Approximation Theory and Finite Element Analysis

Edward Small 10786391

April 2021

Contents

0.1 Preamble

Given an open bounded domain $\Omega \subset \mathbb{R}^2$ with boundary conditions $\partial\Omega$ consisting of two different, non-overlapping pieces $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$, let $u : \Omega \to \mathbb{R}$ be the solution to the boundary value problem

$$\begin{aligned} -\nabla^2 u &= 0 \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega_D, \qquad \frac{\partial u}{\partial n} + u &= 1 \text{ on } \partial\Omega_N \end{aligned} \tag{1}$$

You can assume that $\partial\Omega_N$ has non-zero length so that the *Robin* boundary condition holds on some part of the boundary. You may use the Cauchy– Schwarz inequality or the Poincare–Friedrichs inequality in answering any of the following questions without giving a proof.

1 Interpretation of the problem

Question: Explain what is meant by a *classical solution* of (1). What is the physical relevance of the boundary condition on $\partial\Omega_N$?

1.1 Classical Solution

A classical solution of (1) would be a solution of u that is smooth enough to satisfy the conditions prescribed inside the domain Ω and the boundary conditions $\partial\Omega$, which we will label as $\overline{\Omega} = \Omega \cup \partial\Omega$. We require that u is suitably smooth in each specified domain. More specifically in this case, we require that

- $u \in C^2(\Omega)$, since the Laplacian is equal to a constant (in this case 0).
- $u \in C^1(\overline{\Omega})$

Therefore the classical solution for u for this mixed boundary condition case needs to be (at least) twice differentiable inside the domain Ω and once differentiable on the boundary, or

$$u \in C^2(\Omega) \cap C^1(\overline{\Omega}) \tag{2}$$

1.2 Physical Interpretation

Before we continue exploring this problem, it is sometimes best to give a physical intuition before pressing on. The open domain (Ω) is the area of interest, and it has been defined to be 2-dimensional ($\mathbb{R}^2$). We could potentially see this as a flat sheet, where a unique point in the domain requires 2 coordinates (possibly x and y) to define. The total problem ($\overline{\Omega}$) is split into 3 parts, with the boundary of the domain ($\partial\Omega$) is split into two parts.

- The bounded open domain (Ω) where the Laplacian of u is defined as fixed at 0 (also known as the Laplace equation). Simply put, this means that each point in the domain takes the average of the neighbouring points. If we saw this problem as the temperature across a flat sheet, this would mean that we have no "spikes" in temperature relative to points infinitesimally close to the point we are analysing.
- The Dirichlet boundary ($\partial\Omega_D$) where u is defined and fixed at 0. Following the problem as temperature across a flat sheet, we could interpret this to mean that this part of the boundary is dipped in ice, and is fixed at 0.

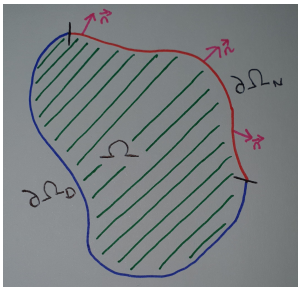


Figure 1: Picture of the domain with corresponding boundary conditions. The open bounded domain is the green centre section, with Dirichlet boundary being blue and the Robin boundary being red. Some normals to the Robin boundary are drawn in pink

- The Robin boundary ($\partial\Omega_N$) where u is not fixed, but the normal derivative is defined, where n is the normal of the boundary (shown in pink in figure 1). It could be seen as a weighted combination of the Dirichlet conditions (described above) and Neumann conditions (where the derivative of u with respect to the normal is specified by some constant or function). Basically, we can gain or lose "energy" through this surface through some amount of flux which is proportional to u itself, plus some forcing term (in this case, 1). This type of boundary follows reasonably intuitively from the heat example. If the boundary is very very hot (relative to its surroundings), we would expect it to cool quickly, and so the change in u will be greater, proportional to itself. The hotter the boundary is at this point, the more heat it can radiate out (also known as a convective boundary condition), and so the change in temperature is directly linked to the current temperature. We also have a small forcing term of 1, which we could see as the ambient temperature surrounding the boundary heating it slightly.

As a fully fledged example, we could see this problem as a 2-dimensional sheet of metal, taking some shape where the temperature on one boundary is fixed at 0, and on the other boundary we have leaky insulation where we lose heat.

It is worth noting here that the boundary does not have to specifically be in two distinct parts, just that we only have two ways to define values on the boundary - either Robin or Dirichlet. We could have the boundary being split into, say 4 sections, where (going around the boundary) we have a Robin section, a Dirichlet section, then another Robin, and another Dirichlet. The parts of the boundary, however, cannot overlap, which means $\partial\Omega_N \cap \partial\Omega_D = \emptyset$.

2 Test Space

Question: Given the test space $X := \{v | v \in \mathcal{H}^1(\Omega), v = 0 \text{ on } \partial\Omega_D\}$, where $\mathcal{H}^1$ is the standard Sobolev space, show that u solving (1) also satisfies the variational formula: find $u \in X$ such that

$$\int_\Omega \nabla u \cdot \nabla v + \int_{\partial\Omega_N} uv ds = \int_{\partial\Omega_N} v ds \qquad \forall v \in X \tag{3}$$

Define the Galerkin approximation $u_h \in X_h = \text{ span}\{\phi_j\}_{j=1}^k \subset X$ and show that the Galerkin approximation leads to a $k \times k$ matrix system

$$A\mathbf{x} = \mathbf{f} \tag{4}$$

Identify explicitly the entries A_{ij} of the matrix A and f_i of the vector $\mathbf{f}$.

2.1 Variational Formula

The test space X contains functions v that are defined in the same way as u. In this case $v : \Omega \to \mathbb{R}$, where the domain $\Omega \subset \mathbb{R}^2$. We require that each $v \in \mathcal{H}^1(\Omega)$, which is to say that they have a square integrable first derivative inside the domain, and that it is 0 on the Dirichlet boundary. Any function in the test space must be residually orthogonal to to the PDE, such that

$$\int_\Omega \{\nabla^2 u\} v = 0 \tag{5}$$

Using this, we can reduce the required differentiability of u. We start by taking the PDE and multiplying it by a function from the test space, which gives

$$-(\nabla^2 u)v = 0 \tag{6}$$

From here, we can integrate by parts over the domain to get

$$-\int_\Omega (\nabla^2 u)v \tag{7}$$

When separating by parts, we can employ Green's first identity, which states

$$\int_\Omega g\nabla^2 f = \int_{\partial\Omega} g\frac{\partial f}{\partial n} - \int_\Omega \nabla g \cdot \nabla f \tag{8}$$

to get boundary integrals. This leaves us with

$$-\int_\Omega (\nabla \cdot \nabla u)v = -\int_\Omega \nabla v \cdot \nabla u + \int_{\partial_\Omega} v\frac{\partial u}{\partial n}ds = 0 \tag{9}$$

and so

$$\int_\Omega \nabla u \cdot \nabla v - \int_{\partial_\Omega} v\frac{\partial u}{\partial n}ds = 0 \tag{10}$$

However, the boundary is split into two parts, so

$$\int_{\partial\Omega} \frac{\partial u}{\partial n}ds = \int_{\partial\Omega_D} \frac{\partial u}{\partial n}ds + \int_{\partial\Omega_N} \frac{\partial u}{\partial n}ds \tag{11}$$

The temperature is fixed to 0 on the Dirichlet boundary, and so we expect to have no flux through here. Not only that, but the partial derivative with respect to the normal is defined on the Robin boundary. Allowing for the Dirichlet boundary flux to be 0, and using the conditions at the Robin boundary as a substitution, we get that

$$\begin{aligned}\int_{\partial\Omega} \frac{\partial u}{\partial n}ds &= \int_{\partial\Omega_N} \frac{\partial u}{\partial n}ds \\ &= \int_{\partial\Omega_N} 1ds - \int_{\partial\Omega_N} uds\end{aligned} \tag{12}$$

Substituting (12) into (10) gives

$$\int_\Omega \nabla v \cdot \nabla u - \int_{\partial\Omega_N} vds + \int_{\partial\Omega_N} uvds = 0 \tag{13}$$

and therefore, rearranged, we get

$$\int_\Omega \nabla u \cdot \nabla v + \int_{\partial\Omega_N} uvds = \int_{\partial\Omega_N} vds \tag{14}$$

which must be true for all functions v inside the test space X. This gives us a *weak* formulation of the problem because the u that satisfies (14) does not need to satisfy the Robin condition at every point, but needs to do so averagely across the boundary.

2.2 Galerkin Approximation

Now that we have a weak solution that satisfy the Dirichlet condition, we can now introduce a subset of the test space X. Currently, X is an infinite dimensional space. We want to create a finite-dimensional approximation of the space X^h, which will be a subset of the space X, so

$$u_h \in X_h \subset X \text{ with } X_h = \text{span}\{\phi_j(x,y)\}_{j=1}^k \tag{15}$$

where

- x and y are the two variables needed to define a unique point on the domain
- $k = n + n_\partial$ are the number of interpolation points the domain has been split into, with k points defined for $\Omega \cup \partial\Omega_N$ and n_∂ points defined on the boundary $\partial\Omega_D$

- ϕ_j satisfy interpolation conditions

$$\phi_j = \begin{cases} 1 & \text{at vertex } j \\ 0 & \text{at vertex } i \neq j \end{cases} \tag{16}$$

We can then define u_h to be a finite element approximation to the weak solution u that satisfies (14), so

$$\int_\Omega \nabla u_h \cdot \nabla v_h + \int_{\partial\Omega_N} u_h v_h ds = \int_{\partial\Omega_N} v_h ds \tag{17}$$

which is the same as (14), except the problem is now posed on finite subspaces. Since we have k points to interpolate between, we require k functions, hence (15). u_h is then constructed in such a way that

$$u_h(x,y) = \sum_{j=1}^{n} u_j \phi_j(x,y) + \sum_{j=n+1}^{k} \overline{u}_j \phi_j(x,y) \tag{18}$$

where

- $\sum_{j=1}^{n} u_j \phi_j(x,y) = 0$ on $\partial\Omega_D$
- $\sum_{j=n+1}^{k} \overline{u}_j \phi_j(x,y) = 0$ on $\partial\Omega_D$
- u_j are unknown coefficients
- $\overline{u}_j$ are fixed values, and represent the value of u at the jth boundary vertex

Following the temperature example given in the previous section, each value of u_j would be the temperature at the jth vertex. Since each ϕ_i form a basis for the approximation test space X^h we can rewrite (17) to get

$$\int_\Omega \nabla u_h \cdot \nabla \phi_i + \int_{\partial\Omega_N} u_h \phi_i ds = \int_{\partial\Omega_N} \phi_i ds \tag{19}$$

And so substituting in (18) gives

$$\begin{gathered} \int_\Omega \nabla\Big\{ \sum_{j=1}^{n} u_j\phi_j + \sum_{j=n+1}^{k} \overline{u}_j\phi_j \Big\} \cdot \nabla\phi_i + \int_{\partial\Omega_N} \Big\{ \sum_{j=1}^{n} u_j\phi_j + \sum_{j=n+1}^{k} \overline{u}_j\phi_j \Big\}\phi_i ds = \int_{\partial\Omega_N} \phi_i ds \\ \sum_{j=1}^{n} u_j \Big\{ \int_\Omega \nabla\phi_j \cdot \nabla\phi_i + \int_{\partial\Omega_N} \phi_j\phi_i ds \Big\} + \sum_{j=n+1}^{k} \overline{u}_j \Big\{ \int_\Omega \nabla\phi_j \cdot \nabla\phi_i + \int_{\partial\Omega_N} \phi_j\phi_i ds \Big\} = \int_{\partial\Omega_N} \phi_i ds \end{gathered} \tag{20}$$

Rearranging gives

$$\sum_{j=1}^{n} u_j \Big\{ \int_\Omega \nabla\phi_j \cdot \nabla\phi_i + \int_{\partial\Omega_N} \phi_j\phi_i ds \Big\} = \int_{\partial\Omega_N} \phi_i ds - \sum_{j=n+1}^{k} \overline{u}_j \Big\{ \int_\Omega \nabla\phi_j \cdot \nabla\phi_i + \int_{\partial\Omega_N} \phi_j\phi_i ds \Big\} \tag{21}$$

which is the same as solving the $k \times k$ system

$$A\mathbf{x} = \mathbf{f} \tag{22}$$

where

- $A_{ij} = \int_\Omega \nabla\phi_j \cdot \nabla\phi_i$ (since $\int_{\partial\Omega_N} \phi_j\phi_i ds = 0$)
- $f_i = \int_{\partial\Omega_N} \phi_i ds$ (since $\overline{u} = 0$)

3 Unique Solution

Question: Discuss whether or not a solution of (14) is a unique solution in the special case $\partial\Omega = \partial\Omega_N$.

3.1 Unique Strong Solution

If a solution to u is not unique, then there must exist (at least) one other solution. Let us assume that there are two solutions and label them u_1 and u_2. We can then define the difference between the solutions to be

$$w = u_1 - u_2 \tag{23}$$

where w will satisfy the homogeneous Robin boundary ($\frac{\partial w}{\partial n} + w = 0$) and the Laplace equation. Since $\nabla^2 w = 0$ and $f\nabla^2 f = \nabla \cdot (f\nabla f) - (\nabla f)^2$

$$\begin{aligned} \int_\Omega w\nabla^2 w &= 0 \\ \int_\Omega \left\{\nabla \cdot (w\nabla w) - (\nabla w)^2\right\} &= 0 \end{aligned} \tag{24}$$

We can then split this into two integrals so that

$$\int_\Omega \nabla \cdot (w\nabla w) - \int_\Omega (\nabla w)^2 = 0 \tag{25}$$

Next, we use the divergence theorem

$$\int_\Omega \nabla f = \int_{\partial\Omega} \frac{\partial f}{\partial n} ds \tag{26}$$

to get

$$\int_{\partial\Omega} w\frac{\partial w}{\partial n} ds = \int_\Omega (\nabla w)^2 \tag{27}$$

Since w satisfies the homogeneous Robin boundary, then

$$\frac{\partial w}{\partial n} = -w \tag{28}$$

and so (27) is

$$\int_{\partial\Omega} w(-w) ds = \int_\Omega (\nabla w)^2 \tag{29}$$

However

$$\int_{\partial\Omega} w(-w) ds \leq 0 \qquad \int_\Omega (\nabla w)^2 \geq 0 \tag{30}$$

which means

$$w = 0 \Longleftrightarrow u_1 = u_2 \tag{31}$$

and so therefore u is unique.

3.2 Unique weak solution

If the entire boundary is set to be the Robin boundary, we instead set the conditions for the test space to be that the average value of the function is 0, so

$$\frac{1}{|\partial\Omega|} \int_{\partial\Omega} \left(\frac{\partial v_h}{\partial n} + v_h \right) = 0 \tag{32}$$

where $|\partial\Omega|$ is the length of the domain. Once again, assume u has two solutions and label them u_1 and u_2. We create a new function w defined by

$$w = u_1 - u_2 \tag{33}$$

which is in the test space. This w function passes all requirements to be in the test space as performing the average integration over the same domain for two different solutions will give the same constant, say C. $C - C = 0$, which is the requirement to be in the test space, and so w is in the test space.

From (14) we can then set up two equations

$$\begin{aligned} \int_\Omega \nabla u_1 \cdot \nabla v + \int_{\partial\Omega_N} u_1 v ds &= \int_{\partial\Omega_N} v ds \\ \int_\Omega \nabla u_2 \cdot \nabla v + \int_{\partial\Omega_N} u_2 v ds &= \int_{\partial\Omega_N} v ds \end{aligned} \tag{34}$$

and take the away from each other to get

$$\begin{aligned} \int_\Omega \nabla(u_1 - u_2) \cdot \nabla v + \int_{\partial\Omega_N} (u_1 - u_2) v ds &= 0 \\ \int_\Omega \nabla w \cdot \nabla v + \int_{\partial\Omega_N} w v ds &= 0 \end{aligned} \tag{35}$$

Since v is also in the test space, we can set $v = w$ to get

$$\begin{aligned} \int_\Omega \nabla w \cdot \nabla w + \int_{\partial\Omega_N} w w ds &= 0 \\ \sqrt{\int_\Omega \nabla w \cdot \nabla w + \int_{\partial\Omega_N} w w ds} &= 0 \\ \sqrt{\int_\Omega (\nabla w)^2} \sqrt{\int_\Omega (w)^2} &= 0 \\ ||\nabla w||_{L^2(\Omega)} ||w||_{L^2(\Omega)} &= 0 \end{aligned} \tag{36}$$

However, Poincaré-Friedrichs inequality states that

$$||w||_{L^2(\Omega)} \leq \alpha ||\nabla w||_{L^2(\Omega)} \tag{37}$$

and the rule of norms state that $||.|| \geq 0$. Therefore, (36) only holds if $||\nabla w||_{L^2(\Omega)} = 0$ or $||w||_{L^2(\Omega)} = 0 \iff w = 0$. For $w = 0$ to be true, $u_1 = u_2$, which means that there is only one unique solution.

4 Best Approximation

Question: Prove that the Galerkin solution u_h is the best approximation to $u \in X$ when measured in the energy norm $||u||_E$, that is

$$||u - u_h||_E \leq ||u - v_h||_E \qquad \forall v_h \in X_h \tag{38}$$

where $||u||_E^2 := \int_\Omega \nabla u \cdot \nabla u + \int_{\partial\Omega_N} u^2 ds$

4.1 Galerkin Orthogonality

Consider equations (14) and (17). Pick an arbitrary function from the test space z_h and set $v = v_h = z_h$. We then perform (14)-(17) to get

$$\begin{aligned} \int_\Omega \nabla u \cdot \nabla z_h + \int_{\partial\Omega_N} u z_h ds - \int_\Omega \nabla u_h \cdot \nabla z_h - \int_{\partial\Omega_N} u_h z_h ds = 0 \\ \int_\Omega \nabla(u - u_h) \cdot \nabla z_h + \int_{\partial\Omega_N} (u - u_h) z_h ds = 0, \qquad \forall z_h \in X_h \end{aligned} \tag{39}$$

This result will be useful in proving the best approximation

4.2 Best Approximation in the Energy Norm

From the formulation of the question, we have

$$||u||_E^2 := \int_\Omega \nabla u \cdot \nabla u + \int_{\partial\Omega_N} u^2 ds \tag{40}$$

or

$$||u||_E = \left\{ \int_\Omega \nabla u \cdot \nabla u + \int_{\partial\Omega_N} u^2 ds \right\}^{\frac{1}{2}} \tag{41}$$

What we want to show is that the difference between the general solution u and the approximate solution u_h is smaller than any other function in the test space (shown in equation (32)).

From (40) we can say that

$$\begin{aligned} ||(u - u_h)||_E^2 &= \int_\Omega \nabla(u - u_h) \cdot \nabla(u - u_h) + \int_{\partial\Omega_N} (u - u_h)^2 ds \\ &= \int_\Omega \nabla(u - u_h) \cdot \nabla(u - u_h) + \int_{\partial\Omega_N} (u - u_h)(u - u_h) ds \end{aligned} \tag{42}$$

Since $v_h - v_h = 0$, we can say that

$$\begin{aligned} ||(u - u_h)||_E^2 &= \int_\Omega \nabla(u - u_h) \cdot \nabla(u - u_h + v_h - v_h) + \int_{\partial\Omega_N} (u - u_h)(u - u_h + v_h - v_h) ds \\ &= \int_\Omega \nabla(u - u_h) \cdot \nabla(u - v_h) + \int_\Omega \nabla(u - u_h) \cdot \nabla(v_h - u_h) \\ &+ \int_{\partial\Omega_N} (u - u_h)(u - v_h) ds + \int_{\partial\Omega_N} (u - u_h)(u_h - v_h) ds \end{aligned} \tag{43}$$

Since $u_h - v_h \in X^h$ (because it is just a linear combination of functions in the test space, and the subtraction allows them to still meet the conditions to be in the test space) we can appeal to orthogonality in (39) to get

$$||(u - u_h)||_E^2 = \int_\Omega \nabla(u - u_h) \cdot \nabla(u - v_h) + \int_{\partial\Omega_N} (u - u_h)(u - v_h) ds \tag{44}$$

We can use Cauchy-Schwartz to say that (44) is

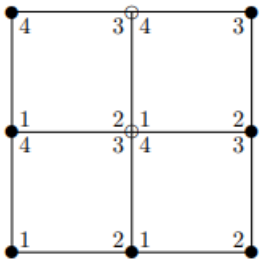
$$\begin{aligned} &\leq \left(||\nabla(u - u_h)||_{L^2(\Omega)}^2 + ||(u - u_h)||_{L^2(\partial\Omega_N)}^2 \right) \left(||\nabla(u - v_h)||_{L^2(\Omega)}^2 + ||(u - v_h)||_{L^2(\partial\Omega_N)}^2 \right) \\ &= \left\{ \int_\Omega \nabla(u - u_h) \cdot \nabla(u - u_h) + \int_{\partial\Omega_N} (u - u_h)^2 ds \right\} \\ &\cdot \left\{ \int_\Omega \nabla(u - v_h) \cdot \nabla(u - v_h) + \int_{\partial\Omega_N} (u - v_h)^2 ds \right\} \\ &= ||(u - u_h)||_E ||(u - v_h)||_E \end{aligned} \tag{45}$$

Since $||(u-u_h)||_E^2 \leq ||(u-u_h)||_E||(u-v_h)||_E$, it therefore shows that

$$||(u-u_h)||_E \leq ||(u-v_h)||_E \qquad \forall v_h \in X_h \tag{46}$$

5 Square Domain

Suppose that Ω is the square domain $(-1,1)\times(-1,1)$ and that $\partial\Omega_N$ is the top boundary: that is that $y=1$ with $-1<x<1$. Suppose further that $u_h \in X_h$ is the piece wise bilinear approximation to u satisfying (3) that is associated with the uniform grid of square elements with $h=1$ shown below



5.1 General Square Element

Question: Consider the general square element $\square_k$, with edge length h with nodal basis functions numbered anticlockwise (as shown). Show that the Jacobian matrix associated with the mapping to $\square_k$ from $\square_* = [-1,1]\times[-1,1]$ is the diagonal matrix

$$J_k = \frac{1}{2}\begin{bmatrix} h & 0 \\ 0 & h \end{bmatrix} \tag{47}$$

By mapping the derivatives of the reference element basis functions, show that the 4×4 element matrix associated with the negative Laplacian operator is given by

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} & \frac{2}{3} \end{bmatrix} \tag{48}$$

5.2 Mapping

To create the Jacobian, we first need to describe what the mapping process will describe.

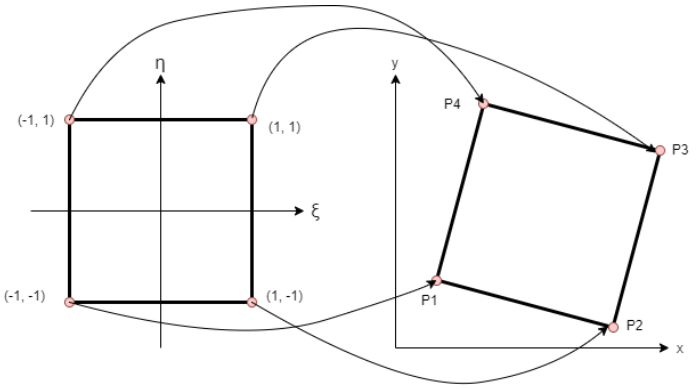


Figure 2: Picture that shows the process of mapping $\square_*$ onto an element in the grid $\square_k$

We therefore need to set up a system that defines the mapping described in figure 2, where each P_n is a vertex with (x_n, y_n) coordinates. This mapping must be defined for all points inside $\square_k$.

Here we will say that the square element $\square_k$ exists in (x, y) space, and that $\square_*$ exists in (ξ, η) space. We there fore define a set of functions

$$\begin{aligned}
\chi_1(\xi,\eta) &= \frac{(\xi-1)(\eta-1)}{4} \\
\chi_2(\xi,\eta) &= \frac{-(\xi+1)(\eta-1)}{4} \\
\chi_3(\xi,\eta) &= \frac{(\xi+1)(\eta+1)}{4} \\
\chi_4(\xi,\eta) &= \frac{-(\xi-1)(\eta+1)}{4}
\end{aligned} \tag{49}$$

We can then define the mapping function such that

$$\begin{aligned}
x(\xi,\eta) &= \sum_{n=1}^{4} x_n \chi_n \\
y(\xi,\eta) &= \sum_{n=1}^{4} y_n \chi_n
\end{aligned} \tag{50}$$

5.3 Jacobian

We then define the Jacobian to be

$$J_k = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} \tag{51}$$

In the previous section we defined both x and y as functions of ξ and η. Therefore we can differentiate each function by the respective variable to find the partial derivatives.

Starting with x with respect to ξ, we get

$$\begin{aligned}
x &= \frac{1}{4}\big\{x_1(\xi-1)(\eta-1) - x_2(\xi+1)(\eta-1) + x_3(\xi+1)(\eta+1) - x_4(\xi-1)(\eta+1)\big\} \\
\frac{\partial x}{\partial \xi} &= \frac{1}{4}\big\{\eta(x_1 - x_2 + x_3 - x_4) - x_1 + x_2 + x_3 - x_4\big\}
\end{aligned} \tag{52}$$

A crucial thing to note at this stage is the formulation of the grid we are mapping onto. We can see that the points are labelled clockwise, and from the definitions we know that each line is parallel with

the x or y axis. This means that

$$x_1 = x_4 \qquad \text{and} \qquad x_2 = x_3 \tag{53}$$

and so

$$\begin{aligned} \frac{\partial x}{\partial \xi} &= \frac{1}{4}\big\{ - x_1 + x_2 + x_2 - x_1\big\} \\ &= \frac{1}{2}\big\{x_2 - x_1\big\} \end{aligned} \tag{54}$$

But if the element is a square, then $x_2 - x_1$ is the height of the square h, so

$$\frac{\partial x}{\partial \xi} = \frac{1}{2}h \tag{55}$$

Following similar logic for y with respect to ξ

$$\begin{aligned} y &= \frac{1}{4}\big\{y_1(\xi-1)(\eta-1) - y_2(\xi+1)(\eta-1) + y_3(\xi+1)(\eta+1) - y_4(\xi-1)(\eta+1)\big\} \\ \frac{\partial y}{\partial \xi} &= \frac{1}{4}\big\{\eta(y_1 - y_2 + y_3 - y_4) - y_1 + y_2 + y_3 - y_4\big\} \end{aligned} \tag{56}$$

However, this time $y_1 = y_2$ and $y_3 = y_4$, and so

$$\frac{\partial y}{\partial \xi} = 0 \tag{57}$$

Again, the same logic shows that for x with respect to η

$$\begin{aligned} x &= \frac{1}{4}\big\{x_1(\xi-1)(\eta-1) - x_2(\xi+1)(\eta-1) + x_3(\xi+1)(\eta+1) - x_4(\xi-1)(\eta+1)\big\} \\ \frac{\partial x}{\partial \eta} &= \frac{1}{4}\big\{\xi(x_1 - x_2 + x_3 - x_4) - x_1 - x_2 + x_3 + x_4\big\} \\ &= 0 \end{aligned} \tag{58}$$

And, finally, taking y with respect to η and knowning that $y_4 - y_1 = h$ yields

$$\begin{aligned} \frac{\partial y}{\partial \eta} &= \frac{1}{4}\big\{\xi(y_1 - y_2 + y_3 - y_4) - y_1 - y_2 + y_3 + y_4\big\} \\ &= \frac{1}{2}\big\{y_4 - y_1\big\} \\ &= \frac{1}{2}h \end{aligned} \tag{59}$$

This proves that

$$J_k = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}h & 0 \\ 0 & \frac{1}{2}h \end{bmatrix} = \frac{1}{2}\begin{bmatrix} h & 0 \\ 0 & h \end{bmatrix} \tag{60}$$

5.4 Negative Laplacian Operator

To begin we need to solve the following linear system

$$\begin{bmatrix} 1 \\ x \\ y \\ xy \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ x_1y_1 & x_2y_2 & x_3y_3 & x_4y_4 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{bmatrix} \tag{61}$$

so that the contribution to the Gelerkin matrix can be calculated. We will refer to this linear system as $\mathbf{x} = B\mathbf{l}$. To do this, we need to emply Cramer's rule to find the values of L_i. Cramers rule states that

$$L_i = \frac{\det(B|i)}{\det(B)} \tag{62}$$

where $\det(B|i)$ is the determinant of the B matrix with the ith column replaced with $\mathbf{x}$. After the $\mathbf{l}$ vector has been determined, we can calculate the contribution matrix A where

$$A_{ij} = \int_\Omega \nabla L_i \cdot \nabla L_j \tag{63}$$

Since this matrix is 4×4, this is a little cumbersome. Luckily, computers can come to the rescue, and the code for this section is kept in appendix A. However, I will walk through one entry in A to show understanding.

First, there are a few things to notice from the formation of the grid. Since we are using square elements, where each square element as side length 1 and each boundary is parallel to an axis, this means that

$$\begin{array}{llllll} x_1 = x_1 & x_2 = x_1 + 1 & x_3 = x_1 + 1 & x_4 = x_1 \\ y_1 = y_1 & y_2 = y_1 & y_3 = y_1 + 1 & y_4 = y_1 + 1 \end{array} \tag{64}$$

From substituting this into (55), using Cramer's rule from (56) gives

$$\begin{aligned} L_1 &= x_1 - x - y + y_1 + xy - xy_1 - x_1 y + x_1 y_1 + 1 \\ L_2 &= x - x_1 - xy + xy_1 + x_1 y - x_1 y_1 \end{aligned} \tag{65}$$

which, when partially differentiated, gives

$$\begin{array}{ll} \dfrac{\partial L_1}{\partial x} = y - y_1 - 1 & \dfrac{\partial L_1}{\partial y} = x - x_1 - 1 \\ \dfrac{\partial L_2}{\partial x} = y_1 - y + 1 & \dfrac{\partial L_2}{\partial y} = x_1 - x \end{array} \tag{66}$$

So, to get $A_{1,2}$ we just need to use (63). However, we need to define our integral. Since we are integrating over a square element, we can split the domain integral into two sections - a y integral and an x integral. And we need to integrate between the width and the height of the element. Therefore (63) becomes

$$\begin{aligned} A_{1,2} &= \int_{y_1}^{y_1+1} \int_{x_1}^{x_1+1} \left\{ \frac{\partial L_1}{\partial x}\frac{\partial L_2}{\partial x} + \frac{\partial L_1}{\partial y}\frac{\partial L_2}{\partial y} \right\} dxdy \\ &= \int_{y_1}^{y_1+1} \int_{x_1}^{x_1+1} \left\{ (x - x_1)(x_1 - x + 1) - (y_1 - y + 1)^2 \right\} dxdy \\ &= \int_{y_1}^{y_1+1} \left\{ -y^2 + 2yy_1 + 2y - y_1^2 - 2y_1 - \frac{5}{6} \right\} dy \\ &= -\frac{1}{6} \end{aligned} \tag{67}$$

Following this algorithm for all entries in A gives

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} & \frac{2}{3} \end{bmatrix} \tag{68}$$

Due to the construction of each element, clearly $A_{i,i} \geq 0$ and $A_{i,j} = A_{j,i}$. The code in the appendix calculates and displays the full matrix by running the above algorithm for all element entries.

6 Assembled Galerkin System

Since we have 2 unknown nodes in the domain, the system we will need to solve is 2×2. Label the two nodes as u_1 being the interior node and u_2 being the node on the boundary. We need to find the contribution from each element for each node to create the system.

To begin with we need to calculate $\mathbf{f}$, which is

$$f_i = \int_{\partial\Omega_N} \phi_i ds \tag{69}$$

f_1 therefore is very simple. The node is not on the Robin boundary, and so $f_1 = 0$. However, u_2 is on the boundary, and so we need to find the integral of the function in this case. At the node, $\phi = 1$ and at the other two nodes $\phi = 0$. We also know that the total length of the boundary is 2, as each element has length 1. So the integral, visually, looks something like

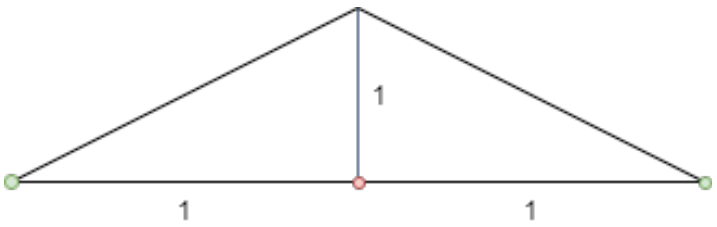


Figure 3: Image of the values that ϕ can take over the Robin boundary. Green nodes are Dirichlet, and red are Robin

This means that $f_2 = \frac{1}{2}2 \times 1 = 1$ (the area of the triangle), and so

$$\mathbf{f} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{70}$$

To calculate the contributions, we need to observe how each element contributes to each node. For u_1 we have four contributions, one from each element. They are

$$A_{33} + A_{44} + A_{11} + A_{22} = \frac{2}{3} + \frac{2}{3} + \frac{2}{3} + \frac{2}{3} = \frac{8}{3} \tag{71}$$

For the edge between u_1 and u_2 we only have contributions from two elements, the top two squares.

$$A_{14} + A_{23} = -\frac{1}{6} - \frac{1}{6} = -\frac{1}{3} \tag{72}$$

At u_2 we only have entries, again, from two square elements, and so

$$A_{44} + A_{33} = \frac{2}{3} + \frac{2}{3} = \frac{4}{3} \tag{73}$$

This gives the Galerkin system to be

$$\begin{bmatrix} \frac{8}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{74}$$

We can then easily solve this system with linear equations, so

$$\begin{aligned} \frac{8}{3}u_1 - \frac{1}{3}u_2 &= 0 \\ 8u_1 &= u_2 \end{aligned} \tag{75}$$

From the other equation

$$\begin{aligned} \frac{4}{3}u_2 - \frac{1}{3}u_1 &= 1 \\ \frac{4}{3}8u_1 - \frac{1}{3}u_1 &= 1 \\ u_1 &= \frac{3}{31} \end{aligned} \tag{76}$$

Giving $u_2 = \frac{24}{31}$. Therefore the entire Gelerkin approximation is

$$u_h = \frac{3}{31}\phi_1 + \frac{24}{31}\phi_2 \tag{77}$$

which is defined for all x and y in the domain. Each of the ϕ functions are 1 at the node they correspond with, 0 at the other nodes. This matches the definition set in equation (18), as we have $n = 2$ (one interior node, one Robin boundary node).

7 Sobolev Space

Question: Show that $u(r,\theta) = r^{\frac{2}{3}} sin((2\theta + \pi/3))$ satisfies Laplace's equation expressed in polar coordinates

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0 \tag{78}$$

Given that $u \in \mathcal{H}^s(\Omega) \iff ||D^s u|| < \infty$, show that $u(r,\theta)$ defined on the pie shaped domain Ω where $0 \leq r \leq 1$ and $-\pi/2 \leq \theta \leq \pi$ is in $\mathcal{H}^1(\Omega)$, but is not in $\mathcal{H}^2(\Omega)$.

7.1 Satisfaction of Laplace in Polar Coordinates

If

$$u(r,\theta) = r^{\frac{2}{3}} sin((\frac{2\theta + \pi}{3})) \tag{79}$$

then

$$\frac{\partial u}{\partial r} = \frac{2sin(\frac{2\theta+\pi}{3})}{3r^{\frac{1}{3}}} \tag{80}$$

$$\frac{\partial^2 u}{\partial r^2} = -\frac{2sin(\frac{2\theta+\pi}{3})}{9r^{\frac{4}{3}}} \tag{81}$$

$$\frac{\partial^2 u}{\partial \theta^2} = -\frac{4r^{\frac{2}{3}} sin(\frac{2\theta+\pi}{3})}{9} \tag{82}$$

This means that

$$\begin{aligned} \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} &= \frac{-2sin(\frac{2\theta+\pi}{3})}{9r^{\frac{4}{3}}} + \frac{1}{r}\frac{2sin(\frac{2\theta+\pi}{3})}{3r^{\frac{1}{3}}} - \frac{1}{r^2}\frac{4r^{\frac{2}{3}} sin(\frac{2\theta+\pi}{3})}{9} \\ &= 2sin(\frac{2\theta+\pi}{3})\left\{ -\frac{1}{9r^{\frac{4}{3}}} + \frac{1}{3r^{\frac{4}{3}}} - \frac{2}{9r^{\frac{4}{3}}} \right\} \\ &= 2sin(\frac{2\theta+\pi}{3})\left\{ \frac{3}{9r^{\frac{4}{3}}} - \frac{3}{9r^{\frac{4}{3}}} \right\} \\ &= 0 \end{aligned} \tag{83}$$

Therefore, the Laplace equation is satisfied.

7.2 Member of kth Sobolev Space

For a function to be a member of $\mathcal{H}^k$ the kth differential must be square integrable over the domain. In otherwords

$$u \in \mathcal{H}^k(\Omega) \iff \int_\Omega u^2 < \infty, \int_\Omega (u')^2 < \infty, ..., \int_\Omega (u^{(k)})^2 < \infty \tag{84}$$

Therefore we will also need

$$\frac{\partial u}{\partial \theta} = \frac{2r^{\frac{2}{3}} cos(\frac{2\theta+\pi}{3})}{3} \tag{85}$$

and finally

$$\frac{\partial^2 u}{\partial \theta \partial r} = \frac{4r cos(\frac{2\theta+\pi}{3})}{9r^{\frac{1}{3}}} \tag{86}$$

7.2.1 $\mathcal{H}^1(\Omega)$

If $u \in \mathcal{H}^1(\Omega)$ then

$$\int_\Omega \left(\frac{\partial u}{\partial \theta} + \frac{\partial u}{\partial r} \right)^2 < \infty \tag{87}$$

Substitute in the equation and boundary conditions to get

$$\int_0^1 \int_{-\frac{\pi}{2}}^{\pi} \left(\frac{2r^{\frac{2}{3}} cos(\frac{2\theta+\pi}{3})}{3} + \frac{2sin(\frac{2\theta+\pi}{3})}{3r^{\frac{1}{3}}} \right)^2 r d\theta dr \tag{88}$$

Solve the first inner integral to get

$$\int_0^1 \frac{13\pi r^{7/3}}{12} dr \tag{89}$$

Next, solve the outer integral to get

$$\frac{13\pi}{40} \tag{90}$$

Clearly this answer is less than ∞, and therefore the function is a member of $\mathcal{H}^1(\Omega)$.

7.2.2 $\mathcal{H}^2(\Omega)$

If $u \in \mathcal{H}^2(\Omega)$ then

$$\int_\Omega \left(\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial r^2} + \frac{\partial^2 u}{\partial \theta \partial r} \right)^2 < \infty \tag{91}$$

Substitute in the equation and boundary conditions to get

$$\int_0^1 \int_{-\frac{\pi}{2}}^{\pi} \left(-\frac{4r^{\frac{2}{3}} sin(\frac{2\theta+\pi}{3})}{9} - \frac{2sin(\frac{2\theta+\pi}{3})}{9r^{\frac{4}{3}}} + \frac{4r cos(\frac{2\theta+\pi}{3})}{9r^{\frac{1}{3}}} \right)^2 r d\theta dr \tag{92}$$

Solve the first inner integral to get

$$\int_0^1 \frac{\pi(8r^4 + 4r^2 + 1)}{27r^{\frac{5}{3}}} dr \tag{93}$$

which gives the sum

$$\left[\frac{\pi(8r^4 + 10r^2 - 5)}{90r^{\frac{2}{3}}} \right]_0^1 \tag{94}$$

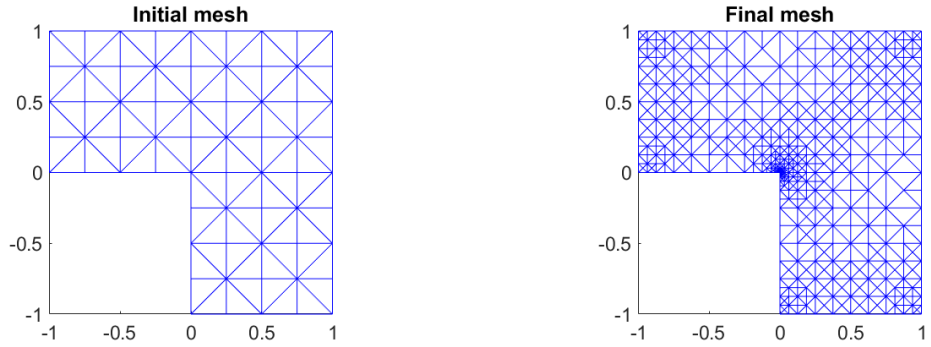
Since one of the bounds is 0 in the above sum one of the terms will be divided by 0, which gives ∞. Therefore this function is not square integrable in the described domain, and is not a member of $\mathcal{H}^2(\Omega)$.

8 Computational Exercises

8.1 Adaptive refinement and Errors

Question: The aim of this exercise is to assess the effectiveness of the adaptive refinement strategy that is built into T-IFISS by looking at a problem with a singular solution.

Running the **Run8DigitChallenge** solves a problem on an L-shaped domain with specified boundary conditions. The adaptive mesh routine ensures that we minimise the global error of the system in an efficient way. Specifically, it solves the problem over a uniform grid, and estimates the error for each element. Elements with a high error are "refined", meaning that the mesh in this area is made to be finer. The problem is solved again, with this process repeating until some tolerance has been reached. The aim is to ensure that the local error of each element is roughly equal.



(a) Initial mesh

(b) Mesh after 15 iterations

Figure 4: Difference between initial and final mesh. We can clearly see that the inside corner is an area of complexity

Running the algorithm with a tolerance set to $\epsilon = 0.005$ had the code converge in 15 iterations, with 1901 vertices.

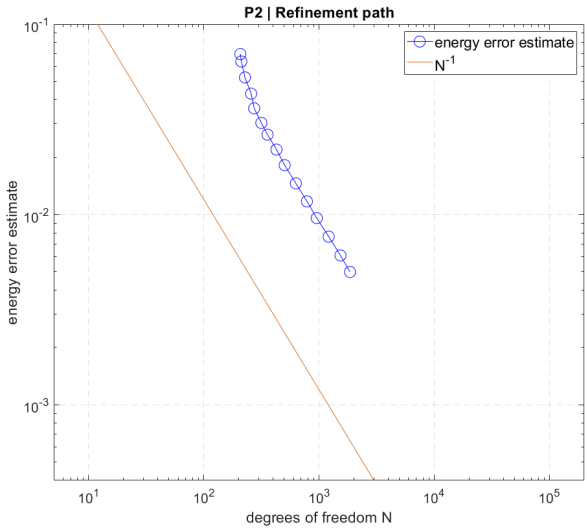


Figure 5: Plot to show the energy error estimate dropping every iteration

Clearly, as the degrees of freedom increase, the energy norm reduces by order $1/N$. We would like to compare the conversion rate of a linear approximation against a quadratic approximation.

Grid Parameter	number of vertices	Linear $\|\|e_{ref}\|\|_E$	Quadratic $\|\|e_{ref}\|\|_E$
4	225	1.599×10^{-1}	3.8918×10^{-2}
5	833	8.971×10^{-2}	2.3848×10^{-2}
6	3201	5.067×10^{-2}	1.4941×10^{-2}
7	12545	2.9108×10^{-2}	9.4021×10^{-3}
8	49665	1.7072×10^{-2}	N/A

Table 1: Table to show how uncreasing the elements in a uniform grid with different interpolations decreases the energy error. N/A means that matlab kept freezing!

Taking a uniform grid and increasing the mesh density decreases the error energy norm, as shown in table 1. We can see a direct comparison by graphing the convergence.

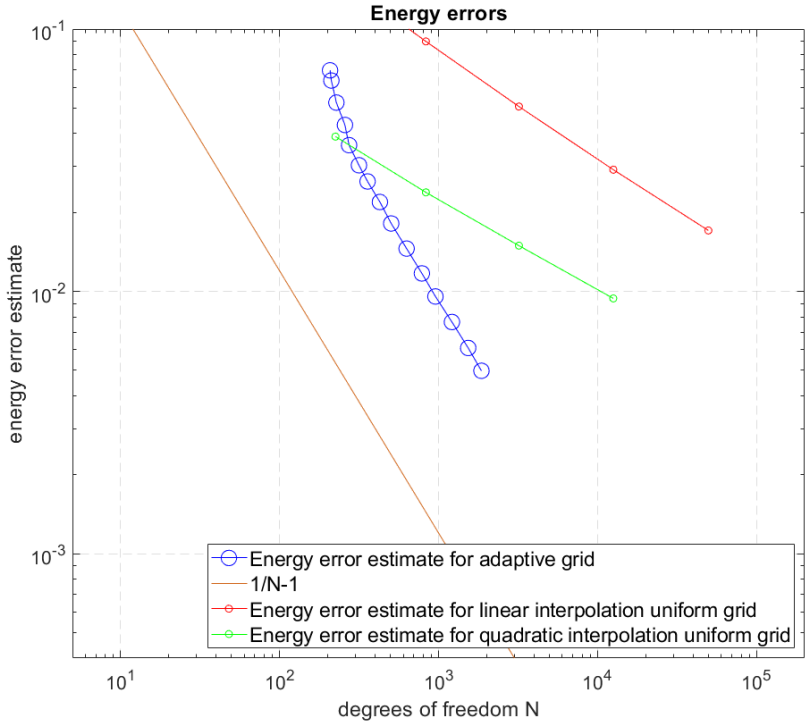


Figure 6: Plot to show the enery error estimate dropping every iteration

Clearly the adaptive grid method converges more quickly, giving a lower error for fewer grids. This is intuitive, as the density of elements is highest in areas of high error, so the algorithm focuses on reducing the error in these areas. The quadratic interpolation gives a lower energy error than the linear interpolation. However, they converge at the same rate, $\sim N^{-\frac{1}{2}}$. This is because the weak solution of u is not smooth enough to satisfy

1.
$$||u - u_h||_E \leq C_1 h ||f||_{L^2(\Omega)} \tag{95}$$
where C_1 is a constant and h is the largest side of an element in the mesh

2.

$$||u - u_h||_E \leq C_2 h^2 \tag{96}$$

where C_2 is a constant and h is the largest side of an element in the mesh

which are the priori error estimates for linear and quadratic interpolation respectively. If the weak solution was smooth enough, we would expect to see $u_h \to u$ as the number of elements $\to \infty$.

8.2 Eigenvalues

Question: Write a MATLAB function that calls eigs and generates an estimate of the three smallest eigenvalues and associated eigenfunctions of the negative Laplacian operator on the L-shaped domain featured in the previous exercise.

To solve this problem, we need to solve the following system

$$A\mathbf{x} = \lambda_h M\mathbf{x} \tag{97}$$

where A is the stiffness matrix, M is the mass matrix, and λ_h are the eigenvalues. The matrices are calculated when running the L-shaped plate problem for various meshes using the *ell_adiff* command. Running the command *[eigvec, eigval] = eigs(A, M, 3, 'smallestabs')* gives the first 3 eigen vectors and corresponding eigen values. We can estimate these values over finer and finer meshes to see an approximation of the true eigen values of the PDE.

Grid Parameter	number of vertices	λ_1	λ_2	λ_3
2	21	5.033×10^{-15}	1.6585	3.7633
3	65	2.0132×10^{-14}	1.5389	3.5961
4	225	8.3785×10^{-14}	1.4991	3.5501
5	833	3.3092×10^{-13}	1.4846	3.5381
6	3201	1.2963×10^{-12}	1.4791	3.5351
7	12545	5.2184×10^{-12}	1.4770	3.5343
8	49665	2.0766×10^{-11}	1.4762	3.5341

Table 2: Table of estimated eigenvalues

We can also plot the eigen vectors of the problem to see interesting behaviours.

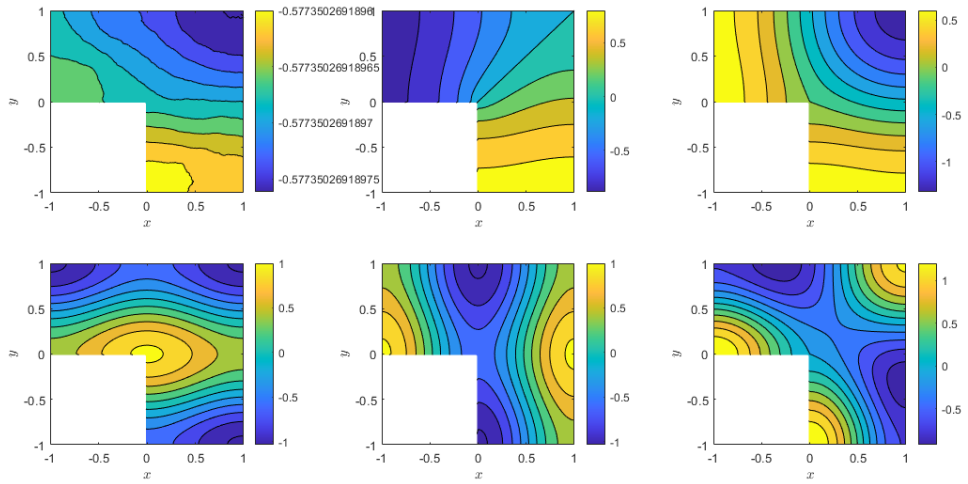


Figure 7: Contour plot to show behaviour for the smallest 6 eigen values

The first eigenvalue is incredibly small, and is testing the limitations of my computer! However, the interior corner of the L-shaped domain seems to be an area that has some interesting behaviour. It seems to be an area of "change" - sometimes the eigenfunction changes from positive to negative, sometimes it changes from a positive gradient to a negative gradient.

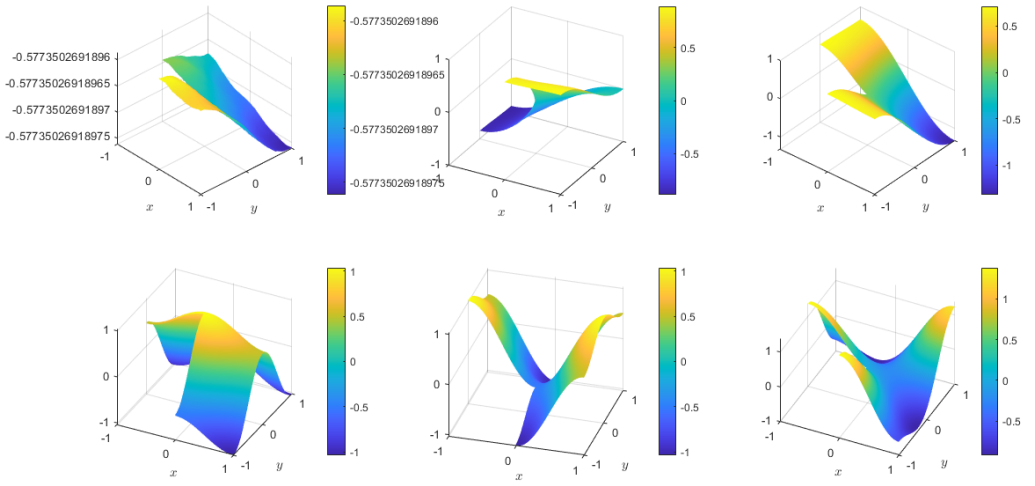


Figure 8: Surface plot to show behaviour for the smallest 6 eigen values

A Contribution Matrix Routine

```
% Set up symbols and Ab=c linear system
syms y1 x1 L1 L2 L3 L4 x y

A = [1, 1, 1, 1;
    x1, (x1+1), (x1+1), x1;
    y1, y1, (y1+1), (y1+1);
    y1*x1, y1*(x1+1), (y1+1)*(x1+1), (y1+1)*x1];

c = [1; x; y; x*y];

b = sym('a',[1 4]);

% Calculate determinant of A
Adet = det(A);

% Use Cramers rule to find values of Li

for n=1:4
    B=A;
    B(:,n)=c;
    b(n)=det(B)/Adet;
end

% find partial derivatives of Li
bdiff = sym('a', [4 2]);
for n=1:4
    bdiff(n, 1) = diff(b(n), x);
    bdiff(n, 2) = diff(b(n), y);
end

% Calculate stiffness matrix
stiff = zeros(4, 4);
for i=1:4
    for j=1:4
        eq = bdiff(i, 1) *bdiff(j, 1) + bdiff(i, 2) * bdiff(j, 2);
        int1 = int(eq, x, x1, (x1+1));
        int2 = int(int1, y, y1, (y1+1));
        stiff(i, j) = int2;
    end
end

stiff
```

B Bilinear Interpolation Code for Single Element

```
x1=0; x2=1; y1=0; y2=1;
z11=0; z12=1; z21=1; z22=0.5;

phi = @(x, y) (1/(x2-x1))*[x2 - x, x- x1]*[z11, z12; z21, z22]*[y2 - y; y-y1];

x = [0:0.01:1];
y = [0:0.01:1];

z = zeros(length(x), length(y));

for i=1:length(x)
    for j=1:length(y)
        z(i, j) = phi(x(i), y(j));
    end
end

```

```
[X, Y] = meshgrid(x, y);
surf(X, Y, z);
shading interp
view(2)
```

C Energy error plot

```
verts = [225, 833, 3201, 12545, 49665];
lin_e = [1.599e-1, 8.971e-2, 5.067e-2, 2.9108e-2, 1.7072e-2];
lin_q = [3.8918e-2, 2.3848e-2, 1.4941e-2, 9.4021e-3, NaN]
openfig('refinementpath.fig')
hold on
x=[0:1:10e5];
loglog(verts, lin_e, 'r-o')
hold on
loglog(verts, lin_q, 'g-o')

title('Energy errors')
legend('Energy error estimate for adaptive grid', '1/N-1', ...
    'Energy error estimate for linear interpolation uniform grid', ...
    'Energy error estimate for quadratic interpolation uniform grid', ...
    'Location', 'Southeast')
```

D Eigenfunction plot

```
%% run problem solver
ell_adiff;
%% get eigen values and vectors
[eigvec, eigval]  = eigs(A, M, 5, 'smallestabs');
eigval(1, 1)
eigval(2, 2)
eigval(3, 3)
length(x_gal)


%% plot eigen functions
xv = min(xy(:, 1)):0.01:max(xy(:, 1))
yv = min(xy(:, 2)):0.01:max(xy(:, 2))
[X, Y] = meshgrid(xv, yv);
for n=1:3
    figure(1)
    subplot(3, 1, n)
    u = griddata(xy(:, 1), xy(:, 2), eigvec(:, n), X, Y, 'cubic');
    u(1:99, 1:99) = NaN;
    contourf(X, Y, u)
    xlabel('$x$', 'interpreter', 'Latex')
    ylabel('$y$', 'interpreter', 'Latex')
    colorbar
end
subplot(3, 1, 1)
title("Eigenfunction plots for first 3 Eigenvalues")
```