

# Numerics Exercise 3 - Hannah Franklin

## Ex 25-3 FEM Sheet (Poisson):

Poisson system:  $\begin{cases} -\nabla^2 u = f \text{ on } (x,y) \in [0,1]^2 \\ f(x,y) = 2\pi^2 \sin(\pi x) \cos(\pi y) \\ u(0,y) = u(1,y) = 0 \\ \frac{\partial u}{\partial y}|_{y=0} = \frac{\partial u}{\partial y}|_{y=1} = 0 \end{cases}$

Claim: exact solution is  $u_e(x,y) = \sin(\pi x) \cos(\pi y)$

$$-\nabla^2 u = -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2}$$

$$\begin{aligned} -\nabla^2 u_e &= \pi^2 \sin(\pi x) \cos(\pi y) + \pi^2 \sin(\pi x) \cos(\pi y) \\ &= 2\pi^2 \sin(\pi x) \cos(\pi y) \\ &= f \quad \text{as required.} \end{aligned}$$

$$\begin{aligned} u_e(0,y) &= \sin(0) \cos(\pi y) = 0 \\ u_e(1,y) &= \sin(\pi) \cos(\pi y) = 0 \\ \frac{\partial u_e}{\partial y}|_{y=0} &= -\pi \sin(\pi x) \sin(0) = 0 \\ \frac{\partial u_e}{\partial y}|_{y=1} &= -\pi \sin(\pi x) \sin(\pi) = 0 \end{aligned} \quad \left. \right\} \therefore \text{satisfies required BCs}$$

Hence  $u_e$  is exact soln of Poisson system (1).

Step 1: Ritz-Galerkin principle for (1).

$$\Omega := [0,1]^2, \Gamma_1 := \{x=0,1\}, \Gamma_2 := \{y=0,1\}, \partial\Omega := \Gamma_1 \cup \Gamma_2$$

soln to system (1) minimises the functional:  $I[u] := \iint_{\Omega} (\frac{1}{2} |\nabla u|^2 - u f) d\Omega$

Riesz' method:  $\delta I := \frac{dI}{d\varepsilon} = \lim_{\varepsilon \rightarrow 0} \left( \frac{I[u + \varepsilon \delta u] - I[u]}{\varepsilon} \right) = 0 \left( (\delta u)(x,y) \right)$

$$\begin{aligned} I[u + \varepsilon \delta u] - I[u] &= \iint_{\Omega} \left( \frac{1}{2} |\nabla(u + \varepsilon \delta u)|^2 - (u + \varepsilon \delta u) f \right) d\Omega - \iint_{\Omega} \left( \frac{1}{2} |\nabla u|^2 - u f \right) d\Omega \\ &= \iint_{\Omega} \left( \frac{1}{2} |\nabla u + \varepsilon \nabla(\delta u)|^2 - (\delta u) f - \frac{1}{2} |\nabla u|^2 \right) d\Omega \\ &= \iint_{\Omega} \left( \frac{1}{2} |\nabla u|^2 + \varepsilon \nabla u \cdot \nabla(\delta u) + \frac{\varepsilon^2}{2} |\nabla(\delta u)|^2 - (\delta u) f - \frac{1}{2} |\nabla u|^2 \right) d\Omega \\ &= \iint_{\Omega} \left( \varepsilon \nabla u \cdot \nabla(\delta u) + \frac{\varepsilon^2}{2} |\nabla(\delta u)|^2 - (\delta u) f \right) d\Omega \end{aligned}$$

$$\Rightarrow \delta I = \lim_{\varepsilon \rightarrow 0} \left( \iint_{\Omega} [\nabla u \cdot \nabla(\delta u) + \frac{\varepsilon^2}{2} |\nabla(\delta u)|^2 - (\delta u) f] d\Omega \right) \\ = \iint_{\Omega} (\nabla u \cdot \nabla(\delta u) - \delta u f) d\Omega$$

$$\delta I = 0 \Rightarrow \iint_{\Omega} (\nabla u \cdot \nabla(\delta u) - \delta u f) d\Omega = 0 \rightarrow \text{weak formulation}$$

$$\text{IBPs} \Rightarrow -\iint_{\Omega} (\delta u \nabla^2 u + \delta u f) d\Omega + \int_{\partial\Omega} (\delta u \nabla u \cdot \vec{n}) dS = 0$$

(Integration by parts)  $-\iint_{\Omega} \delta u (\nabla^2 u + f) d\Omega + \int_{\Gamma_L} \delta u (\nabla u \cdot \vec{n}) dS + \int_{\Gamma_R} \delta u (\nabla u \cdot \vec{n}) dS = 0$

$$BC_2 \Rightarrow \nabla u \cdot \mathbf{g} = 0 \text{ on } \Gamma_2$$

$\Rightarrow$  require variation  $\delta u(x, y) = 0$  on  $\Gamma_1$

$\Rightarrow -\int_{\Omega} \delta u (\nabla^2 u + f) d\Omega = 0$  variations  $\delta u$  that satisfy  $\delta u = 0$  on  $\Gamma_1$ .

$\Rightarrow -\nabla^2 u - f = 0 \Rightarrow -\nabla^2 u = f$  on  $\Omega$  w/  $\nabla u \cdot \mathbf{g} = 0$  on  $\Gamma_2$

$\therefore$  yields system (1) w/ condition that variation  $\delta u = 0$  on  $\Gamma_1$ .  
 $(\delta u(0, y) = \delta u(1, y) = 0)$

weak formulation:

let  $\omega(x, y)$  be test func.

multiply system (1) by  $\omega$  and integrate over  $\Omega$ :

$$\int_{\Omega} \omega (-\nabla^2 u - f) d\Omega = 0$$

$$IBPs \Rightarrow \int_{\Omega} (\nabla u \cdot \nabla \omega - \omega f) d\Omega + \int_{\partial\Omega} \omega (\nabla u \cdot \mathbf{n}) dS = 0$$

$$\nabla u \cdot \mathbf{n} = 0 \text{ on } \Gamma_2 \Rightarrow \int_{\Omega} (\nabla u \cdot \nabla \omega - \omega f) d\Omega + \int_{\Gamma_1} \omega (\nabla u \cdot \mathbf{n}) dS = 0$$

$\Rightarrow$  test func  $\omega(x, y) = 0$  on  $\Gamma_1$ .

$\Rightarrow \int_{\Omega} (\nabla u \cdot \nabla \omega - \omega f) d\Omega = 0$  is weak formulation.  
w/ test func  $\omega(x, y) = 0$  on  $\Gamma_1$ .

this is the same form as obtained through Ritz' method w/ the same condition on test func  $\omega(x, y)$  as on variation  $\delta u(x, y)$ .

hence  $\delta u = \omega$ .

Step 2: algebraic or discrete Ritz-Galerkin principle.

$u(x, y) \approx u_h(x, y) = u_j \varphi_j(x, y)$  for  $\varphi_j(x, y)$  compact global basis funcs  
 $\omega(x, y) \approx \omega_h(x, y) = \omega_j \varphi_j(x, y)$

take care where  $\omega_j = \delta_{ij} \Rightarrow \omega_h(x, y) = \varphi_i(x, y)$  ( $\varphi_i(0, y) = \varphi_i(1, y) = 0$ )

weak formulation:  $\int_{\Omega} (\nabla u_h \cdot \nabla \omega_h - \omega_h f) d\Omega = 0$  this also guarantees BC on  $\Gamma_1$  for  $u_h$ .

substitute in to weak formulation  $\Rightarrow \int_{\Omega} (\nabla u_h \cdot \nabla \omega_h - \omega_h f) d\Omega = 0$   
( $u_j$  constants)  $\Rightarrow \int_{\Omega} (u_j \nabla \varphi_j \cdot \nabla \varphi_i - \varphi_i f) d\Omega = 0$

$$\Rightarrow u_j \int_{\Omega} (\nabla \varphi_j \cdot \nabla \varphi_i) d\Omega = \int_{\Omega} \varphi_i f d\Omega$$

$\Rightarrow$  weak formulation becomes  $A_{ij} u_j = b_i$

$$\text{where } A_{ij} := \int_{\Omega} (\nabla \varphi_j \cdot \nabla \varphi_i) d\Omega, b_i := \int_{\Omega} \varphi_i f d\Omega$$

where global basis func  $\varphi_i(x, y)$  is s.t.  $\varphi_i(0, y) = \varphi_i(1, y) = 0 \forall i$ .

now consider variation of  $u_h(x, y)$  w/ variation func  $\delta u_h(x, y)$

$$\delta I = \frac{dI}{d\epsilon} = \lim_{\epsilon \rightarrow 0} \left( \frac{I[u_h + \epsilon \delta u_h] - I[u_h]}{\epsilon} \right)$$

$$= \lim_{\epsilon \rightarrow 0} \left( \frac{1}{\epsilon} \left[ \iint_{\Omega} \left( \frac{1}{2} |\nabla(u_h + \epsilon \delta u_h)|^2 - (u_h + \epsilon \delta u_h) f \right) d\Omega \right. \right. \\ \left. \left. - \iint_{\Omega} \left( \frac{1}{2} |\nabla u_h|^2 - u_h f \right) d\Omega \right] \right)$$

$$= \lim_{\epsilon \rightarrow 0} \left( \frac{1}{\epsilon} \left[ \iint_{\Omega} \left( \epsilon \nabla u_h \cdot \nabla \delta u_h + \frac{\epsilon^2}{2} |\nabla \delta u_h|^2 - \epsilon \delta u_h f \right) d\Omega \right] \right)$$

$$= \lim_{\epsilon \rightarrow 0} \left( \iint_{\Omega} [\nabla u_h \cdot \nabla \delta u_h + \frac{\epsilon}{2} |\nabla \delta u_h|^2 - \delta u_h f] d\Omega \right)$$

$$= \iint_{\Omega} (\nabla u_h \cdot \nabla \delta u_h - \delta u_h f) d\Omega \quad \text{this is weak formulation}$$

require  $\delta u_h(0, y) = \delta u_h(1, y) = 0$  from step 1

provided  $\delta u_h = \omega_h$

Variation of discrete RG principle missing; i and i' indexing difference missing. -0.5

$$\Rightarrow \iint_{\Omega} (u_j \nabla \psi_j \cdot \nabla \psi_i - \psi_i f) d\Omega = \delta I = 0$$

$$\Rightarrow \iint_{\Omega} (\nabla \psi_j \cdot \nabla \psi_i) d\Omega u_j = \iint_{\Omega} \psi_i f d\Omega$$

$$\Rightarrow A_{ij} u_j = b_i \quad \text{w/ } A_{ij}, b_i \text{ as defined previously.}$$

hence the variation of  $u_h$  yields the algebraic eq. as required.

Step 3: local coordinate system and reference coordinates.

quadrilateral elements:  $N=4$  nodes  $\underline{x}_K, \alpha = \underline{\alpha} \in \alpha : 0, 1, 2, 3$

reference element  $\underline{\xi} = (\xi_1, \xi_2) \in (-1, 1)^2$  element  $K$

want to map  $(x, y) \mapsto (\xi_1, \xi_2)$

$\underline{x} \mapsto \underline{\xi}$  | Element  $K \mapsto$  reference element  $\underline{\xi}$

map coordinates of element  $K$  w/ shape funcs  $\chi_{\alpha}(\underline{\xi})$  s.t.:

$$\underline{x} = \sum_{\alpha=0}^3 [\underline{x}_{\alpha, \alpha} \chi_{\alpha}(\underline{\xi})], \text{ where } \underline{x}_{\alpha, \alpha} \text{ local node coords of element } K.$$

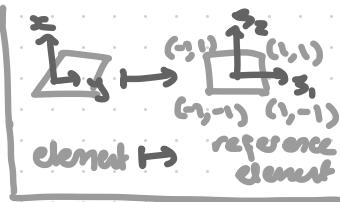
for quadrilaterals shape funcs:

$$\chi_0(\underline{\xi}) = \frac{1}{4} (1-\xi_1)(1-\xi_2)$$

$$\chi_1(\underline{\xi}) = \frac{1}{4} (1+\xi_1)(1-\xi_2)$$

$$\chi_2(\underline{\xi}) = \frac{1}{4} (1+\xi_1)(1+\xi_2)$$

$$\chi_3(\underline{\xi}) = \frac{1}{4} (1-\xi_1)(1+\xi_2)$$



These are also Galerkin basis funcs.

$$\text{Jacobian of transformation } \underline{x} \mapsto \underline{\xi} : J = \begin{pmatrix} \frac{\partial \underline{x}}{\partial \xi_1}, \frac{\partial \underline{x}}{\partial \xi_2} \\ \frac{\partial \underline{y}}{\partial \xi_1}, \frac{\partial \underline{y}}{\partial \xi_2} \end{pmatrix}$$

$$\underline{x} = \underline{x}_0 + \underline{x}_1 \xi_1 + \underline{x}_2 \xi_2 + \underline{x}_3 \xi_3$$

$$\Rightarrow J = \frac{1}{4} \begin{pmatrix} (1-\xi_2)(x_1-x_0) + (1+\xi_2)(x_2-x_0) & (1-\xi_1)(x_3-x_0) + (1+\xi_1)(x_1-x_0) \\ (1-\xi_2)(y_1-y_0) + (1+\xi_2)(y_2-y_0) & (1-\xi_1)(y_3-y_0) + (1+\xi_1)(y_1-y_0) \end{pmatrix}$$

from which  $\det J$  and  $(J^T)^{-1}$  can be found

$$\underline{\xi} = J \underline{x} \Rightarrow \underline{v} \mapsto (J^T)^{-1} \begin{pmatrix} \frac{\partial}{\partial \xi_1} \\ \frac{\partial}{\partial \xi_2} \end{pmatrix}$$

$$d\underline{\xi} = \frac{1}{|\det J|} d\underline{v} \Rightarrow d\underline{v} = |\det J| d\underline{\xi}$$

considering  $A_{ij} u_j = b_i$ ,  $A_{ij} = \int_{\Omega} (\nabla \underline{v}_j \cdot \nabla \underline{v}_i) d\underline{v}$ ,  $b_i = \int_{\Omega} \underline{f} \cdot \underline{v}_i d\underline{v}$   
 $\underline{x}_i$  are basis funcs  $\Rightarrow \hat{A}_{\alpha\beta} = \int_K \left[ (J^T)^T \left( \frac{\partial}{\partial \xi_1} \right) \underline{x}_\alpha \cdot (J^T)^T \left( \frac{\partial}{\partial \xi_2} \right) \underline{x}_\beta \right] |\det J| d\xi$

$$\hat{b}_\alpha = \int_K (\underline{f}(\underline{\xi}) \cdot \underline{x}_\alpha |\det J|) d\xi$$

for  $\alpha, \beta = 0, 1, 2, 3$  on each reference element  $K$ .

$\hat{A}_{\alpha\beta}$ ,  $\hat{b}_\alpha$  are elemental matrices and vectors.

global matrix A, global vector b

- Set all elements of A, b to be 0. ( $A_{ij} = b_i = 0 \forall i, j$ )
- loop over all elements K :  $K=0, N_e-1$  ( $N_e = \text{no. elements}$ )  
 for  $\alpha = 0, 3$  :

i = Index(K,  $\alpha$ )

for  $\beta = 0, 3$  :

j = Index(K,  $\beta$ )

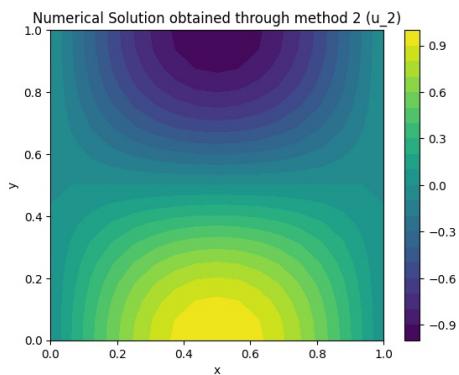
$A_{ij} = A_{ij} + \hat{A}_{\alpha\beta}$

$b_i = b_i + \hat{b}_\alpha$

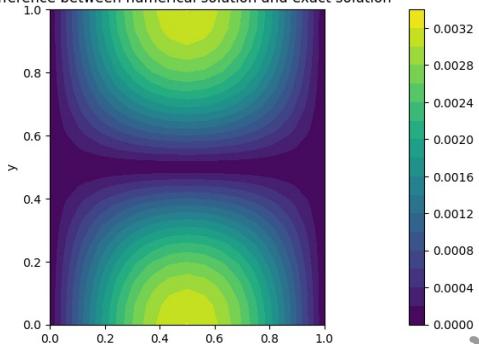
to define the global matrix using the elemental matrices.

#### 4. Step 4 : solve the system

$\{h, p\} = \{\frac{1}{16}, 1\}$  contours of solution  $u_2$  and  $|u_2 - u_{\text{ex}}|$ :

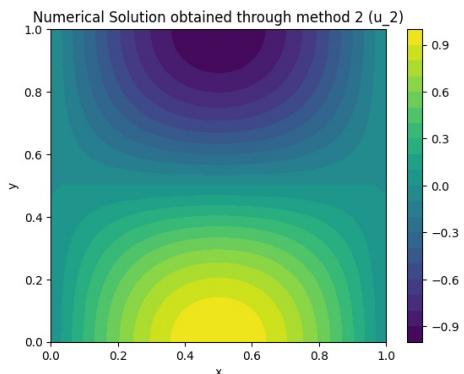


Difference between numerical solution and exact solution

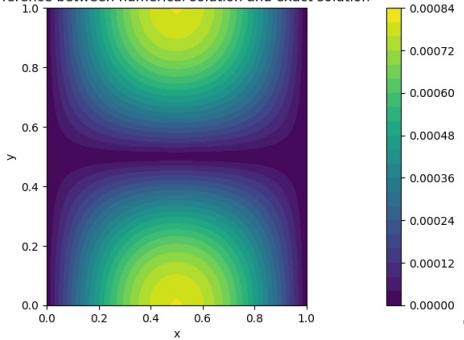


max error  $\div 4$

$\{h, p\} = \{\frac{1}{32}, 1\}$  contours:

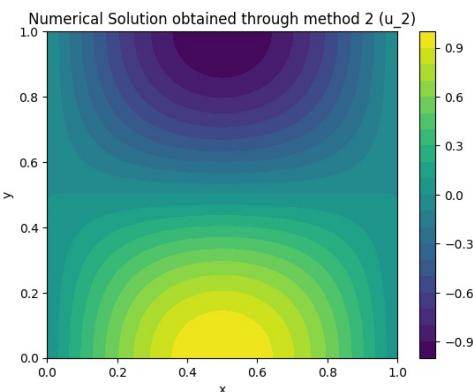


Difference between numerical solution and exact solution

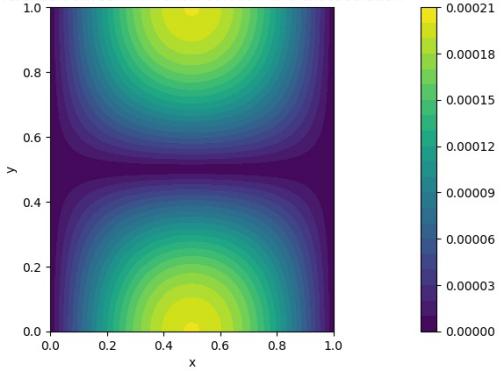


max error  $\div 4$

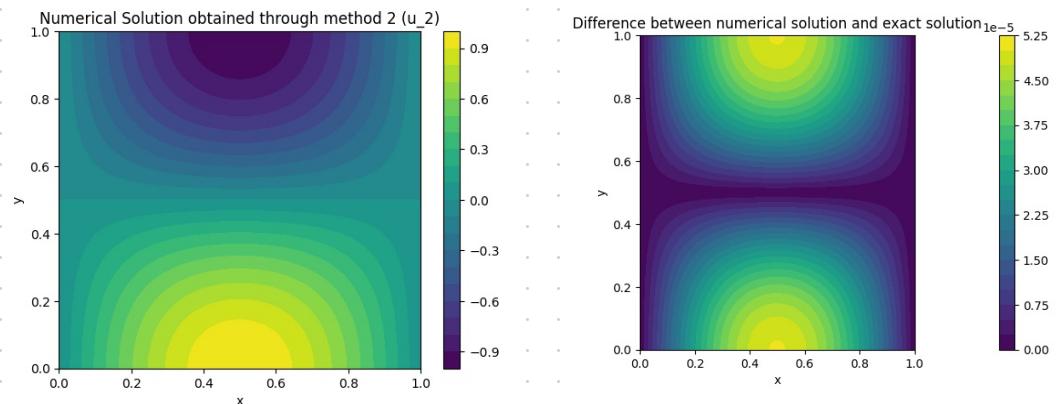
$\{h, p\} = \{\frac{1}{64}, 1\}$  contours:



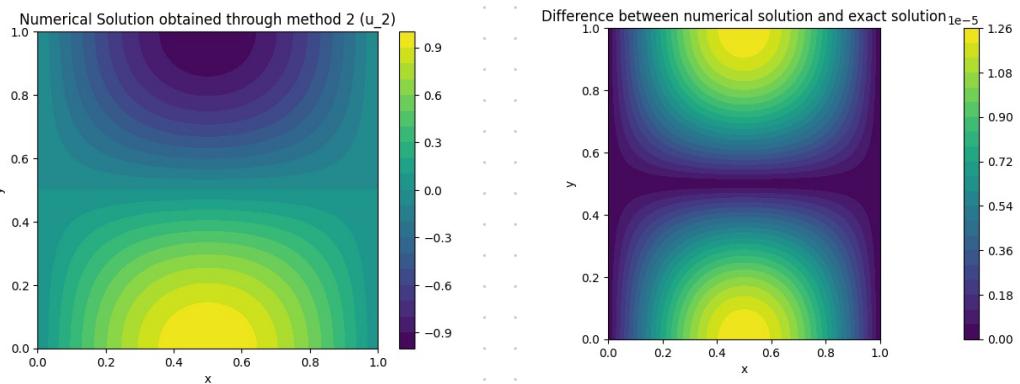
Difference between numerical solution and exact solution



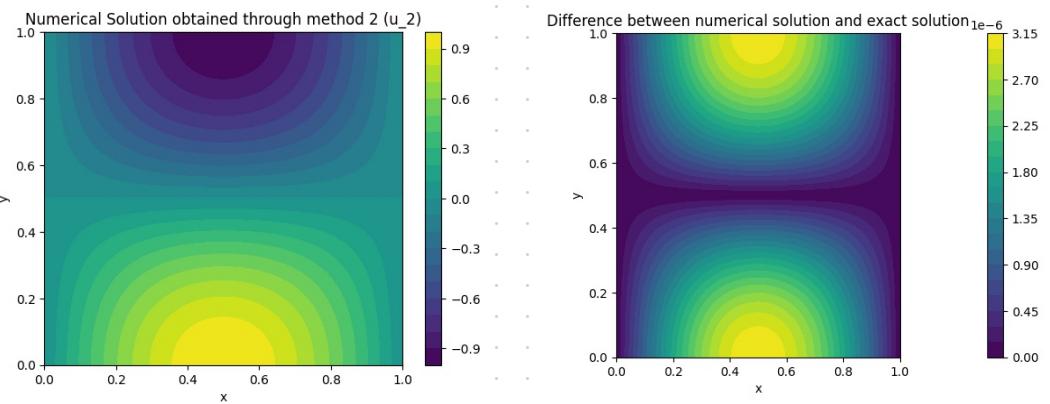
$\{h, p\} = \{\frac{1}{128}, 1\}$  contours:



$\{h, p\} = \{\frac{1}{256}, 1\}$  contours:

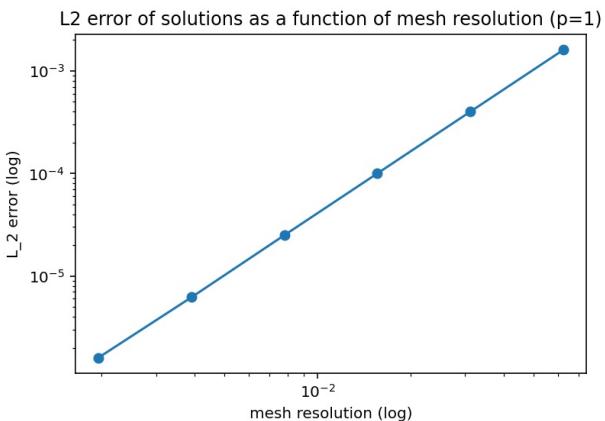


$\{h, p\} = \{\frac{1}{512}, 1\}$  contours:



$L_2$  error for different mesh resolutions ( $h$  values) for  $p=1$ :

$h$	$L_2$ error
$\frac{1}{16}$	$1.6 \times 10^{-3}$
$\frac{1}{32}$	$4.0 \times 10^{-4}$
$\frac{1}{64}$	$1.0 \times 10^{-4}$
$\frac{1}{128}$	$2.5 \times 10^{-5}$
$\frac{1}{256}$	$6.2 \times 10^{-6}$
$\frac{1}{512}$	$1.6 \times 10^{-6}$



Plotting the  $L_2$  error of the solution as a function of the mesh resolution  $h$  for order  $p=1$  allows us to find the dependence of the solution on  $h$  for  $p=1$ .

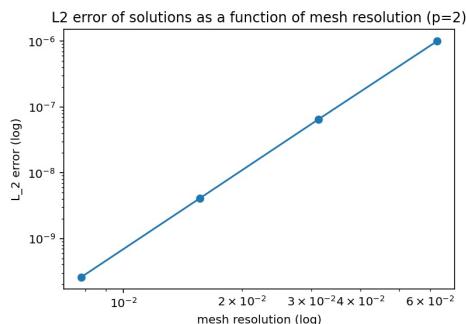
if  $L_2 \propto h^\alpha$ , taking the logscale of the plot allows us to calculate the value of  $\alpha$ .

find that  $\alpha = 1.996 \approx 2$

thus  $L_2 \propto h^2$  for  $p=1$ .  $\Rightarrow$  second order convergence in space for  $p=1$ .

Same analysis for  $p=2$ :

$h$	$L_2$ error
$\frac{1}{16}$	$1.0 \times 10^{-6}$
$\frac{1}{32}$	$6.5 \times 10^{-8}$
$\frac{1}{64}$	$4.1 \times 10^{-9}$
$\frac{1}{128}$	$2.6 \times 10^{-10}$

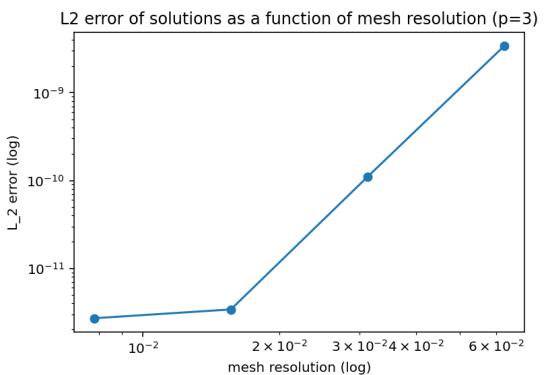


This time it is found that  $\alpha = 3.97 \approx 4$   
 thus for  $p=2$  :  $L_2 \propto h^4 \Rightarrow$  4th order convergence  
 in space for  $p=2$ .

This suggests that  $L_2 \propto h^{2p}$  ( $p=1 \Rightarrow 2p=2$ ,  $p=2 \Rightarrow 2p=4$ )

For  $p=3$ :

$h$	$L_2$ error
$\frac{1}{16}$	$3.4 \times 10^{-9}$
$\frac{1}{32}$	$1.1 \times 10^{-10}$
$\frac{1}{64}$	$3.4 \times 10^{-12}$
$\frac{1}{128}$	$2.7 \times 10^{-12}$



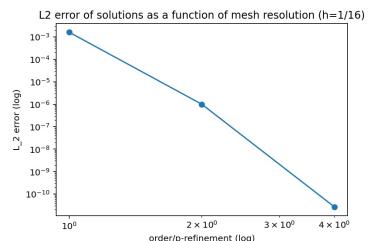
However this trend does not continue for  $p=3$ .

The linear portion of the graph has a gradient of  $\approx 5$  which is closer to expected, but not the expected value of 6.

However at  $p=3$  the errors are small enough that round-off error when generating the numerical solution becomes a concern.

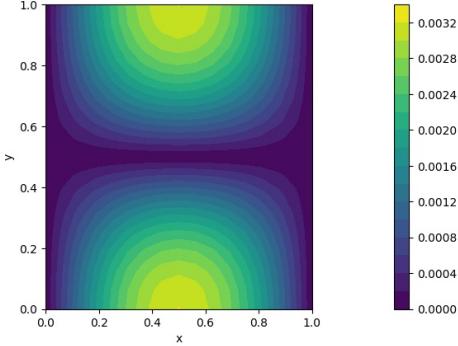
With this in mind, consider different values of  $p$  when  $h=\frac{1}{16}$  to minimize the effect of round-off error.

$p$	$L_2$ error
1	$1.6 \times 10^{-3}$
2	$1.0 \times 10^{-6}$
4	$2.7 \times 10^{-11}$



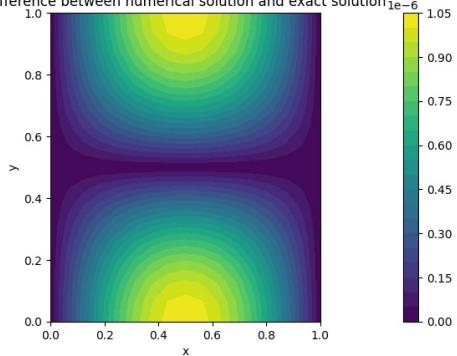
$$\{h, p\} = \{\frac{1}{16}, 1\}$$

Difference between numerical solution and exact solution



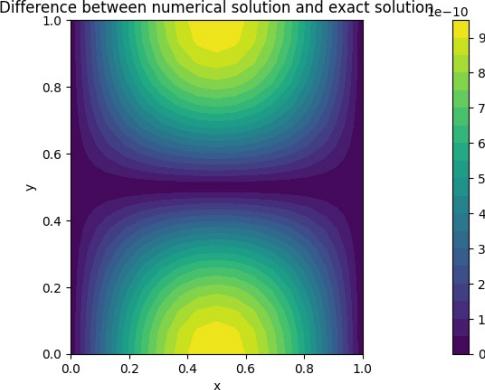
$$\{h, p\} = \{\frac{1}{16}, 2\}$$

Difference between numerical solution and exact solution  $10^{-6}$



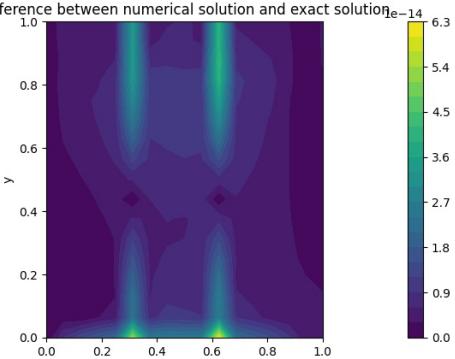
$$\{h, p\} = \{\frac{1}{16}, 3\}$$

Difference between numerical solution and exact solution  $10^{-10}$



$$\{h, p\} = \{\frac{1}{16}, 4\}$$

Difference between numerical solution and exact solution  $10^{-14}$



( $p=3$  not included in analysis since consider doubling  $p$ )

if  $L_2 \propto h^{2p}$  would expect that doubling  $p$  would square the  $L_2$  error.  
 $p \rightarrow 2p \Rightarrow L_2 \propto h^{4p} = (h^{2p})^2$

$(1.6 \times 10^{-3})^2 = 2.56 \times 10^{-6} \sim 1.0 \times 10^{-6}$  if consider potential effects of round-off error.

$p=1, 2 \quad h = \frac{1}{32} : \quad p=1 : L_2 = 4.0 \times 10^{-4}$

$(4.0 \times 10^{-4})^2 = 1.6 \times 10^{-7} \sim 6.5 \times 10^{-8}$

This is approximately the case, thus it can be concluded that the error scales w/  $h^{2p}$

$L_2 \sim h^{2p} \Rightarrow \{h, p\}$  combinations w/ similar  
 $h^{2p}$  values are equivalent.

$$h = 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}, 2^{-9} \quad (\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512})$$

$$p = 1, 2, 3, 4$$

$$(2^{-4})^2 = 2^{-8} \Rightarrow \left\{ \frac{1}{16}, 2 \right\} \sim \left\{ \frac{1}{256}, 1 \right\}$$

$$L_2: 1.0 \times 10^{-6} \sim 6.3 \times 10^{-6}$$

same order  $\Rightarrow$  roughly equivalent

example of equivalent  $\{h, p\}$  pairs w/ same value of  $h^{2p}$ .

5. Explain how first four steps are implemented in Firedrake.

Uploaded code file is fully commented to identify where these steps occur in the code.

6. Change and implement BCs for different  $f(x, y)$  and exact solutions  $u(x, y)$ .

$$-\nabla^2 u = f(x, y)$$

$$\text{consider } u_{\text{exact}} = \sin(\frac{\pi}{2}x) \cos(\pi y)$$

$$\text{then } u(0, y) = 0$$

$$\begin{aligned} u(1, y) &= \sin(\frac{\pi}{2}) \cos(\pi y) \\ &= \cos(\pi y) \end{aligned}$$

$$-\nabla^2 u = -\left(-\frac{\pi^2}{4} \sin(\frac{\pi}{2}x) \cos(\pi y) - \pi^2 \sin(\frac{\pi}{2}x) \cos(\pi y)\right)$$

$$= \frac{5\pi^2}{4} \sin(\frac{\pi}{2}x) \cos(\pi y)$$

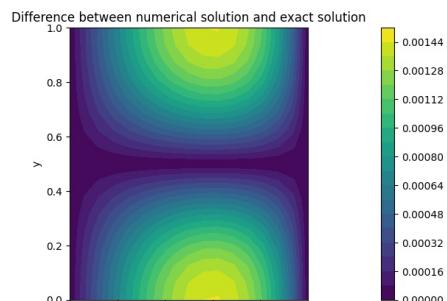
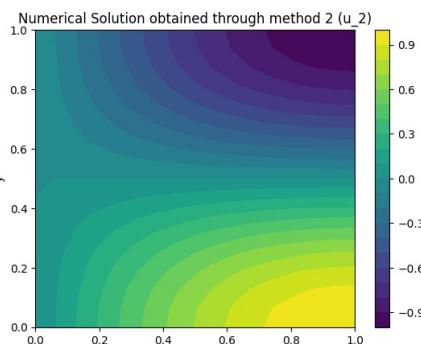
$$= f(x, y)$$

note: neumann BCs imposed, same as before.

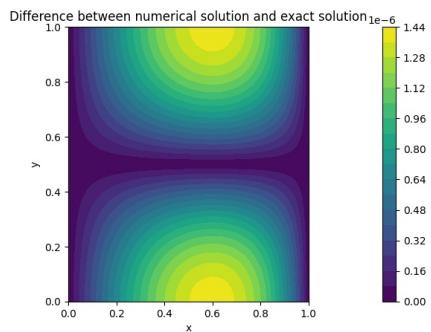
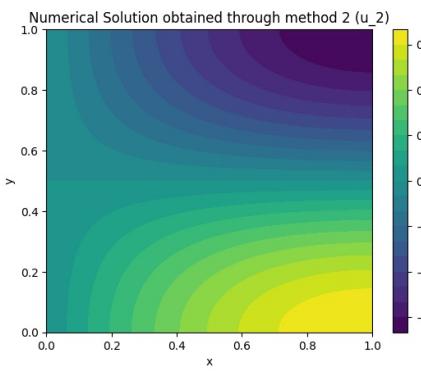
Steps taken to change the code:

1. change function  $f$  to  $\frac{5\pi^2}{4} \sin(\frac{\pi}{8}x) \cos(\pi y)$
2. change BCs  $bc\_x_0, bc\_x$ .  
 $bc\_x_0$  stays the same for this chosen function  
 $bc\_x$ , no longer a constant, so function  $\cos(\pi y)$  must be implemented instead.
3. change exact solution to  $\sin(\frac{\pi}{8}x) \cos(\pi y)$
4. run code over chosen  $\{h, p\}$  to verify it works.

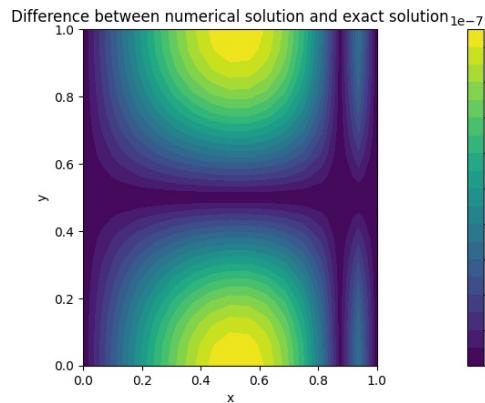
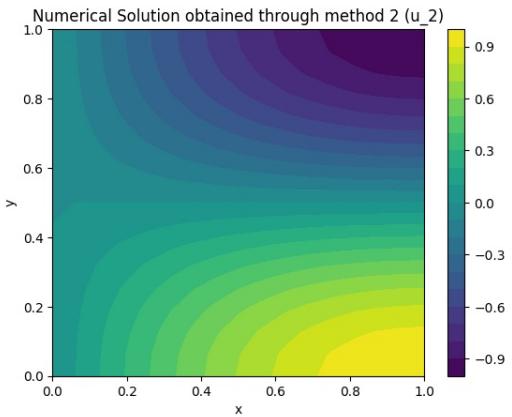
$$\{h, p\} = \left\{ \frac{1}{16}, 1 \right\}$$



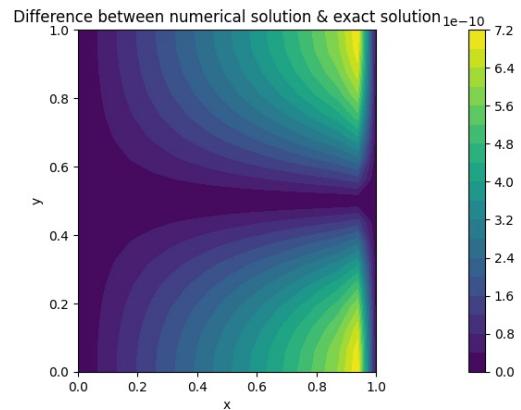
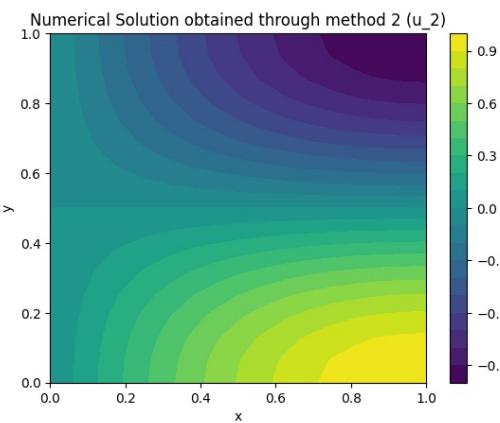
$$\{h, p\} : \left\{ \frac{1}{512}, 1 \right\}$$



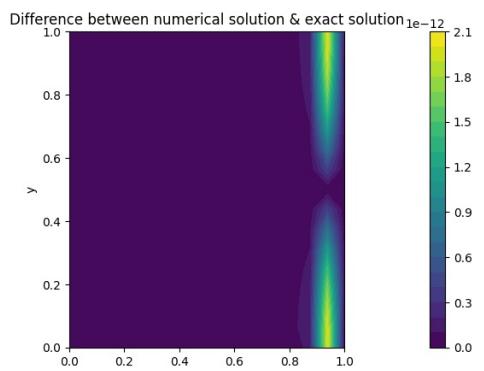
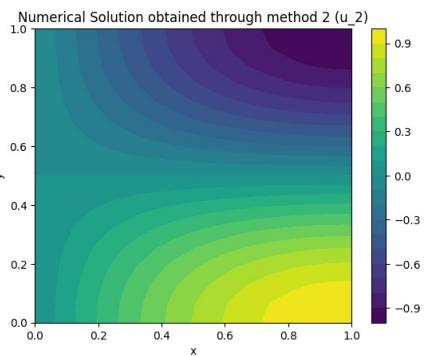
$$\{h, p\} = \left\{\frac{1}{16}, 2\right\}$$



$$\{h, p\} = \left\{\frac{1}{16}, 3\right\}$$



$$\{h, p\} = \left\{\frac{1}{16}, 4\right\}$$



The error  $\|u_2 - u_{\text{exact}}\|$  decreases w/ decreasing  $h$   
& increasing  $p$

The error converges more quickly w/ changing  $p$ , as  
for the previously used  $f$ .

The contours for  $\|u_2 - u_{\text{exact}}\|$  for the chosen  $\{h, p\}$   
pairs suggests that the new function  $f(x, y)$  and BCs  
have been successfully implemented.

Well done; especially in Q6.