [7], using the following bounds:

$$\begin{aligned} &(a \cdot \nabla(u - \tilde{u}_h), u_h - \tilde{u}_h) = -(a \cdot \nabla(u_h - \tilde{u}_h), u - \tilde{u}_h) \\ &\leq \left(\tau^{1/2} \|a \cdot \nabla(u_h - \tilde{u}_h)\| + k^{1/2} \|\nabla(u_h - \tilde{u}_h)\| \right) \min\left(\tau^{-1/2}, \frac{|a|}{k^{1/2}} \right) \|u - \tilde{u}_h\| \\ &\lesssim |||u_h - \tilde{u}_h||| \min\left(\tau^{-1/2}, \frac{|a|}{k^{1/2}} \right) \|u - \tilde{u}_h\| \end{aligned}$$

For p fixed, when k is large (small Péclet number), the previous analysis leads to

$$E_0^2 \lesssim k + |a| \frac{h}{p} + \frac{|a|^2}{k} \frac{h^2}{p^2} \lesssim k$$

so that optimality in the diffusion dominated case is recovered for p fixed. Nevertheless, for k, a fixed, $\min\left(\tau^{-1/2}, \frac{|a|}{k^{1/2}}\right) \to \tau^{-1/2}$ as $p \to \infty$, and we are back to the lack of optimality observed earlier. This can be made more precise as follows. In general, for $\alpha_1(p) \sim p^4$ we have that

$$E_0^2 \lesssim k + \frac{\mathrm{Pe}}{p}k + \min\left(p^2, \frac{\mathrm{Pe}^2}{p^2}\right)k$$

The first two terms are optimal. *Lack of optimality will occur if the last term dominates the other two.* Let us check when this happens:

- $p^2 \leq \frac{\mathrm{Pe}^2}{p^2} \iff \mathrm{Pe} \geq p^2$. Then $E_0^2 \lesssim k + \frac{\mathrm{Pe}}{p}k + p^2 k$. This is suboptimal for $p^2 \geq \frac{\mathrm{Pe}}{p} \iff \mathrm{Pe} \leq p^3$. So, the error estimate is suboptimal for $p^2 \leq \mathrm{Pe} \leq p^3$.
- $p^2 \geq \frac{\mathrm{Pe}^2}{p^2} \iff \mathrm{Pe} \leq p^2$. Then $E_0^2 \lesssim k + \frac{\mathrm{Pe}}{p}k + \frac{\mathrm{Pe}^2}{p^2}k$. This is suboptimal for $\frac{\mathrm{Pe}}{p} > 1 \iff \mathrm{Pe} \geq p$. So, the error estimate is suboptimal for $p \leq \mathrm{Pe} \leq p^2$. This case however is different to the previous one, since for a fixed k the method is optimal in terms of h and p.

Combining the two results, we observe that *the error estimate is suboptimal for* $p \leq \mathrm{Pe} \leq p^3$ (or at least for $p^2 \leq \mathrm{Pe} \leq p^3$, in view of the last observation).

In the previous analysis we have assumed that the stabilization parameter τ is given by (5). We are going to show that in fact the lack of optimality holds *for any* τ. To this end, let us analyze which are the conditions for stability and optimal accuracy when, up to constants, τ is given by

$$\begin{aligned}\tau^{-1} &= \frac{k}{h^2}p^2\varphi(\mathrm{Pe}, p) + \frac{|a|}{h}p \\ &= \frac{k}{h^2}p^2\left[\varphi(\mathrm{Pe}, p) + \frac{\mathrm{Pe}}{p}\right] =: \frac{k}{h^2}p^2\psi(\mathrm{Pe}, p)\end{aligned}$$

where $\varphi(\mathrm{Pe}, p) \gtrsim 1$ is a function to be determined (bounded from below by a constant).

Repeating the analysis carried out before, it is easily seen that the stability condition is

$$\begin{aligned}&1 - \frac{kp^4}{h^2}\tau \gtrsim 1, \quad \frac{kp^4}{h^2} \lesssim \tau^{-1} = \frac{k}{h^2}p^2\varphi(\mathrm{Pe}, p) + \frac{|a|}{h}p \\ &p^2 \lesssim \varphi(\mathrm{Pe}, p) + \frac{\mathrm{Pe}}{p} = \psi(\mathrm{Pe}, p)\end{aligned}$$

whereas the optimal convergence condition is

$$\begin{aligned}E_0^2 &\lesssim k + \tau k^2\frac{p^2}{h^2} + \tau|a|^2 + \min\left(\tau^{-1}, \frac{|a|^2}{k}\right)\frac{h^2}{p^2} \\ &\lesssim k + k\left(\varphi + \frac{\mathrm{Pe}}{p}\right)^{-1} + k\frac{\mathrm{Pe}}{p}\left(1 + \frac{p}{\mathrm{Pe}}\varphi\right)^{-1} + k\min\left(\varphi + \frac{\mathrm{Pe}}{p}, \frac{\mathrm{Pe}^2}{p^2}\right) \\ &\lesssim k + k\frac{\mathrm{Pe}}{p} + k\min\left(\varphi + \frac{\mathrm{Pe}}{p}, \frac{\mathrm{Pe}^2}{p^2}\right)\end{aligned}$$

The condition for optimal accuracy is

$$\min\left(\varphi + \frac{\mathrm{Pe}}{p}, \frac{\mathrm{Pe}^2}{p^2}\right) = \min\left(\psi, \frac{\mathrm{Pe}^2}{p^2}\right) \lesssim \max\left(1, \frac{\mathrm{Pe}}{p}\right)$$

It turns out that *the conditions for stability* $p^2 \lesssim \psi$ *and optimal accuracy* $\min\left(\psi, \frac{\mathrm{Pe}^2}{p^2}\right) \lesssim \max\left(1, \frac{\mathrm{Pe}}{p}\right)$ *are impossible to achieve in the range*

$$p \leq \mathrm{Pe} \leq p^3$$

Therefore, the error estimates presented for classical residual-based stabilized methods cannot be hp-optimal in the whole range of Pe, *regardless of the expression of the stabilization parameter*. Nevertheless, the range of lack of optimality is not observed in practice; it is open whether better a priori estimates could be obtained without relying on inverse estimates.

4 Non-residual based stabilization

The objective of this section is to show that the lack of (theoretical) optimality described before is not restricted to residual-based methods, but it also appears in term-by-term stabilized FEMs. We will motivate and analyze one method of this type, but the conclusion to be drawn carries over to other term-by-term stabilized formulations such as local projection stabilization (LPS).

To start the motivation of the method to be considered, let us obtain an improved stability estimate for the original Galerkin method (2), sharper than just (3). Let $P_{h,0}$ the projection onto the finite element space V_h, which incorporates boundary conditions. If $v_{h,0} = \tau P_{h,0}(a \cdot \nabla u_h)$, we have that

$$\begin{aligned} B(u_h, v_{h,0}) \gtrsim{}& \tau \| P_{h,0}(a \cdot \nabla u_h)\|^2 \\ & - k\|\nabla u_h\| \frac{C_{\mathrm{inv}}}{h} p^2 \tau \| P_{h,0}(a \cdot \nabla u_h)\| \\ & - s\|u_h\| \tau \| P_{h,0}(a \cdot \nabla u_h)\| \end{aligned}$$

If $\tau \leq \min\left(\frac{h^2}{C_{\mathrm{inv}}^2 p^2 k}, \frac{1}{s}\right)$ then

$$B(u_h, v_{h,0}) \gtrsim \tau \| P_{h,0}(a \cdot \nabla u_h)\|^2 - k\|\nabla u_h\|^2 - s\|u_h\|^2$$

The last term can be controlled by the terms bounded by the Galerkin method taking $v_h = u_h$, and therefore *only control on* $\tau \| P_{h,0}^{\perp}(a \cdot \nabla u_h)\|^2$ *is missing*. This suggests that the subscales can be chosen as orthogonal to the finite element space, a possibility that we consider later.

Nevertheless, we can obtain a stability condition that is independent of the space of subscales, and therefore inherent to the two-scale splitting. Regardless of the expression of $\tilde{u}$, we have:

$$\begin{aligned} & B(u_h, u_h) + (\tilde{u}, \mathcal{L}^* u_h)_h + (\mathcal{L} u_h, \tilde{u})_h + \tau^{-1}(\tilde{u}, \tilde{u}) \\ & \quad = k\|\nabla u_h\|^2 + s\|u_h\|^2 + 2(\tilde{u}, -k\Delta u_h + s u_h)_h + \tau^{-1}\|\tilde{u}\|^2 \\ & \quad \gtrsim k\|\nabla u_h\|^2 + s\|u_h\|^2 - 2\left(\tau \frac{k}{h^2} C_{\mathrm{inv}}^2 p^4\right) k\|\nabla u_h\|^2 \\ & \qquad - 2(\tau s)s\|u_h\|^2 - \frac{\tau^{-1}}{2}\|\tilde{u}\|^2 + \tau^{-1}\|\tilde{u}\|^2 \\ & \quad \gtrsim k\|\nabla u_h\|^2 + s\|u_h\|^2 + \tau^{-1}\|\tilde{u}\|^2 \end{aligned} \tag{12}$$

Therefore, *the stability condition* $\tau \frac{k}{h^2} C_{\text{inv}}^2 p^4 \lesssim 1$ *is always needed for the type of stability estimates we wish to obtain*.

Remember that we have shown that residual-based stabilized FEMs have a range of Pe for which there is no optimality in p in the error estimates obtained, and this is independent of the stabilization parameter τ. Thus, we will keep expression (5) for this parameter from now onwards.

Let us move now to two particular cases for the choice of the subscale space. Let us start noting that the general equation for the subscales can be written as:

$$(\mathcal{L}u_h, \tilde{v})_h + \tau^{-1}(\tilde{u}, \tilde{v}) = \langle f, \tilde{v} \rangle \quad \Rightarrow \quad \tilde{u} = \tau \tilde{P}(f - \mathcal{L}u_h)$$

where $\tilde{P}$ stands for the L^2 projection onto $\tilde{V}$. The first choice for $\tilde{P}$, and therefore for $\tilde{V}$, is to consider that $\tilde{P}$ is the identity when applied to element-wise finite element residuals, which results in (8). The danger of this approach is that $V_{hp} \cap \tilde{V} = \{0\}$ may not hold, and therefore the direct sum property of V_{hp} and $\tilde{V}$ may be lost.

The second choice consists of taking the space of subscales as L^2-orthogonal to the finite element space, that is to say

$$\tilde{V} = V_{hp}^{\perp} \iff \tilde{u} = \tau P_{h,0}^{\perp}(f - \mathcal{L}u_h) \tag{13}$$

This leads to the orthogonal subscale stabilization method introduced in [2]. The discrete variational equation to be solved is

$$\begin{aligned} &B(u_h, v_h) + \tau(P_{h,0}^{\perp}(\mathcal{L}u_h), -\mathcal{L}^* v_h)_h \\ &\quad = \langle f, v_h \rangle + \tau(P_{h,0}^{\perp}(f), -\mathcal{L}^* v_h)_h \quad \forall v_h \in V_{hp} \end{aligned} \tag{14}$$

and we immediately obtain the stability estimate

$$\begin{aligned} &B(u_h, u_h) + \tau(P_{h,0}^{\perp}(\mathcal{L}u_h), P_{h,0}^{\perp}(-\mathcal{L}^* u_h))_h \\ &\quad \gtrsim k\|\nabla u_h\|^2 + s\|u_h\|^2 + \tau\|P_{h,0}^{\perp}(a \cdot \nabla u_h)\|^2 \end{aligned}$$

As we have seen, the term $\tau\|P_{h,0}^{\perp}(a \cdot \nabla u_h)\|^2$ is precisely what the Galerkin method lacks. Therefore, it is easy to understand that this method is going to be stable and convergent in the same norm as the classical residual based stabilization analyzed in the previous section.

However, rather than analyzing the method obtained with orthogonal subscales, let us introduce a simplified version that shares several features in common with LPS methods, and in particular the stability and convergence requirements for p-refinement.

Let P_h be the projection onto the finite element space without boundary conditions. For any function g with appropriate regularity, $\|P_h^{\perp}(g)\|$ will converge to zero as $h \to 0$ at the same rate as the interpolation error. This suggests to neglect several terms in (14) and keep only the one that contributes to stability. We may thus consider the *non-residual* based method which consists of finding $u_h \in V_{hp}$ such that

$$B(u_h, v_h) + \tau(P_h^{\perp}(a \cdot \nabla u_h), P_h^{\perp}(a \cdot \nabla v_h)) = \langle f, v_h \rangle \quad \forall v_h \in V_{hp} \tag{15}$$

When replacing u_h by the solution of the continuous problem u, it is seen that this method is not consistent, but the consistency error arising from $P_h^{\perp}(a \cdot \nabla u)$ is of optimal order with respect to h and p. Obviously, the projection P_h cannot be replaced by $P_{h,0}$ if the consistency error has to be optimal. It is also seen that the structure of the method resembles that of LPS methods, simply replacing $P_h^{\perp}$ by a so-called fluctuation operator, which is defined as the identity minus a certain (local) projection . This method can be considered a term-by-term stabilization method, since only what lacks stability is added to the terms of the Galerkin approximation. In the problem we consider, only the convective term causes instabilities.

Stability and convergence happen to hold in the norm

$$k\|\nabla v_h\|^2 + s\|v_h\|^2 + \tau\|P_h^{\perp}(a\cdot\nabla v_h)\|^2$$

In this norm, it can be easily proved that the method is stable and optimally convergent for all ranges of Pe and for all p. However, the analysis in this norm (and also in the 'natural' norm of LPS methods) is incomplete, since there is no control on part of the convective term, namely, that belonging to the finite element space (without boundary conditions).

Let us sketch how to perform the numerical analysis of (15) in a finer norm. Let us consider:

$$I = P_{h,0} + P_{h,\partial} + P_h^{\perp}$$

where $P_{h,0}$ is the projection onto V_{hp} already introduced, P_h the projection onto the finite element space $\hat{V}_{hp}$ constructed before imposing the boundary conditions, and $P_{h,\partial}$ the projection onto the boundary subspace that completes V_{hp} in $\hat{V}_{hp}$. All these are L^2 projections.

Stability of the convective term can be obtained as follows:

- Control over $\tau\|P_h^{\perp}(a\cdot\nabla u_h)\|^2$ is provided by the stabilization terms in (15), as we have seen before.
- The term $\tau\|P_{h,0}(a\cdot\nabla u_h)\|^2$ can be controlled by the Galerkin method (improved stability analyzed before).
- The term $\tau\|P_{h,\partial}(a\cdot\nabla u_h)\|^2$ can be controlled in terms of the other two if the finite element mesh satisfies a (very weak) regularity assumption, which was discussed first in [5] (see also [3]).

Following these steps it is possible to prove stability in the same norm as for residual based methods, namely,

$$|||v_h|||^2 = k\|\nabla v_h\|^2 + s\|v_h\|^2 + \tau\|a\cdot\nabla v_h\|^2 \tag{16}$$

This stability can be proved in the form of an inf-sup condition.

At this point one could wonder whether this simplified stabilization method is optimal in p for all Pe, and the answer is *no, it is not*. Let us explain why without detailing all the steps of the analysis.

Concerning the stability analysis, as it has been explained before, control over $\tau\|P_{h,0}(a\cdot\nabla u_h)\|^2$ is obtained from the Galerkin terms when the test function is taken as $v_h = \tau P_{h,0}(a\cdot\nabla u_h)$. The convective term provides stability, but the other terms (the diffusive, reactive and stabilization terms) need to controlled by using inverse estimates, *and therefore we are back to the requirement* (11). Convergence analysis follows exactly the same steps as for residual-based stabilization, and it is found that the optimality conditions are also exactly the same as for residual-based methods. Therefore, *the (theoretical) problem of lack of optimal convergence in p for the range of Péclet numbers $p \le \mathrm{Pe} \le p^3$ also holds for this non-residual method.*

Similar arguments could be applied to LPS methods.

5 A spectral method for Burguers' equation and H^1-orthogonality of the subscales

In this section we start by a detour from finite element methods and describe the formulation presented in [1], where the properties of a VMS spectral method for the one-dimensional

Burguers equation are analyzed. This motivates the introduction of a subscale space in finite elements that is both L^2- and H^1-orthogonal to the finite element space. The practical accuracy and numerical performance of the resulting stabilized FEM have not been tested, but the numerical analysis reveals that the inverse estimate (10) is not needed, and that therefore one can prove optimal convergence in p for all Péclet numbers for a proper choice of the stabilization parameter.

Let us consider the problem of finding a scalar function $u(x, t)$ of position x and time t such that

$$\partial_t u + u\partial_x u - \nu\partial_{xx} u = f, \quad x \in \Omega = (0, 2\pi), \quad t > 0 \tag{17}$$

with an initial condition $u(x, 0) = u^0(x)$ in Ω and periodic boundary conditions at $x = 0$ and $x = 2\pi$, $t \geq 0$. The differential equation (17) can be written in the form

$$\partial_t u + \mathcal{L}(u; u) = f, \qquad x \in (0, 2\pi), \quad t > 0$$

by introducing the bilinear operator $\mathcal{L}(u; v) = u\partial_x v - \nu\partial_{xx} v$. The reason for dealing with a transient and nonlinear problem instead of a linear one as we have done before will be clear later.

To write the variational form of the problem, let $W = H^1_{\text{per}}(\Omega)$ be the subspace of periodic functions in $H^1(\Omega)$. Let also

$$A(u; w, v) := -\frac{1}{2}(u\partial_x w, v) + \nu(\partial_x w, \partial_x v)$$

Then, the variational form of the problem can be written as: find $u : \mathbb{R}^+ \longrightarrow W$ such that

$$\begin{aligned} (w, \partial_t u) + A(u; w, u) &= (w, f), \quad t > 0 \\ (w, u) &= (w, u^0), t = 0 \end{aligned}$$

for all $w \in W$.

We will proceed to design now a VMS method based on a spectral approximation of the problem. Let W_N be a subspace of dimension N of W. The Galerkin approximation is: find $u_N : \mathbb{R}^+ \longrightarrow W_N$ such that

$$\begin{aligned} (w_N, \partial_t u_N) + A(u_N; w_N, u_N) &= (w_N, f), \quad t > 0 \\ (w_N, u_N) &= (w_N, u^0), \quad t = 0 \end{aligned}$$

for all $w_N \in W_N$. Let now $W = W_N \oplus \tilde{W}$, where $\tilde{W}$ is in principle any space to complete W_N in W. The original continuous problem can be equivalently written as

$$(w_N, \partial_t u) + A(u; w_N, u) = (w_N, f) \quad \forall w_N \in W_N, \quad t > 0 \tag{18}$$

$$(\tilde{w}, \partial_t u) + A(u; \tilde{w}, u) = (\tilde{w}, f) \quad \forall \tilde{w} \in \tilde{W}, \quad t > 0 \tag{19}$$

Initial conditions have to be added to these equations.

Before defining the way subspace W_N is constructed, we can design the same formulations as for the convection-diffusion equation and, in particular, the classical residual-based method and the method based on orthogonal subscales, given respectively for the convection-diffusion problem by (9) and (14). If W_N is made of smooth functions, Eq. (19) can be written as

$$(\tilde{w}, \partial_t \tilde{u}) + (\tilde{w}, \mathcal{L}(u; \tilde{u})) = (\tilde{w}, f - \partial_t u_N - \mathcal{L}(u; u_N))$$

We now proceed to make approximations to render the problem numerically solvable. First, we approximate:

$$\mathcal{L}(u; \tilde{u}) \approx \tau^{-1}(u)\tilde{u}$$

where τ is defined below. Because of the nonlinearity it may depend on the unknown u. The residual of the resolved scale is

$$R_N(u; u_N) = f - \partial_t u_N - \mathcal{L}(u; u_N)$$

For this nonlinear problem we may further approximate:

$$\tau(u) \approx \tau(u_N), \quad R_N(u; u_N) \approx R_N(u_N; u_N)$$

The equation for the subscales is then

$$\left(\tilde{w}, \partial_t \tilde{u} + \tau^{-1}(u_N)\tilde{u}\right) = (\tilde{w}, R_N(u_N; u_N)) \quad \forall \tilde{w} \in \tilde{W}, \ \ t > 0$$

Now we have to choose the space for the subscales. If the choice is to take it as the space of residuals of functions in W_N the result is

$$\partial_t \tilde{u} + \tau^{-1}(u_N)\tilde{u} = R_N(u_N; u_N), \ \ t > 0 \tag{20}$$

whereas if subscales are taken as L^2-orthogonal to W_N we obtain

$$\partial_t \tilde{u} + \tau^{-1}(u_N)\tilde{u} = P_N^{\perp}[R_N(u_N; u_N)], \ \ t > 0 \tag{21}$$

where P_N is now the L^2-projection onto W_N.

Let us consider now a particular way to construct space W_N, namely, a spectral method with standard Fourier basis. The approximation space is now

$$W_N := \text{span}\{e^{ikx} : k \in [-N/2, N/2 - 1] \subset \mathbb{N}\}$$

and the approximation to the unknown u can be written as

$$u_N(x, t) = \sum_{k=-N/2}^{N/2-1} \hat{u}_k(t) e^{ikx}$$

with the amplitudes $\hat{u}_k(t)$ to be determined. Now k is the wavenumber (not to be confused with the diffusion of the previous sections).

Consider first problem (20). We may assume that f belongs to W_N for each t, since this does not decrease the order of accuracy. Therefore, in the residual

$$R_N(u_N; u_N) = f - \partial_t u_N - \nu \partial_{xx} u_N - u_N \partial_x u_N$$

all terms but the last one belong to W_N, and the last one belongs to W_{2N}. Note that if we replace the convective term $u_N \partial_x u_N$ by $a\partial_x u_N$, with a constant, the last term would also belong to W_N, and therefore the subscale $\tilde{u}$ would belong to W_N, in contradiction with the assumption that the intersection of the approximation space and the space of subscales is the trivial element. Classical residual-based methods for this problem do not make any sense. This can be understood noting that when a is constant the Galerkin method does not produce oscillations. Indeed, we may take as test function $v_N = a\partial_x u_N \in W_N$ and obtain control over the convective term.

Consider now problem (21). We have that $P_N^{\perp}(R_N) = -P_N^{\perp}(u_N \partial_x u_N)$ belongs to the complement of W_N in W_{2N}, which is now the space of subscales:

$$\tilde{W} := \text{span}\{\mathrm{e}^{\mathrm{i}rx} : r \in [-N, N-2] \setminus [-N/2, N/2-1] \subset \mathbb{N}\}$$

Recalling the orthogonality relations

$$\left(\mathrm{e}^{\mathrm{i}kx}, \mathrm{e}^{\mathrm{i}lx}\right) = 2\pi\delta_{kl}, \quad \left(\partial_x \mathrm{e}^{\mathrm{i}kx}, \partial_x \mathrm{e}^{\mathrm{i}lx}\right) = -2\pi k^2 \delta_{kl}$$

we have that in this case W_N and $\tilde{W}$ *are not only L^2-orthogonal, but also H^1-orthogonal*. This is an important property that we will exploit below.

In this particularly simple problem we can obtain the final expression of the equations to be solved for the amplitudes $\hat{u}_k(t)$. First, note that for the transient and diffusive terms we have

$$\left(\mathrm{e}^{\mathrm{i}kx}, \partial_t \sum_l \hat{u}_l \mathrm{e}^{\mathrm{i}lx}\right) = 2\pi\,\partial_t \hat{u}_k, \quad \left(\nu\partial_x \mathrm{e}^{\mathrm{i}kx}, \partial_x \sum_l \hat{u}_l \mathrm{e}^{\mathrm{i}lx}\right) = -2\pi\nu k^2 \hat{u}_k,$$

The nonlinear Galerkin term results in

$$\left(\frac{1}{2}\sum_q \hat{u}_q \mathrm{e}^{\mathrm{i}qx}\partial_x \mathrm{e}^{\mathrm{i}kx}, \sum_l \hat{u}_l \mathrm{e}^{\mathrm{i}lx}\right) = -\pi \mathrm{i}k \sum_{q+l=k} \hat{u}_q \hat{u}_l$$

The variational form of the equation for the subscales is

$$\left(\tilde{w}, \partial_t \tilde{u} + \tau^{-1}(u_N)\tilde{u}\right) = (\tilde{w}, -u_N\partial_x u_N + P_N(u_N\partial_x u_N))$$

The term $u_N\partial_x u_N$ can be developed as

$$u_N\partial_x u_N = \sum_q \mathrm{e}^{\mathrm{i}qx}\hat{u}_q \sum_l \mathrm{i}l\mathrm{e}^{\mathrm{i}lx}\hat{u}_l = \sum_{q,l} \mathrm{i}l\mathrm{e}^{\mathrm{i}(q+l)x}\hat{u}_q\hat{u}_l$$

from where we get that the equations for the amplitudes of the subscales are

$$\partial_t \hat{\tilde{u}}_r + \tau^{-1}(u_N)\hat{\tilde{u}}_r = \sum_{q+l=r} \mathrm{i}l\hat{u}_q\hat{u}_l \tag{22}$$

To close the approximation we need to give an expression for τ, which we take as

$$\tau(u_N) = \left[3\pi\nu^2\left(\frac{4}{h^2}\right)^2 + \frac{4}{h^2}\|u_N\|^2\right]^{-\frac{1}{2}}$$

where $h = \pi/N$ (see [1] for motivation and details).

The final expression for the amplitudes of the resolvable scales of the Burguers equation is

$$\partial_t \hat{u}_k + \frac{\mathrm{i}k}{2}\sum_{q+l=k} \hat{u}_q\hat{u}_l + \mathrm{i}k\sum_{q+r=k} \hat{u}_q\hat{\tilde{u}}_r + \frac{\mathrm{i}k}{2}\sum_{r+p=k} \hat{\tilde{u}}_r\hat{\tilde{u}}_p + \nu k^2\hat{u}_k = f_k$$

where the subscales are computed from Eq. (22).

Using the fact that we have allowed the subscales to be time dependent and using also the L^2- and H^1-orthogonality of the resolvable scales (those in W_N) and the subscales, one can prove an energy estimate in the form

$$\|u_N(T)\|^2 + \nu\int_0^T \|\partial_x u_N\|^2 \mathrm{d}t + \tau\int_0^T \|u_N\partial_x u_N\|^2 \mathrm{d}t \lesssim \mathcal{N}$$

where $\mathcal{N}$ depends on the forcing term f and the initial conditions.

However, our objective is not to analyze in detail this spectral approximation to Burgers' problem, but to use it to motivate a stabilized finite element in which the subscales are both L^2- and H^1-orthogonal to the finite element space, in a sense that needs to be made precise.

Let us start introducing the space

$$L_{hp} := \{\phi_h : \Omega \longrightarrow \mathbb{R} \mid \phi_h|_K = \Delta v_h|_K,\ K \in \mathcal{T}_h,\ v_h \in V_{hp}\}$$

that is to say, the space of functions that are Laplacians of finite element functions inside the element domains. Note that for linear elements ($p = 1$) we would have $L_{h1} = \{0\}$. The method we now propose consists in taking

$$\tilde{V} = V_{hp}^{\perp} \cap L_{hp}^{\perp} \tag{23}$$

instead of (13), again considering orthogonality in the sense of L^2.

The reason for choice (23) becomes clear if we take a look at derivation (12), which is valid for any space of subscales. If $\tilde{u}$ belongs to the space given by (23) we have that

$$(\tilde{u}, \Delta u_h)_h = 0, \quad (\tilde{u}, u_h)_h = 0 \tag{24}$$

and therefore *condition* (11) *is not required to obtain the stability result* (12). The important consequence of this is that we may choose the stabilization parameter in (5) with $\alpha_1(p)$ of order p^2 and $\alpha_2(p)$ of order p, which are the choices that allow us to obtain *optimal stability and error estimates in p for all Péclet numbers in the norm* (16). The proof of this fact follows the same lines as the analysis sketched before, and we will not detail it here, but let us explain the key points of it. First, note that (12) in the case we consider can be written as

$$\begin{aligned} B(u_h, u_h) &+ (\tilde{u}, \mathcal{L}^* u_h)_h + (\mathcal{L} u_h, \tilde{u})_h + \tau^{-1}(\tilde{u}, \tilde{u}) \\ &\gtrsim k\|\nabla u_h\|^2 + s\|u_h\|^2 + \tau\|P_h^{\perp} P_L^{\perp}(a \cdot \nabla u_h)\|^2 \end{aligned}$$

where $P_L^{\perp} = I - P_L$ and P_L is the L^2 projection onto L_h. As we have shown before, control on $\tau\|P_{h,0}(a \cdot \nabla u_h)\|^2$ can be obtained from the Galerkin terms and on $\tau\|P_{h,\partial}(a \cdot \nabla u_h)\|^2$ by using a mild condition on the finite element mesh. To have full control over the norm of the convective term it remains to bound $\tau\|P_L(a \cdot \nabla u_h)\|^2$. This can be achieved by proving that for any piecewise polynomial function g_h there holds that $\|P_L(g_h)\| \lesssim \|P_h(g_h)\|$, again under mild assumptions over the finite element mesh. Altogether, this allows one to prove stability in the norm (16) in the form of an inf-sup condition. Optimal convergence is then easy to fulfill as soon as we may take $\alpha_1(p)$ of order p^2 in the stabilization parameter.

Let us comment on the sense of H^1-orthogonality employed here. For the Burgers problem that has motivated the formulation we are describing, this orthogonality is exact. Now strictly speaking the first condition in (24) can be considered as orthogonality in H^1 (assuming the second condition holds) if $\tilde{u}$ vanishes on the element boundaries, i.e., it is a bubble function (which is obviously not the case with the approximations performed). If this were the case, we would have

$$(\tilde{u}, \Delta u_h)_h = -\sum_K (\nabla\tilde{u}, \nabla u_h)_K = 0$$

Finally, let us discuss how projection P_L can be computed for any piecewise polynomial function g_h. The first point is to construct a basis for L_{hp}. Let this basis be $\{\psi_1, \ldots, \psi_m\}$, with $m = \dim(L_{hp}) \le M = \dim(V_{hp})$. If the shape functions of V_{hp} are $\{\eta_1, \ldots, \eta_M\}$, a possible algorithmic way to compute $\{\psi_1, \ldots, \psi_m\}$ is through a singular value decomposition (SVD) of $\{\Delta\eta_1, \ldots, \Delta\eta_M\}$. Once the basis is constructed, we may express

$$P_L(g_h) = \sum_{i=1}^{m} \hat{g}_i \psi_i$$

and compute the coefficients $\hat{g}_i$ by solving

$$\sum_{i=1}^{m} (\psi_j, \psi_i)\hat{g}_i = (\psi_j, g_h), \quad j = 1, \ldots, m$$

6 Conclusions

In this paper we have revisited some stabilized FEMs for the convection-diffusion equation, both classical residual-based and some non-residual-based methods based on projections. The main conclusions of the analysis presented are the following:

- The error estimates obtained for standard residual-based stabilized finite element methods *are not optimal* in the range $p \leq \mathrm{Pe} \leq p^3$ (although this does not seem to have practical consequences).
- Non-residual-based methods based on L^2 projections do not solve the problem. They require an inverse estimate for full stability. The same can be said about LPS techniques.
- Subscales *both* L^2*-and* H^1*- orthogonal* to the resolvable scales lead to optimal finite element solutions for all Pe and p. A method has been proposed in this line motivated by the analysis of the 1D Burguers' equation using a spectral method. The reasons that make this method optimal have been explained.

Acknowledgements The author would like to acknowledge the support received from the ICREA Acadèmia Program, from the Catalan Government.

References

1. Bayona, C., Baiges, J., Codina, R.: Variational multi-scale approximation of the one-dimensional forced Burguers' equation: the role of orthogonal sub-grid scales in turbulence modelling. Int. J. Numer. Methods Fluids **86**, 313–328 (2018)
2. Codina, R.: Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods. Comp. Methods Appl. Mech. Eng. **190**, 1579–1599 (2000)
3. Codina, R.: Analysis of a stabilized finite element approximation of the Oseen equations using orthogonal subscales. Appl. Numer. Math. **58**, 264–283 (2008)
4. Codina, R., Badia, S., Baiges, J., Principe, J.: Variational Multiscale Methods in Computational Fluid Dynamics. In: Stein, E., Borst, R., Hughes, T.J. (eds.) Encyclopedia of Computational Mechanics, 2nd edn. (2017). https://doi.org/10.1002/9781119176817.ecm2117
5. Codina, R., Blasco, J.: A finite element formulation for the Stokes problem allowing equal velocity-pressure interpolation. Comput. Methods Appl. Mech. Eng. **143**, 373–391 (1997)
6. Codina, R., Principe, J., Baiges, J.: Subscales on the element boundaries in the variational two-scale finite element method. Comput. Methods Appl. Mech. Eng. **198**, 838–852 (2009)
7. Houston, P., Süli, E.: Stabilised *hp*-finite element approximation of partial differential equations with nonnegative characteristic form. Computing **66**, 99–119 (2001)
8. Hughes, T., Feijóo, G., Mazzei, L., Quincy, J.: The variational multiscale method—a paradigm for computational mechanics. Comput. Methods Appl. Mech. Eng. **166**, 3–24 (1998)
9. Lube, G., Rapin, G.: Residual-based stabilized higher-order FEM for advection-dominated problems. Comput. Methods Appl. Mech. Eng. **195**, 4124–4138 (2006)

10. Matthies, G., Skrzypacz, P., Tobiska, L.: A unified convergence analysis for local projection stabilisations applied to the Oseen problem. ESAIM Math Model. Numer. Anal. **41**, 713–742 (2007)
11. Schwab, C.: *p*- and *hp*-Finite Element Methods. Theory and Application to Solid and Fluid Mechanics. Oxford University Press, Oxford (1998)