A STABLE, DIRECT SOLVER FOR THE GRADIENT EQUATION

ROB STEVENSON

Abstract. A new finite element discretization of the equation \( \text{grad} \, p = g \) is introduced. This discretization gives rise to an invertible system that can be solved directly, requiring a number of operations proportional to the number of unknowns. We prove an optimal error estimate, and furthermore show that the method is stable with respect to perturbations of the right-hand side \( g \).

We discuss a number of applications related to the Stokes equations.

1. Introduction and applications

This paper concerns a finite element discretization of the following problem: On some bounded, connected domain \( \Omega \subset \mathbb{R}^2 \) with Lipschitz continuous boundary \( \Gamma \), and for some right-hand side \( g = (g_1, g_2)^T \) with \( 0 = \text{rot} \, g := \partial_1 g_2 - \partial_2 g_1 \), find \( p \) with \( \int_\Omega p \, dx = 0 \), such that
\[
\text{grad} \, p = g.
\]

More precisely, we consider this problem in its variational form: Let \( L^2_0(\Omega) := \{ q \in L^2(\Omega) : \int_\Omega q \, dx = 0 \} \), \( H_0(\text{div}; \Omega) := \{ v \in L^2(\Omega)^2 : \text{div} \, v \in L^2(\Omega), v \cdot n \mid_\Gamma = 0 \} \),
\[
b(v, p) := -\int_\Omega p \, \text{div} \, v \, dx,
\]
and let \( V \subset H_0(\text{div}; \Omega) \) be a Hilbert space that will be specified below. Given \( g \in V' \) with \( g(v) = 0 \) when \( \text{div} \, v = 0 \), we search for \( p \in L^2_0(\Omega) \) such that
\[
b(v, p) = g(v) \quad (v \in V). \tag{1.1}
\]

It is well-known that (1.1) has a unique solution \( p \), with \( \|p\|_{L^2} \lesssim \|g\|_{V'} \), when
\[
|b(v, q)| \lesssim \|v\|_V \|q\|_{L^2} \quad (v \in V, q \in L^2_0(\Omega)) \tag{1.2}
\]
and
\[
\|q\|_{L^2} \lesssim \sup_{0 \neq v \in V} \frac{|b(v, q)|}{\|v\|_V} \quad (q \in L^2_0(\Omega)), \tag{1.3}
\]
see e.g. [Brez]. (Here and subsequently, by \( C \lesssim D \) we mean that \( C \) can be bounded by a multiple of \( D \), independently of parameters which \( C \) and \( D \) may depend on. Obviously, \( C \gtrsim D \) is defined as \( D \lesssim C \), and \( C \approx D \) as \( C \lesssim D \) and \( C \gtrsim D \).) It is known (see [GRS, Corollary 2.4]) that (1.3) is valid for \( V = H_0^1(\Omega)^2 \), and thus also for \( V = H_0^1(\Omega)^2 \).

Received by the editor April 28, 1998.
2000 Mathematics Subject Classification. Primary 65N30, 65F05, 42C40, 76D05, 35Q30.
Key words and phrases. LBB-stability, Stokes equations, multiscale bases, direct solver.
With $V = H^1_0(\Omega)^2$ or $V = H_0(\text{div}; \Omega)$ respectively, problem (1.1) naturally arises in connection with the (Navier)-Stokes problem or a mixed formulation of the Poisson equation. Our discussion of applications will be focussed on the Stokes problem.

Assuming that $\Omega$ is a polygon, starting from some conforming initial triangulation $\tau_0$ of $\Omega$, we consider a sequence of triangulations $(\tau_k)_{k \geq 0}$, where $\tau_{k+1}$ is constructed from $\tau_k$ by subdividing each triangle from $\tau_k$ into four congruent subtriangles. Although generalizations to piecewise polynomials of higher order are possible, for ease of presentation we consider trial spaces $Q_k$ of the lowest order. That is, we take $Q_k$ as the space of piecewise constants with respect to $\tau_k$ with zero mean. In Section 2, we will construct test spaces $V_k \subset H^1_0(\Omega)^2$ such that

$$\dim V_k = \dim Q_k,$$

and

$$\gamma := \inf_k \inf_{q_k \in Q_k} \sup_{v_k \in V_k} \frac{|b(v_k, q_k)|}{\|v_k\|_{H^1(\Omega)^2} \|q_k\|_{L^2}} > 0,$$

which latter property is known as the Ladyshenskaja-Babuška-Brezzi (LBB) stability condition.

Because of (1.4) and (1.5), for any $g_k \in V_k'$, the problem of finding $p_k \in Q_k$ such that

$$b(v_k, p_k) = g_k(v_k) \quad (v_k \in V_k)$$

has a unique solution. Moreover, it will be shown that for usual approximations $g_k$ of the right-hand side $g$ of (1.1), the square system (1.6) can be constructed and solved in $O(\dim Q_k)$ operations.

**Remark 1.1.** In fact the results from this paper can be generalized to certain types of locally refined triangulations and corresponding spaces $Q_k$. More precisely, those triangulations are covered where

- a triangle from $\tau_{k+1}$ either is a triangle from $\tau_k$ or it is generated by subdividing a triangle from $\tau_k$ into four congruent subtriangles,
- a triangle that is contained in both $\tau_k$ and $\tau_{k+1}$ is part of $\tau_\ell$ for any $\ell \geq k$,
- two triangles from $\tau_k$ that have a nonempty intersection have comparable diameters, uniformly in $k$.

Yet, since it requires some technicalities to show that in these local refinement cases the resulting system can be constructed (and solved) in $O(\dim Q_k)$ operations, for ease of presentation in the remainder of this paper we restrict ourselves to the uniform refinement case.

From (1.4), (1.5) we obtain the following optimal error estimate:

**Theorem 1.2.** For $g_k, g \in H^{-1}(\Omega)^2$, such that $g(v) = 0$ when $\text{div} \, v = 0$, let $p_k \in Q_k$, $p \in L^2_0(\Omega)$ be the solutions of

$$b(v_k, p_k) = g_k(v_k) \quad (v_k \in V_k),$$

$$b(v, p) = g(v) \quad (v \in H^1(\Omega)^2).$$

Then

$$\|p - p_k\|_{L^2} \leq (1 + \frac{\sqrt{2}}{\bar{\gamma}}) \inf_{q_k \in Q_k} \|p - q_k\|_{L^2} + \frac{1}{\bar{\gamma}}\|g - g_k\|_{H^{-1}(\Omega)^2}.$$
Proof. With \( \tilde{\psi}_k \in Q_k \) being the solution of \( b(\mathbf{v}_k, \tilde{\psi}_k) = g(\mathbf{v}_k) \) \((\mathbf{v}_k \in V_k)\), from (1.5) we infer that

\[
\| \tilde{\psi}_k - p_k \|_{L^2} \leq \frac{1}{\gamma} \sup_{0 \neq \mathbf{v}_k \in V_k} \frac{|g(\mathbf{v}_k) - g_k(\mathbf{v}_k)|}{\| \mathbf{v}_k \|_{(H^1)^2}} \leq \frac{1}{\gamma} \| g - g_k \|_{(H^{-1})^2}.
\]

Let

\[
\sigma := \inf_k \inf_{0 \neq q_k \in Q_k} \sup_{0 \neq \mathbf{v}_k \in V_k} \frac{|b(\mathbf{v}_k, q_k)|}{\| \text{div} \mathbf{v}_k \|_{L^2} \| q_k \|_{L^2}}.
\]

For arbitrary \( q_k \in Q_k \), we have \( \| p - \tilde{\psi}_k \|_{L^2} \leq \| p - q_k \|_{L^2} + \| q_k - \tilde{\psi}_k \|_{L^2} \), whereas

\[
\| q_k - \tilde{\psi}_k \|_{L^2} \leq \frac{1}{\sigma} \sup_{0 \neq \mathbf{v}_k \in V_k} \left( \frac{|b(\mathbf{v}_k, q_k - \tilde{\psi}_k)|}{\| \text{div} \mathbf{v}_k \|_{L^2}} \right) \leq \frac{1}{\sigma} \| q_k - \tilde{\psi}_k \|_{L^2},
\]

so that \( \| p - \tilde{\psi}_k \|_{L^2} \leq \left( 1 + \frac{1}{\sigma} \right) \inf_{q_k \in Q_k} \| p - q_k \|_{L^2} \). The proof is completed by noting that \( \| \text{div} \mathbf{v}_k \|_{L^2} \leq \sqrt{2} \| \mathbf{v}_k \|_{(H^1)^2} \) and thus \( \frac{1}{\sigma} \leq \frac{\sqrt{2}}{\gamma} \).

Remark 1.3. In the above theorem, suppose that also \( g_k \) satisfies \( g_k(\mathbf{v}) = 0 \) when \( \text{div} \mathbf{v} = 0 \). Then the continuous problem of finding \( \tilde{\psi} \in L^2(\Omega) \) with \( b(\mathbf{v}, \tilde{\psi}) = g_k(\mathbf{v}) \) \((\mathbf{v} \in H^1_0(\Omega))\) has a solution, and the stability of this problem shows that there exists a \( C > 0 \) such that \( \| p - \tilde{\psi} \|_{L^2} \leq C \| g - g_k \|_{(H^{-1})^2} \). The above proof shows that \( \| p - p_k \|_{L^2} \leq \left( 1 + \frac{1}{\sigma} \right) \inf_{q_k \in Q_k} \| p - q_k \|_{L^2} \), and so

\[
\| p - p_k \|_{L^2} \leq \left( 1 + \frac{1}{\sigma} \right) \inf_{q_k \in Q_k} \| p - q_k \|_{L^2} + \| g - g_k \|_{(H^{-1})^2}.
\]

In other words, for this case it is already sufficient when \( \sigma > 0 \). Yet, a crucial point of this paper is that the much stronger LBB condition \((1.5)\) is valid, which yields stability of the gradient solver with respect to perturbations of \( g \) in any direction in \( H^{-1}(\Omega)^2 \).

We now discuss some applications of our gradient solver and comment on some existing alternative approaches. Consider the Stokes equations in its primitive variables: Given \( \mathbf{f} = (f_1, f_2)^T \), find \( \mathbf{u} = (u_1, u_2)^T \) and \( p \), with \( \int_{\Omega} pdx = 0 \), satisfying

\[
-\Delta \mathbf{u} + \text{grad} \ p = \mathbf{f} \quad \text{in } \Omega, \\
\text{div} \ \mathbf{u} = 0 \quad \text{in } \Omega, \\
\mathbf{u} = 0 \quad \text{on } \partial\Omega.
\]

For \( \Omega \) being a bounded, convex polygon and \( \mathbf{f} \in L^2(\Omega)^2 \), it is known that the unique solution \((\mathbf{u}, p)\) is in \((H^2(\Omega) \cap H^1_0(\Omega))^2 \times H^1(\Omega)\) with \( \| \mathbf{u} \|_{(H^2)^2} + \| p \|_{H^1} \lesssim \| \mathbf{f} \|_{(L^2)^2} \).

Standard mixed finite element discretizations yield approximations for the velocities which are only discretely divergence-free. Approaches to obtain approximate velocities that are exactly divergence-free are based on approximating a “stream-function” \( \psi \), which is a function that satisfies \( \mathbf{u} = \text{curl} \ \psi := (\partial_2 \psi, -\partial_1 \psi)^T \). Indeed, note that an approximation \( \psi_k \in H^1(\Omega) \) of \( \psi \) yields an approximate velocity vector \( \mathbf{u}_k := \text{curl} \ \psi_k \in \mathbf{H}(\text{div};\Omega) \) with \( \text{div} \mathbf{u}_k = 0 \).

A computation of an approximation \( \psi_k \) of \( \psi \) can be based on the biharmonic equation

\[
\Delta^2 \psi = \text{rot} \ \mathbf{f} \quad \text{in } \Omega, \\
\psi = \partial_n \psi = 0 \quad \text{on } \partial\Omega.
\]
Yet, discretizing this equation requires $C^1$, or, in case of nonconforming approximations, “nearly” $C^1$ finite elements.

An alternative approach (cf. [GR86], Ch.III, §§2-3) is to discretize the problem of finding $\psi \in \mathcal{H}^1_0(\Omega)$, $\omega \in \mathcal{H}^1(\Omega)$ such that

\begin{align}
(\text{curl } \omega, \text{curl } \phi)(L^2)^2 &= f(\text{curl } \phi) \quad (\phi \in \mathcal{H}^1_0(\Omega)),
(\text{curl } \psi, \text{curl } \mu)(L^2)^2 &= (\omega, \mu)(L^2) \quad (\mu \in \mathcal{H}^1(\Omega)),
\end{align}

(1.7)

where $\omega = \text{rot } \mathbf{u}$ is called the vorticity.

Instead of solving for $\psi$ and $\omega = \nabla \psi$, it is also possible to set up equations for $\psi$ and all its second derivatives $\partial^2_{ij} \psi$, leading to the so-called Hellan-Herrmann-Johnson scheme (cf. [GR86], Ch.III, §4).

The above formulations have in common that the pressure approximation, “nearly” in finite element space $\hat{\mathcal{V}}_k$ and all its second derivatives $\psi_k$, with a test space $\hat{\mathcal{V}}_k$ can compute finite element space with respect to $\tau_k$ contains an internal vertex, then $\dim \hat{\mathcal{V}}_k > \dim Q_k$. Yet, if $\omega_k$ is such that $f(\hat{\mathbf{v}}_k) = (\text{curl } \omega_k, \hat{\mathbf{v}}_k)(L^2)^2$ for all divergence-free $\hat{\mathbf{v}}_k \in \hat{\mathcal{V}}_k$, then the system (1.8)
with $V_k$ replaced by $\tilde{V}_k$ has a unique solution $\hat{\nu}_k \in Q_k$. To handle the case of a perturbed $\omega_k$, analogously to the construction that will be set out in Section 2, a subspace of $\tilde{V}_k$ can be selected with dimension equal to that of $Q_k$ such that (1.11) is retained. Yet, in any case, since $\tilde{V}_k \not\subset H^1(\Omega)^2$, (1.11) with $\|\tilde{\nu}_k\|_{H^1(\Omega)^2}$ replaced by $\|\tilde{\nu}_k\|_{(H^1)^2}$ is not valid, and so a bound on $\|p - \hat{p}_k\|_{L^2}$ similar to (1.9) will depend on a norm of $\omega - \omega_k$ which is stronger than the $L^2$-norm.

To discuss the final application of our gradient solver, we consider the Stokes point problem (1.14). Let

$$a(u, v) + b(v, p) = f(v) \quad (v \in H^1(\Omega)^2),$$

$$b(u, q) = 0 \quad (q \in L^2(\Omega)),$$

where

$$a(w, v) := \int_{\Omega} \nabla w : \nabla v dx.$$

We describe the usual mixed finite element discretization. For $k \in \mathbb{N}$, let $S_k$ and $\tilde{Q}_k$ be finite element spaces that serve as (increasingly better) approximations of $H^1_0(\Omega)^2$ and $L^2_0(\Omega)$ respectively. It is no restriction to assume that $\tilde{Q}_k \subset L^2_0(\Omega)$, but we do allow nonconforming finite element spaces $S_k$, i.e., $S_k \not\subset H^1_0(\Omega)^2$. As a consequence, we generally need extensions of $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$ to a scalar product $a_k(\cdot, \cdot)$ and a bilinear form $b_k(\cdot, \cdot)$ on $(H^1_0(\Omega)^2 + S_k) \times (H^1_0(\Omega)^2 + S_k)$ and $(H^1_0(\Omega)^2 + S_k) \times L^2_0(\Omega)$ respectively. We equips $H^1_0(\Omega)^2 + S_k$ with the energy-norm $\| \cdot \|_{1,k} = \sqrt{a_k(\cdot, \cdot)}$. We assume that $b_k(\cdot, \cdot)$ is uniformly bounded, and that the following LBB-condition is valid:

$$\inf \{a_k(u_k, v_k) + b_k(v_k, p_k) : (u_k, v_k, p_k) \in S_k \times S_k \times \tilde{Q}_k\} > 0.$$

Assuming that $f \in S'_k$, we arrive at the following approximation scheme: Find

$$u_k \in S_k, \quad p_k \in \tilde{Q}_k$$

such that

$$a_k(u_k, v_k) + b_k(v_k, p_k) = f(v_k) \quad (v_k \in S_k),$$

$$b_k(u_k, q_k) = 0 \quad (q_k \in \tilde{Q}_k).$$

Because of (1.13), this system has a unique solution, and, depending on the possible consistency error, the approximation properties of $S_k$ and $\tilde{Q}_k$, and the regularity of $u$ and $p$, appropriate bounds on $\|u - u_k\|_{1,k}$ and $\|p - p_k\|_{L^2}$ are known.

Defining the space of discretely divergence-free velocities

$$Z_k = \{v_k \in S_k : b_k(v_k, q_k) = 0 \forall q_k \in \tilde{Q}_k\},$$

the velocity component $u_k$ of the solution of (1.14) can be characterized as the unique solution of the problem: Find $u_k \in Z_k$ such that

$$a_k(u_k, v_k) = f(v_k) \quad (v_k \in Z_k).$$

For some pairs $(S_k, \tilde{Q}_k)$ a local basis for $Z_k$ is known. This opens a way to compute $u_k$ by solving the elliptic problem (1.13) only, instead of solving the original saddle-point problem (1.14).

An example of a pair for which such a basis is available (cf. [Cro72, Tho81]) is given by the case that, with respect to some conforming triangulation $\tau_k$ of $\Omega$,
S_k = S^2_k with S_k being the nonconforming P_1 finite element space, i.e.,

S_k = \{ v \in \prod_{T \in \tau_k} P_1(T), \; v \text{ is continuous at the midpoints } m_e \text{ of the interelement boundaries } e, \; \text{and it vanishes at the midpoints } m_e \text{ of edges along } \partial \Omega \},

and \( \tilde{Q}_k (= Q_k) \) is the space of piecewise constant functions with zero mean value.

In this case

\[ a_k(w_k, v_k) := \sum_{T \in \tau_k} \int_T \nabla w_k : \nabla v_k \, dx, \quad b_k(v_k, q_k) := -\sum_{T \in \tau_k} \int_T q_k \, \text{div} \, v_k \, dx. \]

For this pair, optimal multi-grid, domain decomposition and cascade multi-level methods for solving \( u_k \) from (1.15) were proposed and analyzed in [Bre90], [Bre96] and [Ste99] respectively.

Other examples of pairs, all involving nonconforming spaces \( S_k \), for which a local basis for the resulting \( Z_k \) has been constructed can be found in [CSS86] [Tur94]. Constructions based on wavelets were discussed in [Urb96].

Knowing \( u_j \), we are left with the problem of finding an approximation for the pressure. The obvious approach is to solve \( p_k \in \tilde{Q}_k \) from (1.16)

\[ b_k(v_k, p_k) = f(v_k) - a_k(u_k, v_k) \quad (v_k \in S_k). \]

Indeed, existence and uniqueness of this \( p_k \) are already known, and suitable error estimates are available.

The number of equations in (1.16) exceeds the number of unknowns. However, since (1.16) is trivially valid for \( v_k \in Z_k \), it is sufficient to satisfy (1.17)

\[ b_k(v_k, p_k) = f(v_k) - a_k(u_k, v_k) \quad (v_k \in R_k), \]

where \( R_k \) is some subspace of \( S_k \) satisfying \( R_k \cap Z_k = \{ 0 \} \) and \( \dim R_k = \dim \tilde{Q}_k \), or equivalently, \( S_k = R_k \oplus Z_k \).

In [CSS86] [Tur94] [Urb96] we find similar choices of \( R_k \) which give rise to direct solvers that can be implemented efficiently. In the following, we describe the idea for the nonconforming P_1, piecewise constant finite element pair.

For all pairs \( T, \hat{T} \in \tau_k \), such that \( e := T \cap \hat{T} \) is an edge, let

\[ \mathbf{w}_e = |e|^{-1} g_e \mathbf{n}_e \in S_k, \]

where \( \mathbf{n}_e \) is a unit vector normal to \( e \), and \( g_e \in S_k \) is the standard basis function defined by \( g_e(m_e) = \begin{cases} 1 & \text{if } \check{e} = e, \\ 0 & \text{if } \check{e} \neq e, \end{cases} \) see Figure 1. It is easily verified that, assuming \( \mathbf{n}_e \) points into \( T \),

\[ \text{div} \, \mathbf{w}_e = \partial_{n_e} g_e = \begin{cases} -(\text{vol } T)^{-1} & \text{on } T, \\ (\text{vol } \hat{T})^{-1} & \text{on } \hat{T}, \\ 0 & \text{elsewhere}. \end{cases} \]

**Figure 1.** \( g_e \) and \( \mathbf{w}_e \).
Now on some $T_0 \in \tau_k$, fix $p_k$, and let $T_1 \in \tau_k$ be such that $e = T_0 \cap T_1$ is an edge. We will call such triangles neighbours. Then (1.16) for $v_k = w_e$ determines $p_k|_{T_1}$ uniquely. In this way, by marching from neighbour to a still unvisited neighbour, $p_k \in \prod_{T \in \tau_k} P_0(T)$ can be fixed completely. In the end, by subtracting a suitable constant, $p_k$ is mapped into $Q_k$. Clearly, this procedure for computing $p_k$ is equivalent to solving (1.17), where $R_k = \text{span}\{w_e\}$ with $e$ running over all edges that were crossed in the marching process.

In view of Remark 1.3, a potential pitfall with this approach of solving (1.17) is that in practice, instead of the exact solution $u_k$, only an approximation $\tilde{u}_k \in Z_k$ will be at one’s disposal, since (1.15) will have been solved by an iterative method. With $u_k$ replaced by $\tilde{u}_k \neq u_k$, the system (1.16) does not have a solution, but because of (1.13) and $S_k = R_k \oplus Z_k$, the system (1.17) does, which solution we denote by $\tilde{p}_k$. With $T_k : Z_k \rightarrow \tilde{Q}_k$ being the linear operator defined by

\[ b_k(v_k, T_k w_k) = -a_k(w_k, v_k) \quad (v_k \in R_k), \]

we have

\[ p_k - \tilde{p}_k = T_k(u_k - \tilde{u}_k). \]

Remark 1.4. It is easily verified that

\[ \|T_k\|_{L^2(\Omega)} \leq \left( \inf_{0 \neq q_k \in Q_k} \sup_{0 \neq w_k \in R_k} \frac{|b_k(v_k, q_k)|}{\|v_k\|_{1,k} \|q_k\|_{L^2}} \right)^{-1}, \]

so that LBB-stability of $(\tilde{Q}_k, R_k)$ would imply stability of the above procedure for finding the pressure.

For the nonconforming $P_1$, piecewise constant finite element pair, we have computed $\|T_k\|_{L^2(\Omega)}$ numerically in the following situation: $\Omega = [0, 1]^2$, $\tau_k$ is a uniform partition of $\Omega$ into right-angled isosceles triangles of which the equal sides have length $h_k := 2^{-k}$, and $R_k = \text{span}\{w_e\}$, where $e$ runs over all edges corresponding to the dotted lines as indicated in Figure 2 for the case $k = 2$. The results given in Table 1 indicate that $\|T_k\|_{L^2(\Omega)} \approx h_k^2$. In the case of $f \in L^2(\Omega)^2$, and thus

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|T_k|_{L^2(\Omega)}$</td>
<td>2.3</td>
<td>10</td>
<td>41</td>
<td>160</td>
<td>650</td>
<td>2600</td>
<td>10000</td>
</tr>
</tbody>
</table>

Figure 2. Crossed edges in the marching process defining $R_k$.
\[ \|u - u_k\|_{1,k} = O(h_k) \text{ and } \|p - p_k\|_{L^2} = O(h_k), \]

this means that the algebraic error \(u_k - \hat{u}_k\) in \(\|1,k\)-norm should be of order \(h_k^2\) to be sure that the resulting algebraic error in the pressure will not dominate the discretization error. That is, thinking of a linearly convergent iterative method, and an initial error that is of order 1, one should triple the number of iterations sufficient for solving the velocities from \((1.13)\), if one wants to use the outcome to compute the pressure afterwards using this marching process.

An alternative approach to solve for the pressure can be based on our gradient solver: Given some approximation \(\hat{u}_k \in S_k\) of \(u\), solve \(\tilde{p}_k \in Q_k\) from

\[ b(v, \tilde{p}_k) = f(v) - a_k(\hat{u}_k, v_k) \quad (v_k \in V_k). \]

Since \(u\) and \(p\) satisfy

\[ b(v, p) = f(v) - a_k(u, v) \quad (v \in H^1_0(\Omega)^2), \]

an application of Theorem \((1.2)\) shows the optimal error estimate

\[ \|p - \tilde{p}_k\|_{L^2} \leq (1 + \frac{\sqrt{2}}{\gamma}) \inf_{q_k \in Q_k} \|p - q_k\|_{L^2} + \frac{1}{\gamma} \|u - \hat{u}_k\|_{1,k}, \]

where now \(\|p - \tilde{p}_k\|_{L^2}\) and \(\|u - \hat{u}_k\|_{1,k}\) are the “total” errors. In particular, with this approach an algebraic error in the approximate velocities is not blown up.

2. Construction and implementation of the gradient solver

Let \(\tau_0\) be some conforming triangulation of a polygon \(\Omega \subset \mathbb{R}^2\), and for \(k \geq 0\), let \(\tau_{k+1}\) be constructed from \(\tau_k\) by subdividing each triangle from \(\tau_k\) into four congruent subtriangles. For each \(k\), \(Q_k\) is defined as the space of piecewise constants with respect to \(\tau_k\) with zero mean. We will construct spaces \(V_k\) satisfying both \(\dim V_k = \dim Q_k\) \((1.3)\) and the LBB condition \((1.5)\).

We recall the marching process discussed in \((1.4)\), which however here will be applied on the coarsest level only. Starting from some \(T \in \tau_0\), until we have been in all triangles in \(\tau_0\), we travel from already visited triangles to yet unvisited neighbours, putting the edges that were crossed between such neighbours in a set called \(E_0\). In case \(\tau_0\) does not contain internal vertices and \(\Omega\) is simply-connected, \(E_0\) will be the set of all internal edges in \(\tau_0\), but otherwise \(E_0\) will be a proper subset of that set. In any case the number of elements in \(E_0\) will be equal to \(\dim Q_0\). For each \(e \in E_0\), \(e = T \cap \tilde{T}\) with \(T, \tilde{T} \in \tau_0\), let \(w_e \in H^1_0(T \cup \tilde{T})^2\) be some function such that \(\int_T w_e \cdot n_e ds \neq 0\), where \(n_e\) is a unit vector normal to \(e\). We define \(V_0 = \text{span}\{w_e : e \in E_0\}\).

For each \(0 \neq q \in Q_0\), there exists an \(e \in E_0\), \(e = T \cap \tilde{T}\) with \(T, \tilde{T} \in \tau_0\), such that \(q|_T \neq q|_{\tilde{T}}\). From \(|b(w_e, q)| = |(q|_T - q|_{\tilde{T}}) \int_T w_e \cdot n_e ds| \neq 0\), we conclude that

\[ \inf_{0 \neq q \in Q_0} \sup_{0 \neq v \in V_0} \frac{|b(v, q)|}{\|v\|_{(H^1)^2} \|q\|_{L^2}} > 0. \]

For \(m \geq 1\), \(m \in \mathbb{N}\), we define \(E_m\) as the set of new edges in \(\tau_m\), that is, all edges that were added to refine \(T \in \tau_{m-1}\). For each \(e \in E_m\), \(e = T \cap \tilde{T}\) with \(T, \tilde{T} \in \tau_m\), let \(\ell\) be the line connecting both vertices of \(T\) and \(\tilde{T}\) which are not on \(e\). Since, because of the refinement procedure, \(T \cup \tilde{T}\) is a parallelogram, \(\ell\) intersects \(e\) at its midpoint \(m_e\). Let \(g_e \in H^1_0(T \cup \tilde{T})\) be the function that is \(1\) at \(m_e\), and that is linear on the four triangles generated by adding the line \(\ell\). For a nonzero vector \(s_e\) in \(m_e\)
pointing along $\ell$, and with $\mathbf{n}_e$ a unit vector normal to $e$, say pointing into $T$, we put
\[ \mathbf{w}_e = \frac{2g_ese}{|e|} \mathbf{s}_e \cdot \mathbf{n}_e \in H^1_0(T \cup \hat{T})^2, \]
see Figure 3. By construction, $\text{div } \mathbf{w}_e = \frac{2 |e| s_e \cdot n_e}{|e|} \partial_n g_e$ is constant on both $T$ and $\hat{T}$, and in particular
\[ \text{div } \mathbf{w}_e = \begin{cases} 
- (\text{vol } T)^{-1} & \text{on } T, \\
(\text{vol } \hat{T})^{-1} & \text{on } \hat{T}, \\
0 & \text{elsewhere}. 
\end{cases} \]

We infer that
\[ (2.2) \quad \text{div } \mathbf{w}_e \in Q_m, \quad \text{div } \mathbf{w}_e \perp L^2 Q_{m-1}, \]
and, with $W_k := \text{span} \bigcup_{m=1}^{k} \{ \mathbf{w}_e : e \in E_m \}$, that
\[ (2.3) \quad \text{div } W_k = Q_k \ominus L^2 Q_0. \]

Finally, using the fact that $\text{supp } \mathbf{w}_e \cap \text{supp } \mathbf{w}_{\mathbf{h}} = 0$ for all $e, \mathbf{h} \in E_m$ that are not contained in a common $T \in \tau_{m-1}$, a homogeneity argument shows that
\[ (2.4) \quad 2^n \| \cdot \|_{(L^2)^2} \lesssim \| \text{div } \cdot \|_{L^2} \quad \text{on } \text{span} \{ \mathbf{w}_e : e \in E_m \}. \]

Note that the latter relation is valid uniformly in all triangulations $\tau_0$ that satisfy some minimal angle condition.

Defining $V_k = V_0 + W_k$, we see that
\[ \dim V_k = \# \left( \bigcup_{m=0}^{k} E_m \right) = \dim Q_k. \]

In Theorem 2.4 we will prove that
\[ (2.5) \quad \inf_k \inf_{0 \neq q \in Q_k} \sup_{0 \neq \mathbf{v} \in W_k} \frac{|b(\mathbf{v}, q)|}{\|\mathbf{v}\|_{(H^1)^2} \|q\|_{L^2}} > 0. \]

Since by (2.3), $b(\mathbf{W}_k, Q_0) = 0$, an application of Lemma 2.1 given below now shows that (2.1) and (2.5) imply that
\[ \inf_k \inf_{0 \neq q \in Q_k} \sup_{0 \neq \mathbf{v} \in V_k} \frac{|b(\mathbf{v}, q)|}{\|\mathbf{v}\|_{(H^1)^2} \|q\|_{L^2}} > 0, \]
i.e., both (1.4) and (1.5) are valid.
Lemma 2.1. Let $Q = Q_1 + Q_2$ and $V = V_1 + V_2$ be normed linear spaces, $b$ a bilinear form on $V \times Q$ for which

$$b(V_2, Q_1) = 0,$$

and $C, \gamma_1, \gamma_2 > 0$ constants such that

$$\sup_{0 \neq q \in Q, 0 \neq v \in V} \frac{|b(v, q)|}{\|v\|_V \|q\|_Q} \leq C, \quad \inf_{0 \neq q \in Q} \sup_{0 \neq v \in V} \frac{|b(v, q)|}{\|v\|_V \|q\|_Q} \geq \gamma_i \quad (i \in \{1, 2\}).$$

Then there exists a $\gamma > 0$, depending only on $C, \gamma_1, \gamma_2$, such that

$$\inf_{0 \neq q \in Q} \sup_{0 \neq v \in V} \frac{|b(v, q)|}{\|v\|_V \|q\|_Q} \geq \gamma.$$

Proof. Let $q = q_1 + q_2$, where $q_1 \in Q_1$, $q_2 \in Q_2$.

In case $\|q_2\|_Q \leq \frac{2\gamma_1}{\gamma_2} \|q_1\|_Q$, and thus $\|q\|_Q \leq (1 + \frac{2\gamma_1}{\gamma_2}) \|q_1\|_Q$, let $v_1 \in V_1$ be such that $|b(v_1, q_1)| \geq \frac{3}{4} \gamma_1 \|v_1\|_V \|q_1\|_Q$. Then

$$|b(v_1, q_1 + q_2)| \geq \frac{3}{4} \gamma_1 \|v_1\|_V \|q_1\|_Q - C \|v_1\|_V \|q_2\|_Q \geq \frac{1}{4} \gamma_1 \|v_1\|_V \|q_1\|_Q \geq \left(\frac{1}{4} \gamma_1 / (1 + \frac{2\gamma_1}{\gamma_2})\right) \|v_1\|_V \|q\|_Q.$$

Otherwise, when $\|q_2\|_Q \geq \frac{2\gamma_1}{\gamma_2} \|q_1\|_Q$, and so $\|q\|_Q \leq (\frac{2\gamma_1}{\gamma_2} + 1) \|q_2\|_Q$, let $v_2 \in V_2$ be such that $|b(v_2, q_2)| \geq \frac{3}{4} \gamma_2 \|v_2\|_V \|q_2\|_Q$. Then

$$|b(q_1 + q_2, v_2)| = |b(q_2, v_2)| \geq \frac{3}{4} \gamma_2 \|v_2\|_V \|q_2\|_Q \geq \left(\frac{3}{4} \gamma_2 / (\frac{2\gamma_1}{\gamma_2} + 1)\right) \|v_2\|_V \|q\|_Q.$$

It remains to prove (2.6). At first for theoretical purposes, but later also for constructing an efficient implementation of the gradient solver, for $m \in \mathbb{N}$, we define $\tilde{S}_m$ as the conforming $P_1$ finite element space with respect to a refined triangulation $\tilde{\tau}_m$ defined below, i.e., $\tilde{S}_m = C(\Omega) \cap H^1_0(\Omega) \cap \prod_{T \in \tilde{\tau}_m} P_1(T)$. The triangulations $\tilde{\tau}_m$ are constructed from $\tau_m$ by subdividing each $T \in \tau_m$ into 6 subtriangles by connecting the vertices with the midpoints on the opposite edges, see Figure 4. The resulting spaces $\tilde{S}_m$ are nested, i.e. $\tilde{S}_m \subset \tilde{S}_{m+1}$, and for $m \geq 1$,

$$\{w_e : e \in E_m\} \subset \tilde{S}_m := \tilde{S}_m^2.$$

Remark 2.2. The construction of $w_e \in H^1_0(T \cup \tilde{T})^2$ directly generalizes to any pair of triangles $T, \tilde{T}$ that share an edge $e$, and for the line $\ell$ connecting both vertices of $T$ and $\tilde{T}$ that are not on $e$ intersects $e$ in an interior point. The reason why we only considered cases where $\ell$ intersects $e$ at its midpoint is the property (2.6).

Theorem 2.3.

$$\inf_{0 \neq q \in Q_k} \inf_{0 \neq v \in V_k} \sup_{0 \neq q \in Q_0} \frac{|b(v, q)|}{\|v\|_{(H^1)^2} \|q\|_{L^2}} > 0.$$

Figure 4. Construction of the triangulation $\tilde{\tau}_m$ underlying $\tilde{S}_m$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Proof. A so-called strengthened Cauchy-Schwarz inequality is valid on the sequence \((\hat{S}_m)\), i.e., there exists a \(\delta < 1\) such that for \(n > m\),
\[
|\langle \mathbf{v}_m, \mathbf{v}_n \rangle_{H^1}^2| \lesssim \delta^{n-m} \| \mathbf{v}_m \|_{H^1}^2 \| \mathbf{v}_n \|_{L^2}^2 \quad (\mathbf{v}_m, \mathbf{v}_n \in \hat{S}_n).
\]
Combined with the inverse inequality,
\[
\| \mathbf{v}_m \|_{H^1}^2 \lesssim 2^m \| \mathbf{v}_m \|_{L^2}^2 \quad (\mathbf{v}_m \in \hat{S}_m),
\]
we derive that
\[
(2.7) \quad \| \sum_{m=0}^{k} \mathbf{v}_m \|_{H^1}^2 \lesssim \sum_{m=0}^{k} 2^m \| \mathbf{v}_m \|_{L^2}^2 \quad (\mathbf{v}_m \in \hat{S}_m).
\]
Substituting \(\mathbf{v}_m = \sum_{e \in E_m} c_e \mathbf{w}_e\) in (2.7), by (2.4) and (2.2) we find that
\[
\| \sum_{m=1}^{k} \sum_{e \in E_m} c_e \mathbf{w}_e \|_{H^1}^2 \lesssim \sum_{m=1}^{k} \| \sum_{e \in E_m} c_e \text{div} \mathbf{w}_e \|_{L^2}^2 = \| \sum_{m=1}^{k} \sum_{e \in E_m} c_e \text{div} \mathbf{w}_e \|_{L^2}^2,
\]
or \(\| \cdot \|_{H^1} \approx \| \text{div} \cdot \|_{L^2}\) on \(\mathbf{W}_k\) (uniformly in \(k\)).

Since by (2.3) any \(q \in Q_k \cap \mathcal{L}^2\) \(Q_0\) can be written as \(q = \text{div} \mathbf{v}\) for some \(\mathbf{v} \in \mathbf{W}_k\), we arrive at
\[
\| q \|_{L^2} = \frac{| b(\mathbf{v}, q) |}{\| \text{div} \mathbf{v} \|_{L^2}} \approx \frac{| b(\mathbf{v}, q) |}{\| \mathbf{v} \|_{H^1}^2},
\]
which completes the proof. \(\Box\)

Finally, for given \(\mathbf{g}_k \in \mathbf{V}_k^t\), we discuss the implementation of setting up and solving a system corresponding to the problem of finding \(p_k \in Q_k\) satisfying
\[
(2.8) \quad b(\mathbf{v}_k, p_k) = \mathbf{g}_k(\mathbf{v}_k) \quad (\mathbf{v}_k \in \mathbf{V}_k).
\]

Let \(\Phi_0\) be some basis on \(Q_0\). If we equip \(\mathbf{V}_k\) and \(\mathbf{Q}_k\) with bottom-to-top level-wise ordered bases \(\bigcup_{m=0}^{k} \{ \mathbf{w}_e : e \in E_m \}\) and \(\Phi_0 + \bigcup_{m=1}^{k} \{ \text{div} \mathbf{w}_e : e \in E_m \}\) respectively, then (2.8) results in a matrix-vector system
\[
\mathbf{B}_k \mathbf{p}_k = \mathbf{g}_k,
\]
where \(\mathbf{P}_k\) is the representation of \(p_k\) with respect to above basis of \(Q_k\), \(\mathbf{g}_k = [\mathbf{g}_k(\mathbf{w}_e)]_{e \in E_m, 0 \leq m \leq k}\), and \(\mathbf{B}_k\) is the matrix having as elements the application of \(\text{div}\) to all pairs of basis functions from \(\mathbf{V}_k\) and \(\mathbf{Q}_k\) respectively. The multi-level ordering of these bases induces a block partitioning \(\mathbf{B}_k = ([\mathbf{B}_{k}]_{mn})_{0 \leq m, n \leq k}\), with the size of \(\mathbf{B}_{k}\) being \#\(E_m \times \#E_n\). The property (2.2) now implies that \(\mathbf{B}_{k}\) has a block diagonal matrices, with blocks of size \(3 \times 3\). We conclude that \(\mathbf{B}_{k}\) can be inverted in \(O(\text{dim } Q_k)\) operations.

Remark 2.4. If for all \(e \in E_0\), \(e = T \cap \tilde{T}\) with \(T, \tilde{T} \in \tau_0\), the line \(\ell\) connecting both vertices of \(T\) and \(\tilde{T}\) which are not on \(e\) intersects \(e\) at its midpoint, then, just as on levels \(> 0\), \(\mathbf{w}_e\) can be selected in \(\mathbf{S}_0\) with \(\text{div } \mathbf{w}_e \in Q_0\). In this case, \(\mathbf{B}_{k}(e) = 0\) for \(n > 0\), or \(\mathbf{B}_{k}\) will be a block diagonal matrix.

Otherwise, a reasonable approach is to take \(\mathbf{w}_e = \frac{2e}{|e|} n_e\), where \(f_e \in H^1(T \cup \tilde{T})\) is defined by \(f_e(m_e) = 1\), and \(f_e\) is linear on all \(T \in \tau_1\). Note that \(\mathbf{w}_e \in \mathbf{S}_1\), \(\int_{T} \mathbf{w}_e \cdot n_e ds = 1\), and that \(\text{div } \mathbf{w}_e \in Q_1\), which means that \(\mathbf{B}_{k}(e) = 0\) for \(n > 1\).
Since the diameters of the supports of the basis functions $w_e$ of $V_k$ are not all of order $2^{-k}$, but instead range from order 1 to order $2^{-k}$, a straightforward computation of $g_k$, or a sufficiently accurate approximation of this vector involving numerical quadrature, can be expected to demand a number of operations of order $k \dim Q_k$.

Therefore, let us equip $S_k$ with the standard nodal basis $\{v_{m,x} : x \in N_k\}$, where $N_k$ is the set of interior vertices of $\bar\tau_k$. Since diam$(\text{supp} \nu_{k,x}) \approx 2^{-k}$, we may expect that $\tilde g_k = [(g_k^{(1)}(v_{k,x}), g_k^{(2)}(v_{k,x}))]_{x \in N_k}$, or a sufficiently accurate approximation of this vector, can be computed in $O(\dim S_k) = O(\dim Q_k)$ operations.

In view of Remark 2.3, we assume that $V_0 = \text{span} \{w_e : e \in E_0\} \subset S_1$, and so $V_k \subset S_k$ for $k \geq 1$. For $k \geq 1$, let $I_k$ be the representation of the embedding of $V_k$ into $S_k$. Then $g_k = I_k^T \tilde g_k$. With span$\{w_e : e \in E_m\}$ being equipped with the basis $\{w_e : e \in E_m\}$, for $m \geq 1$ let the uniformly sparse matrices $q_m$ and $p_{m+1}$ be the representations of the embeddings span$\{w_e : e \in E_m\} \rightarrow S_m$ and $S_m \rightarrow S_{m+1}$ respectively, and let $q_0$ be the representation of the embedding

$$\text{span}\{w_e : e \in E_0\} \rightarrow S_1.$$ 

With these definitions, the mappings $I_k$ satisfy

$$I_{k+1} = \begin{bmatrix} p_{k+1} & q_{k+1} \end{bmatrix} (k \geq 1), \quad I_1 = \begin{bmatrix} q_0 & q_1 \end{bmatrix}.$$ 

So for the transpose we get

$$I_{k+1}^T = \begin{bmatrix} I_k^T & p_{k+1} \\ q_{k+1} \end{bmatrix} (k \geq 1), \quad I_1^T = \begin{bmatrix} q_0^T \\ q_1^T \end{bmatrix},$$

which induces a top-to-bottom recursive procedure to evaluate $I_k^T$ times vector, in particular to compute $g_k = I_k^T \tilde g_k$, in $O(\dim Q_k)$ operations.

As a result of the computation described above, one obtains a vector $P_k$ that represents the solution $p_k$ with respect to the multi-level basis $\Phi_0 + \bigcup_{m=1}^k \{\text{div} w_e : e \in E_m\}$ of $Q_k$. Yet, one often prefers to have a representation, denoted by $\hat p_k$, of $p_k$ with respect to the canonical basis of $\prod_{T \in \tau_k} P_0(T)$. We follow an analogous procedure as described above. We equip $\prod_{T \in \tau_k} P_0(T)$ with its canonical basis. Let $I_k$ be the representation of the embedding of $Q_k$ into $\prod_{T \in \tau_k} P_0(T)$. Equipping span$\Phi_0$ with the basis $\Phi_0$, and for $m \geq 1$, span$\{\text{div} w_e : e \in E_m\}$ with the basis $\{\text{div} w_e : e \in E_m\}$, let $q_0$ be the representation of the embedding span$\Phi_0 \rightarrow \prod_{T \in \tau_m} P_0(T)$, and for $m \geq 1$ let the uniformly sparse matrices $q_m$ and $p_m$ be the representations of the embeddings span$\{\text{div} w_e : e \in E_m\} \rightarrow \prod_{T \in \tau_m} P_0(T)$ and $\prod_{T \in \tau_{m-1}} P_0(T) \rightarrow \prod_{T \in \tau_m} P_0(T)$ respectively. Then the mappings $I_k$ satisfy

$$I_{k+1} = \begin{bmatrix} p_{k+1} & q_{k+1} \end{bmatrix} (k \geq 0), \quad I_0 = q_0,$$

which yield a bottom-to-top recursive procedure to evaluate $I_k$ times vector, in particular to compute $\hat P_k = I_k P_k$, in $O(\dim \prod_{T \in \tau_k} P_0(T)) = O(\dim Q_k)$ operations.

Summarizing: Assuming that $\tilde g_k$ is available, we can compute $\hat P_k = I_k B_k^{-1} I_k^T \tilde g_k$ taking $O(\dim Q_k)$ operations.
A STABLE, DIRECT SOLVER FOR THE GRADIENT EQUATION

REFERENCES


DEPARTMENT OF MATHEMATICS, UTRECHT UNIVERSITY, P.O. BOX 80.010, NL-3508 TA UTRECHT, THE NETHERLANDS

E-mail address: stevenso@math.uu.nl