# Contrast problem in PDE
## a computational study with Finite Element Method

Simiao Zuo[1] and Ivo Babuska[2]

[1]Department of Computer Science, The University of Texas at Austin

[2]Institute for Computational Engineering and Sciences, The University of Texas at Austin

September 5, 2017

# 1 Introduction

The Finite Element Method (FEM) is today without doubt one of the major numerical tools in computational science. This position of pre-eminence has arisen as a result of a number of complementary factors. These include: the requirement for engineers, scientists, financiers and clinicians to be able to simulate and solve numerically ever more complex problems; the results on the theory of finite elements that have been produced by mathematicians since the 1960s, particularly regarding error estimation; and last but by no means least, the inexorable increase in computer power over the last half century and the corresponding increase in the expectations of users.

Many problems in engineering are PDEs with large contrast. These could lead to the system of linear equations with a large scaled condition number. The report analyzes the condition number in connection with different boundary conditions of such problems.

# 2 One-dimensional interface problem

## 2.1 The interface problem

Consider the following interface problem:

$$-\frac{d}{dx}(a(x)\frac{du(x)}{dx}) = f(x), 0 \leq x \leq 1$$

$$\begin{cases} u(0) = 0 & or \quad u'(0) = 0 \\ u(1) = 0 & or \quad u'(1) = 0 \end{cases}$$

where

$$a(x) = \begin{cases} 1 & 0 \le x < \beta \\ A & \beta \le x \le 1 \end{cases}$$

In the interface problem, $x = \beta$ is the position of the interface, and $A$ is the contrast. Homogeneous Dirichlet or Neumann boundary is imposed at each end, i.e. $x = 0$ and $x = 1$.

For simplification, we take $f(x) = 1$ for all the problems.

## 2.2 Construction of the exact solution

### 2.2.1 Dirichlet-Dirichlet boundary problem

By solving the equations

$$\begin{cases} -u''(x) = 1 & 0 \le x < \beta \\ -Au''(x) = 1 & \beta \le x \le 1 \end{cases}$$

we have a general solution

$$\begin{cases} u_1(x) = -\frac{x^2}{2} + c_1 x + c_2 & 0 \le x < \beta \\ u_2(x) = -\frac{x^2}{2A} + c_3 x + c_4 & \beta \le x \le 1 \end{cases} \tag{1}$$

Homogeneous Dirichlet boundary conditions are imposed at both endpoints, and we impose $u(x)$ to be continuously differentiable at $x = \beta$, thus

$$\begin{cases} u_1(0) = 0, & u_2(1) = 0 \\ u_1(\beta) = u_2(\beta), & u_1'(\beta) = A u_2'(\beta) \end{cases}$$

By solving the equations, we have

$$u(x) = \begin{cases} -\frac{x^2}{2} + \frac{\beta^2(A-1)+1}{2(\beta(A-1)+1)} x & 0 \le x < \beta \\ -\frac{x^2}{2A} + \frac{\beta^2(A-1)+1}{2A(\beta(A-1)+1)} x + \frac{(\beta-\beta^2)(A-1)}{2A(\beta(A-1)+1)} & \beta \le x \le 1 \end{cases}$$

Obviously when the contrast $A$ goes to infinity

$$u_{A \to \infty}(x) = \begin{cases} -\frac{x^2}{2} + \frac{\beta}{2} x & 0 \le x < \beta \\ 0 & \beta \le x \le 1 \end{cases}$$

2

### 2.2.2   Dirichlet-Neumann boundary problem

In the same way, we have equation 1, then we impose homogeneous Dirichlet boundary condition at $x = 0$ and homogeneous Neumann boundary condition at $x = 1$.

By solving

$$\begin{cases} u_1(0) = 0, & u_2'(1) = 0 \\ u_1(\beta) = u_2(\beta), & u_1'(\beta) = Au_2'(\beta) \end{cases}$$

we have

$$u(x) = \begin{cases} -\frac{x^2}{2} + x & 0 \leq x < \beta \\ -\frac{x^2}{2A} + \frac{1}{A}x + \frac{(2\beta - \beta^2)(A-1)}{2A} & \beta \leq x \leq 1 \end{cases}$$

Obviously when the contrast $A$ goes to infinity

$$u_{A \to \infty}(x) = \begin{cases} -\frac{x^2}{2} + x & 0 \leq x < \beta \\ \frac{2\beta - \beta^2}{2} & \beta \leq x \leq 1 \end{cases}$$

### 2.2.3   Neumann-Dirichlet boundary problem

Similarly to above, we impose homogeneous Neumann boundary at $x = 0$ and homogeneous Dirichlet boundary at $x = 1$ on equation 1.

By solving

$$\begin{cases} u_1'(0) = 0, & u_2(1) = 0 \\ u_1(\beta) = u_2(\beta), & u_1'(\beta) = Au_2'(\beta) \end{cases}$$

we have

$$u(x) = \begin{cases} -\frac{x^2}{2} + \frac{\beta^2(A-1)+1}{2A} & 0 \leq x < \beta \\ -\frac{x^2}{2A} + \frac{1}{2A} & \beta \leq x \leq 1 \end{cases}$$
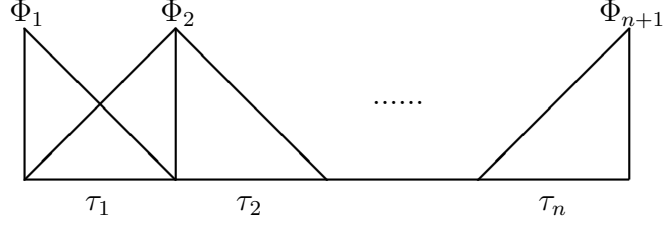
When the contrast $A$ goes to infinity, we obtain

$$u_{A \to \infty}(x) = \begin{cases} -\frac{x^2}{2} + \frac{\beta^2}{2} & 0 \leq x < \beta \\ 0 & \beta \leq x \leq 1 \end{cases}$$

## 2.3 Construction of the FEM solution

### 2.3.1 Construction of the system of equations

Consider a uniform mesh on $\Omega = [0, 1]$:



where

$$\Phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{h} & x_{i-1} \leq x < x_i \\ \frac{x_{i+1} - x}{h} & x_i \leq x < x_{i+1} \\ 0 & otherwise \end{cases}$$

are the shape functions, and the FEM solution $u(x) = \sum_i u_i \Phi_i(x)$.

Denote $h$ the size of each element. Since we are taking a uniform mesh, $h$ is a constant, then $N = \frac{1}{h}$ is the number of elements. Denote function $B(u, v) = \int_0^1 a(x) u'(x) v'(x) dx$.

We can construct the system of equations

$$\begin{bmatrix} B(\Phi_1, \Phi_1) & B(\Phi_1, \Phi_2) & \dots & B(\Phi_1, \Phi_{n+1}) \\ B(\Phi_2, \Phi_1) & B(\Phi_2, \Phi_2) & \dots & B(\Phi_2, \Phi_{n+1}) \\ \vdots & \vdots & \ddots & \vdots \\ B(\Phi_{n+1}, \Phi_1) & B(\Phi_{n+1}, \Phi_2) & \dots & B(\Phi_{n+1}, \Phi_{n+1}) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} \int_0^1 f \Phi_1 dx \\ \int_0^1 f \Phi_2 dx \\ \vdots \\ \int_0^1 f \Phi_{n+1} dx \end{bmatrix}$$

Construct the global stiffness matrix element by element. For the elements that don't contain the interface, notice that each element $\tau_i$ is only connected to $\Phi_i$ and $\Phi_{i+1}$, thus we have
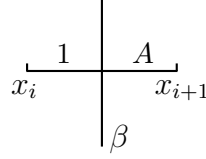
$$B(\Phi_i, \Phi_j) = 0 \quad |i - j| > 1$$

$$B(\Phi_i, \Phi_i) = B(\Phi_{i+1}, \Phi_{i+1}) = \int_\tau \frac{a}{h^2} dx = \frac{a}{h}$$

$$B(\Phi_i, \Phi_{i+1}) = B(\Phi_{i+1}, \Phi_i) = \int_\tau -\frac{a}{h^2} dx = -\frac{a}{h}$$

Therefore, for each element that don't contain the interface, the local stiffness matrix is

$$\mathbf{K} = \begin{bmatrix} \frac{a}{h} & -\frac{a}{h} \\ -\frac{a}{h} & \frac{a}{h} \end{bmatrix}$$

Now consider the element that contains the interface

4

Here obviously $x_i = \lfloor \beta \rfloor$ and $i = \lceil N\beta \rceil$. Then

$$B(\Phi_i, \Phi_i) = B(\Phi_{i+1}, \Phi_{i+1}) = \int_{x_i}^{\beta} \frac{1}{h^2} dx + \int_{\beta}^{x_{i+1}} \frac{A}{h^2} dx = \frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2}$$

$$B(\Phi_i, \Phi_{i+1}) = B(\Phi_{i+1}, \Phi_i) = \int_{x_i}^{\beta} -\frac{1}{h^2} dx + \int_{\beta}^{x_{i+1}} -\frac{A}{h^2} dx = -\frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2}$$

Therefore, for the element that contains the interface, the local stiffness matrix is

$$\mathbf{K} = \begin{bmatrix} \frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} & -\frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} \\ -\frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} & \frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} \end{bmatrix}$$

Now construct the load vector of each element

$$\mathbf{f} = \begin{bmatrix} \int_{\tau} \frac{x - x_{i-1}}{h} dx \\ \int_{\tau} \frac{x_{i+1} - x}{h} dx \end{bmatrix} = \begin{bmatrix} \frac{h}{2} \\ \frac{h}{2} \end{bmatrix}$$

Then we can construct the unconstrained system of equations (with Neumann-Neumann boundary conditions)

$$\begin{bmatrix} \frac{1}{h} & -\frac{1}{h} & & & & & & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & -\frac{1}{h} & \frac{2}{h} & & -\frac{1}{h} & & & \\ & & & -\frac{1}{h} & (\frac{1}{h} + \frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2}) & -\frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} & & & \\ & & & & -\frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2} & (\frac{A}{h} + \frac{\beta - x_i + A(x_{i+1} - \beta)}{h^2}) & -\frac{A}{h} & & \\ & & & & & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} & \\ & & & & & & \ddots & \ddots & \ddots \\ & & & & & & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\ & & & & & & & -\frac{A}{h} & \frac{A}{h} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ u_n \\ u_{n+1} \end{bmatrix} = \begin{bmatrix} \frac{h}{2} \\ h \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ h \\ \frac{h}{2} \end{bmatrix}$$

### 2.3.2 FEM solution of Dirichlet-Dirichlet problem

When imposing homogeneous Dirichlet boundaries at both endpoints, the system of equations become

$$
\begin{bmatrix}
\frac{2}{h} & -\frac{1}{h} & & & \\
-\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\
& \ddots & \ddots & \ddots & \\
& & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\
& & & -\frac{A}{h} & \frac{2A}{h}
\end{bmatrix}
\begin{bmatrix}
u_2 \\
u_3 \\
\vdots \\
u_{n-1} \\
u_n
\end{bmatrix}
=
\begin{bmatrix}
h \\
h \\
\vdots \\
h \\
h
\end{bmatrix}
$$

with $u_1 = 0$ and $u_{n+1} = 0$.

By solving the system of equations, we obtain the FEM solution.
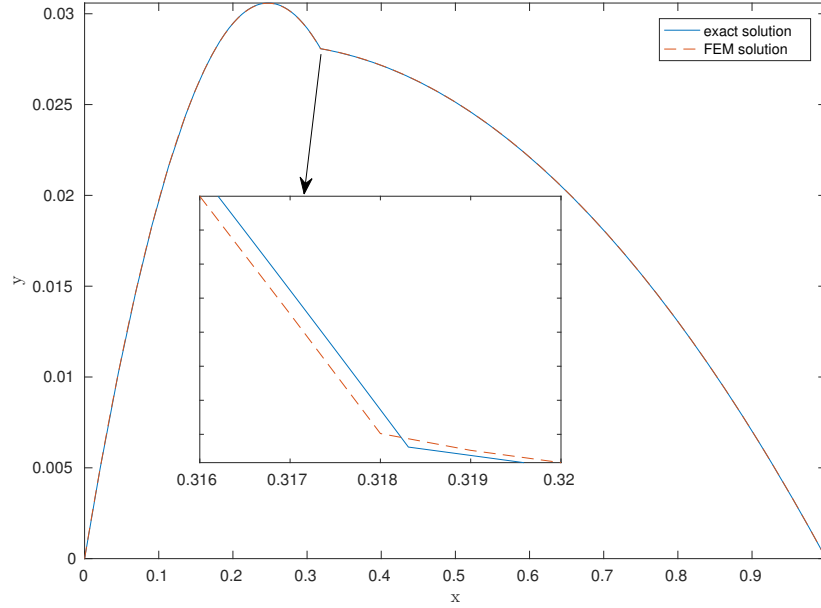


Figure 1: FEM solution, $h = 10^{-3}, \beta = \frac{1}{\pi}, A = 10$, Dirichlet-Dirichlet boundary
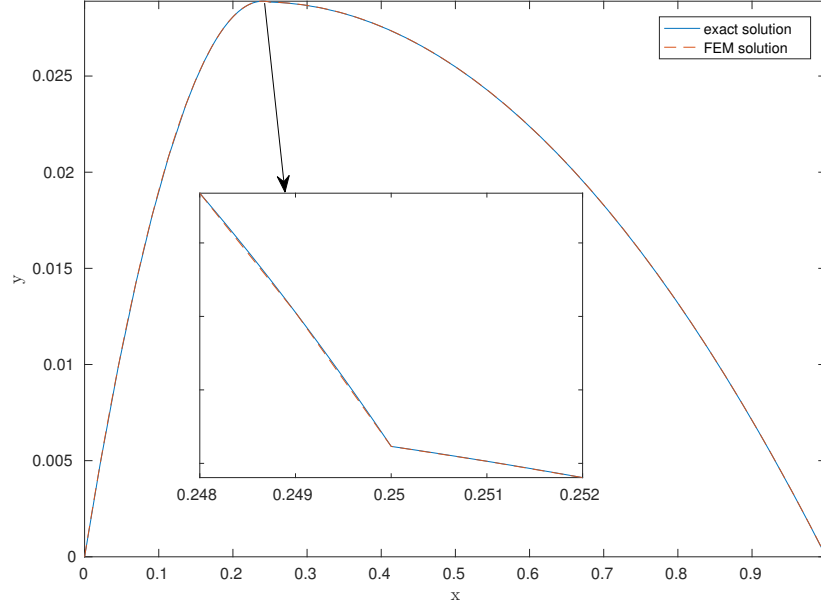
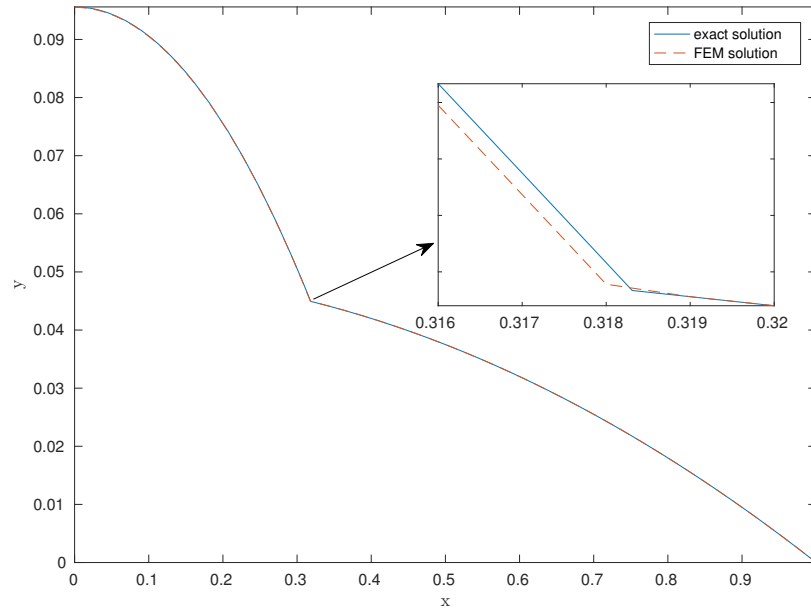Figure 2: FEM solution, $h = 10^{-3}, \beta = \frac{1}{4}, A = 10$, Dirichlet-Dirichlet boundary

We can see the different behavior of the FEM solution when the position of the interface changes, i.e. from $\frac{1}{\pi}$ to $\frac{1}{4}$. This is due to the fact that when taking $\beta = \frac{1}{\pi}$, the interface is always between two nodal points, but when taking $\beta = \frac{1}{4}$, the interface can be on a nodal point. The same happens when imposing other types of boundary conditions.

### 2.3.3 FEM solution of Dirichlet-Neumann problem

Similarly, the system of equations are

$$
\begin{bmatrix}
\frac{2}{h} & -\frac{1}{h} & & & \\
-\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\
& \ddots & \ddots & \ddots & \\
& & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\
& & & -\frac{A}{h} & \frac{A}{h}
\end{bmatrix}
\begin{bmatrix}
u_2 \\
u_3 \\
\vdots \\
u_n \\
u_{n+1}
\end{bmatrix}
=
\begin{bmatrix}
h \\
h \\
\vdots \\
h \\
\frac{h}{2}
\end{bmatrix}
$$

with $u_1 = 0$. We can then construct the FEM solution
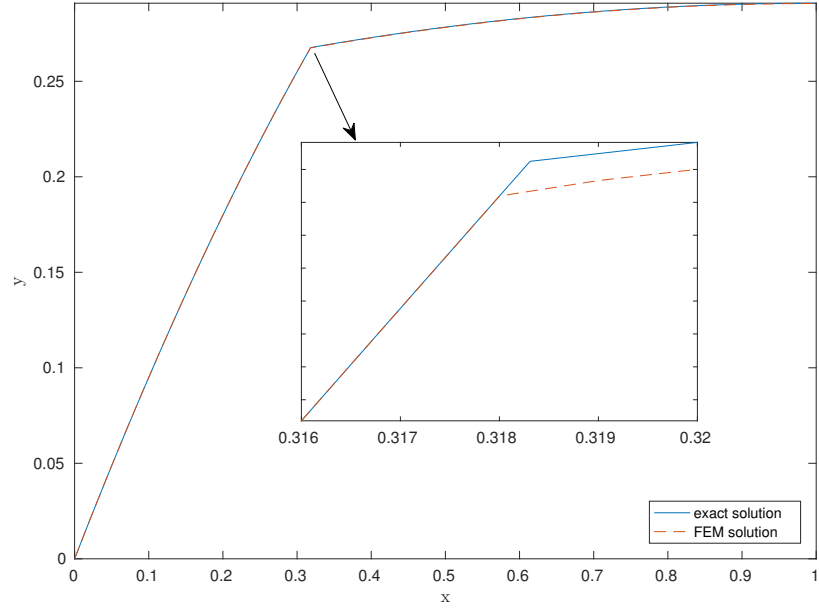
Figure 3: FEM solution, $h = 10^{-3}, \beta = \frac{1}{\pi}, A = 10$, Dirichlet-Neumann boundary
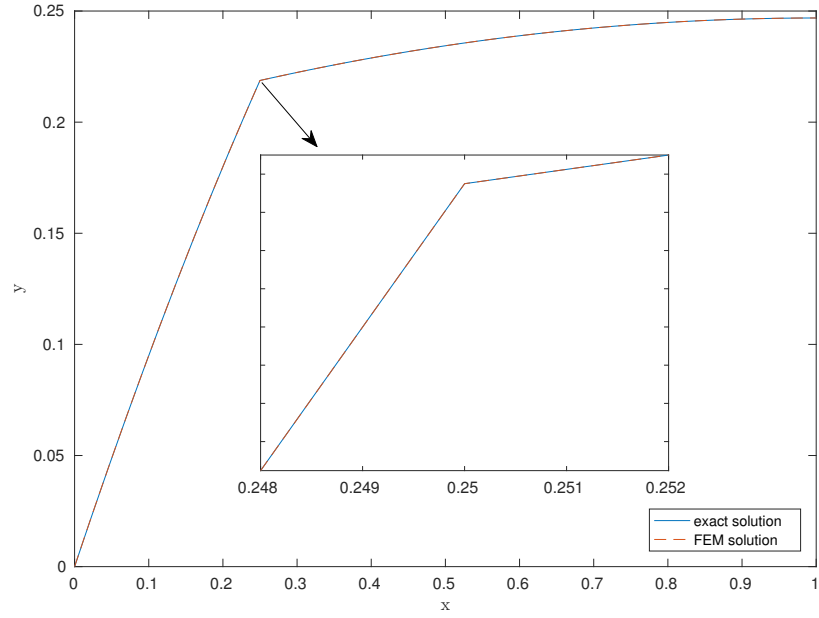


Figure 4: FEM solution, $h = 10^{-3}, \beta = \frac{1}{4}, A = 10$, Dirichlet-Neumann boundary

### 2.3.4 FEM solution of Neumann-Dirichlet problem

Now the system of equations are

$$
\begin{bmatrix}
\frac{1}{h} & -\frac{1}{h} & & & \\
-\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\
& \ddots & \ddots & \ddots & \\
& & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\
& & & -\frac{A}{h} & \frac{2A}{h}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_{n-1} \\
u_n
\end{bmatrix}
=
\begin{bmatrix}
\frac{h}{2} \\
h \\
\vdots \\
h \\
h
\end{bmatrix}
$$

with $u_{n+1} = 0$.

The FEM solution is



Figure 5: FEM solution, $h = 10^{-3}, \beta = \frac{1}{\pi}, A = 10$, Neumann-Dirichlet boundary
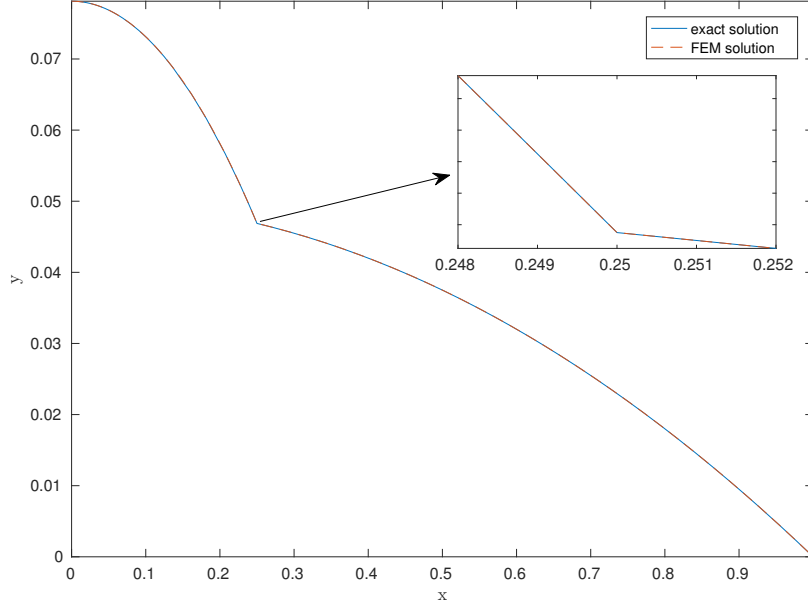
Figure 6: FEM solution, $h = 10^{-3}, \beta = \frac{1}{4}, A = 10$, Neumann-Dirichlet boundary

## 2.4 Error analysis of FEM solution

### 2.4.1 Error of Dirichlet-Dirichlet problem

The error is computed piecewise in energy norm

$$e^2 = B(u - u_h, u - u_h) = \sum_i \int_{x_i}^{x_{i+1}} a(x)((u - u_h)')^2 dx$$

where $u$ is the exact solution, and $u_h$ is the FEM solution.

We can see that the error of FEM solution behaves differently for interface $\beta = \frac{1}{\pi}$ and $\beta = \frac{1}{4}$, which is intuitively shown in the comparison of the FEM solution with the exact solution. The reason is that for the elements that don't contain the interface, $error \propto h$, and for the element that contains the interface, $error \propto \sqrt{h}$. Notice that in figure 8, the number of elements $N$ is always a multiplication of 4 and we take the interface $\beta = \frac{1}{4}$. Thus the interface is always on a nodal point, therefore $e \propto h$. On the other hand, when taking the interface $\beta = \frac{1}{\pi}$, it's always contained in one of the elements, the combined consequence is that $e \propto \sqrt{h}$.

As the graphs illustrated, when taking $\beta = \frac{1}{4}$, the error of FEM solution converges linearly to 0 at the rate of $h$. And when taking $\beta = \frac{1}{\pi}$, the error converges to 0 at the rate of $\sqrt{h}$. Worth noticing that when taking $\beta = \frac{1}{\pi}$, the error does not converges linearly, and even not converging locally, as shown by the small graph in figure 7. This is also true for other types of boundary conditions.

10
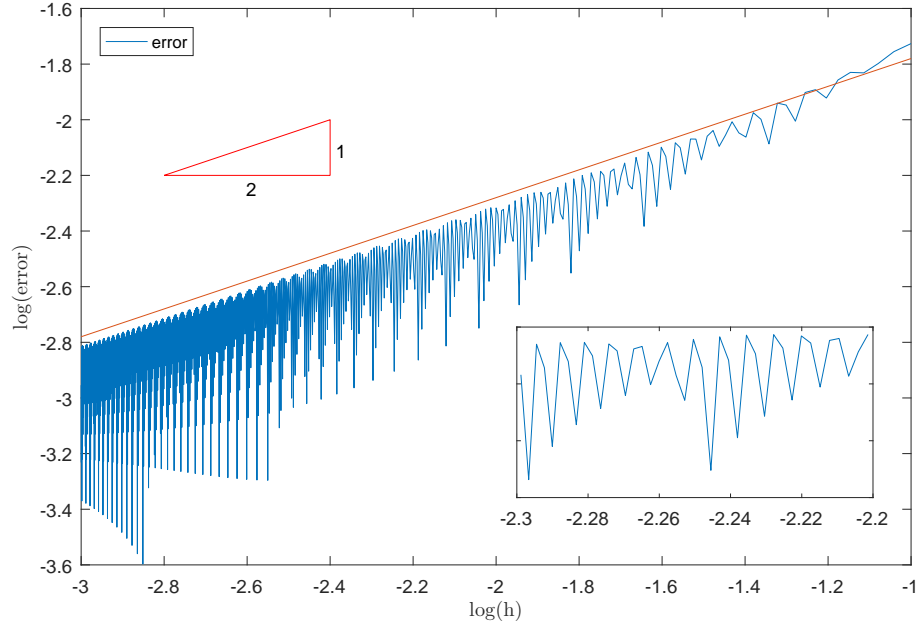
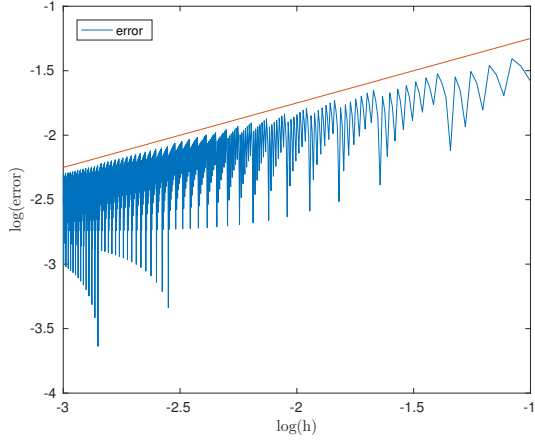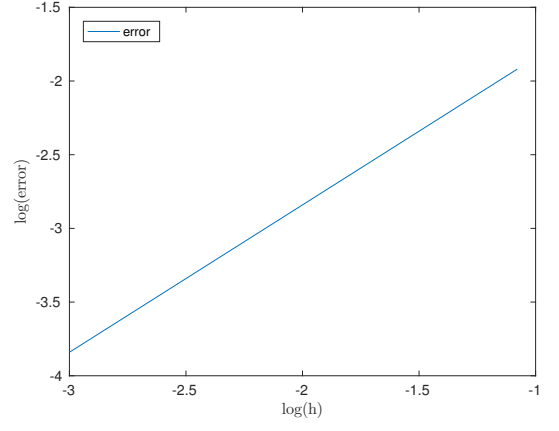Figure 7: Error of FEM solution, $\beta = \frac{1}{\pi}, A = 10$, Dirichlet-Dirichlet boundary
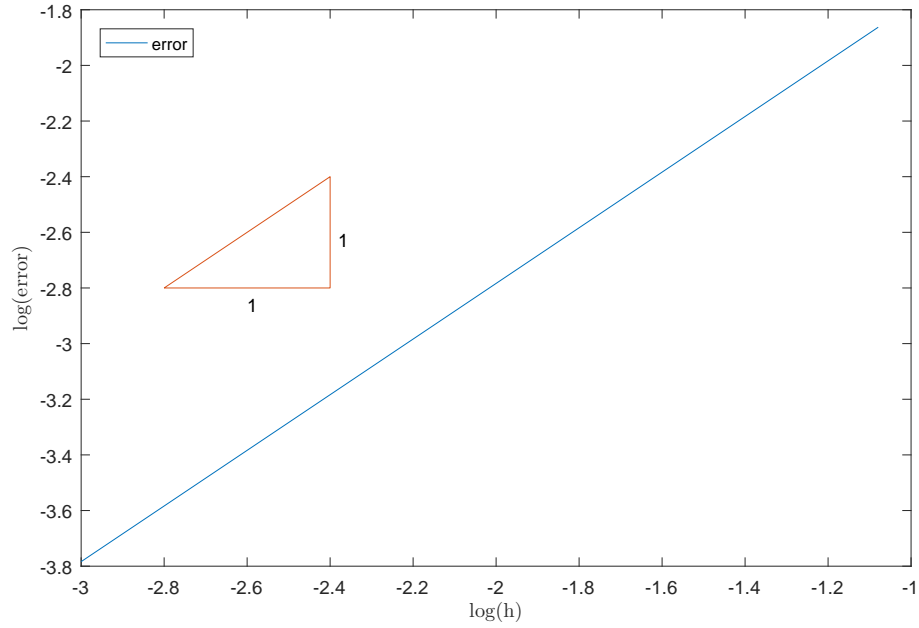


Figure 8: Error of FEM solution, $h = \frac{1}{4k}, \beta = \frac{1}{4}, A = 10$, Dirichlet-Dirichlet boundary

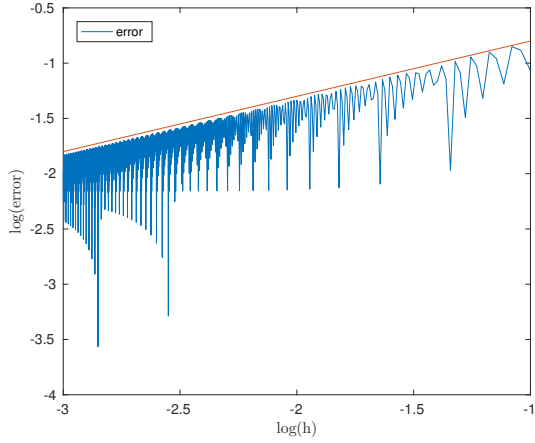Now consider a large contrast.

(a) $\beta = \frac{1}{\pi}$        (b) $h = \frac{1}{4k}, \beta = \frac{1}{4}$

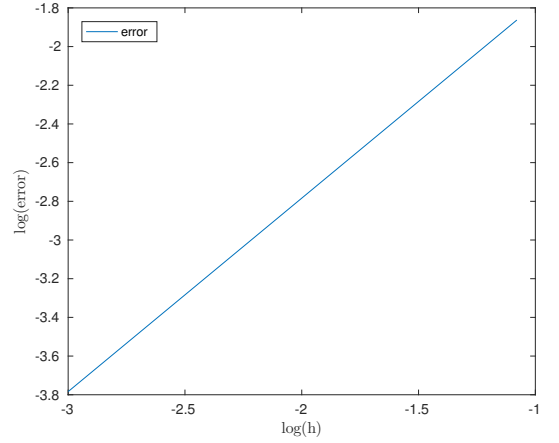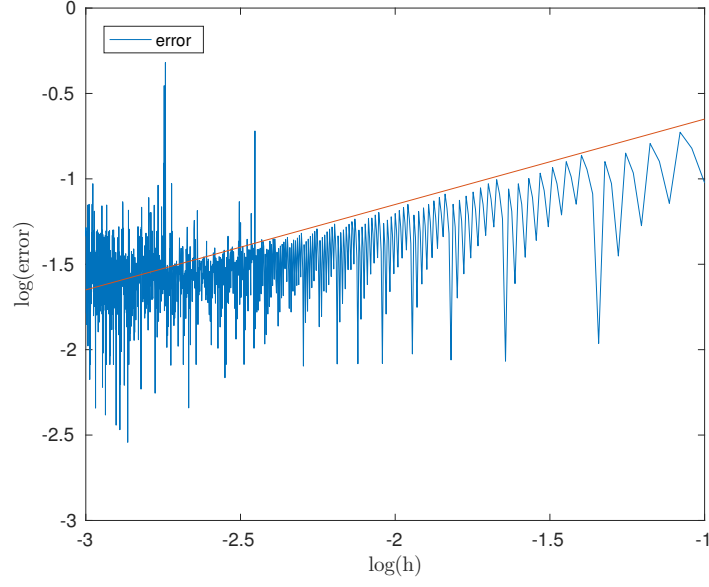Figure 9: Error of FEM solution,$A = 10^{12}$, Dirichlet-Dirichlet boundary

Notice that increasing the contrast does not have a significant influence on the behavior of error.

### 2.4.2 Error of Dirichlet-Neumann problem

In the same way we compute the error of FEM solution in energy norm.



(a) $\beta = \frac{1}{\pi}$        (b) $h = \frac{1}{4k}, \beta = \frac{1}{4}$

Figure 10: Error of FEM solution,$A = 10$, Dirichlet-Neumann boundary

In comparison, error is also computed with a large contrast.

Figure 11: Error of FEM solution, $\beta = \frac{1}{\pi}$, $A = 10^{12}$, Dirichlet-Neumann boundary
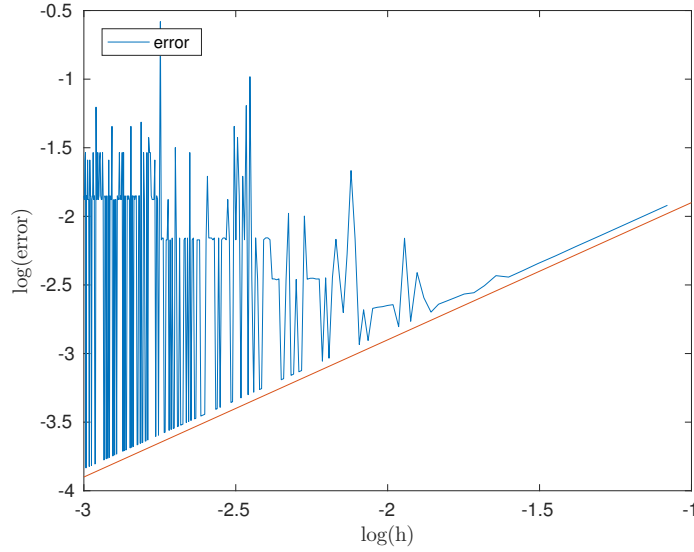


Figure 12: Error of FEM solution, $h = \frac{1}{4k}$, $\beta = \frac{1}{4}$, $A = 10^{12}$, Dirichlet-Neumann boundary

For a small contrast, i.e. $A = 10$, the error behaves similarly to the one of Dirichlet-Dirichlet problem. When taking $\beta = \frac{1}{\pi}$, the error converges to 0 at the rate of $\sqrt{h}$. And when taking $\beta = \frac{1}{4}$, the error converges to 0 at the rate of $h$. However, for a large contrast, i.e. $A = 10^{12}$, the error is unpredictable for $h$ smaller than a certain number, as illustrated in figure 11 and 12.

This is the effect of the round-off error in computations. Note that this error is of order $10^{-2}$ when $h = \frac{1}{300}$ although the computation is in double precision.

### 2.4.3 Error of Neumann-Dirichlet problem

Similarly, the error of FEM solution, both with small contrast and with large contrast, is computed in energy norm.



(a) $\beta = \frac{1}{\pi}$

(b) $h = \frac{1}{4k}, \beta = \frac{1}{4}$
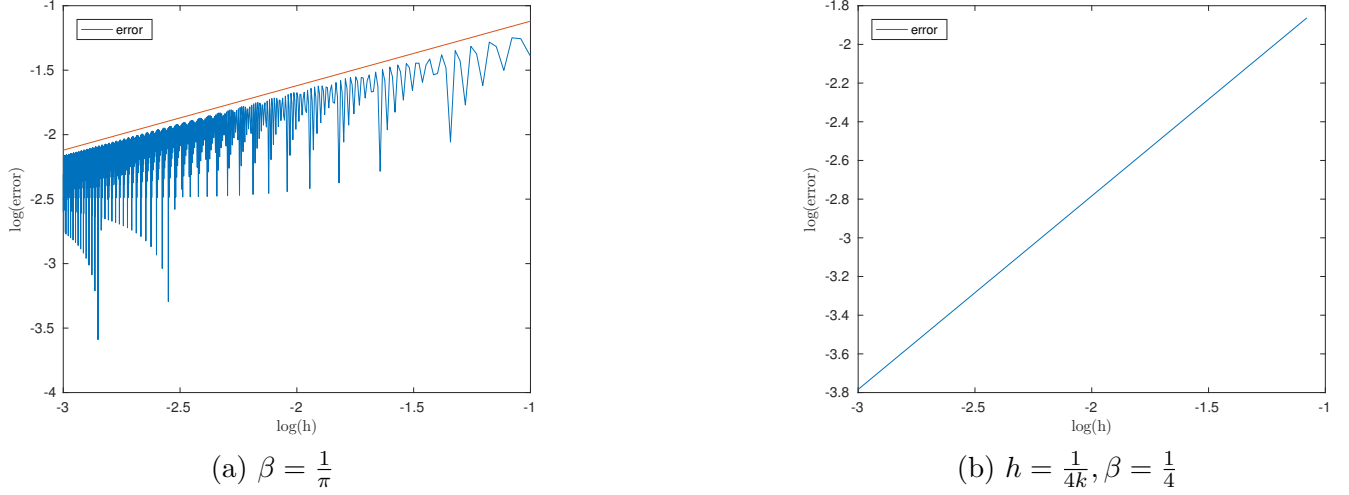
Figure 13: Error of FEM solution,$A = 10$, Neumann-Dirichlet boundary



(a) $\beta = \frac{1}{\pi}$

(b) $h = \frac{1}{4k}, \beta = \frac{1}{4}$
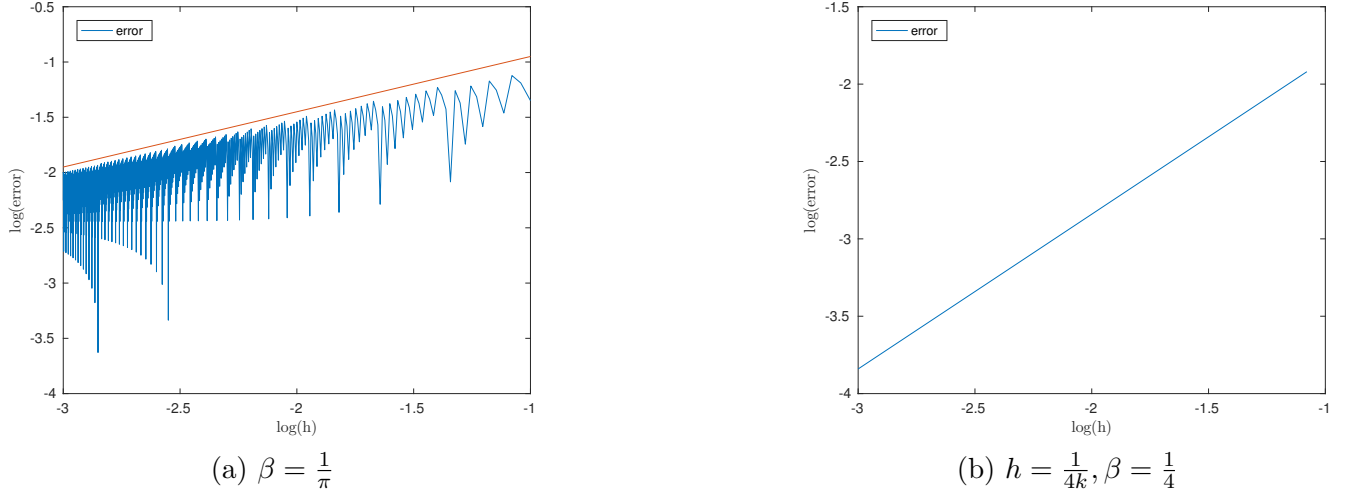
Figure 14: Error of FEM solution,$A = 10^{12}$, Neumann-Dirichlet boundary

Like in Dirichlet-Dirichlet problem, even with large contrast, the error still converges to 0 without any disturbance.

## 2.5 Scaled condition number of global stiffness matrix

As shown above, when having a large contrast, i.e. $A = 10^{12}$, the error of FEM solution is unpredictable. The reason is that we cannot solve the system of equations accurately, numerical experiments are conducted to show the reason.

First we scaled the global stiffness matrix.

$$\mathbf{K}^* = \mathbf{DKD}$$

where

$$\mathbf{D}_{i,i} = \frac{1}{\sqrt{\mathbf{K}_{i,i}}}$$

$$\mathbf{D}_{i,j} = 0 \quad i \neq j$$

so that in this way, all the diagonal elements $\mathbf{K}_{i,i}^* = 1$.

After scaling, the system

$$\mathbf{Ku} = \mathbf{f}$$

becomes

$$\mathbf{D}^{-1}\mathbf{K}^*\mathbf{D}^{-1}\mathbf{u} = \mathbf{f}$$

and thus

$$\mathbf{K}^*\mathbf{y} = \mathbf{Df} \qquad \mathbf{u} = \mathbf{Dy}$$

Notice that by solving the new system, we obtain basically the same result as solving the system $\mathbf{Ku} = \mathbf{f}$. As a matter of fact, if every $\mathbf{D}_{i,j}$ is power of 2, then solving the scaled system and the unscaled system will yield the exact same results. Therefore, scaling the global stiffness matrix doesn't change the FEM solution nor the error, but in this way, we significantly decreases the condition number of the global stiffness matrix.

For simplicity, consider the global stiffness matrix with interface $\beta = \frac{1}{4}$

$$\mathbf{K} = \begin{bmatrix} \frac{1}{h} & -\frac{1}{h} & & & & & & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & & \\ & & & -\frac{1}{h} & \frac{1+A}{h} & -\frac{A}{h} & & & \\ & & & & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} & & \\ & & & & & \ddots & \ddots & \ddots & \\ & & & & & & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\ & & & & & & & -\frac{A}{h} & \frac{A}{h} \end{bmatrix}$$

then take such a $\mathbf{D}$ to scale the global stiffness matrix

$$
\mathbf{D} = \begin{bmatrix}
\sqrt{h} & & & & & & & \\
& \sqrt{\frac{h}{2}} & & & & & & \\
& & \ddots & & & & & \\
& & & \sqrt{\frac{h}{2}} & & & & \\
& & & & \sqrt{\frac{h}{1+A}} & & & \\
& & & & & \sqrt{\frac{h}{2A}} & & \\
& & & & & & \ddots & \\
& & & & & & & \sqrt{\frac{h}{2A}} \\
& & & & & & & & \sqrt{\frac{h}{A}}
\end{bmatrix}
$$

Consider only the elements that contain the interface and the global stiffness matrix of those elements

$$
\tau_i \qquad \tau_{i+1}
$$

$$
\beta
$$
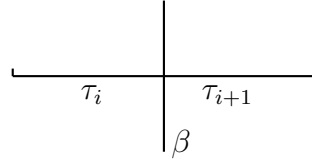
$$
\mathbf{K}_{small} = \begin{bmatrix}
\frac{2}{h} & -\frac{1}{h} & \\
-\frac{1}{h} & \frac{1+A}{h} & -\frac{A}{h} \\
& -\frac{A}{h} & \frac{2A}{h}
\end{bmatrix}
$$

this matrix is scaled by multiplying

$$
\mathbf{D}_{small} = \begin{bmatrix}
\sqrt{\frac{h}{2}} & & \\
& \sqrt{\frac{h}{1+A}} & \\
& & \sqrt{\frac{h}{2A}}
\end{bmatrix}
$$

then

$$
\mathbf{K}^*_{small} = \mathbf{D}_{small}\mathbf{K}_{small}\mathbf{D}_{small} = \begin{bmatrix}
1 & -\frac{1}{\sqrt{2(1+A)}} & \\
-\frac{1}{\sqrt{2(1+A)}} & 1 & -\frac{A}{\sqrt{2(1+A)}} \\
& -\frac{A}{\sqrt{2(1+A)}} & 1
\end{bmatrix}
$$

obviously when the contrast $A$ goes to infinity

$$\lim_{A\to\infty} \mathbf{K}^*_{small} = \begin{bmatrix} 1 & 0 & \\ 0 & 1 & -\frac{1}{\sqrt{2}} \\ & -\frac{1}{\sqrt{2}} & 1 \end{bmatrix}$$

Put it in the global stiffness matrix of all the elements, we have

$$\lim_{A\to\infty} \mathbf{K}^* = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & & & & & & \\ -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{2} & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & -\frac{1}{2} & 1 & & & & \\ & & & & 1 & -\frac{1}{\sqrt{2}} & & \\ & & & & -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{2} & \\ & & & & & \ddots & \ddots & \ddots \\ & & & & & & -\frac{1}{2} & 1 & -\frac{1}{\sqrt{2}} \\ & & & & & & & -\frac{1}{\sqrt{2}} & 1 \end{bmatrix}$$

we can see that the scaled matrix $\mathbf{K}^*_{A\to\infty}$ is divided into two separate sub-matrices

$$\mathbf{K}^*_{A\to\infty} = \left[ \begin{array}{c|c} \mathbf{K}^*_1 & \\ \hline & \mathbf{K}^*_2 \end{array} \right]$$

and the joint of two sub-matrices (both sides of the interface), is behaving as a Neumann boundary



So if we impose Neumann boundary at $x = 1$, then the condition number of $\mathbf{K}^*_2$ is unbounded, and thus the scaled condition number of the global stiffness matrix is unbounded. On the other hand, if we impose Dirichlet boundary at $x = 1$, then the scaled condition number of the global stiffness matrix is bounded. This result is expendable to problems with interface not on nodal point, i.e. $\beta = \frac{1}{\pi}$.

The result is illustrated by the following graphs, notice that the scaled condition number always increases when the size of each element $h$ decreases. This is because the condition number depends on the size of the matrix as well as the contrast. Without the contrast the growth of the condition number is at the rate of $h^{-2}$.
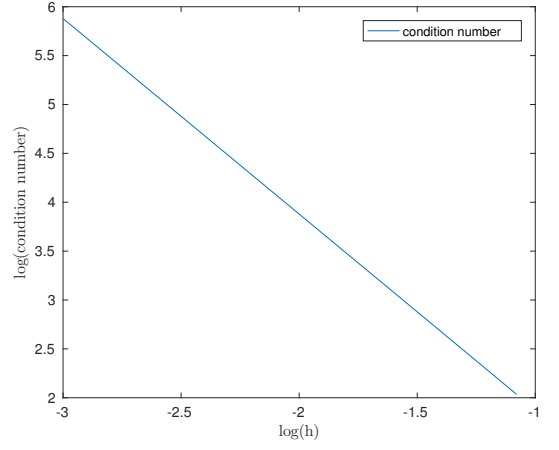
17
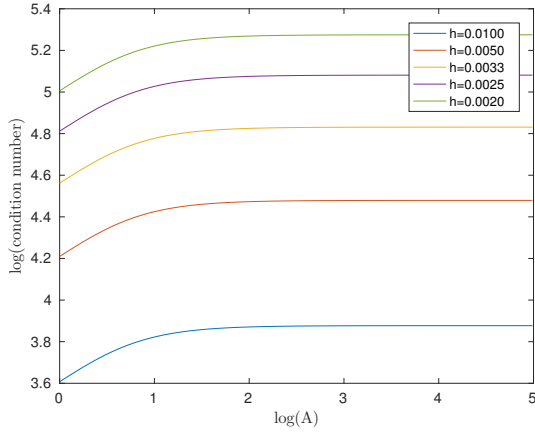
(a)
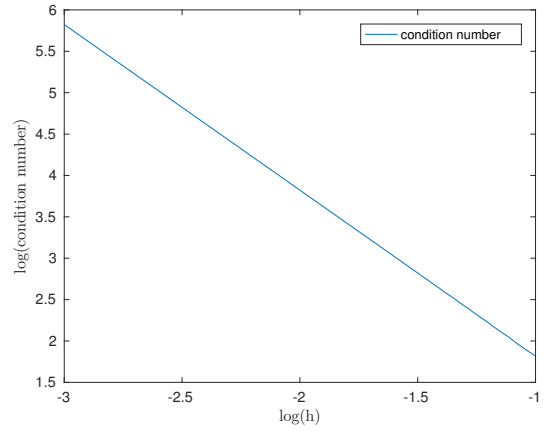
(b) $A = 10$

Figure 15: Scaled condition number, $\beta = \frac{1}{\pi}$, Dirichlet-Dirichlet boundary
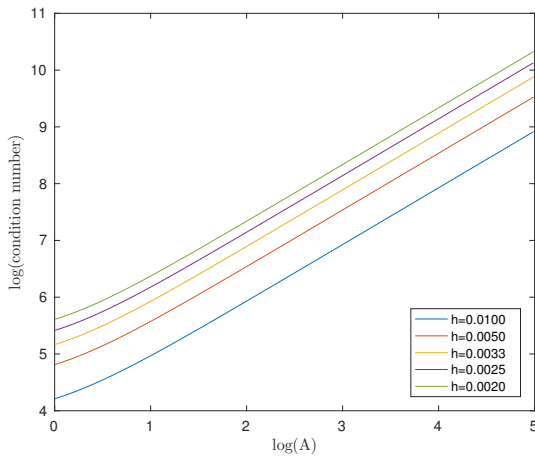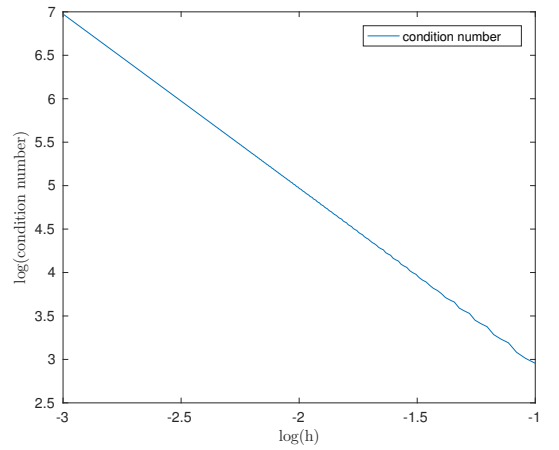


(a)
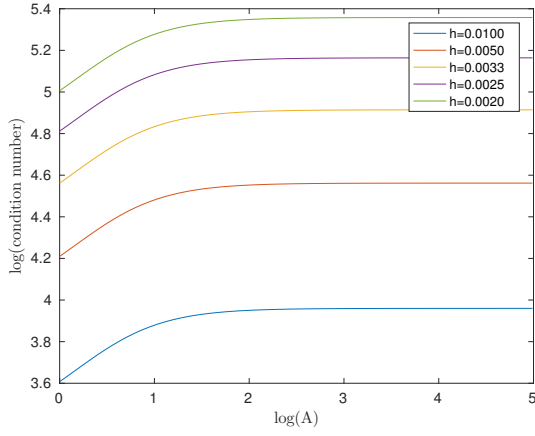
(b) $h = \frac{1}{4k}, A = 10$

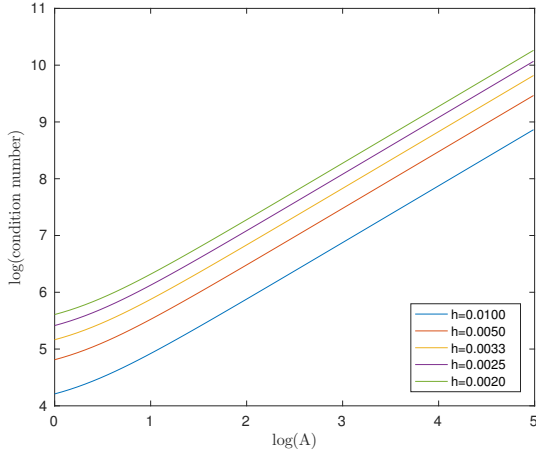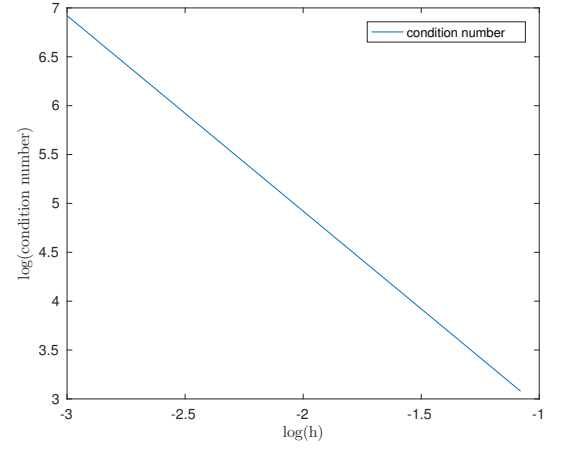Figure 16: Scaled condition number, $\beta = \frac{1}{4}$, Dirichlet-Dirichlet boundary



(a)

(b) $A = 10$

Figure 17: Scaled condition number, $\beta = \frac{1}{\pi}$, Dirichlet-Neumann boundary
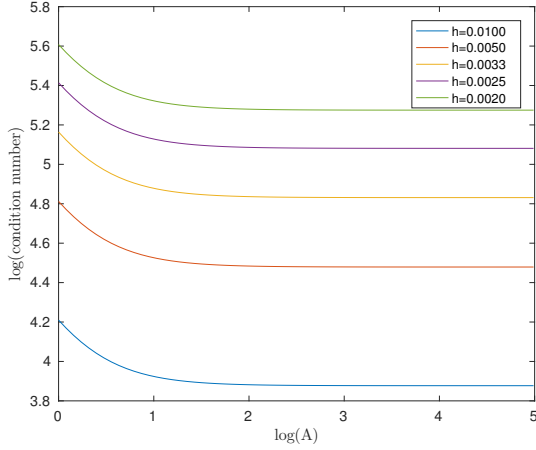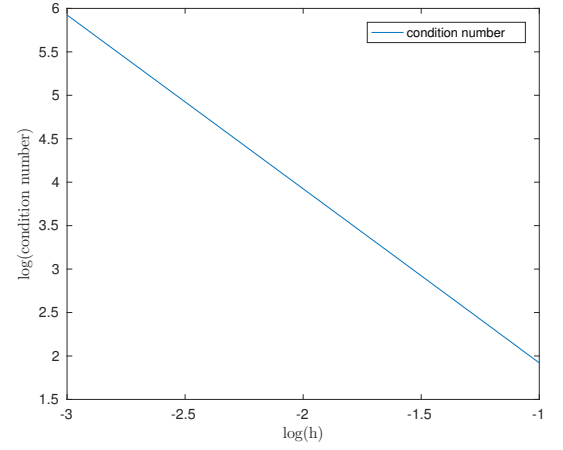
(a)

(b) $h = \frac{1}{4k}, A = 10$

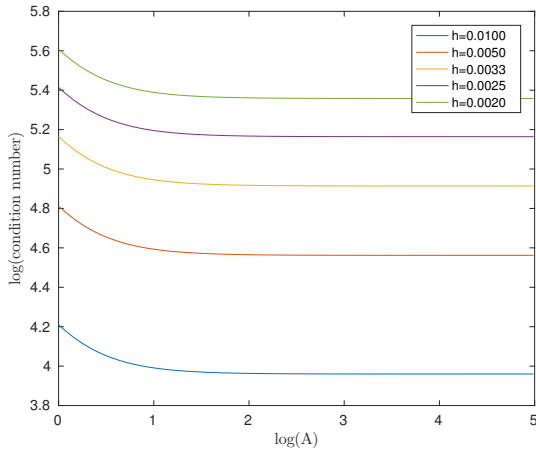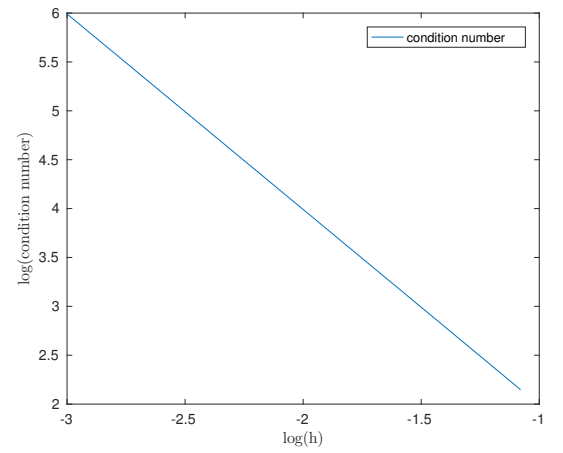Figure 18: Scaled condition number, $\beta = \frac{1}{4}$, Dirichlet-Neumann boundary



(a)

(b) $A = 10$

Figure 19: Scaled condition number, $\beta = \frac{1}{\pi}$, Neumann-Dirichlet boundary



(a)

(b) $h = \frac{1}{4k}, A = 10$

Figure 20: Scaled condition number, $\beta = \frac{1}{4}$, Neumann-Dirichlet boundary

## 2.6 Correction for error disturbance

We know that the unbounded scaled condition number of the global stiffness matrix caused the disturbance in error of FEM solution. Below is a way to fix the disturbance.

Consider the global stiffness matrix and load vector in Dirichlet-Neumann boundary

$$
\mathbf{K} = \begin{bmatrix}
\frac{2}{h} & -\frac{1}{h} & & & & \\
-\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & & \\
& \ddots & \ddots & \ddots & & \\
& & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} & \\
& & & -\frac{A}{h} & \frac{A}{h} &
\end{bmatrix}
\qquad
\mathbf{f} = \begin{bmatrix}
h \\ h \\ \vdots \\ h \\ \frac{h}{2}
\end{bmatrix}
$$

impose Dirichlet boundary at $x = 1$ and denote

$$
\mathbf{A} = \begin{bmatrix}
\frac{2}{h} & -\frac{1}{h} & & & \\
-\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & & \\
& \ddots & \ddots & \ddots & \\
& & -\frac{A}{h} & \frac{2A}{h} & -\frac{A}{h} \\
& & & -\frac{A}{h} & \frac{2A}{h}
\end{bmatrix}
\qquad
\mathbf{b} = \begin{bmatrix}
h \\ h \\ \vdots \\ h \\ h
\end{bmatrix}
\qquad
\mathbf{c} = \begin{bmatrix}
0 \\ 0 \\ \vdots \\ 0 \\ \frac{A}{h}
\end{bmatrix}
$$

Solve the systems

$$
\mathbf{A}\mathbf{x} = \mathbf{b} \qquad \mathbf{A}\mathbf{z} = \mathbf{c}
$$

for $\mathbf{x}$ and $\mathbf{z}$, then the FEM solution

$$
\mathbf{u}_h = \mathbf{x} + k\mathbf{z}
$$

Consider the energy. We know that for any $u$,

$$
E(u) = B(u, u) - 2 \int_0^1 f(x)u(x)dx
$$

thus,

$$
E(\mathbf{u}_h) = B(\mathbf{x}, \mathbf{x}) + 2kB(\mathbf{x}, \mathbf{z}) + k^2 B(\mathbf{z}, \mathbf{z}) - 2 \int_0^1 f(x)\mathbf{x}dx - 2k \int_0^1 f(x)\mathbf{z}dx
$$

to minimize energy, we take

$$
\frac{\partial E(\mathbf{u}_h)}{\partial k} = 0
$$

therefore,

$$
k = \frac{\int_0^1 f(x)\mathbf{z}dx - B(\mathbf{x}, \mathbf{z})}{B(\mathbf{z}, \mathbf{z})}
$$

To compute, we use

$$B(\mathbf{x}, \mathbf{x}) = \sum_i \int_\tau a(x)(\frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{h})^2 dx = \sum_i a(x)\frac{(\mathbf{x}_{i+1} - \mathbf{x}_i)^2}{h}$$

$$B(\mathbf{z}, \mathbf{z}) = \sum_i \int_\tau a(x)(\frac{\mathbf{z}_{i+1} - \mathbf{z}_i}{h})^2 dx = \sum_i a(x)\frac{(\mathbf{z}_{i+1} - \mathbf{z}_i)^2}{h}$$

$$B(\mathbf{x}, \mathbf{z}) = \sum_i \int_\tau a(x)\frac{(\mathbf{x}_{i+1} - \mathbf{x}_i)(\mathbf{z}_{i+1} - \mathbf{z}_i)}{h^2} dx = \sum_i a(x)\frac{(\mathbf{x}_{i+1} - \mathbf{x}_i)(\mathbf{z}_{i+1} - \mathbf{z}_i)}{h}$$

$$\int_0^1 f(x)\mathbf{x}dx = \sum_i \frac{h(\mathbf{x}_{i+1} - \mathbf{x}_i)}{2}$$

$$\int_0^1 f(x)\mathbf{z}dx = \sum_i \frac{h(\mathbf{z}_{i+1} - \mathbf{z}_i)}{2}$$

The energy difference $E(\mathbf{u}_h) - E(u)$ is computed as opposed to error in energy norm. Here $u$ is the exact solution. As illustrated by the graphs, disturbance of error for Dirichlet-Neumann problem vanished.
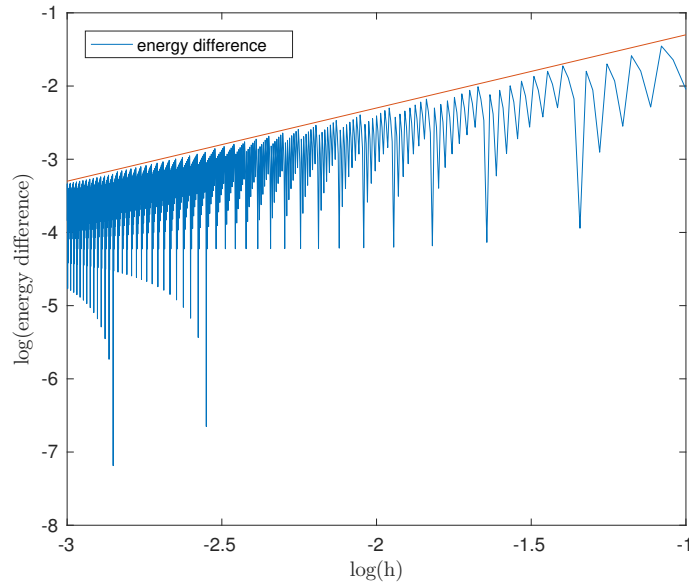


Figure 21: Energy Difference, $\beta = \frac{1}{\pi}$, $A = 10^{12}$, Dirichlet-Neumann boundary
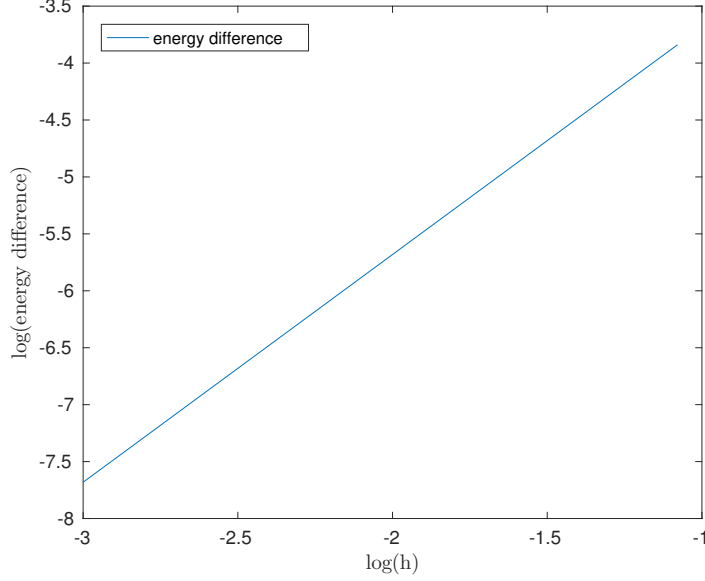
Figure 22: Energy Difference, $h = \frac{1}{4k}, \beta = \frac{1}{4}, A = 10^{12}$, Dirichlet-Neumann boundary

## 2.7 Position of the constraint

So far the constraint of the global stiffness matrix is on the endpoints, i.e. $x = 0$ or $x = 1$. The constraint can also be imposed elsewhere. We consider a Neumann-Neumann boundary problem with constraint put somewhere other than the endpoints.

Still consider the scaled global stiffness matrix for interface $\beta = \frac{1}{4}$

$$
\lim_{A \to \infty} \mathbf{K}^* =
\begin{bmatrix}
1 & -\frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}} & 1 & -\frac{1}{2} \\
& \ddots & \ddots & \ddots \\
& & -\frac{1}{2} & 1 \\
& & & & 1 & -\frac{1}{\sqrt{2}} \\
& & & & -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{2} \\
& & & & & \ddots & \ddots & \ddots \\
& & & & & & -\frac{1}{2} & 1 & -\frac{1}{\sqrt{2}} \\
& & & & & & & -\frac{1}{\sqrt{2}} & 1
\end{bmatrix}
\qquad
\mathbf{K}^*_{A \to \infty} =
\left[
\begin{array}{c|c}
\mathbf{K}^*_1 & \\
\hline
& \mathbf{K}^*_2
\end{array}
\right]
$$

if we put the constraint right on the interface, i.e. at $x = \frac{1}{4}$, then the scaled condition number is bounded. If we put the constraint in $\mathbf{K}^*_1$, i.e. at $x = 0.1$, then the problem is analog to a Dirichlet-Neumann problem; and if we put the constraint in $\mathbf{K}^*_2$, i.e. at $x = 0.9$, then the problem will become a Neumann-Dirichlet problem.
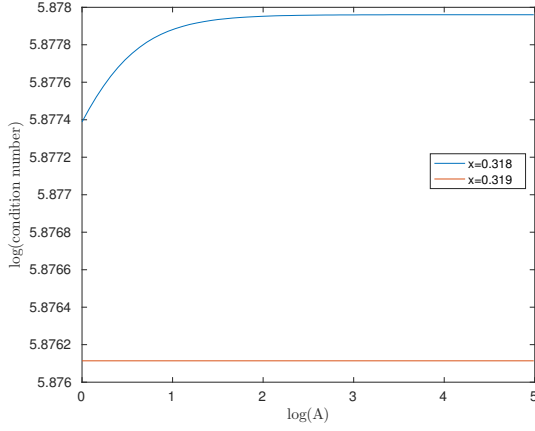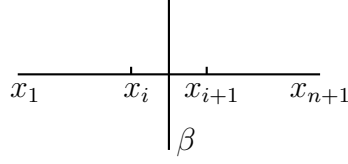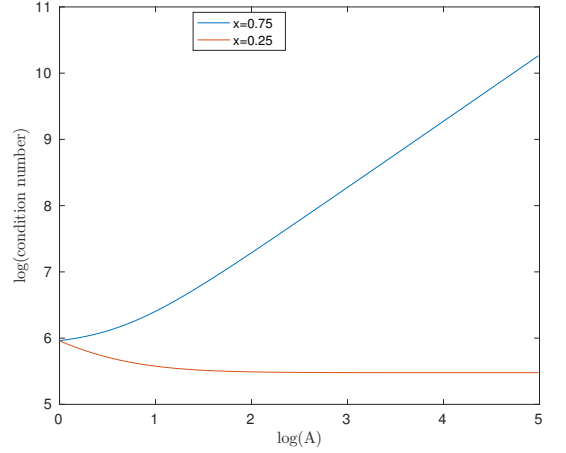
22
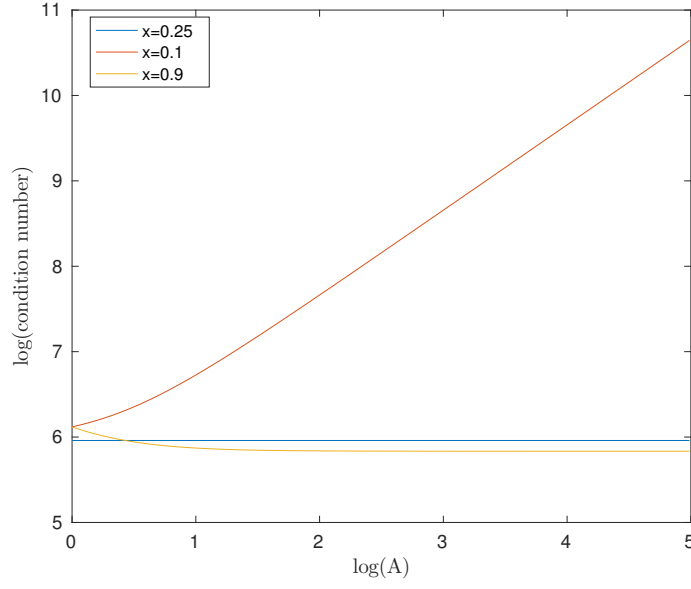
Figure 23: Scaled condition number, $h = 10^{-3}, \beta = \frac{1}{4}$, Neumann-Neumann boundary

The result is similar for interface $\beta = \frac{1}{\pi}$. But instead of putting the constraint on the interface, we put it near the interface, i.e. at $x_i$ or $x_{i+1}$.
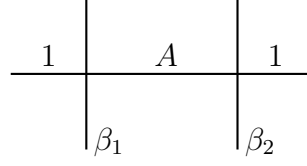


(a) Constraint near the interface

(b) Constraint away from the interface

Figure 24: Scaled condition number, $h = 10^{-3}, \beta = \frac{1}{\pi}$, Neumann-Neumann boundary

## 2.8 Multiple interfaces

Consider a uniform mesh on $[0, 1]$ with two interfaces, $\beta_1 = \frac{1}{4}$ and $\beta_2 = \frac{3}{4}$



As before, we first scale the global stiffness matrix and then let the contrast $A$ goes to infinity. We obtain

$$\lim_{A \to \infty} \mathbf{K}^* = \begin{bmatrix} 1 & -\frac{1}{\sqrt{2}} & & & & & & \\ & \ddots & \ddots & & & & & \\ & -\frac{1}{2} & 1 & & & & & \\ & & & 1 & -\frac{1}{\sqrt{2}} & & & \\ & & & & \ddots & \ddots & & \\ & & & & -\frac{1}{2} & 1 & & \\ & & & & & & 1 & -\frac{1}{2} \\ & & & & & & & \ddots & \ddots \\ & & & & & & & -\frac{1}{2} & 1 \end{bmatrix}$$

Similar to the case with one interface, the scaled global stiffness matrix is divided into three completely separated sub-matrices

$$\mathbf{K}^*_{A \to \infty} = \begin{bmatrix} \mathbf{K}^*_1 & & \\ \hline & \mathbf{K}^*_2 & \\ \hline & & \mathbf{K}^*_3 \end{bmatrix}$$

In this case, if we only put constraints on the endpoints $x = 0$ and $x = 1$, the sub-matrix $\mathbf{K}^*_2$ will always be a Neumann-Neumann problem, and thus the scaled condition number of $\mathbf{K}^*_2$ is unbounded, and hence the scaled condition number of the global stiffness matrix goes to infinity, as illustrated by the following graphs.
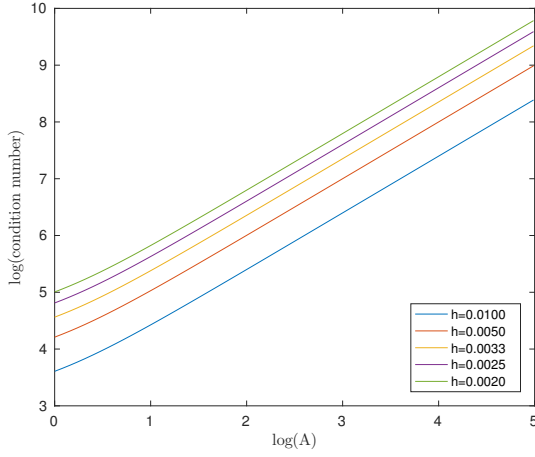
(a) Dirichlet-Dirichlet boundary
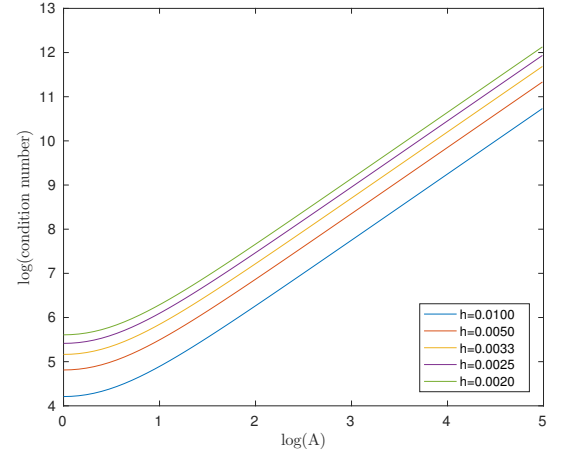


(b) Dirichlet-Neumann boundary
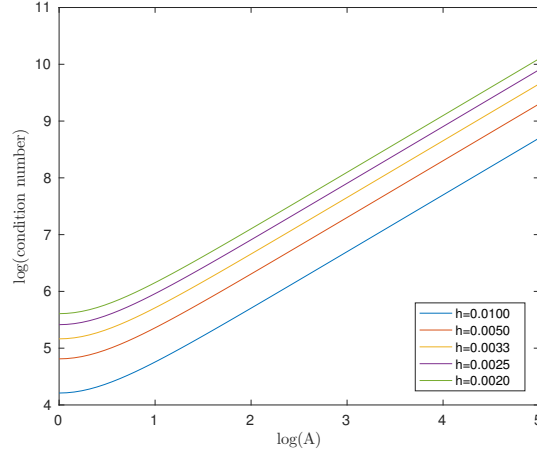


(c) Neumann-Dirichlet boundary

Figure 25: Scaled condition number, $\beta_1 = \frac{1}{4}$, $\beta_2 = \frac{3}{4}$

If we put constraints on the endpoints ($x = 0$ and $x = 1$), either Dirichlet or Neumann, then an additional constraint has to be imposed at $\beta_1 = \frac{1}{4}$ in order to obtain an analog with previous problems.

To obtain a bounded scaled condition number, we need to put an additional constraint on one of the interfaces. In the following graph, constraints are put on both the endpoints and on one of the interfaces, i.e. $x = 0$, $x = 1$, and $x = \frac{1}{4}$.
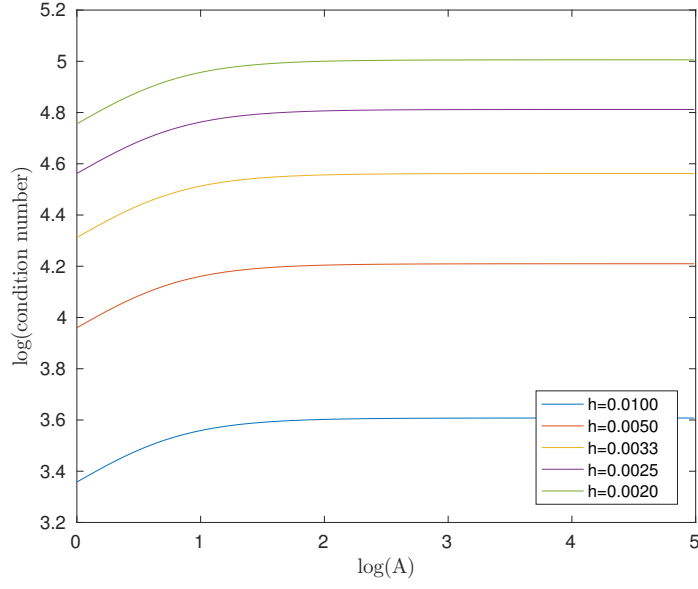
Figure 26: Scaled condition number, $\beta_1 = \frac{1}{4}$, $\beta_2 = \frac{3}{4}$

# 3 Two-dimensional interface problem

## 3.1 Construction of global stiffness matrix

Consider a uniform mesh on $\Omega = [0,1] \times [0,1]$
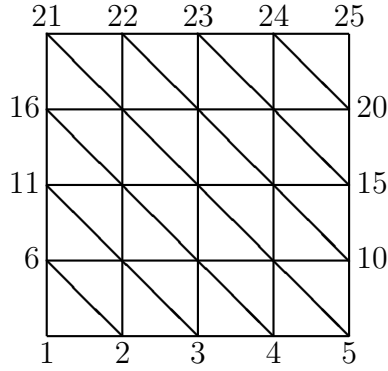


In this case, the region $\Omega$ is divided into 32 elements and 25 nodal points, with $h = 0.2$. The boundary $\partial\Omega$ is the boundary of the square, namely

$$\partial\Omega = \begin{cases} x = 0 & 0 \leq y \leq 1 \\ x = 1 & 0 \leq y \leq 1 \\ y = 0 & 0 \leq x \leq 1 \\ y = 1 & 0 \leq x \leq 1 \end{cases}$$

In theory, we should put Dirichlet or Neumann boundary on all the points on $\partial\Omega$, but in order to show the growth of the scaled condition number, several points are enough.

Consider a master element on $\Omega$
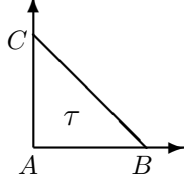


where

$$N_A = 1 - x - y$$

$$N_B = x$$

$$N_C = y$$

Obviously,

$$N_A N_A = \iint_\tau (\frac{\partial N_A}{\partial x})^2 + (\frac{\partial N_A}{\partial y})^2 dA = 1$$

$$N_A N_B = \iint_\tau \frac{\partial N_A}{\partial x}\frac{\partial N_B}{\partial x} + \frac{\partial N_A}{\partial y}\frac{\partial N_B}{\partial y} dA = -\frac{1}{2}$$

Similarly, compute all the other values $N_A N_C = -\frac{1}{2}$, $N_B N_A = -\frac{1}{2}$, $N_B N_B = \frac{1}{2}$, $N_B N_C = 0$, $N_C N_A = -\frac{1}{2}$, $N_C N_B = 0$, and $N_C N_C = \frac{1}{2}$.

Thus, we can construct the local stiffness matrix

$$\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

From here, it is easy to construct the global stiffness matrix (see reference 1 for more details).
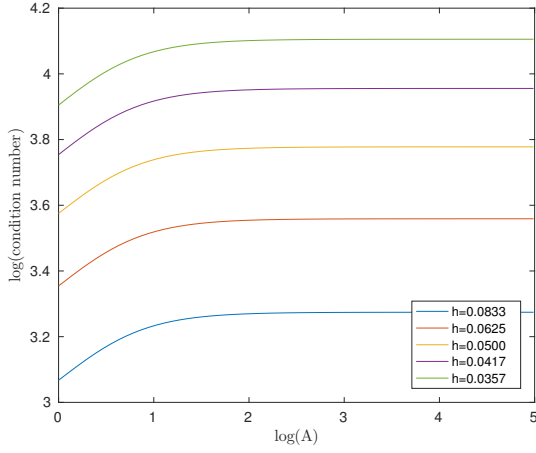
## 3.2 Linear interface

Consider a uniform mesh on $\Omega$ with a linear interface $\beta = \frac{1}{4}$. We take $a(x, y) = 1$ when $(x, y)$ is on the left of the interface (excluding); and $a(x, y) = A$ otherwise.
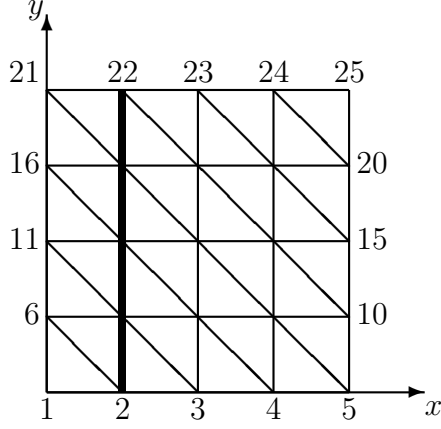
Like in one-dimensional problem, we consider the scaled condition number of the global stiffness matrix. First we put Neumann boundary on all $\partial\Omega$, i.e. nodal points 1, 2, 3, 4, 5, 6, 10, 11, 15, 16, 20, 21, 22, 23, 24, 25 in the above figure. Then we impose constraints on some points. By saying Dirichlet-Dirichlet boundary, we mean putting constraints on points $(0,0)$ and $(1,1)$. Likewise, we put a constraint on point $(0,0)$ to simulate Dirichlet-Neumann boundary, and we put a constraint on point $(1,1)$ to simulate Neumann-Dirichlet boundary.

As we can see in the following figure, the two-dimensional problem is a complete analog to the one-dimensional problem. Also, in order to obtain a bounded scaled condition number, we can put the constraint on the interface (only one point is enough).
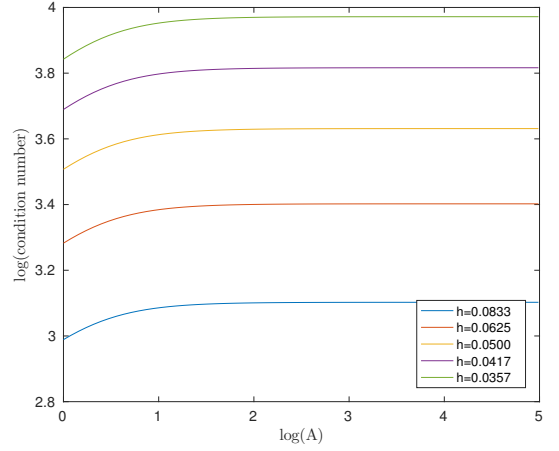
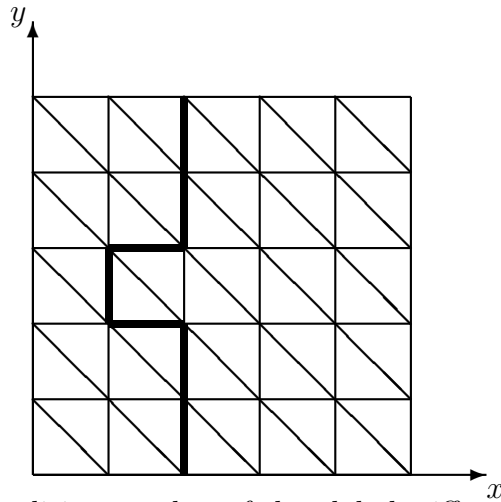(a) Dirichlet-Dirichlet boundary
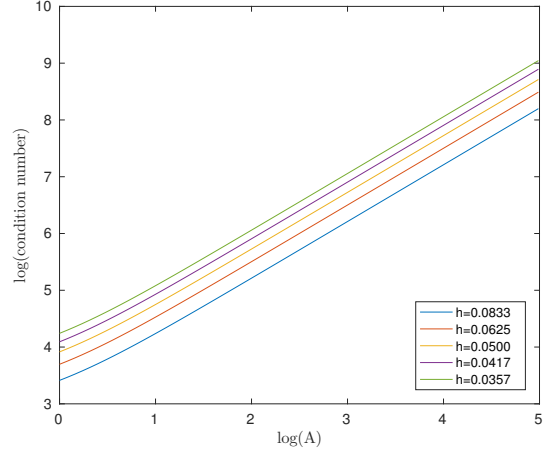


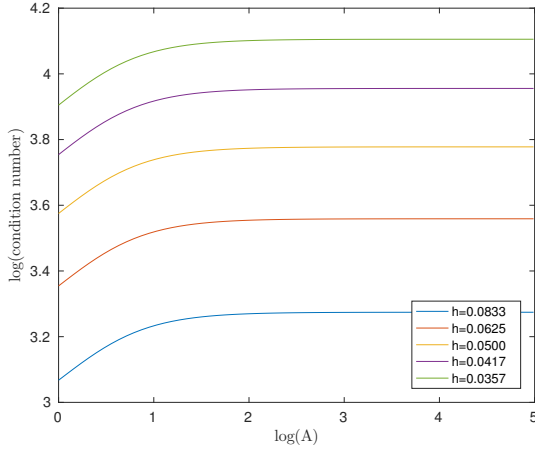(b) Dirichlet-Neumann boundary
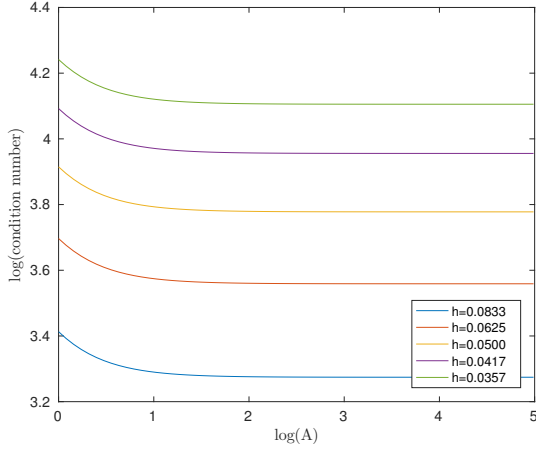


(c) Neumann-Dirichlet boundary



(d) Constraint on $\left(\frac{1}{4}, \frac{1}{4}\right)$

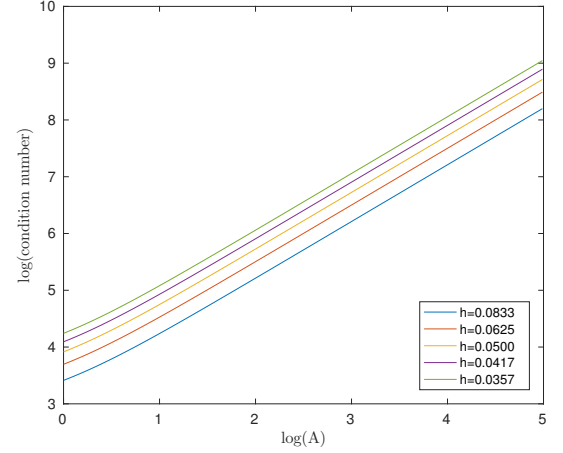Figure 27: Scaled condition number, linear interface $\beta = \frac{1}{4}$

Likewise, consider a pseudo-linear interface where there is a distortion in the middle of $\beta$.
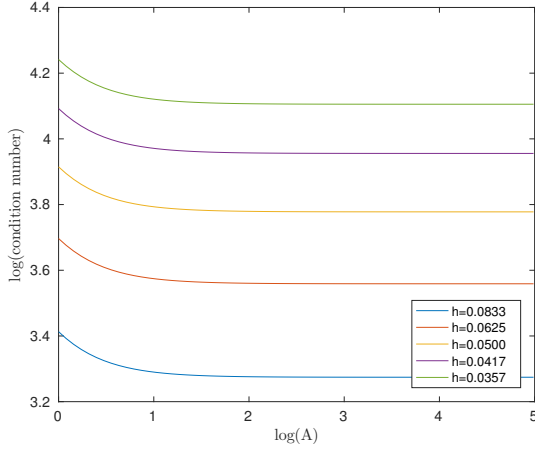


Still, we consider the scaled condition number of the global stiffness matrix.
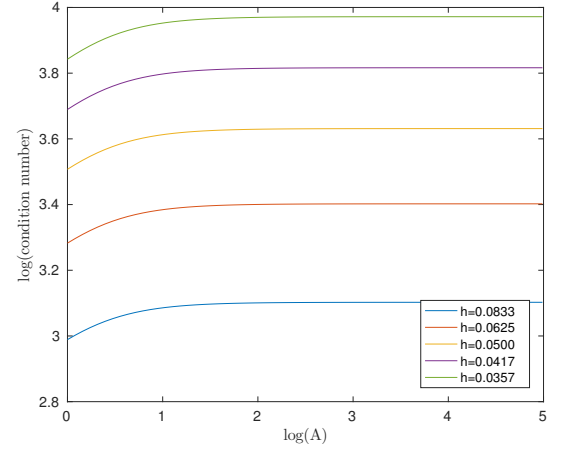
(a) Dirichlet-Dirichlet boundary

(b) Dirichlet-Neumann boundary

(c) Neumann-Dirichlet boundary

(d) Constraint on interface

Figure 28: Scaled condition number, pseudo-linear interface

Easy to spot that the scaled condition number behaves the same as a completely linear interface, and thus a complete analog with the one-dimensional case.
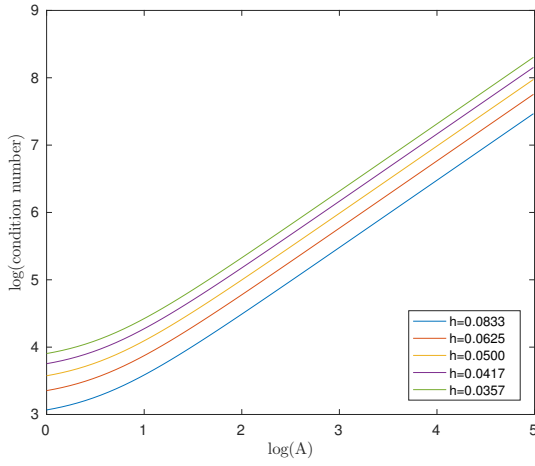
## 3.3  Square interface

Consider a square interface on $\Omega$. Here the interface $\beta$ is defined as

$$\beta = \begin{cases} x = \frac{1}{4} & \frac{1}{4} \le y \le \frac{3}{4} \\ x = \frac{3}{4} & \frac{1}{4} \le y \le \frac{3}{4} \\ y = \frac{1}{4} & \frac{1}{4} \le x \le \frac{3}{4} \\ y = \frac{3}{4} & \frac{1}{4} \le x \le \frac{3}{4} \end{cases}$$

As illustrated by the following figure, consider a uniform mesh on $\Omega$. We take $a(x, y) = A$ when $(x, y)$ is inside the interface (excluding); and $a(x, y) = 1$ otherwise.
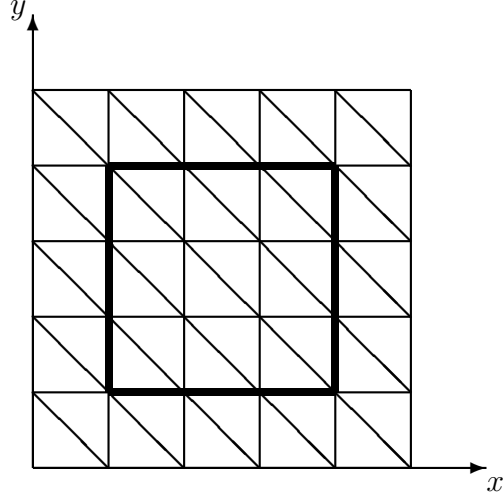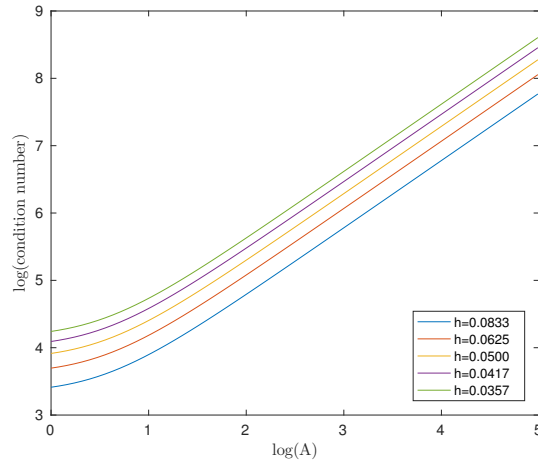
Consider the scaled condition number of the global stiffness matrix. Similar to the linear interface problem, we first put Neumann boundary on all the boundary nodal points, then we impose some constraints.



(a) Dirichlet-Dirichlet boundary
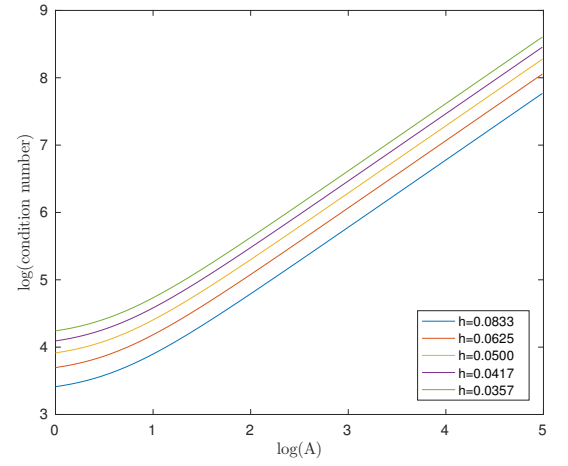


(b) Dirichlet-Neumann boundary



(c) Neumann-Dirichlet boundary

Figure 29: Scaled condition number, square interface $\beta$

Obviously if we only put constraints on the endpoints, i.e. on points $(0,0)$ or $(1,1)$, then the scaled condition number will be unbounded. Again, by saying Dirichlet-Dirichlet boundary, we mean putting constraints on points $(0,0)$ and $(1,1)$. Likewise, we put a constraint on point $(0,0)$ to simulate Dirichlet-Neumann boundary, and we put a constraint on point $(1,1)$ to simulate Neumann-Dirichlet boundary.

However, like in the linear interface two-dimensional problem, if we put a constraint only on the interface, the scaled condition number is bounded.
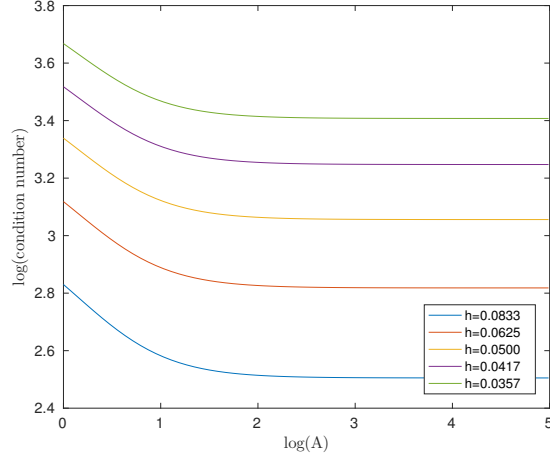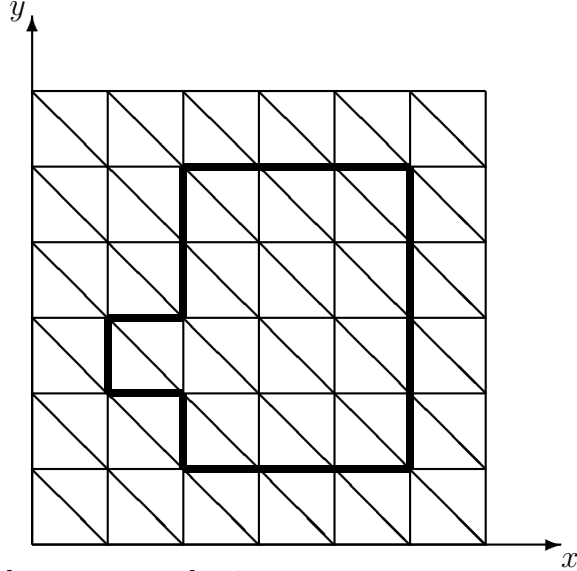


Figure 30: Scaled condition number, square interface $\beta$, constraint on $(\frac{1}{4}, \frac{1}{4})$

Similarly, consider a pseudo-square interface that has a distortion on $\beta$.



Like in the completely circular case, we obtain

(a) Dirichlet-Dirichlet boundary

(b) Dirichlet-Neumann boundary

(c) Neumann-Dirichlet boundary

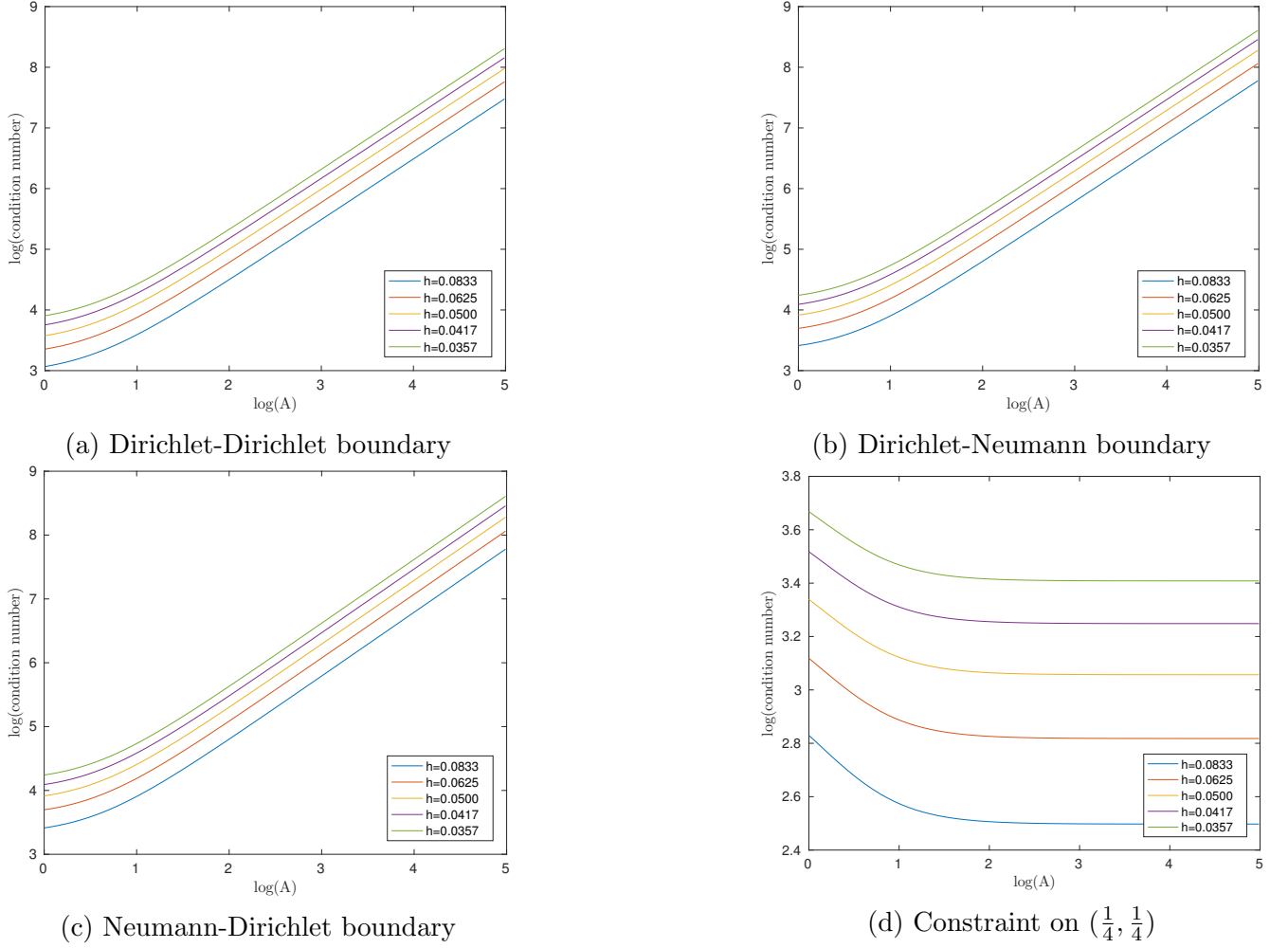(d) Constraint on $\left(\frac{1}{4}, \frac{1}{4}\right)$

Figure 31: Scaled condition number, pseudo-square interface

# 4 Conclusion

From the discussion above, we can see that for the contrast problem, the behavior of the FEM solution depends significantly on the boundary conditions we impose, i.e. Dirichlet boundary or Neumann boundary.

Specifically, for an equation with large contrast, imposing Neumann boundary can cause the scaled condition number of the global stiffness matrix to be unbounded, and consequently cause the FEM solution to be inaccurate. On the other hand, imposing Dirichlet boundary or simply put a constraint on the interface will result in a bounded condition number.

# Acknowledgment

The author would like to thank Dr.Ivo Babuska for his patience and great help throughout this project.

# References

Babuska, Ivo. *Finite Elements, An Introduction to the Method and Error Estimation*