TREATMENTS OF DISCONTINUITY AND BUBBLE FUNCTIONS IN THE MULTIGRID METHOD

SHANGYOU ZHANG AND ZHIMIN ZHANG

Abstract. When multilevel finite element spaces are not nested, different intergrid transfer operators would lead to different multigrid algorithms. It is proposed in this paper that discontinuous functions be averaged to continuous functions and that the bubble functions be discarded in the multigrid transferring. Applications of the techniques to various problems are presented with convergence analysis. Numerical comparisons with other existing methods are provided.

1. Introduction

The multigrid method provides optimal-order algorithms for solving large linear systems of finite element and finite difference equations (cf. [3]). The multigrid theory is well established (cf. [12], [16], [3], [15] and [5]). However, in many situations, the multi-level discrete spaces are nonnested due to the nature of the underlying finite elements ([4], [7]–[8], [18], [25]) or due to the special structures of grids (cf. [6], [9] [20], [26], [27]). Some special treatments are then needed to extend the standard multigrid method.

Brenner [7] and Braess-Verfürth [4] have studied \( P_1 \) nonconforming multigrid methods. The intergrid transfer operator in [7] averages the midpoint values of coarse-level functions on the two sides of each edge. This operator is generalized in [4], where the averaging is weighted by the areas of two neighboring triangles, such that the nonconforming multigrid method can be covered by the standard multigrid theory. However, the analysis of [4] is not valid for variable coefficient equations. Brenner [8] and Verfürth [25] also have treated bubble functions in multigrid methods. They have two different averaging operators to transfer a coarse-level bubble (a cubic polynomial vanishing at three edges of a triangle) to four small bubbles on the four subtriangles. In this paper, numerical tests on [7], [4], [8], [25] will be presented along with some tests on several new algorithms.

Different from [7], [8], [4], [25], we average a discontinuous finite element function to a continuous one which is naturally in the higher-level nonconforming finite element (NCF) space. We also discard bubble functions in transferring them to higher levels. By using averaging functions and discarding bubbles we would still retain the smooth components of coarse-level solutions. It is shown that the new multigrid methods converge with constant rates and retain the optimal order of computation for several model problems. The direct estimate of strengthened Cauchy inequalities in [25] would limit the extension of the theory there, while the analysis here...
Let \( C \) be the generic constant independent of the multigrid construction. The rest of the paper is organized as follows. In §2, the \( P_1 \) nonconforming multigrid method is defined. The constant rate of convergence and the optimal order of convergence for the nonconforming multigrid method are proved in §3. Raviart-Thomas mixed triangular elements and their nonconforming finite element correspondence are introduced in §4. Section 5 is devoted to solving Raviart-Thomas finite element equations via nonconforming multigrid methods. A brief discussion on the multigrid method for the Wilson element is given in §6. A multigrid iteration for solving mini-element equations arising from discretizing stationary Stokes equations is defined in §7 and the convergence analysis is presented in §8. We present some numerical results in §9 with discussions on implementing the multigrid method. In some numerical computations, the cubic bubble functions are replaced by linear macro-bubbles to avoid higher order quadratures.

2. \( P_1 \) Nonconforming Elements

For simplicity, we consider the following Poisson equation in a convex, bounded, polygonal domain in \( \mathbb{R}^2 \):

\[
-\Delta u = f \quad \text{in } \Omega, \quad \quad u = 0 \quad \text{on } \partial \Omega,
\]

and its weak formulation: Find \( u \in H_0^1(\Omega) \), such that

\[
a(u, v) = (f, v) \quad \forall v \in H_0^1(\Omega),
\]

where \( a(u, v) := \int_\Omega \nabla u \cdot \nabla v \, dx \) and \( H_0^1(\Omega) \) is the space of Sobolev \( W^{1,2}(\Omega) \) functions with zero trace on \( \partial \Omega \). We note that the analysis here covers general self-adjoint elliptic problems.

Let \( \{ T_k \} \) be a nested family of triangulations on \( \Omega \), where \( T_k \) is obtained by refining each triangle in \( T_{k-1} \) into four congruent subtriangles by connecting midpoints on edges. The mesh sizes of these quasiuniform grids have the relation \( h_k = h_{k-1}/2 \). The work here can be extended to the case of nonquasiuniform meshes as in [27]. Let \( V_k \) be the \( P_1 \) nonconforming space on grid \( T_k \), i.e., the space of all piecewise linear functions on \( T_k \) which are continuous at midpoints of all edges and vanish at midpoints of all boundary edges. The finite element problems arising from discretizing (2.1) read: Find \( u_k \in V_k \), such that

\[
a_k(u_k, v) = (f, v) \quad \forall v \in V_k,
\]

where \( a_k(u, v) := \sum_{K \in T_k} \int_K \nabla u \cdot \nabla v \, dx \) for all \( u, v \in V_k + H_0^1(\Omega) \). It is known from [22] that

\[
\| u - u_k \|_{L^2(\Omega)} + h_k |u - u_k|_{1,k} \leq C h_k^2 \| u \|_{H^2(\Omega)},
\]

where \( u \) and \( u_k \) are solutions of (2.1) and (2.2) respectively, and \( |v|_{1,k} := \sqrt{a_k(v, v)} \). Here and later, \( C \) stands for a generic constant independent of the multigrid index \( k \).

Since \( V_{k-1} \not\subset V_k \), we introduce auxiliary, \( P_1 \) conforming finite element spaces \( \{ \tilde{V}_k \} \) (cf. [10]). We note that \( \tilde{V}_k = V_k \cap H_0^1(\Omega) \). It holds that ([10])

\[
\inf_{v \in \tilde{V}_k \subset V_k} \{ \| u - v \|_{L^2(\Omega)} + h_k |u - v|_{1,k} \} \leq C h_k^2 \| u \|_{H^2(\Omega)} \quad \forall u \in H^2(\Omega) \cap H_0^1(\Omega).
\]
Noting that \( \tilde{V}_{k-1} \) is a subspace of both \( V_{k-1} \) and of \( V_k \) (in fact \( \tilde{V}_{k-1} = V_{k-1} \cap V_k \)), we define an intergrid transfer operator \( \tilde{I}_k : V_{k-1} \rightarrow \tilde{V}_{k-1} \subset V_k \): For any \( v \in V_{k-1} \),

\[
(\tilde{I}_k v)(a) = \frac{1}{j_a} \sum_{j=1}^{j_a} v|_{K_j}(a)
\]

and \( (\tilde{I}_k v)(a) = 0 \) if \( a \in \partial \Omega \), where \( a \) is a vertex in \( T_{k-1} \) and \( \{K_j\}_{j=1}^{j_a} \) are all the triangles in \( T_{k-1} \) which have \( a \) as a vertex. We will introduce a few more intergrid transfer operators in §9.

Let \( A_k : V_k \rightarrow V_k \) be defined by \( (A_k u, v) = a_k(u, v) \) \( \forall u, v \in V_k \), and

\[
|v|_{s,k} := \sqrt{(A_k^s v, v)} \quad \forall v \in V_k, 0 \leq s \leq 2.
\]

We note that \( |v|_{0,k} = ||v||_{L^2(\Omega)} \) and that \( |v|_{1,k} = \sqrt{a_k(v, v)} \). Let \( \Lambda_k \) be the spectral radius of \( A_k \). By a simple calculation (cf. [10]) on each triangle, we have the inverse estimate: \( \Lambda_k \leq Ch_k^{-2} \). We now define a multigrid scheme for solving (2.2) (cf. [3]).

**Definition 2.1** (one \( k \)th-level multigrid iteration).

1. For \( k = 1 \), the original problem (2.2) or the residual problem (2.8a) defined below is solved exactly.
2. For \( k > 1 \), \( w_{m+1} \) will be generated from \( w_0 \) by the following two steps. We do \( m \) smoothings:

\[
(w_l - w_{l-1}, v) = \Lambda_k^{-1}(f, v) - a_k(w_{l-1}, v) \quad \forall v \in V_k, \quad l = 1, 2, ..., m,
\]

and a coarse-level correction: \( w_{m+1} = w_m + \tilde{I}_k \epsilon \), where \( \epsilon \in V_{k-1} \) is obtained by doing \( p \) (> 1) \((W\)-cycle iterations) \((k - 1)\)st-level multigrid iterations starting with 0 guess for the following residual problem: Find \( \epsilon_a \in V_{k-1} \), such that

\[
a_{k-1}(\epsilon_a, v) = (f, \tilde{I}_k v) - a_k(w_m, \tilde{I}_k v) = (f_{\text{new}}, v) \quad \forall v \in V_{k-1}.
\]

In nested multigrid methods, the prolongation operator \( \tilde{I}_k \) in Definition 2.1 is simply the identity operator. Indeed, noting \( \tilde{V}_k \subset V_k \), we can define such a coarse-level correction: Find \( \bar{\epsilon} = \bar{\epsilon}_l \in \tilde{V}_{k-1} \) (or \( \bar{\epsilon} = \epsilon_c \in \tilde{V}_{k-1} \), resp.), such that

\[
a(\bar{\epsilon}, v) = (f, v) - a_k(w_m, v) = (f_{\text{new}}, v) \quad \forall v \in V \subset V_k,
\]

where in (2.8b) \( \bar{V} = \tilde{V}_{k-1} \) (in (2.8c) \( \bar{V} = \tilde{V}_k \), resp.) We note that the method with (2.8b) or (2.8c) is covered by a standard multigrid analysis (see the proof for Theorem 3.3 below or [3]). To use (2.8b) or (2.8c), one needs to implement both conforming and nonconforming elements. But for the purpose of comparison, we still include methods (2.8b,c) in our tests in §9.

3. **Convergence of P1 nonconforming multigrid methods**

The central task is to estimate the perturbation to the orthogonal-projection property caused by the intergrid transfer operator in the coarse-level correction (2.8a). The perturbation relates strongly to the stability of \( \tilde{I}_k \), which will be studied in the next lemma.

**Lemma 3.1.** It holds that

\[
|\tilde{I}_k v|_{l,k} \leq C|v|_{l,k-1} \quad \forall v \in V_{k-1} \quad \text{for } l = 0, 1.
\]
Proof. The proof uses a straightforward triangle by triangle argument. Let $K \in T_{k-1}$ have vertices $\{a_1, a_2, a_3\}$, $\{K_{ij}\}_{j=1}^{j_i} \subset T_{k-1}$ having $a_i$ as a vertex, and let $v_{ij} = v\mid_{K_{ij}}(a_i)$.

\begin{equation}
\|\bar{I}_k v\|^2_{L^2(K)} \leq |K| \sum_{i=1}^{3} (\bar{I}_k v)^2(a_i) \leq C h_{k-1}^2 \sum_{i=1}^{3} \sum_{j=1}^{j_i} v_{ij}^2 j_i^2 \leq C\|v\|^2_{L^2(\bigcup_{i,j} K_{ij})}.
\end{equation}

Summing (3.2) over all triangles in $T_{k-1}$, (3.1) is proved for $l = 0$ where $C$ depends only on the smallest angle (the same on all levels) in $T_1$. Let $w_{ij}$ be the nodal value of $v$ at the midpoint of the edge shared by two triangles $K_{ij}$ and $K_{i,j+1}$ (we let $K_{i,j+1} = K_{ij}$). Here we list $\{K_{ij}\}$ such that $w_{11} = w_{3j}$ and $w_{1j} = w_{2j}$. As $\bar{I}_k$ is an averaging operator, we get a telescoping sum in the estimate below.

\begin{equation}
\|\nabla \bar{I}_k v\|^2_{L^2(\Omega)} \leq C h_{k-1}^2 \left( (\bar{I}_k v(a_2) - \bar{I}_k v(a_1))^2 + (\bar{I}_k v(a_3) - \bar{I}_k v(a_1))^2 \right)
\end{equation}

\begin{equation}
= C h_{k-1}^2 \left( \sum_{j=1}^{j_1-1} \frac{j}{2} (v_{2j} - w_{2j} + w_{2,j+1}) + (v_{2j} - w_{2j}) \right.
\end{equation}

\begin{equation}
+ w_{1j} - w_{1j+1}) - \sum_{j=1}^{j_1-1} \frac{j}{2} (v_{1j} - w_{1j} + w_{1,j+1}) \right)^2
\end{equation}

\begin{equation}
+ C h_{k-1}^2 \left( \sum_{j=1}^{j_3-1} \frac{j}{3} (v_{3j} - w_{3j} + w_{3,j+1}) + (v_{3j} - w_{3j}) \right. \right.
\end{equation}

\begin{equation}
+ w_{11} - w_{11}) - \sum_{j=1}^{j_3-1} \frac{j}{3} (v_{1j} - w_{1j} + w_{1,j+1}) \right)^2
\end{equation}

\begin{equation}
\leq C\|\nabla v\|^2_{L^2(\bigcup_{i,j} K_{ij})}.
\end{equation}

Summing over all triangles, (3.1) is proved for $l = 1$.

To analyze the coarse-level corrections (2.8a–c), we introduce the following operators:

\begin{equation}
P_a : V_k \rightarrow V_{k-1}, \quad a_{k-1}(P_a v, w) = a_k(v, \bar{I}_k w) \quad \forall w \in V_{k-1},
\end{equation}

\begin{equation}
P_b : D(P_b) \rightarrow \bar{V}_{k-1}, \quad a(P_b v, w) = a_k(v, w) \quad \forall w \in \bar{V}_{k-1},
\end{equation}

where $D(P_b) = V_k + V_{k-1} + H^1_0(\Omega)$ in (3.3b). We note that $P_b$ is an $a_k(\cdot, \cdot)$ orthogonal-projection operator. In particular, by (2.4) $|u - P_b u|_{1,k-1} \leq C h_{k-1} \|u\|_{H^2(\Omega)}$.

Lemma 3.2. For any $v \in V_l$, $l = k - 1$ or $l = k$, it holds that

\begin{equation}
|v - P_b v|_{0,l} \leq C h_l |v - P_b v|_{1,l}.
\end{equation}

Proof. We use a standard duality argument. Let $u$ be the solution of (2.1) with $f = v - P_b v$. Let $u_l$ be the corresponding finite element solution of (2.2). Noting $P_b v \in \bar{V}_{k-1} \subset V_l$, the lemma is proved by the following estimate:

\begin{equation}
|v - P_b v|_{0,l}^2 = a_l(u_l, v - P_b v) = a_l(u_l - P_b u_l, v - P_b v)
\end{equation}

\begin{equation}
\leq |u_l - P_b u_l|_{1,l} \|v - P_b v\|_{1,l} \leq (|u_l - u_{1,l}| + |u - P_b u_{1,l}|) \|v - P_b v\|_{1,l}
\end{equation}

\begin{equation}
\leq C h_l \|u\|_{H^2(\Omega)} \|v - P_b v\|_{1,l} \leq C h_l |v - P_b v|_{0,l} |v - P_b v|_{1,l},
\end{equation}

\begin{equation}
\leq C h_l |v - P_b v|_{1,l},
\end{equation}
where (2.3)–(2.4) and \( \|u\|_{H^2} \leq C\|f\|_{L^2} = C|v - P_b v|_{0,1} \) (since \( \Omega \) is convex, cf. [11]) are applied.

\[ \text{Theorem 3.3 (The two-level method). Let } \epsilon = \epsilon_a \text{ in Definition 2.1. Given any } 0 < \gamma < 1, \text{ there exists an integer } m \text{ independent of } k \text{ such that} \]

\[ |u_k - w_{m+1}|_{1,k} \leq \gamma |u_k - w_0|_{1,k}, \]

where \( u_k \) is defined in (2.2) and \( w_l \) in Definition 2.1.

\[ \text{Proof. Let the errors be denoted by } e_l = u_k - w_l \text{ for } 0 \leq l \leq m + 1. \text{ We need to show } |e_{m+1}|_{1,k} \leq \gamma |e_0|_{1,k}. \text{ By Definition 2.1 and (3.3a–b), we have } \epsilon_a = P_a \epsilon_m \text{ and consequently we get that} \]

\[ |e_{m+1}|_{1,k} = |e_m - I_k \epsilon_a|_{1,k} = |e_m - I_k P_a \epsilon_m|_{1,k} \]

\[ \leq |e_m - P_b \epsilon_m|_{1,k} + |P_b \epsilon_m - I_k P_a \epsilon_m|_{1,k}. \]

The first term in (3.5) can be estimated by Lemma 3.2 as follows, noting \( P_b \) is an \( a_k(\cdot, \cdot) \) orthogonal-projection operator.

\[ |e_m - P_b \epsilon_m|_{1,k}^2 \leq |e_m - P_b \epsilon_m|_{0,k} |e_m|_{2,k} \leq Ch_k |e_m - P_b \epsilon_m|_{1,k} |e_m|_{2,k}. \]

For the second term in (3.5), we derive by Lemma 3.1 that

\[ |P_b \epsilon_m - I_k P_a \epsilon_m|_{1,k} = |I_k (P_b \epsilon_m - P_a \epsilon_m)|_{1,k} \leq C |P_b \epsilon_m - P_a \epsilon_m|_{1,k-1}. \]

Noting that \( I_k w = w \) for all \( w \in V_{k-1} \) and that

\[ a(P_b P_a \epsilon_m, w) = a_{k-1}(P_a \epsilon_m, w) = a_k(e_m, I_k w) = a_k(e_m, \epsilon_m) = a(P_b \epsilon_m, w), \]

we have \( P_b P_a \epsilon_m = P_b \epsilon_m \). Therefore, we can apply Lemmas 3.1 and 3.2 to estimate the right-hand side of (3.7) as follows. For any \( v \in V_{k-1} \),

\[ a_{k-1}(P_b P_a \epsilon_m - P_b \epsilon_m, v) = a_{k-1}(P_a \epsilon_m, P_b v - v) \]

\[ = a_k(e_m, I_k (P_b v - v)) \leq |e_m|_{2,k} |I_k (P_b v - v)|_{0,k} \]

\[ \leq |e_m|_{2,k} |P_b v - v|_{0,k-1} \leq Ch_k |e_m|_{2,k} |P_b v - v|_{1,k-1} \]

\[ \leq Ch_k |e_m|_{2,k} |v|_{1,k-1}. \]

Therefore, the second term in (3.5) can be estimated by (3.7) and (3.8) as

\[ |P_b \epsilon_m - I_k P_a \epsilon_m|_{1,k} \leq C |P_b P_a \epsilon_m - P_a \epsilon_m|_{1,k-1} \leq Ch_k |e_m|_{2,k}. \]

For the smoothing iteration (2.7) we have (cf. [3] for example)

\[ |e_m|_{2,k} \leq C m^{-1/2} A_k^{-1/2} |e_0|_{1,k} \leq C m^{-1/2} h_k^{-1} |e_0|_{1,k}. \]

Combining (3.5)–(3.6) and (3.9)–(3.10), (3.4) is proved by letting \( m > (C/\gamma)^2 \):

\[ |e_{m+1}| \leq C m^{-1/2} |e_0|_{1,k} \leq \gamma |e_0|_{1,k}. \]

By a standard argument in [3] and Theorem 3.3, one can show the convergence for \( W \)-cycle iterations and the theorem of optimal-order computation for the multigrid method.
4. The lowest order Raviart-Thomas elements

We consider first to solve (2.1) by the triangular Raviart-Thomas elements of the lowest order (cf. [19]):

\[ \text{RT}_0(T_k) := \{ \tau : \tau|_K = (a + bx, c + by) \ \forall K \in T_k, \text{ with normal components continuous on edges of elements in } T_k \}, \]

\[ M_0^{-1}(T_k) := \{ v : v \text{ is piecewise constant on elements in } T_k \}. \]

We omit the mixed formulation and the mixed element equation (cf. [1]). We will study the multigrid method for the equivalent (established in [1]) nonconforming finite element problems: Find \( \phi_k \in U_k \) such that

\[ \sum_{K \in T_k} \int_K P_0^0(\text{RT}_0(T_k)) \nabla \phi_k \cdot \nabla v \ dx = \int_{\Omega} P_0^0 M_0^{-1}(T_k) f \cdot v \ dx \quad \forall v \in U_k, \tag{4.1} \]

where \( P_0^0 \) denotes the \( L^2 \) orthogonal-projection onto \( V \) and \( \text{RT}_0^{-1}(T_k) := \{ \tau : \tau|_K = (a + bx, c + by) \ \forall K \in T_k \}. \) In (4.1) the multilevel nonconforming spaces are

\[ U_k := V_k + B_k, \]

where \( V_k \) is the \( P_1 \) nonconforming space, and \( B_k \) is the space of piecewise cubic polynomials vanishing on all edges of \( T_k \). We remark that \( v = v^l + v^b \) is unique where \( v^l \in V_k \) and \( v^b \in B_k \). Because \( P_0^0(\text{RT}^{-1})_0(T_k) \nabla u^l = \nabla u^l \) and because \( \int_K \nabla u^l \cdot \nabla v^b \ dx = 0 \), (4.1) can be decoupled as

\[ \sum_{K \in T_k} \int_K \nabla \phi_k^l \cdot \nabla v \ dx = \int_{\Omega} P_0^0 M_0^{-1}(T_k) f \cdot v \ dx \quad \forall v \in V_k, \]

\[ \sum_{K \in T_k} \int_K P_0^0(\text{RT}^{-1})_0(T_k) \nabla \phi_k^b \cdot \nabla v \ dx = \int_{\Omega} P_0^0 M_0^{-1}(T_k) f \cdot v \ dx \quad \forall v \in B_k. \]

The first equation is exactly the \( P_1 \) nonconforming equation (2.2), which can be solved by the multigrid method defined in §2. The second equation has a diagonal coefficient matrix. The equation is solved trivially. Because the two equations are uncoupled, and because we can recover the Raviart-Thomas solution from \( \phi_k \) with a linear order of computation (see [1]), we get the following theorem by Theorem 3.3 (cf. [3]).

**Theorem 4.1** (Optimal order MG in solving R-T equations). The algorithm described in Definition 2.1 and above can produce Raviart-Thomas mixed element solutions up to the order of truncation error by order \( \dim(U_k) \) (optimal order) computations.

However, for a general-coefficient problem \( \text{div}(-A(x)\nabla u) = f \), the above uncoupling of the linear and cubic elements in the linear system is not possible. We will consider next (cf. [1]) to solve

\[ \sum_{K \in T_k} \int_K P_k^k(A \nabla \phi_k) \cdot \nabla v \ dx = \int_{\Omega} P_0^k M_0^{-1}(T_k) f v \ dx \quad \forall v \in U_k, \tag{4.2} \]
where $A(x) \in W^{1,\infty}(\Omega)^{2\times 2}$ is a symmetric matrix, uniformly positive definite, and $P^k_{RT} : L^2(\Omega) \to RT^k(\mathcal{T}_k)$ is a weighted $L^2$ orthogonal-projector:

$$
\int_{\Omega} A^{-1}(P^k_{RT}\sigma) \cdot \tau \, dx = \int_{\Omega} A^{-1}\sigma \cdot \tau \, dx \quad \forall \tau \in RT^k(\mathcal{T}_k).
$$

In addition to the discontinuity, bubble functions also cause nonnestedness ($B_{k-1} \not\subset B_k$). In transferring bubble functions, different from [8] and [25], we simply discard those bubble functions by defining $I_k : U_{k-1} \to U_k$ as follows:

$$(4.3) \quad I_k u := I_k(u^l + u^b) = \tilde{I}_k u^l + I_k u^b = \tilde{I}_k u^l \quad \forall u = u^l + u^b \in U_{k-1},$$

where $\tilde{I}_k$ is defined in (2.5). We now define the bilinear forms for (4.2) and the discrete norms

$$
a^b_k(u,v) := \sum_{k \in \mathcal{T}_k} \int_K P^k_{RT}(A\nabla u) \cdot \nabla v \, dx \quad \forall u,v \in U_k,
$$

$$
|u|_{s,k,b} := \sqrt{(A^* u,u)} \quad \forall u \in U_k, \quad 0 \leq s \leq 2,
$$

where $(A_k u,v) := a^b_k(u,v)$. Let $\Lambda_{k,b}$ be the spectral radius of $A_k u$.

**Definition 4.2** (A multigrid method for (4.2)). The same as Definition 2.1 except (2.7) and (2.8a) are replaced by (4.4) and (4.5a) respectively.

$$(4.4) \quad (w_l - w_{l-1}, v) = \Lambda_{k,b}^{-1}((P^0_{M_{l+1}}(\mathcal{T}_k) f, v) - a^b_k(w_{l-1}, v)) \quad \forall v \in U_k,$n

$$(4.5a) \quad a^b_{k-1}(\epsilon, v) = (P f, I_k v) - a^b_k(w_m, I_k v) =: (f_{\new}, v) \quad \forall v \in U_{k-1},$$

where in (4.5a) $P$ is either $P^0_{M_{l+1}}(\mathcal{T}_k)$ if the $k$-th level is the highest level, or just the identity operator, i.e., depending on if the $k$-th level problem at the iteration is a finite element problem, or a coarse-level residual problem.

Noting $\tilde{V}_{k-1} \subset \tilde{V}_k \subset \tilde{V}_k \subset U_k$, the coarse-level problem could be defined by

$$
(4.5b,c) \quad a^b_k(\epsilon, v) = (P f, v) - a^b_k(w_m, v) =: (f_{\new}, v) \quad \forall v \in \tilde{V},
$$

where $\tilde{V}$ can be chosen either as $\tilde{V}_{k-1}$ (4.5b) or $\tilde{V}_k$ (4.5c). Because of the same reason mentioned after (2.8b,c), one has to program both conforming and nonconforming elements (two sets of data structures) if one uses (4.5b) or (4.5c) instead of (4.5a) in practice.

**5. Multigrid convergence for Raviart-Thomas elements**

The existence and uniqueness of a solution to (4.2) can be obtained from the formulation of Raviart-Thomas mixed method with Lagrange multipliers (cf [1]), or from the following lemma on the coercivity of $a^b_k(\cdot, \cdot)$.

**Lemma 5.1** (Lemma 2.1 in [8]). There exists a $C > 0$ such that

$$
|u|_{1,k,b} \leq C|u|_{1,k} \quad \forall u \in H^1_0(\Omega) \cup U_k,
$$

$$
|u|_{1,k,b} \geq C^{-1}|u|_{1,k} \quad \forall u \in U_k.
$$

**Lemma 5.2.** It holds that

$$
|I_k v|_{j,k,b} \leq C|v|_{j,k-1,b} \quad \forall v \in U_k, \quad \text{for } j = 0, 1.
$$
Proof. By a triangle wise calculation, we have (cf. Lemma 2.3 in [8]),
(5.1)\[ C(\|u^1\|_{j,k} + \|u^2\|_{j,k}) \leq \|u^1 + u^2\|_{j,k} \quad \forall u = u^1 + u^2 \in U_k, \quad \text{for } j, 0, 1. \]
In fact $\|u^1\|_{j,k} + \|u^2\|_{j,k} = \|u^1 + u^2\|_{j,k}$ (the case that $A(x)$ is the identity matrix). We now apply Lemma 5.1, Lemma 3.1 and the strengthened Cauchy inequalities (5.1) to get
\[ |I_k v|_{j,k,b} \leq C|I_k v|_{j,k} = C\|I_k v^1\|_{j,k} \leq C\|v^1\|_{j,k-1} \leq C\|v\|_{j,k-1,b}. \]

Similar to (3.3a–b), we define the following operators:
\begin{align}
(5.2a) & \quad P^b_j: U_k \to U_{k-1}, \quad a^b_{k-1}(P^b_j v, w) = a^b_k(v, I_k w) \quad \forall w \in U_{k-1}, \\
(5.2b) & \quad P^b_j: U_k \to \tilde{V}_{k-1}, \quad a^b_{k-1}(P^b_j v, w) = a^b_k(v, w) \quad \forall w \in \tilde{V}_{k-1}, \\
(5.2c) & \quad \tilde{P}^b_j: U_j \to \tilde{V}_{k-1}, \quad a^b_j(\tilde{P}^b_j v, w) = a^b_j(v, w) \quad \forall w \in \tilde{V}_{k-1},
\end{align}
where in (5.2c) $j = k - 1$ or $j = k$. We note that, unlike the $P_j$ in (3.3b), $P^b_j$, $P^b_{k-1}$ and $\tilde{P}^b_j$ are different operators caused by $P^b_{H^2}$.

Lemma 5.3. It holds that, for $j = k - 1$ and $j = k$,
(5.3)\[ |v - P^b_j v|_{0,j,b} \leq C h_j |v - P^b_j v|_{1,j,b} \quad \forall v \in U_j. \]

Proof. The proof is similar to that for Lemma 3.2. But here we have variational crimes (first Strang lemma) caused by $P^b_{H^2}$ in the bilinear forms. Let $\phi_j = v - P^b_j v$. Note that $\phi_j$ is $a^b_j(\cdot, \cdot)$-orthogonal in $\tilde{V}_{k-1}$. We define three solutions using $(v - P^b_j v)$ as the right-hand side function in the following three equations:
\begin{align}
& a^b_j(\tilde{u}_j, w) = (v - P^b_j v, w) \quad \forall w \in \tilde{V}_{k-1}, \\
& a^b_j(u_j, w) = (v - P^b_j v, w) \quad \forall w \in U_j, \\
& a^b_j(u, w) = (v - P^b_j v, w) \quad \forall w \in H^1_0(\Omega),
\end{align}
where $\tilde{u}_j \in \tilde{V}_{k-1}$, $u_j \in U_j$, $u \in H^2 \cap H^1_0$, and $a^b_j(u, w) := \int_\Omega A \nabla u \cdot \nabla w$. We note that both finite element solutions $\tilde{u}_j$ and $u_j$ approximate $u$ in the order of $O(h_j)$ in the norm $\|\text{grad}_j(\cdot)\|_{L^2(\Omega)}$, where $\text{grad}_j$ denotes the elementwise gradient operator (cf. for example, Proposition 2.1 in [8]). We now apply the standard duality analysis.
\begin{align}
|v - P^b_j v|_{0,j,b}^2 & = a^b_j(u_j, v - P^b_j v) = a^b_j(u_j - \tilde{u}_j, v - P^b_j v) \\
& \leq |u_j - \tilde{u}_j|_{1,j,b} \|v - P^b_j v|_{1,j,b} \leq C\|\text{grad}_j(u_j - \tilde{u}_j)\|_{L^2(\Omega)} \|v - P^b_j v|_{1,j,b} \\
& \leq C \left(\|\text{grad}_j(u_j - u)\|_{L^2} + |u - \tilde{u}_j|_{H^1} \right) \|v - P^b_j v|_{1,j,b} \\
& \leq C h_j \|u\|_{H^2} \|v - P^b_j v|_{1,j,b} \leq C h_j |v - P^b_j v|_{1,j,b} \quad \forall v \in U_j. \]
\end{align}

Theorem 5.4 (The two-level method for R-T equations). Let $\epsilon = \epsilon_\alpha$ in Definition 4.2. Given any $0 < \gamma < 1$, there exists an integer $m$ independent of the level number $k$ such that
\[ |\phi_k - w_m|_{1,k,b} \leq \gamma |\phi_k - w_0|_{1,k,b}. \]

Proof. By Definition 4.2 and the triangle inequality (see Theorem 3.3 for notations), we have
(5.5)\[ |\epsilon_{m+1}|_{1,k,b} \leq |\epsilon_m - P^b_k \epsilon_m|_{1,k,b} + |P^b_k \epsilon_m - P^b_k \epsilon_{m}|_{1,k,b} + |P^b_k \epsilon_m - I_k P^b_k \epsilon_m|_{1,k,b}. \]
The first and the third term in (5.5) can be analyzed as in Theorem 3.3. To estimate the second term in (5.5), we define \( f_k \in U_k \) and \( u \in H^1_0(\Omega) \cap H^2(\Omega) \) by
\[
(f_k, v) = a_k(e_m, v) \quad \forall v \in U_k
\]
and
\[
-\nabla(A(x)\nabla u) = f_k \quad \text{weakly.}
\]
We have \( \|u\|_{H^2(\Omega)} \leq C|f_k|_{0, k, b} \leq C|e_m|_{2, k, b} \) since \( \Omega \) is convex (cf. [11]). Comparing (5.6)–(5.7) and (5.2b,c), we realize that both \( P^b_k e_m \) and \( P^b_b e_m \) are finite element approximations (with variational crimes) of \( (5.6)–(5.7) \) and \( (5.2b,c) \), we obtain
\[
|P^b_k e_m - P^b_b e_m|_{1, k, b} \leq |P^b_k e_m - u|_{1, k, b} + |u - P^b_b e_m|_{1, k, b} \leq C h_k |e_m|_{2, k, b}.
\]
The rest of the argument in the proof of Theorem 3.3 remains the same.

By Theorem 5.4, Theorem 4.1 holds for the general coefficient problem (4.2).

6. Multigrid methods for Wilson's element

In this section, we consider multigrid methods for 2D Wilson elements on general quadrilateral meshes. This falls into the framework for \( P_1 \) nonconforming elements. It is straightforward to generalize the analysis here to the case of 3-D Wilson bricks on rectangular brick meshes, or to the case of the modified Wilson elements [14], or to the case of solving linear elasticity problems (cf. [13]).

Let \( K \) be a convex quadrilateral with vertices \( \{p_i = (x'_i, x''_i), 1 \leq i \leq 4\} \) and \( \hat{K} = [-1, 1] \times [-1, 1] \) the reference square with vertices \( \{\hat{p}_i\} \). There is a unique bilinear mapping such that
\[
F_K(\hat{K}) = K, \quad F_K(\hat{p}_i) = p_i, \quad 1 \leq i \leq 4.
\]
On the reference square \( \hat{K} \) the shape function has the following form:
\[
\hat{u}(\hat{x}_1, \hat{x}_2) = \frac{1}{4}\left\{(1 + \hat{x}_1)(1 + \hat{x}_2)u^1 + (1 - \hat{x}_1)(1 + \hat{x}_2)u^2 + (1 - \hat{x}_1)(1 - \hat{x}_2)u^3 + (1 + \hat{x}_1)(1 - \hat{x}_2)u^4\right\} + \frac{1}{8}\left\{(1 - \hat{x}^2_1)u^5 + (1 - \hat{x}^2_2)u^6\right\}.
\]
On \( K \) the shape function is given by \( u = \hat{u} \circ F_K^{-1} \). The last two terms in (6.2) can be called bubbles, which causes the discontinuity (and nonnestedness). A Wilson element function is determined by its values at the vertices \( p_i \) of \( K \) and the mean values of the second derivatives:
\[
u^5 = \int_K J_K^{-1} \frac{\partial^2 \hat{u}}{\partial x_1^2} \circ F_K^{-1} \, dx_1 \, dx_2, \quad u^6 = \int_K J_K^{-1} \frac{\partial^2 \hat{u}}{\partial x_2^2} \circ F_K^{-1} \, dx_1 \, dx_2,
\]
where \( J_K^{-1} \) is the determinant of the Jacobian for \( F_K^{-1} \).

We consider solving (2.1) by the Wilson element. Let \( \Omega \) be covered by a convex, quadrilateral grid \( T_1 \) initially. The multilevel grids on \( \Omega \) are defined by linking the midpoints on opposite edges of previous level quadrilaterals. The family of subdivisions \( \{T_k, \text{ size}(T_k) = h_k\} \) is regular and satisfies the following condition of Shi [20]:

**Condition (B).** There is a constant \( C \) depending only on the initial subdivision \( T_1 \), such that \( d_k \leq Ch_k^2 \) for level \( k \).
Here $d_k$ denotes the maximal distance between the two midpoints of two diagonals of quadrilaterals on level $k$. $d_kq = 0$ if $T_k$ contains only parallelograms. The Wilson spaces are denoted by $W_k := \{ u \mid u = 0 \text{ at all boundary vertices and } u|_K = \hat{u} \circ F_K^{-1} \forall K \in T_k \}$, where $F_K$ and $\hat{u}$ are defined in (6.1)–(6.2) respectively. The convergence of Wilson finite elements on quadrilateral grids is proved by Shi [21] under the Condition (B). The Wilson elements are nonconforming since they are not continuous on edges (but at vertices). We define $a_k(u,v)$ as in (2.2), but $K$ stands for a quadrilateral now. As $W_{k-1} \not\subset W_k$, again, we discard bubble functions when doing the intergrid transfer. We introduce auxiliary (conforming) spaces $\tilde{W}_k$, the bilinear elements on $T_k$ (cf. [10]). Each function $u \in W_{k-1}$ can be uniquely written as $w + z$ where $w \in \tilde{W}_{k-1}$ and $z$ is a bubble (last two terms in (6.2)). Noting $\tilde{W}_{k-1} \subset \tilde{W}_k \subset W_k$, the intergrid transfer operator $\tilde{I}_k$ is defined by

\begin{equation}
\tilde{I}_k : W_{k-1} \rightarrow W_k, \quad \tilde{I}_k u = \tilde{I}_k (w + z) = w \in \tilde{W}_{k-1} \subset W_{k-1} \cap W_k.
\end{equation}

The other definitions in §2 remain the same. We now prove Lemma 3.1 for Wilson elements.

**Lemma 6.1.** It holds that

\[ |\tilde{I}_k u|_{l,k} \leq C |u|_{l,k-1} \quad \forall u \in W_{k-1}, \quad l = 0, 1. \]

**Proof.** To prove the lemma, by the equivalence of norms, it suffices to prove the following two estimates on one quadrilateral $K \in T_{k-1}$:

\begin{align}
(6.4) \quad & \int_K \left| \nabla \tilde{I}_k u \right|^2 \, dx \leq C \int_K |\nabla u|^2 \, dx, \\
(6.5) \quad & \int_K \left| \tilde{I}_k u \right|^2 \, dx \leq C \int_K u^2 \, dx.
\end{align}

Let $u = w + z$ be decomposed as (6.3). Mapping $K$ to the reference element, (6.4) is shown as follows:

\[ C \int_K |\nabla w|^2 \, dx \leq \int_K |\nabla \hat{w}|^2 \, d\hat{x} \leq \int_K |\nabla \hat{w}|^2 \, d\hat{x} + \int_K |\nabla \hat{z}|^2 \, d\hat{x} \]

\[ = \int_K |\nabla (\hat{w} + \hat{z})|^2 \, d\hat{x} \leq C \int_K |\nabla (w + z)|^2 \, dx. \]

We note that the Wilson bubbles are $(\nabla \cdot, \nabla \cdot)$-orthogonal to bilinear functions on squares. For the general coefficients problem, (6.4) can be easily proven by a strengthened Cauchy inequality (see (6.6) below), as we shall do next in proving (6.5). Since

\begin{equation}
\left( \int_K \hat{w} \hat{z} \, d\hat{x} \right)^2 \leq \gamma_0 \int_K \hat{w}^2 \, d\hat{x} \int_K \hat{z}^2 \, d\hat{x}
\end{equation}

for some constant $0 \leq \gamma_0 < 1$, (6.5) follows the following argument:

\begin{align}
\int_K (\hat{w} + \hat{z})^2 \, d\hat{x} \geq & \left( 1 - \gamma_0 \right) \int_K \hat{w}^2 \left( \int_K \hat{w}^2 \right)^{1/2} - \left( \int_K \hat{z}^2 \right)^{1/2} \\
\geq & \left( 1 - \gamma_0 \right) \int_K \hat{w}^2. \quad \square
\end{align}
The remaining analysis in §3 stays the same. In particular we remark that the
conforming space \( \widetilde{W}_k \) has the same order approximation property as that of \( W_k \),
like the situation for the pair \((\widetilde{V}_k, \overline{V}_k)\). We conclude that the multigrid iteration
has a constant convergence rate for Wilson’s elements, and is of optimal order (see
[3] and Theorem 3.3 for the details).

7. Mini elements for stationary Stokes equations

For brevity, we will restrict our analysis to the mini elements for 2-D Stokes
equations only, although the work can be extended in several directions, for exam-
ple, higher space-dimensional problems and higher-degree “mini elements” (cf. [2]).
We consider the following 2D stationary Stokes equations on a convex polygonal
domain:

\[
\begin{aligned}
-\Delta \mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega, \\
\text{div } \mathbf{u} &= 0 \quad \text{in } \Omega, \\
\mathbf{u} &= 0 \quad \text{on } \partial \Omega.
\end{aligned}
\]

(7.1)

We follow the notations of Verfürth in [24], [25]. The variational problem for (7.1)
reads: Find \([\mathbf{u}, p] \in \mathcal{H} := (H^1_0(\Omega))^2 \times (L^2(\Omega)/\mathbb{R})^1\) such that

\[
\mathcal{L}([\mathbf{u}, p]; [v, q]) = (f, v) \quad \forall [v, q] \in \mathcal{H},
\]

(7.2)

where \((\cdot, \cdot)\) is the \(L^2\)-inner product, \(\mathcal{L}([\mathbf{u}, p]; [v, q]) := a(\mathbf{u}, v) + b(v, p) + b(\mathbf{u}, q),
\)
a(\(\mathbf{u}, v\)) := \int_{\Omega} \nabla \mathbf{u} : \nabla v \, dx, and \(b(\mathbf{u}, p) := -\int_{\Omega} \text{div} \mathbf{u} p \, dx\). Let \(\{T_k\}\) be as in
§2. Define \(\overline{V}_k\) and \(B_k\) as in §2 and §4, respectively. The multilevel mini (mixed)
element spaces are (cf. [2])

\[
\mathcal{H}_k := (\overline{V}_k \oplus B_k)^2 \times S^0_k,
\]

where \(S^0_k\) is the space of \(P_1\) conforming functions (\(\overline{V}_k\) functions without the boundary
condition) on \(T_k\) with mean-value zero. The mini elements are stable for Stokes
problems: there exists a constant \(C\) independent of \(k\) such that (cf. [2])

\[
\sup_{v \in (\overline{V}_k \oplus B_k)^2} \frac{b(v, p)}{\|v\|_{H^1(\Omega)}^2} \geq C \|p\|_{L^2} \quad \forall p \in S^0_k.
\]

(7.3)

Again the multilevel spaces are nonnested, \(\mathcal{H}_{k-1} \not\subset \mathcal{H}_k\), due to bubble functions.
As in §4, we discard bubbles when doing the intergrid transfer:

\[
I_k : \mathcal{H}_{k-1} \to \overline{V}^2_{k-1} \times S^0_{k-1} =: \mathcal{H}_{k-1,l} \subset \mathcal{H}_k,
\]

(7.4)

\[
I_k[v, q]_{k-1} = I_k[v_{k-1,l} \oplus v_{k-1,b}, q_{k-1}] = [v_{k-1,l}, q_{k-1}].
\]

In (7.4) and later, we use notations like

\[
[v, q]_k = [v_k, q_k] \quad \text{and} \quad I_k v_{k-1} = I_k(v_{k-1,l} \oplus v_{k-1,b}) = v_{k-1,l}.
\]

We define a mesh-dependent inner product as in [24]:

\[
([\mathbf{u}, p]; [v, q])_{0,k} := ([\mathbf{u}_k, v_k]) + h^2_k (p_k, q_k).
\]

(7.5)

Let \(L_k : \mathcal{H}_k \to \mathcal{H}_k\) (symmetric but not positive definite) be defined by

\[
(L_k([\mathbf{u}, p]; [v, q])_{0,k} = \mathcal{L}([\mathbf{u}, p]; [v, q]) \quad \forall [v, q]_k \in \mathcal{H}_k.
\]
We order its eigenvalues as $0 < |\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_{N_k}|$. We define a family of mesh-dependent norms:

\[(7.6) \quad \|\mathbf{u}, p\|_{k,s} := ((L_k L_k)^{s/2}[\mathbf{u}, p]_k; [\mathbf{u}, p]_k)_{0,k} \quad \forall [\mathbf{u}, p]_k \in \mathcal{H}_k, \quad -2 \leq s \leq 2.\]

We now define a multigrid scheme (cf. [24]) for solving the following discretized version of equation (7.2):

\[(7.7) \quad \mathcal{L}(\mathbf{u}, p)_k; [v, q]) = (f, v) \quad \forall [v, q] \in \mathcal{H}_k.\]

**Algorithm 7.1** (One $k$-th level MG iteration). One iteration will produce a solution $[\mathbf{u}, p]_k^{m+1} \in \mathcal{H}_k$ from a given initial guess $[\mathbf{u}, p]_k^0$ as follows.

1. **Smoothing.** This is done on the normal equation. For $l = 1, 2, \ldots, m$, $[\mathbf{u}, p]_k^l$ is defined by

   \[
   ([w, r]_k^l; [v, q])_{0,k} = (f, v) - \mathcal{L}([u, p]_k^{l-1}; [v, q]) \quad \forall [v, q] \in \mathcal{H}_k,
   \]

   \[
   ([u, p]_k^l - [u, p]_k^{l-1}; [v, q])_{0,k} = \lambda_{N_k}^{-1} \mathcal{L}([w, r]_k^l; [v, q]) \quad \forall [v, q] \in \mathcal{H}_k.
   \]

2. **Correction.** Let $[\mathbf{u}, p]_{k-1} \in \mathcal{H}_{k-1}$ solve

\[(7.9) \quad \mathcal{L}([\mathbf{u}, p]_{k-1}; [v, q]) = (f, \mathcal{I}_k v) - \mathcal{L}([\mathbf{u}, p]_k^m; \mathcal{I}_k [v, q]) \quad \forall [v, q] \in \mathcal{H}_{k-1}.
\]

The operator $\mathcal{I}_k$ in (7.9) and (7.10) below is defined in (7.4). If $k = 2$, let $[\mathbf{u}, p]_0^\sim = [\mathbf{u}, p]_1^1$. If $k \geq 2$, let $[\mathbf{u}, p]_k^{\sim} \in \mathcal{H}_{k-1}$ be the approximation of $[\mathbf{u}, p]_{k-1} \in \mathcal{H}_{k-1}$ obtained by doing $\mu$ ($k-1$)-st level multigrid iterations with starting guess zero for some $\mu \geq 2$. Then the new iterative solution after one $k$-th level iteration is defined by

\[(7.10) \quad [\mathbf{u}, p]_{k+1}^m := [\mathbf{u}, p]_k^m + \mathcal{I}_k [\mathbf{u}, p]_{k-1}^{\sim}.
\]

8. **Convergence analysis of multigrid mini elements**

Verf"urth has set up in [24] a framework for analyzing the multigrid mixed-element method defined in Algorithm 7.1. Our main work is to prove the stability of the intergrid transfer operator which throws away coarse-level bubbles. But this has been proved in Lemma 5.2. Here we only need to change a few minor points of the analysis of Verf"urth [24] because of the nonnestedness. Let $[\mathbf{u}_k^*, p_k^*] \in \mathcal{H}_k$ denote the solution of (7.7) and

\[(8.1) \quad [\mathbf{e}_k^l, \epsilon_k^l] := [\mathbf{u}_k^* - \mathbf{u}_k^l, p_k^* - p_k^l] \]

be the error of the $l$-th iterate defined in Algorithm 7.1, $0 \leq l \leq m + 1$. The following fine-level smoothing property is shown in [24]:

\[(8.2) \quad \|\mathbf{e}_k^m, \epsilon_k^m\|_2 \leq C h_k^{-2} m^{-1/2}\|\epsilon_k^0, \epsilon_k^0\|_0.
\]

The next lemma is used for analyzing the coarse-level correction (7.9)–(7.10).

**Lemma 8.1.** Let $[\mathbf{u}_k, p_k] \in \mathcal{H}_k$ be $L^2$ orthogonal to $\mathcal{H}_{k-1,l}$, i.e.

\[(\mathbf{u}_k, v) + h_k^2 (p_k, q) = 0 \quad \forall [v, q] \in \mathcal{H}_{k-1,l},\]

where $\mathcal{H}_{k-1,l}$ is defined in (7.4). Then it holds that

\[
\|\mathbf{u}_k, p_k\|_{L^2} \leq C h_k^2 \|\mathbf{u}_k, p_k\|_0.
\]

**Proof.** This lemma is almost identical to Lemma 4.2 in [24] except the space $\mathcal{H}_{k-1,l}$ here is replaced by $\mathcal{H}_{k-1}$. But the proof in [24] does not need the inf-sup condition (7.3) on the subspace. It only requires the approximation property. We note that the space $\mathcal{H}_{k-1,l}$ has the same order of approximation as that for $\mathcal{H}_{k-1}$ and that $\mathcal{H}_{k-1,l} \subset \mathcal{H}_k$. Therefore the proof in [24] remains the same. \(\square\)
Theorem 8.2 (Convergence of the two-level mini-element multigrid method). For any $0 < \gamma < 1$ there is a number $m$ independent of $k$, such that
\[
\| [e_k^{m+1}, e_k^m] \|_0 \leq \gamma \| [e_k^0, e_k^0] \|_0,
\]
where $[e_k^{m+1}, e_k^m]$ are defined in (8.1).

Proof. We will apply the framework of Verfürth [24]. Let $[w_{k-1}, r_{k-1}]$ be the $L^2$ orthogonal projection of the iterative error $[e_k^m, e_k^m]$ in $H_{k-1,t}$. Let $[u_k^*, p_k^*]$ be the exact solution of the residual problem (7.9). Since $H_{k-1,t} \subset H_{k-1}$, it follows by a duality argument that
\[
\| [e_k^m, e_k^m] - [u_k^*, p_k^*] \|_0 \leq C h_k \inf_{[v,q] \in H_{k-1}} \{ \| e_k^m - v \|_1 + \| e_k^m - q \|_0 \}
\leq C h_k \{ \| e_k^m - w_{k-1} \|_1 + \| e_k^m - r_{k-1} \|_1 \}
\leq C \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0,
\]
where in the last step we used an inverse inequality on P1 elements. Because of the nonnestedness of spaces, the $(m + 1)$-st iterative error is actually
\[
\| [e_k^{m+1}, e_k^m] \|_0 = \| [e_k^m, e_k^m] - I_k [u_{k-1}, p_{k-1}] \|_0
\leq \| [e_k^m, e_k^m] - I_k [w_{k-1}, r_{k-1}] \|_0 + \| I_k [w_{k-1}, r_{k-1}] - I_k [u_{k-1}, p_{k-1}] \|_0
= \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0 + \| I_k ([w_{k-1}, r_{k-1}] - [u_{k-1}, p_{k-1}]) \|_0
\leq \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0 + C \| [w_{k-1}, r_{k-1}] - [u_{k-1}, p_{k-1}] \|_0,
\]
where Lemma 5.2 is applied in the last step. Combining the above two estimates we obtain
\[
\| [e_k^{m+1}, e_k^m] \|_0 \leq C \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0.
\]
Applying (8.2) and Lemma 8.1, we can get now (cf. [24]) that
\[
\| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0 \leq C \| [e_k^m, e_k^m] \|_0 \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0 \leq C \sqrt{m} \| [e_k^m, e_k^m] \|_0 \| [e_k^m, e_k^m] - [w_{k-1}, r_{k-1}] \|_0.
\]
Therefore, the proof is completed by choosing $m$ large enough as
\[
\| [e_k^{m+1}, e_k^m] \|_0 \leq C \sqrt{m} \| [e_k^0, e_k^0] \|_0.
\]

9. Numerical tests

In our numerical tests, we solve (2.1) where
\[
f(x, y) = 2y(1 - y) + 2x(1 - x) \quad \text{and} \quad \Omega = (0, 1) \times (0, 1).
\]
$\Omega$ is uniformly triangulated as in Figure 1. The energy norm squares of the conforming finite element (CFE) and the nonconforming finite element (NCFE) solutions are listed in Table 1. We note that for CFE, $\| u - u_k \|^2_{H^1} = \| u \|^2_{H^1} - \| u_k \|^2_{H^1}$. The relative nodal errors on level 5 are about 0.5% for CFE and 0.15% for NCFE, but $\dim(V_k) \sim \dim(V_k)/3$. This is interesting. The solution in $V_k$ is better than that in $\tilde{V}_k$, but the cost for NCFE is not necessarily more than that of CFE. As a matter of fact, in our code, the NCFE consumes a little less computing time, partially because a node is shared by 6 triangles in CFE but only 2 in NCFE. We remark that the stiffness matrices are not generated in our code, and that the matrix-vector product is evaluated triangle by triangle.
Figure 1. Nodal errors before (A) and after doing a coarse-level correction by \( I_{k-1} \) (B), or by \( I_k \) (C), or by \( \tilde{I}_k \) (D), or by P1 CFE nested correction in \( \tilde{V}_k \) (E).

Table 1. Comparison of conforming and nonconforming finite elements.

<table>
<thead>
<tr>
<th>level</th>
<th>number of unknowns</th>
<th>( u_k )</th>
<th>( u_{k-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>CFE 81</td>
<td>NCFE 208</td>
<td>CFE 0.0207626450</td>
</tr>
<tr>
<td>5</td>
<td>289</td>
<td>800</td>
<td>0.0218489246</td>
</tr>
<tr>
<td>6</td>
<td>1,089</td>
<td>3,136</td>
<td>0.0221283623</td>
</tr>
<tr>
<td>7</td>
<td>4,225</td>
<td>12,416</td>
<td>0.0221987236</td>
</tr>
<tr>
<td>8</td>
<td>16,641</td>
<td>49,408</td>
<td>0.0222163455</td>
</tr>
<tr>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>0.0222222222 \ldots</td>
</tr>
</tbody>
</table>

We then test several “prolongation” (intergrid transfer) operators for NCFE. In Figure 1, the nodal errors before and after doing a coarse-level correction are plotted. The operator of Brenner [7] and Braess-Verfürth [4] \( I_{k-1} \) is used in
Figure 1B, while $\tilde{I}_k$ (defined in (2.5)) is used in Figure 1C. It seems the difference between the two is small. In Figure 1D, another averaging operator $\hat{I}_k$ is introduced:

$$
\hat{I}_k: V_{k-1} \to \tilde{V}_k \subset V_k; \quad \hat{I}_k v(n) = \begin{cases} 
\tilde{I}_k v(n) & \text{if } n \text{ is a vertex in } T_{k-1}, \\
v(n) & \text{if } n \text{ is a midpoint in } T_{k-1}.
\end{cases}
$$

We remark that the operator $\hat{I}_k$ is also used by Vassilevski and Wang in [23] recently.

The graph shows that $\hat{I}_k$ is better than both $I^k_{k-1}$ and $\tilde{I}_k$. If we use $P_1$ CFE as the coarse-level subspaces, then both higher- and lower-frequency components of the error are reduced as shown in Figure 1E.

The operation counts for various “prolongation” operators are listed below (note that $\dim(V_k) \approx 12 \dim(\tilde{V}_{k-1})$):

<table>
<thead>
<tr>
<th>Operators</th>
<th>$I_k$</th>
<th>$I^k_{k-1}$</th>
<th>$\tilde{I}_k$</th>
<th>$\hat{I}_k$</th>
<th>$I$ (2.8c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operations (times $\dim(V_k)$)</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Here we introduce a simpler prolongation operator $\bar{I}_k$: Let $c$ be the midpoint of an edge $\overline{ab}$ in $T_{k-1}$. $\bar{I}_k$ simply takes the value $v(c)$ as the values of $I_k v$ at the midpoints of $\overline{ac}$ and $\overline{cb}$. It seems that this operator transfers NCFE roughly. But the convergence rate of the multigrid method defined by this operator is not worse than that with $\tilde{I}_k$ (see Table 2).

In Table 2, we list the number of $V$-cycles needed for a multigrid method with $p=1$ and $m=8$ in Definition 2.1. By Figure 1D–E, the multigrid method seems to have a better convergence rate using coarse-level correction (2.8c) than using (2.8a) with $\tilde{I}_k$ replaced by $\hat{I}_k$. But it turns out that the two rates are about the same by the data in Table 2. In the last column of Table 2, we list the data for CFE. By Table 2, one may conclude that the NCFE multigrid with various perturbations in (2.8a) is almost as good as the CFE multigrid.

**Table 2. Numbers of $V$-cycle iterations.**

<table>
<thead>
<tr>
<th>Level</th>
<th>$I_k$</th>
<th>$I^k_{k-1}$</th>
<th>$\tilde{I}_k$</th>
<th>$\hat{I}_k$</th>
<th>$I$ (2.8c)</th>
<th>CFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

In Table 3, the spectral radii are presented for solving (4.2). The smoothing and correction operators are defined respectively by $S := (I - A_k^{-1} A_k)$ and $C := (I - I_k A_{k-1}^{-1} I_k^k A_k)$ (cf. (2.7–2.8)). Here, $I_k$ is either defined by (4.3) or the $I^k_{k-1}$ defined in [7] and [4]. In matrix form, the $A_k$ for (4.2) is a 2-block diagonal matrix. The matrix in the first block is the $A_k$ for $P_1$ NCFE (2.2). The second block is diagonal, generated by bubble functions. We scale bubbles such that each has a nodal value $\sqrt{60}/3$ at its barycenter to get basis functions. Therefore the spectrum of the multigrid operator associated with this lower block is very close to zero, no matter that the bubbles are thrown away (our method) or projected to fine-level bubbles ([8]).
Table 3. Spectral radii of 2-level multigrid operators for (4.2).

<table>
<thead>
<tr>
<th>grid (# unknowns)</th>
<th>interpolate $I_{k-1}$</th>
<th>interpolate $I_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho(S^C)$</td>
<td>$\rho(\text{SC})$</td>
</tr>
<tr>
<td>$4 \times 4$ (40)</td>
<td>0.5422</td>
<td>0.8516</td>
</tr>
<tr>
<td>$8 \times 8$ (176)</td>
<td>0.5708</td>
<td>1.0745</td>
</tr>
<tr>
<td>$12 \times 12$ (408)</td>
<td>0.6295</td>
<td>1.1957</td>
</tr>
<tr>
<td>$16 \times 16$ (736)</td>
<td>0.6548</td>
<td>1.2445</td>
</tr>
<tr>
<td>$20 \times 20$ (1160)</td>
<td>0.6672</td>
<td>1.2684</td>
</tr>
</tbody>
</table>

As we know, the orthogonal-projection would be lost in (2.8a) when the multilevel spaces are nonnested. The number of fine-level smoothings has to be sufficiently large in this case. This is verified in column 4 of Table 3. The V-cycle using Brenner’s method with one step of smoothing diverges according to the data in column 3 of Table 3. Surprisingly, one smoothing is enough to ensure the convergence of nonnested multigrids when our new intergrid transfer operator $\tilde{I}_k$ is in use. Columns 2 and 5 of Table 3 show that Brenner’s method is better (consistent with Table 2). But Columns 3 and 6 of Table 3 imply that our method is better.

It tells us that $\tilde{I}_k$ is more stable than $I_{k-1}$. In fact, it can be seen in the analysis of Oswald [17] that $|v|_1 < |I_{k-1}v|_1 \leq C|v|_1$.

We now present the convergence rate of a multigrid method for solving (9.1) via Wilson’s elements on uniform square grids. As we pointed out in the proof for Lemma 6.1, Wilson’s “bubbles” are $(\nabla \cdot, \nabla \cdot)$-orthogonal to bilinear functions. The stiffness matrix has 2 diagonal blocks where the block for bubbles is diagonal. The rate of convergence of multigrid using Wilson elements is the same as that for bilinear elements. We list here the spectral radii of 2-level multigrid methods with one step of presmoothing:

<table>
<thead>
<tr>
<th>grid (# unknowns)</th>
<th>spectral radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>$6 \times 6$</td>
<td>0.4302</td>
</tr>
<tr>
<td>$10 \times 10$</td>
<td>0.4751</td>
</tr>
<tr>
<td>$14 \times 14$</td>
<td>0.4874</td>
</tr>
<tr>
<td>$18 \times 18$</td>
<td>0.4924</td>
</tr>
<tr>
<td>$22 \times 22$</td>
<td>0.4949</td>
</tr>
<tr>
<td>$26 \times 26$</td>
<td>0.4963</td>
</tr>
<tr>
<td>$30 \times 30$</td>
<td>0.4973</td>
</tr>
</tbody>
</table>

Next we consider the multigrid method in §7 for mini elements. We replace the cubic bubbles in the standard mini element by macro linear-bubbles. This is to avoid using high-order numerical quadratures. Let $O$ be the barycenter of a triangle $T = \triangle ABC \in T_k$. The spaces of macro linear-bubble functions are defined by

$$\overline{B}_k := \text{span} \{ \phi \mid \phi \text{ is linear on } \triangle ABO, \triangle ACO \text{ and } \triangle BCO, \phi \text{ is continuous on } T \text{ and } \phi|_{\partial T} = 0 \forall T \in T_k \}.$$ 

Following [2], we can show that the stability condition (7.3) holds for the modified mini element:

$$\mathcal{P}_k := \left( \overline{V}_k \oplus \overline{B}_k \right)^2 \times S^0_k.$$ 

The multigrid analysis remains valid. We construct all matrices in the multigrid method (uniform grids on the unit square domain) and use MATLAB to find the spectral radii of iterative operators.

Columns 3 and 4 of Table 4 consist of the data of the multigrid method when the bubbles are discarded or averaged by the Verfürth method [25], respectively.
Table 4. Spectral radii of 2–level multigrid mini-elements.

<table>
<thead>
<tr>
<th>grid ( # unknowns)</th>
<th>smoothings</th>
<th>MG iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho(S^8)$</td>
<td>$\rho(CZ S^8)$</td>
</tr>
<tr>
<td>$4 \times 4 (106)$</td>
<td>0.99981616</td>
<td>0.8867</td>
</tr>
<tr>
<td>$6 \times 6 (242)$</td>
<td>0.99995724</td>
<td>0.9033</td>
</tr>
<tr>
<td>$8 \times 8 (434)$</td>
<td>0.99999154</td>
<td>0.9211</td>
</tr>
<tr>
<td>$10 \times 10 (682)$</td>
<td>0.99999688</td>
<td>0.9219</td>
</tr>
<tr>
<td>$12 \times 12 (986)$</td>
<td>0.99999870</td>
<td>0.9219</td>
</tr>
</tbody>
</table>

Table 5. Maximal errors at vertices and barycenters.

<table>
<thead>
<tr>
<th>grid ( # unknowns)</th>
<th>$L^{\infty}(\Omega)$</th>
<th>$\ell^{\infty}$-error at barycenters(c), nodes(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$</td>
<td>u_k,h</td>
</tr>
<tr>
<td>$2 \times 2 (26)$</td>
<td>0.1070</td>
<td>0.8865</td>
</tr>
<tr>
<td>$4 \times 4 (106)$</td>
<td>0.1027</td>
<td>0.5607</td>
</tr>
<tr>
<td>$8 \times 8 (434)$</td>
<td>0.0308</td>
<td>0.1535</td>
</tr>
<tr>
<td>$16 \times 16 (1762)$</td>
<td>0.0590</td>
<td>0.0533</td>
</tr>
<tr>
<td>$32 \times 32 (7106)$</td>
<td>0.0028</td>
<td>0.0101</td>
</tr>
</tbody>
</table>

Both methods have constant rates of convergence. We remark that the original algorithm of [25] contains a minor error. It is necessary to include the intergrid transfer operator into the coarse-level residual equation (7.9) (cf. [4] [6], [7], [26] and [27]). The idea of discarding bubble functions is based on the fact that bubble functions only provide the stability but not the approximation property. We next numerically check it. The test problem is the Stokes equations (7.1) on the unit square with $f = -\Delta \mathbf{curl} g$ where $g = 100x^2(1-x)^2y^2(1-y)^2$. The exact solution is $\mathbf{curl} g$. In column 2 of Table 5, the maximal norms of bubble functions are listed. The maximal nodal errors of the mini–element solutions are shown in columns 3–5. We can see that the convergence of the mini element is of the same order as that of the linear part of the mini element, both $O(h_k^2)$ in $L^{\infty}$-norm.

To conclude, we would remark again that the way to treat the discontinuity and bubbles in the multigrid method is not unique. Each method may have its advantages and disadvantages, depending on the specific implementation.

Dedication

We dedicate this paper to Professor James H. Bramble. An early version of this paper [28] was dedicated to Professor Bramble at the workshop on Mathematics of Computation in Partial Differential Equations held at Cornell University, January 25–27, 1991, in honor of his 60th birthday.

References