Finite element approximation of the Stokes problem

Ramon Codina

1. Introduction and problem statement
2. Basic theory for the continuous problem
3. Mixed approximations
4. Optimization problem and penalty methods
5. Stabilized formulations

1. **Introduction and problem statement**
   1.1 Variational formulation
   1.2 Associated optimization problem
2. Basic theory for the continuous problem
3. Mixed approximations
4. Optimization problem and penalty methods
5. Stabilized formulations

Boundary value problem

The Stokes problem in differential form consists in finding a velocity (or displacement) $\boldsymbol{u}$ and a pressure p such that

$$
\begin{aligned}
-\nu\Delta\boldsymbol{u} + \nabla p &= \boldsymbol{f} && \text{in } \Omega \\
\nabla\cdot\boldsymbol{u} &= 0 && \text{in } \Omega \\
\boldsymbol{u} &= \boldsymbol{0} && \text{on } \partial\Omega
\end{aligned}
$$

where:

$\boldsymbol{f}$: Vector of body forces

ν : Kinematic viscosity (fluids) or shear modulus (solids)

$\Omega \subset \mathbb{R}^d$: Computational domain

Weak form 1

Let

$V = H_0^1(\Omega)^d$	Velocity space
$Q = L^2(\Omega)/\mathbb{R}$	Pressure space
$(\cdot,\cdot)$	Inner product in Q
$\langle\cdot,\cdot\rangle$	Duality pairing in $V \times V'$, $V' = H^{-1}(\Omega)^d$

The weak form of the problem is: Find $\boldsymbol{u} \in V$, $p \in Q$ such that

$$\begin{aligned} \nu(\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) - (p, \nabla\cdot\boldsymbol{v}) &= \langle \boldsymbol{f}, \boldsymbol{v}\rangle \qquad && \forall \boldsymbol{v} \in V \\ (q, \nabla\cdot\boldsymbol{u}) &= 0 && \forall q \in Q \end{aligned}$$

Defining

$$\begin{aligned} a &: V \times V \longrightarrow \mathbb{R}, & a(\boldsymbol{u}, \boldsymbol{v}) &= \nu(\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) \\ b &: Q \times V \longrightarrow \mathbb{R}, & b(q, \boldsymbol{v}) &= (q, \nabla \cdot \boldsymbol{v}) \\ l &: V \longrightarrow \mathbb{R}, & l(v) &= \langle \boldsymbol{f}, \boldsymbol{v} \rangle \end{aligned}$$

the problem can be written as: Find $\boldsymbol{u} \in V$, $p \in Q$ such that

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) - b(p, \boldsymbol{v}) &= l(\boldsymbol{v}) & \forall \boldsymbol{v} \in V \\ b(q, \boldsymbol{u}) &= 0 & \forall q \in Q \end{aligned}$$

Not that this problem is not symmetric (but can be symmetrized).

Defining

$$B : (V \times Q) \times (V \times Q) \longrightarrow \mathbb{R}$$
$$B([\boldsymbol{u}, p], [\boldsymbol{v}, q]) := a(\boldsymbol{u}, \boldsymbol{v}) - b(p, \boldsymbol{v}) + b(q, \boldsymbol{u})$$
$$L : V \times Q \longrightarrow \mathbb{R}$$
$$L([\boldsymbol{v}, q]) = l(\boldsymbol{v})$$

it can also be written as: Find $[\boldsymbol{u}, p] \in V \times Q$ such that

$$B([\boldsymbol{u}, p], [\boldsymbol{v}, q]) = L([\boldsymbol{v}, q]) \quad \forall [\boldsymbol{v}, q] \in V \times Q$$

Weak form 2

Define the space

$$J := \{\boldsymbol{v} \in V \mid \nabla \cdot \boldsymbol{v} = 0\}$$

The weak form of the problem can also be stated as: Find $\boldsymbol{u} \in J$ such that

$$\nu(\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle \qquad \forall \boldsymbol{v} \in J$$

or

$$a(\boldsymbol{u}, \boldsymbol{v}) = l(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in J$$

Optimization associated to weak form 2

Let us define the quadratic functional

$$
\begin{aligned}
F :& J \longrightarrow \mathbb{R} \\
F(\boldsymbol{v}) :=& \frac{1}{2}\nu(\nabla \boldsymbol{v}, \nabla \boldsymbol{v}) - \langle \boldsymbol{f}, \boldsymbol{v} \rangle = \frac{1}{2}a(\boldsymbol{u}, \boldsymbol{v}) - l(\boldsymbol{v})
\end{aligned}
$$

If $\boldsymbol{u} \in J$ is the solution of the Stokes problem, then

$$
\boldsymbol{u} = \arg\inf_{\boldsymbol{v} \in J} F(\boldsymbol{v})
$$

since it is immediately checked that

$$
\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}\right|_{\epsilon=0} F(\boldsymbol{u} + \epsilon \boldsymbol{v}) &= a(\boldsymbol{u}, \boldsymbol{v}) - l(\boldsymbol{v}) = 0 \\
\left.\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\right|_{\epsilon=0} F(\boldsymbol{u} + \epsilon \boldsymbol{v}) &= a(\boldsymbol{v}, \boldsymbol{v}) \geq 0
\end{aligned}
$$

Optimization associated to weak form 1

The Lagrangian associated to the condition $\nabla \cdot \boldsymbol{u} = 0$ is

$$L : V \times Q \longrightarrow \mathbb{R}$$

$$L(\boldsymbol{v}, q) := \frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v}) - b(q, \boldsymbol{v}) - l(\boldsymbol{v})$$

If $[\boldsymbol{u}, p]$ is the solution of weak form 1, then

$$[\boldsymbol{u}, p] = \arg \inf_{\boldsymbol{v} \in V} \sup_{q \in Q} L(\boldsymbol{v}, q)$$

since it is immediately checked that

$$\left.\frac{\partial}{\partial \epsilon_1}\right|_{\epsilon_1 = 0, \epsilon_2 = 0} L(\boldsymbol{u} + \epsilon_1 \boldsymbol{v}, p + \epsilon_2 q) = a(\boldsymbol{u}, \boldsymbol{v}) - b(p, \boldsymbol{v}) - l(\boldsymbol{v}) = 0$$

$$\left.\frac{\partial}{\partial \epsilon_2}\right|_{\epsilon_1 = 0, \epsilon_2 = 0} L(\boldsymbol{u} + \epsilon_1 \boldsymbol{v}, p + \epsilon_2 q) = b(q, \boldsymbol{u}) = 0$$

Penalized "Lagrangian"

Instead of introducing a Lagrange multiplier, one can penalize the constraint $\nabla \cdot \boldsymbol{u} = 0$. Define

$$\begin{aligned} L_1^\epsilon &: V \longrightarrow \mathbb{R} \\ L_1^\epsilon(\boldsymbol{v}) &:= \frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v}) - l(v) + \frac{1}{2\epsilon} \|\nabla \cdot \boldsymbol{v}\|^2 \end{aligned}$$

with $\epsilon > 0$. The associated variational problem is: Find $\boldsymbol{u}^\epsilon \in V$ such that

$$a(\boldsymbol{u}^\epsilon, \boldsymbol{v}) + \frac{1}{\epsilon}(\nabla \cdot \boldsymbol{u}^\epsilon, \nabla \cdot \boldsymbol{v}) = l(\boldsymbol{v}) \quad \forall \boldsymbol{v} \in V$$

and the associated boundary value problem:

$$\begin{aligned} -\nu \Delta \boldsymbol{u}^\epsilon - \frac{1}{\epsilon} \nabla(\nabla \cdot \boldsymbol{u}^\epsilon) &= \boldsymbol{f} && \text{in } \Omega \\ \boldsymbol{u}^\epsilon &= \boldsymbol{0} && \text{on } \partial\Omega \end{aligned}$$

Perturbed Lagrangian

In fact, the optimization of the Lagrangian with respect to the multiplier is not defined (neither positive nor negative). One can define the perturbed Lagrangian:

$$L_2^\epsilon : V \times Q \longrightarrow \mathbb{R}$$

$$L_2^\epsilon(\boldsymbol{v}, q) := \frac{1}{2} a(\boldsymbol{v}, \boldsymbol{v}) - b(q, \boldsymbol{v}) - l(\boldsymbol{v}) - \epsilon \|q\|^2$$

with $\epsilon > 0$. Now it holds:

$$\left.\frac{\partial^2}{\partial \epsilon_1^2}\right|_{\epsilon_1=0,\epsilon_2=0} L_2^\epsilon(\boldsymbol{u} + \epsilon_1 \boldsymbol{v}, p + \epsilon_2 q) = a(\boldsymbol{v}, \boldsymbol{v}) \geq 0$$

$$\left.\frac{\partial^2}{\partial \epsilon_2^2}\right|_{\epsilon_1=0,\epsilon_2=0} L_2^\epsilon(\boldsymbol{u} + \epsilon_1 \boldsymbol{v}, p + \epsilon_2 q) = -\epsilon \|q\|^2 \leq 0$$

so the problem is really a saddle point problem.

The associated variational problem is: Find $[\boldsymbol{u}^\epsilon, p^\epsilon] \in V \times Q$ such that

$$\begin{aligned} a(\boldsymbol{u}^\epsilon, \boldsymbol{v}) - b(p^\epsilon, \boldsymbol{v}) &= l(\boldsymbol{v}) && \forall \boldsymbol{v} \in V \\ b(q, \boldsymbol{u}^\epsilon) + \epsilon(p^\epsilon, q) &= 0 && \forall q \in Q \end{aligned}$$

and the associated boundary value problem:

$$\begin{aligned} -\nu \Delta \boldsymbol{u}^\epsilon + \nabla p^\epsilon &= \boldsymbol{f} && \text{in } \Omega \\ \epsilon p^\epsilon + \nabla \cdot \boldsymbol{u}^\epsilon &= 0 && \text{in } \Omega \\ \boldsymbol{u}^\epsilon &= \boldsymbol{0} && \text{on } \partial\Omega \end{aligned}$$

Remarks

- Note that if q is constant:

$$\epsilon \int_\Omega p^\epsilon + \int_\Omega \nabla \cdot \boldsymbol{u}^\epsilon = \epsilon \int_\Omega p^\epsilon = 0$$

  and thus the pressure space is

$$Q = \left\{ q \in L^2(\Omega) \mid \int_\Omega q = 0 \right\} \cong L^2(\Omega)/\mathbb{R}$$

- At the differential level, the penalized and the perturbed Lagrangian lead to the same BVP:

$$\left. \begin{aligned} -\nu \Delta \boldsymbol{u}^\epsilon + \nabla p^\epsilon &= \boldsymbol{f} \\ \epsilon p^\epsilon + \nabla \cdot \boldsymbol{u}^\epsilon &= 0 \end{aligned} \right\} \Leftrightarrow -\nu \Delta \boldsymbol{u}^\epsilon - \frac{1}{\epsilon} \nabla (\nabla \cdot \boldsymbol{u}^\epsilon) = \boldsymbol{f}$$

- In elasticity

$$\frac{1}{\epsilon} = \frac{E}{1 - 2\nu} = K \quad \text{Bulk modulus}$$

Objectives

- Theoretical: existence and uniqueness of $\boldsymbol{u}$ and p.
- Theoretical: convergence of $\boldsymbol{u}^\epsilon$ and p^ϵ to $\boldsymbol{u}$ and p as $\epsilon \to 0$.
- Approximation of V by $V_h \subset V$ and Q by $Q_h \subset Q$. Conditions on Q_h and V_h.
- Is it possible to avoid the conditions on Q_h and V_h?

Other non-touched (and important!) issues:

- Is it possible to approximate J by $J_h \subset J$? Solenoidal velocity interpolations.
- What about approximations where $V_h \not\subset V$ and/or $Q_h \not\subset Q$? Non-conforming approximations and, in particular, discontinuous Galerkin methods.

1. Introduction and problem statement
2. **Basic theory for the continuous problem**
3. Mixed approximations
4. Optimization problem and penalty methods
5. Stabilized formulations

Existence and uniqueness in J

Theorem. There exists a unique $\boldsymbol{u} \in J$ such that

$$a(\boldsymbol{u}, \boldsymbol{v}) = l(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in J$$

PROOF: Use Lax-Milgram's Lemma (a is coercive). □

REMARK. Sometimes it is convenient to define $a(\boldsymbol{u}, \boldsymbol{v})$ as

$$a(\boldsymbol{u}, \boldsymbol{v}) = \nu \int_\Omega \nabla^S \boldsymbol{u} : \nabla^S \boldsymbol{v}, \quad \nabla^S \boldsymbol{v} := \frac{1}{2}\left[\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T\right]$$

In this case, coercivity of a follows from Korn's inequality instead of Poincaré-Friedrics inequality.

Helmholtz decomposition

Theorem. Given a vector field $\boldsymbol{w} \in L^2(\Omega)^d$, there exists a unique decomposition

$$
\begin{aligned}
&\boldsymbol{w} = \boldsymbol{v} + \nabla q \\
&\boldsymbol{v} \in H(\mathrm{div}, \Omega),\ \nabla \cdot \boldsymbol{v} = 0 \text{ in } \Omega,\ \boldsymbol{v} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \\
&q \in H^1(\Omega)/\mathbb{R},\ \int_\Omega \boldsymbol{v} \cdot \nabla q = 0
\end{aligned}
$$

PROOF: If the decomposition exists, it is easily checked that it is orthogonal:

$$\int_\Omega \boldsymbol{v} \cdot \nabla q = \int_\Omega \nabla \cdot (\boldsymbol{v} q) = \int_{\partial\Omega} \boldsymbol{v} \cdot \boldsymbol{n} q = 0$$

and from this it follows that it is unique:

$$\boldsymbol{w} = \boldsymbol{v}_1 + \nabla q_1 = \boldsymbol{v}_2 + \nabla q_2 \Rightarrow \boldsymbol{v}_1 - \boldsymbol{v}_2 = \nabla q_1 - \nabla q_2$$
$$\Rightarrow \boldsymbol{v}_1 = \boldsymbol{v}_2, \; q_1 = q_2 + \text{constant}$$

Next, it can be seen that q and $\boldsymbol{v}$ can be found as the solution of

$$-\Delta q = \nabla \cdot \boldsymbol{w} \quad \text{in } \Omega, \qquad \frac{\partial q}{\partial n} = 0 \quad \text{on } \partial\Omega$$
$$\boldsymbol{v} = \boldsymbol{w} - \nabla q$$

Existence and uniqueness in $V \times Q$

Theorem. If $N : V \longrightarrow \mathbb{R}$ is linear and continuous and $N(\boldsymbol{v}) = 0$ for all $\boldsymbol{v} \in J$ then there exists $p \in L^2(\Omega)/\mathbb{R}$ such that $N(\boldsymbol{v}) = (p, \nabla \cdot \boldsymbol{v})$.

PROOF: Use a representation theorem to show that $N(\boldsymbol{v}) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle$ and the Helmholtz decomposition of $\boldsymbol{w}$. □

Theorem. There exists a unique $[\boldsymbol{u}, p] \in V \times Q$ solution of the Stokes problem in velocity-pressure form.

PROOF: Define $N : V \longrightarrow \mathbb{R}$ by $N(\boldsymbol{v}) = a(\boldsymbol{u}, \boldsymbol{v}) - l(\boldsymbol{v})$. Then $N(\boldsymbol{v}) = 0$ for $\boldsymbol{v} \in J$ and thus there exists $p \in L^2(\Omega)/\mathbb{R}$ such that $a(\boldsymbol{u}, \boldsymbol{v}) - l(\boldsymbol{v}) = (p, \nabla \cdot \boldsymbol{v})$. □

Abstract problem and proof by regularization

Theorem. Let V, Q be Hilbert spaces, $a : V \times V \longrightarrow \mathbb{R}$ continuous and V-elliptic:

$$a(\boldsymbol{u}, \boldsymbol{v}) \leq N_a \|\boldsymbol{u}\|_V \|\boldsymbol{v}\|_V, \quad a(\boldsymbol{v}, \boldsymbol{v}) \geq K_a \|\boldsymbol{v}\|_V^2$$

$b : Q \times V \longrightarrow \mathbb{R}$ continuous and satisfying the inf-sup condition

$$\inf_{q \in Q} \sup_{\boldsymbol{v} \in V} \frac{b(q, \boldsymbol{v})}{\|q\|_Q \|\boldsymbol{v}\|_V} \geq K_b$$

and $l : V \longrightarrow \mathbb{R}$ continuous: $l(\boldsymbol{v}) \leq N_l \|\boldsymbol{v}\|_V$. Then, there exists a unique solution to the problem of finding $[\boldsymbol{u}, p] \in V \times Q$ such that

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) - b(p, \boldsymbol{v}) &= l(\boldsymbol{v}) & \forall \boldsymbol{v} \in V \\ b(q, \boldsymbol{u}) &= 0 & \forall q \in Q \end{aligned}$$

PROOF: Consider the regularized problem

$$\begin{aligned} a(\boldsymbol{u}^\epsilon, \boldsymbol{v}) - b(p^\epsilon, \boldsymbol{v}) &= l(\boldsymbol{v}) && \forall \boldsymbol{v} \in V \\ \epsilon(p^\epsilon, q) + b(q, \boldsymbol{u}^\epsilon) &= 0 && \forall q \in Q \end{aligned}$$

Step 1. Lax Milgram's Theorem guarantees that there exists a unique $[\boldsymbol{u}^\epsilon, p^\epsilon]$ for all $\epsilon > 0$.
Step 2. It is easily shown that $\|\boldsymbol{u}^\epsilon\|_V$ and $\|p^\epsilon\|_Q$ are bounded independently of ϵ (use is made of the inf-sup condition).
Step 3. As a consequence, $\boldsymbol{u}^\epsilon \rightharpoonup \boldsymbol{u}^*$, $p^\epsilon \rightharpoonup p^*$, as $\epsilon \to 0$, and $[\boldsymbol{u}^*, p^*]$ turns out to be solution of the problem with $\epsilon = 0$.
Step 4. Uniqueness is easy (again, the inf-sup condition is needed). □

1. Introduction and problem statement
2. Basic theory for the continuous problem
3. **Mixed approximations**
   3.1 The inf-sup condition
   3.2 Discrete problem
   3.3 Admissible elements and convergence theorem
   3.4 Mixed stress-velocity-pressure formulation
4. Optimization problem and penalty methods
5. Stabilized formulations

Babuška's Theorem (Babuška-Banach-Nečas)

Theorem. Let $B : V \times W \longrightarrow \mathbb{R}$, $L : W \longrightarrow \mathbb{R}$ and consider the abstract variational problem: Find $u \in V$ such that

$$B(u, v) = L(v) \quad \forall v \in W$$

Consider also the following conditions:

$$B(u, v) \leq N_B \|u\|_V \|v\|_W \qquad \text{Continuity of } B$$

$$L(v) \leq N_L \|v\|_W \qquad \text{Continuity of } L$$

$$\inf_{u \in V} \sup_{v \in W} \frac{B(u, v)}{\|u\|_V \|v\|_W} \geq K_B \qquad V\text{-Stability of } B$$

Then, these are necessary and sufficient conditions for the variational problem to have a unique solution that depends continuously on the data as $\|u\|_V \leq \frac{N_L}{K_B}$

Coercivity vs V-stability

If $B : V \times V \longrightarrow \mathbb{R}$ and $B(v, v) \geq K_B \|v\|_V^2$, the previous theorem reduces to Lax-Milgram's Lemma. Coercivity is sufficient, but not necessary for the problem to be well posed. The following analogy holds:

Finite dimensional problem:		Variational problem:
$v^T B u = v^T f$		$B(u, v) = L(v)$
Matrix B positive definite	$\leftrightarrow$	Bilinear form B coercive
Matrix B non-singular	$\leftrightarrow$	Bilinear form B V-stable

Why Ladyzhenskaya-Babuška-Brezzi ?

Consider the abstract form of the Stokes problem

$$B([\boldsymbol{u},p],[\boldsymbol{v},q]) = L([\boldsymbol{v},q]) \quad \forall [\boldsymbol{v},q] \in V \times Q$$
$$B([\boldsymbol{u},p],[\boldsymbol{v},q]) := a(\boldsymbol{u},\boldsymbol{v}) - b(p,\boldsymbol{v}) + b(q,\boldsymbol{u}), \quad L([\boldsymbol{v},q]) = l(\boldsymbol{v})$$

- If a is coercive in V, B is $V \times Q$-stable if, and only if (Brezzi):

$$\inf_{q \in Q} \sup_{\boldsymbol{v} \in V} \frac{b(q,\boldsymbol{v})}{\|q\|_Q \|\boldsymbol{v}\|_V} \geq K_b > 0$$

- If $b(q,\boldsymbol{v}) = (p, \nabla \cdot \boldsymbol{v})$, $Q = L^2(\Omega)/\mathbb{R}$, $V = H^1_0(\Omega)^d$, the previous condition holds true (Ladyzhenskaya).

Discrete variational form

Let $\mathcal{T} = \{K\}$ be a finite element partition of Ω of size h and consider the finite element spaces:

$$V_h := \left\{ \boldsymbol{v}_h \in C^0(\Omega)^d \mid \boldsymbol{v}_h|_K \in R_k(K)^d \right\}$$
$$Q_h := \left\{ q_h \in L^2(\Omega)/\mathbb{R} \mid q_h|_K \in R_l(K) \right\}$$

where $R_p(K)$ is the space of complete polynomials of degree p on K. By construction

$$V_h \subset V, \quad Q_h \subset Q \quad \text{Conforming approximation}$$

The discrete Stokes problem is: Find $\boldsymbol{u}_h \in V_h$, $p_h \in Q_h$ such that

$$\begin{aligned} a(\boldsymbol{u}_h, \boldsymbol{v}_h) - b(p_h, \boldsymbol{v}_h) &= l(\boldsymbol{v}_h) & \forall \boldsymbol{v}_h \in V_h \\ b(q_h, \boldsymbol{u}_h) &= 0 & \forall q_h \in Q_h \end{aligned}$$

Matrix version

The matrix form of this problem is:

$$\begin{aligned} \mathsf{K}\mathsf{U} - \mathsf{D}^T\mathsf{P} &= \mathsf{F} \\ \mathsf{D}\mathsf{U} &= 0 \end{aligned}$$

where U and P are the arrays of nodal velocities and pressures, respectively and

$$\begin{aligned} \mathsf{K}_{ij}^{ab} &= \nu(\nabla N_u^a, \nabla N_u^b)\delta_{ij} \\ \mathsf{D}_j^{cb} &= (N_p^c, \partial_j N_u^b) \end{aligned}$$

N_u and N_p denote the velocity and pressure shape functions, respectively, and the superscript the node to which they are associated.

Stability and discrete inf-sup condition

Taking $\boldsymbol{v}_h = \boldsymbol{u}_h$, $q_h = p_h$ it follows that:

$$K_a \|\boldsymbol{u}_h\|_V^2 \leq a(\boldsymbol{u}_h, \boldsymbol{u}_h) = l(\boldsymbol{u}_h) \leq N_l \|\boldsymbol{u}_h\|_V \Rightarrow \|\boldsymbol{u}_h\|_V \leq \frac{N_l}{K_a}$$

Pressure stability can only be obtained from the only place where pressure occurs:

$$b(p_h, \boldsymbol{v}_h) = l(\boldsymbol{v}_h) - a(\boldsymbol{u}_h, \boldsymbol{v}_h) \leq \left(N_l + N_a \frac{N_l}{K_a} \right) \|\boldsymbol{v}_h\|_V$$

The only way to control the pressure is to guarantee that

$$\forall p_h \in Q_h \; \exists \; \boldsymbol{v}_h \in V_h \; \text{ such that } \; K_b \|p_h\| \|\boldsymbol{v}_h\|_V \leq b(p_h, \boldsymbol{v}_h)$$

$$\iff \inf_{p_h \in Q_h} \sup_{\boldsymbol{v}_h \in V_h} \frac{b(p_h, \boldsymbol{v}_h)}{\|p_h\|_Q \|\boldsymbol{v}_h\|_V} \geq K_b$$

Compatible velocity-pressure pairs

Main problem: if $V_h \subset V$

$$\inf_{p\in Q}\sup_{\mathbf{v}\in V}\frac{b(p,\mathbf{v})}{\|p\|_Q\|\mathbf{v}\|_V}\geq K_b \not\Rightarrow \inf_{p_h\in Q_h}\sup_{\mathbf{v}_h\in V_h}\frac{b(p_h,\mathbf{v}_h)}{\|p_h\|_Q\|\mathbf{v}_h\|_V}\geq K_b'$$

Pairs V_h-Q_h satisfying the inf-sup condition will be called compatible, admissible or, in the present context, div-stable.

Methods to check compatibility:

- Verfürth (1984)
- Boland-Nicolaides (1983)
- Stenberg (1984). Macroelement technique.

Example to check the inf-sup condition: Fortin's trick

Theorem. If there exists an operator $\Pi_h : V \longrightarrow V_h$ such that

- $\forall \boldsymbol{v} \in V, \forall q_h \in Q_h, b(q_h, \boldsymbol{v} - \Pi_h \boldsymbol{v}) = 0$
- $\|\Pi_h \boldsymbol{v}\|_V \leq C \|\boldsymbol{v}\|_V$

then the pair V_h-Q_h satisfies the inf-sup condition.

PROOF:

$$\sup_{\boldsymbol{v}_h \in V_h} \frac{b(q_h, \boldsymbol{v}_h)}{\|\boldsymbol{v}_h\|_V} \geq \sup_{\boldsymbol{v} \in V} \frac{b(q_h, \Pi_h \boldsymbol{v})}{\|\Pi_h \boldsymbol{v}\|_V} = \sup_{\boldsymbol{v} \in V} \frac{b(q_h, \boldsymbol{v})}{\|\Pi_h \boldsymbol{v}\|_V}$$

$$\geq \frac{1}{C} \sup_{\boldsymbol{v} \in V} \frac{b(q_h, \boldsymbol{v})}{\|\boldsymbol{v}\|_V} \geq \frac{1}{C} K_b \|q_h\|_Q$$

Counting variables

If the discrete problem has a unique solution:

$$\mathsf{U} = \mathsf{K}^{-1}(\mathsf{F} + \mathsf{D}^T\mathsf{P})$$
$$\mathsf{D}\mathsf{K}^{-1}\mathsf{F} + \mathsf{D}\mathsf{K}^{-1}\mathsf{D}^T\mathsf{P} = 0$$

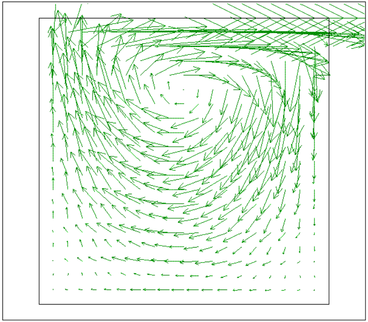
from where

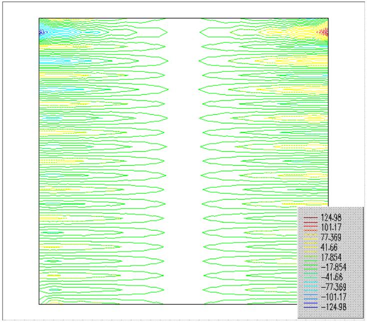
$$\mathrm{rank}(\mathsf{D}\mathsf{K}^{-1}\mathsf{D}^T) = n_p \leq \mathrm{rank}(\mathsf{K}^{-1}) = n_u$$

If there is a partition for which $n_p > n_u$, it is impossible to fulfill the inf-sup condition with K_b independent of h.

$n_p \leq n_u$ for all partitions is a necessary, but not sufficient condition for the inf-sup condition to hold.

Example: Galerkin method for cavity flow problem (Q_1/Q_1 interpolation)





Abstract convergence theorem

Consider the abstract variational problem: Find $u \in V$ such that

$$B(u, v) = L(v) \quad \forall v \in W,$$

Theorem. Suppose that the conditions of Babuška's theorem hold and, moreover,

- $\inf_{u_h \in V_h} \sup_{v_h \in W_h} \frac{B(u_h, v_h)}{\|u_h\|_V \|v_h\|_W} \geq K_B$: V_h-Stability of B
- $B(u - u_h, v_h) \leq C\varepsilon(h)\|v_h\|_W$: Consistency error
- $\inf_{\hat{u}_h \in V_h} \|u - \hat{u}_h\|_V \leq C\varepsilon(h)$: Interpolation error

for a certain error function $\varepsilon(h)$ (with $\varepsilon(h) \to 0$ as $h \to 0$). Then

$$\|u - u_h\|_V \leq C\varepsilon(h)$$

PROOF: For any $\hat{u}_h \in V_h$ there holds:

$$\begin{aligned} K_B \|\hat{u}_h - u_h\|_V \|v_h\|_W &\leq B(\hat{u}_h - u_h, v_h) \\ &= B(\hat{u}_h - u, v_h) + B(u - u_h, v_h) \\ &\leq N_B \|\hat{u}_h - u\|_V \|v_h\|_W + C\varepsilon(h)\|v_h\|_W \end{aligned}$$

On the other hand:

$$\begin{aligned} \|u - u_h\|_V &\leq \|u - \hat{u}_h\|_V + \|\hat{u}_h - u_h\|_V \\ &\leq \|u - \hat{u}_h\|_V + \frac{N_B}{K_B}\|\hat{u}_h - u\|_V + \frac{C}{K_B}\varepsilon(h) \end{aligned}$$

The result follows taking the inf over $\hat{u}_h \in V_h$. □

Application to the Stokes problem

In this case:

- $u_h = [\boldsymbol{u}_h, p_h]$
- $V_h = W_h \equiv V_h \times Q_h$ (Recall: $V \times Q = H_0^1(\Omega)^d \times L^2(\Omega)/\mathbb{R}$)
- $B(u - u_h, v_h) =$
  $a(\boldsymbol{u} - \boldsymbol{u}_h, \boldsymbol{v}_h) - b(p - p_h, \boldsymbol{v}_h) + b(q_h, \boldsymbol{u} - \boldsymbol{u}_h) = 0$
- $\inf_{\hat{u}_h \in V_h} \|u - \hat{u}_h\|_V = \inf_{\hat{\boldsymbol{u}}_h \in V_h, \hat{q}_h \in Q_h} (\|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_V + \|p - \hat{p}_h\|_Q)$.
  For $\boldsymbol{u}$ and p regular enough:
  $\inf_{\hat{u}_h \in V_h} \|u - \hat{u}_h\|_V \leq C \left(h^k \|\boldsymbol{u}\|_{H^{k+1}(\Omega)} + h^{l+1} \|p\|_{H^l(\Omega)} \right)$

The previous theorem reads:

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_V + \|p - p_h\|_Q \leq C \left(h^k \|\boldsymbol{u}\|_{H^{k+1}(\Omega)} + h^{l+1} \|p\|_{H^l(\Omega)} \right) \equiv \varepsilon(h)$$

L^2 estimate for the velocity via a duality argument

Theorem. Suppose that the domain Ω is regular enough, so that the Stokes problem

$$\begin{aligned} -\nu\Delta \boldsymbol{z} + \nabla\xi &= \boldsymbol{f} && \text{in } \Omega \\ \nabla\cdot\boldsymbol{z} &= 0 && \text{in } \Omega \\ \boldsymbol{z} &= \boldsymbol{0} && \text{on } \partial\Omega \end{aligned}$$

is such that $\|\boldsymbol{z}\|_{H^2} \leq C\|\boldsymbol{f}\|_{L^2}$, $\|\xi\|_{H^1} \leq C\|\boldsymbol{f}\|_{L^2}$. Then

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^2} \leq Ch\,\varepsilon(h)$$

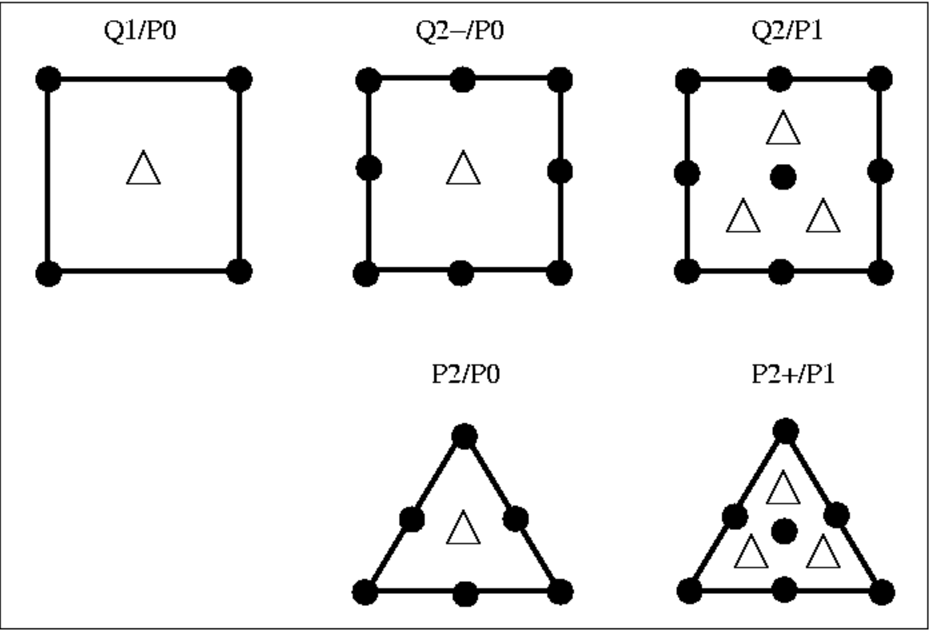
PROOF: Take $\boldsymbol{f} = \boldsymbol{u} - \boldsymbol{u}_h$ in the previous Stokes problem. Then

$$
\begin{aligned}
\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^2}^2 &= a(\boldsymbol{z}, \boldsymbol{u} - \boldsymbol{u}_h) - b(\xi, \boldsymbol{u} - \boldsymbol{u}_h) \\
&= a(\boldsymbol{z} - \boldsymbol{z}_h, \boldsymbol{u} - \boldsymbol{u}_h) + a(\boldsymbol{z}_h, \boldsymbol{u} - \boldsymbol{u}_h) - b(\xi - \xi_h, \boldsymbol{u} - \boldsymbol{u}_h) \\
&= a(\boldsymbol{z} - \boldsymbol{z}_h, \boldsymbol{u} - \boldsymbol{u}_h) + b(p - p_h, \boldsymbol{z}_h - \boldsymbol{z}) - b(\xi - \xi_h, \boldsymbol{u} - \boldsymbol{u}_h) \\
&\leq N_a \|\boldsymbol{z} - \boldsymbol{z}_h\|_{H^1} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{H^1} + \|p - p_h\|_{L^2} \|\boldsymbol{z} - \boldsymbol{z}_h\|_{H^1} \\
&\quad + \|\xi - \xi_h\|_{L^2} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{H^1} \\
&\leq C \left(h\|\boldsymbol{z}\|_{H^2} + h\|\xi\|_{H^1} \right) \varepsilon(h) \\
&\leq Ch\|\boldsymbol{u} - \boldsymbol{u}_h\|_{L^2} \varepsilon(h)
\end{aligned}
$$

□

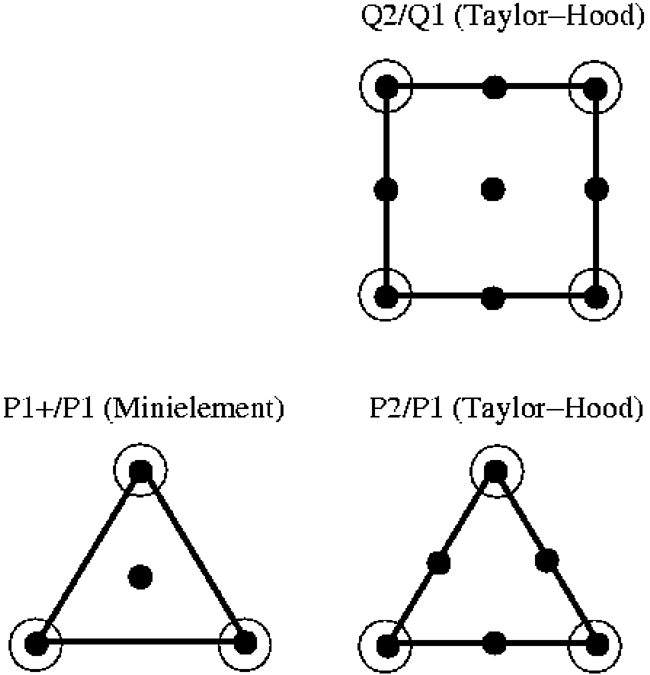
Examples of admissible pairs I: discontinuous pressures

- Q_1/P_0. Multilinear velocity, piecewise constant pressure. It is believed that, except in some particular situations, this element satisfies the inf-sup condition. It gives good results except for the possible appearance of a spurious pressure mode which can be filtered out.
- Q_2^-/P_0. Serendipid second order velocity, piecewise constant pressure.
- Q_2/P_1. Multiquadratic velocity, piecewise linear pressure.
- P_2/P_0. Quadratic velocity, piecewise constant pressure.
- P_2^+/P_0. Quadratic velocity enriched with an internal bubble, piecewise linear pressure.


Q1/P0
Q2–/P0
Q2/P1
P2/P0
P2+/P1

Examples of admissible pairs II: continuous pressures

- P_1^+/P_1. Linear velocity enriched with an internal bubble, linear pressure (minielement).
- P_2/P_1. Quadratic velocity, linear pressure (Taylor-Hood).
- Q_2/Q_1. Multiquadratic velocity, multilinear pressure (Taylor-Hood).


Q2/Q1 (Taylor–Hood)
P1+/P1 (Minielement)
P2/P1 (Taylor–Hood)

Boundary value problem

If $\boldsymbol{\sigma}$ is the the deviatoric component of the stress field, the field equations to be solved in the domain Ω can be written as

$$
\begin{aligned}
-\nabla\cdot\boldsymbol{\sigma}+\nabla p &= \boldsymbol{f}\\
\nabla\cdot\boldsymbol{u} &= 0\\
\boldsymbol{\sigma}-2\mu\boldsymbol{\varepsilon} &= 0
\end{aligned}
$$

where $\boldsymbol{\varepsilon}$ the symmetrical part of $\nabla\boldsymbol{u}$.

This is an important model problem in cases where $\boldsymbol{\sigma}$ needs to be interpolated independently.

Variational form

Let $\mathcal{V} = (H^1_0(\Omega))^d$, $\mathcal{Q} = L^2(\Omega)/\mathbb{R}$ and $\mathcal{T} = (L^2(\Omega))^{d\times d}$. If we call $U = (\boldsymbol{u}, p, \boldsymbol{\sigma})$, $\mathcal{X} = \mathcal{V} \times \mathcal{Q} \times \mathcal{T}$, the weak form of the problem consists in finding $U \in \mathcal{X}$ such that

$$B(U, V) = L(V)$$

for all $V = (\boldsymbol{v}, q, \boldsymbol{\tau}) \in \mathcal{X}$, where

$$B(U, V) = (\nabla^S \boldsymbol{v}, \boldsymbol{\sigma}) - (p, \nabla \cdot \boldsymbol{v}) + (q, \nabla \cdot \boldsymbol{u}) + \frac{1}{2\mu}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - (\varepsilon, \boldsymbol{\tau})$$

$$L(V) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle$$

Stability of the Galerkin finite element discretization

Let us consider a finite element spaces $\mathcal{V}_h \subset \mathcal{V}$, $\mathcal{Q}_h \subset \mathcal{Q}$ and $\mathcal{T}_h \subset \mathcal{T}$. If $\mathcal{X}_h = \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{T}_h$ and $U_h = (\boldsymbol{u}_h, p_h, \boldsymbol{\sigma}_h)$, the Galerkin finite element approximation consists in finding $U_h \in \mathcal{X}_h$ such that

$$B(U_h, V_h) = L(V_h)$$

for all $V_h = (\boldsymbol{v}_h, q_h, \boldsymbol{\tau}_h) \in \mathcal{X}_h$.

If we take $V_h = U_h$, it is found that

$$B(U_h, U_h) = \frac{1}{2\mu} \|\boldsymbol{\sigma}_h\|^2_{L^2}$$

It is seen that B is not coercive in $\mathcal{X}_h$.

Moreover, the inf-sup condition

$$\inf_{U_h \in \mathcal{X}_h} \sup_{V_h \in \mathcal{X}_h} \frac{B(U_h, V_h)}{\|U_h\|_{\mathcal{X}} \|V_h\|_{\mathcal{X}}} \geq K_B$$

which would lead to a well posed problem, is not satisfied for any positive constant K_B unless the two conditions

$$\inf_{q_h \in \mathcal{Q}_h} \sup_{\boldsymbol{v}_h \in \mathcal{V}_h} \frac{(q_h, \nabla \cdot \boldsymbol{v}_h)}{\|q_h\|_{\mathcal{Q}} \|\boldsymbol{v}_h\|_{\mathcal{V}}} \geq C_1 > 0$$

$$\inf_{\boldsymbol{v}_h \in \mathcal{V}_h} \sup_{\boldsymbol{\tau}_h \in \mathcal{T}_h} \frac{(\boldsymbol{\tau}_h, \nabla^S \boldsymbol{v}_h)}{\|\boldsymbol{\tau}_h\|_{\mathcal{T}} \|\boldsymbol{v}_h\|_{\mathcal{V}}} \geq C_2 > 0$$

hold for positive constants C_1 and C_2.

These inf-sup conditions pose stringent requirements on the choice of the finite element spaces.

1. Introduction and problem statement
2. Basic theory for the continuous problem
3. Mixed approximations
4. **Optimization problem and penalty methods**
   4.1 Weak penalization
   4.2 Strong penalization
5. Stabilized formulations

Strong and weak penalization

Strong penalization:

$$\text{Optimize } L_1^\epsilon(\boldsymbol{v}) := \frac{1}{2}a(\boldsymbol{v},\boldsymbol{v}) - l(\boldsymbol{v}) + \frac{1}{2\epsilon}\|\nabla\cdot\boldsymbol{v}\|^2$$

$$\text{Variational problem: } a(\boldsymbol{u}^\epsilon,\boldsymbol{v}) + \frac{1}{\epsilon}(\nabla\cdot\boldsymbol{u}^\epsilon,\nabla\cdot\boldsymbol{v}) = l(\boldsymbol{v}) \quad \forall\boldsymbol{v}\in V$$

$$\text{PDE: } -\nu\Delta\boldsymbol{u}^\epsilon - \frac{1}{\epsilon}\nabla(\nabla\cdot\boldsymbol{u}^\epsilon) = \boldsymbol{f}$$

Weak penalization:

$$\text{Optimize } L_2^\epsilon(\boldsymbol{v},q) := \frac{1}{2}a(\boldsymbol{v},\boldsymbol{v}) - b(q,\boldsymbol{v}) - l(\boldsymbol{v}) - \epsilon\|q\|^2$$

$$\text{Variational problem: } a(\boldsymbol{u}^\epsilon,\boldsymbol{v}) - b(p^\epsilon,\boldsymbol{v}) = l(\boldsymbol{v}) \quad \forall\boldsymbol{v}\in V$$

$$b(q,\boldsymbol{u}^\epsilon) + \epsilon(p^\epsilon,q) = 0 \quad \forall q\in Q$$

$$\text{PDE: } -\nu\Delta\boldsymbol{u}^\epsilon + \nabla p^\epsilon = \boldsymbol{f}, \quad \epsilon p^\epsilon + \nabla\cdot\boldsymbol{u}^\epsilon = 0$$

Convergence

Theorem. Given $\epsilon > 0$, let $[\boldsymbol{u}^\epsilon, p^\epsilon] \in V \times Q$ the solution of the variational problem

$$a(\boldsymbol{u}^\epsilon, \boldsymbol{v}) - b(p^\epsilon, \boldsymbol{v}) = l(\boldsymbol{v}) \quad \forall \boldsymbol{v} \in V$$
$$b(q, \boldsymbol{u}^\epsilon) + \epsilon(p^\epsilon, q) = 0 \quad \forall q \in Q$$

and let $[\boldsymbol{u}, p] \in V \times Q$ the solution for $\epsilon = 0$. Then

$$\|\boldsymbol{u} - \boldsymbol{u}^\epsilon\|_V \leq C\epsilon, \quad \|p - p^\epsilon\|_Q \leq C'\epsilon$$

for certain constants C and C'. Likewise, if $[\boldsymbol{u}_h^\epsilon, p_h^\epsilon]$ is a stable finite element approximation to $[\boldsymbol{u}^\epsilon, p^\epsilon]$ and $[\boldsymbol{u}_h, p_h]$ to $[\boldsymbol{u}, p]$,

$$\|\boldsymbol{u}_h - \boldsymbol{u}_h^\epsilon\|_V \leq C\epsilon, \quad \|p_h - p_h^\epsilon\|_Q \leq C'\epsilon$$

PROOF: It is enough to prove the result in the continuous case:

$$a(\boldsymbol{u} - \boldsymbol{u}^\epsilon, \boldsymbol{v}) - b(p - p^\epsilon, \boldsymbol{v}) = 0 \quad \forall \boldsymbol{v} \in V$$
$$0 \leq (p - p^\epsilon, p - p^\epsilon) \Rightarrow (p^\epsilon, p - p^\epsilon) \leq (p, p - p^\epsilon)$$

From these two expressions:

$$\begin{aligned} K_a \|\boldsymbol{u} - \boldsymbol{u}^\epsilon\|_V^2 &\leq a(\boldsymbol{u} - \boldsymbol{u}^\epsilon, \boldsymbol{u} - \boldsymbol{u}^\epsilon) = b(p - p^\epsilon, \boldsymbol{u} - \boldsymbol{u}^\epsilon) \\ &= -b(p - p^\epsilon, \boldsymbol{u}^\epsilon) = \epsilon(p^\epsilon, p - p^\epsilon) \\ &\leq \epsilon(p, p - p^\epsilon) \leq \epsilon \|p\|_Q \|p - p^\epsilon\|_Q \end{aligned}$$

From the inf-sup condition, there exists $\boldsymbol{v} \in V$ such that

$$\begin{aligned} K_b \|p - p^\epsilon\|_Q \|\boldsymbol{v}\|_V &\leq b(p - p^\epsilon, \boldsymbol{v}) = a(\boldsymbol{u} - \boldsymbol{u}^\epsilon, \boldsymbol{v}) \\ &\leq N_a \|\boldsymbol{u} - \boldsymbol{u}^\epsilon\|_V \|\boldsymbol{v}\|_V \end{aligned}$$

$$\Rightarrow \quad K_a \|\boldsymbol{u} - \boldsymbol{u}^\epsilon\|_V^2 \leq \epsilon \|p\|_Q \frac{N_a}{K_b} \|\boldsymbol{u} - \boldsymbol{u}^\epsilon\|_V$$

Discretization

Let $Q_h \subset Q$, $V_h \subset V$ be such that the discrete inf-sup condition holds. If functions in Q_h are discontinuous, we may take as test function for the continuity equation

$$q_h = N_p^a \begin{cases} \neq 0 & \text{in } K \text{ if node } a \in K \\ = 0 & \text{outside } K \end{cases}$$

The continuity equation becomes:

$$\epsilon \sum_b \left(\int_K N_p^a N_p^b \right) \mathrm{P}^b + \sum_c \left(\int_K N_p^a \partial_i N_u^c \right) \mathrm{U}_i^c = 0$$

which has the form

$$\epsilon \mathrm{M}_{p,K} \mathrm{P}_K + \mathrm{D}_K \mathrm{U}_K = 0$$

where subindex K denotes that all matrices and arrays refer only to element K.

Since $\mathsf{M}_{p,K}$ is easily inverted:

$$\mathsf{P}_K = -\epsilon^{-1}\mathsf{M}_{p,K}^{-1}\mathsf{D}_K\mathsf{U}_K$$

The momentum equation becomes:

$$\begin{aligned}\mathsf{KU} - \sum_K \mathsf{D}_K^T\mathsf{P}_K &= \mathsf{KU} + \epsilon^{-1}\sum_K \mathsf{D}_K^T\mathsf{M}_{p,K}^{-1}\mathsf{D}_K\mathsf{U}_K \\ &= \left[\mathsf{K} + \epsilon^{-1}\mathbb{A}_K\left(\mathsf{D}_K^T\mathsf{M}_{p,K}^{-1}\mathsf{D}_K\right)\right]\mathsf{U} = \mathsf{F}\end{aligned}$$

which can be solved for U without relying on P. It has to be noted that

- Matrix $\mathsf{K}_\epsilon := \mathsf{K} + \epsilon^{-1}\mathbb{A}_K\left(\mathsf{D}_K^T\mathsf{M}_{p,K}^{-1}\mathsf{D}_K\right)$ is symmetric and positive definite. No pivoting will be needed.
- The condition number of K_ϵ is $\chi = \mathcal{O}(\epsilon^{-1})$. If ϵ is small, iterative solvers are unfeasible.

Problem formulation

Variational formulation: Find $\boldsymbol{u}_h^\epsilon \in V_h$ such that

$$a(\boldsymbol{u}_h^\epsilon, \boldsymbol{v}_h) + \frac{1}{\epsilon}(\nabla \cdot \boldsymbol{u}_h^\epsilon, \nabla \cdot \boldsymbol{v}_h) = l(\boldsymbol{v}_h) \quad \forall \boldsymbol{v}_h \in V_h$$

Associated pressure space:

$$p_h^\epsilon = -\frac{1}{\epsilon}\nabla \cdot \boldsymbol{u}_h^\epsilon \iff Q_h = \nabla \cdot V_h$$

Matrix form:

$$\left(\mathrm{K} + \frac{1}{\epsilon}\mathrm{K}_v\right)\mathrm{U}^\epsilon = \mathrm{F}, \quad {\mathrm{K}_v}_{ij}^{ab} = (\partial_i N_u^a, \partial_j N_u^b)$$

Locking: if matrix K_v is nonsingular (after prescribing BC's)

$$\lim_{\epsilon \longrightarrow 0} \|\mathrm{U}^\epsilon\| = 0$$

Underintegration

Original idea: integrate K_v with a low order rule that yields a rank-defficient matrix (Zienkiewicz, Taylor, Too (1973) in plates, Fried (1974) in elasticity, formalization by Oden (1982)).

Reduced-integration-penality (RIP) method: Find $\boldsymbol{u}_h^\epsilon \in V_h$ such that

$$a(\boldsymbol{u}_h^\epsilon, \boldsymbol{v}_h) + \frac{1}{\epsilon}(\nabla \cdot \boldsymbol{u}_h^\epsilon, \nabla \cdot \boldsymbol{v}_h)_* = l(\boldsymbol{v}_h) \quad \forall \boldsymbol{v}_h \in V_h$$

where

$$(f, g)_* = \sum_K \sum_{i=1}^{n_{\text{int}}} w_K^i f(\xi_K^i) g(\xi_K^i)$$

where n_{int} is the number of integration points per element.

Equivalence with weak penalization

Theorem. Let Q_h the space of piecewise polynomials that are n_{int}-unisolvent. If

$$(q_h, q_h')_* = (q_h, q_h') \quad \forall q_h, q_h' \in Q_h$$
$$(q_h, \nabla \cdot \boldsymbol{v}_h)_* = (q_h, \nabla \cdot \boldsymbol{v}_h) \quad \forall q_h \in Q_h, \forall \boldsymbol{v}_h \in V_h$$

then RIP method $\iff$ weak penalization in $V_h \times Q_h$

PROOF:

$$0 = \epsilon(q_h, p_h^\epsilon) + (q_h, \nabla \cdot \boldsymbol{u}_h^\epsilon) = (q_h, \epsilon p_h^\epsilon + \nabla \cdot \boldsymbol{u}_h^\epsilon)_*$$
$$= \sum_K \sum_{i=1}^{n_{\text{int}}} w_K^i q_h(\xi_K^i) \left(\epsilon p_h^\epsilon(\xi_K^i) + \nabla \cdot \boldsymbol{u}_h^\epsilon(\xi_K^i) \right)$$

Since Q_h is n_{int}-unisolvent, $\epsilon p_h^\epsilon(\xi_K^i) = -\nabla \cdot \boldsymbol{u}_h^\epsilon(\xi_K^i)$. It can be checked that the starting weak penalty method reduces to the RIP method.

Example: Q_1 element

Consider:

- Method A: Weak penalty method with Q_1/P_0 interpolation
- Method B: RIP method, with 1 point integration for $\epsilon^{-1}(\nabla \cdot \boldsymbol{u}_h^\epsilon, \nabla \cdot \boldsymbol{v}_h)$

Then

$$\text{Method A} \iff \text{Method B}$$

since 1 point integration on quads is exact for functions of the form

(q_h, q_h')	product of two constants
$(q_h, \nabla \cdot \boldsymbol{v}_h)$	product of constant and linear function

1. Introduction and problem statement
2. Basic theory for the continuous problem
3. Mixed approximations
4. Optimization problem and penalty methods
5. **Stabilized formulations**
   5.1 The Galerkin/least-squares (GLS) method
   5.2 Stabilization through bubble functions
   5.3 A general framework: the subgrid scale concept

Motivation of stabilized finite element methods

- In general:
  - Singular perturbation problems
  - Non-definite bilinear forms with non-trivial inf-sup conditions
- For the Stokes problem:
  - Equal velocity-pressure interpolation
  - Stability of convenient elements such as Q_1/P_0 and P_1/P_0
- Other examples that “need” stabilization:
  - Convection-diffusion, reaction-diffusion
  - Plate problem
  - Darcy’s problem
  - Waves...

The Stokes BVP reformulated

The Stokes problem can be written in vector form as

$$\mathcal{L}(\boldsymbol{u}, p) = \begin{bmatrix} \mathcal{L}_1(\boldsymbol{u}, p) \\ \mathcal{L}_2(\boldsymbol{u}, p) \end{bmatrix} := \begin{bmatrix} -\nu \Delta \boldsymbol{u} + \nabla p \\ \nabla \cdot \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f} \\ 0 \end{bmatrix} =: \mathcal{F}$$

Matrix form of the “convective” (first order) terms in 2D:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \frac{\partial}{\partial x_1} \begin{bmatrix} u_1 \\ u_2 \\ p \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \frac{\partial}{\partial x_2} \begin{bmatrix} u_1 \\ u_2 \\ p \end{bmatrix} = \begin{bmatrix} \nabla p \\ \nabla \cdot \boldsymbol{u} \end{bmatrix}$$

Original idea: extension of SUPG

Since the first order terms can be understood as "convective", the idea is to use a SUPG-like strategy (diffusion is 0 for the pressure). This means to multiply the original problem by $[\boldsymbol{v}_h, q_h]$ to obtain the Galerkin terms and by

$$\boldsymbol{\tau} \begin{bmatrix} \nabla q_h \\ \nabla \cdot \boldsymbol{v}_h \end{bmatrix}$$

to obtain the stabilizing terms. Here $\boldsymbol{\tau}$ is a matrix that can be taken in 2D as

$$\boldsymbol{\tau} = \mathrm{diag}(\tau_1, \tau_1, \tau_2)$$

Original method

The stabilized finite element problem based on SUPG is: Find $\boldsymbol{U}_h := [\boldsymbol{u}_h, p_h]$ such that:

$$\nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) - (p_h, \nabla \cdot \boldsymbol{v}_h) + (q_h, \nabla \cdot \boldsymbol{u}_h) - \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle + \sum_K \int_K \tau_1 \nabla q_h \cdot (-\nu \Delta \boldsymbol{u}_h + \nabla p_h - \boldsymbol{f}) + \sum_K \int_K \tau_2 (\nabla \cdot \boldsymbol{v}_h)(\nabla \cdot \boldsymbol{u}_h) = 0$$

for all $\boldsymbol{V}_h := [\boldsymbol{v}_h, q_h]$. The stabilizing term is of the form

$$\sum_K \int_K [\boldsymbol{\tau} \mathcal{P}(\boldsymbol{V}_h)] \cdot [\mathcal{L}(\boldsymbol{U}_h) - \mathcal{F}]$$

where $\mathcal{P}(\boldsymbol{V}_h)$ is part of the Stokes operator applied to $\boldsymbol{V}_h$.

The GLS method

The next idea is to apply all the differential operator to the test functions to obtain the stabilization term. This yields the term

$$
\begin{aligned}
&\sum_K \int_K [\boldsymbol{\tau}\mathcal{L}(\boldsymbol{V}_h)] \cdot [\mathcal{L}(\boldsymbol{U}_h) - \mathcal{F}] \\
&= \sum_K \int_K \tau_1 \mathcal{L}_1(\boldsymbol{v}_h, q_h)[\mathcal{L}_1(\boldsymbol{u}_h, p_h) - \boldsymbol{f}] \\
&\quad + \sum_K \int_K \tau_2 \mathcal{L}_2(\boldsymbol{v}_h, q_h)\mathcal{L}_2(\boldsymbol{u}_h, p_h) \\
&= \sum_K \int_K \tau_1 \left(-\nu\Delta\boldsymbol{v}_h + \nabla q_h\right) \cdot \left(-\nu\Delta\boldsymbol{u}_h + \nabla p_h - \boldsymbol{f}\right) \\
&\quad + \sum_K \int_K \tau_2 (\nabla \cdot \boldsymbol{v}_h)(\nabla \cdot \boldsymbol{u}_h)
\end{aligned}
$$

The reason for the name

If A is a matrix and x and b are vectors, we have:

$$\|Ax - b\|^2 = (Ax - b, Ax - b)$$
$$\text{is minimum} \iff$$
$$\frac{\partial}{\partial x}\|Ax - b\|^2 = 0 \iff (Ay, Ax - b) = 0 \quad \forall y$$

Thus, the stabilizing term

$$\sum_K \int_K [\boldsymbol{\tau}\mathcal{L}(\boldsymbol{V}_h)] \cdot [\mathcal{L}(\boldsymbol{U}_h) - \mathcal{F}]$$

can be thought of as a least square form (weighted by $\boldsymbol{\tau}$) of the elementwise finite element residual.

System of discrete variational equations

Find $[\boldsymbol{u}_h, p_h]$ such that:

First (momentum) equation:

$$\nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) - (p_h, \nabla \cdot \boldsymbol{v}_h) - \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle + \sum_K \int_K \tau_1 \, (-\nu \Delta \boldsymbol{v}_h) \cdot (-\nu \Delta \boldsymbol{u}_h + \nabla p_h - \boldsymbol{f}) + \sum_K \int_K \tau_2 (\nabla \cdot \boldsymbol{v}_h)(\nabla \cdot \boldsymbol{u}_h) = 0$$

Second (continuity) equation:

$$(q_h, \nabla \cdot \boldsymbol{u}_h) + \sum_K \int_K \tau_1 \nabla q_h \cdot (-\nu \Delta \boldsymbol{u}_h + \nabla p_h - \boldsymbol{f}) = 0$$

for all $[\boldsymbol{v}_h, q_h]$.

Matrix form

The matrix structure is the following:

$$\begin{bmatrix} \mathrm{K} + \bar{\mathrm{K}} & -\mathrm{D}^t - \bar{\mathrm{D}}^t \\ \mathrm{D} - \bar{\mathrm{D}} & \mathrm{L} \end{bmatrix} \begin{bmatrix} \mathrm{U} \\ \mathrm{P} \end{bmatrix} = \begin{bmatrix} \mathrm{F} + \bar{\mathrm{F}} \\ \bar{\bar{\mathrm{F}}} \end{bmatrix}$$

where the arrays with an overbar come from the stabilization terms and the rest from the Galerkin contribution:

$$\mathrm{K} \longleftarrow \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h)$$

$$\bar{\mathrm{K}} \longleftarrow \sum_K \int_K \tau_1 \left(-\nu \Delta \boldsymbol{v}_h\right) \cdot \left(-\nu \Delta \boldsymbol{u}_h\right) + \sum_K \int_K \tau_2 (\nabla \cdot \boldsymbol{v}_h)(\nabla \cdot \boldsymbol{u}_h)$$

$$\mathrm{D} \longleftarrow (q_h, \nabla \cdot \boldsymbol{u}_h)$$

$$\bar{\mathrm{D}} \longleftarrow \sum_K \int_K \tau_1 \nabla q_h \cdot (-\nu \Delta \boldsymbol{u}_h)$$

$$\mathsf{L} \longleftarrow \sum_K \int_K \tau_1 \nabla q_h \cdot \nabla p_h$$

$$\bar{\mathsf{F}} \longleftarrow \sum_K \int_K \tau_1 \left(-\nu \Delta \boldsymbol{v}_h\right) \cdot \boldsymbol{f}$$

$$\bar{\bar{\mathsf{F}}} \longleftarrow \sum_K \int_K \tau_1 \nabla q_h \cdot \boldsymbol{f}$$

Observe that the final algebraic system is not symmetric, but it is positive-definite.

The stabilizing effect of the GLS method on the pressure comes from the term involving L, which provides a pressure Laplacian.

Main convergence theorem

Theorem. Suppose that the Stokes problem is solved using the GLS method with

$$\tau_1 = C_\tau \frac{h^2}{\nu}, \quad \tau_2 = 0$$

Let B_{GLS} the bilinear form associated to this method and define the mesh dependent norm $|||\cdot|||$ by

$$|||\boldsymbol{V}_h|||^2 = \nu \|\boldsymbol{v}_h\|^2 + \sum_K \int_K \tau_1 \|\nabla q_h\|^2$$

Then, for a small enough value of the constant C_τ,

$$B_{\mathrm{GLS}}(\boldsymbol{U}_h, \boldsymbol{U}_h) \geq C|||\boldsymbol{U}_h|||^2 \quad \text{(Stability)}$$

$$|||\boldsymbol{U} - \boldsymbol{U}_h||| \leq \inf_{\boldsymbol{v}_h \in \mathcal{W}_0} |||\boldsymbol{U} - \boldsymbol{V}_h||| \quad \text{(Convergence)}$$

Remarks

- The method presented is not what was originally called GLS in the case of the Stokes problem. This name was given to a method yielding a symmetric problem which we will recover from a different point of view later on.
- It is also possible to obtain the same result with the norm

$$|||\boldsymbol{V}_h|||'^2 = \nu\|\boldsymbol{v}_h\|^2 + \sum_K \int_K \tau_1 \| -\nu\Delta\boldsymbol{v}_h + \nabla q_h\|^2$$

- If $\tau_2 > 0$, the norm in which stability and convergence is proven is

$$|||\boldsymbol{V}_h|||^2 + \sum_K \int_K \tau_2 \|\nabla \cdot \boldsymbol{v}_h\|^2$$

Inclusion of jumps for discontinuous pressures

The Q_1/P_0 element (for example) does not satisfy the BB condition:

- There are spurious pressure modes
- $K_b = K_b(h)$ in the inf-sup condition, at least for regular meshes.

Let us consider a partition of Ω into patches of 2×2 Q_1/P_0 elements (in 2D). Let Γ_p be the collection of interior edges of patch number p and $[\![\cdot]\!]$ is the jump operator.

The Silvester-Kechkar stabilization consists of replacing the incompressibility condition $(q_h, \nabla \cdot \boldsymbol{u}_h) = 0$ by

$$(q_h, \nabla \cdot \boldsymbol{u}_h) + \sum_p \alpha h \int_{\Gamma_p} [\![q_h]\!] [\![p_h]\!] = 0$$

Generalized convergence result

Theorem. Consider the GLS method with the associated bilinear form enlarged with the term $\sum_E \alpha h \int_E [\![q_h]\!][\![p_h]\!]$, where the sum extends to all the internal edges (2D) or faces (3D) of the finite element partition. Define the mesh dependent norm $|||\boldsymbol{V}_h|||^2 = \nu\|\boldsymbol{v}_h\|^2 + \sum_K \int_K \tau_1 \|\nabla q_h\|^2 + \sum_E \int_E \alpha h [\![q_h]\!]^2$. Then, for a small enough value of the constant C_τ, the method is stable and optimally convergent in this norm if at least one of the following conditions is verified:

- Pressures are continuous (the jump disappears)
- $\alpha > 0$
- The order of velocity interpolation is $k \geq d$

GLS with P_1 elements

Let us consider the GLS method using P_1 finite element interpolations for both the velocity and the pressure.

If $\tau_2 = 0$ and $\Delta \boldsymbol{u}_h = \Delta \boldsymbol{v}_h = \boldsymbol{0}$ within each element, the GLS method reduces to

$$\nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) - (p_h, \nabla \cdot \boldsymbol{v}_h) = \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle$$

$$(q_h, \nabla \cdot \boldsymbol{u}_h) + \sum_K \int_K \tau_1 \nabla q_h \cdot (\nabla p_h - \boldsymbol{f}) = 0$$

for all $[\boldsymbol{v}_h, q_h]$.

Galerkin with the minielement

Consider the P_1^+/P_1 interpolation (div-stable) and the standard Galerkin method.

Within each element, the velocity (which we denote $\boldsymbol{u}_h^b(\boldsymbol{x})$) can be expanded as:

$$\boldsymbol{u}_h^b(\boldsymbol{x})|_K = \boldsymbol{u}_h(\boldsymbol{x})|_K + \sum_j \boldsymbol{B}_j^K(\boldsymbol{x}) u_j^{K,b}$$

$\boldsymbol{u}_h(\boldsymbol{x})$: linear part of the velocity.
$\boldsymbol{B}_j^K(\boldsymbol{x})$: vector that has the bubble function $B^K(\boldsymbol{x})$ in the jth component and zero the rest.
$u_j^{K,b}$: j-th component of the velocity nodal value associated to the bubble of element K.

Condensation of the bubble velocity

Using the first equation to be solved and taking $\boldsymbol{v}_h = \boldsymbol{B}_j^K$ we have, for each element,

$$\nu(\nabla \boldsymbol{u}_h^b, \nabla \boldsymbol{B}_j^K) - (p_h, \nabla \cdot \boldsymbol{B}_j^K) = \langle \boldsymbol{f}, \boldsymbol{B}_j^K \rangle$$

$$u_j^{K,b} \nu(\nabla \boldsymbol{B}_j^K, \nabla \boldsymbol{B}_j^K) = \langle \boldsymbol{f}, \boldsymbol{B}_j^K \rangle + (p_h, \nabla \cdot \boldsymbol{B}_j^K) - \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{B}_j^K)$$

$$= \int_K \boldsymbol{B}_j^K \cdot (\boldsymbol{f} - \nabla p_h + \nu \Delta \boldsymbol{u}_h) + \int_{\partial K} \boldsymbol{B}_j^K \cdot (p\boldsymbol{n} - \nu \boldsymbol{n} \cdot \nabla \boldsymbol{u}_h)$$

$$= \int_K \boldsymbol{B}_j^K \cdot (\boldsymbol{f} - \nabla p_h),$$

$$u_j^{K,b} = \frac{1}{\nu(\nabla B^K, \nabla B^K)} \int_K \boldsymbol{B}_j^K \cdot (\boldsymbol{f} - \nabla p_h)$$

Modified continuity equation

For the second equation to be solved:

$$(q_h, \nabla \cdot \boldsymbol{u}_h^b) = (q_h, \nabla \cdot \boldsymbol{u}_h) + \sum_K \sum_j u_j^{K,b} (q_h, \nabla \cdot \boldsymbol{B}_j^K)$$

$$= (q_h, \nabla \cdot \boldsymbol{u}_h)$$

$$- \sum_K \sum_j \frac{1}{\nu(\nabla B^K, \nabla B^K)} \left[\int_K \boldsymbol{B}_j^K \cdot (\boldsymbol{f} - \nabla p_h) \right] \left[\int_K \boldsymbol{B}_j^K \cdot \nabla q_h \right]$$

$$= (q_h, \nabla \cdot \boldsymbol{u}_h) - \sum_K \int_K \nabla q_h \cdot \hat{\tau} (\boldsymbol{f} - \nabla p_h) = 0$$

where we have used the fact that ∇q_h and ∇p_h are piecewise constants and we have defined

$$\hat{\tau} := \left[\nu(\nabla B^K, \nabla B^K) \right]^{-1} (\text{meas} K)^{-1} \left(\int_K B^K \right)^2$$

Comparison of GLS with P_1 and Galerkin with P_1^+/P_1

Momentum equation: If $\boldsymbol{B}_j^K$ is a bubble function, it follows that

$$\int_K \nabla \boldsymbol{B}_j^K : \nabla \boldsymbol{v}_h = -\int_K \boldsymbol{B}_j^K \cdot \Delta \boldsymbol{v}_h + \int_{\partial K} \boldsymbol{B}_j^K \cdot (\boldsymbol{n} \cdot \nabla \boldsymbol{v}_h) = 0$$

so the momentum equation for Galerkin with P_1^+/P_1 is

$$\nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) - (p_h, \nabla \cdot \boldsymbol{v}_h) = \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle$$

This equation is the same for GLS (and for Galerkin) with P_1

Continuity equation: It can be written as

$$(q_h, \nabla \cdot \boldsymbol{u}_h) + \sum_K \int_K \nabla q_h \cdot \hat{\tau}(\nabla p_h - \boldsymbol{f}) = 0$$

This equation is the same for GLS with P_1, with an appropriate definition of τ, which in both cases behave as $\frac{h^2}{\nu}$.

Remarks

- For elements of higher order, the analogy GLS = Galerkin with bubble functions does not hold exactly.
- However, the previous development motivated the design of stabilized methods by introducing bubble functions.
- The problem of estimating the stabilization parameters is transformed into the problem of selecting the bubble. The classical cubic bubble is only a possibility.
- This framework was particularly exploited in the context of "residual free bubbles" introduced later.

Model problem

System of convection-diffusion-reaction equations:

$$\mathcal{L}(\boldsymbol{U}) = \boldsymbol{F} \quad \text{in} \quad \Omega$$
$$\mathcal{L}(\boldsymbol{U}) := \sum_{i=1}^{d} \frac{\partial}{\partial x_i}(\boldsymbol{A}_i \boldsymbol{U}) - \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}}{\partial x_j}\right) + \boldsymbol{S}\boldsymbol{U}$$
$$\boldsymbol{U} = \boldsymbol{0} \quad \text{on } \partial\Omega$$

$\boldsymbol{U}$ and $\boldsymbol{F}$ are vectors of D unknowns and $\boldsymbol{A}_i$, $\boldsymbol{K}_{ij}$ and $\boldsymbol{S}$ are $D \times D$ matrices ($i, j = 1, ..., n_{\rm sd}$).

Let $\mathcal{W}$ be the space where the problem is posed. The weak form is: Find $\boldsymbol{U} \in \mathcal{W}$ such that

$$B(\boldsymbol{U}, \boldsymbol{V}) = L(\boldsymbol{V}) \qquad \forall \boldsymbol{V} \in \mathcal{W}$$

Integrating by parts the second-order terms and assuming that the boundary contributions vanish:

$$B(\boldsymbol{U},\boldsymbol{V}) := \int_\Omega \boldsymbol{V}^{\mathrm{t}} \frac{\partial}{\partial x_i}(\boldsymbol{A}_i\boldsymbol{U}) + \int_\Omega \frac{\partial \boldsymbol{V}^{\mathrm{t}}}{\partial x_i}\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}}{\partial x_j} + \int_\Omega \boldsymbol{V}^{\mathrm{t}}\boldsymbol{S}\boldsymbol{U}$$

$$L(\boldsymbol{V}) := \int_\Omega \boldsymbol{V}^{\mathrm{t}}\boldsymbol{F}$$

If $\mathcal{W}_h \subset \mathcal{W}$ is a finite element approximation to $\mathcal{W}$, the Galerkin finite element problem is: find $\boldsymbol{U}_h \in \mathcal{W}_h$ such that

$$B(\boldsymbol{U}_h,\boldsymbol{V}_h) = L(\boldsymbol{V}_h) \qquad \forall \boldsymbol{V}_h \in \mathcal{W}_h$$

Description of the method

Let $\mathcal{W} = \mathcal{W}_h \oplus \tilde{\mathcal{W}}$. The continuous problem is equivalent to: find $\boldsymbol{U}_h \in \mathcal{W}_h$ and $\tilde{\boldsymbol{U}} \in \tilde{\mathcal{W}}$ such that

$$
\begin{aligned}
B(\boldsymbol{U}_h, \boldsymbol{V}_h) + B(\tilde{\boldsymbol{U}}, \boldsymbol{V}_h) &= L(\boldsymbol{V}_h) \qquad && \forall \boldsymbol{V}_h \in \mathcal{W}_h \\
B(\boldsymbol{U}_h, \tilde{\boldsymbol{V}}) + B(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}) &= L(\tilde{\boldsymbol{V}}) \qquad && \forall \tilde{\boldsymbol{V}} \in \tilde{\mathcal{W}}
\end{aligned}
$$

Function $\tilde{\boldsymbol{U}} \in \tilde{\mathcal{W}}$ is called the subgrid-scale or subscale, and $\tilde{\mathcal{W}}$ the is thus the space of subscales.

Let us introduce the notation:

$$
\int_{\Omega'} := \sum_K \int_K, \quad \int_{\partial\Omega'} := \sum_{\partial K} \int_{\partial K}
$$

Integrating by parts within each element:

$$B(\boldsymbol{U}_h, \boldsymbol{V}_h) + \int_{\partial\Omega'} \tilde{\boldsymbol{U}}^{\mathrm{t}} n_i \boldsymbol{K}_{ij} \frac{\partial \boldsymbol{V}_h}{\partial x_j} \, \mathrm{d}\Gamma + \int_{\Omega'} \tilde{\boldsymbol{U}}^{\mathrm{t}} \mathcal{L}^*(\boldsymbol{V}_h) = L(\boldsymbol{V}_h)$$

$$\int_{\partial\Omega'} \tilde{\boldsymbol{V}}^{\mathrm{t}} n_i \boldsymbol{K}_{ij} \frac{\partial}{\partial x_j} (\boldsymbol{U}_h + \tilde{\boldsymbol{U}}) \, \mathrm{d}\Gamma + \int_{\Omega'} \tilde{\boldsymbol{V}}^{\mathrm{t}} \mathcal{L}(\tilde{\boldsymbol{U}}) = \int_{\Omega'} \tilde{\boldsymbol{V}}^{\mathrm{t}} [\boldsymbol{F} - \mathcal{L}(\boldsymbol{U}_h)]$$

Adjoint of the convection-diffusion-reaction operator:

$$\mathcal{L}^*(\boldsymbol{V}_h) = -\boldsymbol{A}_i^{\mathrm{t}} \frac{\partial \boldsymbol{V}_h}{\partial x_i} - \frac{\partial}{\partial x_i} \left(\boldsymbol{K}_{ij}^{\mathrm{t}} \frac{\partial \boldsymbol{V}_h}{\partial x_j} \right) + \boldsymbol{S}^{\mathrm{t}} \boldsymbol{V}_h$$

Problem for the subscales

Since diffusive fluxes must be continuous across inter-element boundaries, the second equation is equivalent to:

$$\mathcal{L}(\tilde{\boldsymbol{U}}) = \boldsymbol{F} - \mathcal{L}(\boldsymbol{U}_h) + \boldsymbol{V}_{h,\mathrm{ort}} \quad \text{in } K$$
$$\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{U}}_{\mathrm{ske}} \quad \text{on } \partial K$$

for all K. We call $\tilde{\boldsymbol{U}}_{\mathrm{ske}}$ the *skeleton* of $\boldsymbol{U}$.

Main problems:

- How to choose $\boldsymbol{V}_{h,\mathrm{ort}}$?
- How to approximate $\tilde{\boldsymbol{U}}_{\mathrm{ske}}$?
- How to solve for $\tilde{\boldsymbol{U}}$?

Approximation of the subscales

In the following, we will make the approximations:

- $\boldsymbol{V}_{h,\mathrm{ort}} = \boldsymbol{0}$ (this defines the space for $\tilde{\boldsymbol{U}}$).
- $\tilde{\boldsymbol{U}}_{\mathrm{ske}} = \boldsymbol{0}$

It remains to define how to solve for $\tilde{\boldsymbol{U}}$. Different approaches, yielding different stabilization methods, are possible. In any case, the problem to be solved is:

$$B(\boldsymbol{U}_h, \boldsymbol{V}_h) + \int_{\Omega'} \tilde{\boldsymbol{U}}^{\mathrm{t}} \mathcal{L}^*(\boldsymbol{V}_h) = L(\boldsymbol{V}_h) \quad \forall \boldsymbol{V}_h$$

$$\mathcal{L}(\tilde{\boldsymbol{U}}) = \boldsymbol{F} - \mathcal{L}(\boldsymbol{U}_h) \quad \text{in } K$$

$$\tilde{\boldsymbol{U}} = \boldsymbol{0} \quad \text{on } \partial K$$

Approximation to the Green's function

The problem for the subscales is

$$\mathcal{L}(\tilde{\boldsymbol{U}}) = \boldsymbol{F} - \mathcal{L}(\boldsymbol{U}_h) \quad \text{in } K$$
$$\tilde{\boldsymbol{U}} = \boldsymbol{0} \quad \text{on } \partial K$$

The solution can be written using the formula:

$$\tilde{\boldsymbol{U}}(\boldsymbol{x}) = \int_K \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \left[\boldsymbol{F} - \mathcal{L}(\boldsymbol{U}_h) \right] (\boldsymbol{y}) d\boldsymbol{y}$$

where $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ is the Green's function of the problem (which now is a matrix) and the integration arguments have been explicitly indicated.

The problem now is to approximate $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$

Algebraic approximation to the Green's function

Let us consider the approximation

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \approx \boldsymbol{\tau}(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$$

where $\delta(\boldsymbol{x}-\boldsymbol{y})$ is the Dirac delta function. Using this approximation:

$$\begin{aligned}\tilde{\boldsymbol{U}}(\boldsymbol{x}) &\approx \int_K \delta(\boldsymbol{x}-\boldsymbol{y})\boldsymbol{\tau}(\boldsymbol{x})\left[\boldsymbol{F}-\mathcal{L}(\boldsymbol{U}_h)\right](\boldsymbol{y})d\boldsymbol{y} \\ &= \boldsymbol{\tau}(\boldsymbol{x})\left[\boldsymbol{F}-\mathcal{L}(\boldsymbol{U}_h)\right](\boldsymbol{x})\end{aligned}$$

A possible way to determine $\boldsymbol{\tau}$ is to impose that moments of $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})$ and $\boldsymbol{\tau}(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{y})$ coincide. In particular, if $\boldsymbol{\tau}(\boldsymbol{x})$ is taken constant,

$$\boldsymbol{\tau} = \frac{1}{\mathrm{meas}(K)}\int_K\int_K \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y}$$

In general, $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ is unknown and $\boldsymbol{\tau}$ must be approximated by other means.

If $\boldsymbol{\tau}$ is taken constant, we call it algebraic approximation to the subscales. In this case, the stabilized method is:

$$B(\boldsymbol{U}_h, \boldsymbol{V}_h) + \int_{\Omega'} [-\mathcal{L}^*(\boldsymbol{V}_h)]^{\mathrm{t}} \boldsymbol{\tau} [\mathcal{L}(\boldsymbol{U}_h) - \boldsymbol{F}] = L(\boldsymbol{V}_h) \quad \forall \boldsymbol{V}_h$$

Equivalently, the stabilized problem can be written:

$$B_{\mathrm{asgs}}(\boldsymbol{U}_h, \boldsymbol{V}_h) = L_{\mathrm{asgs}}(\boldsymbol{V}_h) \quad \forall \boldsymbol{V}_h$$

where

$$B_{\mathrm{asgs}}(\boldsymbol{U}_h, \boldsymbol{V}_h) = B(\boldsymbol{U}_h, \boldsymbol{V}_h) + \int_{\Omega'} [-\mathcal{L}^*(\boldsymbol{V}_h)]^{\mathrm{t}} \boldsymbol{\tau} [\mathcal{L}(\boldsymbol{U}_h)]$$

$$L_{\mathrm{asgs}}(\boldsymbol{V}_h) = L(\boldsymbol{V}_h) + \int_{\Omega'} [-\mathcal{L}^*(\boldsymbol{V}_h)]^{\mathrm{t}} \boldsymbol{\tau} [\boldsymbol{F}]$$

Approximation to the subscales using residual free bubbles

To simplify the notation, let us consider an abstract problem of the form $a(u, v) = l(v)$. The problem for the subscales is

$$\mathcal{L}(\tilde{u}) = f - \mathcal{L}(u_h) \quad \text{in } K$$
$$\tilde{u} = 0 \quad \text{on } \partial K$$

If $\tilde{W}_K$ is the space of subscales within each element, the problem can also be written as

$$a(\tilde{u}, \tilde{v}) = l(\tilde{v}) - a(u_h, \tilde{v}) \quad \forall \tilde{\boldsymbol{v}} \in \tilde{W}_K$$

Let us consider the finite element subspace of $\tilde{W}_K$:

$$B_h := \mathrm{Span}\{B_1(\boldsymbol{x}), ..., B_{n_{\mathrm{bub}}}(\boldsymbol{x})\} \subset \tilde{W}_K$$

We approximate:

$$\tilde{u} \approx \tilde{u}_h = \sum_{i=1}^{n_{\text{bub}}} B_i(\boldsymbol{x})\tilde{u}_i \quad \text{in } K$$

Then, for each element:

$$\sum_{i=1}^{n_{\text{bub}}} a(B_j, B_i)\tilde{u}_i = l(B_j) - a(u_h, B_j)$$

$$= \int_K B_j(\boldsymbol{x})\left[f - \mathcal{L}(u_h)\right] \quad \forall j = 1, \ldots, n_{\text{bub}}$$

If we call $a_{ij} = a(B_j, B_i)$ and a_{ij}^{-1} are the components of the inverse matrix, we have, in each element K:

$$\tilde{u}_h(\boldsymbol{x}) = \sum_{i,j=1}^{n_{\text{bub}}} \int_K B_i(\boldsymbol{x}) a_{ij}^{-1} B_j(\boldsymbol{y})\left[\mathcal{L}(u_h) - f\right](\boldsymbol{y})d\boldsymbol{y}$$

We note that:

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \approx \boldsymbol{g}_h(\boldsymbol{x},\boldsymbol{y}) = \sum_{i,j=1}^{n_{\text{bub}}} B_i(\boldsymbol{x}) a_{ij}^{-1} B_j(\boldsymbol{y})$$

It remains to specify which are the bubble functions that span B_h within each element. Ideally:

$$\mathcal{L}(\tilde{u}) = f - \mathcal{L}(u_h) \iff \sum_{i=1}^{n_{\text{bub}}} \mathcal{L}(B_i)\tilde{u}_i = f - \sum_{j=1}^{n_{\text{nod}}} \mathcal{L}(N_j) u_j$$

where n_{nod} is the number of nodes in the element under consideration, the nodal shape functions are N_j and the nodal values of the finite element unknown are u_j.

Let $n_{\text{bub}} = n_{\text{nod}} + 1$ and let us construct the B_i's such that

$$\mathcal{L}(B_i) = -\mathcal{L}(N_i), \quad i = 1, ..., n_{\text{nod}}$$
$$\mathcal{L}(B_{n_{\text{bub}}}) = f$$

Then

$$\tilde{u}_i = u_i, \quad i = 1, ..., n_{\text{nod}}, \quad \text{and} \quad \tilde{u}_{n_{\text{bub}}} = 1 \iff$$
$$\tilde{u}(\boldsymbol{x}) = \sum_{i=1}^{n_{\text{nod}}} B_i(\boldsymbol{x})u_i + B_{n_{\text{bub}}}(\boldsymbol{x})$$

The drawback is that one has to solve $n_{\text{nod}} + 1$ problems of the form

$$\mathcal{L}(w) = \phi \quad \text{in} \quad K$$
$$w = 0 \quad \text{on} \quad \partial K$$

The bubble constructed this way are called residual free.

Application to the Stokes problem

$$\mathcal{L}(\boldsymbol{u}, p) = \begin{bmatrix} -\nu\Delta\boldsymbol{u} + \nabla p \\ \nabla\cdot\boldsymbol{u} \end{bmatrix}, \quad \mathcal{L}^*(\boldsymbol{v}, q) = \begin{bmatrix} -\nu\Delta\boldsymbol{v} - \nabla q \\ \nabla\cdot\boldsymbol{v} \end{bmatrix}$$

Taking $\boldsymbol{\tau} = \mathrm{diag}(\tau_1, \tau_1, \tau_2)$ (in 2D) it follows that the stabilizing terms are

$$\begin{aligned} &\sum_K \int_K [-\boldsymbol{\tau}\mathcal{L}^*(\boldsymbol{V}_h)] \cdot [\mathcal{L}(\boldsymbol{U}_h) - \mathcal{F}] \\ &= \sum_K \int_K \tau_1 \left(\nu\Delta\boldsymbol{v}_h + \nabla q_h\right) \cdot \left(-\nu\Delta\boldsymbol{u}_h + \nabla p_h - \boldsymbol{f}\right) \\ &\quad + \sum_K \int_K \tau_2 (\nabla\cdot\boldsymbol{v}_h)(\nabla\cdot\boldsymbol{u}_h) \end{aligned}$$

These terms are the same as for GLS, except for the sign of the viscous term. However, the same convergence results apply.