Math5453M FEM Numerical Exercises 3

Joel Brown

Question 1

Consider the Poisson system on domain $(x, y) \in [0, 1]^2$:

$$\begin{cases} -\nabla^2 u = f(x, y) & (1.1) \\ f(x, y) = 2\pi^2 \sin(\pi x)\cos(\pi y) & (1.2) \\ u(0, y) = u(1, y) = 0 & (1.3) \\ \frac{\partial u}{\partial y}\Big|_{y=0} = \frac{\partial u}{\partial y}\Big|_{y=1} = 0 & (1.4) \end{cases} \tag{1}$$

The Ritz-Galerkin principle for this system is found by considering a variation on the energy functional, given by

$$\mathfrak{J}[u] = \int_\Omega \left[\frac{1}{2}|\nabla u|^2 - fu\right] d\underline{x} = \int_0^1 \int_0^1 \left[\frac{1}{2}|\nabla u|^2 - fu\right] dxdy. \tag{2}$$

Consider variation δ defined as $u \to u + \epsilon\delta u$ with ϵ small, the Ritz Galerkin principle is:

$$\begin{aligned} &\delta \int_0^1 \int_0^1 \left[\frac{1}{2}|\nabla u|^2 - fu\right] dxdy = 0 \\ \Rightarrow &\int_0^1 \int_0^1 \left[\frac{1}{2}\left|\nabla(u + \epsilon\delta u)\right|^2 - f(u + \epsilon\delta u)\right] dxdy = 0 \\ \Rightarrow &\int_0^1 \int_0^1 \left[\frac{1}{2}|\nabla u|^2 + \frac{1}{2}|\epsilon\nabla\delta u|^2 + \epsilon\nabla u \cdot \nabla\delta u - fu - f\epsilon\delta u\right] dxdy = 0 \end{aligned}$$

Differentiate this equation with respect to ϵ,

$$\int_0^1 \int_0^1 \left[\epsilon\left|\nabla\delta u\right|^2 + \nabla u \cdot \nabla\delta u - f\delta u\right] dxdy = 0$$

Since we are considering infinitesimal variations, let $\epsilon \to 0$, and we are left with the following weak formulation

$$\int_0^1 \int_0^1 \left[\nabla u \cdot \nabla\delta u - fu\right] dxdy = 0 \tag{3}$$

Now let us consider a space of square integrable functions and corresponding Hilbert space in which functions satisfy the Dirichlet boundary conditions:

$$\begin{aligned} L^2\left([0, 1]^2\right) &= \left\{w \Big| \int_0^1 \int_0^1 |w|^2 \, dxdy < \infty\right\} \\ H_0^2\left([0, 1]^2\right) &= \left\{w \in L^2([0, 1]^2) \,|\, \partial_y w, \partial_x w \in L^2\left([0, 1]^2\right),\, w|_{x=0} = w|_{x=1} = 0\right\} \end{aligned}$$

Multiply both sides of equation (1.1) by test function $w(x, y) \in H_0^2([0, 1]^2)$, and integrate over domain $[0, 1]^2$,

$$-\int_0^1 \int_0^1 w\nabla^2 u \, dxdy = \int_0^1 \int_0^1 wf \, dxdy$$

Using the identity $a\nabla^2 b = \nabla \cdot (a\nabla b) - \nabla a \cdot \nabla b$, where a and b are scalar functions, the left hand side of the equation is equivalent to

$$\int_0^1 \int_0^1 (\nabla w \cdot \nabla u - \nabla \cdot (w\nabla u)) \, dx \, dy$$

.

Let $\underline{\hat{u}}_1, \underline{\hat{u}}_2, \underline{\hat{u}}_3, \underline{\hat{u}}_4$ be the outward facing normals on each edge of the boundary $\partial[0, 1]^2$, such that

- Edge $(x,y) \in [0] \times [0,1]$ has boundary condition $u(0,y) = 0$ and normal $\underline{\hat{u}}_1 = (-1, 0)$
- Edge $(x,y) \in [0,1] \times [1]$ has boundary condition $\partial_y u|_{y=1} = 0$ and normal $\underline{\hat{u}}_2 = (0, 1)$
- Edge $(x,y) \in [1] \times [0,1]$ has boundary condition $u(1,y) = 0$ and normal $\underline{\hat{u}}_3 = (1, 0)$
- Edge $(x,y) \in [0,1] \times [0]$ has boundary condition $\partial_y u|_{y=0} = 0$ and normal $\underline{\hat{u}}_4 = (0, -1)$

Therefore, by the divergence theorem, the left hand side is equal to:

$$\int_0^1 \nabla w \cdot \nabla u \, dx \, dy + \int_0^1 \left(w \frac{\partial u}{\partial x} \right)\Big|_{x=0} dy - \int_0^1 \left(w \frac{\partial u}{\partial y} \right)\Big|_{y=1} dx + \int_0^1 \left(w \frac{\partial u}{\partial x} \right)\Big|_{x=1} dy - \int_0^1 \left(w \frac{\partial u}{\partial y} \right)\Big|_{y=0} dx$$

Since $w = 0$ on $x = 0, 1$ and $\frac{\partial u}{\partial y} = 0$ on $y = 0, 1$, all terms except the first disappear, so

$$\int_0^1 \int_0^1 \nabla w \cdot \nabla u \, dx \, dy = \int_0^1 \int_0^1 w f \, dx \, dy$$
$$\Rightarrow \int_0^1 \int_0^1 \left[\nabla u \cdot \nabla w - f w \right] dx \, dy = 0$$

Compare this to the weak form (3) found earlier, we see that the test function $w(x,y)$ and variation $\delta u(x,y)$ are equivalent.

Question 2

So far we have the following Ritz-Galerkin principle and weak formulation:

$$\begin{cases} \delta \int_0^1 \int_0^1 \left[\frac{1}{2} |\nabla u|^2 - f u \right] dx \, dy = 0 & (4.1) \\ \int_0^1 \int_0^1 \left[\nabla u \cdot \nabla w - f w \right] dx \, dy = 0 & (4.2) \end{cases} \tag{4}$$

Let us introduce an FEM approximation of u given by

$$u(x,y) \approx u_h(x,y) = \sum_{j=1}^{N} \hat{u}_j \phi_j(x,y) \tag{5}$$

where $\hat{u}_j$ are coefficients, ϕ_j are basis functions and N is the number of degrees of freedom of the mesh. Substituting $u = u_h$ into the Ritz-Galerkin principle (4.1):

$$\delta \int_0^1 \int_0^1 \left[\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{u}_i \hat{u}_j \nabla \phi_i \cdot \nabla \phi_j - \sum_{j=1}^{N} f \hat{u}_j \phi_j \right] dx \, dy = 0$$
$$\Rightarrow \delta \sum_{j=1}^{N} \left(\frac{1}{2} \sum_{i=1}^{N} A_{ij} \hat{u}_i \hat{u}_j - b_j \hat{u}_j \right) = 0$$

where $A_{ij} = \int_0^1 \int_0^1 \nabla \phi_i \cdot \nabla \phi_j \, dx \, dy$ and $b_j = \int_0^1 \int_0^1 f \phi_j \, dx \, dy$.
Differentiating with respect to $\hat{u}_j$,

$$\delta \sum_{j=1}^{N} \left(\sum_{i=1}^{N} A_{ij} \hat{u}_i - b_j \right) = 0$$

Thus, writing in index notation, we have the matrix equation

$$\boxed{A_{ij} \hat{u}_i = b_j} \tag{6}$$

Now let us do the same with the weak formulation (4.2), substituting in $u = u_h$ and $w = \phi_j$

$$\int_0^1 \int_0^1 \left[\nabla \left(\sum_{i=1}^{N} \hat{u}_i \phi_i \right) \cdot \nabla \phi_j - f \phi_j \right] dx \, dy = 0 \quad j = 1, 2, ..., N$$
$$\Rightarrow \int_0^1 \int_0^1 \left[\sum_{i=1}^{N} \hat{u}_i \nabla \phi_i \cdot \nabla \phi_j - f \phi_j \right] dx \, dy = 0$$
$$\Rightarrow \sum_{i=1}^{N} A_{ij} \hat{u}_i - b_j = 0$$

This gives us the exact same result, equation (6).

Question 3

Consider quadrilateral mesh element K with vertices (x_i, y_i), $i = 1, 2, 3, 4$ and introduce a reference coordinate system $(\xi, \eta) \in [-1, 1]^2$ such that

$$x(\xi, \eta) = \sum_{i=1}^{N} x_i \phi_i(\xi, \eta), \quad y(\xi, \eta) = \sum_{i=1}^{N} y_i \phi_i(\xi, \eta) \tag{7}$$

Letting g be some function of x and y, we transform an integral of the function g over mesh element K according to the identity:

$$\int_K g(x, y)\, dx\, dy = \int_{-1}^{1} \int_{-1}^{1} g\left(x(\xi, \eta), y(\xi, \eta)\right) |J|\, d\xi\, d\eta \tag{8}$$

where $J = \begin{pmatrix} \partial_\xi x & \partial_\eta x \\ \partial_\xi y & \partial_\eta y \end{pmatrix}$ is the Jacobian.

Let $A_{ij}^K = \int_K \nabla\phi_i(\xi, \eta) \cdot \nabla\phi_j(\xi, \eta)\, dx\, dy = \int_{-1}^{1} \int_{-1}^{1} \nabla\phi_i(\xi, \eta) \cdot \nabla\phi_j(\xi, \eta) |J|\, d\xi\, d\eta$,

$$\therefore \boxed{A_{ij} = \sum_K \int_{-1}^{1} \int_{-1}^{1} \nabla\phi_i(\xi, \eta) \cdot \nabla\phi_j(\xi, \eta) \left(\frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta} \frac{\partial y}{\partial \xi} \right) \Big|_{x=x(\xi,\eta), y=y(\xi,\eta)} d\xi\, d\eta} \tag{9}$$

where $\nabla\phi_i = J^{-T} \begin{pmatrix} \partial_\xi \\ \partial_\eta \end{pmatrix} \phi_i$.

Let $b_j^K = \int_K f\phi_j\, dx\, dy = \int_{-1}^{1} \int_{-1}^{1} f\left(x(\xi, \eta), y(\xi, \eta)\right) \phi_j(\xi, \eta) |J|\, d\xi\, d\eta$

$$\therefore \boxed{b_j = \sum_K \int_{-1}^{1} \int_{-1}^{1} f\left(x(\xi, \eta), y(\xi, \eta)\right) \phi_j(\xi, \eta) \left(\frac{\partial x}{\partial \xi} \frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta} \frac{\partial y}{\partial \xi} \right) \Big|_{x=x(\xi,\eta), y=y(\xi,\eta)} d\xi\, d\eta} \tag{10}$$

We construct A_{ij} and b_j by summing over all elements K of the mesh.

Question 4

The Poisson system we are considering has exact solution $f_e(x, y) = \sin(\pi x) \cos(\pi y)$. In the provided Firedrake code, a solution $u_h \approx u_e$ is found numerically using the finite element Continuous Galerkin (CG) method. Two different methods for computing the weak form of the equation are used. In the first, it is constructed manually by multiplying and manipulating the Poisson equation, giving solution $u_{h,1}$. In the second, it is constructed by differentiating or variating the Ritz-Galerkin principle, giving solution $u_{h,2}$. The code also computes the L^2-errors: $L^2(u_{h,1}, u_e), L^2(u_{h,2}, u_e)$ and $L^2(u_{h,1}, u_{h,2})$, where

$$L^2(u_i, u_j) = \sqrt{\int_0^1 \int_0^1 |u_i(x) - u_j(x)|^2\, dx\, dy}$$

These solutions and errors can be computed for different combinations $\{h, p\}$, where $h = \frac{1}{N_x}$ where N_x is the number of elements the x-axis is split into, and p is the order of the method (i.e. $p = 1$ for the piecewise linear CG1 method). I ran the code for all 24 combinations of $h = 1/16, 1/32, 1/64, 1/128$ and $p = 1, 2, 3, 4, 5, 6$.

The solution $u_{h,1}$ for $\{h, p\} = \{1/16, 1\}$ is shown as a contour plot in Figure 1 below. By comparing it to a contour plot of the exact solution, we see that we get an accurate result, and we would expect the L^2-error to be small. To check, the errors are plotted for different values of h and p in Figure 2. We see that the combination $\{1/16, 1\}$ has an error of magnitude 10^{-3}.

In general, the L^2-error for the combination $\{h, p\}$ converges according to:

$$L^2(u_{h,1}, u_e) \leq Ch^{p+1} \tag{11}$$

for some constant C. As an example, consider CG1 ($p = 1$). The estimated values of C are given for different h-values in Table 1 where we see that $C \approx 0.41$.

We thus can say that the L^2-errors of two combinations $\{h_1, p_1\}, \{h_2, p_2\}$ are (roughly) equivalent if $h_1^{p_1+1} \approx h_2^{p_2+1}$. We can see by simple calculation that we have six such cases:

- $\{1/64, 1\}$ & $\{1/16, 2\}$: $\left(\frac{1}{64}\right)^{1+1} = \left(\frac{1}{16}\right)^{2+1}$

h	$L^2(u_{h,1}, u_e)$ (4 s.f.)	$C \approx L^2/h^2$ (4 s.f.)
1/16	1.593×10^{-3}	0.4078
1/32	4.008×10^{-4}	0.4104
1/64	1.003×10^{-4}	0.4108
1/128	$2.510 \times 10^{-}5$	0.4112

Table 1: Estimated C-values for the CG1 method.

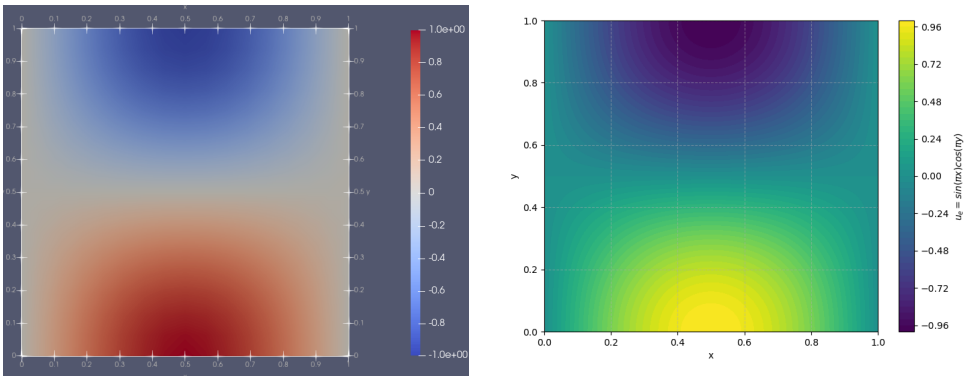

Figure 1: Left: Contour plot of numerical solution for the first weak form method, with $h = 1/16$ and $p = 1$, using Paraview. Right: Contour plot of exact solution u_e, using Python code

.

- $\{1/32, 3\}$ & $\{1/16, 4\}$: $\left(\frac{1}{32}\right)^{3+1} = \left(\frac{1}{16}\right)^{4+1}$
- $\{1/64, 3\}$ & $\{1/16, 5\}$: $\left(\frac{1}{64}\right)^{3+1} = \left(\frac{1}{16}\right)^{5+1}$
- $\{1/64, 4\}$ & $\{1/32, 5\}$: $\left(\frac{1}{64}\right)^{4+1} = \left(\frac{1}{32}\right)^{5+1}$
- $\{1/128, 4\}$ & $\{1/32, 5\}$: $\left(\frac{1}{128}\right)^{4+1} = \left(\frac{1}{32}\right)^{5+1}$
- $\{1/128, 5\}$ & $\{1/164, 6\}$: $\left(\frac{1}{128}\right)^{5+1} = \left(\frac{1}{64}\right)^{6+1}$

I have plotted the difference $|u_{h,1}(x, y) - u_e|$ for both $\{1/64, 1\}$ and $\{1/16, 2\}$ as contour maps in Paraview. These are shown in Figure 3. Though the two methods have similar convergence patterns, we can see that the CG2 method on the left gives a much smaller maximal error (magnitude 10^{-6}) than the CG1 method on the right (magnitude 10^{-4}). However the estimated solution $u_{1/16,1}$, with the domain only split into $16^2 = 256$ elements gives a much less smooth and accurate plot, as opposed to $u_{1/64,1}$ with $64^2 = 4096$ elements.
As well as the increased time needed for Firedrake to complete these methods as the order p increases, it is important to balance the order and mesh element size, to ensure that the L^2-error is not too high and the solution is sufficiently smooth.

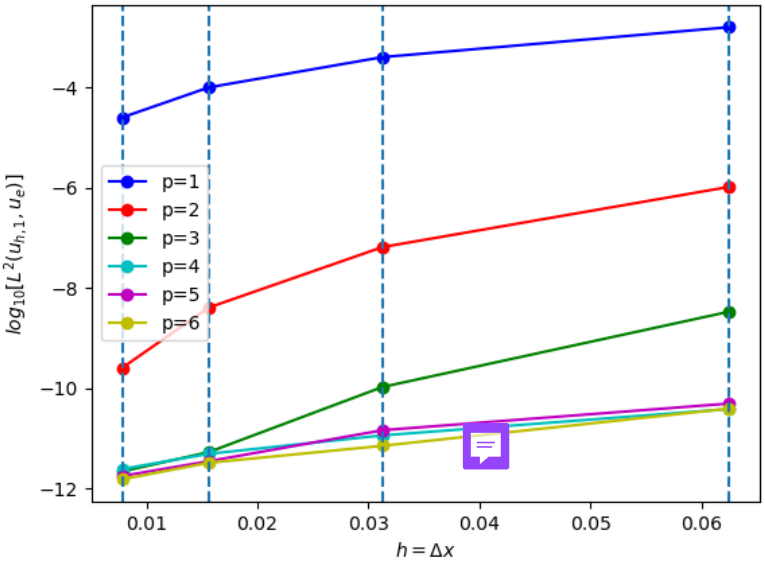

Figure 2: Logarithmic plot of error $L^2(u_{h,1}, u_e)$ for different $\{h, p\}$-values
. The vertical lines from left to right represent $h = 1/128, 1/64, 1/32$ and $1/16$ respectively.

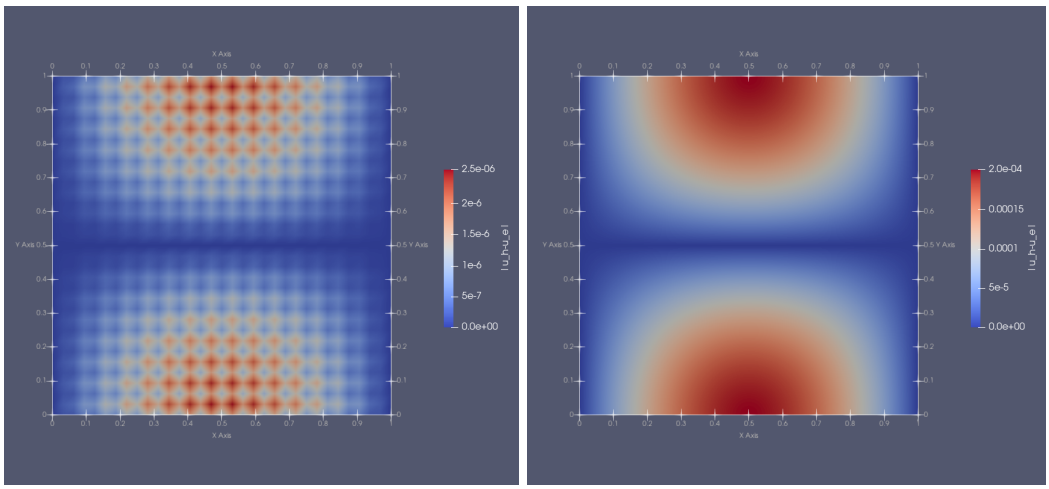

Figure 3: Contour maps of the difference between the numerical solution $u_{h,1}$ and exact solution u_e for CG2 with $h = 1/16$ (left) and CG1 with $h = 1/64$ (right).