3. Lecture 3: Adaptivity II: General Operators and Extensions

In this lecture we discuss new ideas to prove convergence of AFEM for general operator (1.3) in sections 3.1-3.4, along with extensions in section 3.5. We use the notation

$$e_h := u - u_h, \qquad e_H := u - u_H, \qquad \varepsilon_H := u_h - u_H.$$

3.1. Quasi-Orthogonality.
If $\mathbf{b} \neq 0$ in (1.7), then the bilinear form $\mathcal{B}$ is no longer symmetric, and thus a scalar product. Therefore, the orthogonality relation (2.23) between $u - u_H$ and $u_h - u_H$, the so-called Pythagoras equality, fails to hold. We have instead a perturbation result referred to as quasi-orthogonality provided that the initial mesh is fine enough.

Lemma 3.1 (Quasi-orthogonality). *There exists a constant $C^* > 0$, solely depending on the shape regularity constant γ^* and a number $0 < s \leq 1$ dictated only by the interior angles of $\partial\Omega$, such that if the meshsize h_0 of the initial mesh satisfies $C^* h_0^s \|\mathbf{b}\|_{L^\infty} < 1$, then*

$$(3.1) \qquad \|u - u_h\|^2 \leq \Lambda_0 \|u - u_H\|^2 - \|u_h - u_H\|^2,$$

where $\Lambda_0 := (1 - C^* h_0^s \|\mathbf{b}\|_{L^\infty})^{-1}$. The equality holds provided $\mathbf{b} = 0$ in Ω.

Proof. In view of Galerkin orthogonality (1.13), namely $\mathcal{B}[e_h, v] = 0$, for all $v \in \mathbb{V}_h$, we have

$$\|e_H\|^2 = \|e_h\|^2 + \|\varepsilon_H\|^2 + \mathcal{B}[\varepsilon_H, e_h].$$

If $\mathbf{b} = 0$, then $\mathcal{B}$ is symmetric and $\mathcal{B}[\varepsilon_H, e_h] = \mathcal{B}[e_h, \varepsilon_H] = 0$. For $\mathbf{b} \neq 0$, instead, $\mathcal{B}[\varepsilon_H, e_h] \neq 0$, and we must account for this term. It is easy to see that div $\mathbf{b} = 0$ and integration by parts yield

$$\mathcal{B}[\varepsilon_H, e_h] = \mathcal{B}[e_h, \varepsilon_H] + \langle \mathbf{b} \cdot \nabla \varepsilon_H, e_h \rangle - \langle \mathbf{b} \cdot \nabla e_h, \varepsilon_H \rangle = 2 \langle \mathbf{b} \cdot \nabla \varepsilon_H, e_h \rangle.$$

Hence

$$\|e_h\|^2 = \|e_H\|^2 - \|\varepsilon_H\|^2 - 2 \langle \mathbf{b} \cdot \nabla \varepsilon_H, e_h \rangle.$$

Using Cauchy-Schwarz inequality and replacing the $H^1(\Omega)$-norm by the energy norm we have, for any $\delta > 0$ to be chosen later,

$$-2 \langle \mathbf{b} \cdot \nabla \varepsilon_H, e_h \rangle \leq \delta \|e_h\|_{L^2}^2 + \frac{\|\mathbf{b}\|_{L^\infty}^2}{\delta c_B} \|\varepsilon_H\|^2.$$

We then realize the need to relate $L^2(\Omega)$ and energy norms to replace $\|e_h\|_{L^2}$ by $\|e_h\|$. This requires a standard duality argument whose proof is reported in Ciarlet [15].

Lemma 3.2 (Duality). *Let $f \in L^2(\Omega)$ and $u \in H^{1+s}(\Omega)$ for some $0 < s \leq 1$ be the solution of (1.6), where s depends on the interior angles of $\partial\Omega$ ($s = 1$ if Ω is convex). Then, there exists a constant C_D, depending only on the shape regularity constant γ^* and the data of (1.3) such that*

$$(3.2) \qquad \|e_h\|_{L^2} \leq C_D h^s \|e_h\|_{\mathbb{V}}.$$

Inserting this estimate in the preceding two bounds, and using $h \leq h_0$, the meshsize of the initial mesh, in conjunction with $c_B \|e_h\|_{\mathbb{V}}^2 \leq \|e_h\|^2$ we deduce

$$\left(1 - \delta C_D^2 c_B^{-1} h_0^{2s}\right) \|e_h\|^2 \leq \|e_H\|^2 - \left(1 - \|\mathbf{b}\|_{L^\infty}^2 (\delta c_B)^{-1}\right) \|\varepsilon_H\|^2.$$

We now choose $\delta = \frac{\|\mathbf{b}\|_{L^\infty}}{C_D h_0^s}$ to equate both parenthesis, as well as h_0 sufficiently small for $\delta C_D^2 h_0^{2s} c_B^{-1} = C^* h_0^s \|\mathbf{b}\|_{L^\infty} < 1$ with $C^* := C_D/c_B$. We end up with

$$\|e_h\|^2 \leq \frac{1}{1 - C^* h_0^s \|\mathbf{b}\|_{L^\infty}} \|e_H\|^2 - \|\varepsilon_H\|^2.$$

This implies (3.1) and concludes the proof. $\square$

3.2. **Error and Oscillation Reduction.** In order to extend the theory of section 2.4 to the operator $\mathcal{L}$ of (1.3), we first notice that the local lower bound of Lemma 2.9 is still valid in this context. This is because its proof does not depend on the specific structure of the oscillation osc_H. However, osc_H cannot be related directly to osc_h as in Lemma 2.10, because osc_H and osc_h depend on u_H and u_h respectively. The next lemma accounts for this dependence.

Lemma 3.3 (Oscillation Reduction). *There exist constants $0 < \rho_1 < 1$ and $0 < \rho_2$, only depending on γ^* and θ_0, such that*

$$\mathrm{osc}_h(\Omega)^2 \leq \rho_1 \mathrm{osc}_H(\Omega)^2 + \rho_2 \|u_h - u_H\|^2. \tag{3.3}$$

Proof. The proof hinges on the Marking Strategy O and the Interior Node Property. We recall that if $T \in \mathcal{T}_h$ is contained in $T' \in \widehat{\mathcal{T}}_H$, then REFINE gives a reduction factor $\gamma_0 < 1$ of element size:

$$h_T \leq \gamma_0 H_{T'}. \tag{3.4}$$

The proof proceeds in three steps as follows.

1. *Relation between Oscillations.* We would like to relate $\mathrm{osc}_h(T')$ and $\mathrm{osc}_H(T')$ for any $T' \in \mathcal{T}_H$. To this end, we note that for all $T \in \mathcal{T}_h$ contained in T', we can write

$$R_T(u_h) = R_T(u_H) - \mathcal{L}_T(\varepsilon_H) \qquad \text{in } T,$$

where $\varepsilon_H = u_h - u_H$ as before and

$$\mathcal{L}_T(\varepsilon_H) := -\,\mathrm{div}\,(\mathbf{A}\nabla\varepsilon_H) + \mathbf{b} \cdot \nabla\varepsilon_H + c\,\varepsilon_H \qquad \text{in } T.$$

By Young's inequality, we have for all $\delta > 0$

$$\mathrm{osc}_h(T)^2 = h_T^2 \left\| R_T(u_h) - \overline{R_T(u_h)} \right\|_{L^2(T)}^2$$

$$\leq (1+\delta)h_T^2 \left\| R_T(u_H) - \overline{R_T(u_H)} \right\|_{L^2(T)}^2 + (1+\delta^{-1})h_T^2 \left\| \mathcal{L}_T(\varepsilon_H) - \overline{\mathcal{L}_T(\varepsilon_H)} \right\|_{L^2(T)}^2,$$

where $\overline{R_T(u_h)}$, $\overline{R_T(u_H)}$, and $\overline{\mathcal{L}_T(\varepsilon_H)}$ are L^2-projections of $R_T(u_h)$, $R_T(u_H)$, and $\mathcal{L}_T(\varepsilon_H)$ onto polynomials of degree $\leq n - 1$ on T. We next observe that

$$\left\| \mathcal{L}_T(\varepsilon_H) - \overline{\mathcal{L}_T(\varepsilon_H)} \right\|_{L^2(T)} \leq \|\mathcal{L}_T(\varepsilon_H)\|_{L^2(T)}$$

and that, according to (3.4),

$$h_T \leq \gamma_{T'} H_{T'}$$

provided $\gamma_{T'} = \gamma_0$ if $T' \in \widehat{\mathcal{T}}_H$ and $\gamma_{T'} = 1$ otherwise. Therefore, if $\mathcal{T}_h(T')$ denotes all $T \in \mathcal{T}_h$ contained in T',

$$\mathrm{osc}_h(T')^2 = \sum_{T \in \mathcal{T}_h(T')} \mathrm{osc}_h(T)^2$$

$$\leq (1 + \delta)\gamma_{T'}^2 \mathrm{osc}_H(T')^2 + (1 + \delta^{-1}) \sum_{T \in \mathcal{T}_h(T')} h_T^2 \|\mathcal{L}_T(\varepsilon_H)\|_{L^2(T)}^2, \tag{3.5}$$

since $R_T(u_H) = R_{T'}(u_H)$ and $\overline{R_T(u_H)}$ is the best L^2-approximation of $R_{T'}(u_H)$ in T.

2. *Estimate of $\mathcal{L}_T(\varepsilon_H)$.* In order to estimate $\|\mathcal{L}_T(\varepsilon_H)\|_{L^2(T)}$ in terms of $\|\varepsilon_H\|_{H^1(T)}$, we first split it as follows

$$\|\mathcal{L}_T(\varepsilon_H)\|_{L^2(T)} \leq \|\mathrm{div}(\mathbf{A}\nabla\varepsilon_H)\|_{L^2(T)} + \|\mathbf{b} \cdot \nabla\varepsilon_H\|_{L^2(T)} + \|c\,\varepsilon_H\|_{L^2(T)}$$

and denote these terms N_A, N_B, and N_C, respectively. Since

$$N_A \leq \|(\mathrm{div}\,\mathbf{A}) \cdot \nabla\varepsilon_H\|_{L^2(T)} + \|\mathbf{A} : H(\varepsilon_H)\|_{L^2(T)}$$

where $H(\varepsilon_H)$ is the Hessian of ε_H in T, invoking the Lipschitz continuity of $\mathbf{A}$ together with an inverse estimate in T, we infer that

$$N_A \leq C_A \left(\|\nabla\varepsilon_H\|_{L^2(T)} + h_T^{-1} \|\nabla\varepsilon_H\|_{L^2(T)} \right),$$

where C_A depends on $\mathbf{A}$ and the shape regularity constant γ^*. Besides, we readily have

$$N_B \leq C_B \|\nabla \varepsilon_H\|_{L^2(T)}, \qquad N_C \leq C_C \|\varepsilon_H\|_{L^2(T)},$$

where C_B, C_C depend on $\mathbf{b}$, c. Combining these estimates, we arrive at

(3.6) $$h_T^2 \|\mathcal{L}_T(\varepsilon_H)\|_{L^2(T)}^2 \leq C_* \|\varepsilon_H\|_{H^1(T)}^2.$$

3. *Choice of δ.* We insert (3.6) into (3.5) and add over $T' \in \mathcal{T}_H$. Recalling the definition of $\gamma_{T'}$ and utilizing (2.18), we deduce

$$\sum_{T' \in \mathcal{T}_H} \gamma_{T'}^2 \mathrm{osc}_H(T')^2 = \gamma_0^2 \sum_{T' \in \widehat{\mathcal{T}}_H} \mathrm{osc}_H(T')^2 + \sum_{T' \in \mathcal{T}_H \setminus \widehat{\mathcal{T}}_H} \mathrm{osc}_H(T')^2$$

$$= \mathrm{osc}_H(\Omega)^2 - (1 - \gamma_0^2) \sum_{T' \in \widehat{\mathcal{T}}_H} \mathrm{osc}_H(T')^2$$

$$\leq (1 - (1 - \gamma_0^2)\hat{\theta}^2))\mathrm{osc}_H(\Omega)^2,$$

where $theta_0$ is the user's parameter in (2.18). Moreover, since $C_* \|\varepsilon_H\|_{H^1}^2 \leq C_o \|\varepsilon_H\|^2$ with $C_o = C_* c_B^{-1}$ in light of the equivalence between energy and H^1 norms, we end up with

$$\mathrm{osc}_h(\Omega)^2 \leq (1 + \delta)(1 - (1 - \gamma_0^2)\hat{\theta}^2)\mathrm{osc}_H(\Omega)^2 + (1 + \delta^{-1})C_o \|\varepsilon_H\|^2.$$

We finally choose δ sufficiently small so that

$$\rho_1 = (1 + \delta)(1 - (1 - \gamma_0^2)\hat{\theta}^2) < 1, \qquad \rho_2 = (1 + \delta^{-1})C_o.$$

to complete the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

3.3. **Convergence for General Operators.** Since error and oscillation are now coupled, in order to prove convergence we need to handle them together. This leads to a novel argument and result, the contraction property (3.7) below, according to which both error and oscillation decrease together.

Theorem 3.4 (Convergence of AFEM)**.** *Let $\{u_k\}_{k \in \mathbb{N}_0}$ be a sequence of finite element solutions corresponding to a sequence of nested finite element spaces $\{\mathbb{V}^k\}_{k \in \mathbb{N}_0}$ produced by AFEM. There exist constants $\sigma, \gamma > 0$, and $0 < \xi < 1$, depending solely on the mesh regularity constant γ^*, data, parameters θ_E and θ_0, and a number $0 < s \leq 1$ dictated by interior angles of $\partial\Omega$, such that if the initial meshsize h_0 satisfies $h_0^s \|\mathbf{b}\|_{L^\infty} c_B^{-1} < \sigma$, then for any two consecutive iterates k and $k+1$, we have*

(3.7) $$\|u - u_{k+1}\|^2 + \gamma\,\mathrm{osc}_{k+1}(\Omega)^2 \leq \xi^2 \left(\|u - u_k\|^2 + \gamma\,\mathrm{osc}_k(\Omega)^2 \right).$$

Therefore AFEM converges with a linear rate ξ, namely,

$$\|u - u_k\|^2 + \gamma\,\mathrm{osc}_k(\Omega)^2 \leq C_0\,\xi^{2k},$$

where $C_0 := \|u - u_0\|^2 + \gamma\,\mathrm{osc}_0(\Omega)^2$.

Proof. We just prove the contraction property (3.7), which obviously implies the decay estimate. For convenience, we introduce the notation

$$e_k := \|u - u_k\|, \qquad \varepsilon_k := \|u_{k+1} - u_k\|, \qquad \mathrm{osc}_k := \mathrm{osc}_k(\Omega).$$

The idea is to use the quasi-orthogonality (3.1) and replace the term $\|u_{k+1} - u_k\|^2$ using new results of error and oscillation reduction estimates (2.21) and (3.3). We proceed in three steps as follows.

1. We first get a lower bound for ε_k in terms of e_k. To this end, we use Marking Strategy E and the upper bound (2.5) to write

$$\theta_E^2 e_k^2 \leq C_1 \theta_E^2 \eta_k(\Omega)^2 \leq C_1 \sum_{T \in \widehat{\mathcal{T}}_k} \eta_k(T)^2.$$

Adding (2.21) of Lemma 2.9 over all marked elements $T \in \widehat{\mathcal{T}_k}$, and observing that each element can be counted at most $D := d + 2$ times due to overlap of the sets ω_T, together with $\|v\|_{\mathbb{V}}^2 \le c_B^{-1} \|v\|^2$ for all $v \in H_0^1(\Omega)$, we arrive at

$$\theta_E^2 e_k^2 \le \frac{DC_1}{C_2 c_B}\, \varepsilon_k^2 + \frac{DC_1}{C_2}\, \mathrm{osc}_k^2.$$

If $\Lambda_1 := \frac{c_B C_2}{DC_1}\theta_E^2, \Lambda_2 := c_B$, then this implies the lower bound for ε_k^2,

(3.8)
$$\varepsilon_k^2 \ge \Lambda_1 e_k^2 - \Lambda_2 \mathrm{osc}_k^2.$$

2. If h_0 is sufficiently small so that the quasi-orthogonality (3.1) of Lemma 3.1 holds with $\Lambda_0 = (1 - C^* h_0^s \|\mathbf{b}\|_{L^\infty} c_B^{-1})^{-1}$, then

$$e_{k+1}^2 \le \Lambda_0 e_k^2 - \varepsilon_k^2.$$

Replacing the fraction $\beta\varepsilon_k^2$ of ε_k^2 via (3.8) we obtain

$$e_{k+1}^2 \le (\Lambda_0 - \beta\Lambda_1)e_k^2 + \beta\Lambda_2 \mathrm{osc}_k^2 - (1 - \beta)\varepsilon_k^2,$$

where $0 < \beta < 1$ is a constant to be chosen suitably. We now assert that it is possible to chose h_0 compatible with Lemma 3.1 and also that

$$0 < \alpha := \Lambda_0 - \beta\Lambda_1 < 1.$$

A simple calculation shows that this is the case provided

$$h_0^s \|\mathbf{b}\|_{L^\infty} c_B^{-1} < \frac{\beta\Lambda_1}{C^*(1 + \beta\Lambda_1)} < \frac{1}{C^*},$$

i.e., $h_0^s \|\mathbf{b}\|_{L^\infty} c_B^{-1} < \sigma$ with $\sigma := \frac{\beta\Lambda_1}{C^*(1+\beta\Lambda_1)}$. Consequently

(3.9)
$$e_{k+1}^2 \le \alpha e_k^2 + \beta\Lambda_2 \mathrm{osc}_k^2 - (1 - \beta)\varepsilon_k^2.$$

3. To remove the last term of (3.9) we resort to the oscillation reduction estimate of Lemma 3.3

$$\mathrm{osc}_{k+1}^2 \le \rho_1 \mathrm{osc}_k^2 + \rho_2 \varepsilon_k^2.$$

We multiply it by $(1 - \beta)/\rho_2$ and add it to (3.9) to deduce

$$e_{k+1}^2 + \frac{1 - \beta}{\rho_2}\mathrm{osc}_{k+1}^2 \le \alpha\, e_k^2 + \left(\beta\Lambda_2 + \frac{\rho_1}{\rho_2}(1 - \beta)\right)\mathrm{osc}_k^2.$$

If $\gamma := \frac{1-\beta}{\rho_2}$, then we would like to choose $\beta < 1$ in such a way that

$$\beta\Lambda_2 + \rho_1\gamma = \mu\gamma$$

for some $\mu < 1$. A simple calculation yields

$$\beta = \frac{\frac{\mu-\rho_1}{\rho_2}}{\Lambda_2 + \frac{\mu-\rho_1}{\rho_2}},$$

and shows that $\rho_1 < \mu < 1$ guarantees that $0 < \beta < 1$. Therefore,

$$e_{k+1}^2 + \gamma\, \mathrm{osc}_{k+1}^2 \le \alpha\, e_k^2 + \mu\gamma\, \mathrm{osc}_k^2,$$

and the asserted estimate (3.7) follows upon taking $\xi = \max(\alpha, \mu) < 1$. □

Remark 3.5 (Comparison with Theorem 2.13). In Thereom 2.13, and [34, 35], the oscillation is independent of discrete solutions, i.e. $\rho_2 = 0$, and is reduced by the factor $\rho_1 < 1$ in (3.3). Consequently, Step 3 above is avoided by setting $\beta = 1$ and the decay of e_k and osc_k is monitored separately. Since this is no longer possible, e_k and osc_k are now combined and decreased together.

Remark 3.6 (Splitting of ε_k). The idea of splitting ε_k is already used by Chen and Jia [13] in examining one time step for the heat equation. This is because a mass (zero order) term naturally occurs, which did not take place in [34, 35]. The elliptic operator is just the Laplacian in [13].

Remark 3.7 (Effect of Convection)**.** Assuming that $h_0^s \|\mathbf{b}\|_{L^\infty} c_B^{-1} < \sigma$ implies that the local Peclet number is sufficiently small for the Galerkin method not to exhibit oscillations. This appears to be essential for u_0 to contain relevant information and guide correctly the adaptive process. This restriction is difficult to verify in practice because it involves unknown constants.However, starting from coarser meshes than needed in theory does not seem to be a problem in practice (see simulations in §3.4.2).

Remark 3.8 (Vanishing Convection)**.** If $\mathbf{b} = 0$, then Theorem 3.4 has no restriction on the initial mesh. This thus extends the convergent result of Morin et al. [34, 35] to variable diffusion coefficient and zero order terms.

Remark 3.9 (Optimal β)**.** The choice of β can be optimized. In fact, we can easily see that

$$\alpha = \Lambda_0 - \beta\Lambda_1, \qquad \mu = \rho_1 + \frac{\beta}{1-\beta}\rho_2\Lambda_2$$

yields a unique value $0 < \beta_* < 1$ for which $\alpha = \mu$ and the contraction constant ξ of Theorem 3.4 is minimal. This β_* depends on geometric constant $\Lambda_0, \Lambda_1, \Lambda_2$ as well on θ_E, θ_0 and h_0, but it is not computable.

3.4. Numerical Experiments.

For convenience of presentation, we introduce the following notation:
- DOF := number of elements in a given mesh, which is comparable with number of degree of freedoms;
- EOC := $\frac{\log(e(k-1)/e(k))}{\log(\mathsf{DOF}(k)/\mathsf{DOF}(k-1))}$, experimental order of convergence, $e(k) := \|u - u_k\|_{H^1}$;
- EOC(η) := $\frac{\log(\eta_{k-1}/\eta_k)}{\log(\mathsf{DOF}(k)/\mathsf{DOF}(k-1))}$, experimental order of convergence of estimator, $\eta_k := \eta_k(\Omega)$;
- $Z_e := e(k)/e(k-1)$ and $Z_o := \mathrm{osc}_k/\mathrm{osc}_{k-1}$, reduction factors of error and oscillation;
- Eff := e_k/η_k, effectivity index, i.e. the ratio between the error and the estimator. This ratio does in fact give an estimate to constant C_1 for upper bound (2.5).
- M_E and M_O are the number of marked elements due to Marking Strategy E and the additional marked elements due to Marking Strategy O, respectively.

The experimental order of convergence (EOC) measures how the error (or estimator) decreases as the number of DOF increases. In fact we have $\mathsf{error}_k \approx C\,\mathsf{DOF}_k^{-\mathsf{EOC}}$.

3.4.1. *Example: Oscillatory Coefficients and Nonconvex Domain.* We consider the PDE (1.3) with Dirichlet boundary condition $u = g$ on the nonconvex L-shape domain $\Omega := (-1,1)^2 \setminus [0,1] \times [-1,0]$. We also take the exact solution

$$u(r) = r^{\frac{2}{3}} \sin(\frac{2}{3}\theta),$$

where $r^2 := x^2 + y^2$ and $\theta := \tan^{-1}(y/x) \in [0, 2\pi)$. We deal with variable coefficients $\mathbf{A}(x,y) = a(x,y)\mathbf{I}$, $\mathbf{b}(x,y) = \mathbf{0}$ and $c(x,y)$ defined by

$$(3.10) \qquad\qquad a(x,y) = \frac{1}{4 + P(\sin(\frac{2\pi x}{\epsilon}) + \sin(\frac{2\pi y}{\epsilon}))}\ ,$$

$$(3.11) \qquad\qquad c(x,y) = A_c(\cos^2(lx) + \cos^2(lx))\ ,$$

where P, ϵ, A_c, and l are parameters. The functions f in (1.3) and g are defined accordingly. The results are shown in Tables 3.1 and 3.2 and Figure 3.1. The observations and conclusions of this experiment are as follows:
- AFEM gives an optimal rate of convergence of order ≈ 0.5, while standard uniform refinement achieves the suboptimal rate of 0.3 as expected from theory.
- Both AFEM and FEM with uniform refinement perform with effectivity index Eff ≈ 2.0, which give the estimate of constant $C_1 \approx 0.5$ for upper bound (2.5); no weights have been used in (2.4). For AFEM, the reduction factors of error and oscillation are approximately 0.7 and 0.5 as DOF increases (Table 3.1). The oscillation thus decreases faster than the error and becomes insignificant asymptotically for k large. In addition, AFEM outperforms FEM in terms of CPU time vs energy error.
- Figure 3.1 depicts the effect of a corner singularity and rapid variation of diffusion coefficient $a(x,y)$ in mesh grading; c does not play much role.

k	DOF_k	$\|u - u_k\|$	EOC_e	RF_E	RF_O	Eff	M_E	M_O
-	24	2.181e-01	–	–	–	4.504	3	0
1	65	1.481e-01	0.388	0.679	0.446	2.994	10	0
2	229	1.056e-01	0.268	0.713	0.558	2.475	11	0
3	423	8.812e-02	0.295	0.834	0.652	2.222	13	0
4	651	5.083e-02	1.276	0.577	0.314	2.053	37	0
5	1156	3.305e-02	0.750	0.650	0.444	2.028	89	0
6	2299	2.206e-02	0.588	0.668	0.408	1.980	253	0
7	5148	1.445e-02	0.525	0.655	0.658	1.965	771	0
8	12678	7.991e-03	0.657	0.553	0.175	1.957	1833	0
9	29979	4.911e-03	0.566	0.615	0.426	2.032	-	-

TABLE 3.1. Example 3.4.1 (Oscillatory coefficients and nonconvex domain): The parameters of AFEM are $\theta_E = \theta_0 = 0.5$, and those controlling the oscillatory coefficients are $P = 1.8, \epsilon = 0.4, A_c = 4.0, l = 1.0$, as described in (3.10) and (3.11). The experimental order of convergence EOC_e is close to the optimal rate of 0.5, which indicates quasi-optimal meshes. The oscillation reduction factor RF_O is smaller than the error reduction factor RF_E, which confirms that oscillation decreases faster than error. The effectivity index Eff is approximately around 2.0. There are no additional marked elements from oscillation for this $\theta_E = 0.5$ i.e. $M_O = 0$. However, this is not the case if $\theta_E < 0.3$

DOF_k	$\|u - u_k\|$	EOC_e	RF_E	RF_O	Eff
384	1.005e-01	0.400	0.574	0.300	2.398
1536	4.809e-02	0.532	0.478	0.195	2.127
6144	2.597e-02	0.444	0.540	0.182	1.984
24576	1.551e-02	0.372	0.597	0.242	1.845
98304	9.585e-03	0.347	0.618	0.264	1.745

TABLE 3.2. Example 3.4.1 (Oscillatory coefficients and nonconvex domain): Standard uniform refinement is performed using the same values for parameters P, ϵ, A_c, and l as that of AFEM given in Table 3.1 above. EOC_e is now suboptimal and close to the expected value $1/3$ ($\approx s/2$ according to (1.26)). The effectivity index Eff is around 2 which is about the same as AFEM. We need about 10^5 DOFs to get the error around 10^{-2} whereas for AFEM we only need 10^4 DOFs.
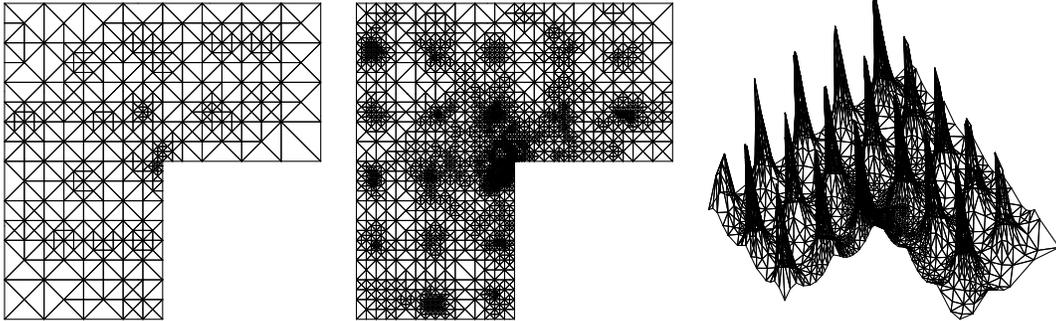

FIGURE 3.1. Example 3.4.1 (Oscillatory coefficients and nonconvex domain): Parameters of AFEM are $\theta_E = \theta_0 = 0.5$, those of oscillatory coefficients are $P = 1.8, \epsilon = 0.4, A_c = 1.0, l = 1.0$. The sequence of graded meshes after 4 and 7 iterations shows that mesh refinement is dictated by geometric (corner) singularities as well as periodic variations of the diffusion coefficient but not much from the zero order term. Also on the right, 3-D plot of diffusion coefficient $a(x, y)$ of (3.10) interpolated onto the mesh of iteration 7. This shows the combined effect of rapidly varying $a(x, y)$ and exact solution $u = r^{\frac{2}{3}} \sin(\frac{2}{3}\theta)$: meshes are refined more where $a(x, y)$ has large gradient.

• The number of additional marked elements M_O due to Marking strategy O depends on parameters θ_E and θ_0. For this example, $M_O = 0$ because the parameter θ_E is sufficiently big, hence the condition for Marking strategy O is automatically satisfied. Similar experiments for $\theta_E < 0.3$ and $\theta_0 = 0.5$ yield $M_O \neq 0$ and M_O becomes even dominant for $\theta_E = 0.1$; see Example 3.4.2 for more details.

3.4.2. *Example: Convection-Dominated Convection.* We consider the convection dominated-diffusion elliptic model problem (1.3) with Dirichlet boundary condition $u = g$ on convex domain $\Omega := (0,1)^2$, with isotropic diffusion coefficient $\mathbf{A} = \epsilon \mathbf{I}$, $\epsilon = 10^{-3}$, convection velocity $\mathbf{b} = (y, \frac{1}{2} - x)$ and $c = f = 0$; note that $\operatorname{div} \mathbf{b} = 0$. The Dirichlet boundary condition $g(x,y)$ on $\partial \Omega$, a pulse, is the continuous piecewise linear function given by

$$(3.12) \qquad g(x,y) = \begin{cases} 1 & \{.2 + \tau \leq x \leq .5 - \tau; \ y = 0\}, \\ 0 & \partial\Omega \setminus \{.2 \leq x \leq .5; \ y = 0\}, \\ \text{linear} & \{(.2 \leq x \leq .2 + \tau) \text{ or } (.5 - \tau \leq x \leq .5); y = 0\}, \end{cases}$$

where τ is a parameter. This problem models the transport of a pulse from $\partial\Omega$ inside Ω and back to $\partial\Omega$. Results are reported in Table 3.3 and Figures 3.2, 3.3 for parameters $\theta_E = 0.3, \theta_0 = 0.6, \tau = 0.005$, starting from a coarser mesh than what we would need in theory. To see weather oscillation plays any role in AFEM, Table 3.4 shows results of AFEM without using Marking Strategy O. Observations and conclusions follow:

$\mathsf{DOF_k}$	$\eta_k(\Omega)$	EOC_η	$\mathsf{RF_O}$	$\mathsf{M_E}$	$\mathsf{M_O}$
64	1.74e-1	-	-	2	5
147	9.48e-2	0.73	0.27	8	7
360	2.35e-2	1.55	0.33	4	9
500	1.68e-2	1.02	0.50	5	15
762	1.12e-2	0.95	0.43	10	23
1170	8.58e-3	0.62	0.52	15	70
2173	6.10e-3	0.55	0.48	22	137
3862	4.75e-3	0.43	0.48	30	298
7149	3.45e-3	0.51	0.50	80	600
13981	2.60e-3	0.42	0.51	-	-

TABLE 3.3. Example 3.4.2: AFEM with parameters $\theta_E = 0.3, \theta_0 = 0.6$, and $\tau = 0.005$. The optimal decay ≈ 0.5 of estimator $\eta(\Omega)$ is computational evidence of optimal meshes. The reduction factor of oscillation $\mathsf{RF_O} := \mathrm{osc}_k/\mathrm{osc}_{k-1}$ gives an estimate of constant $\rho_1 \approx 0.5$ in Lemma 3.3. In contrast to Experiment 1, the additional marking $\mathsf{M_O}$ due to oscillation dominates $\mathsf{M_E}$ from Marking Strategy E. This controls $\mathsf{RF_O}$, the decay of oscillations, which decrease together with the error according to Theorem 3.4.

$\mathsf{DOF_k}$	$\eta_k(\Omega)$	EOC_η	$\mathsf{RF_O}$
64	1.74e-1	-	-
95	1.02e-1	1.34	0.59
244	3.81e-2	1.31	0.86
414	1.75e-2	4.09	0.62
654	9.42e-3	1.18	0.70
834	9.05e-3	0.16	0.59
1577	5.43e-3	0.89	0.93
2970	3.56e-3	0.51	0.92
4250	2.84e-3	0.62	0.82
6502	2.15e-3	0.65	0.59
10209	1.66e-3	0.57	0.62

TABLE 3.4. Example 3.4.2: AFEM performance without Marking Strategy O, using the same parameters as for Table 3.3. The reduction factor of oscillation $\mathsf{RF_O}$ is not as stable as our AFEM shown in Table 3.3. The estimator still reduces at the optimal rate but requires a few more iterations to reach the same level as that of our AFEM.

• Tables 3.3 and 3.4 document the role of oscillation in AFEM. Without marking due to oscillation $\mathsf{M_O} = 0$, estimator $\eta(\Omega)$ still reduces at optimal rate but oscillation reduction $\mathsf{RF_O}$ is not stable. The factor $\mathsf{RF_O}$ approximates ρ_1 of Lemma 3.3 and thus controls the oscillation decay between consecutive iterations. In fact Table 3.4 indicates that lack of control of $\mathsf{RF_O}$ leads to more iterations for the same estimator. Tables 3.3 and 3.4 illustrate the need of Marking Strategy O to control the reduction rate of oscillations and confirm the convergence theory of AFEM. Our experiments show that the ratio $\mathsf{M_E}/\mathsf{M_O}$ depends inversely on the ratio θ_E/θ_0. If $\theta_E = \theta_0$, then $\mathsf{M_E}$ dominates $\mathsf{M_O}$.
• Comparison of computational cost is measured using CPU time used by each procedure. In average, about 80% of total CPU time is used by SOLVE; the other procedures ESTIMATE, MARK and REFINE use about 5-10%.
• In theory, the initial meshsize h_0 must satisfy

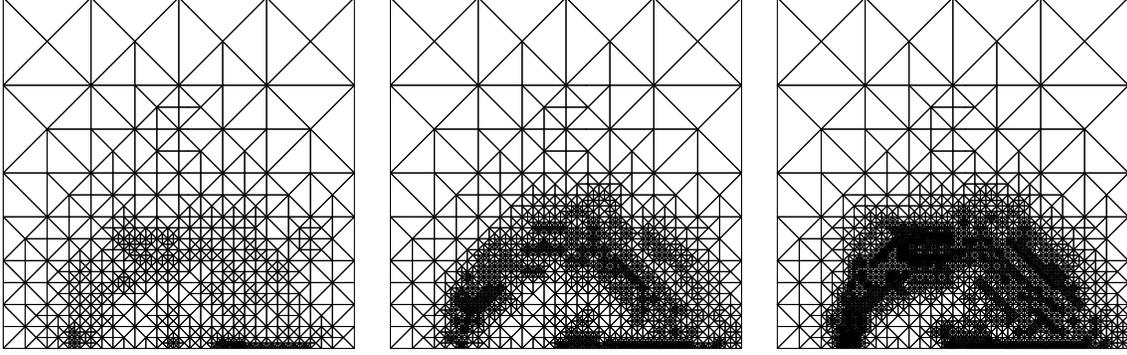$$C^* B h_0 < \frac{\beta \Lambda_1}{1 + \beta \Lambda_1} = \beta_0,$$

FIGURE 3.2. Example 3.4.2 (Convection-Dominated Diffusion with $\epsilon = 10^{-3}, \mathbf{b} = (y, \frac{1}{2} - x)$): Adaptively refined meshes after 5, 7, and 8 iterations corresponding to Table 3.3 starting from a uniform mesh coarser than required in theory. After a few iterations, AFEM detects the region of rapid variation (circular transport of a pulse) and boundary layer in the outflow, whereas the rest of the mesh remains unchanged. Refinement in smooth region is caused by early oscillations.
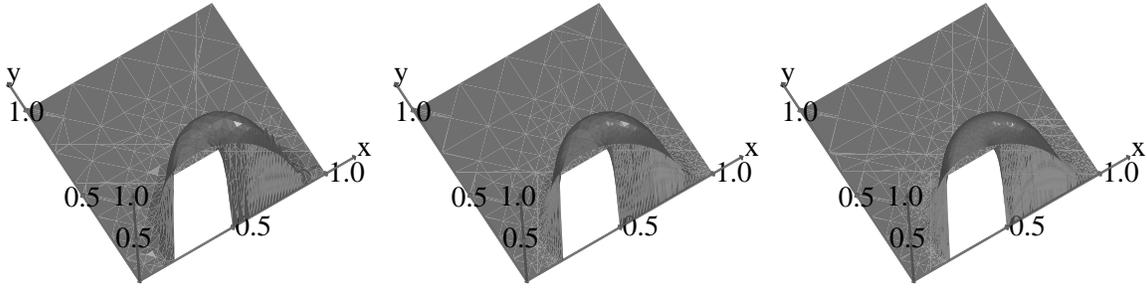

FIGURE 3.3. Example 3.4.2 (Convection-Dominated Diffusion with $\epsilon = 10^{-3}, \mathbf{b} = (y, \frac{1}{2} - x)$): plots of solutions after 5, 7, and 8 iterations. No oscillations (of Galerkin solutions) are detected after a few iterations even though AFEM is not stabilized.

where $B = \|\mathbf{b}\|_{L^\infty}$, $\beta_0 = O(1)$, and C^* is the constant from Lemma 3.2. In this particular case, we can express C^* in terms of ϵ and B quite explicitly. We first observe that H^2-regularity theory gives [28]

$$\begin{cases} \mathcal{L}\varphi &= \zeta \quad \text{in } \Omega \\ \varphi &= 0 \quad \text{on } \partial\Omega \end{cases} \implies \|\varphi\|_{H^2(\Omega)} \leq C_R B^{1/2} \epsilon^{-3/2} \|\zeta\|_{L^2(\Omega)}$$

with $C_R > 0$ independent of data. We also note that C_D of Lemma 3.2 satisfies

$$C_I C_R \left(\frac{B}{\epsilon}\right)^{\frac{3}{2}} h_0 \leq \frac{1}{2} \implies C_D = 2 C_I C_R \left(\frac{B}{\epsilon}\right)^{\frac{1}{2}},$$

where C_I is an interpolation constant solely dependent on shape regularity. This results from the usual duality argument and the fact that div $\mathbf{b} = 0$, namely

$$|\langle e_h, \zeta \rangle| = |\mathcal{B}[e_h, \varphi]| \leq C_I h_0 \left(\epsilon \|\nabla e_h\|_{L^2} + B \|e_h\|_{L^2}\right) \|\varphi\|_{H^2}.$$

We finally recall that $C^* = C_D/\epsilon$ (see proof of Lemma 3.1) to arrive at

$$h_0 < \frac{\beta_0}{2 C_I C_R} \left(\frac{\epsilon}{B}\right)^{3/2},$$

which is consistent with the previous restriction on h_0. We stress that this implies $h_0 \approx 10^{-4}$ in theory, whereas $h_0 \approx 10^{-1}$ works in our examples; see Figures 3.2-3.3.

• The local Peclet number $P_e = \frac{h_0 B}{\epsilon}$ is about 10^2 at the beginning. Since $P_e > 1$, and the Galerkin method is not stabilized, oscillations are observed in the first few iterations but cured later by AFEM

via local refinement; see Figure 3.3, which displays solutions without oscillations for iterations 7 and 8. Figure 3.2 depicts several graded meshes and confirms that mesh refinement is localized around the pulse location and outflow boundary layer. Minor refinement in the smooth region is caused by early oscillations.

3.5. The Stokes Operator. In this section we show convergence of an adaptive Uzawa FEM for the Stokes problem. The chief difficulty in this setting is the lack of a minimization principle and thus the failure of Lemma 2.11, which allows for the quantification of error reduction. Our approach is instead based on [3], which in turn exploits an idea introduced in [17] in the context of wavelet approximations; see also [16]. We refer to [3] for details, extensive computations, and complexity considerations.

3.6. The Stokes problem. Let $\mathbb{V} := (H_0^1(\Omega))^d$ and $\mathbb{P} := L_0^2(\Omega)$ be the subspace of $L^2(\Omega)$ of functions with zero mean value. Let $a : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ and $b : \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ be the following continuous bilinear forms

$$(3.13) \qquad a(v,\, w) := \langle \nabla v,\, \nabla w \rangle, \qquad b(v,\, q) := - \langle q,\, \mathrm{div}\, v \rangle, \qquad \forall\, v, w \in \mathbb{V}, q \in \mathbb{P}.$$

Then there exists a unique solution $(u, p) \in \mathbb{V} \times \mathbb{P}$ of the following saddle point problem:

$$(3.14) \qquad\qquad a(u,\, v) + b(v,\, p) = \langle f,\, v \rangle \qquad \forall\, v \in \mathbb{V},$$

$$(3.15) \qquad\qquad b(u,\, q) = 0 \qquad \forall\, q \in \mathbb{P};$$

see [7, Chapter II]. Let $A : \mathbb{V} \to \mathbb{V}^*$, $B : \mathbb{V} \to \mathbb{P}^* = \mathbb{P}$, and $B^* : \mathbb{P} \to \mathbb{V}^*$, the adjoint of B, be the operators $A := -\Delta$, $B := -\,\mathrm{div}$ and $B^* := \nabla$. The system (3.14)–(3.15) can be equivalently written in operator form as follows [7]:

$$(3.16) \qquad\qquad Au + B^*p = f, \qquad Bu = 0.$$

3.6.1. Uzawa AFEM. If $S := BA^{-1}B^* : \mathbb{P} \to \mathbb{P}$ denotes the Schur complement operator, which turns out to be positive definite, selfadjoint, and bounded, then p satisfies the equation

$$(3.17) \qquad\qquad Sp = BA^{-1}f.$$

Since $\alpha := \|I - \rho S\| < 1$ provided the relaxation parameter $0 < \rho < 2/\|S\|_{\mathcal{L}(\mathbb{P},\mathbb{P})}$, the following *Uzawa* iteration converges

$$(3.18) \qquad\qquad p_k = p_{k-1} - \rho\big(Sp_{k-1} - BA^{-1}f\big) = (I - \rho S)p_{k-1} + \rho BA^{-1}f.$$

We stress that $\rho = 1$ is an admissible choice for the Stokes problem, and that this iteration is carried out at the infinite dimensional level. In weak form it reads

$$(3.19) \qquad\qquad a(u_k,\, v) = \langle f,\, v \rangle - b(v,\, p_{k-1}) \qquad \forall\, v \in \mathbb{V}$$

$$(3.20) \qquad\qquad \langle p_k,\, q \rangle = \langle p_{k-1},\, q \rangle + \rho b(u_k,\, q) \qquad \forall\, q \in \mathbb{P}.$$

Suppose now that a procedure ELLIPTIC for the operator A, such as AFEM of Lecture 2, approximates the solution u_k of (3.19) to any desired tolerance ε_k:

$$(3.21) \qquad\qquad (\mathcal{T}_k, U_k) \leftarrow \mathsf{ELLIPTIC}(\mathcal{T}_{k-1}, P_{k-1}, \varepsilon_k, f).$$

In other words, given $P_{k-1} \in \mathbb{P}_{k-1}$ over the triangulation $\mathcal{T}_{k-1}$, ELLIPTIC constructs a refinement $\mathcal{T}_k$ of $\mathcal{T}_{k-1}$ and computes

$$(3.22) \qquad U_k \in \mathbb{V}_k : \qquad a(U_k,\, V) = \langle f,\, V \rangle - b(V,\, P_{k-1}) \qquad \forall\, V \in \mathbb{V}_k,$$

such that $\|U_k - u_k\|_{\mathbb{V}} \le \varepsilon_k$, where u_k is the *continuous* solution of

$$(3.23) \qquad u_k \in \mathbb{V} : \qquad a(u_k,\, v) = \langle f,\, v \rangle - b(v,\, P_{k-1}) \qquad \forall\, v \in \mathbb{V}.$$

We also assume that a procedure UPDATE for the operator B, namely

$$(3.24) \qquad\qquad P_k \leftarrow \mathsf{UPDATE}(\mathcal{T}_k, P_{k-1}, U_k, \rho),$$

computes a discrete solution of (3.20)

$$(3.25) \qquad P_k \in \mathbb{P}_k : \qquad \langle P_k,\, Q \rangle = \langle P_{k-1},\, Q \rangle + \rho b(U_k,\, Q) \qquad \forall\, Q \in \mathbb{P}_k.$$

If $\Pi_k : \mathbb{P} \to \mathbb{P}_k$ denotes the L^2-projection operator, then (3.25) reads equivalently

$$(3.26) \qquad P_k = P_{k-1} + \rho \Pi_k B U_k.$$

If j is the polynomial degree for velocity, and ℓ is that for pressure, it turns out that the pairs of *continuous* finite element spaces

$$(3.27) \qquad \mathbb{V}_k = \mathcal{P}^j(\mathcal{T}_k) \cap \mathbb{V}, \quad \mathbb{P}_k = \mathcal{P}^\ell(\mathcal{T}_k) \cap \mathbb{P}, \qquad \ell = j, j-1 \geq 1,$$

as well as the *discontinuous* finite element spaces

$$(3.28) \qquad \mathbb{V}_k = \mathcal{P}^j(\mathcal{T}_k) \cap \mathbb{V}, \quad \mathbb{P}_k = \mathcal{P}_d^{j-1}(\mathcal{T}_k) \cap \mathbb{P}, \qquad j \geq 1,$$

are of interest. Hereafter, $\mathcal{P}_d^j(\mathcal{T}_k)$ denotes the space of —scalar-valued as well as vector-valued— (possibly *discontinuous*) functions that restricted to an element T are polynomials of degree $\leq j$ for all $T \in \mathcal{T}_k$, and $\mathcal{P}^j(\mathcal{T}_k)$ denotes the subspace of *continuous* functions of $\mathcal{P}_d^j(\mathcal{T}_k)$. We observe that $\ell = j-1$ in (3.27) corresponds to the popular Taylor-Hood family of finite elements. Any other choice in either (3.27) or (3.28) yields an *unstable* pair of spaces. These spaces $(\mathbb{V}_k, \mathbb{P}_k)$ satisfy [3]:

$$(3.29) \qquad \|\Pi_k B U_k - B U_k\|_\mathbb{P} \leq C_* \varepsilon_k.$$

We then have the following convergence result for the Stokes problem, which improves upon the original one in [3].

Theorem 3.10 (Convergence of Uzawa AFEM). *Let $\rho > 0$ be such that $\alpha = \|I - \rho S\| < 1$. Given $\varepsilon_0 > 0$ and $0 < \gamma < 1$, define $\varepsilon_k := \gamma \varepsilon_{k-1}$. Let the procedures* ELLIPTIC *and* UPDATE *satisfy (3.22)–(3.26) and (3.29), and let $\alpha_0 := \max(\alpha, \gamma)$. Then, for every β such that $\alpha_0 < \beta < 1$, the iterates $(U_k, P_k) \in \mathbb{V}_k \times \mathbb{P}_k$ satisfy*

$$\|p - P_k\|_\mathbb{P} \leq C_p \beta^k, \qquad \|u - U_k\|_\mathbb{V} \leq C_u \beta^k,$$

with

$$C_p := \|p - P_0\|_\mathbb{P} + \frac{\rho(1 + C_*)\varepsilon_0}{\beta - \alpha_0}, \qquad C_u := \frac{C_p}{\beta} + \varepsilon_0.$$

Proof. In view of (3.22) and (3.26), we see that

$$P_k = P_{k-1} + \rho B u_k + \rho B(U_k - u_k) + \rho(\Pi_k - I) B U_k$$
$$= (I - \rho S) P_{k-1} + \rho B A^{-1} f + \rho B(U_k - u_k) + \rho(\Pi_k - I) B U_k.$$

Using (3.17), we readily get the error equation

$$p - P_k = (I - \rho S)(p - P_{k-1}) - \rho B(U_k - u_k) - \rho(\Pi_k - I) B U_k.$$

We now use the fact that $\|Bv\|_\mathbb{P} \leq \|v\|_\mathbb{V}$ for all $v \in \mathbb{V}$, and invoke (3.29), to arrive at

$$\|p - P_k\|_\mathbb{P} \leq \alpha \|p - P_{k-1}\|_\mathbb{P} + \rho(1 + C_*)\varepsilon_k,$$

whence, by recursion, we end up with

$$\|p - P_k\|_\mathbb{P} \leq \alpha^k \|p - P_0\|_\mathbb{P} + \rho(1 + C_*)\varepsilon_0 \sum_{j=0}^{k-1} \alpha^j \gamma^{k-j}.$$

Arguing similarly as in Theorem 2.13, we readily obtain

$$\|p - P_k\|_\mathbb{P} \leq \beta^k \left(\|p - P_0\|_\mathbb{P} + \frac{\rho(1 + C_*)\varepsilon_0}{\beta - \alpha_0} \right).$$

This is the asserted estimate for $\|p - P_k\|_\mathbb{P}$. To obtain a similar bound for $\|u - U_k\|_\mathbb{V}$, we first observe that

$$a(u - u_k, v) = b(v, P_{k-1} - p), \qquad \forall v \in \mathbb{V},$$

whence $\|u - u_k\|_\mathbb{V} \leq \|p - P_{k-1}\|_\mathbb{P}$. Since

$$\|u - U_k\|_\mathbb{V} \leq \|u - u_k\|_\mathbb{V} + \|u_k - U_k\|_\mathbb{V} \leq \|p - P_{k-1}\|_\mathbb{P} + \varepsilon_k,$$

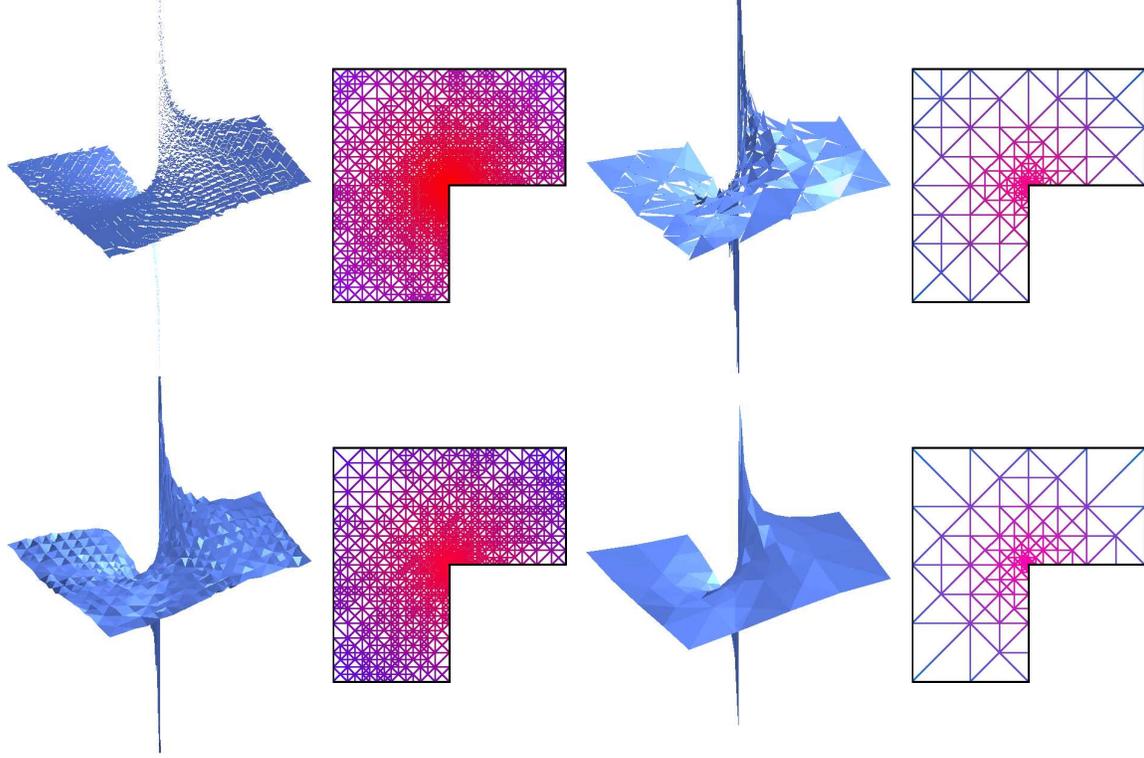the remaining estimate for $\|u - U_k\|_\mathbb{V}$ follows immediately and finishes the proof. $\qquad \square$

FIGURE 3.4. Example 3.6.2: Pressures and meshes for tolerance of 5% and finite element pairs (resp. outer iteration number/DOFs); $\mathcal{P}^1$-$\mathcal{P}^0_d$ (50/9680) , $\mathcal{P}^2$-$\mathcal{P}^1_d$ (35/1940), $\mathcal{P}^1$-$\mathcal{P}^1$ (50/4971), $\mathcal{P}^2$-$\mathcal{P}^1$ (50/1200). The oscillations for unstable elements do not persist under further selective refinement.

3.6.2. *Example: L-shaped domain.* We present a simulation on an unit-size L-shaped domain with exact solutions $u \approx r^{1/2}$ and $p \approx r^{-1/2}$ [3],[46]; see also Example 1.1. We only depict the pressure P_k in Figure 3.4, which is the most sensitive variable to instabilities. We point out that the only stable pair is the Taylor-Hood element $\mathcal{P}^2 - \mathcal{P}^1$ (Figure 3.4-right bottom), but that oscillations for unstable pairs do not persist under selective refinement. We refer to [3] for extensive computations showing quasi-optimal meshes for all combinations (3.27)-(3.28) in 2d and 3d.

The computational results corroborate the assertion of Theorem 3.10. We may rephrase this as follows: *the addition of least amount of selective refinement by adaptivity has a stabilizing effect.* This is quite different from stabilization techniques based on global refinement for velocity [7], and a surprising outcome of this approach.

3.7. Exercises.

3.7.1. *Exercise: Gårding's Inequality.* Suppose that $\operatorname{div} \mathbf{b} \neq 0$. Prove

$$(3.30) \qquad \qquad \|v\|^2 - \gamma_G \|v\|^2_{L^2(\Omega)} \leq \mathcal{B}[v,v] \qquad \forall\, v \in H^1_0(\Omega),$$

where $\gamma_G = \|\operatorname{div} \mathbf{b}\|_\infty /2$.

3.7.2. *Exercise: Duality.* Prove Lemma 3.2.

3.7.3. *Exercise: Interpolation Estimate for Discontinuous Polynomials.* Prove the estimate (3.29).

3.7.4. *Exercise: Complexity of* ELLIPTIC. Suppose the tolerance reduction factor γ of Theorem 3.10 satisfies $\gamma > \alpha$. Show that the number of iterations of ELLIPTIC is bounded by a constant which depends only on f, the initial pressure guess P_0, the initial triangulation $\mathcal{T}_0$, the ratio α/γ, and the parameters θ_E and θ_0 of ELLIPTIC, but on the adaptive counter k.