RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Frères Mentouri Constantine 1
Faculté des Sciences Exactes
Département de Mathématiques

N^o d'ordre : .....
Série : .....

THÈSE

Présentée pour l'obtention du diplôme de
Doctorat en Sciences

Option
Analyse numérique

Stabilized numerical schemes for iterative solution of the generalized Stokes problem

---

Par
CHIBANI Alima

Devant le jury :

Président :		
	DEGHDAK Messaoud	Prof. Université Frères Mentouri, Constantine1
Rapporteur :		
	KECHKAR Nasserdine	Prof. Université Frères Mentouri, Constantine1
Examinateurs :		
	ABADA Nadjet	M. C. A. École Normale Supérieure Assia Djebbar, Constantine
	DIAR Ahmed	M. C. A. Université Larbi Ben M'hidi, Oum El Bouaghi

Soutenue le : 30 / 11 / 2020

To the memories of
my father, my mother and my sister Zouleikha.

Acknowledgements

Al-Hamdou Li ALLAH.

I would like to adress my deep gratitude and my sincere thanks to my supervisor Prof. Nasserdine KECHKAR for his guidances and encouragements during the running of this thesis. I am also grateful for his help and efforts along the way. I would also like to thank the rest of my dissertation committee members; Prof. Messaoud DEGHDAK, Dr. Nadjet ABADA and Dr. Ahmad DIAR for their valuable comments and discussion. I would like especially to express my deep gratitude to Dr. N. ABADA for attending the commitee online despite her desease at that time.

I am most grateful to Prof. Andrew John WATHEN and Dr. Kamal BENTAHAR who provided me an opportunity to attend the workshop which was funded by the *King AbdULLAH University of Science and Technology (KAUST)* to the *Oxford Center for Collaborative Applied Mathematics, University of Oxford (OCCAM).*
Although it has been years since I have joined their team as trainee, I still take their diligence, continuous persiverance and encouragement lessons with me. I deeply thank them for their kindness, generosity and all facilities I have recieved.
Certainly, similar thanks extends to OCCAM−KAUST and to Prof. Mohamed DENCHE and Prof. Salah DJEZZAR for their recommandation letters that were a precious support. I would also like to express my heartfelt thanks to Prof. David J. SILVESTER for his valuable guidances and encouragements.

Likewise, I am extremely thankful to all my teachers for their efforts, encoragements and kindness throughout my whole studies.

Also, I profoundly appreciate the sumpathy and assistance I have received from my colleagues, labmates and the administrative staff. I would particularly like to deeply thank Prof. Samir BENHADID, Dr. Bassem MEKNANI and Dr. Faiza SAADI for their valuable services.

Additionally, I would like to deeply thank my nephews Azeddine AGOUNE and Youcef NAILI for their help to carry out this work.

Similarly, my sincere thanks go to all my family and friends for their encouragements and aid. Among them, I would like to mention Miss Leyla BENTOUIL, Dr. Ilham LAROUSSI, Dr. Nadia LEKRINE, Dr. Souheila LOUAR, Dr. Fatiha MESLOUB and Dr. AbdERRAHMANE ZARAI as well as my cusin Hicham CHIBANI and my nefews Mohamed As-Saleh NAILI and Salaheddine CHIBANI.

Abstract

In this thesis, some novel discrete formulations for stabilizing the mixed finite element method *Q1-Q0* (bilinear velocity and constant pressure approximations) are introduced and discussed for the generalized Stokes problem. These are based on stabilizing discontinuous pressure approximations via local jump operators. The developing idea consists in a reduction of terms in the local jump formulation, introduced earlier, in such a way that stability and convergence properties are preserved.
Moreover, some iterative methods of conjugate gradient type are discussed and their algorithms are presented. These are used for the iterative solution of the algebraic systems of linear equations which arise from the spatial discretization of the continuous problem. The computer implementation aspects and numerical evaluation of the stabilized discrete formulations are also considered. For illustrating the numerical performance of the proposed approaches and comparing the three versions of the local jump methods against the global jump setting, some obtained results for two test generalized Stokes problems are presented. Numerical tests confirm the stability and accuracy characteristics of the resulting approximations. Likewise, the numerical reliability of the discussed iterative solvers is assessed.

Keywords : Stokes problem, Finite elements, Stabilization, Iterative methods

Résumé

Dans cette thèse, quelques nouvelles formulations discrètes pour stabiliser la méthode des éléments finis mixtes *Q1-Q0* (approximations de vitesse bilinéaire et pression constante) sont introduites et discutées pour le problème généralisé de Stokes. Elles sont basées sur la stabilisation des approximations de pression discontinue à travers les opérateurs du saut local. L'idée principale est de réduire les tèrmes dans la formulation du saut local, introduits avant, tels que la stabilité et les propriétés de convergence soient conservées.
De plus, quelques méthodes de type gradient conjugué sont discutées et leurs algorithmes sont présentés. Celles-ci sont utilisées pour la résolution itérative des systèmes d'équations linéaires obtenues de la discrétisation spaciale du problème continu. La mise en oeuvre des calculs sur l'ordinateur et l'évaluation numérique des formulations discrètes stabilisées sont aussi considérées. Pour illustrer la performance numérique des approximations proposées et comparer les trois versions des méthodes du saut local avec celle du saut global, quelques résultats obtenus pour deux problêmes tests généralisés de Stokes sont présentés. Les tests numériques confirment la stabilité et l'exactitude des approximations résultantes. De même, la fiabilité numérique des méthodes itératives discutées est éstimée.

Mots clés: Problème de Stokes, Eléments finis, Stabilisation, Méthodes itératives

ملخص

في هذه الأطروحة، تم تقديم ومناقشة بعض الصيغ المنفصلة الجديدة من أجل استقرار طريقة العناصر المنتهية المختلطة (السرعة ثنائية الخطية و الضغط الثابت) لمشكلة ستوكس المعممة. وهي تستند إلى استقرار تقريب الضغط المتقطع من خلال القفز المحلي. ترتكز الفكرة الرئيسية على تخفيض الحدود في الصياغة المنفصلة للقفز المحلي المقدمة سابقا، بحيث يتم الحفاظ على الاستقرار وخصائص التقارب.
بالإضافة إلى ذلك، تمت مناقشة بعض الطرق التكرارية من نمط التدريج المرافق وعرض الخوارزميات الخاصة بها و ذلك من أجل الحل التكراري لجمل المعادلات الخطية التي تم الحصول عليها. كما تم تنفيذ الحسابات على الكمبيوتر و التقييم العددي للتركيبات المنفصلة المستقرة. لتوضيح الأداء العددي للتقريبات المقترحة و مقارنة الصيغ الثلاثة للقفز المحلي مع القفز العام، تم تقديم بعض النتائج المحصل عليها. تؤكد الاختبارات العددية التي تم تنفيذها على مشكلتي اختبار لستوكس ثبات و دقة التقريبات المحصل عليها كما تم تقدير كفاءة الطرق التكرارية لحل جمل المعادلات الخطية التي تمت مناقشتها.

الكلمات المفتاحية مشكل ستوكس، العنلصر المنتهية، الاستقرار، الطرق التكرارية

Contents

List of Figures

List of Tables

Introduction

The mixed finite element methods are widely used for the numerical solution of incompressible flow problems. Many of them involve the use of approximations for the unknown primitive variables (velocity and pressure) in the Galerkin methodology. However, it is widely known that the discrete velocity and pressure spaces cannot be chosen independently of each other. There is a compatibility condition, commonly called the Babuška-Brezzi stability condition, that needs to be satisfied if the resulting mixed approximation is to be effective.
For simplicity, low-order mixed approximation methods are preferred. Nevertheless, many of them are unstable in the standard Babuška-Brezzi sense. Among these, the mixed methods referred to as *Q1-Q0, P1-P0, P1-P1* and *Q1-Q1* are notorious. Hence, Boland and Nicolaïdes [4] have shown that the mixed finite element *Q1-Q0* does not satisfy the stability condition. In addition, Sani et al. [28] and [29] have demonstrated that for certain boundary conditions the method generates spurious pressure modes (called *checkerboard modes*) resulting in numerical instabilities in the approximate pressure.

As a result of that, several researchers have been interested in overcoming the need of satisfying the Babuška-Brezzi stability condition. The idea of such called stabilization was initially proposed in the pioneering work of Brezzi and Pitkäranta [7]. Later, Hughes and Franca [21] constructed a Stokes discrete formulation which ensures convergence of discrete solutions for any mixed approximation. For a discontinuous pressure approximation, the called universal stability can be achieved by the introduction of pressure jump terms into the standard Galerkin discrete formulation. However, for achieving the universal stability, these jump terms must control pressure jumps across all internal inter-element edges. In the early 1990s, Silvester and Kechkar [31] suggested that a more robust way of stabilizing a mixed method based on discontinuous pressure consists in restricting the global jump operator of Hughes and Franca locally to a macro-element partitioning of the solution domain. Furthermore, Kechkar and Silvester in [23] showed that the local jump stabilization can restore optimal interpolation rates of convergence for the *Q1-Q0* and *P1-P0* methods. A key feature of the local jump stabilization is that a conventional macro-element implementation is possible, so that the new stabilized discrete formulation can be implemented into element-by-element iterative solution techniques. The stabilized mixed *Q1-Q0* and *P1-P0* methods introduced by Silvester and Kechkar [31] and then analyzed in [23] have been applied to Stokes equations in many applications (see [8], [18],[20] and [24]) while stability with respect to aspect ratio has been restored in [1] using minimal constraints on the pressure space.

In this work, some modifications of the local jump stabilization technique, discussed in [31] and [23], are proposed for the *Q1-Q0* mixed element when used for the discretization of the generalized Stokes problem. The latter is obtained through the introduction of an additional term to the classical Stokes problem. It occurs in the refined numerical modeling of most industrial incompressible fluid flows (see [9], [15],[30]). The added term can be taken as a Darcy term or may represent the time discretization of the evolution term in the unsteady-state Stokes problem. The present techniques consist in reducing the number of local jump terms in the discrete formulation to two jumps, and even to one jump in each 2×2 macro-element. Furthermore, the local jump framework can be more easily implemented into existing software codes. The well-posedness and convergence of the two new stabilized discrete formulations are theoretically discussed, whereas the robustness properties are exhibited through some computational test problems.

Solving the generalized Stokes problem by means of the stabilized finite element method *Q-Q0* leads us to deal with the arising large sparse linear systems. Although there are effective direct methods (called *sparse elimination methods*) based on the Gaussian elimination that exploit the sparsity of the coefficient matrix and reduce computational requirements, they work well only for moderate dimensions. For larger dimensions, they are very expensive and require infeasible computational resources. To reduce these costs, many people have switched to iterative methods. The developpement of these became a rigorous branch of numerical analysis. It should be mentioned that basic iterative methods do not converge for any linear system, or usually converge very slowly. By preconditioning, the convergence can be faster. A companion comparative discussion of the efficiency of some iterative solvers (in particular, Krylov subspace methods) is hence presented for the discrete linear systems arising from the stabilization via the different techniques.

The thesis is organized as follows. In the next chapter, some preliminaries that will be used are recalled. In the second chapter, the generalized Stokes problem is presented along with its weak formulation. Then, the standard Galerkin formulation is derived and discussed. In the following chapter, the local jump stabilized formulation by the way of the *Q1-Q0* mixed method is reviewed and the new formulations (local two-jump and one-jump) are introduced and analyzed. Discussion of some iterative solvers for the discrete linear systems is addressed to in the fourth chapter. In the last chapter, some implementation aspects of the proposed stabilization techniques are discussed in such a way that stability and convergence are proved while numerical performance of the new stabilization techniques is assessed on some test problems and compared to those of the earlier (local and global) jump stabilization techniques. A comparative study of the efficiency of the iterative solvers is also discussed for the different stabilized linear systems. Finally, some concluding remarks are drawn.

Chapter 1

Preliminaries

In this chapter, we give an overview of the relevant concepts that will be used next. We introduce elementary notations and basic analysis theory that will be used throughout the thesis

1.1 Functional spaces

Given a bounded domain Ω ($\subset \mathbb{R}^2$), with a polygonal boundary $\partial\Omega$, we denote by $L^2(\Omega)$ the Lebesgue space of functions that are square-integrable over Ω with respect to the inner product:

$$(q, r)_{0,\Omega} = \int_\Omega q\, r \; d\Omega \tag{1.1}$$

for all $q, r \in L^2(\Omega)$, and the corresponding norm

$$\|q\|_{0,\Omega} = \sqrt{(q, q)_{0,\Omega}}\,. \tag{1.2}$$

Denote by $L^2_0(\Omega)$, the subspace of functions in $L^2(\Omega)$ with zero mean over Ω, i.e.

$$L^2_0(\Omega) = \left\{ q \in L^2(\Omega) \; ; \; \int_\Omega q \; d\Omega = 0 \right\}. \tag{1.3}$$

The Sobolev space of square-integrable functions with square−integrable first derivatives is defined by

$$H^1(\Omega) = \left\{ v \in \; L^2(\Omega) \; ; \; \frac{\partial\, v}{\partial\, x_1} \; , \; \frac{\partial\, v}{\partial\, x_2} \in L^2(\Omega) \; \right\}. \tag{1.4}$$

The subspace of $H^1(\Omega)$ of functions that vanish on $\partial\Omega$ is given by

$$H^1_0(\Omega) = \left\{ v \in \; H^1(\Omega) \; ; \, v = 0 \text{ on } \partial\Omega \right\}. \tag{1.5}$$

The space $H^1(\Omega)$ is equiped with the inner product:

$$(v, w)_{1,\Omega} = \int_\Omega v\, w \, d\Omega + \int_\Omega \nabla v \cdot \nabla w \; d\Omega, \tag{1.6}$$

where

$$\nabla v \cdot \nabla w = \frac{\partial v}{\partial x_1}\frac{\partial w}{\partial x_1} + \frac{\partial v}{\partial x_2}\frac{\partial w}{\partial x_2}. \tag{1.7}$$

The vector

$$\nabla v = \begin{pmatrix} \dfrac{\partial v}{\partial x_1} \\ \\ \dfrac{\partial v}{\partial x_2} \end{pmatrix} \tag{1.8}$$

is called *the gradient* of v, whereas the Laplacian of v, denoted by Δv, is given by

$$\Delta v = \frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2} \tag{1.9}$$

and *the divergence* of v is given by:

$$div\, v = \frac{\partial v}{\partial x_1} + \frac{\partial v}{\partial x_2}. \tag{1.10}$$

The space $H^1(\Omega)$ can be equipped with the norm

$$\|v\|_{1,\Omega} = \sqrt{(v,v)_{1,\Omega}} = \sqrt{\|v\|_{0,\Omega}^2 + \|\nabla v\|_{0,\Omega}^2}\,. \tag{1.11}$$

For vector two valued functions $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$, the inner product can be generalized as follows

$$(\mathbf{v}, \mathbf{w})_{1,\Omega} = (v_1, w_1)_{1,\Omega} + (v_2, w_2)_{1,\Omega}\,, \tag{1.12}$$

with the corresponding norm

$$\|\mathbf{v}\|_{1,\Omega} = \sqrt{\|v_1\|_{1,\Omega}^2 + \|v_2\|_{1,\Omega}^2}\,. \tag{1.13}$$

Likewise, we define the gradient of the vector $\mathbf{v} = (v_1, v_2)$ by

$$\nabla \mathbf{v} = \begin{pmatrix} \dfrac{\partial v_1}{\partial x_1} & \dfrac{\partial v_2}{\partial x_1} \\ \\ \dfrac{\partial v_1}{\partial x_2} & \dfrac{\partial v_2}{\partial x_2} \end{pmatrix}. \tag{1.14}$$

1.2 Abstract mixed formulation

Let Z and Q be two Hilbert spaces with corresponding norms $\|.\|_Z$ and $\|.\|_Q$ respectively and denote their topological dual spaces by $Z^{'}$ and $Q^{'}$. In addition, let

$$a(.,.) : Z \times Z \to \mathbb{R} \tag{1.15}$$

and

$$b(.,.) : Q \times Z \to \mathbb{R} \tag{1.16}$$

be two bounded bilinear forms. i.e.

$$\exists\, C_a, C_b > 0 \quad \text{(finite) such that} \quad \begin{cases} |a(u,v)| \leq C_a \|u\|_Z \|v\|_Z & \forall u, v \in Z \\ |b(q,v)| \leq C_b \|q\|_Q \|v\|_Z & \forall (q,v) \in Q \times Z. \end{cases} \tag{1.17}$$

Now, consider the following problem:

$$\text{Given} \;\; l \in Z^{'}, \; \text{find} \, (u,p) \in Z \times Q \; \text{such that}$$

$$\begin{cases} a\,(u,v) + b\,(p,v) = l\,(v) & \forall v \in Z \\ b\,(q,u) = 0 & \forall q \in Q. \end{cases} \tag{1.18}$$

Moreover, let $A : Z \to Z^{'}$ and $B : Z \to Q^{'}$ be two continuous linear operators associated with respect to $a(.,.)$ and $b(.,.)$, defined respectively by:

$$\begin{cases} \langle Au, v \rangle = a(u,v) & \forall v \in Z \\ \langle Bv, q \rangle = b(q,v) & \forall q \in Q. \end{cases} \tag{1.19}$$

Similarly, let $B^{'} : Q \to Z^{'}$ be the dual of the operator B:

$$\langle B^{'} q, v \rangle = \langle q, Bv \rangle \quad \forall v \in Z. \tag{1.20}$$

The problem (1.18) can then be reformulated as follows:

$$\text{Find} \, (u,p) \in Z \times Q \; \text{satisfying}$$

$$\begin{cases} Au + B^{'} p = l & \text{in } Z^{'} \\ Bu = 0 & \text{in } Q^{'}. \end{cases} \tag{1.21}$$

By defining the null space of the operator B

$$M = KerB = \{ v \in Z; \; b(q,v) = 0 \quad \forall q \in Q \}, \tag{1.22}$$

the associated problem to (1.18) which is restricted to M is then given by

$$\text{Find} \, u \in M \; \text{such that}$$

$$a\,(u,v) = l\,(v) \qquad \forall v \in M \;. \tag{1.23}$$

Note that if $(u,p) \in Z \times Q$ is a solution of (1.18), then u is a solution of (1.23). Now, the question is: what are suitable conditions ensuring the converse? The following result can give an answer.

Theorem 1.1. *(Existence and uniqueness)*
Assume that the following hypotheses are satisfied.

H1 (M-ellipticity of a)

There exists a constant $c_1 > 0$ *such that*

$$a(v,v) \geq c_1 \|v\|_Z^2 \qquad \forall v \in M. \tag{1.24}$$

H2 (LBB condition)

There exists a constant $c_2 > 0$ *such that*

$$\sup_{0 \neq v \in Z} \frac{b(q,v)}{\|v\|_Z} \geq c_2 \|q\|_Q \qquad \forall\, q \in Q. \tag{1.25}$$

Then, the problem (1.23) *has a unique solution* $u \in M$ *and there exists a unique* $p \in Q$ *such that the pair* (u,p) *is the unique solution of problem* (1.18).

Proof. (see [6] or [16])

It is also instructive to rewrite Problem (1.18) in the following equivalent form:

$$\begin{gathered} \text{Given} \;\; l \in Z', \; \text{find} \, (u,p) \in Z \times Q \; \text{such that} \\ \mathcal{B}\left((u,p)\,;(v,q)\right) = \mathcal{L}\,(v,q) \qquad \forall\,(v,q) \in Z \times Q, \end{gathered} \tag{1.26}$$

where

$$\begin{aligned} \mathcal{B}\left((u,p)\,;(v,q)\right) &= a(u,v) + b(p,v) + b(q,u) \\ \mathcal{L}\,(v,q) &= l(v), \end{aligned} \tag{1.27}$$

This problem is used essentially for the following theorem.

Theorem 1.2.
Let W_1 *and* W_2 *be two Hilbert spaces with norms* $\|.\|_1$ *and* $\|.\|_2$ *respectively. Further, let* $\mathcal{B}(.,.)$ *be a bilinear form on* $W_1 \times W_2$ *such that*

$$|\mathcal{B}(z,t)| \leq C_1 \|z\|_1 \|t\|_2 \quad \forall\,(z,t) \in W_1 \times W_2,$$

$$\sup_{z \in W_1} \frac{|\mathcal{B}(z,t)|}{\|z\|_1} \geq C_2 \|t\|_2 \quad \forall t \in W_2,$$

$$\sup_{t \in W_2} \frac{|\mathcal{B}(z,t)|}{\|t\|_2} \geq C_3 \|z\|_1 \quad \forall z \in W_1,$$

with $C_1 < \infty$ *and* $C_2 > 0$, $C_3 > 0$. *Further, let* $\mathcal{L}(.)$ *be a linear functional on* W_2, *i.e.* $\mathcal{L} \in W_2'$. *Then, there exists exactly one element* $z_0 \in W_1$ *such that*

$$\mathcal{B}(z_0,t) = \mathcal{L}\,(t)$$

for all $t \in W_2$ *and*

$$\|z_0\|_1 \leq \frac{\|\mathcal{L}\|_\infty}{C_3},$$

where

$$\|\mathcal{L}\|_\infty = \sup_{r \in W_2} \frac{\mathcal{L}(r)}{\|r\|_2}.$$

Proof. (see [2]) □

It should be mentioned that this theorem is very useful in the analysis of mixed finite element approximations of the continuous generalized Stokes problem.

1.3 Approximation of mixed problems

In this section, we focus on the approximation of the abstract mixed problems discussed above. Keeping the same notations as in the previous section, let h denote a discretization parameter tending to zero, and let $Z_h \subset Z$ and $Q_h \subset Q$ be two finite dimensional subspaces. The closed linear subspace M_h of the linear space Z_h analogue to (1.22) is defined by

$$M_h = \{v_h \in Z_h;\ b(q_h, v_h) = 0 \quad \forall q_h \in Q_h\}. \tag{1.28}$$

Obviously, M_h is nonempty since $0 \in M_h$. Further, it should be noted that in general, M_h is not a subset of M. Hence, even though the bilinear form $b(.,.)$ featured in (1.18) satisfies the stability condition (1.25), it does not necessarily satisfy the discrete stability condition with respect to Z_h and Q_h. This makes the construction of finite space approximations to mixed variational problems difficult. The validity of a discrete stability condition must, hence, be independently verified for each particular choice of spaces Z_h and Q_h. Now, following the Galerkin methodology the problem (1.18) can be approximated by the following:

$$\text{Find } (u_h, p_h) \in Z_h \times Q_h \text{ such that}$$

$$\begin{cases} a(u_h, v_h) + b(p_h, v_h) = l(v_h) & \forall v_h \in Z_h \\ b(q_h, u_h) = 0 & \forall q_h \in Q_h, \end{cases} \tag{1.29}$$

with the associated restricted discrete problem:

$$\begin{gathered} \text{Find } u_h \in M_h \text{ such that} \\ a(u_h, v_h) = l(v_h) \quad \forall v_h \in M_h. \end{gathered} \tag{1.30}$$

Since $M_h \not\subseteq M$ in general, the problem (1.30) may be considered as an external approximation of (1.23). As it was seen above, the first component u_h of any solution (u_h, p_h) of (1.29) is also a solution of (1.30). For treating the converse, the following discrete version of Theorem 1.1 is crucial.

Theorem 1.3. *(Existence and uniqueness)*
Assume the following hypotheses.

H1 (M_h-ellipticity of a)

There exists a constant $c_1^* > 0$ such that

$$a(v_h, v_h) \geq c_1^* \|v_h\|_Z^2 \qquad \forall v_h \in M_h. \tag{1.31}$$

H2 (Discrete LBB condition)

There exists a constant $c_2^* > 0$ such that

$$\sup_{0 \neq v_h \in Z_h} \frac{b(q_h, v_h)}{\|v_h\|_Z} \geq c_2^* \|q_h\|_Q \qquad \forall q_h \in Q_h. \tag{1.32}$$

Then, the problem (1.30) has a unique solution $u_h \in M_h$ and there exists a unique $p_h \in Q_h$ such that the pair (u_h, p_h) is the unique solution of problem (1.29). Moreover, there exists a constant C dependent on c_1^* and c_2^* such that

$$\|u - u_h\|_Z + \|p - p_h\|_Q \leq C \left(\inf_{v_h \in Z_h} \|u - v_h\|_Z + \inf_{q_h \in Q_h} \|p - q_h\|_Q \right). \tag{1.33}$$

Proof. (see [16]) □

1.4 Some aspects from linear algebra

In this section, we assume that the reader is familiar with the basic notions of matrices and their corresonding operations. Now, we denote by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{w}^T \mathbf{v}, \tag{1.34}$$

where $\mathbf{w}^T$ is the transpose of $\mathbf{w}$, the Euclidean inner product on $\mathbb{R}^n$ with the associated norm:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \tag{1.35}$$

Definition 1.1.
A square matrix A is said to be :

- *symmetric if*
$$A^T = A. \tag{1.36}$$
- *positive definite if*
$$\langle A\mathbf{v}, \mathbf{v} \rangle > 0 \qquad \forall\, \mathbf{v} \neq 0, \tag{1.37}$$
- *positive semi-definite if*
$$\langle A\mathbf{v}, \mathbf{v} \rangle \geq 0 \qquad \forall\, \mathbf{v} \in \mathbb{R}^n, \tag{1.38}$$
- *negative definite if*
$$\langle A\mathbf{v}, \mathbf{v} \rangle < 0 \qquad \forall\, \mathbf{v} \neq 0, \tag{1.39}$$

- *negative semi-definite if*
$$\langle A\mathbf{v}, \mathbf{v}\rangle \leq 0 \qquad \forall\, \mathbf{v} \in \mathbb{R}^n, \tag{1.40}$$
- *indefinite if it is neither positive semi-definite nor negative semi-definite.*

If A is a symmetric positive definite matrix of order n, then the bilinear form given by

$$\langle \mathbf{v}, \mathbf{w}\rangle_A = \mathbf{w}^T A\mathbf{v} \tag{1.41}$$

defines an inner product on $\mathbb{R}^n$ with the associated norm:

$$\|\mathbf{v}\|_A = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle_A}. \tag{1.42}$$

Definition 1.2.
Symmetric matrices Y and Z are said to be congruent if $Y = X\,Z\,X^T$ for some non-singular matrix X.

Theorem 1.4. *(Sylvester law of inertia)*
Any congruent matrices have the same number of negative, zero and positive eigenvalues.

Definition 1.3.
The condition number of a non singular matrix A is given by

$$\kappa(A) = \|A\|\|A^{-1}\|, \tag{1.43}$$

where

$$\|A\| = \max_{v\neq 0} \frac{\|Av\|}{\|v\|}.$$

If A is symmetric and positive definite then, $\|A\| = \lambda_{max}(A)$ and $\|A^{-1}\| = \dfrac{1}{\lambda_{min}(A)}$, where λ denotes the eigenvalue of the matrix A. In this case, we have that

$$\kappa(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}. \tag{1.44}$$

Chapter 2

The generalized Stokes problem and its approximation

Usually, it is not easy to get analytic solutions to the Stokes problems type, which leads us to think of approximate solutions instead. To this end, numerical methods can play an important role and hence, be strongly recommended. In this chapter, the generalized Stokes equations are presented along with the weak formulation that is needed to approximate the problem using the low-order mixed finite elements method in the third section.

2.1 The generalized Stokes equations

Let Ω be a bounded two-dimensional domain with a polygonal boundary $\partial\Omega$. Consider the incompressible generalized Stokes problem, also called the Brinkman model: Given a body force $\mathbf{f}$, find functions $\mathbf{u} = (u_1, u_2)$ and p defined in Ω such that

$$\begin{aligned} \alpha\mathbf{u} - \mu\Delta\mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega, \\ div\ \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= \mathbf{g} \quad \text{on } \partial\Omega, \end{aligned} \tag{2.1}$$

where $\mathbf{u}$ is the fluid velocity, p the pressure, $\mu > 0$ the kinematic viscosity coefficient, $\mathbf{g}$ is a prescribed velocity on $\Gamma = \partial\Omega$, and α a positive real number that may come from the time discretization of the evolution term $\frac{\partial \mathbf{u}}{\partial t}$ in the unsteady-state Stokes equations (cf. [9]). Typically, we have $\alpha >> 1$.

Following the well-known monograph [16], there exists $\mathbf{u}_0 \in H^1(\Omega)$ such that $\mathbf{u}_0 = \mathbf{g}$ on Γ and $div\,\mathbf{u}_0 = 0$ in Ω. Therefore, setting $\mathbf{U} = \mathbf{u} - \mathbf{u}_0$ gives the problem:

$$\begin{aligned} &\text{Find functions } \mathbf{U} \text{ and } p \text{ defined in } \Omega \text{ such that} \\ \alpha\mathbf{U} - \mu\Delta\mathbf{U} + \nabla p &= \mathbf{f} - \alpha\mathbf{u}_0 + \mu\Delta\mathbf{u}_0 \quad \text{in } \Omega, \\ div\ \mathbf{U} &= 0 \quad \text{in } \Omega, \\ \mathbf{U} &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{2.2}$$

The problem can then be stated as follows.

$$\begin{aligned} &\text{Find functions } \mathbf{u} = (u_1, u_2) \text{ and } p \text{ defined in } \Omega \text{ such that} \\ &\qquad\qquad \alpha\mathbf{u} - \mu\Delta\mathbf{u} + \nabla p = \mathbf{f} \qquad \text{in } \Omega, \\ &\qquad\qquad div\ \mathbf{u} = 0 \qquad \text{in } \Omega, \\ &\qquad\qquad \mathbf{u} = 0 \qquad \text{on } \Gamma, \end{aligned} \tag{2.3}$$

where $\mathbf{u}$ is used instead of $\mathbf{U}$. Throughout the analysis, the homogeneous Dirichlet condition in (2.3), called no-slip boundary condition, is considered here only for simplicity of presentation. Other boundary conditions can also be taken as it will be the case below in the numerical experiments. Using definitions (1.8), (1.9) and (1.10) given above, the problem (2.3) is then equivalent to the follows:

$$\begin{aligned} &\text{Find functions } \mathbf{u} = (u_1, u_2) \text{ and } p \text{ defined in } \Omega \text{ such that} \\ &\alpha u_1 - \mu\left(\tfrac{\partial^2 u_1}{\partial x_1^2} + \tfrac{\partial^2 u_1}{\partial x_2^2}\right) + \tfrac{\partial p}{\partial x_1} = f_1 \qquad \text{in } \Omega, \\ &\alpha u_2 - \mu\left(\tfrac{\partial^2 u_2}{\partial x_1^2} + \tfrac{\partial^2 u_2}{\partial x_2^2}\right) + \tfrac{\partial p}{\partial x_2} = f_2 \qquad \text{in } \Omega, \\ &\qquad\qquad \tfrac{\partial u_1}{\partial x_1} + \tfrac{\partial u_2}{\partial x_2} = 0 \qquad \text{in } \Omega, \\ &\qquad\qquad u_1 = 0 \ \text{ and } u_2 = 0 \qquad \text{on } \Gamma, \end{aligned} \tag{2.4}$$

where $\mathbf{f} = (f_1, f_2)$.

2.2 Weak formulation of the generalized Stokes problem

First, let us consider the function spaces:

$$P = L_0^2(\Omega) \quad \text{and} \quad \mathbf{V} = [H_0^1(\Omega)]^2, \tag{2.5}$$

being the usual Lebesgue and Sobolev spaces, defined in (1.3) and (1.5) respectively. The choice of the pressure function space in P is needed to ensure the uniqueness since it is clear from (2.3) that the pressure can be determined only up to an additive constant. Then, a weak formulation of the generalized Stokes problem (2.3) is given as follows:

$$\begin{aligned} &\text{Find } (\mathbf{u}, p) \in \mathbf{V} \times P \text{ such that:} \\ &\alpha\textstyle\int_\Omega \mathbf{u}\cdot\mathbf{v}\ d\Omega + \mu\int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{v}\, d\Omega \quad -\int_\Omega p\ div\,\mathbf{v}\ d\Omega \\ &\qquad\qquad\qquad\qquad = \textstyle\int_\Omega \mathbf{f}\cdot\mathbf{v}\, d\Omega \qquad \forall\ \mathbf{v} \in \mathbf{V}, \\ &\qquad\qquad -\textstyle\int_\Omega q\ div\ \mathbf{u}\ d\Omega \quad = 0 \qquad \forall\ q \in P. \end{aligned} \tag{2.6}$$

Note that we have multiplied the second equation in (2.6) by minus one. The purpose of this is to get a symmetric formulation which will greatly simplify the discussion. Further,

we can take the right-hand side $\mathbf{f}$ in $[L^2(\Omega)]^2$, although this space is not the largest function space for the data $\mathbf{f}$ such that (2.6) makes sense.

Define the bilinear forms: for all $(\mathbf{v}, \mathbf{w}) \in \mathbf{V} \times \mathbf{V}$ and $q \in P$,

$$a(\mathbf{v}, \mathbf{w}) = \alpha \int_\Omega \mathbf{v} \cdot \mathbf{w} \, d\Omega + \mu \int_\Omega \nabla \mathbf{v} \cdot \nabla \mathbf{w} \, d\Omega, \tag{2.7}$$

$$b(q, \mathbf{w}) = -\int_\Omega q \, div \, \mathbf{w} \, d\Omega \tag{2.8}$$

and the linear form

$$L(\mathbf{v}) = \int_\Omega \mathbf{f} \cdot \mathbf{v} \, d\Omega \; . \tag{2.9}$$

The weak formulation (2.6) can then be written in the following form.

Find $(\mathbf{u}, p) \in \mathbf{V} \times P$ such that:

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) + b(p, \mathbf{v}) &= L(\mathbf{v}) \quad \forall \, \mathbf{v} \in \mathbf{V}, \\ b(q, \mathbf{u}) &= 0 \qquad \forall \, q \in P. \end{aligned} \tag{2.10}$$

Following standard arguments from the classical theory (cf. [16]), it can be shown that there is a unique solution $(\mathbf{u}, p)$ to the weak formulation (2.6) by applying Theorem 1.1 (cf. [13] for more details).

2.3 Approximation using mixed finite elements

Denote by $h\,(> 0)$ the mesh parameter. Adopting a conform mixed finite element method by using the finite-dimensional subspaces $\mathbf{V}_h \subset \mathbf{V}$ and $P_h \subset P$, the standard Galerkin methodology yields the following approximate problem:

Find $(\mathbf{u}_h, p_h) \in \mathbf{V}_h \times P_h$ such that

$$\begin{aligned} \alpha \int_\Omega \mathbf{u}_h \cdot \mathbf{v} d\Omega + \mu \int_\Omega \nabla \mathbf{u}_h \cdot \nabla \mathbf{v} \, d\Omega \quad & - \int_\Omega p_h \, div \, \mathbf{v} \, d\Omega \\ & = \int_\Omega \mathbf{f} \cdot \mathbf{v} \, d\Omega \qquad \forall \, \mathbf{v} \in \mathbf{V}_h, \end{aligned} \tag{2.11}$$

$$- \int_\Omega q \, div \, \mathbf{u}_h \, d\Omega = 0 \qquad \forall \, q \in P_h. \tag{2.12}$$

The domain Ω is subdivided into convex quadrilaterals such that the resulting partitioning τ_h is regular in the usual sense, i.e. for some positive constants $\sigma > 1$ and $0 < \varepsilon < 1$ we have

$$h_K \; \leq \; \sigma \, \rho_K \quad \text{and} \quad |\, cos \, \theta_{i,K}| \; \leq \varepsilon \quad \forall K \; \in \tau_h \tag{2.13}$$

where h_K is the diameter of the element K, ρ_K is the diameter of the inscribed circle of K and $\theta_{i,K}$ $(i = 1, 2, 3, 4)$ are the angles of K. The mesh parameter h is explicitly given by

$$h = \max_{K \, \in \, \tau_h} \; h_K.$$

Furthermore, the *Q1-Q0* mixed method is the lowest order conforming quadrilateral approximation method and is characterized by the pair of finite-dimensional spaces $\{V_h, P_h\}$ defined by

$$V_h = \left\{ v \in V \cap C^0(\Omega) \; ; \; v|_K \in [Q_1(K)]^2 \quad \forall K \in \tau_h \right\} \tag{2.14}$$

and

$$P_h = \{ q \in P \; ; \; q|_K \in Q_0(K) \quad \forall K \in \tau_h \} \tag{2.15}$$

where $Q_1(K)$ is the space of iso-parametrically transformed bilinear functions in each K and $Q_0(K)$ is the space of constant functions in each K.

As it was pointed out above, it has been shown in [4] that the finite element space pair given by (2.14) and (2.15) does not satisfy the key discrete *LBB* stability condition:

$$\exists \, \omega > 0 \qquad \sup_{0 \neq v \in V_h} \frac{(q, div \mathbf{v})}{|\mathbf{v}|_1} \geq \omega \|q\|_0 \qquad \forall q \in P_h. \tag{2.16}$$

In this respect, in [28] and [29] it has also been demonstrated that the method develops spurious pressure modes resulting in numerical instabilities in the approximate pressure for certain boundary conditions. This suggests that for one or a few, but not all, $q \in P_h$ we have:

$$\int_\Omega q \; div \; \mathbf{v} \; d\Omega = 0 \qquad \forall \, \mathbf{v} \in \mathbf{V}_h. \tag{2.17}$$

On the other hand, Boland and Nicolaïdes have also shown in [5] that in this case there is a more important failure of (2.16). In particular, they established the existence of some $q \in P_h$ such that

$$C_1 \; h \; \|q\|_0 \; \leq \; \sup_{0 \neq v \in V_h} \frac{(q, div \mathbf{v})}{|\mathbf{v}|_1} \; \leq \; C_2 \; h \|q\|_0, \tag{2.18}$$

which implies that $\omega = 0$ in (2.16). Hence, on the contrary to the weak formulation (2.6), the unique solution to the approximate problem (2.11), (2.12) is not garanteed since the second hypothesis of Theorem 1.3 is not satisfied.

However, it should also be worth noting that $Q1$-Q0 mixed approximation is stable if some non-rectangular meshes are used (cf. [14]). Because of the complexity of these latters and despite (2.17) and (2.18), the use of $Q1$-Q0 approximation on rectangular meshes is motivated by its computational convenience. For simplicity of presentation, the rectangular axis-parallel meshes of the domain Ω will be considered in the remainder. The results can be easily extended to more general quadrilateral meshes using standard isoparametric transformations (see Figure 2.1):

$$\begin{cases} x(\xi, \eta) = & \displaystyle\sum_{j=1}^{4} x_j N_j(\xi, \eta), \\ \\ y(\xi, \eta) = & \displaystyle\sum_{j=1}^{4} y_j N_j(\xi, \eta), \end{cases} \tag{2.19}$$

where on the reference element $\hat{K} = [-1.1] \times [-1.1]$, the basis functions are given by

$$\begin{cases} N_1(\xi,\eta) = \frac{1}{4}(1-\xi)\,(1-\eta), \\ N_2(\xi,\eta) = \frac{1}{4}(1+\xi)\,(1-\eta), \\ N_3(\xi,\eta) = \frac{1}{4}(1+\xi)\,(1+\eta), \\ N_4(\xi,\eta) = \frac{1}{4}(1-\xi)\,(1+\eta). \end{cases} \tag{2.20}$$

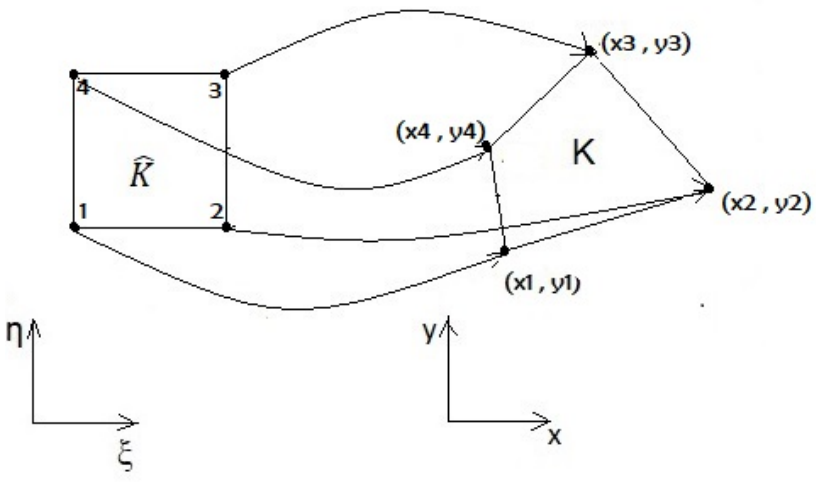

Figure 2.1: Isoparametric transformation.

Chapter 3

Local stabilizations of the *Q1-Q0* mixed finite element

In this chapter, some modifications of the local jump stabilization technique, are proposed for the *Q1-Q0* mixed element for the discretization of the generalized Stokes problem. The present techniques consist in reducing the number of local jump terms in the discrete formulation to two jumps, and even to only one jump in each 2x2 macro-element. Most of the materials presented in this chapter have been the object of a recently published paper [10].

3.1 Earlier jump stabilizations

In order to overcome the major difficulty mentioned above, the discrete problem (2.11), (2.12) can be stabilized by introducing into the equation (2.12) a bounded symmetric bilinear form $C_h(.,.)$ which is positive semi-definite over $P_h \times P_h$. In so doing, this induces the modified discrete incompressibility constraint:

$$-\int_\Omega q \; div \; \mathbf{u}_h \; d\Omega \; - C_h(p_h, q) = 0 \quad \forall \; q \in P_h. \tag{3.1}$$

The motivation behind introducing the form $C_h(.,.)$ is the following result which provides a sufficient condition for the well-posedness of the new discrete problem (2.11), (3.1).

Theorem 3.1.
Assume that the form $C_h(.,.)$ satisfies the condition:

$$\left\{ \begin{array}{l} \textit{For any } p_m \in P_h \textit{ such that } \quad \int_\Omega p_m div \; \boldsymbol{v} \; d\Omega = 0 \; \forall \boldsymbol{v} \in V_h \quad \textit{we have} \\ \\ \qquad\qquad C_h(p_m, p_m) = 0 \quad \Longrightarrow \; p_m = 0 \; . \end{array} \right. \tag{3.2}$$

Then, the solution $(\boldsymbol{u}_h, p_h)$ of (2.11), (3.1) is uniquely determined in $\boldsymbol{V}_h \times P_h$.

Proof. (see [31]) □

One such way of stabilizing the $Q1$-Q0 mixed method was introduced in [21] through the so-called global jump stabilization. It consists in introducing the bilinear form $C_h(.,.)$

in the discrete incompressibility equation such that:

$$C_h(p_h, q) = \beta \sum_{e=1}^{N_e} h^{(e)} \int_{\Gamma^{(e)}} [|p_h|]_{\Gamma^{(e)}} \, [|q|]_{\Gamma^{(e)}} ds. \tag{3.3}$$

Here, $[|.|]_{\Gamma^{(e)}}$ is the jump operator across $\Gamma^{(e)}$ and $\beta(> 0)$ is a stabilizing parameter. The summation runs over all interior inter-element edges $\{\Gamma^{(e)}; e = 1, 2, ..., N_e\}$ with lengths $h^{(e)}$. The modified discrete incompressibility constraint (3.1) is then:

$$-\int_\Omega q \; div \; \mathbf{u}_h \; d\Omega - \beta \sum_{e=1}^{N_e} h^{(e)} \int_{\Gamma^{(e)}} [|p_h|]_{\Gamma^{(e)}} \, [|q|]_{\Gamma^{(e)}} ds = 0 \qquad \forall q \in P_h. \tag{3.4}$$

A general theoretical analysis of the global jump stabilized formulation is given in [21]. Later, it was demonstrated in [31] that the global jump stabilization can be effective in practice. However, a careful choice of the parameter β is required to keep the accuracy in the solution.

As it was discussed in [31], the global jump method could be simplified by modifying the discrete bilinear form (3.3) by using macro-elements. For this end, assume that the elements in τ_h can be assembled into disjoint macro-elements (element paths) so that a macro-element partitioning $\mathcal{M}_h$ is constructed. Moreover, the notion of the equivalence macro-element classes which are topologically equivalent to a reference macro-element $\widehat{M}$ (cf. [31]) leads to the following macro-element internal regularity condition: there exists a constant $\omega(\widehat{M}) > 0$ such that

$$K_M \; \geq \; \omega_{\widehat{M}} \, G_M, \tag{3.5}$$

where

$$G_M = \max_{K \subset M} \; |K|, \; K_M = \min_{K \subset M} \; |K|$$

and $|K|$ represents the area of K. In addition, suppose that the common boundary of any two neighboring macro-elements M_1, M_2 in $\mathcal{M}_h$ contains a node strictly in the interior of this boundary (connectivity macro-element condition). Then, the bilinear stabilization term can be given, instead of (3.3), by

$$C_h(p_h, q) = \beta \sum_{M=1}^{N_M} \sum_{i=1}^{e_M} h_M^{(i)} \int_{\Gamma_M^{(i)}} [|p_h|]_{\Gamma_M^{(i)}} \, [|q|]_{\Gamma_M^{(i)}} \, ds, \tag{3.6}$$

where the first summation is over all macro-elements, whereas the second summation runs over all inter-element edges strictly within each macro-element (see Figure 3.1 for a 2x2 rectangular macro-element). This stabilization technique will be referred to as the local jump stabilization.

Provided some usual smoothness and regularity assumptions on the solution $(\mathbf{u}, p)$, the following optimal error estimates for the Stokes problem ($\alpha = 0$) are theoretically established in [23]:

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 \leq \; Ch \, (|\mathbf{u}|_2 + |p|_1) \tag{3.7}$$

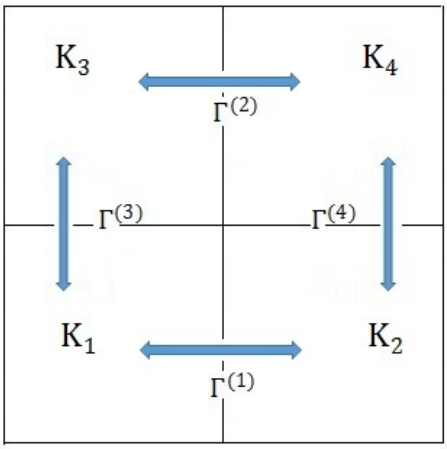

Figure 3.1: A 2x2 rectangular macro-element with four pressure jumps.

$$\|\mathbf{u} - \mathbf{u}_h\|_0 \leq \; Ch^2 \left(|\mathbf{u}|_2 + |p|_1\right). \tag{3.8}$$

The arguments developed in [31] suggest that the local jump stabilization can be preferred to the global jump one because of its special features:

- (i) The implementation is more straightforward since the local stabilization matrix obtained from (3.6) is block diagonal.
- (ii) The local mass conservation is preserved.
- (iii) The discrete velocity solution is less sensitive to the size of the stabilization parameter β.

Clearly, the two pressure jump stabilization techniques can be extended to the generalized Stokes problem (2.3) to get the error estimates (3.7) and (3.8) with the constant C depending on the parameter α as well.

In the remainder, for ensuring the connectivity macro-element condition mentioned above we assume that a coarser mesh $\mathcal{M}_h$ is given and that the latter is refined by joining the opposed element mid-edge points to get the grid τ_h. Next, two approaches for reduced local jump formulations will be presented.

3.2 New local jump schemes

3.2.1 Reduced local two-jump stabilizations

With preserved consistency and without apparently losing stability and convergence properties, the choice of (3.6) can be changed into a first reduced version by considering

the pressure jumps only on one direction (horizontal or vertical) over inter-element edges strictly within each macro-element (see Figure 3.2). The two new stabilized methods,

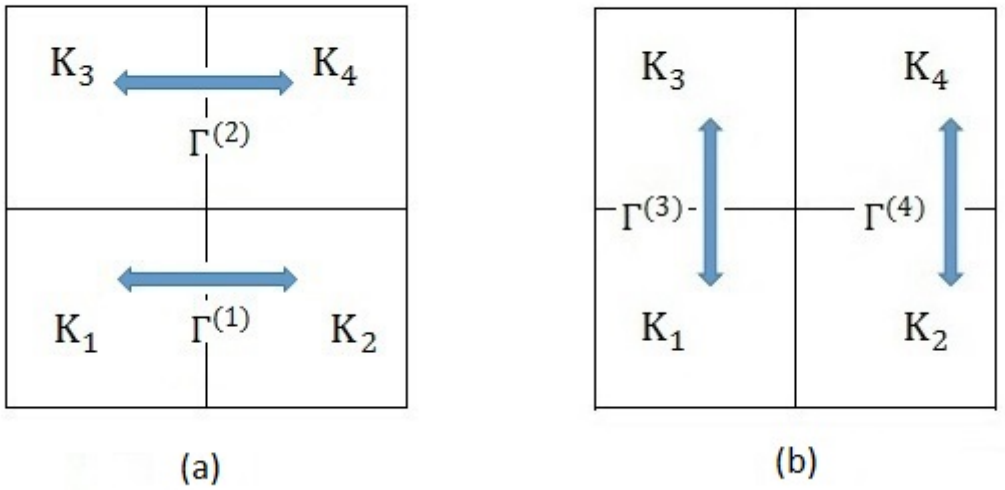

Figure 3.2: A 2x2 rectangular macro-element with two pressure jumps.

which will be referred to as the reduced local two-jump stabilizations, are given by the bilinear form:

$$C_h^{(2)}(p_h,q) = \begin{cases} \beta \displaystyle\sum_{M=1}^{N_M} \sum_{i=1}^{2} h_M^{(i)} \int_{\Gamma_M^{(i)}} [|p_h|]_{\Gamma_M^{(i)}} \, [|q|]_{\Gamma_M^{(i)}} ds \\ \text{or} \\ \beta \displaystyle\sum_{M=1}^{N_M} \sum_{i=3}^{4} h_M^{(i)} \int_{\Gamma_M^{(i)}} [|p_h|]_{\Gamma_M^{(i)}} \, [|q|]_{\Gamma_M^{(i)}} ds \, . \end{cases} \tag{3.9}$$

This leads to the following perturbed discrete incompressibility constraint:

$$-\int_\Omega q \; div \; \mathbf{u}_h \; d\Omega - C_h^{(2)}(p_h,q) = 0 \qquad \forall q \in P_h. \tag{3.10}$$

The well posedness of the so-obtained discrete problem is given by the next result.

Theorem 3.2.
The reduced local two-jump formulation (2.11), (3.10) *has a unique solution.*

Proof.
The proof follows exactly, that is used for the Stokes problem using the local jump stabilization technique (see [22]).
First, assume that the horizontal version (see Figure 3.2 (a)) is adopted. As it is clear from Theorem 3.1, in order to establish the assertion of the theorem, it is sufficient to prove the criterion (3.2) for the choice (3.9). We will proceed by contradiction. Namely, for any non-constant $p_m \in P_h$ (in particular, non-zero since $P_h \subset L_0^2(\Omega)$) such that

$$\int_\Omega p_m \; div \; \mathbf{v} \; d\Omega = 0 \qquad \forall \mathbf{v} \in \mathbf{V}_h \tag{3.11}$$

we have

$$C_h^{(2)}(p_m, p_m) \neq 0. \tag{3.12}$$

So, let $p_m \in P_h$ be a non-constant function with the property (3.11). From the definition (2.15), there exist at least two adjoining elements K_1 , K_2 $\in \tau_h$ such that $p_m|_{K_1} \neq p_m|_{K_2}$. Set $e_0 = K_1 \cap K_2$, $a = p_m|_{K_1}$ and $b = p_m|_{K_2}$. This yields

$$a \neq b \quad \text{and} \quad [|p_m|]_{e_0} \neq 0. \tag{3.13}$$

Three cases can occur for the inter-element boundary e_0, as depicted in Figure 3.3

- (1) There exists a macro-element $M \in \mathcal{M}_h$ such that $K_1, K_2 \subset M$ and the edge e_0 is in the horizontal direction (Figure 3.3 (a)). Then,

$$C_h^{(2)}(p_m, p_m) \geq \beta h_{e_0} \int_{e_0} [|p_m|]_{e_0}^2 ds > 0 \tag{3.14}$$

in virtue of (3.13).

- (2) There exists a macro-element $M \in \mathcal{M}_h$ such that $K_1, K_2 \subset M$ but on the contrary, the edge e_0 is in the vertical direction (Figure 3.3 (b)). Then, there exists a node P_0 in the interior of M and the corresponding basis function $\mathbf{v}_0 \in \mathbf{V}_h$ such that

$$supp\ \mathbf{v}_0 \subset M \text{ and } \int_\Sigma \mathbf{v}_0 \cdot \mathbf{n}\ ds\ \neq 0$$

where $\Sigma = (K_1 \cup K_3) \cap (K_2 \cup K_4)$ is the common doundary. Taking $\mathbf{v} = \mathbf{v}_0$ in (3.11) gives

$$\int_M p_m\ div\ \mathbf{v}_0\ d\Omega\ = 0.$$

It follows that

$$\begin{aligned} 0 = &\quad a\ \textstyle\int_{K_1 \cup K_3}\ div\ \mathbf{v}_0\ d\Omega\ +\ b\ \int_{K_2 \cup K_4}\ div\ \mathbf{v}_0\ d\Omega \\ = &\quad a\ \textstyle\int_{\partial(K_1 \cup K_3)}\ \mathbf{v}_0 \cdot \mathbf{n}_1\ ds\ +\ b\ \int_{\partial(K_2 \cup K_4)}\ \mathbf{v}_0 \cdot \mathbf{n}_2\ ds \end{aligned}$$

where $\partial\,(K_1 \cup K_3)$ is the boundary of $K_1 \cup K_3$ with the outward normal $\mathbf{n}_1$, and $\partial\,(K_2 \cup K_4)$ is the boundary of $K_2 \cup K_4$ with the outward normal $\mathbf{n}_2$. On the other hand, since $supp\ \mathbf{v}_0 \subset M$, we have

$$\int_{\partial(K_1 \cup K_3)} \mathbf{v}_0 \cdot \mathbf{n}_1\ ds\ = -\ \int_{\partial(K_2 \cup K_4)} \mathbf{v}_0 \cdot \mathbf{n}_2\ ds.$$

Substituting the latter into the preceding equation yields $a = b$ which is in contradiction with (3.13). That means that this situation never happens.

- (3) There are two neighboring macro-elements $M_1, M_2 \in \mathcal{M}_h$ such that $K_1 \subset M_1$ and $K_2 \subset M_2$ with $e_0 = K_1 \cap K_2 \subset M_1 \cap M_2$ (Figure 3.3 (c_1) or (c_2)). Therefore, $\int_{M_1 \cap M_2} ds \neq 0$.

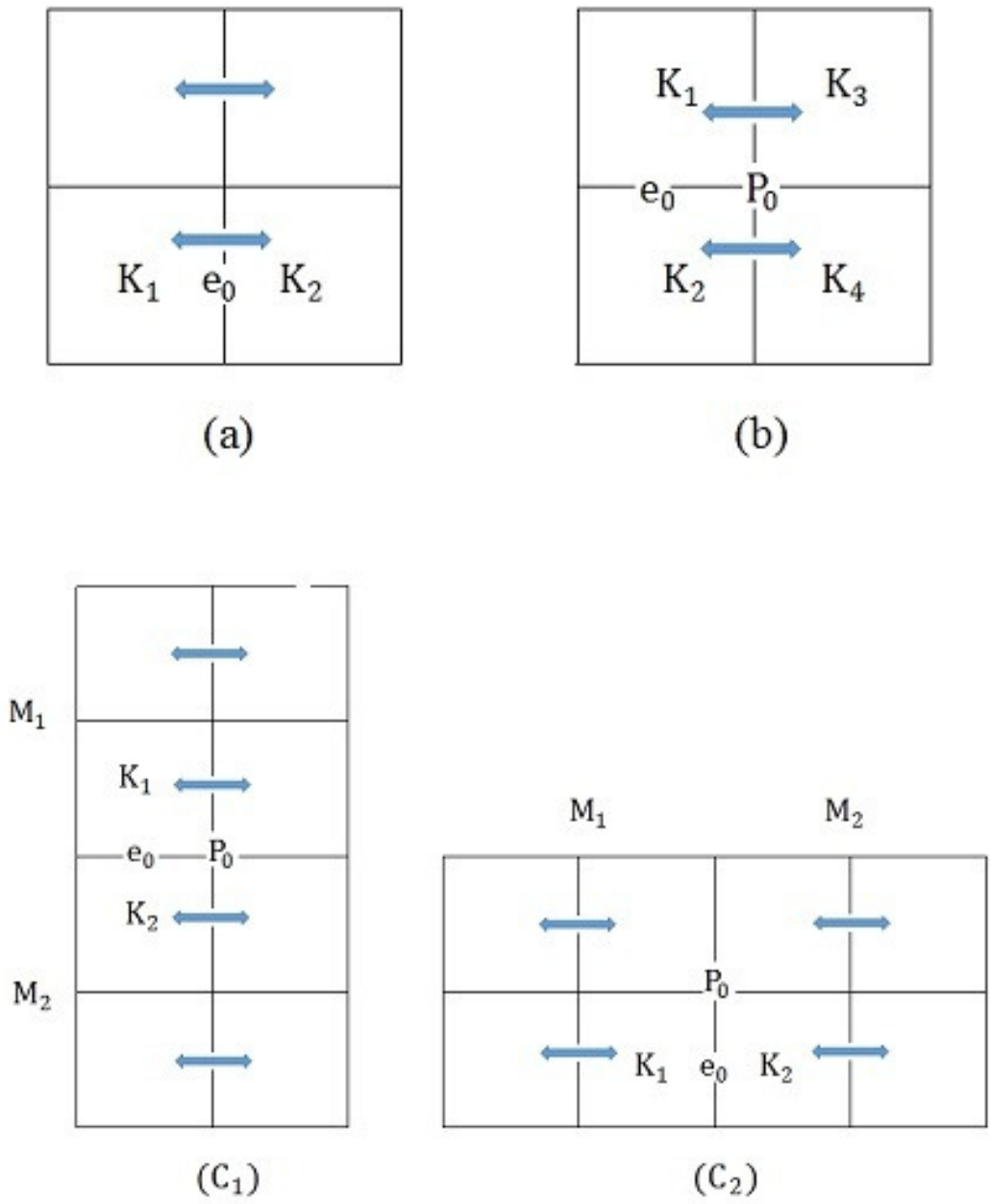

Figure 3.3: Three cases for the inter-element boundary e_0.

Again, there exist a node P_0 in the common boundary of M_1 and M_2 and a basis function $\mathbf{v}_0 \in \mathbf{V}_h$ satisfying :

$$supp\ \mathbf{v}_0 \subset M_1 \cup M_2 \text{ and } \int_{M_1 \cap M_2} \mathbf{v}_0 \cdot \mathbf{n}\ ds\ \neq 0.$$

p_m must be constant in M_1 and M_2 respectively, since, otherwise, we have the previous case.,
Of course, $a = p_m|_{M_1}$ and $b = p_m|_{M_2}$. Taking $\mathbf{v} = \mathbf{v}_0$ in (3.11) yields

$$\begin{aligned} 0 = &\ \int_{M_1 \cup M_2} p_m\ div\ \mathbf{v}_0\ d\Omega \\ = &\ a\ \int_{M_1}\ div\ \mathbf{v}_0\ d\Omega\ +\ b\ \int_{M_2}\ div\ \mathbf{v}_0\ d\Omega \\ = &\ a\ \int_{\partial M_1}\ \mathbf{v}_0 \cdot \mathbf{n}_1\ ds\ +\ b\ \int_{\partial M_2}\ \mathbf{v}_0 \cdot \mathbf{n}_2\ ds, \end{aligned}$$

where ∂M_i is the boundary of M_i and $\mathbf{n}_i$ is the outward normal to ∂M_i $(i = 1, 2)$. On the other hand, since $supp\ \mathbf{v}_0 \subset M_1 \cup M_2$, we have

$$\int_{\partial M_i}\ \mathbf{v}_0 \cdot \mathbf{n}_i\ ds = \pm \int_{M_1 \cap M_2}\ \mathbf{v}_0 \cdot \mathbf{n}\ ds \quad i = 1, 2,$$

where $\mathbf{n}$ denotes a direction normal to $M_1 \cap M_2$. The appropriate sign depends on whether $\mathbf{n}$ is oriented from M_1 to M_2 , or conversely. Thus,

$$(a - b) \int_{M_1 \cap M_2} \mathbf{v}_0 \cdot \mathbf{n}\ ds = 0$$

which implies by the choice of $\mathbf{v}_0$ that $a = b$. This is also in contradiction with (3.13), meaning that the third case never happens.

Therefore, p_m must be constant. Moreover, p_m is a zero function since it belongs to $L^2_0(\Omega)$. The analysis of the vertical version (Figure 3.2, (b)) is similar. Consequently, the bilinear form (3.9) verifies the hypothesis (3.2) of Theorem 3.1, so that the reduced local two-jump formulation (2.11), (3.10) has a unique solution.

□

It is worthwhile to note that any non-vanishing $p_m \in P_h$ satisfying (3.11) is called *a pressure mode*. Moreover, a pressure mode is said to be spurious if it is not constant.

3.2.2 Reduced local one-jump stabilizations

Using 2x2 macro-elements, the reduced local two-jump formulations can themselves be simplified to the so-called reduced local one-jump formulations given by the stabilization bilinear form:

$$C_h^{(1)}(p_h, q) = \beta \sum_{M=1}^{N_M} h_M \int_{\Gamma_M} [|p_h|]_{\Gamma_M}\ [|q|]_{\Gamma_M}\ ds, \tag{3.15}$$

where Γ_M can be one of the four inter-element boundaries interior to the M^{th} macro-element. That is, the number of jumps within each macro-element is only one. Likewise,

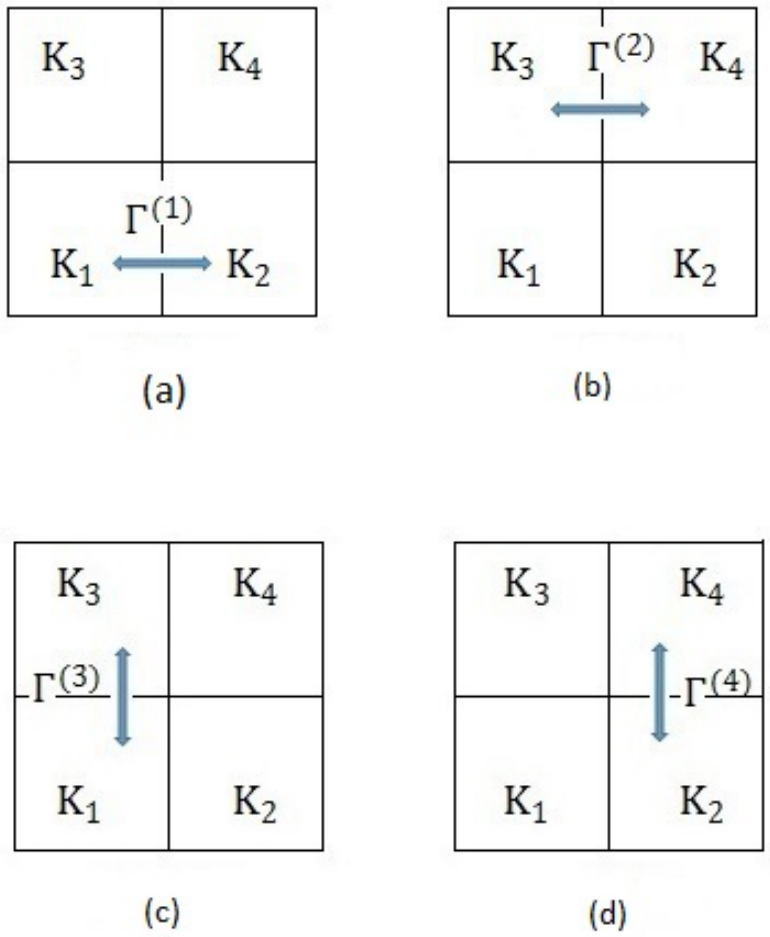

Figure 3.4: Four 2x2 rectangular macro-elements with one pressure jump.

there is no jump over the macro-element boundaries (see Figure 3.4). This leads to the following perturbed discrete incompressibility constraint:

$$-\int_\Omega q \; div \; \mathbf{u}_h \; d\Omega - C_h^{(1)}(p_h, q) = 0 \qquad \forall q \in P_h. \tag{3.16}$$

The well-posedness of the so-obtained discrete problem is given by the following result.

Theorem 3.3.
The reduced local one-jump formulation (2.11), (3.16) *has a unique solution* $(\boldsymbol{u}_h, p_h)$ *in* $\boldsymbol{V}_h \times P_h$.

First, it should be mentioned that the proof of this theorem cannot be directly established by similar arguments to those used in the proof of Theorem 3.2 since (3.14) is not assured for $C_h^{(1)}(p_m, p_m)$.
Indeed, if the number of jumps within each macro-element is reduced to one, then the

pressure jump could be considered only between the remaining elements K_3 and K_4 (Figure 3.4 (b)), where it is possible for $p_m|_{K_3}$ and $p_m|_{K_4}$ to be equal. In this case, $C_h^{(1)}(p_m, p_m)$ vanishes although $p_m|_{K_1} \neq p_m|_{K_2}$ because the pressure jump $[|p_m|]_{e_0}$ defined in (3.13) does not exist. In other words, the pressure values could be different in two elements where no pressure jump is considered. The latter could be considered just between the two other elements having the same pressure value and then the pressure jump vanishes. This does not ensure the strict positivity of $C_h^{(1)}(p_m, p_m)$.

There is another way for establishing the existence and uniqueness of the discrete solution using linear algebra arguments. To this end, let us postpone the proof of Theorem 3.3 and give some useful notations.

By introducing $\{\Phi_j\}_{j=1}^{n_u}$ and $\{\psi_k\}_{k=1}^{n_p}$, the two sets of basis functions for the discrete spaces $\mathbf{V}_h$ and P_h defined in (2.14) and (2.15) respectively, the discrete solutions $\mathbf{u}_h$ and p_h can be expressed as follows:

$$\mathbf{u}_h = \sum_{j=1}^{n_u} u_j \, \Phi_j, \quad p_h = \sum_{k=1}^{n_p} p_k \, \psi_k. \tag{3.17}$$

Now, by taking $\mathbf{v} = \Phi_i$ and $q = \psi_l$ ($i = 1, ..., n_u$ and $l = 1, ..., n_p$) in (2.11), (3.16), the stabilized algebraic problem that corresponds to the reduced local one-jump formulation (2.11), (3.16) becomes:

$$\begin{gathered} \text{Find } (\mathbf{U},\mathbf{P}) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \quad \text{such that} \\ \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\beta\, \mathbf{C}^{(1)} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{P} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix}, \end{gathered} \tag{3.18}$$

where $\mathbf{U}$ and $\mathbf{P}$ are the unknown vectors of n_u velocity nodal values and n_p pressure nodal values respectively.

The matrix $\mathbf{A} = [a_{ij}]$ with entries:

$$a_{ij} \quad = \alpha \int_\Omega \Phi_i \cdot \Phi_j \, d\Omega + \mu \int_\Omega \nabla\Phi_i \cdot \nabla\Phi_j \, d\Omega \tag{3.19}$$

is clearly symmetric and positive definite, and $\mathbf{B} = [b_{kj}]$ is the so-called divergence matrix with entries:

$$b_{kj} \quad = - \int_\Omega \psi_k \, div \, \Phi_j \, d\Omega \tag{3.20}$$

for $i, j = 1, ..., n_u$ and $k = 1, ..., n_p$.

Moreover, $\mathbf{C}^{(1)} = [c_{kl}]$ represents the stabilization matrix with entries:

$$c_{kl} \quad = - \sum_{M=1}^{N_M} h_M \int_{\Gamma_M} [|\psi_k|]_{\Gamma_M} \, [|\psi_l|]_{\Gamma_M} \, ds \tag{3.21}$$

for $k, l = 1, ..., n_p$, whereas the right-hand side vector $\mathbf{F}$ of length n_u is given by:

$$\mathbf{F} = [f_i], \quad f_i = \int_\Omega \mathbf{f} \cdot \Phi_i \, d\Omega \tag{3.22}$$

Consequently, the unique solvability of (3.18) leads to that of the discrete formulation (2.11), (3.16).

Proof. (of Theorem 3.3)
Following Elman, Silvester and Wathen [14], the unique solvability of the matrix system (3.18) can be determined by studying the corresponding homogeneous system:

$$\begin{cases} \mathbf{A}\mathbf{U} + \mathbf{B}^T\mathbf{P} = \mathbf{0} \\ \\ \mathbf{B}\mathbf{U} - \beta\, \mathbf{C}^{(1)}\mathbf{P} = \mathbf{0}. \end{cases} \tag{3.23}$$

By pre-multiplying the first equation of (3.23) by $\mathbf{U}^T$ and the second equation by $\mathbf{P}^T$ then substituting $\mathbf{P}^T\mathbf{B}\mathbf{U}$ in the first equation gives

$$\mathbf{U}^T\mathbf{A}\mathbf{U} + \beta\, \mathbf{P}^T\mathbf{C}^{(1)}\mathbf{P} = 0. \tag{3.24}$$

Thus, $\mathbf{U} = \mathbf{0}$ since $\mathbf{A}$ is positive definite, $\mathbf{C}^{(1)}$ is positive semi-definite and $\beta > 0$. This implies the unique solvability with respect to the velocity.

Unique solvability with respect to the pressure is more delicate. Substituting $\mathbf{U} = \mathbf{0}$ into (3.23) gives

$$\begin{cases} \mathbf{B}^T\mathbf{P} = \mathbf{0}, \\ \\ \mathbf{C}^{(1)}\mathbf{P} = \mathbf{0}, \end{cases} \tag{3.25}$$

implying that any pressure solution of (3.23) is only unique up to the null space of the matrices $\mathbf{B}^T$ and $\mathbf{C}^{(1)}$. Again, pre-multiplying the first equation of (3.25) by $\mathbf{V}^T$ and the second equation by $\mathbf{P}^T$ produces the algebraic corresponding condition of Theorem 3.1:

$$\begin{cases} \mathbf{V}^T\mathbf{B}^T\mathbf{P} = \mathbf{0} \qquad \forall\, \mathbf{V}^T \in \mathbb{R}^{n_u} \\ \\ \mathbf{P}^T\mathbf{C}^{(1)}\mathbf{P} = \mathbf{0}. \end{cases} \tag{3.26}$$

Thus, in order to get an algebraic equivalent result to the criterion (3.2), it is sufficient to show that any solution $\mathbf{P}$ of (3.25) must vanish. Moreover, the first equation of (3.23) gives $\mathbf{U} = -\mathbf{A}^{-1}\mathbf{B}^T\mathbf{P}$. By substituting in the second equation of (3.23), we get

$$\mathbf{S}_\beta^{(1)}\mathbf{P} = \mathbf{0}, \tag{3.27}$$

where

$$\mathbf{S}_\beta^{(1)} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T + \beta\, \mathbf{C}^{(1)}. \tag{3.28}$$

The equation (3.27) represents the homogeneous system associated to

$$\mathbf{S}_{\beta}^{(1)}\,\mathbf{P} = \mathbf{B}\mathbf{A}^{-1}\mathbf{F}, \tag{3.29}$$

that the pressure solution $\mathbf{P}$ of (3.18) satisfies. The matrix $\mathbf{S}_{\beta}^{(1)}$ is symmetric positive semi-definite (a property which will be proved in the fourth chapter); it is called "stabilized pressure Schur complement". Therefore, it is sufficient to show that any solution $\mathbf{P}$ of (3.27) must be constant. In particular, such $\mathbf{P}$ must be zero vector, i.e. the nullspace of the Schur complement matrix should be reduced to the zero vector. In other words, the eigenvectors corresponding to the eigenvalue $\lambda = 0$ of the matrix $\mathbf{S}_{\beta}^{(1)}$ must have zero components.

On the other hand, it is clear that a well-posed local discrete problem defined on each single macro-element ensures well-posedness of the discrete problem on the whole domain Ω since standard macro-element theory developed by Boland and Nicolaïdes [3], and Stenberg [32] can then be used. To this end, the stabilized pressure matrix $\mathbf{S}_{\beta}^{(1)}$ will be considered locally. Consider the 2x2 rectangular macro-element illustrated in Figure 3.4 and set $\mathbf{P} = [p_i]_{i=1}^{4}$. The local matrix $\mathbf{A}$ is the 2x2 diagonal matrix defined from (3.19) with entries

$$a_{11} = a_{22} = d = \frac{4\alpha}{9}\,h_x h_y + \frac{4\mu}{3}\,\left(\frac{h_x}{h_y} + \frac{h_y}{h_x}\right). \tag{3.30}$$

Construction of the local divergence matrix yields

$$\mathbf{B}^T = \tfrac{1}{2}\begin{bmatrix} -h_y & h_y & h_y & -h_y \\ -h_x & -h_x & h_x & h_x \end{bmatrix}, \tag{3.31}$$

where h_x and h_y are element dimensions. Next, adopting, for instance, the stabilization bilinear form (3.15) for the local one-jump case in Figure 3.4 (a), gives the local one-jump stabilization matrix:

$$\mathbf{C}^{(1)} = h_y^2\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.32}$$

Computing the stabilized pressure matrix $\mathbf{S}_{\beta}$ from (3.28) produces

$$\mathbf{S}_{\beta}^{(1)} = \tfrac{1}{4d}\begin{bmatrix} h_x^2+h_y^2+4\beta d h_y^2 & h_x^2-h_y^2-4\beta d h_y^2 & -h_x^2-h_y^2 & -h_x^2+h_y^2 \\ h_x^2-h_y^2-4\beta d h_y^2 & h_x^2+h_y^2+4\beta d h_y^2 & -h_x^2+h_y^2 & -h_x^2-h_y^2 \\ -h_x^2-h_y^2 & -h_x^2+h_y^2 & h_x^2+h_y^2 & h_x^2-h_y^2 \\ -h_x^2+h_y^2 & -h_x^2-h_y^2 & h_x^2-h_y^2 & h_x^2+h_y^2 \end{bmatrix}. \tag{3.33}$$

It can be seen that the eigenvalues of $\mathbf{S}_\beta$ and their associated eigenvectors are

$$\left\{ 0, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{q_1} \right\}, \qquad \left\{ \frac{h_x^2}{d}, \underbrace{\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}}_{q_2} \right\},$$

$$\left\{ \frac{h_y^2}{2d}\left(2d\beta + 1 - \sqrt{4d^2\beta^2+1}\right), \underbrace{\begin{bmatrix} 1 \\ -1 \\ 2d\beta + \sqrt{4d^2\beta^2+1} \\ -2d\beta - \sqrt{4d^2\beta^2+1} \end{bmatrix}}_{q_3} \right\}, \tag{3.34}$$

$$\left\{ \frac{h_y^2}{2d}\left(2d\beta + 1 + \sqrt{4d^2\beta^2+1}\right), \underbrace{\begin{bmatrix} 1 \\ -1 \\ 2d\beta - \sqrt{4d^2\beta^2+1} \\ -2d\beta + \sqrt{4d^2\beta^2+1} \end{bmatrix}}_{q_4} \right\}.$$

This shows that for all $\beta > 0$ there is only one zero eigenvalue and its corresponding eigenvectors are constant pressures as it is required. Moreover, the fact that $\mathbf{p}$ with zero mean over M leads to $\mathbf{p} = \mathbf{0}$, which ensures the unique solvability of the discrete pressure. Similar arguments can be used to establish the same result for the three other one-jump cases in Figure 3.4. □

It is important to note that the local stabilization matrices analogous to (3.32) that correspond to these reduced local one-jump stabilizations are respectively:

$$h_y^2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \; h_x^2 \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \; h_x^2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}. \tag{3.35}$$

Remark 3.1.
Keeping in mind the local numbering of Figure 3.4, the local stabilization matrix corresponding to the local jump stabilization term (3.6) *is of the form:*

$$\begin{bmatrix} h_x^2+h_y^2 & -h_y^2 & -h_x^2 & 0 \\ -h_y^2 & h_x^2+h_y^2 & 0 & -h_x^2 \\ -h_x^2 & 0 & h_x^2+h_y^2 & -h_y^2 \\ 0 & -h_x^2 & -h_y^2 & h_x^2+h_y^2 \end{bmatrix}, \tag{3.36}$$

whereas that of the reduced local two-jump stabilization term (3.9) *is given by:*

$$h_y^2 \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad or \quad h_x^2 \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \tag{3.37}$$

depending on the horizontal or vertical choice.

Remark 3.2.
The choice of the stabilization parameter β requires a balance between stability and accuracy. For stability, β should not be too small since the spurious pressure mode has to be removed. On the other hand, to keep accuracy, β should not be large because the right-hand side of (3.29) *is a linear combination of the three eigenvectors q_2, q_3 and q_4. Further, since the solution $\boldsymbol{P}$ is defined by multiplying each of these components by the inverse of the associated eigenvalues, the eigenvector q_4 will not figure in the solution (see Elman, Silvester and Wathen [14]).*

Remark 3.3.
It should be mentioned that the idea of reducing the number of interior jumps in macro-elements can be naturally extended to three-dimensional Q1-Q0 brick elements.

Now, the convergence of the approximate solutions $(\mathbf{u}_h, p_h)$ towards the exact solution when the mesh parameter h decreases can be discussed in the following section, since having proved the existence and uniqueness of $(\mathbf{u}_h, p_h)$ gives no idea about this significant topic.

3.3 Stability and convergence theory

In this section, the stability framework of Babuška [2] is used and then error estimates in the natural norms will be derived. First, let $B(.,.)$ be the generalized bilinear form

$$B\left((\mathbf{u},p);(\mathbf{v},q)\right)=\alpha\left(\mathbf{u},\mathbf{v}\right)+\mu\left(\nabla\mathbf{u},\nabla\mathbf{v}\right)-\left(p,div\ \mathbf{v}\right)-\left(q,div\ \mathbf{u}\right) \tag{3.38}$$

and $L(.,.)$ be the linear functional

$$L\left(\mathbf{v},q\right)=\left(\mathbf{f},\mathbf{v}\right) \tag{3.39}$$

where $(.,.)$ denotes the usual L^2-inner product defined by (1.1). Thus, the problem (2.6) can be equivalently stated as:

$$\begin{aligned}&\text{Find } (\mathbf{u},p)\in\mathbf{V}\times P \text{ such that :}\\ &B\left((\mathbf{u},p);(\mathbf{v},q)\right)=L\left(\mathbf{v},q\right)\qquad\forall\left(\mathbf{v},q\right)\in\mathbf{V}\times P.\end{aligned} \tag{3.40}$$

Next, by defining

$$\begin{cases} B_h\left((\mathbf{u}_h,p_h);(\mathbf{v},q)\right) &= \alpha\left(\mathbf{u}_h,\mathbf{v}\right)+\mu\left(\nabla\mathbf{u}_h,\nabla\mathbf{v}\right)-\left(p_h,div\ \mathbf{v}\right)\\ &\quad -\left(q,div\,\mathbf{u}_h\right)\\ L_h\left(\mathbf{v},q\right) &= \left(\mathbf{f},\mathbf{v}\right)\qquad\forall\left(\mathbf{v},q\right)\in\mathbf{V}_h\times P_h,\end{cases} \tag{3.41}$$

and by introducing the bilinear form

$$C_h^{(t)}\left(r,q\right)=\begin{cases} C_h^{(2)}\left(r,q\right) & \text{for the local 2-jump}\\ C_h^{(1)}\left(r,q\right) & \text{for the local 1-jump}\end{cases}\qquad\forall r,q\in P_h, \tag{3.42}$$

where $C_h^{(2)}\left(.,.\right)$ and $C_h^{(1)}\left(.,.\right)$ are defined in (3.9) and (3.15) respectively, the reduced locally stabilized formulations (2.11), (3.10) and (2.11), (3.16) become:

$$\begin{aligned}&\text{Find } (\mathbf{u}_h,p_h)\in\mathbf{V}_h\times P_h \text{ such that :}\\ &B_h^{(t)}\left((\mathbf{u}_h,p_h);(\mathbf{v},q)\right)=L_h\left(\mathbf{v},q\right)\qquad\forall\left(\mathbf{v},q\right)\in\mathbf{V}_h\times P_h\end{aligned} \tag{3.43}$$

where

$$B_h^{(t)}\left((\mathbf{u}_h,p_h);(\mathbf{v},q)\right)=\quad B_h\left((\mathbf{u}_h,p_h);(\mathbf{v},q)\right)-\beta\,C_h^{(t)}\left(p_h,q\right). \tag{3.44}$$

We also need to introduce some subspaces of the discrete pressure space P_h. Set:

$$R_h = \begin{cases} \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(1)}} = [|q|]_{\Gamma_M^{(2)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 2j-H scheme} \\ \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(3)}} = [|q|]_{\Gamma_M^{(4)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 2j-V scheme} \\ \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(1)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 1j-H1 scheme} \\ \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(2)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 1j-H2 scheme} \\ \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(3)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 1j-V1 scheme} \\ \left\{ q \in P_h;\ [|q|]_{\Gamma_M^{(4)}} = 0 \quad \forall M \in \mathcal{M}_h \right\} & \text{for the 1j-V2 scheme.} \end{cases} \tag{3.45}$$

Now, let $C_M^{(t)}(.,.)$ be the restriction of $C_h^{(t)}(.,.)$ to a macro-element M given by

$$C_M^{(t)}(r,q) = \begin{cases} C_M^{(2)}(r,q) \\ \text{or} \qquad \forall r, q \in P_h \\ C_M^{(1)}(r,q) \end{cases} \tag{3.46}$$

Here, $C_M^{(2)}(.,.)$ and $C_M^{(1)}(.,.)$ are the restrictions to a macro-element M of $C_h^{(2)}(.,.)$ and $C_h^{(1)}(.,.)$ respectively, i.e.

$$C_M^{(2)}(r,q) = \begin{cases} \sum_{i=1}^{2} h_M^{(i)} \int_{\Gamma_M^{(i)}} [|r|]_{\Gamma_M^{(i)}} [|q|]_{\Gamma_M^{(i)}} ds \\ \text{or} \\ \sum_{i=3}^{4} h_M^{(i)} \int_{\Gamma_M^{(i)}} [|r|]_{\Gamma_M^{(i)}} [|q|]_{\Gamma_M^{(i)}} ds \end{cases} \tag{3.47}$$

and

$$C_M^{(1)}(r,q) = \beta\, h_M \int_{\Gamma_M} [|r|]_{\Gamma_M} [|q|]_{\Gamma_M} ds. \tag{3.48}$$

For $t = 1, 2$, we obviously have

$$C_h^{(t)}(r,q) = \sum_{M=1}^{N_M} C_M^{(t)}(r,q) \quad \forall r, q \in P_h. \tag{3.49}$$

Moreover, the restricted pressure spaces for a macro-element M are given by

$$\begin{cases} P_{0,M} = \{ q \in L_0^2(M);\, q|_K \text{ is constant } \forall K \subset M \} \\ R_M = L_0^2(M) \cap R_h, \end{cases} \tag{3.50}$$

accordingly to (3.45).

The stability of the local (two-jump or one-jump) stabilized formulations (3.43) can be analyzed by establishing three interesting lemmas. The first one gives the macro-element positivity of the stabilization bilinear form $C_h^{(t)}(.,.)$.

Lemma 3.1.
Let $\zeta_{\widehat{M}}$ be a class of macro-elements. Then, there exists $\gamma_{\widehat{M}} > 0$ such that

$$C_M^{(t)}(q,q) \geq \gamma_{\widehat{M}} \|q\|_{0,M}^2 \quad \forall q \in P_{0,M} \setminus R_M \tag{3.51}$$

for $t = 1, 2$.

Proof.
Let $M \in \zeta_{\widehat{M}}$ and $q \in P_{0,M} \setminus R_M$. From the definition (3.47) or (3.48) of $C_M^{(t)}$, we note that $C_M^{(t)}(q,q) = 0$ if and only if $q \in R_M$. Hence, the constant γ_M defined through

$$\gamma_M = \inf_{\substack{\mathrm{q} \in P_{0,M} \setminus R_M \\ \|q\|_{0.M} = 1}} C_M^{(t)}(q,q)$$

is positive. By virtue of a scaling argument (cf. [32]), the regularity assumptions (2.13) guarantee the existence of a constant $\gamma_{\widehat{M}}$ such that

$$\gamma_M \geq \gamma_{\widehat{M}} > 0 \quad \forall M \in \zeta_{\widehat{M}}$$

which is equivalent to (3.51). Indeed, let $M \in \zeta_{\widehat{M}}$. We denote by $\{(\hat{x}_i, \hat{y}_i)\}_{i=1}^9$ the vertices of the reference quadrilaterals in $\widehat{M}$. The macroelement M is hence uniquely defined by its vertices $\{(x_i, y_i) = F(\hat{x}_i, \hat{y}_i)\}_{i=1}^9$ where F is an isoparametric transformation defined as in (2.19) (see Figure 3.5).
Using the first inequality in (2.13) leads to the existence of a certain positive constant ν such that

$$h_e^2 \geq \nu k_M$$

for all interelement boundary e, strictly in the interior of M. Now, using (3.5) gives the desired inequality (3.51) (see [22], Lemma 3.1.).

□

Next, let us assume that there is a fixed set of classes $\zeta_{\widehat{M}_i}$, $i = 1, 2, ..., n$ $(n \geq 1)$ such that every macro-element M belongs to one of the equivalence classes. Define Λ_h to be the orthogonal L^2-projection from P_h onto its subspace R_h defined locally by

$$\int_M (q - \Lambda_h q)\, r\, dx = 0 \qquad \forall q \in P_h, \forall r \in R_h, \forall M \in \mathcal{M}_h. \tag{3.52}$$

As a direct consequence of the last lemma is the following global positivity of the form $C_h^{(t)}$

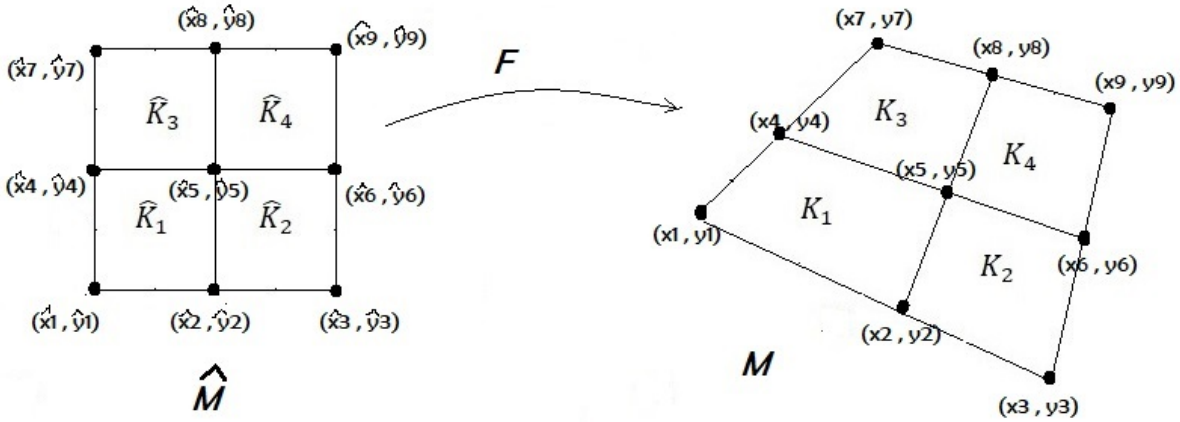

Figure 3.5: Macroelement isoparametric transformation

Lemma 3.2.
There exists $\alpha_1 > 0$ *independent of* h *such that*

$$C_h^{(t)}(q,q) \geq \alpha_1 \| (I - \Lambda_h) q \|_0^2 \quad \forall q \in P_h, \tag{3.53}$$

for $t = 1, 2$.

Proof.
Let $M \in \zeta_{\widehat{M}_i}$. We have

$$C_M^{(t)}(q,q) = C_M^{(t)}((I - \Lambda_h) q, (I - \Lambda_h) q) \quad \forall q \in P_h, \tag{3.54}$$

because the jump $[|\Lambda_h q|]$ vanishes within M since $\Lambda_h q \in R_h$. Moreover,

$$((I - \Lambda_h) q, r) = 0 \quad \forall r \in R_h, \tag{3.55}$$

since $(I - \Lambda_h) q$ is orthogonal to R_h.
By taking $r = 1_M$, it follows that $(I - \Lambda_h) q|_M \in P_{0,M} \backslash R_M$. Hence, by virtue of Lemma 3.1 and using (3.51), we get

$$C_M^{(t)}(q,q) \geq \gamma_{\widehat{M}_i} \| (I - \Lambda_h) q \|_{0,M}^2.$$

Then, the global inequality (3.53) holds for

$$\alpha_1 = \min \left\{ \gamma_{\widehat{M}_i};\ i = 1, 2, ..., n \right\}$$

which is clearly independent of h. □

The third lemma is a global stability result for the discrete pressure space P_h.

Lemma 3.3.
There exists a positive constant α_2 independent of h such that for every $q \in P_h$ there is a $\boldsymbol{g}_h \in \boldsymbol{V}_h$ satisfying

$$(\Lambda_h q, div\ \boldsymbol{g}_h\ ) = \|\Lambda_h q\|_0^2 \quad and \quad \|\boldsymbol{g}_h\|_1 \leq \alpha_2\ \|\Lambda_h q\|_0. \tag{3.56}$$

Proof.

Let $q \in P_h$ be arbitrary. Since $\Lambda_h q \in R_h \subset L_0^2(\Omega)$ there exist $C_1 \geq 0$ and $\mathbf{g} \in \mathbf{V}$ such that (cf. [16])

$$div\ \mathbf{g} = \Lambda_h q \qquad \text{and} \qquad \|\mathbf{g}\|_1 \leq C_1 \|\Lambda_h q\|_0$$

We can now combine some ideas from [11] and [12] with the macro-element methodology of [32] in order to construct an operator $I_h : \mathbf{V} \longrightarrow \mathbf{V}_h$ such that

$$(div\ I_h \mathbf{g}, r) = (div \mathbf{g}, r)\ \forall r \in P_h \quad and \quad \|I_h \mathbf{g}\|_1 \leq C_2 \|\mathbf{g}\|_1.$$

It remains to take $\mathbf{g}_h = I_h \mathbf{g}$ and $\alpha_2 = C_1 C_2$. □

Now, we are in a position to state the following main *inf-sup* result which also confirms the well-posedness of the discrete formulation (3.43).

Theorem 3.4.
There exists a constant $\gamma > 0$ independent of the mesh parameter h such that

$$\sup_{(\boldsymbol{w},r) \in \boldsymbol{V}_h \times P_h} \frac{B_h^{(t)}\left((\boldsymbol{v},q)\,;\,(\boldsymbol{w},r)\right)}{\|\boldsymbol{w}\|_1 \|r\|_0} \geq \gamma\left(\|\boldsymbol{v}\|_1 + \|q\|_0\right)\ \forall\,(\boldsymbol{v},q) \in \boldsymbol{V}_h \times P_h. \tag{3.57}$$

Proof.

Let $(\mathbf{v},q) \in \mathbf{V}_h \times P_h$ and α_1 be as in (3.53). Set $r = -q$, $\mathbf{w} = \mathbf{v} - \delta\ \mathbf{g}_h$ where δ is a positive constant to be determined below. From (3.53) and (3.56), it follows that

$$\begin{aligned}
B_h^{(t)}\left((\mathbf{v},q)\,;\,(\mathbf{w},r)\right) = \quad & \alpha\,(\mathbf{v},\mathbf{v}) - \delta\alpha\,(\mathbf{v},\mathbf{g}_h) + \mu\,(\nabla\mathbf{v},\nabla\mathbf{v}) - \delta\mu\,(\nabla\mathbf{v},\nabla\mathbf{g}_h) \\
& + \delta\,(q, div\ \mathbf{g}_h) + \beta\, C_h^{(t)}(q,q) \\
= \quad & \alpha\|\mathbf{v}\|_0^2 + \mu\|\nabla\mathbf{v}\|_0^2 - \delta\alpha\,(\mathbf{v},\mathbf{g}_h) - \delta\mu\,(\nabla\mathbf{v},\nabla\mathbf{g}_h) \\
& + \delta\,(\Lambda_h q, div\ \mathbf{g}_h) - \delta\,((\Lambda_h - I)\,q, div\ \mathbf{g}_h) + \beta C_h^{(t)}(q,q) \\
\geq \quad & \alpha\|\mathbf{v}\|_0^2 + \mu\|\nabla\mathbf{v}\|_0^2 - \delta\ \alpha\ \alpha_2\|\mathbf{v}\|_0\ \|\Lambda_h q\|_0 - \delta\mu\alpha_2\|\nabla\mathbf{v}\|_0\ \|\Lambda_h q\|_0 \\
& + \delta\ \|\Lambda_h q\|_0^2 - \delta\ \alpha_2\|\,(\Lambda_h - I)\,q\|_0\ \|\Lambda_h q\|_0 + \beta\ \alpha_1\|\,(\Lambda_h - I)\,q\|_0^2 \\
\geq \quad & \alpha\|\mathbf{v}\|_0^2 + \mu\|\nabla\mathbf{v}\|_0^2 - \frac{\alpha}{2}\left(\|\mathbf{v}\|_0^2 + \delta^2\ \alpha_2^2\|\Lambda_h q\|_0^2\right) + \delta\ \|\Lambda_h q\|_0^2 \\
& - \frac{\mu}{2}\left(\|\nabla\mathbf{v}\|_0^2 + \delta^2\ \alpha_2^2\|\Lambda_h q\|_0^2\right) + \beta\ \alpha_1\|\,(\Lambda_h - I)\,q\|_0^2 \\
& - \frac{\beta\ \alpha_1}{2}\left(\|\,(\Lambda_h - I)\,q\|_0^2 + \left(\frac{\delta\ \alpha_2}{\beta\ \alpha_1}\right)^2 \|\Lambda_h q\|_0^2\right).
\end{aligned}$$

By choosing $\delta = \dfrac{1}{\alpha_2^2}\left(\dfrac{1}{\beta\,\alpha_1} + \alpha + \mu\right)^{-1}$, we get

$$B_h^{(t)}\left((\mathbf{v},q)\,;\,(\mathbf{w},r)\right) \geq \frac{\alpha}{2}\|\mathbf{v}\|_0^2 + \frac{\mu}{2}\|\nabla\mathbf{v}\|_0^2 + \frac{\beta\alpha_1}{2}\|\left(\Lambda_h - I\right)q\|_0^2 + \frac{\delta}{2}\|\Lambda_h q\|_0^2,$$

which implies that

$$B_h^{(t)}\left((\mathbf{v},q)\,;\,(\mathbf{w},r)\right) \geq k_1\left(\|\mathbf{v}\|_0 + \|q\|_0\right)^2, \tag{3.58}$$

where $k_1 = \frac{1}{4}\min\left\{\alpha,\mu,\beta\alpha_1,\delta\right\}$.
On the other hand, we have

$$\begin{aligned}\|\mathbf{w}\|_1 + \|r\|_0 &= \|\mathbf{v} - \delta\ \mathbf{g}_h\|_1 + \|q\|_0\\ &\leq \|\mathbf{v}\|_1 + \delta\|\mathbf{g}_h\|_1 + \|q\|_0\\ &\leq \|\mathbf{v}\|_1 + \delta\alpha_2\|\Lambda_h q\|_0 + \|q\|_0\end{aligned}$$

Hence,

$$\|\mathbf{w}\|_1 + \|r\|_0 \leq k_2\left(\|\mathbf{v}\|_1 + \|q\|_0\right), \tag{3.59}$$

where $k_2 = 1 + \delta\alpha_2$.
Finally, combining (3.58) and (3.59) establishes the desired inequality (3.57) with $\gamma = \dfrac{k_1}{k_2}$.

□

Next, let us state two important and required interpolation results (cf. [16]).

Lemma 3.4.
If $\boldsymbol{v} \in \left(H^2(\Omega)\cap H_0^1(\Omega)\right)^2$, *then there exists* $\tilde{\boldsymbol{v}} \in \boldsymbol{V}_h$ *such that*

$$\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|_1 \leq C_1 h|\boldsymbol{v}|_2, \tag{3.60}$$

where C_1 *is a constant indepentent of* h.

Lemma 3.5.
If $q \in H^1(\Omega)\cap L_0^2(\Omega)$, *then there exists* $\tilde{q} \in R_h$ *such that*

$$\|q - \tilde{q}\|_1 \leq C_2 h|q|_1, \tag{3.61}$$

where C_2 *is a constant indepentent of* h.

The convergence of the proposed stabilization schemes is given by the optimal error estimates shown in the following theorem.

Theorem 3.5.
Suppose that the solution of (3.40) *satisfies* $\boldsymbol{u} \in \left(H^2(\Omega)\right)^2$ *and* $p \in H^1(\Omega)$. *Then, there exists a constant* C *independent of* h *such that*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_1 + \|p - p_h\|_0 \leq Ch\left(|\boldsymbol{u}|_2 + |p|_1\right), \tag{3.62}$$

where $(\boldsymbol{u}_h, p_h)$ *is the solution of* (3.43).

Proof.
Applying Lemma 3.4 and Lemma 3.5 with $\mathbf{v} = \mathbf{u}$ and $q = p$ respectively, there exist $\tilde{\mathbf{u}} \in \mathbf{V}_h$ and $\tilde{p} \in R_h$ such that (3.60) and (3.61) hold. Then,

$$\begin{aligned} B_h^{(t)}\left((\mathbf{u}_h - \tilde{\mathbf{u}}, p_h - \tilde{p})\,;\,(\mathbf{v}, q)\right) &= B_h^{(t)}\left((\mathbf{u}_h, p_h)\,;\,(\mathbf{v}, q)\right) - B_h^{(t)}\left((\tilde{\mathbf{u}}, \tilde{p})\,;\,(\mathbf{v}, q)\right) \\ &= B_h^{(t)}\left((\mathbf{u}_h, p_h)\,;\,(\mathbf{v}, q)\right) - B\left((\tilde{\mathbf{u}}, \tilde{p})\,;\,(\mathbf{v}, q)\right), \end{aligned} \tag{3.63}$$

since $\tilde{p} \in R_h$ so that $C_h^{(t)}(\tilde{p}, q) = 0$. Moreover, we have

$$B_h^{(t)}\left((\mathbf{u}_h, p_h)\,;\,(\mathbf{v}, q)\right) = L_h(\mathbf{v}, q) = B\left((\mathbf{u}, p)\,;\,(\mathbf{v}, q)\right) \tag{3.64}$$

because $L_h(\mathbf{v}, q) = L(\mathbf{v}, q)$ for all $(\mathbf{v}, q) \in \mathbf{V}_h \times P_h$. Next, (3.63)) yields

$$\begin{aligned} B_h^{(t)}\left((\mathbf{u}_h - \tilde{\mathbf{u}}, p_h - \tilde{p})\,;\,(\mathbf{v}, q)\right) &= B\left((\mathbf{u} - \tilde{\mathbf{u}}, p - \tilde{p})\,;\,(\mathbf{v}, q)\right) \\ &\leq \alpha\|\mathbf{u} - \tilde{\mathbf{u}}\|_1\|\mathbf{v}\|_1 + \mu\|\mathbf{u} - \tilde{\mathbf{u}}\|_1\|\mathbf{v}\|_1 \\ &\quad + \|\mathbf{u} - \tilde{\mathbf{u}}\|_1\|q\|_0 + \|p - \tilde{p}\|_0\|\mathbf{v}\|_1 \\ &\leq C_3\left(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + \|p - \tilde{p}\|_0\right)\left(\|\mathbf{v}\|_1 + \|q\|_0\right) \end{aligned}$$

for all $(\mathbf{v}, q) \in \mathbf{V}_h \times P_h$, where $C_3 = \max\{\alpha, \mu, 1\}$. Thus,

$$\sup_{(\mathbf{v},q)\in\mathbf{V}_h\times P_h} \frac{B\left((\mathbf{u} - \tilde{\mathbf{u}}, p - \tilde{p})\,;(\mathbf{v}, q)\right)}{\|\mathbf{v}\|_1 + \|q\|_0} \leq C_3\left(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + \|p - \tilde{p}\|_0\right) \tag{3.65}$$

so that (3.64) yields

$$\sup_{(\mathbf{v},q)\in\mathbf{V}_h\times P_h} \frac{B_h^{(t)}\left((\mathbf{u}_h - \tilde{\mathbf{u}}, p_h - \tilde{p})\,;(\mathbf{v}, q)\right)}{\|\mathbf{v}\|_1 + \|q\|_0} \leq C_3\left(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + \|p - \tilde{p}\|_0\right). \tag{3.66}$$

From (3.57), it follows that

$$\|\mathbf{u}_h - \tilde{\mathbf{u}}\|_1 + \|p_h - \tilde{p}\|_0 \leq \frac{C_3}{\gamma}\left(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + \|p - \tilde{p}\|_0\right).$$

Therefore,

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0 \leq \left(1 + \frac{C_3}{\gamma}\right)\left(\|\mathbf{u} - \tilde{\mathbf{u}}\|_1 + \|p - \tilde{p}\|_0\right).$$

Consequently, applying (3.60) and (3.61) gives the desired inequality (3.62), where the constant C is given by $C = \max\{C_1, C_2\}\left(1 + \frac{C_3}{\gamma}\right)$. □

Remark 3.4.
For the Stokes problem, the previous results hold by taking $\alpha = 0$ in the proof.

Remark 3.5.
It is instructive to note that substituting for the value of γ given in the proof of Theorem 3.4, the constant C takes the form

$$C = \max\{C_1, C_2\} \left(1 + 4\max\{\alpha, \mu, 1\} \frac{\left(1 + \dfrac{1}{\alpha_2 \left(\alpha + \mu + \dfrac{1}{\beta\alpha_1} \right)} \right)}{\min\left\{ \alpha, \mu, \beta\alpha_1, \dfrac{1}{\alpha_2^2 \left(\alpha + \mu + \dfrac{1}{\beta\alpha_1} \right)} \right\}} \right). \quad (3.67)$$

This shows, to some extent, that the convergence can be influenced by the values of α and μ. If α is very large and/or μ is too small then this can compromise the approximation accuracy especially for practical values of h .

Remark 3.6.
According to the inequality (3.62)*, the convergence of the discrete solution $(\boldsymbol{u}_h, p_h)$ given by* (3.43) *to $(\boldsymbol{u}, p)$ is then guaranteed since*

$$\lim_{h \to 0} \left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_1 + \|p - p_h\|_0 \right) = 0.$$

In Theorem 3.5, it is shown that $\|\mathbf{u} - \mathbf{u}_h\|_1 = \mathcal{O}(h)$, so that the velocity L^2-norm, i.e.$\|\mathbf{u} - \mathbf{u}_h\|_0$, is at least of the same order. To derive an optimal L^2-estimate, we need to use the corresponding regularity condition:

$$\|\mathbf{u}\|_2 + \|p\|_1 \ \leq \epsilon \|\mathbf{f}\|_0, \quad (3.68)$$

with $\epsilon > 0$ for all $\mathbf{f} \in [L^2(\Omega)]^2$.

Corollary 3.1.
There exists a certain constant $\theta\,(> 0)$, independent of h, such that

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_0 \leq \theta h^2 \left(|\boldsymbol{u}|_2 + |p|_1 \right). \quad (3.69)$$

Proof.
Following [22], let $(\mathbf{z}, r)$ be the solution of the dual problem

$$\begin{aligned} \alpha\mathbf{z} - \mu\Delta\mathbf{z} + \nabla r &= \mathbf{u} - \mathbf{u}_h \quad \text{in } \Omega, \\ div\ \mathbf{z} &= 0 \qquad\ \ \text{in } \Omega, \end{aligned} \quad (3.70)$$

satisfying the condition (3.68). Thus,

$$\|\mathbf{z}\|_2 + \|r\|_1 \ \leq \epsilon \|\mathbf{u} - \mathbf{u}_h\|_0, \quad (3.71)$$

By considering $\tilde{\mathbf{z}}$ and $\tilde{r}$ the interpolants of $\mathbf{z}$ and r in $\mathbf{V}_h$ and R_h respectively (see Lemma 3.4 and Lemma 3.5) and by taking $\mathbf{u} - \mathbf{u}_h$ as a test function in (3.70), we get

$$B\left(\ (\mathbf{u} - \mathbf{u}_h, r - r_h)\,;(\mathbf{z} - \tilde{\mathbf{z}}, r - \tilde{r})\right) = \|\mathbf{u} - \mathbf{u}_h\|_0^2\,, \tag{3.72}$$

using the Galerkin orthogonality property.

On the other hand, we have

$$\begin{aligned} B\left(\ (\mathbf{u} - \mathbf{u}_h, p - p_h)\,;(\mathbf{z} - \tilde{\mathbf{z}}, r - \tilde{r})\right) \leq \quad & (\alpha + \mu)\,\|\mathbf{u} - \mathbf{u}_h\|_1\ \|\mathbf{z} - \tilde{\mathbf{z}}\|_1 \\ & + \|p - p_h\|_0 \|\mathbf{z} - \tilde{\mathbf{z}}\|_1 \\ & + \|\mathbf{u} - \mathbf{u}_h\|_1 \|r - \tilde{r}\|_1. \end{aligned}$$

Using (3.60) and (3.61), we get

$$\begin{aligned} B\left(\ (\mathbf{u} - \mathbf{u}_h, p - p_h)\,;(\mathbf{z} - \tilde{\mathbf{z}}, r - \tilde{r})\right) \leq \quad & (\alpha + \mu)\, C_1\, h\, |\mathbf{z}|_2 \|\mathbf{u} - \mathbf{u}_h\|_1 \\ & + C_1\, h\, |\mathbf{z}|_2 \|p - p_h\|_0 \\ & + C_2\, h\, |r|_1 \|\mathbf{u} - \mathbf{u}_h\|_1, \end{aligned}$$

Therefore,

$$\begin{aligned} B\left(\ (\mathbf{u} - \mathbf{u}_h, p - p_h)\,;(\mathbf{z} - \tilde{\mathbf{z}}, r - \tilde{r})\right) \leq \quad & (\alpha + \mu)\, C_1\, h\, |\mathbf{z}|_2 \|\mathbf{u} - \mathbf{u}_h\|_1 \\ & + C_1\, h\, |\mathbf{z}|_2 \|p - p_h\|_0 \\ & + C_2\, h\, |r|_1 \|\mathbf{u} - \mathbf{u}_h\|_1, \\ & + h |r|_1 \|p - p_h\|_0 \\ \leq \quad & Ch \|\mathbf{u} - \mathbf{u}_h\|_0 \left(\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0\right), \end{aligned}$$

using (3.71) where $C > 0$ is some constant independent of h. Now, it is sufficient to combine this later inequality with (3.72) to get

$$\|\mathbf{u} - \mathbf{u}_h\|_0 \leq Ch \left(\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_0\right). \tag{3.73}$$

Using (3.62), the inequality (3.69) follows. □

Chapter 4

Iterative methods

The numerical solution of the generalized Stokes problem discretized via the low-order mixed finite element method *Q1-Q0* leads to large and sparse systems of algebraic linear equations. The sparsity is an important feature for iterative methods since the storage requirements depend mainly on the nonzero entries of the corresponding matrix. To achieve reasonable convergence rates, the solution stability of Stokes-type problems is required. In this chapter, we discuss some iterative solvers for the discretized systems arising from the different versions of pressure jump stabilization techniques presented previously. Before proceeding, we start with some important features of matrices arising from the discretization discussed above. This will greatly justify the choice of different iterative solvers.

4.1 System matrix properties

As it was mentioned in the third chapter (§3.2.2), the stabilized algebraic problem that corresponds to the stabilized discrete weak formulation of the generalized Stokes problem is as follows:

$$\begin{gathered}\text{Find } (\mathbf{U}, P) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \quad \text{such that} \\ \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\beta\,\mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathrm{P} \end{bmatrix} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix},\end{gathered} \tag{4.1}$$

where $\mathbf{U}$ and P are the unknown vectors of n_u velocity nodal values and n_p pressure nodal values respectively.
The matrices $\mathbf{A}$, $\mathbf{B}$ and the right-hand side $\mathbf{F}$ are given, respectively, in (3.19), (3.20) and (3.22). Recall that the stabilization matrix $\mathbf{C}$ is an assembly of local stabilization matrices given for the local jump, local 2-jump or local 1-jump versions in (3.36), (3.37), (3.32) and (3.35) respectively.
The linear system (4.1) is of the form

$$\mathbf{K}\,\mathbf{x} = \mathbf{b}, \tag{4.2}$$

where

$$\mathrm{K} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\beta\,\mathbf{C} \end{bmatrix}, \tag{4.3}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{U} \\ \mathrm{P} \end{bmatrix} \tag{4.4}$$

and

$$\mathbf{b} = \begin{bmatrix} \mathbf{F} \\ \mathbf{0} \end{bmatrix}. \tag{4.5}$$

Then, the following properties hold.

Proposition 4.1.
The stabilized Schur-complement matrix $\boldsymbol{S}_\beta = \boldsymbol{B}\boldsymbol{A}^{-1}\,\boldsymbol{B}^T + \beta\,\boldsymbol{C}$ *is symmetric and positive definite.*

Proof.
First, $\mathbf{S}_\beta$ is symmetric since $\mathbf{A}^{-1}$ and $\mathbf{C}$ are symmetric. Now, let $\mathbf{q} \in \mathbb{R}^{n_p}$. We have

$$\begin{aligned} \langle \mathbf{S}_\beta\,\mathbf{q}, \mathbf{q}\rangle = &\quad \langle \big(\mathbf{B}\ \mathbf{A}^{-1}\mathbf{B}^T + \beta\ \mathbf{C}\big)\,\mathbf{q}, \mathbf{q}\rangle \\ = &\quad \langle \mathbf{A}^{-1}\mathbf{B}^T\mathbf{q}, \mathbf{B}^T\mathbf{q}\rangle + \beta\ \langle \mathbf{C}\,\mathbf{q}, \mathbf{q}\rangle \end{aligned} \tag{4.6}$$

It is clear that $\langle \mathbf{S}_\beta\,\mathbf{q}, \mathbf{q}\rangle > 0$ for all $\mathbf{q} \neq \mathbf{0}$.
Indeed,

$$\langle \mathbf{S}_\beta\,\mathbf{q}, \mathbf{q}\rangle = 0 \quad \Rightarrow \quad \langle\ \mathbf{A}^{-1}\mathbf{B}^T\mathbf{q}, \mathbf{B}^T\mathbf{q}\rangle = 0 \text{ and } \langle \mathbf{C}\,\mathbf{q}, \mathbf{q}\rangle = 0 \tag{4.7}$$

because $\mathbf{A}^{-1}$ is symmetric and positive definite since $\mathbf{A}$ is so, $\mathbf{C}$ is semi-positive definite and $\beta > 0$. Therefore,

$$\mathbf{B}^T\mathbf{q} = \mathbf{0} \quad \text{and} \quad \langle \mathbf{C}\,\mathbf{q}, \mathbf{q}\rangle = 0. \tag{4.8}$$

It is then, sufficient to apply Theorem 3.1 in its algebraic form to get $\mathbf{q} = \mathbf{0}$. □

We remind that this property has been applied in the proof of Theorem 3.3. The next property concerns the global matrix $\mathbf{K}$ defined in (4.3).

Proposition 4.2.
The matrix $\boldsymbol{K}$ *given in* (4.3) *is symmetric and indefinite.*

Proof.
By construction, $\mathbf{K}$ is clearly symmetric. Concerning its indefiniteness, we follow [14] by defining the matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S}_{\beta,} \end{bmatrix}. \tag{4.9}$$

The matrices $\mathbf{K}$ and $\mathbf{G}$ are then congruent (see Definition 1.2) since the matrix $\mathbf{K}$ can be written as

$$\mathbf{K} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S}_{\beta,} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{4.10}$$

By applying the Sylvester law of inertia, the matrices $\mathbf{K}$ and $\mathbf{G}$ have the same number of positive, zero and negative eigenvalues.
On the other hand, for all $(\mathbf{v},\mathbf{q}) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_p}$, we have

$$(\mathbf{v}, \mathbf{q})^T\ \mathbf{G}\ (\mathbf{v}, \mathbf{q}) = \quad \mathbf{v}^T\,\mathbf{A}\,\mathbf{v}\ -\ \mathbf{q}^T\,\mathbf{S}_\beta\,\mathbf{q}. \tag{4.11}$$

This means that $\mathbf{G}$ has positive and negative eigenvalues since $\mathbf{A}$ and $\mathbf{S}_\beta$ are positive definite with $-\mathbf{S}_\beta$ being negative definite. □

4.2 Iterative solvers

The aim of this section is extending the application of the so called *Krylov subspace methods* to the systems of linear equations arising from the mixed *Q1-Q0* discretization of the stabilized discrete generalized Stokes problem.

As it was seen in the previous section, the associated coefficient matrix is symmetric but indefinite. This leads us to think about the MINimum RESidual method *(MINRES)*, since this later is used for solving any symmetric system.

However, although the Conjugate Gradient method *(CG)* converges for the symmetric and positive definite *(SPD)* matrices, it will be seen in the next chapter that it could also be effective for jump stabilized discrete systems. More details are given below.

4.2.1 The conjugate gradient method

The sparsity of any coefficient matrix, say $\mathbf{M}$, makes the product matrix-vector not expensive. If there is for instance m nonzero entries in a row of the matrix $\mathbf{M}$ that is of order n, then the product of $\mathbf{M}$ with any vector $\mathbf{v}$ requires only mn multiplies and additions. Another product matrix-vector $\mathbf{M}(\mathbf{M}\,\mathbf{v})$ is also cheap. The same occurs for $\mathbf{M}(\mathbf{M}\,\mathbf{v}), \mathbf{M}\left(\mathbf{M}^2\,\mathbf{v}\right), \ldots, \mathbf{M}^i\,\mathbf{v}$ and so on. This makes any linear combination of these later vectors easy to compute and hence leads to the subspace spanned by them, called *Krylov subspace* and denoted by $K_i(\mathbf{M}, \mathbf{v})$. If $\mathbf{x}^{(0)}$ denotes the starting iterate solution of an algebraic system

$$\mathbf{M}\mathbf{x} = \mathbf{f} \tag{4.12}$$

and $\mathbf{r}^{(0)} = \mathbf{f} - \mathbf{M}\,\mathbf{x}^{(0)}$ is the initial residual vector, then the *Krylov subspace* methods are based on calculating the iterate solution in the subspace $K_i(\mathbf{M}, \mathbf{r}^{(0)})$.

One of the best known and simple Krylov subspace methods is the so-called Conjugate Gradient method (CG), which was originally proposed by Hestenes and Stiefel in 1952 (see [19]). If the matrix $\mathbf{M}$ is symmetric and positive definite and $\mathbf{x}^{(i)}$ denotes the i^{th} iterate solution then, the *CG* method consists in minimizing $\|\mathbf{e}^{(i)}\|_{\mathbf{M}}$, the $\mathbf{M} - norm$ of the corresponding error vector $\mathbf{e}^{(i)} = \mathbf{x} - \mathbf{x}^{(i)}$.

Algorithm 4.1. *CONJUGATE GRADIENT METHOD*

Choose $\boldsymbol{x}^{(0)}$, *compute* $\boldsymbol{r}^{(0)} = \boldsymbol{f} - M\ \boldsymbol{x}^{(0)}$, *set* $\boldsymbol{p}^{(0)} = \boldsymbol{r}^{(0)}$.

For $i = 0, 1, 2, \ldots,$ *until convergence, do*

$$\zeta^{(i)} = \frac{\langle \boldsymbol{r}^{(i)},\, \boldsymbol{r}^{(i)} \rangle}{\langle M\boldsymbol{p}^{(i)},\, \boldsymbol{p}^{(i)} \rangle}$$

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \zeta^{(i)}\, \boldsymbol{p}^{(i)}$$

$$\boldsymbol{r}^{(i+1)} = \boldsymbol{r}^{(i)} - \zeta^{(i)}\, M\, \boldsymbol{p}^{(i)}$$

(*Test of convergence*)

$$\varrho^{(i)} = \frac{\langle \boldsymbol{r}^{(i+1)}, \boldsymbol{r}^{(i+1)} \rangle}{\langle \boldsymbol{r}^{(i)}, \boldsymbol{r}^{(i)} \rangle}$$

$$\boldsymbol{p}^{(i+1)} = \boldsymbol{r}^{(i+1)} + \varrho^{(i)}\, \boldsymbol{p}^{(i)}$$

Enddo

From Algorithm 4.1, a breakdown of the method can happen at a division by zero. This can happen when we have, for instance, a zero denominater in the value of $\varrho^{(i)}$. In this case, there is no problem since $\mathbf{r}^{(i)} = 0$ implies $\mathbf{x}^{(i)} = \mathbf{x}$ and there is no sense to continue. This breakdown is called *lucky breakdown*. However, if $\langle\ \mathbf{M}\mathbf{p}^{(i)},\ \mathbf{p}^{(i)} \rangle = 0$, then $\mathbf{p}^{(i)} = 0$ since $\mathbf{M}$ is SPD. This implies that $\mathbf{r}^{(i+1)} = \mathbf{r}^{(i)}$ and also $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)}$, which leads to an infinite loop.

Non realistic solutions can be obtained for certain problems with insufficient accuracy. This fact renders the choice of adequate stopping criterion very important.

4.2.2 Stopping criterion

If we denote by $\kappa(\ \mathbf{M})$ the condition number (see Definition 1.3) of the SPD matrix $\mathbf{M}$ and by $\mathbf{e}^{(i)} = \mathbf{x} - \mathbf{x}^{(i)}$ the error vector in the i^{th} iteration, then the following (classical) convergence result is obtained.

Proposition 4.3.
After i steps of the conjugate gradient method, the iteration error $\boldsymbol{e}^{(i)} = \boldsymbol{u} - \boldsymbol{u}^{(i)}$ *satisfies the bound:*

$$\frac{\|\boldsymbol{e}^{(i)}\|_{\boldsymbol{M}}}{\|\boldsymbol{e}^{(0)}\|_{\boldsymbol{M}}} \leq \sqrt{\kappa(\boldsymbol{M})}\, \frac{\|\boldsymbol{r}^{(i)}\|}{\|\boldsymbol{r}^{(0)}\|}. \tag{4.13}$$

Proof.
Since

$$\begin{aligned} \|\mathbf{r}^{(0)}\|^2 = &\quad \|\mathbf{M}\mathbf{e}^{(0)}\|^2 = \langle \mathbf{M}\mathbf{e}^{(0)}, \mathbf{M}\mathbf{e}^{(0)} \rangle \\ \leq &\quad \|\mathbf{M}\| \, \langle \mathbf{M}\mathbf{e}^{(0)}, \mathbf{e}^{(0)} \rangle \\ = &\quad \|\mathbf{M}\| \, \|\mathbf{e}^{(0)}\|_{\mathbf{M}}^2, \end{aligned} \tag{4.14}$$

we get

$$\frac{1}{\|\mathbf{e}^{(0)}\|_{\mathbf{M}}^2} \leq \|\mathbf{M}\| \, \frac{1}{\|\mathbf{r}^{(0)}\|^2}. \tag{4.15}$$

On the other hand, we have

$$\begin{aligned} \|\mathbf{e}^{(i)}\|_{\mathbf{M}}^2 = &\quad \langle \mathbf{M}\mathbf{e}^{(i)}, \mathbf{e}^{(i)} \rangle = \langle \mathbf{M}\mathbf{e}^{(i)}, \mathbf{M}^{-1}\mathbf{M}\mathbf{e}^{(i)} \rangle \\ \leq &\quad \|\mathbf{M}^{-1}\| \, \langle \mathbf{M}\mathbf{e}^{(i)}, \mathbf{M}\mathbf{e}^{(i)} \rangle \\ = &\quad \|\mathbf{M}^{-1}\| \, \|\mathbf{M}\mathbf{e}^{(i)}\|^2, \end{aligned}$$

so that

$$\|\mathbf{e}^{(i)}\|_{\mathbf{M}}^2 \leq \quad \|\mathbf{M}^{-1}\| \, \|\mathbf{r}^{(i)}\|^2. \tag{4.16}$$

Finally, combining (4.15) and (4.16) and using Definition 1.3 give the estimate (4.13). □

Remark 4.1.
The Euclidean error norm clearly decreases when the residual norm decreases. Moreover, the right side of (4.13) *can be useful since it does not require the exact solution to be known.*

Remark 4.2.
It should be mentioned that in practice, since we have in general no information about the solution. Usually, we take a zero vector as the starting solution. In this case, the stopping criterion concerning the residual norms in (4.13) *can be taken as:*

$$\frac{\|\boldsymbol{r}^{(i)}\|}{\|\boldsymbol{f}\|} \leq \varepsilon, \tag{4.17}$$

where $\varepsilon\,(>0)$ *is a certain precision. Another good stopping criterion that can be taken is the difference between two succesive iterate solutions, i.e.*

$$\|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\| \leq \varepsilon. \tag{4.18}$$

4.2.3 The minimum residual method

If the matrix $\mathbf{M}$ is not positive definite then the $\mathbf{M}$-norm $\|.\|_{\mathbf{M}}$ has no sense and hencce, $\|\mathbf{e}^{(i)}\|_{\mathbf{M}}$ cannot be calculated. Fortunately, an alternative Krylov subspace method called the *MINRES* can be applied to any symmetric system. This iterative method was derived by Paige and Saunders in 1975 (see [26]). It consists in minimizing the Euclidean norm of the residual vector in each step. This leads us to apply the MINRES to solve the

system (4.2) since the corresponding stabilized matrix (4.3) is symmetric and indefinite (see Proposition 4.2).

Before proceeding further, we briefly present another Krylov subspaces method which can be considered as a link between the CG and the MINRES. This is commonly called the *Lanczos* method. The latter consists in calculating an orthonormal basis for the Krylov subspaces $K_i(\mathbf{M}, \mathbf{r}^{(0)})$.

By taking $\mathbf{v}^{(0)} = \mathbf{0}$ and $\mathbf{v}^{(1)}$ a vector such that $\|\mathbf{v}^{(1)}\| = 1$, the basis vectors can be obtained via the following recurrence:

$$\gamma_{j+1}\mathbf{v}^{(j+1)} = \mathbf{M}\,\mathbf{v}^{(j)} - \delta_j\,\mathbf{v}^{(j)} - \gamma_j\,\mathbf{v}^{(j-1)} \qquad \forall\, 1 \leq j \leq i \tag{4.19}$$

where $\delta_j = \langle \mathbf{M}\,\mathbf{v}^{(j)}, \mathbf{v}^{(j)} \rangle$ ($\delta_j > 0$ if $\mathbf{M}$ is positive definite), whereas γ_{j+1} is such that $\|\mathbf{v}^{(j+1)}\| = 1$ of not prescribed sign.
The equations (4.19) lead to the following relation:

$$\mathbf{M}\,\mathbf{V}_i \;= \mathbf{V}_i\,\mathbf{T}_i \;+\; \gamma_{i+1}\,\left[\mathbf{0}, \cdots, \mathbf{0}, \mathbf{v}^{(i+1)}\right] \tag{4.20}$$

where

$$\mathbf{V}_i = \left[\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \cdots, \mathbf{v}^{(i)}\right] \tag{4.21}$$

denotes the matrix containing the vector $\mathbf{v}^{(j)}$ in its j^{th} column, and $\mathbf{T}_i$ is the symmetric tridiagonal matrix given by:

$$\mathbf{T}_i \;=\; \text{tridiag}\,[\gamma_j,\, \delta_j, \gamma_{j+1}] \tag{4.22}$$

for all $j = 1, 2, \cdots, i$. Pre-multiplying (4.20) by the transpose of $\mathbf{V}_i$ produces the following interresting property:

$$\mathbf{V}_i^T\,\mathbf{M}\,\mathbf{V}_i \;=\; \mathbf{T}_i. \tag{4.23}$$

This equation ensures the positive definiteness of the matrix $\mathbf{T}_i$ when the matrix $\mathbf{M}$ is so. Since the CG iterate $\mathbf{x}^{(i)}$ is the unique vector in $\mathbf{x}^{(0)} + K_i(\mathbf{M}, \mathbf{r}^{(0)})$ with residual orthogonal to $K_i(\mathbf{M}, \mathbf{r}^{(0)})$ then, $\mathbf{x}^{(i)}$ can be also recovered from the *Lanczos* vectors (see [14], §2.4 for details). In other words,

$$\mathbf{x}^{(i)} = \mathbf{x}^{(0)} + \mathbf{V}_i\,\mathbf{Y}^{(i)} \tag{4.24}$$

where

$$\mathbf{Y}^{(i)} = \left(y_1^{(i)}, y_2^{(i)}, \cdots, y_i^{(i)}\right)^T \tag{4.25}$$

is a vector of dimension i such that the following orthogonality condition is satisfied

$$\mathbf{V}_i^T\,\mathbf{r}^{(i)} \;= \mathbf{0}. \tag{4.26}$$

Due to (4.20), the residual $\mathbf{r}^{(i)}$ satisfies the following:

$$\mathbf{r}^{(i)} \;= \mathbf{r}^{(0)} - \mathbf{M}\,\mathbf{V}_i\,\mathbf{Y}^{(i)} = \mathbf{V}_i\,\left(\|\mathbf{r}^{(0)}\|\,\mathbf{e}_1 - \mathbf{T}_i\,\mathbf{Y}^{(i)}\right) - \gamma_{i+1}\,y_i^{(i)}\mathbf{v}^{(i+1)} \tag{4.27}$$

where $\mathbf{e}_1 = (1, 0, \cdots, 0)^T$ is the unit vector of dimension i. Orthogonality (4.26) is imposed by choosing $\mathbf{Y}^{(i)}$ in (4.27) satisfying

$$\mathbf{T}_i \, \mathbf{Y}^{(i)} = \|\mathbf{r}^{(0)}\| \, \mathbf{e}_1. \tag{4.28}$$

By solving the system (4.28), the residual in (4.27) satisfies

$$\mathbf{r}^{(i)} = -\gamma_{i+1} \, y_i^{(i)} \mathbf{v}^{(i+1)} \quad with \quad \|\mathbf{r}^{(i)}\| = |\gamma_{i+1} \, y_i^{(i)} \mathbf{v}^{(i+1)}|. \tag{4.29}$$

Due to what was mentioned above, the system (4.28) may not be solved if the matrix $\mathbf{M}$ is not positive definite (according to (4.23)) and the iterate solution in (4.24) can not be calculated.

The MINRES method circumvents this difficulty by minimizing the Euclidean norm of the residual using the least squares solution to (4.27). This leads to Algorithm 4.2 given below (cf. [14], [26] or [27]).

Algorithm 4.2. *MINRES METHOD*

Set $\boldsymbol{v}^{(0)} = 0,\ \boldsymbol{w}^{(0)} = 0,\ \boldsymbol{w}^{(1)} = 0$

Choose $\boldsymbol{x}^{(0)}$, *compute* $\boldsymbol{v}^{(1)} = \boldsymbol{f} - \boldsymbol{M}\,\boldsymbol{x}^{(0)}$, *set* $\gamma^{(1)} = \|\boldsymbol{v}^{(1)}\|$

Set $\eta = \gamma^{(1)}\,,\ s^{(0)} = s^{(1)} = 0\,,\ c^{(0)} = c^{(1)} = 1$

For $i = 1$ *until convergence do*

$$\boldsymbol{v}^{(i)} = \frac{1}{\gamma^{(i)}}\,\boldsymbol{v}^{(i)}$$

$$\delta^{(i)} = \langle \boldsymbol{M}\,\boldsymbol{v}^{(i)}\,,\ \boldsymbol{v}^{(i)} \rangle$$

$$\boldsymbol{v}^{(i+1)} = \boldsymbol{M}\,\boldsymbol{v}^{(i)} - \delta^{(i)}\,\boldsymbol{v}^{(i)} - \gamma^{(i)}\,\boldsymbol{v}^{(i-1)} \qquad (\textit{Lanczos process})$$

$$\gamma^{(i+1)} = \|\,\boldsymbol{v}^{(i+1)}\|$$

$$\alpha_0 = c^{(i)}\,\delta^{(i)} - c^{(i-1)}\,s^{(i)}\,\gamma^{(i)}$$

$$\alpha_1 = \sqrt{\alpha_0^2 + \left(\gamma^{(i+1)}\right)^2}$$

$$\alpha_2 = \delta^{(i)}\,s^{(i)} + c^{(i-1)}c^{(i)}\gamma^{(i)}$$

$$\alpha_3 = s^{(i-1)}\gamma^{(i)}$$

$$c^{(i+1)} = \frac{\alpha_0}{\alpha_1}\ ,\ s^{(i+1)} = \frac{\gamma^{(i+1)}}{\alpha_1}$$

$$\boldsymbol{w}^{(i+1)} = \frac{1}{\alpha_1}\left(\boldsymbol{v}^{(i)} - \alpha_3\,\boldsymbol{w}^{(i-1)} - \alpha_2\,\boldsymbol{w}^{(i)}\right)$$

$$\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i-1)} + \eta\,c^{(i+1)}\,\boldsymbol{w}^{(i+1)}$$

$$\eta = -s^{(i+1)}\eta$$

(*Test of convergence*)

Enddo

As can be seen, the residual reduction profile of the MINRES method is determined by $\alpha^{(0)}$. If the latter vanishes, then no progress is made since $c^{(i+1)}$ depends on it. Thus, we get $\mathbf{x}^{(i)} = \mathbf{x}^{(i-1)}$.

4.3 Preconditioning

In practice, although the above iterative methods converge, they could require lots of executing time to achieve a sufficiently accurate solution, especially, for large systems. Hence, in order to accelerate these iterative solvers, some a priori *preconditioning* would be required. The preconditioning technique consists in reducing the number of iterations as well as the executing time that are required to achieve an accurate solution. It limits the growing under the mesh refinement. The preconditioning is a good technique when the dimension of the discrete system increases. It is based on solving an alternative system in such a way that the corresponding matrix is to be as closer as possible to the identity. Both of the new system and the original one must have the same solution.

To solve (4.12), some symmetric and positive definite matrix $\mathbf{Q}$ has to be found as closer as possible to $\mathbf{M}$ such that (4.12) is equivalent to

$$\mathbf{Q}^{-1}\,\mathbf{M}\,\mathbf{x} = \mathbf{Q}^{-1}\,\mathbf{f}. \tag{4.30}$$

It is clear that the inverse matrix $\mathbf{Q}^{-1}$ should not be calculated unless it is a diagonal matrix. It is only used in theory. Since $\mathbf{Q}$ is symmetric and positive definite, $\mathbf{Q}$-inner product $\langle .,. \rangle_{\mathbf{Q}}$ can have a sense and the new (preconditioned) matrix $\mathbf{Q}^{-1}\,\mathbf{M}$ is symmetric for it. Indeed, for any two vectors $\mathbf{u}$ and $\mathbf{v}$ we have

$$\begin{aligned} \langle \mathbf{Q}^{-1}\,\mathbf{M}\,\mathbf{u}, \mathbf{v}\rangle_{\mathbf{Q}} &= \langle \mathbf{M}\,\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{M}\,\mathbf{v}\rangle \qquad (\mathbf{M}\ \text{b}\textit{eing symmetric}). \\ &= \langle \mathbf{u}, \mathbf{Q}\,\mathbf{Q}^{-1}\mathbf{M}\,\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{Q}^{-1}\mathbf{M}\,\mathbf{v}\rangle_{\mathbf{Q}}. \end{aligned} \tag{4.31}$$

Hence, the $\mathbf{Q}$-inner product can be used in Algorithm 4.1 instead of the usual Euclidean inner product (see [27] for more details).

Denoting the preconditioned residual vector by $\mathbf{z}^{(i)}$ $\left(= \mathbf{Q}^{-1}\,\mathbf{f} - \left(\mathbf{Q}^{-1}\,\mathbf{M}\right)\,\mathbf{x}^{(i)} = \mathbf{Q}^{-1}\,\mathbf{r}^{(i)}\right)$ gives:

- $\langle \mathbf{z}^{(i)},\ \mathbf{z}^{(i)}\rangle_{\mathbf{Q}} = \langle \mathbf{Q}\,\mathbf{z}^{(i)},\ \mathbf{z}^{(i)}\rangle = \langle \mathbf{r}^{(i)},\ \mathbf{z}^{(i)}\rangle$
- $\langle\, \mathbf{Q}^{-1}\,\mathbf{M}\mathbf{p}^{(i)},\ \mathbf{p}^{(i)}\rangle_{\mathbf{Q}} = \langle\, \mathbf{Q}\,\mathbf{Q}^{-1}\,\mathbf{M}\mathbf{p}^{(i)},\ \mathbf{p}^{(i)}\rangle = \langle\, \mathbf{M}\mathbf{p}^{(i)},\ \mathbf{p}^{(i)}\rangle$

Hence, applying the Conjugate Gradient to the preconditioned system (4.30) leads to the following algorithm:

Algorithm 4.3. *PRECONDITIONED CG METHOD*

Choose $\boldsymbol{x}^{(0)}$, *compute* $\boldsymbol{r}^{(0)} = \boldsymbol{f} - M\ \boldsymbol{x}^{(0)}$, *solve* $Q\, \boldsymbol{z}^{(0)} = \boldsymbol{r}^{(0)}$, *set* $\boldsymbol{p}^{(0)} = \boldsymbol{z}^{(0)}$

For $i = 0, 1, 2, \ldots,$ *until convergence, do*

$$\zeta^{(i)} = \frac{\langle \boldsymbol{z}^{(i)},\ \boldsymbol{r}^{(i)} \rangle}{\langle\, M\boldsymbol{p}^{(i)},\ \boldsymbol{p}^{(i)} \rangle}$$

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \zeta^{(i)}\, \boldsymbol{p}^{(i)}$$

$$\boldsymbol{r}^{(i+1)} = \boldsymbol{r}^{(i)} - \zeta^{(i)}\, M\, \boldsymbol{p}^{(i)}$$

(*Test of convergence*)

Solve $Q\, \boldsymbol{z}^{(i+1)} = \boldsymbol{r}^{(i+1)}$

$$\varrho^{(i)} = \frac{\langle \boldsymbol{z}^{(i+1)}, \boldsymbol{r}^{(i+1)} \rangle}{\langle \boldsymbol{z}^{(i)}, \boldsymbol{r}^{(i)} \rangle}$$

$$\boldsymbol{p}^{(i+1)} = \boldsymbol{z}^{(i+1)} + \varrho^{(i)}\, \boldsymbol{p}^{(i)}$$

Enddo

Taking $\mathbf{Q} = \mathbf{I}$ is not a good choice since it means no preconditioning. On the other hand, $\mathbf{Q} = \mathbf{M}$ would be the ideal one since the solution $\mathbf{x}$ will be obtained in only one iteration. Indeed, choosing $\mathbf{Q} = \mathbf{M}$ in Algorithm 4.3 would give

$$\mathbf{z}^{(0)} = \quad \mathbf{M}^{-1}\mathbf{r}^{(0)} = \mathbf{M}^{-1}\left(\mathbf{f} - \mathbf{M}\mathbf{x}^{(0)}\right) = \mathbf{x} - \mathbf{x}^{(0)} = \mathbf{e}^{(0)}. \tag{4.32}$$

Therefore,

$$\zeta^{(0)} = \quad \frac{\langle \mathbf{z}^{(0)},\ \mathbf{r}^{(0)} \rangle}{\langle\, \mathbf{M}\mathbf{p}^{(0)},\ \mathbf{p}^{(0)} \rangle} = \frac{\langle \mathbf{e}^{(0)},\ \mathbf{r}^{(0)} \rangle}{\langle\, \mathbf{M}\mathbf{z}^{(0)},\ \mathbf{z}^{(0)} \rangle} = 1, \tag{4.33}$$

since $\mathbf{z}^{(0)} = \mathbf{e}^{(0)}$ and $\mathbf{M}\,\mathbf{e}^{(0)} = \mathbf{r}^{(0)}$. Consequently, we get

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} + \zeta^{(0)}\mathbf{p}^{(0)} = \mathbf{x}^{(0)} + \mathbf{p}^{(0)} \\ &= \mathbf{x}^{(0)} + \mathbf{z}^{(0)} = \mathbf{x}^{(0)} + \mathbf{e}^{(0)} \\ &= \mathbf{x}. \end{aligned} \tag{4.34}$$

However, this is very expensive. The preconditioning matrix would be strictly between $\mathbf{I}$ and $\mathbf{M}$.
If $\mathbf{Q} = \mathit{diag}(\mathbf{M})$, then we obtain what is called Jacobi preconditioner. This was successfully applied by Kechkar in [22] for the Stokes problem with the local pressure jump stabilization. We should point out that for the reduced local one-jump scheems, zero diagonal coefficients are replaced by βh_x^2 or βh_y^2 accordingly to (3.32), (3.35).

4.3.1 Preconditioned MINRES method

By applying the preconditioning technique, the preconditioned version of Algorithm 4.2 takes the following form (see [14]):

Algorithm 4.4. *PRECONDITIONED MINRES METHOD*

Set $\boldsymbol{v}^{(0)} = 0,\ \boldsymbol{w}^{(0)} = 0,\ \boldsymbol{w}^{(1)} = 0$

Choose $\boldsymbol{x}^{(0)}$, *compute* $\boldsymbol{v}^{(1)} = \boldsymbol{f} - \boldsymbol{M}\,\boldsymbol{x}^{(0)}$

Solve $\boldsymbol{Q}\ \boldsymbol{z}^{(1)} = \boldsymbol{v}^{(1)}$, *set* $\gamma^{(1)} = \sqrt{\langle \boldsymbol{z}^{(1)}, \boldsymbol{v}^{(1)} \rangle}$

Set $\eta = \gamma^{(1)}$, $s^{(0)} = s^{(1)} = 0$, $c^{(0)} = c^{(1)} = 1$

For $i = 1$ *until convergence do*

$$\boldsymbol{z}^{(i)} = \frac{\boldsymbol{z}^{(i)}}{\gamma^{(i)}},\ \delta^{(i)} = \langle \boldsymbol{M}\,\boldsymbol{z}^{(i)}\ ,\ \boldsymbol{z}^{(i)} \rangle$$

$$\boldsymbol{v}^{(i+1)} = \boldsymbol{M}\,\boldsymbol{z}^{(i)}\ -\ \frac{\delta^{(i)}}{\gamma^{(i)}}\,\boldsymbol{v}^{(i)}\ -\ \frac{\gamma^{(i)}}{\gamma^{(i-1)}}\,\boldsymbol{v}^{(i-1)},$$

$$\textit{Solve}\ \boldsymbol{Q}\boldsymbol{z}^{(i+1)} = \boldsymbol{v}^{(i+1)},\ \gamma^{(i+1)} = \sqrt{\langle \boldsymbol{z}^{(i+1)}, \boldsymbol{v}^{(i+1)} \rangle}$$

$$\alpha^{(0)} = c^{(i)}\,\delta^{(i)}\ -\ c^{(i-1)}\,s^{(i)}\,\gamma^{(i)},\ \alpha^{(1)} = \sqrt{\left(\alpha^{(0)}\right)^2\ +\ \left(\gamma^{(i+1)}\right)^2}$$

$$\alpha^{(2)} = \delta^{(i)}\,s^{(i)} + c^{(i-1)}c^{(i)}\gamma^{(i)},\ \alpha^{(3)} = s^{(i-1)}\gamma^{(i)}$$

$$c^{(i+1)} = \frac{\alpha^{(0)}}{\alpha^{(1)}}\ \ s^{(i+1)} = \frac{\gamma^{(i+1)}}{\alpha^{(1)}}$$

$$\boldsymbol{w}^{(i+1)} = \frac{1}{\alpha^{(1)}}\ \left(\boldsymbol{z}^{(i)}\ -\ \alpha^{(3)}\,\boldsymbol{w}^{(i-1)}\ -\ \alpha^{(2)}\,\boldsymbol{w}^{(i)}\right)$$

$$\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i-1)} + \eta\,c^{(i+1)}\,\boldsymbol{w}^{(i+1)}$$

$$\eta = -s^{(i+1)}\eta$$

(*Test of convergence*)

Enddo

4.3.2 Cholesky factorization preconditioner

The *Cholesky factorization* preconditioning is one of the most applied iterative solvers. It is based on calculating a lower triangular matrix, say $\mathbf{L}$, that is the exact Cholesky factor of $\mathbf{M}$. The Cholesky factorization can be obtained using the following algorithm (see [17], [27], [14] for more details).

Algorithm 4.5. *CHOLESKY FACTORIZATION*

For $i = 1$ *until* n *do*

$m = \min\{k\,;\, m_{ik} \neq 0\}$

for $j = m, \ldots,$ *until* $i - 1$ *do*

if $m_{ij} \neq 0$ *then*

$$l_{ij} = m_{ij} - \frac{1}{l_{jj}} \left(\sum_{k=m}^{j-1} l_{ik}\, l_{jk} \right)$$

endif

enddo

$$l_{ii} = \sqrt{m_{ii} - \sum_{k=m}^{i-1} l_{ik}\, l_{ik}}$$

Enddo

Application of Algorithm 4.3 and Algorithm 4.4 with the preconditioner $Q = LL^T$ yields the corresponding algorithms for the *Cholesky factorization preconditioned Conjugate Gradient* and the *Cholesky factorization preconditioned MINimum RESidual* denoted respectively by (CFPCG) and (CFPMINRES).

Chapter 5

Implementation and numerical results

In this chapter, numerical performances of the stabilization techniques and the proposed iterative solvers are assessed using test problems. Many of the numerical results presented here are published in [10].

5.1 Numerical tests

The reduced local jump stabilization procedures are evaluated and compared with the global and local jump methods. We consider two computational test problems specifically chosen to illustrate the different features of the new methods. First, for problem discretization, the domain Ω is uniformly subdivided ($h = h_x = h_y$) into square elements generating the sequence of grids GR0, GR1,...,GR8 with respective sizes $h = 2^{-i}$ $(i = 1, ..., 9)$.
Many fixed values of the stabilization parameter β were considered. However, we present only the representative case $\beta = 1$. The values $\beta = 0.1$, $\beta = 1$ and $\beta = 100$ are considered for the horizontal velocity profiles of the lid-driven cavity problem.

It is worth to mention that more investigation is still underway to determine any optimal value of β. In all present numerical tests, the kinematic viscosity coefficient μ was set to be 1, whereas the parameter α was considered to be 10^n, $n \in \{0, 1, 2, 3\}$.
In addition, the reported results will concern the global and local jump schemes together with the the different reduced local jump schemes. All computations and figure generations concerning the stabilization techniques were performed in Matlab on a Intel(R) Core(TM) i5 PC @ 2.53MHz. For the iterative solvers, figure generations were achieved on a Intel(R) Core(TM) i3 CPU M 380 @ 2.53GHz.

5.1.1 Problem with analytic solution

This first problem is considered as a study of convergence rates and comparison of the different versions of the methods in terms of accuracy. The problem is that of an enclosed flow in the unit square $\Omega = [0, 1] \times [0, 1]$ where the velocity vector and pressure solution fields are given by :

$$\mathbf{u}(x,y) = \begin{bmatrix} x^2\left(\dfrac{x}{3} - \dfrac{1}{2}\right) \\ \\ xy(1-x) \end{bmatrix}, \quad p(x,y) = x^2 - \frac{1}{3}, \tag{5.1}$$

which clearly satisfies the constraint:

$$\int_\Omega p(x,y)\, d\Omega = 0.$$

The body force $\mathbf{f}$ is then chosen to satisfy (2.3). The values of $\mathbf{u}$ on the boundary of Ω are constrained to those given above. To remove the constant pressure mode from the numerical solution, an element pressure was imposed (cf [28]).

The results displayed in Figures 5.1−5.4, show that for all considered values of α, the rates of convergence behave as predicted by the theory, i.e. $O(h)$ order of convergence for $\|\mathbf{u} - \mathbf{u}_h\|_1$ and $\|p - p_h\|_0$ and $O(h^2)$ order for $\|\mathbf{u} - \mathbf{u}_h\|_0$. This occurs for all stabilized schemes exept the global jump one for which we notice no monotonic decrease for the pressure errors. On the other hand, due to the constant C given by (3.67), some of these rates might begin to deteriorate as α takes values above 1000.

Furthermore, in order to show the sensitivity of L^2 and H^1 velocity errors, and L^2 pressure errors to the choice of the parameter α, the results are depicted in Tables 5.1−5.6 for the finest mesh grid with uniform size h^{-9} and the stabilization parameter $\beta = 1$. As may be seen, the error magnitudes are not so affected by the increase of α. It is important to mention that the local 2-jump and 1-jump perform remarkably well. The velocity results of the 1-jump scheme are even better than those of the other local jump methods. When analysing for other considered values of β, not reported here, we also noticed the sensitivity of the global jump scheme with respect to the values of β, confirming what was claimed above. This fact will be more illustrated next.

α	Global jump	Local jump	Local 2j - H	Local 2j - V
0	4.688920×10^{-6}	1.603484×10^{-6}	1.641443×10^{-6}	1.449774×10^{-6}
1	4.687322×10^{-6}	1.603166×10^{-6}	1.641484×10^{-6}	1.449673×10^{-6}
10	4.672607×10^{-6}	1.603220×10^{-6}	1.639725×10^{-6}	1.448819×10^{-6}
100	4.606065×10^{-6}	1.594786×10^{-6}	1.633302×10^{-6}	1.441866×10^{-6}
1000	4.512608×10^{-6}	1.576241×10^{-6}	1.615563×10^{-6}	1.423313×10^{-6}

Table 5.1: Comparison of $\|\mathbf{u} - \mathbf{u}_h\|_0$ results for $\beta = 1$ as α grows (Part 1).

α	Local 1j - H1	Local 1j - H2	Local 1j - V1	Local 1j - V2
0	1.45901×10^{-6}	1.563506×10^{-6}	1.459032×10^{-6}	1.563592×10^{-6}
1	1.458824×10^{-6}	1.563629×10^{-6}	1.458766×10^{-6}	1.563618×10^{-6}
10	1.458598×10^{-6}	1.562656×10^{-6}	1.457984×10^{-6}	1.563014×10^{-6}
100	1.450404×10^{-6}	1.556499×10^{-6}	1.450797×10^{-6}	1.555817×10^{-6}
1000	1.432451×10^{-6}	1.538922×10^{-6}	1.432673×10^{-6}	1.538605×10^{-6}

Table 5.2: Comparison of $\|\mathbf{u} - \mathbf{u}_h\|_0$ results for $\beta = 1$ as α grows (Part 2).

α	Global jump	Local jump	Local 2j - H	Local 2j - V
0	7.574209×10^{-4}	1.088619×10^{-3}	1.128951×10^{-3}	8.425941×10^{-4}
1	7.574209×10^{-4}	1.088624×10^{-3}	1.128951×10^{-3}	8.425942×10^{-4}
10	7.574196×10^{-4}	1.088699×10^{-3}	1.128949×10^{-3}	8.425938×10^{-4}
100	7.574232×10^{-4}	1.088581×10^{-3}	1.128925×10^{-3}	8.425937×10^{-4}
1000	7.574430×10^{-4}	1.088217×10^{-3}	1.128567×10^{-3}	8.426288×10^{-4}

Table 5.3: Comparison of $\|\mathbf{u} - \mathbf{u}_h\|_1$ results for $\beta = 1$ as α grows (Part 1).

α	Local 1j - H1	Local 1j - H2	Local 1j - V1	Local 1j - V2
0	8.450192×10^{-4}	1.049292×10^{-3}	8.450228×10^{-4}	1.049291×10^{-3}
1	8.450232×10^{-4}	1.049293×10^{-3}	8.450183×10^{-4}	1.049291×10^{-3}
10	8.450231×10^{-4}	1.049293×10^{-3}	8.450195×10^{-4}	1.049291×10^{-3}
100	8.450112×10^{-4}	1.049626×10^{-3}	8.450167×10^{-4}	1.049291×10^{-3}
1000	8.450382×10^{-4}	1.049308×10^{-3}	8.450544×10^{-4}	1.049306×10^{-3}

Table 5.4: Comparison of $\|\mathbf{u} - \mathbf{u}_h\|_1$ results for $\beta = 1$ as α grows (Part 2).

α	Global jump	Local jump	Local 2j - H	Local 2j - V
0	2.527735×10^{-4}	3.932066×10^{-4}	6.625991×10^{-4}	6.675800×10^{-4}
1	2.527683×10^{-4}	3.932277×10^{-4}	6.626018×10^{-4}	6.675835×10^{-4}
10	2.527273×10^{-4}	3.936559×10^{-4}	6.626364×10^{-4}	6.676437×10^{-4}
100	2.524362×10^{-4}	3.936708×10^{-4}	6.631283×10^{-4}	6.682224×10^{-4}
1000	2.496844×10^{-4}	3.974746×10^{-4}	6.672416×10^{-4}	6.736131×10^{-4}

Table 5.5: Comparison of $\|p - p_h\|_0$ results for $\beta = 1$ as α grows (Part 1).

α	Local 1j - H1	Local 1j - H2	Local 1j - V1	Local 1j - V2
0	6.799543×10^{-4}	1.806382×10^{-3}	6.762094×10^{-4}	1.810183×10^{-4}
1	6.800084×10^{-4}	1.806391×10^{-3}	6.762154×10^{-4}	1.810192×10^{-4}
10	6.800528×10^{-4}	1.806437×10^{-3}	6.762573×10^{-4}	1.810237×10^{-4}
100	6.805634×10^{-4}	1.808212×10^{-3}	6.768359×10^{-4}	1.810903×10^{-4}
1000	6.861392×10^{-4}	1.813417×10^{-3}	6.824065×10^{-4}	1.817224×10^{-4}

Table 5.6: Comparison of $\|p - p_h\|_0$ results for $\beta = 1$ as α grows (Part 2).

5.1.2 Lid-driven cavity problem

The generalized Stokes problem (2.1) with $\mathbf{f} = \mathbf{0}$ is to be solved in the unit square $[0,1] \times [0,1]$ with the imposed non-leaky boundary conditions:

$$\begin{cases} \mathbf{u}(0,y) = \mathbf{u}(1,y) = \mathbf{u}(x,0) = \mathbf{0} & \text{for } 0 \leq x, y \leq 1 \\ \\ \mathbf{u}(x,1) = [1\ 0]^T & \text{for } 0 < x < 1. \end{cases} \tag{5.2}$$

In order to assess the performance of the local jump stabilization techniques and compare them with the local jump and global jump methods, the mesh GR5 with uniform size $h = 2^{-6}$ is used. Owing to the fact that the lid-driven cavity problem does not possess an analytic solution, it is generally impossible to calculate the exact accuracy and hence the convergence rates for the discrete solutions. Fortunately, there are features that can be exhibited to give an idea of how the computed solutions behave.

First, the convergence of the velocity can be assessed by plotting the profiles of the horizontal velocity u_1 along the centerline $(x = 0.5)$. The obtained profiles are illustrated in Figures 5.5−5.7. As can be seen from these figures, the local 2-jump and 1-jump schemes perform well and reproduce profiles which are almost indistinguishable with those of the local jump. However, as can be seen, the global jump technique can yield very inaccurate velocity solutions when the stabilization parameter β is varied. This demonstrates once more that stabilization via local jump schemes are far less sensitive to the choice of β.

Elevations for the pressure field and the horizontal velocity u_1 are displayed in Figures 5.8−5.13. Computed solutions are comparable to the ones reported in [25].
However, it should be noted that the 1-jump method does behave as satisfactorily for the pressure. Furthermore, there are no pressure oscillations in all presented cases.
In addition, the velocity streamslices, exhibited in Figures 5.14−5.16, indicate that for small values of α, the flow is essentially a Stokes-like flow with small counter rotating recirculations appearing at the bottom two corners which is in agrement with some similar results found in [14] and [25]. Likewise, we observe that high values of α may lead to some oscillations as expected.

5.2 Compared performance of local 2-jump and 1-jump schemes

First of all, it is worthy to admit the difficulty of deciding a priori which of the proposed schemes would perform better than the orthers for a particular problem (geometry and material parameters). This can meanly come from the adopted unified theory. It seems reasonable to reckon that the constants α_1, α_2, C_1 and C_2 appearing in the structure of the convergence constant C (see (3.62) and (3.67)) play a crucial role in the behaviour of the adopted discrete scheme. Now, from the computational side, the undertaken numerical experiments allow us to draw some guidelines with respect to the appropriate number, position and direction of involved pressure jumps in the particular scheme. First, consider separately the two benchmark problems.

5.2.1 Problem 1

In this case, the L^2 and H^1 velocity and L^2 pressure errors are exhibited in Tables 5.7−5.9 for the particular choice $\alpha = 100$ and $\beta = 1$. We observe that the 2-jump schemes behave almost similarly in terms of error magnitudes. However for the 1-jump schemes, the velocity and pressure errors present discrepancies. The errors are quite comparable for the velocity while for the pressure, they are more twice larger for 1-jump H2 and 1-jump V2. Nevertheless, both velocity and pressure errors seem to decrease according to the expected convergence rates for all schemes.

$1/h$	2j-H	2j-V	1j-H1	1j-H2	1j-V1	1j-V2
2	0.038518	0.028126	0.038127	0.047173	0.031932	0.051568
4	0.012144	0.009830	0.011726	0.014879	0.011678	0.014975
8	0.003977	0.003219	0.003554	0.004283	0.003564	0.042401
16	0.001202	0.000978	0.001017	0.001177	0.001020	0.001170
32	0.000335	0.000277	0.000282	0.000318	0.000283	0.000320
64	0.000090	0.000076	0.000077	0.000085	0.000077	0.000085
128	0.000024	0.000020	0.000021	0.000023	0.000021	0.000023
256	0.000006	0.000005	0.000005	0.000006	0.000005	0.000006
512	0.000002	0.000001	0.000001	0.000002	0.000001	0.000002

Table 5.7: Comparison behavior of $\|\mathbf{u} - \mathbf{u}_h\|_0$ for $\beta = 1$ and $\alpha = 100$.

$1/h$	2j-H	2j-V	1j-H1	1j-H2	1j-V1	1j-V2
2	0.205548	0.210018	0.203715	0.246556	0.202647	0.247435
4	0.113239	0.108184	0.105926	0.131609	0.105901	0.131623
8	0.064075	0.053867	0.054042	0.066948	0.053715	0.066809
16	0.034615	0.026905	0.027059	0.033569	0.026980	0.033522
32	0.017823	0.013469	0.013523	0.016788	0.013511	0.016780
64	0.008994	0.006739	0.006760	0.008394	0.006759	0.008393
128	0.004510	0.003370	0.003380	0.004197	0.003380	0.004197
256	0.002257	0.001685	0.001690	0.002099	0.001690	0.002099
512	0.001129	0.000843	0.000845	0.001050	0.000845	0.001049

Table 5.8: Comparison behavior of $\|\mathbf{u} - \mathbf{u}_h\|_1$ for $\beta = 1$ and $\alpha = 100$.

$1/h$	2j- H	2j-V	1j-H1	1j-H2	1j-V1	1j-V2
2	0.839352	0.586335	1.064232	2.209492	1.064232	2.209492
4	0.209152	0.264863	0.37204 0	0.652436	0.371610	0.652866
8	0.055210	0.086260	0.106325	0.181878	0.097789	0.190435
16	0.021657	0.029595	0.033910	0.066582	0.030738	0.069760
32	0.010297	0.012029	0.013135	0.029839	0.012237	0.030737
64	0.005851	0.006132	0.005544	0.014191	0.005308	0.015426
128	0.002612	0.002700	0.002796	0.007217	0.002736	0.007277
256	0.001318	0.001340	0.001372	0.003610	0.001357	0.003626
512	0.000663	0.000668	0.000681	0.001808	0.000677	0.001810

Table 5.9: Comparison behavior of $\|p - p_h\|_0$ for $\beta = 1$ and $\alpha = 100$.

5.2.2 Problem 2

For this problem with non-smooth solution, the proposed schemes are again compared using the grid GR5 with the parameters $\alpha = 100$ and $\beta = 1$. The profiles of the horizontal discrete velocity u_1 along the centerline ($x = 0.5$) are plotted against that of the local jump technique as depicted in Figure 5.17. Likewise, the discrete pressure profiles on the highest computed pressure level (y^*) of the cavity are displayed in Figure 5.18. Note that in every situation, the profiles of the local jump scheme are used as references. All the profiles of the discrete velocity seem to coincide. As for the discrete presure, the behavior of both local 2-jump schemes appears to be similar, whereas the local 1-jump schemes appear to be more affected by the choice of the jump. Some discrepancies occur especially in the close neighborhood of top corners where the considered problem is known to have singularity. Here, the same point is made regarding to the 2-jump schemes.

5.3 Performance of iterative methods

In this section, our interest is focused on the computational behavior of some iterative solvers among those described in the preceeding chapter when applied to the discrete matrix systems generated by the stabilized pressure jump schemes studied in this thesis. In particular, we consider the convergence performance of the CG and MINRES methods as well as that of their preconditioned counterparts according to the Jacobi technique and the Cholesky factorization. The model problem is again that of the lid-driven cavity discretized by the GR4 mesh with uniform size $h = 2^{-5}$. It is important to note that only half of the solution domain was modelled because of the known symmetry of the problem. This gives a total of 1634 degrees of freedom. For all cases, the tolerance of convergence was a reduction of 10^{-8} in the Euclidian norm of residuals.

The obtained convergence plots are given in Figures 5.19-5.27. First, these figures clearly show that all considered iterative solvers converge satisfactorily for all discussed schemes with different values of the parameter α. Secondly, as it was expected, the preconditioned techniques behave better in all situations. Finally, the MINRES method in both unpreconditioned and preconditioned forms needs less iterations and exhibits smoother convergence curves than the corresponding CG method form does. It is also important to note that the Cholesky preconditioning produces by far better results than does the Jacobi preconditioning. Nevertheless, on a behavior/cost scale, the CG algorithms are certainly superior since the computation efforts are not as bulk as for the MINRES algorithms.

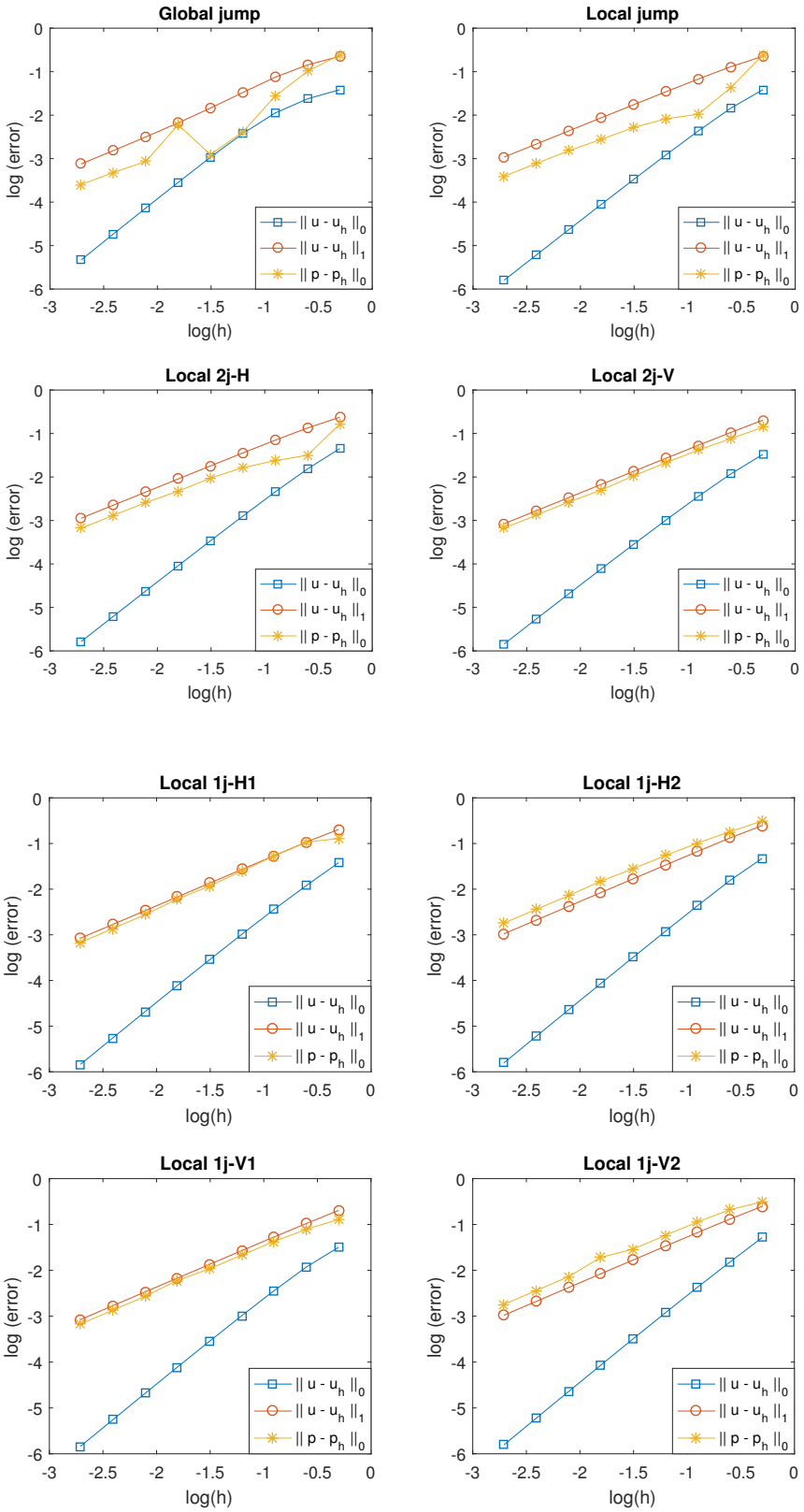

Figure 5.1: Convergence history for $\beta = 1$ and $\alpha = 0$.

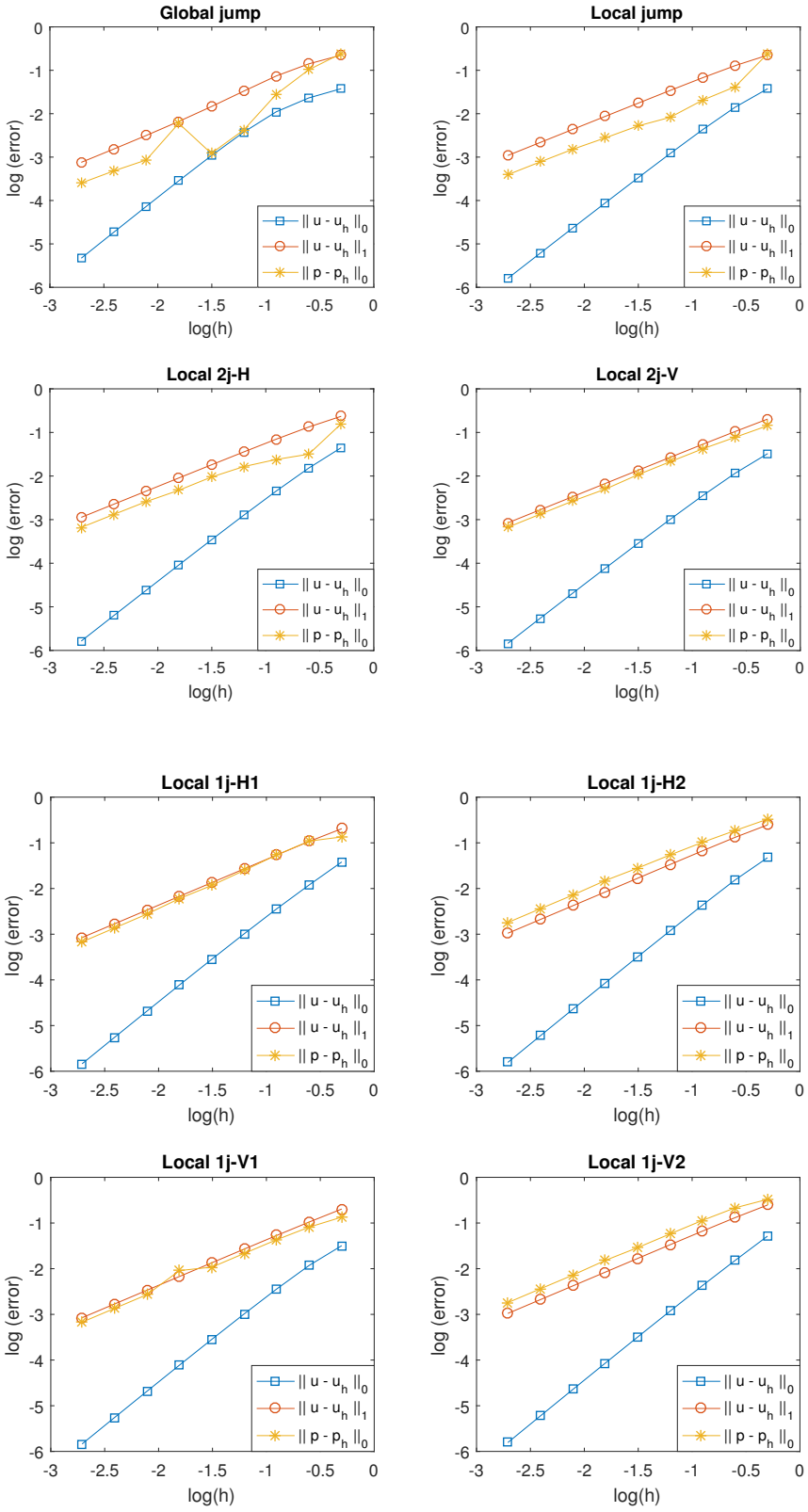

Figure 5.2: Convergence history for $\beta = 1$ and $\alpha = 1$.

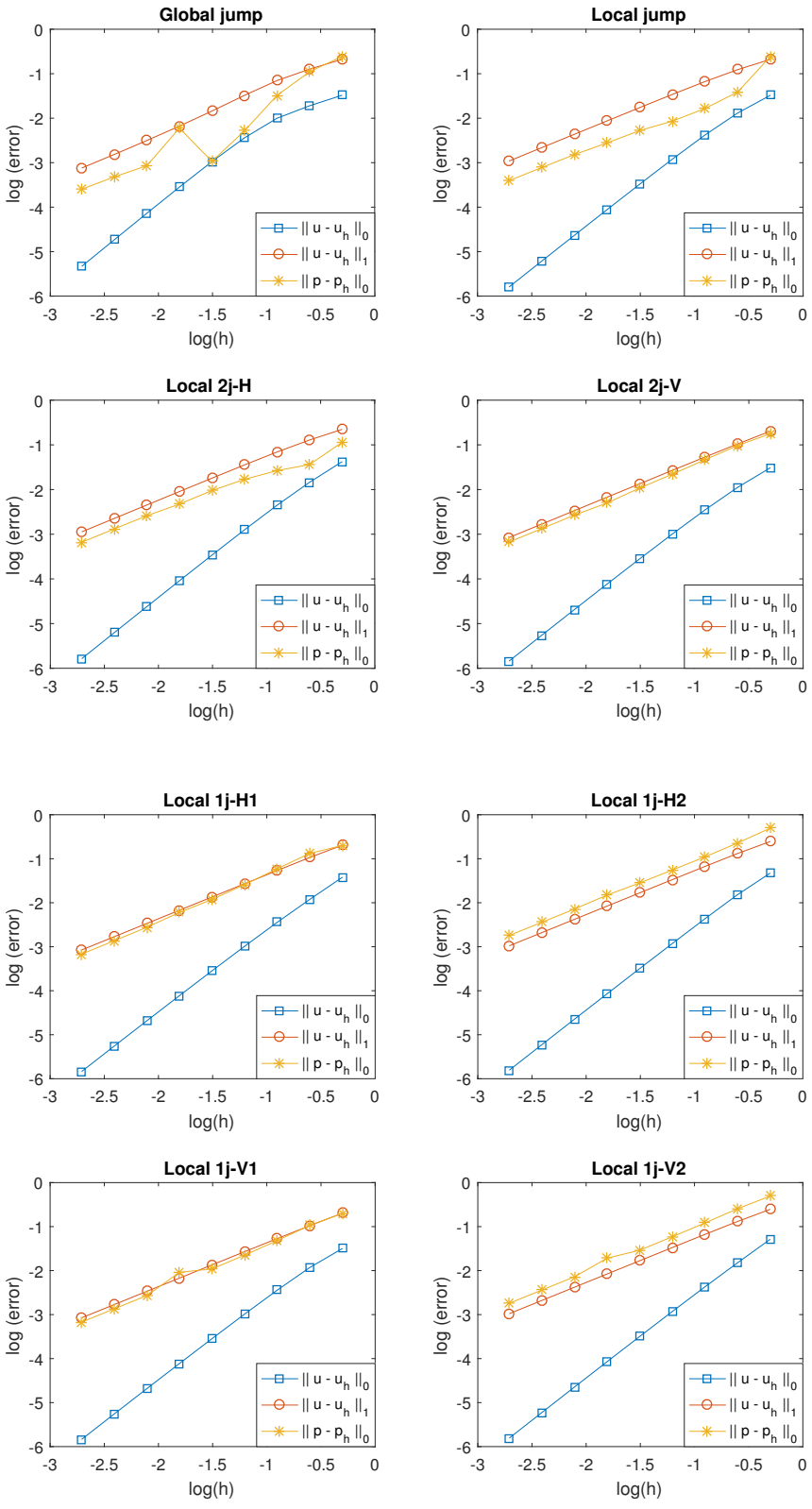

Figure 5.3: Convergence history for $\beta = 1$ and $\alpha = 10$.

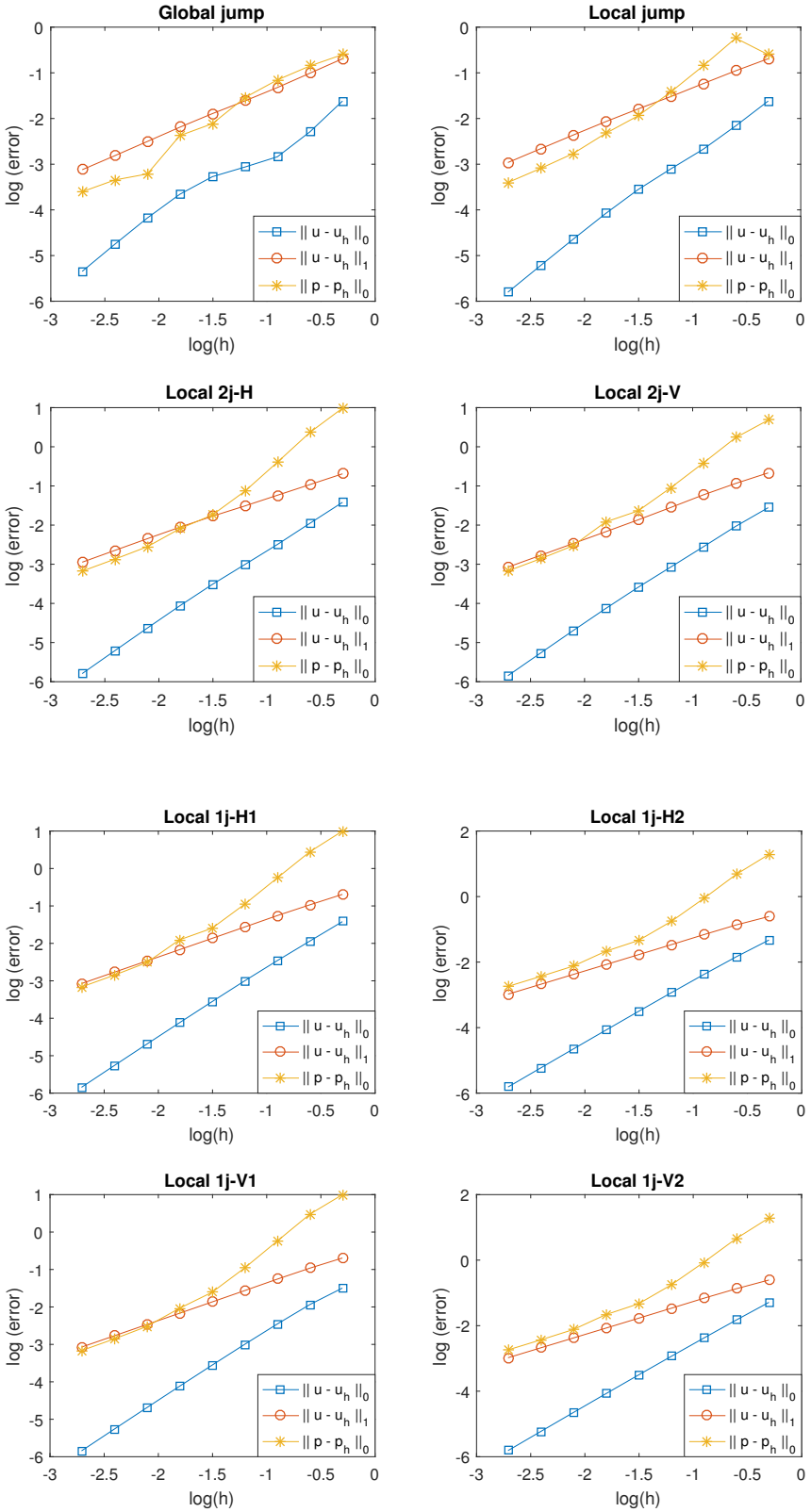

Figure 5.4: Convergence history for $\beta = 1$ and $\alpha = 1000$.

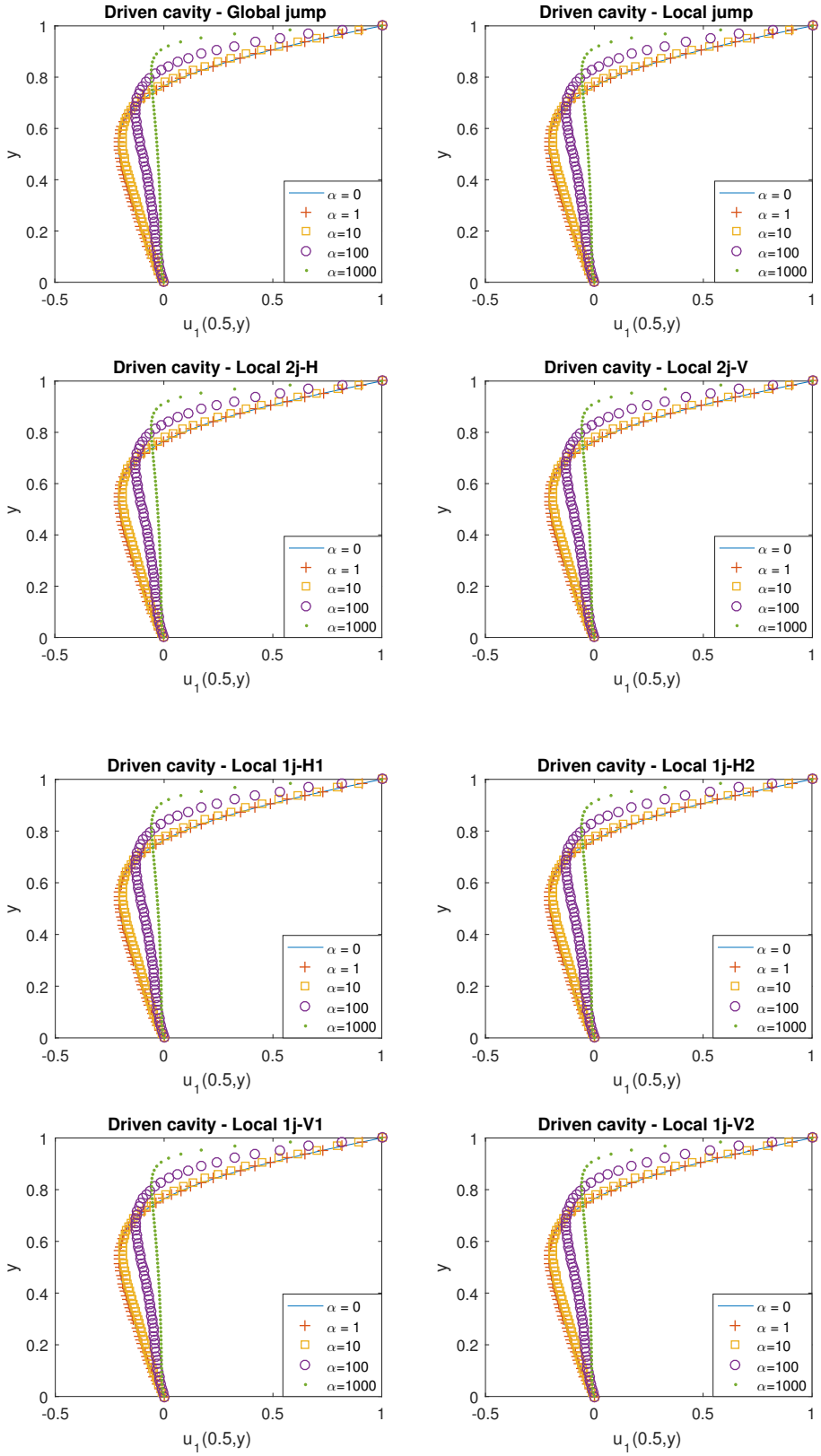

Figure 5.5: Horizontal velocity profiles for $\beta = 0.1$ when α grows.

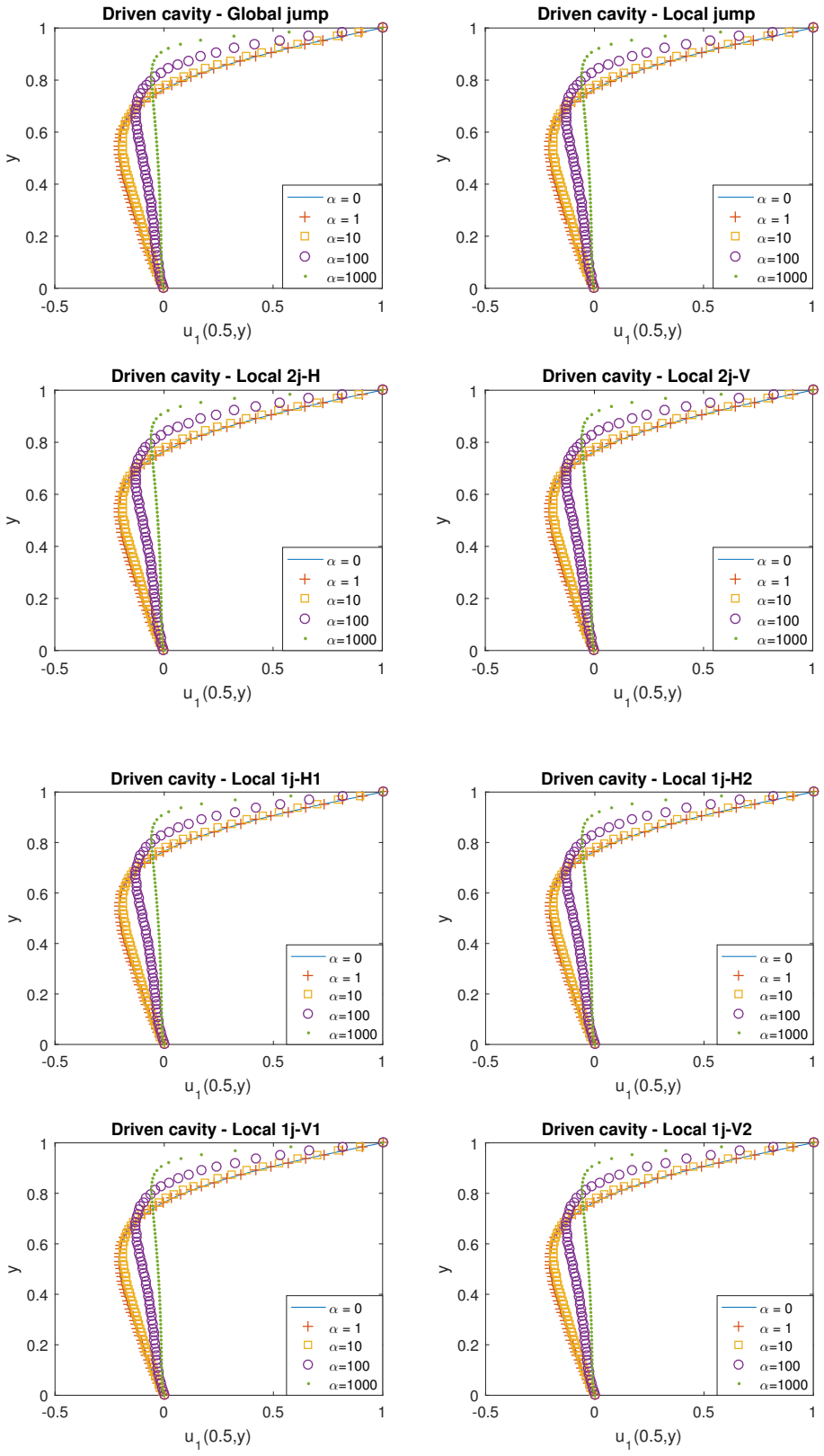

Figure 5.6: Horizontal velocity profiles for $\beta = 1$ when α grows.

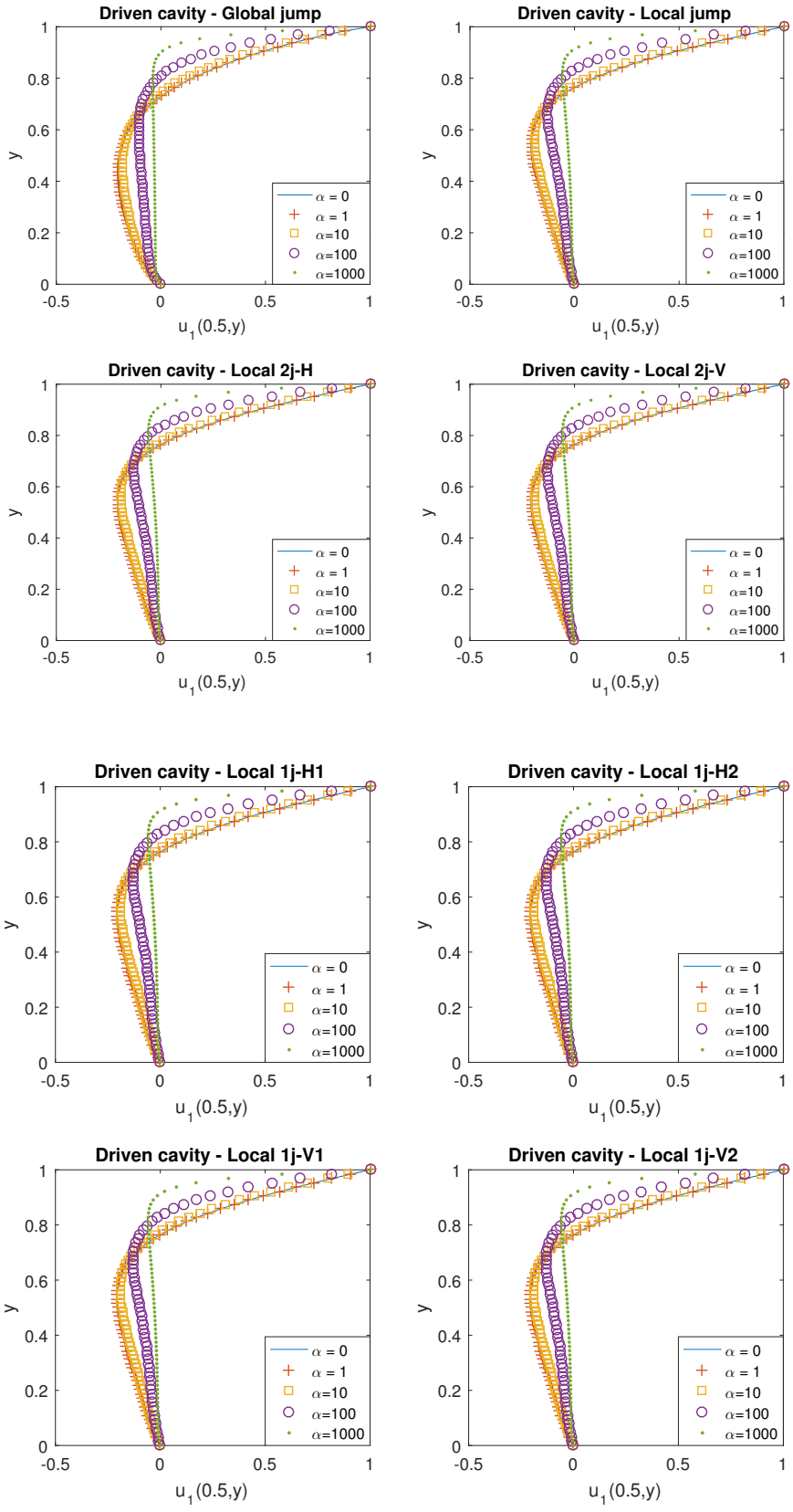

Figure 5.7: Horizontal velocity profiles for $\beta = 100$ when α grows.

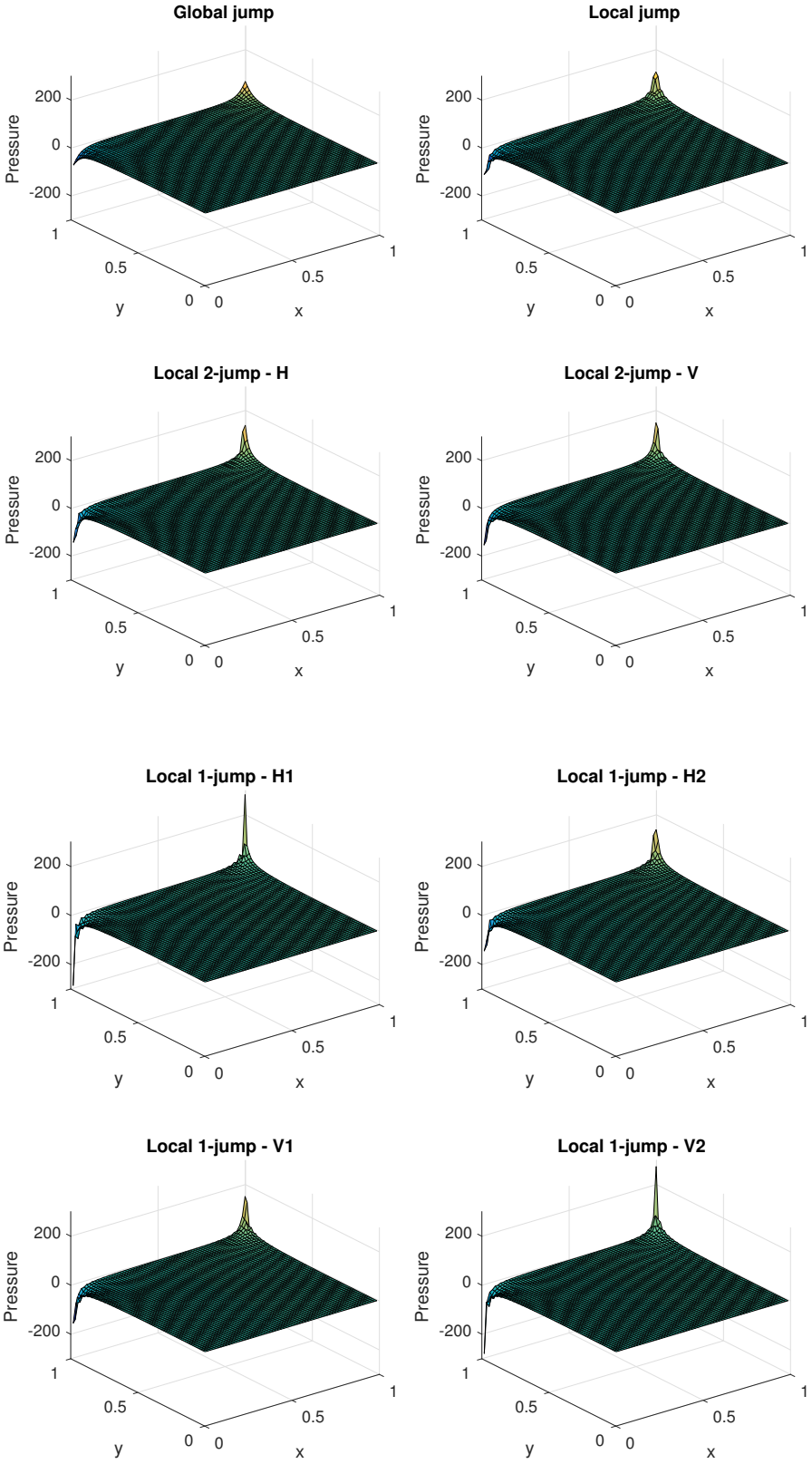

Figure 5.8: Pressure field for $\alpha = 0$.

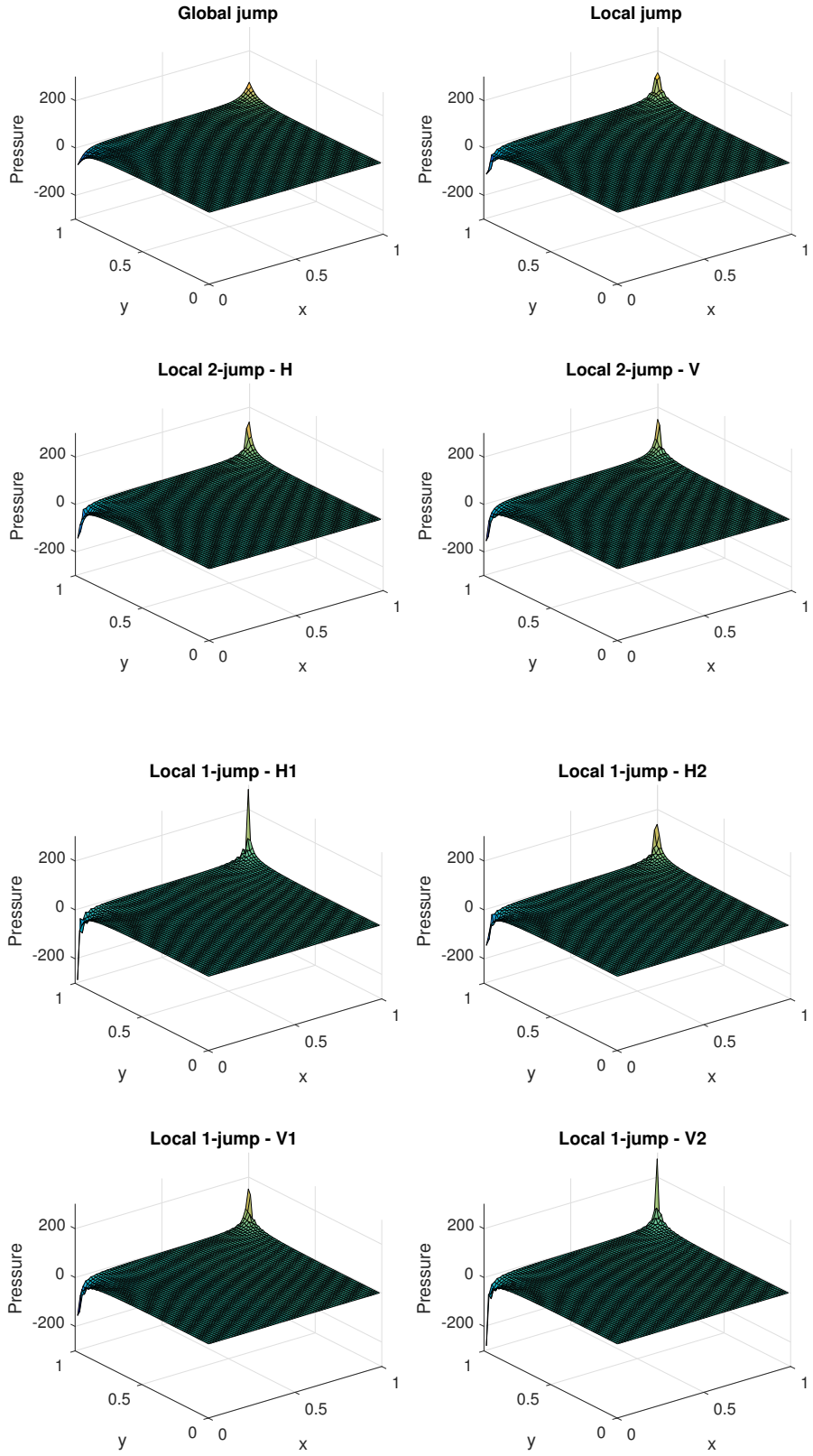

Figure 5.9: Pressure field for $\alpha = 1$.

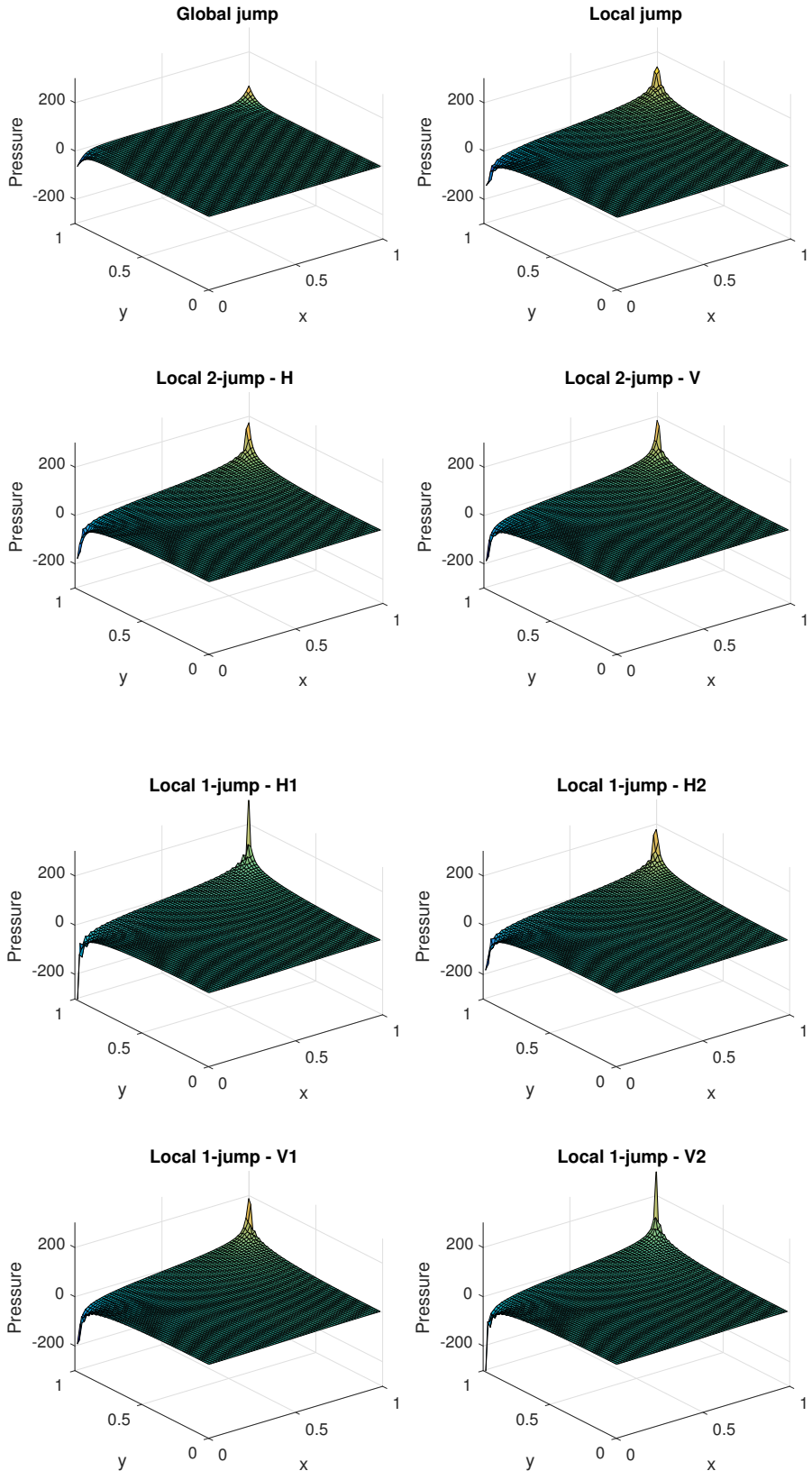

Figure 5.10: Pressure field for $\alpha = 1000$.

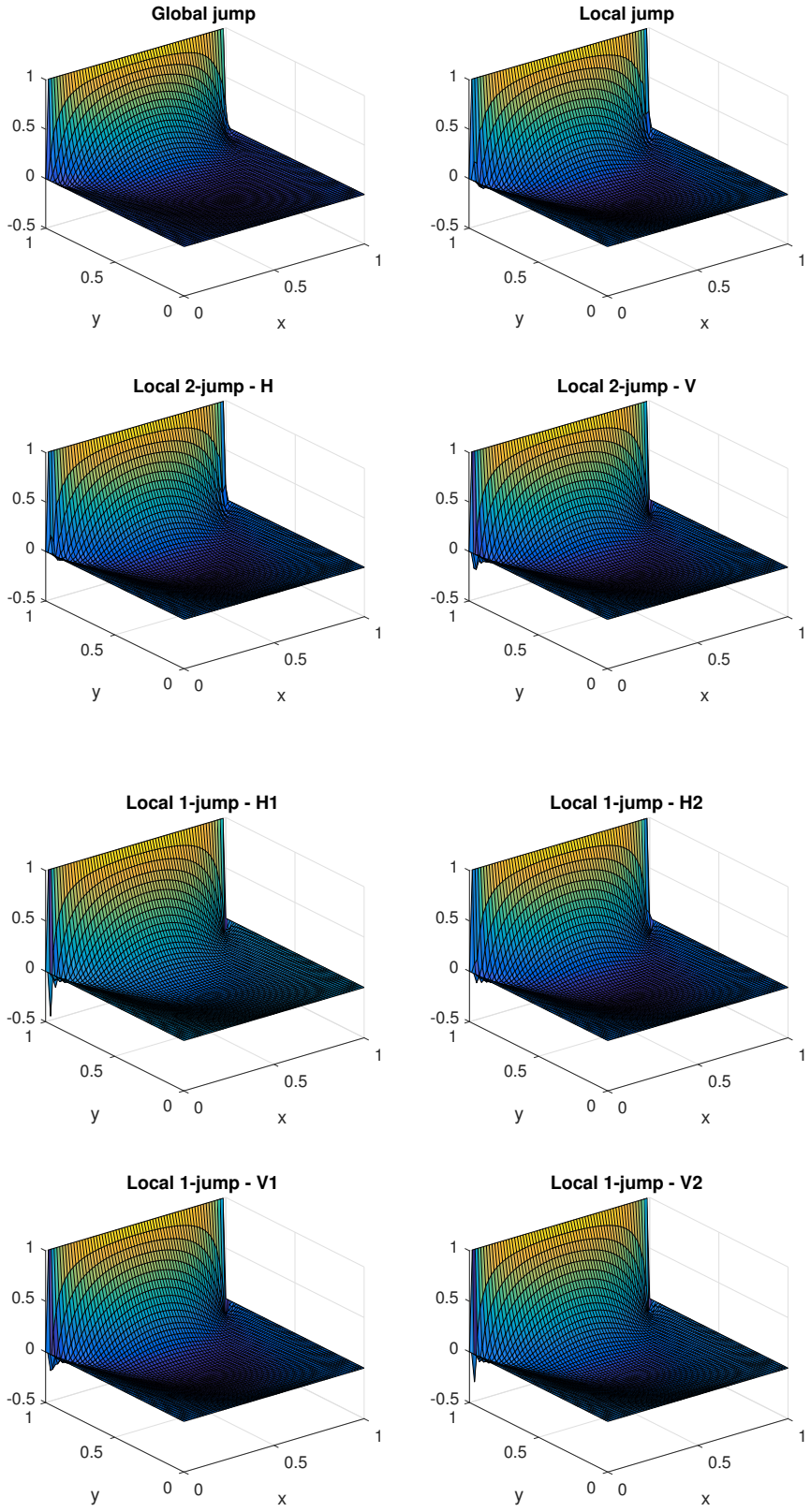

Figure 5.11: Horizontal velocity field for $\alpha = 0$.

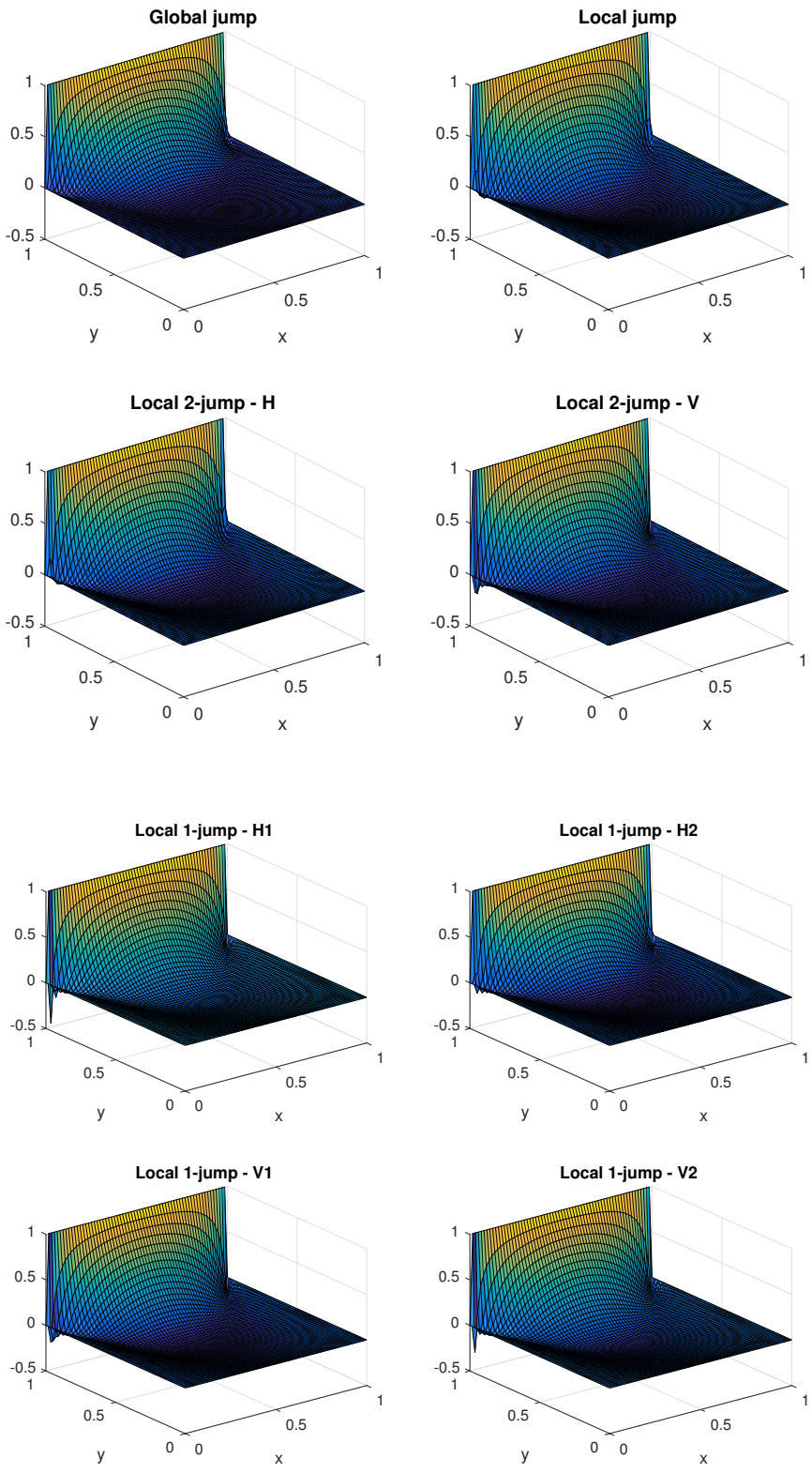

Figure 5.12: Horizontal velocity field for $\alpha = 1$.

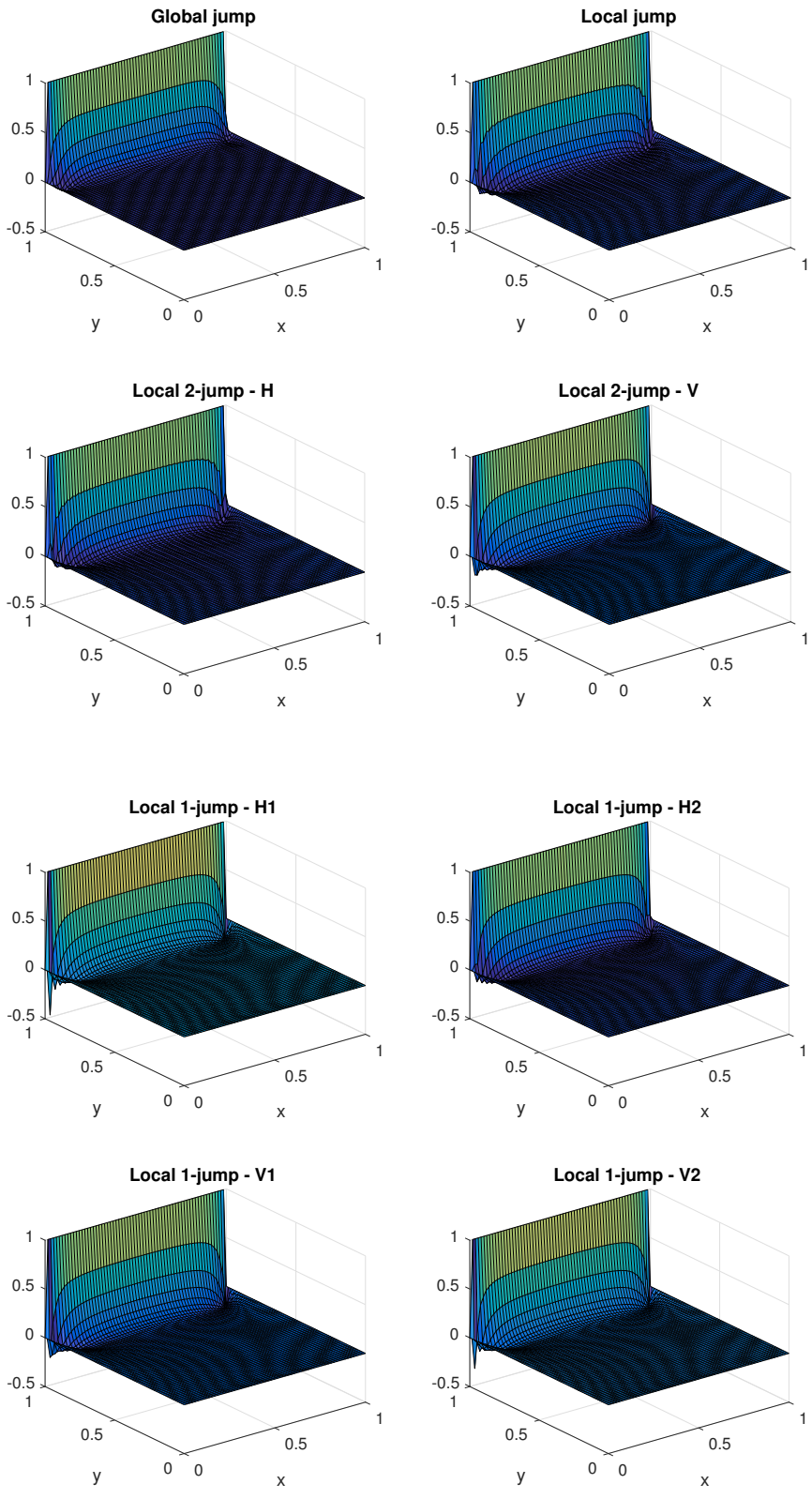

Figure 5.13: Horizontal velocity field for $\alpha = 1000$.

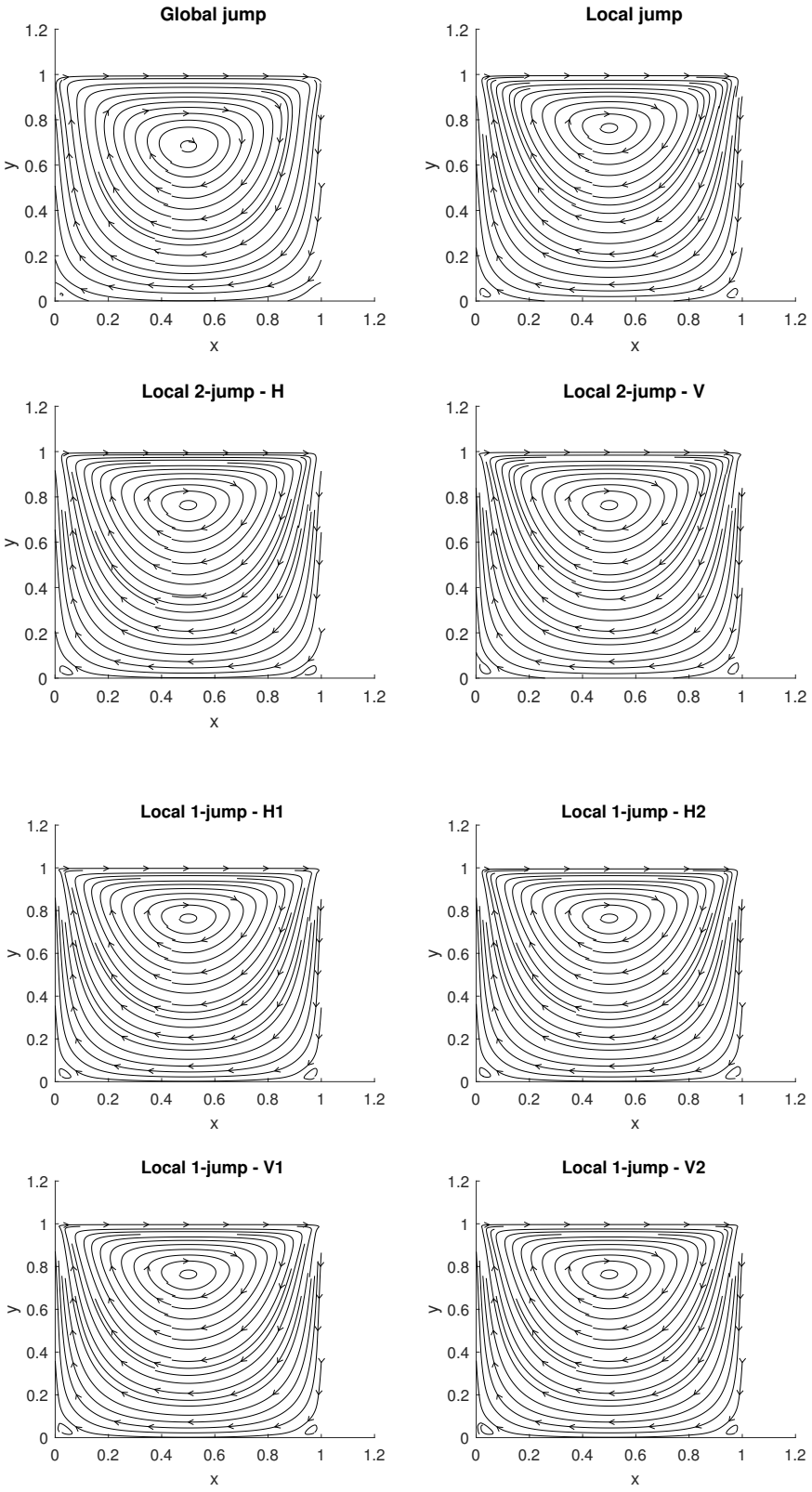

Figure 5.14: Exponential distributed streamslices plot for $\alpha = 0$.

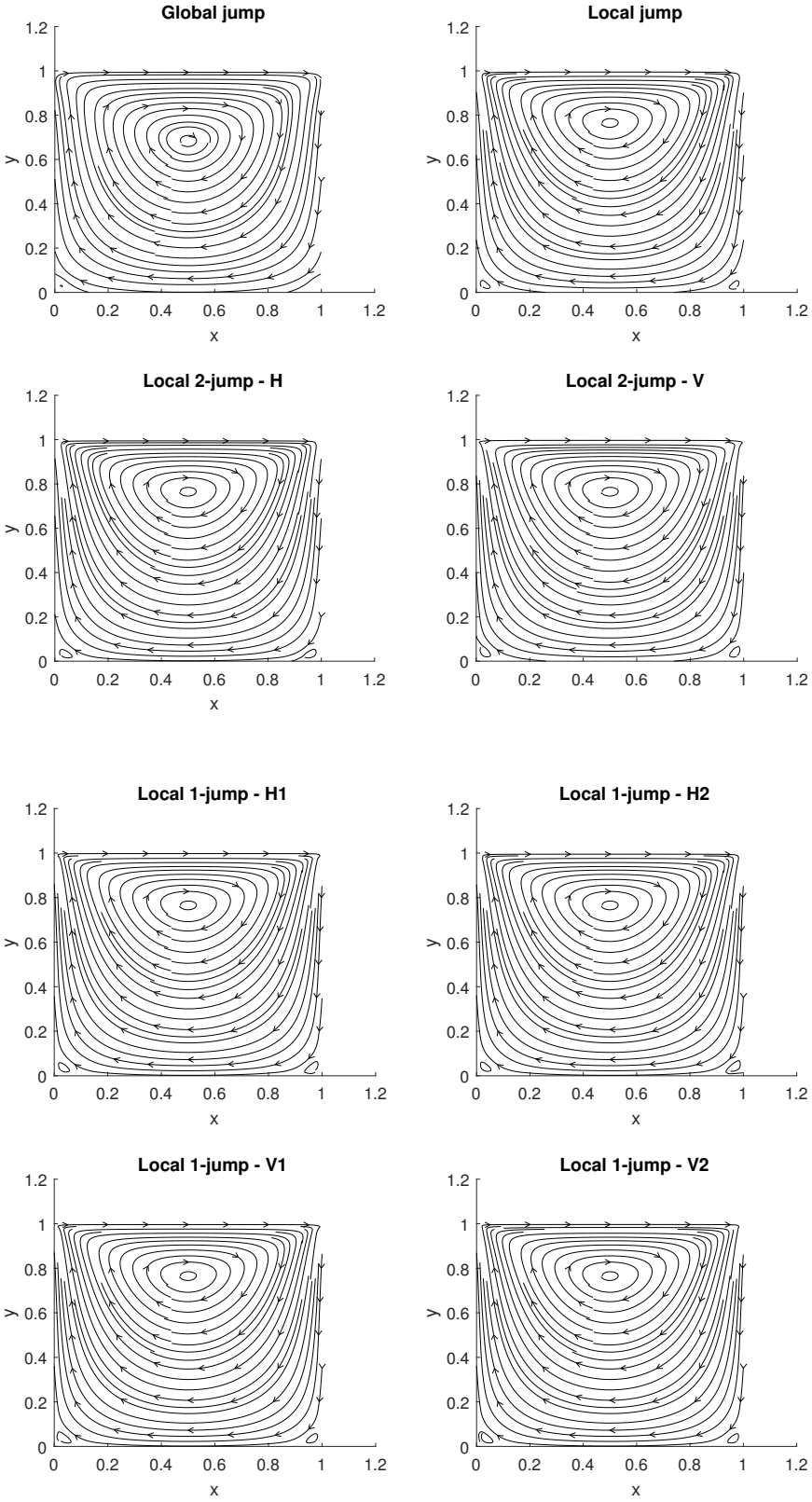

Figure 5.15: Exponential distributed streamslices plot for $\alpha = 1$.

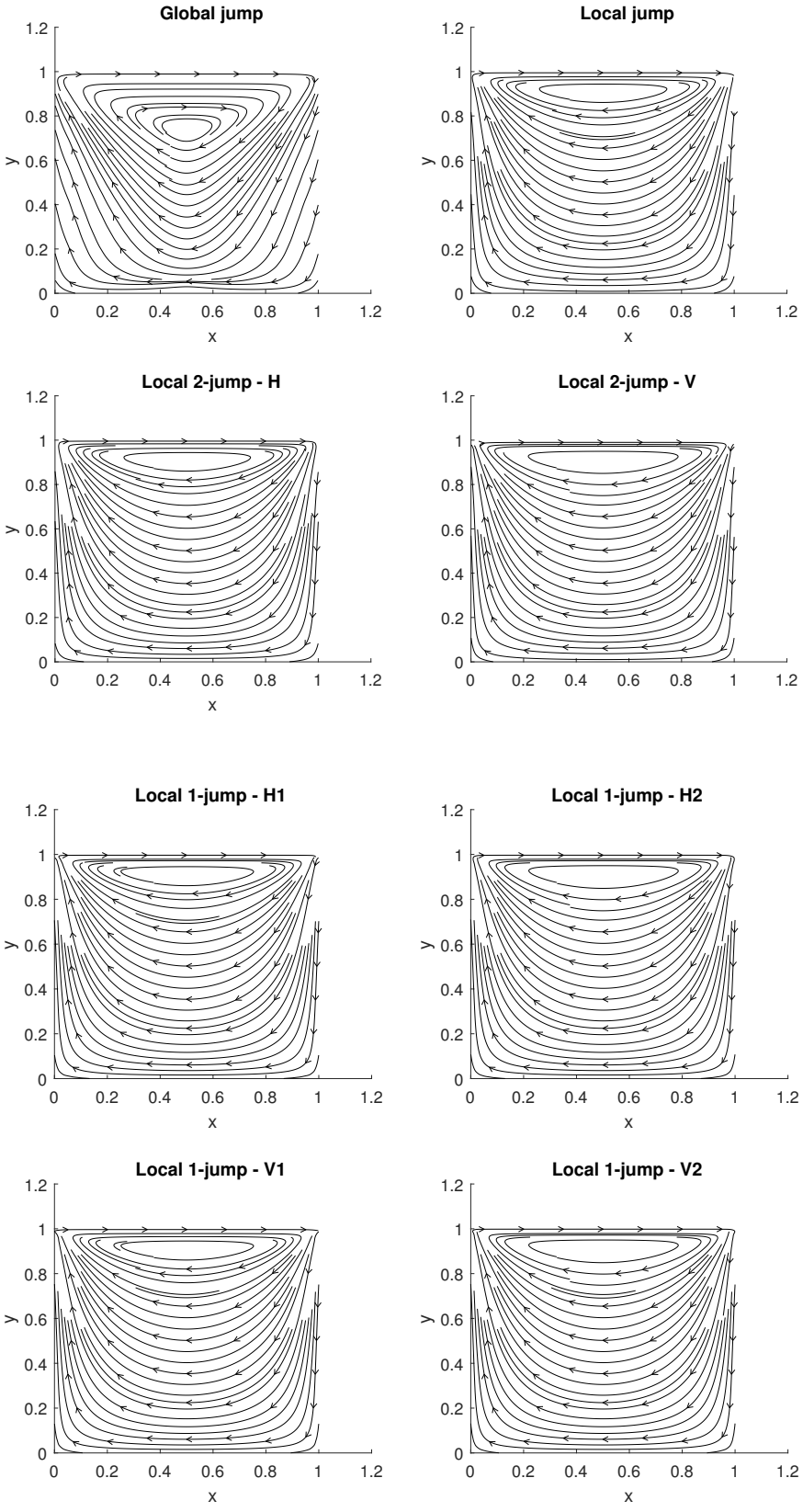

Figure 5.16: Exponential distributed streamslices plot for $\alpha = 1000$.

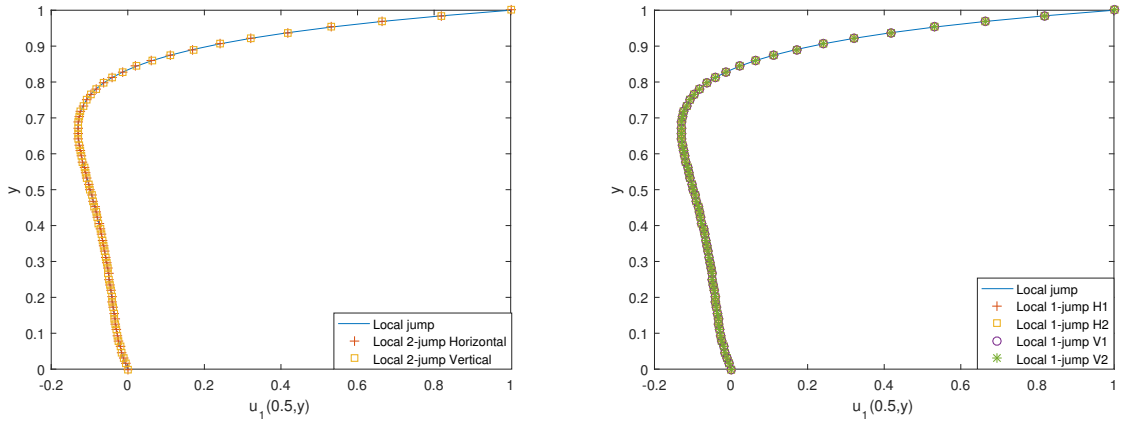

Figure 5.17: Comparative horizontal velocity profiles of the local 2-jump and 1-jump schemes for α=100 and β=1.

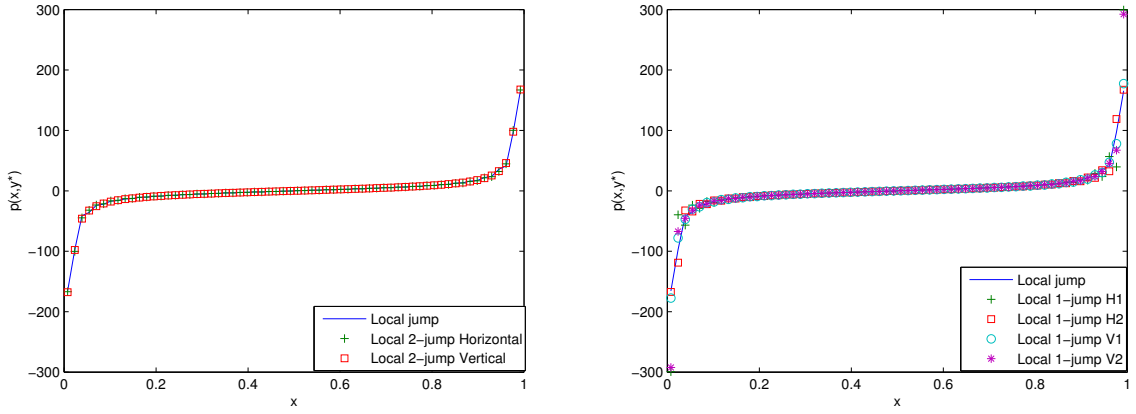

Figure 5.18: Comparative pressure profiles of the local 2-jump and 1-jump schemes for α=100 and β=1.

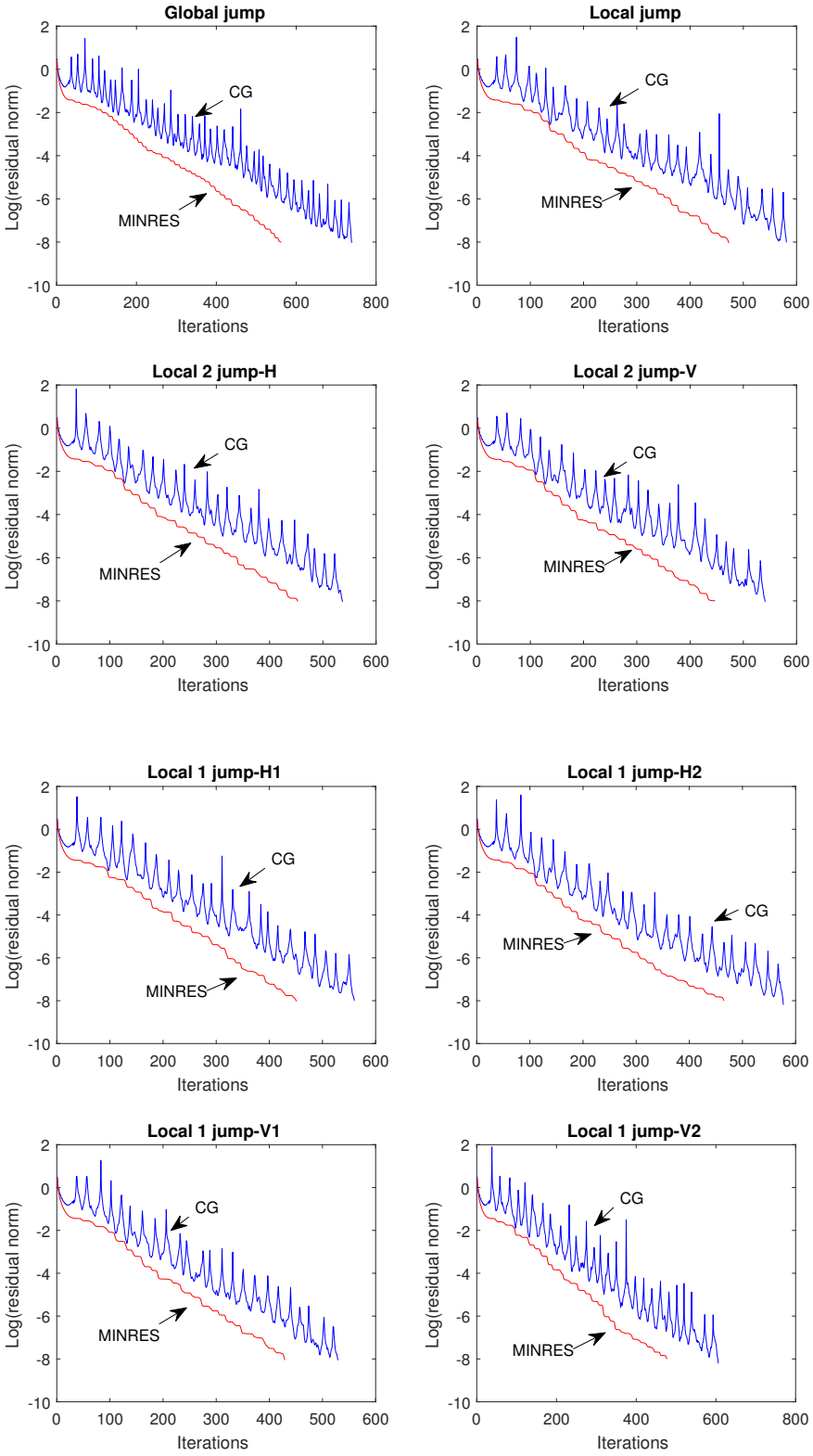

Figure 5.19: Residual reduction history for CG and MINRES algorithms with $\alpha = 0$.

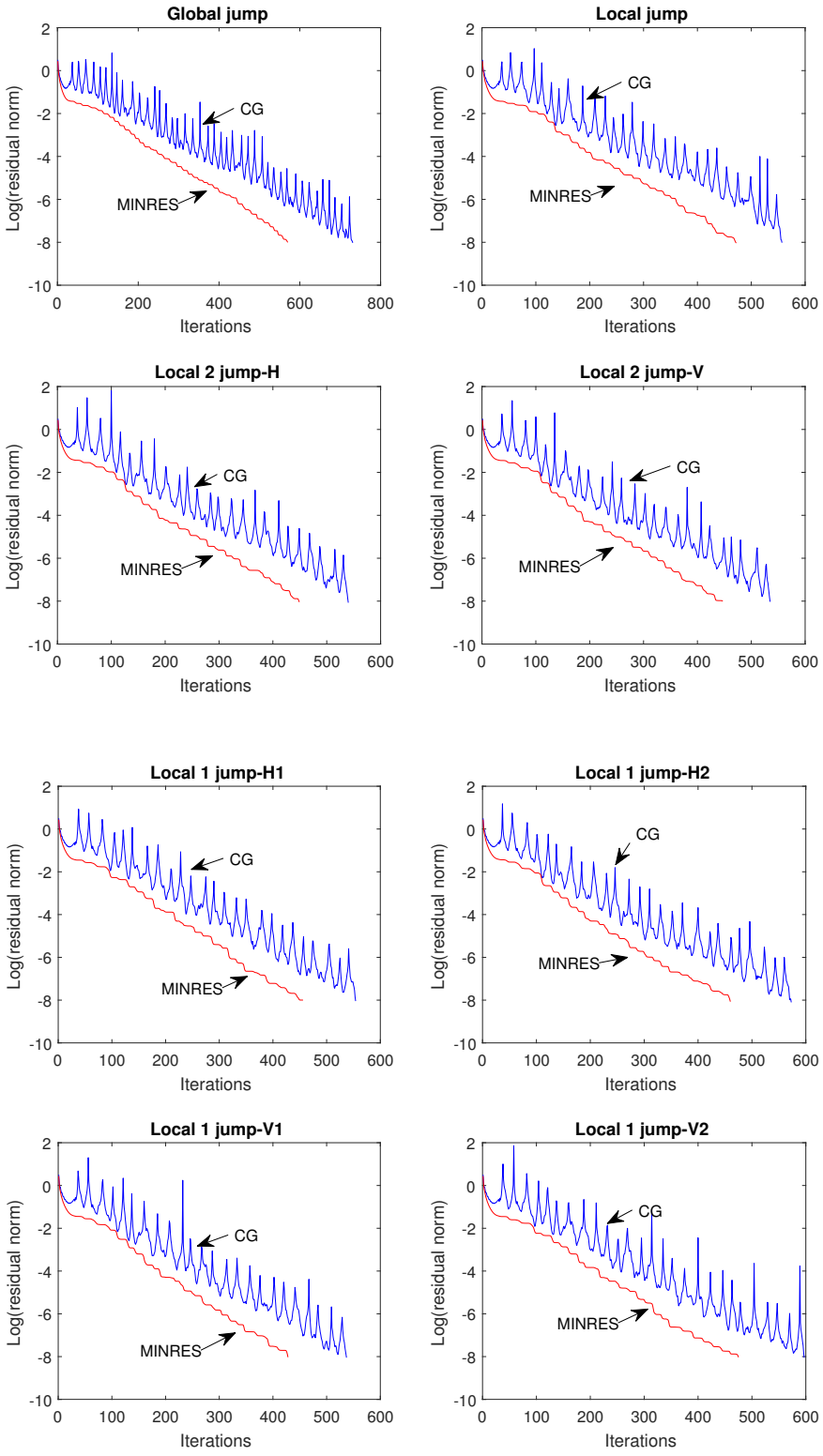

Figure 5.20: Residual reduction history for CG and MINRES algorithms with $\alpha = 1$.

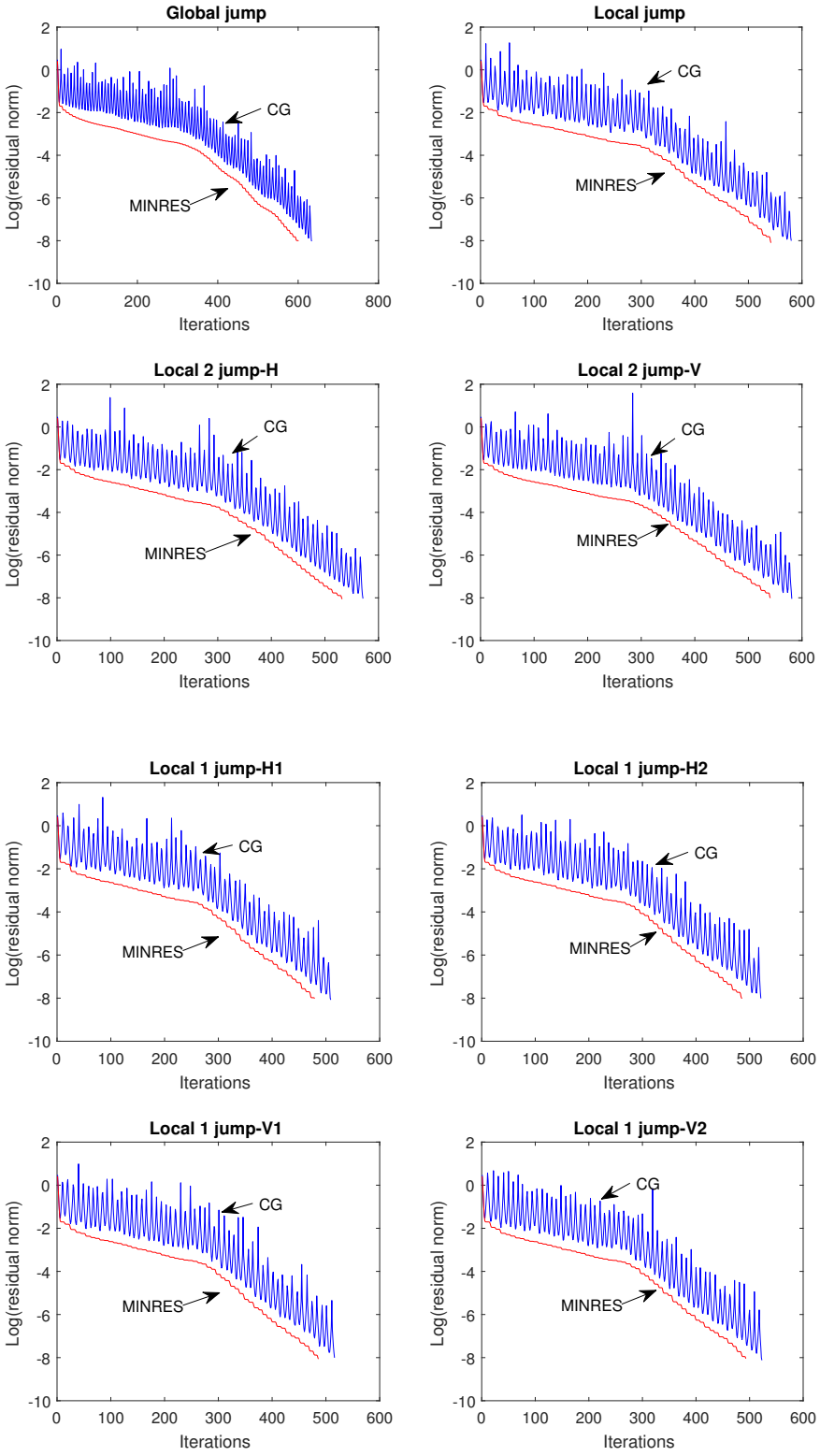

Figure 5.21: Residual reduction history for CG and MINRES algorithms with $\alpha = 1000$.

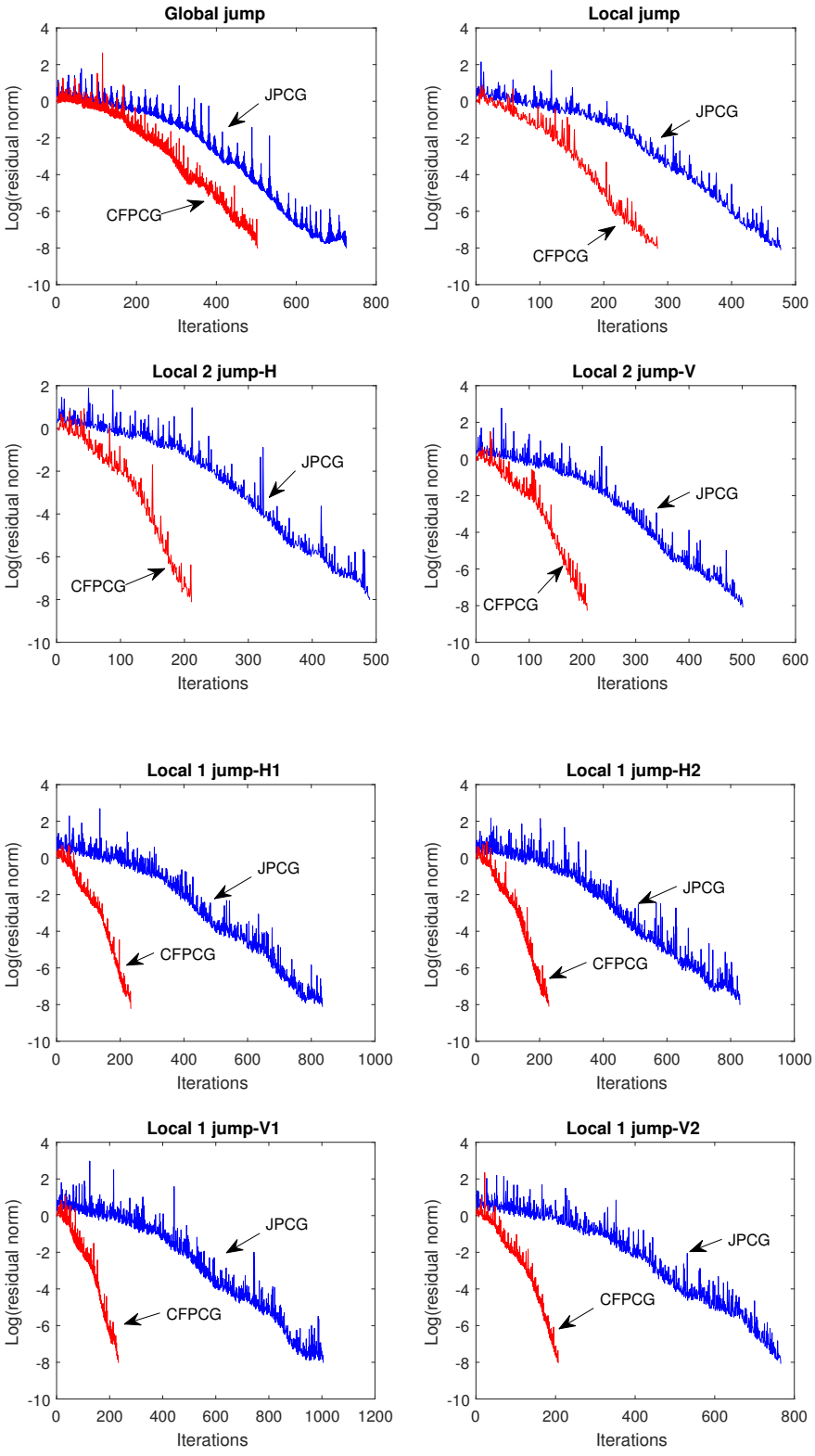

Figure 5.22: Residual reduction history for JPCG and CFPCG algorithms with $\alpha = 0$.

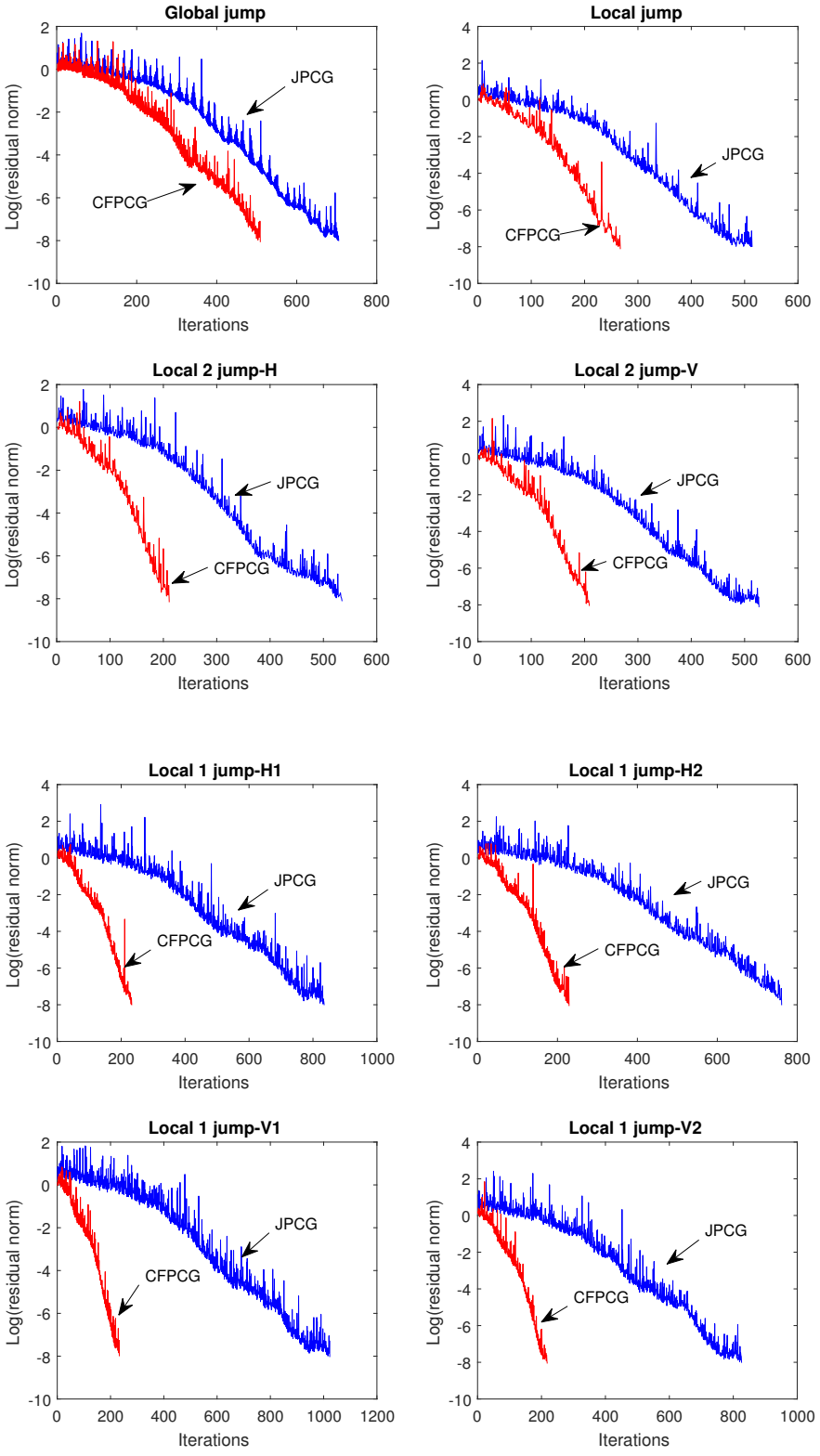

Figure 5.23: Residual reduction history for JPCG and CFPCG algorithms with $\alpha = 1$.

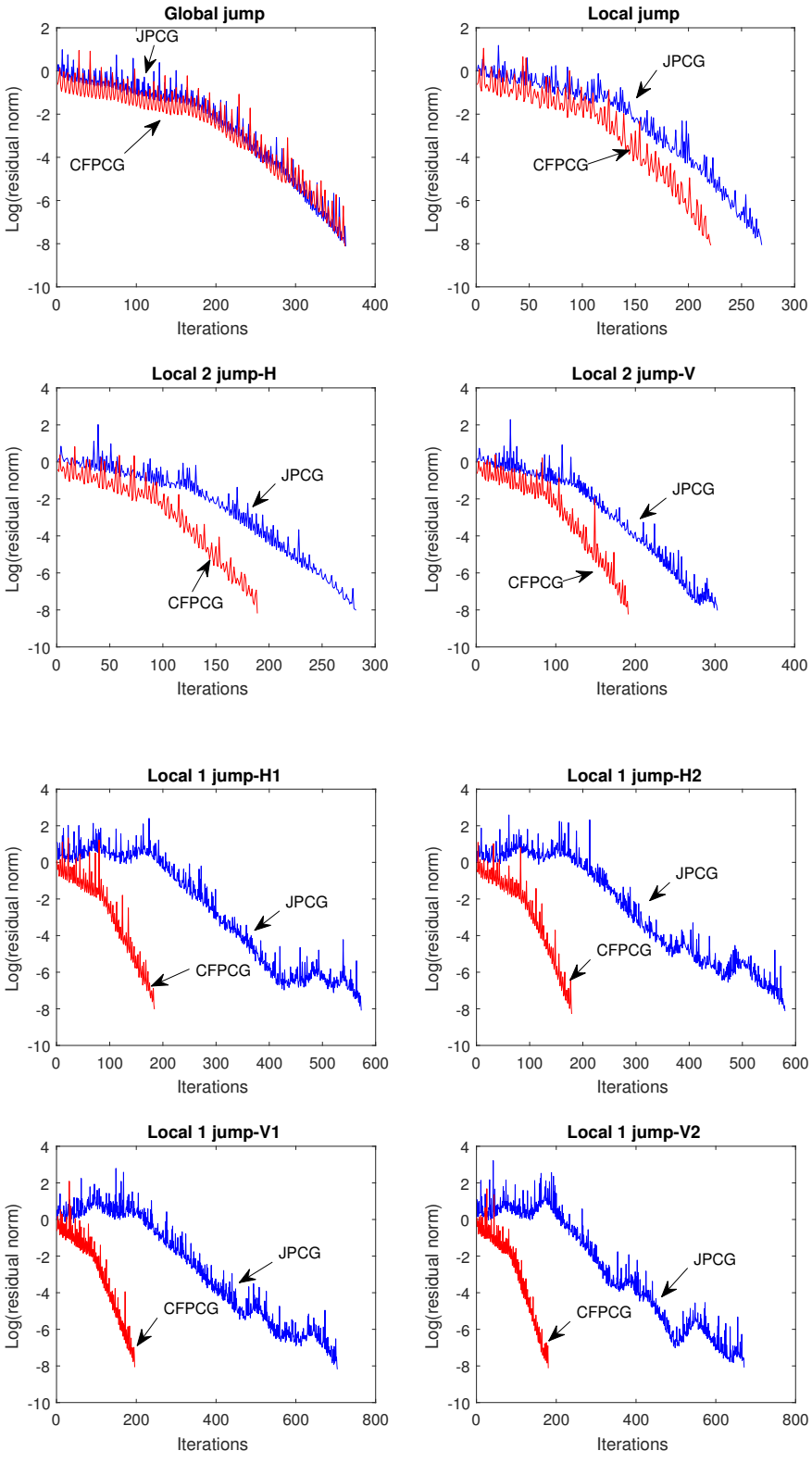

Figure 5.24: Residual reduction history for JPCG and CFPCG algorithms with $\alpha = 1000$.

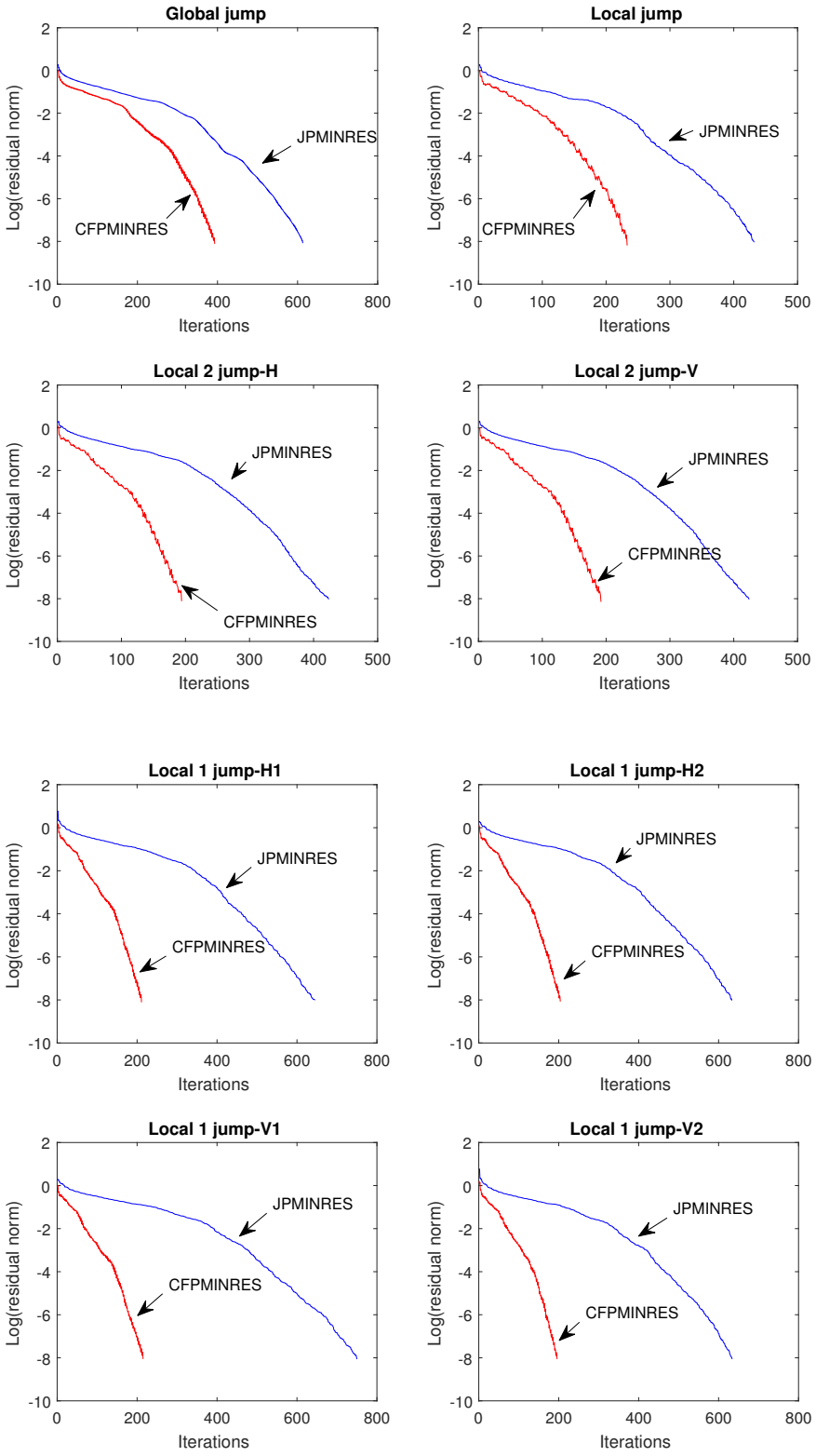

Figure 5.25: Residual reduction history for JPMINRES and CFPMINRES algorithms with $\alpha = 0$.

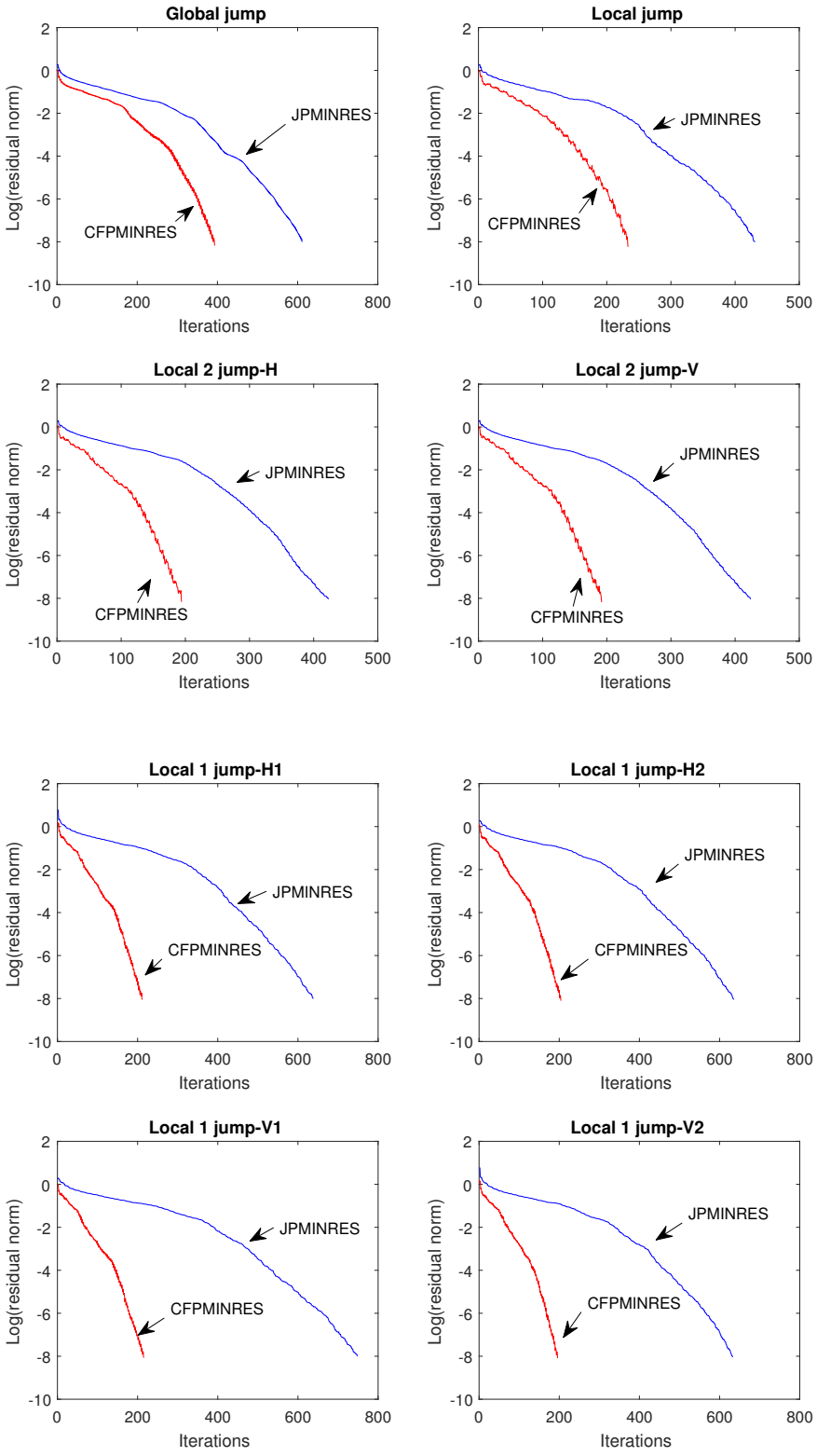

Figure 5.26: Residual reduction history for JPMINRES and CFPMINRES algorithms with $\alpha = 1$.

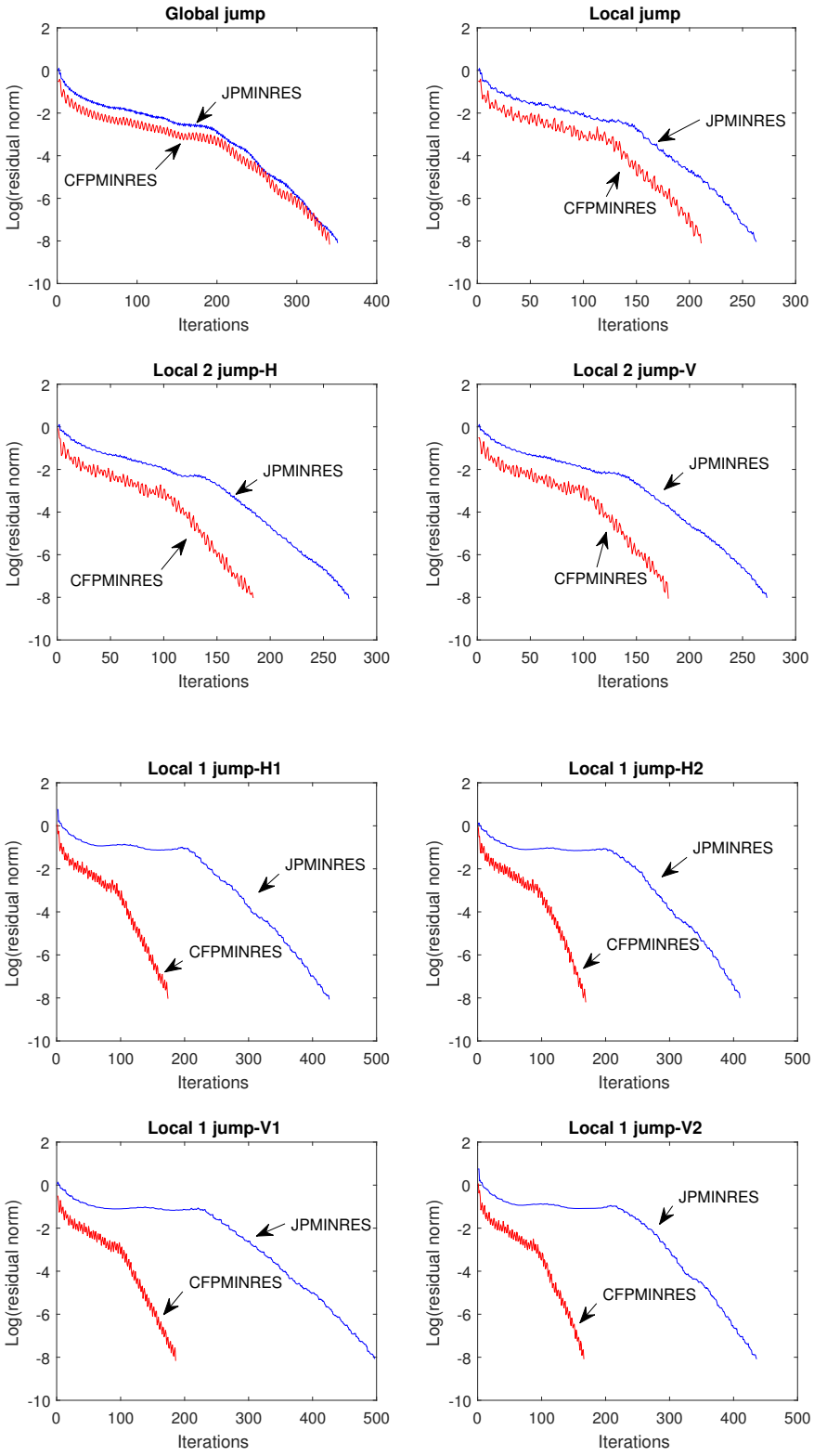

Figure 5.27: Residual reduction history for JPMINRES and CFPMINRES algorithms with $\alpha = 1000$.

Conclusion

In this work, the local jump stabilization method introduced in [23] and [31] has been first extended to the generalized Stokes problem and then reduced computationally attractive methods have been introduced, analyzed and tested on benchmark problems. Although the proposed schemes seem not very competitive for extremely large values of the parameter α, they can constitute good alternatives for moderately large values of this parameter. It is expected that their combined with adaptive grids could be a lot better. Furthermore, there is an obvious generalization to the equivalent three-dimensional case. On the other hand, the algebraic systems resulting from these discretizing procedures seem to be satisfactorily solved by some iterative solvers of conjugate gradient type with a superiority of the preconditioned forms.
As future work, it would be nice to continue the development of the present stabilization techniques so as to merge the procedures within existing fluid flow packages. Research remains also to be done in further extending their applicability to the Oseen and to the fully nonlinear Navier-Stokes equations. Likewise, future attention deserves to be given to the development of fast iterative solvers which exploit the structure of discrete systems generated by the presently studied numerical stabilizing schemes.

Bibliography

[1] M. Ainsworth, G.R. Barrenechea and A.Wachtel, *Stabilization of high aspect ratio mixed finite elements for incompressible flow*, SIAM J. Numer Anal. **53** (2015), no. 2, 1107-1120.

[2] I. Babuška, *Error bounds for finite element method*, Numerische Mathematik. **16** (1971), 322-333.

[3] J. M. Boland and R. A. Nicolaïdes, *Stability of finite elements under divergence constraints*, SIAM J. Numer. Anal. **20** (1983), no. 4, 722-731.

[4] J. M. Boland and R. A. Nicolaïdes, *On the stability of bilinear-constant velocity-pressure finite elements*, Num. Math. **44** (1984), no. 2, 219-222.

[5] J. M. Boland and R. A. Nicolaïdes, *Stable and semi-stable low order finite elements for viscous flows*, SIAM J. Numer. Anal. **22** (1985), no. 3, 474-492.

[6] F. Brezzi, *On the existence, uniqueness and approximation of saddle point problems*, R. A. I. R. O. Analyse Numerique. **8** (1974), 129-151.

[7] F. Brezzi and J. Pitkäranta, *On the stabilization of finite element approximations of the Stokes problem. Efficient solutions of elliptic systems*, edited by W. Hackbusch, Notes on Numerical Fluid Mechanics. **10** (1984), Springer Vieweg, Braunschwei.

[8] E. Burman and P. Hansbo, *A unified stabilized method for Stokes' and Darcy's equations*, J. Comput. Appl. Math. **198** (2007), no. 1, 35-51.

[9] J. Cahouet and J.P. Chabard, *Some fast 3D finite element solvers for the generalized Stokes problem*, Internat. J. Numer. Methods Fluids. **8** (1988), no. 8, 869-895.

[10] A. Chibani and N. Kechkar, *Minimal locally stabilized Q1-Q0 schemes for the generalized Stokes problem*, J. Korean Math. **57** (2020), no. 5, 1239–1266.

[11] P. Clément, *Approximation by finite elements using local regularization*, RAIRO Anal. Numér. **9** (1975), no. R2, 77-84.

[12] M. Crouzeix and P. A. Raviart, *Conforming and nonconforming finite element methods for solving the stationary Stokes equations*, RAIRO Anal. Numér. **7** (1973), no. 3, 33-76.

[13] A. Diar, *Résolution numérique du problème généralisé de Stokes*, Thèse de Magister, Mentouri University, Constantine, (2001).

[14] H. Elman, D. Silvester and A. Wathen, *Finite elements and fast iterative solvers with applications in incompressible fluid dynamics*, Oxford Science Publications, New York (2005).

[15] F. Ghadi, V. Ruas and M. Wakrim, *Finite element solution of a stream function-vorticity system and its application to the Navier Stokes equations*, Appl. Math. **4** (2013), 257-262.

[16] V. Girault and P. A. Raviart, *Finite element methods for Navier-Stokes equations. Theory and algorithms*, Springer-Verlag, Berlin, (1986).

[17] G. H. Golub and C. F. V. Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore, third ed., (1996).

[18] Y. He, J. Li and X. Yang, *Two-level penalized finite element methods for the stationary Navier-Stokes equations*, Int. J. Inf. Syst. Sci. **2** (2006), no. 1, 131–143.

[19] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand, **49**, (1952), no. 6, 409-436.

[20] S. Hong, K. Kim and S. Lee, *Modified cross-grid finite elements for the Stokes problem*, Appl. Math. Lett **16** (2003), 59-64.

[21] T. J. R. Hughes and L. P. Franca, *A new finite element formulation for CFD: VII. The Stokes problem with various well-posed boundary conditions: symmetric formulations that converge for all velocity/pressure spaces*, Comput. Methods Appl. Mech. Engrg. **65** (1987), no. 1, 85-96.

[22] N. Kechkar, *Analysis and application of locally stabilised mixed finite element methods*, PhD thesis, Victoria University, Manchester (1989).

[23] N. Kechkar and D.J. Silvester, *Analysis of locally stabilized mixed finite element methods for the Stokes problem*, Math. Comp. **58** (1992), no. 197, 1-10.

[24] J. Li, J. Wang and X. Ye, *Super convergence by L2-projections for stabilized finite element methods for the Stokes equations*, Int. J. Num. Anal. Model. **6** (2009), no. 4, 711-723.

[25] K. Nafa, *Improved local projection for the generalized Stokes problem*, Adv. Appl. Math. Mech. **1** (2009), no. 6, 862-873.

[26] C. Paige and M. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. **12** (1975), 617-629.

[27] Y. Saad, *Iterative methods for sparse linear systems*, SIAM. (2003).

[28] R. L. Sani, P. M. Gresho, R. L. Lee and D.F. Griffiths, *The cause and cure (?) of the spurious pressures generated by certain fem solutions of the incompressible Navier-Stokes equations: Part 1*, Internat. J. Numer. Methods Fluids **1** (1981), no. 1, 17-43.

[29] R. L. Sani, P. M. Gresho, R. L. Lee, D.F. Griffiths and M. Engelman, *The cause and cure (!) of the spurious pressures generated by certain fem solutions of the incompressible Navier-Stokes equations: Part 2*, Internat. J. Numer. Methods Fluids **1** (1981), no. 2, 171-204.

[30] Y. Shang, *A parallel finite element algorithm for simulation of the generalized Stokes problem*, Bull. Korean Math. Soc. **53** (2016), no. 3, 853-874.

[31] D. J. Silvester and N. Kechkar, *Stabilised bilinear-constant velocity-pressure finite elements for the conjugate gradient solution of the Stokes problem*, Comput. Methods Appl. Mech. Engrg. **79** (1990), no. 1, 71-86.

[32] R. Stenberg, *Analysis of mixed finite elements for the Stokes problem: A unified approach*, Math. Comp. **42** (1984), no. 165, 9-23.