714 Computational Math Homework 2

Varun Gudibanda

here is github: https://github.com/VarunMG/714Homework2

Problem A

a) Since $v \in span\{w_1, w_2, \dots, w_n\}$ then

$$v = \sum_{i=1}^{n} c_i w_i$$

where c_i are currently unknown coefficients. Let's look at $\langle v, w_k \rangle$ where $1 \le k \le n$ so

$$\begin{aligned} \langle v, w_k \rangle &= \langle \sum_{i=1}^{n} c_i w_i, w_k \rangle \\ &= \sum_{i=1}^{n} c_i \langle w_i, w_k \rangle \end{aligned}$$

Using the orthogonality condition, we know that $\langle w_i, w_k \rangle$ is non-zero only when $i = k$ hence

$$\sum_{i=1}^{n} c_i \langle w_i, w_k \rangle = c_k \langle w_k, w_k \rangle = c_k \left\| w_k \right\|^2$$

so we have found $\langle v, w_k \rangle = c_k \left\| w_k \right\|^2 \implies c_k = \frac{\langle v, w_k \rangle}{\left\| w_k \right\|^2}$. Hence

$$v = \sum_{i=1}^{n} c_i w_i = \sum_{i=1}^{n} \frac{\langle v, w_i \rangle}{\left\| w_i \right\|^2} w_i$$

which we wanted to show. ⊠

b)
(i) If our initial guess is the exact solution, then the CG method converges immediately and hence can have $n^* < N$.

(ii) We can use strong induction. For the base case with $n = 1$ Note that we must have then that $j = 0$ hence we want to show $\langle p_1, p_0 \rangle_A = 0$. Since $p_1 = r_1 - \frac{\langle r_1, p_0 \rangle_A}{\|p_0\|_A^2} p_0$. So from this we see

$$
\begin{aligned}
\langle p_1, p_0\rangle_A &= \langle r_1, p_0\rangle_A - \frac{\langle r_1, p_0\rangle_A}{\|p_0\|_A^2}\langle p_0, p_0\rangle_A \\
&= \langle r_1, p_0\rangle_A - \frac{\langle r_1, p_0\rangle_A}{\|p_0\|_A^2}\|p_0\|_A^2 \\
&= \langle r_1, p_0\rangle_A - \langle r_1, p_0\rangle_A = 0
\end{aligned}
$$

so we see that indeed p_1 and p_0 are A-orthogonal. For the induction step we assume that for all $k \leq n$ that $\langle p_k, p_j\rangle = 0$ for $0 \leq j < k \leq n^* - 1$. We know that $p_{n+1} = r_{n+1} - \sum_{j=0}^{n} \frac{\langle r_{n+1}, p_j\rangle_A}{\|p_j\|_A^2} p_j$. Therefore for $0 \leq i < n+1 \leq n^* - 1$,

$$
\langle p_{n+1}, p_i\rangle_A = \langle r_{n+1}, p_i\rangle_A - \sum_{j=0}^{n} \frac{\langle r_{n+1}, p_j\rangle_A}{\|p_j\|_A^2}\langle p_j, p_i\rangle
$$

Note that if $i = n$ then $\langle p_j, p_i\rangle_A = 0$ by the induction hypothesis except for when $j = i = n$ and similarly if $i < n$ then also by the induction hypothesis we know $\langle p_i, p_j\rangle_A = 0$ except for when $j = i$. Therefore we see that in general that $\langle p_i, p_j\rangle_A = 0$ except for when $j = i$ in which case we have $\langle p_i, p_i\rangle_A = \|p_i\|_A^2$ so

$$
\langle p_{n+1}, p_i\rangle_A = \langle r_{n+1}, p_i\rangle_A - \frac{\langle r_{n+1}, p_i\rangle_A}{\|p_i\|_A^2}\|p_i\|_A^2 = \langle r_{n+1}, p_i\rangle_A - \langle r_{n+1}, p_i\rangle_A = 0
$$

Therefore we see that $\langle p_{n+1}, p_i\rangle_A = 0$ for $0 \leq i < n+1 \leq n^* - 1$. So by induction we have shown the claim. ⊠

c)

(i) Let $v = \sum_{i=1}^{n} c_i\phi_i$ and $w = \sum_{i=1}^{n} d_i\phi_i$ (which is possible since $\{\phi_i\}$ are a basis). Note that by orthonormality

$$
\begin{aligned}
\langle v, \phi_k\rangle &= \sum_{i=1}^{n} c_i\langle \phi_k, \phi_i\rangle = \sum_{i=1}^{n} c_i\delta_{ik} = c_k \\
\langle w, \phi_k\rangle &= \sum_{i=1}^{n} d_i\langle \phi_k, \phi_i\rangle = \sum_{i=1}^{n} c_i\delta_{ik} = d_k
\end{aligned}
$$

Furthermore,

$$
Av = \sum_{i=1}^{n} c_i A\phi_i = \sum_{i=1}^{n} c_i\lambda_i\phi_i
$$

Putting this all together we have

$$
\begin{aligned}
\langle Av, w\rangle &= \langle \sum_{i=1}^{n} c_i \lambda_i \phi_i, \sum_{i=1}^{n} d_i \phi_i \rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i c_i d_j \langle \phi_i, \phi_j \rangle \\
&= \sum_{i=1}^{n} \left(\lambda_i c_i \sum_{j=1}^{n} d_j \delta_{ij} \right) \\
&= \sum_{i=1}^{n} \lambda_i c_i d_i \\
&= \sum_{i=1}^{n} \lambda_i \langle v, \phi_i \rangle \langle w, \phi_i \rangle
\end{aligned}
$$

which shows the claim. ⊠

(ii) We know that $A\phi_i = \lambda_i \phi_i$ therefore

$$
\langle \phi_i, A\phi_i \rangle = \lambda_i \langle \phi_i, \phi_i \rangle \implies \langle \phi_i, A\phi_i \rangle = \lambda_i
$$

Since A is symmetric positive definite, then $\langle \phi_i, A\phi_i \rangle > 0$ and therefore $\lambda_i > 0$. ⊠

(iii) We can write $v = \sum_{i=1}^{n} c_i \phi_i$ and using what we had in part (ii) of this very subproblem, we also know $Av = \sum_{i=1}^{n} c_i \lambda_i \phi_i$ so

$$
\begin{aligned}
\langle Av, v\rangle &= \langle \sum_{i=1}^{n} c_i \lambda_i \phi_i, \sum_{i=1}^{n} c_i \phi_i \rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i c_i c_j \langle \phi_i, \phi_j \rangle \\
&= \sum_{i=1}^{n} \left(\lambda_i c_i \sum_{j=1}^{n} c_j \delta_{ij} \right) \\
&= \sum_{i=1}^{n} \lambda_i c_i^2
\end{aligned}
$$

From here, we note that since $\lambda_i \leq \lambda_N$, then $\sum_{i=1}^{n} \lambda_i c_i^2 \leq \lambda_N \sum_{i=1}^{n} c_i^2$ and similarly since $\lambda_i \geq \lambda_1$, then $\sum_{i=1}^{n} \lambda_i c_i^2 \geq \lambda_1 \sum_{i=1}^{n} c_i^2$. And finally, $\|v\|^2 = \sum_{i=1}^{n} c_i^2$ since the ϕ_i are orthonormal hence we have

$$
\lambda_1 \|v\|^2 \leq \langle Av, v \rangle \leq \lambda_N \|v\|^2
$$

and that shows the claim. ⊠

(iv) We can write $v = \sum_{i=1}^{n} c_i \phi_i$ and using what we had in part (ii) of this very subproblem, we also know $Av = \sum_{i=1}^{n} c_i \lambda_i \phi_i$ so

$$
\begin{aligned}
\|Av\|^2 = \langle Av, Av\rangle &= \langle \sum_{i=1}^{n} c_i\lambda_i\phi_i, \sum_{i=1}^{n} c_i\lambda_i\phi_i\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i c_i \lambda_j c_j \langle \phi_i, \phi_j\rangle \\
&= \sum_{i=1}^{n}\left(\lambda_i c_i \sum_{j=1}^{n} \lambda_j c_j \delta_{ij}\right) \\
&= \sum_{i=1}^{n} c_i^2\lambda_i^2
\end{aligned}
$$

Since $\lambda_i \leq \lambda_N$, then $\sum_{i=1}^{n} c_i^2\lambda_i^2 \leq \lambda_N^2 \sum_{i=1}^{n} c_i^2$ and as we noted before, $\|v\|^2 = \sum_{i=1}^{n} c_i^2$ therefore putting this all together

$$
\|Av\|^2 = \sum_{i=1}^{n} c_i^2\lambda_i^2 \leq \lambda_N^2 \sum_{i=1}^{n} c_i^2 = \lambda_N^2 \|v\|^2
$$

Since all quantities are positive then taking the square root gives $\|Av\| \leq \lambda_n \|v\|$ which proves the claim. ⊠

d) We know that $r_{n+1} = r_n - \alpha_n w_n$ and $p_{n+1} = r_{n+1} + \beta_n p_n$ so putting these two together and using $w_n = Ap_n$

$$
p_{n+1} = r_n - \alpha_n A p_n + \beta_n p_n
$$

Now we use that $r_n = p_n - \beta_{n-1}p_{n-1}$ to see that

$$
\begin{aligned}
p_{n+1} &= p_n - \beta_{n-1}p_{n-1} - \alpha_n A p_n + \beta_n p_n \\
&= (1+\beta_n)p_n - \alpha_n A p_n - \beta_{n-1}p_{n-1}
\end{aligned}
$$

which shows the claim. ⊠

e) Cayley-Hamilton tells us that there is a monic polynomial p such that $p(A) = 0$ which can be written as

$$
p(A) = A^n + c_{n-1}A^{n-1} + c_{n-2}A^{n-2} + \cdots + c_1 A + c_0 \det(A) I_n = 0
$$

where the coefficients come from the characteristic polynomial of A and 0 is the zero matrix. Since A is non-singular then $\det(A) \neq 0$ so there is a non-trivial contribution from I. Rearranging this equation gives

$$
A^n = -c_{n-1}A^{n-1} - \cdots - c_1 A - c_0 \det(A) I_n
$$

This shows the claim. ⊠

f)
(i) Note that since u is the true solution then $f - Au = 0$. Using this and also subtracting u from both sides of the scheme gives

$$
\begin{aligned}
u_{n+1} - u &= u_n - u + \alpha(f - Au_N) - \alpha(f - Au) \\
\implies e_{n+1} &= e_n + \alpha\left[(f - Au_n) - (f - Au)\right] \\
\implies e_{n+1} &= e_n + \alpha\left[-A(u_n - u)\right] \\
\implies e_{n+1} &= e_n - \alpha A e_n \\
\implies e_{n+1} &= (I - \alpha A)e_n
\end{aligned}
$$

We have proven the claim. ⊠

(ii) Taking the 2-norm of the result from subpart (i) of this problem, we wee that $\|e_{n+1}\|_2 = \|(I-\alpha A)e_n\|_2 \leq \|I - \alpha A\|_2 \|e_n\|_2$. The $2-norm$ is the spectral radius hence $\|I-\alpha A\|_2 = \rho(I-\alpha A)$ so we see that $\|e_{n+1}\|_2 \leq \rho(I-\alpha A)\|e_n\|_2$. The problem statement defines $\rho = \rho(I-\alpha A)$ so we have $\|e_{n+1}\|_2 \leq \rho\|e_n\|_2$. It is easy to see that if v, λ are an eigenvector/eigenvalue pair then it is clear that $(I-\alpha A)v = (1-\alpha\lambda)v$ so $1-\alpha\lambda$ is an eigenvalue of $I-\alpha A$. So $\rho = \max_{1\leq j\leq N}|1-\alpha\lambda_j|$. This shows the claim. ⊠

(iii) Order the eigenvalues of A as $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Therefore it is clear that $\max_{1\leq j\leq N}|1-\alpha\lambda_j| = \max\{|1-\alpha\lambda_1|, |1-\alpha\lambda_N|\}$. The optimal α for which this works is $\alpha^* = argmin_\alpha \max_{1\leq j\leq N}\{|1-\alpha\lambda_1|, |1-\alpha\lambda_N|\}$. Looking at the function $f(x,y) = \max\{x,y\}$, we know that this is minimized if $y = x$. Hence we need $|1-\alpha\lambda_1| = |1-\alpha\lambda_N|$. This happens either when $1-\alpha\lambda_1 = 1-\alpha\lambda_N$ or if $1-\alpha_1 = -(1-\alpha\lambda_N)$.
In the first case if $1-\alpha\lambda_1 = 1-\alpha\lambda_N$ then we see that we must have $\lambda_1 = \lambda_N$ which in general is not true hence we ignore this.
In the second case if $1-\alpha_1 = -(1-\alpha\lambda_N)$ then we can rearrange to see that $\alpha = \frac{2}{\lambda_1+\lambda_N}$ is the optimal.

So we have that $\alpha = \frac{2}{\lambda_1+\lambda_N}$ is the optimal α and since $\rho = \max\{|1-\alpha\lambda_1|, |1-\alpha\lambda_N|\}$ and α was constructed such that $|1-\alpha\lambda_1| = |1-\alpha\lambda_N|$ then we have that

$$
\rho = |1-\alpha\lambda_1| = \left|1 - \frac{2}{\lambda_1+\lambda_N}\lambda_1\right| = \left|\frac{\lambda_1+\lambda_N}{\lambda_1+\lambda_N} - \frac{2\lambda_1}{\lambda_1+\lambda_N}\right| = \frac{\lambda_N-\lambda_1}{\lambda_1+\lambda_N} = \frac{\lambda_1\left(\frac{\lambda_N}{\lambda_1}-1\right)}{\lambda_1\left(\frac{\lambda_N}{\lambda_1}+1\right)} = \frac{\kappa-1}{\kappa+1}
$$

It is clear that $\kappa - 1 < \kappa + 1$ so $\rho < 1$. This proves the claim. ⊠

(iv) Note that since $C \geq \lambda_N$ and $\frac{1}{c} \geq \frac{1}{\lambda_1}$ then we have $\frac{C}{c} \geq \frac{\lambda_N}{\lambda_1}$. We can start from this and notice that all quantities are strictly positive and do some tricks:

$$
\begin{aligned}
\frac{C}{c} &\geq \frac{\lambda_N}{\lambda_1} \\
C\lambda_1 - c\lambda_N &\geq 0 \\
2C\lambda_1 - 2c\lambda_N &\geq 0 \\
C\lambda_1 - c\lambda_N &\geq -C\lambda_1 + c\lambda_N
\end{aligned}
$$

Now we can add the terms $C\lambda_N$ and $-c\lambda_1$ to both sides

$$
\begin{aligned}
C\lambda_N + C\lambda_1 - c\lambda_N - c\lambda_1 &\geq C\lambda_N - C\lambda_1 - c\lambda_N - c\lambda_1 \\
(C-c)(\lambda_N+\lambda_1) &\geq (C+c)(\lambda_N-\lambda_1) \\
\frac{C-c}{C+c} &\geq \frac{\lambda_N-\lambda_1}{\lambda_N+\lambda_1}
\end{aligned}
$$

So from this we see that indeed $\rho \leq \frac{C-c}{C+c}$ and by very similar work as the previous subpart of this same problem we see that $\frac{C-c}{C+c} = \frac{\kappa'-1}{\kappa'+1} < 1$. This shows the claim. ⊠

g)
(i) We know that $r_k = r_{k-1} - \alpha_{k-1} w_{k-1}$ and since $w_{k-1} = Ap_{k-1}$ then we have $r_k = r_{k=1} - \alpha_{k-1} Ap_{k-1}$. When $k = 1$ we have $r_1 = r_0 - \alpha_0 Ap_0$ and since $p_0 = r_0$ we have that $r_1 = r_0 - \alpha_0 Ar_0$.

(ii) We know that $r_{n+1} = r_n - \alpha_n Ap_n$ and since $p_n = r_n + \beta_{n-1} p_{n-1}$ then we see that

$$r_{n+1} = r_n - \alpha_n A(r_n + \beta_{n-1} p_{n-1}) = r_n - \alpha_n Ar_n - \alpha_N \beta_N Ap_{n-1}$$

Since $r_n = r_{n-1} - \alpha_{n-1} Ap_{n-1}$ then we see that $Ap_{n-1} = -\frac{r_n - r_{n-1}}{\alpha_{n-1}}$. Using this, we see that

$$r_{n+1} = r_n - \alpha_n Ar_n - \alpha_n \beta_N Ap_{n-1} = r_n - \alpha_n Ar_n + \frac{\alpha_n \beta_{n-1}}{\alpha_{n-1}}(r_n - r_{n-1})$$

which shows the claim. ⊠

(iii) Note that $r_n = \|r_n\| q_n$ and $\sqrt{\beta_n} = \frac{\|r_{n+1}\|}{\|r_n\|}$. In the previous sub-part of this problem we saw that $r_1 = r_0 - \alpha_0 Ap_0$. Using this, we note that

$$\begin{aligned} \|r_1\| q_1 &= \|r_0\| q_0 - \alpha_0 \|r_0\| Aq_0 \\ \implies Aq_0 &= \frac{q_0}{\alpha_0} - \frac{1}{\alpha_0} \frac{\|r_1\|}{\|r_0\|} q_1 \\ &= \gamma_0 q_0 - \frac{\sqrt{\beta_1}}{\alpha_0} q_1 \\ &= \gamma_0 q_0 - \delta_0 q_1 \end{aligned}$$

So we see that part is true. For the next part,

$$\begin{aligned} r_{n+1} &= r_n - \alpha_n Ar_n + \frac{\alpha_n \beta_{n-1}}{\alpha_{n-1}}(r_n - r_{n-1}) \\ \implies \|r_{n+1}\| q_{n+1} &= \|r_n\| q_n - \alpha_n \|r_n\| Aq_n + \frac{\alpha_n \beta_{n-1}}{\alpha_{n-1}} (\|r_n\| q_n - \|r_{n-1}\| q_{n-1}) \\ \implies Aq_n &= \frac{q_n}{\alpha_n} - \frac{1}{\alpha_n} \frac{\|r_{n+1}\|}{\|r_n\|} q_{n+1} + \frac{\beta_{n-1}}{\alpha_{n-1}} q_n - \frac{\|r_{n-1}\|}{\|r_n\|} \frac{\beta_{n-1}}{\alpha_{n-1}} q_{n-1} \\ &= \left(\frac{1}{\alpha_n} + \frac{\beta_{n-1}}{\alpha_{n-1}} \right) q_n - \delta_n q_{n+1} - \frac{1}{\sqrt{\beta_{n-1}}} \frac{\beta_{n-1}}{\alpha_{n-1}} q_{n-1} \end{aligned}$$

Now we can use the definitions of quantities given to us in the question and see that

$$Aq_n = -\delta_{n-1} q_{n-1} + \gamma_n q_n - \delta_n q_{n+1}$$

and that shows the claim. ⊠

(iv) Note that

$$
\begin{aligned}
AQ_n &= [Aq_0 \quad Aq_1 \quad \dots \quad Aq_{n-1}] \\
&= [\gamma_0 q_0 - \delta_0 q_1 \quad -\delta_0 q_0 + \gamma_1 q_1 - \delta_1 q_1 \quad \dots \quad -\delta_{n-1} q_{n-1} + \gamma_{n-1} q_{n-1} - \delta_{n-1} q_n] \\
&= [\gamma_0 q_0 - \delta_0 q_1 \quad -\delta_0 q_0 + \gamma_1 q_1 - \delta_1 q_1 \quad \dots \quad -\delta_{n-1} q_{n-1} + \gamma_{n-1} q_{n-1}] \\
&- \delta_{n-1} [0 \quad 0 \quad \dots \quad q_n]
\end{aligned}
$$

Note that when doing matrix multiplication, AB, the i-th column of the product is found by taking the linear combination columns of A using the elements of the i-th row of B as coefficients of the linear combination. Thus, we see that

$$[\gamma_0 q_0 - \delta_0 q_1 \quad -\delta_0 q_0 + \gamma_1 q_1 - \delta_1 q_1 \quad \dots \quad -\delta_{n-1} q_{n-1} + \gamma_{n-1} q_{n-1}] = Q_n T_n$$

With the exact same reasoning, we can write

$$-\delta_{n-1} [0 \quad 0 \quad \dots \quad q_n] = -\delta_{n-1} q_n e_n^T$$

Therefore, we have that $AQ_n = Q_n T_n - \delta_{n-1} q_n e_n^T$.

(v) We can do left multiplication by Q_n^T of the equation $AQ_n = Q_n T_n - \delta_{n-1} q_n e_n^T$ to see that

$$Q_n^T A Q_n = Q_n^T Q_n T_n - \delta_{n-1} Q_n^T q_n e_n^T$$

Since Q_n is orthogonal then $Q_n^T Q_n = I_n$. Note that in the product of a matrix with a vector Av, the i-th entry of the resulting vector is found by taking the dot product the i-th row of A with the vector v. The i-th row of Q_n^T is the vector q_{i-1} for $i = 1, ..., n$ so the i-th entry of the vector $Q_n^T q_n$ is the dot product $q_{i-1} \cdot q_n$ and since $i - 1 \neq n$ for $i = 1, ..., n$ then we know that $q_{i-1} \cdot q_n = 0$ hence the result will be the 0 vector. Therefore $Q_n^T q_N e_n^T = 0$. Therefore we see that

$$Q_n^T A Q_n = T_n - 0 = T_n$$

which proves the claim. ⊠

Problem B

So I calculated the values of f on a very fine grid of 10000 points. Then, I found a linear interpolant for f on a grid of size k and using interp1, interpolated it onto the fine grid of 10000 points and found the max norm of the error. If it did not achieve the desired tolerance, then k was increased. Once it was of proper tolerance, it quit and output the answer of 100.

Problem C

a) We use the following scheme:

$$\frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\Delta t^2} = \frac{u_{i+1,j}^n + u_{i-1,j}^n + u_{i,j+1}^n + u_{i,j-1}^n - 4u_{i,j}^n}{h^2}$$

Which we can solve around to be

$$u_{i,j}^{n+1} = r\left(\frac{u_{i+1,j}^n + u_{i-1,j}^n + u_{i,j+1}^n + u_{i,j-1}^n - 4u_{i,j}^n}{h^2}\right) + 2u_{i,j}^n - u_{i,j}^{n-1}$$

where $r = \frac{\Delta t^2}{h^2}$. We initialize the initial condition as $u_{i,j}^0 = 0$ and since $u_t(x_i, y_j, 0) \approx \frac{u_{i,j}^1 - u_{i,j}^0}{\Delta t} = f(x_i)f(y_j)$.

So

$$u_{i,j}^1 = u_{i,j}^0 + \Delta t f(x_i) f(y_j)$$

where we only update the inner grid points to maintain the dirichlet boundary data. Looking back at the first homework, we can write the discrete Laplacian as $B = kron(I, A) + kron(A, I)$ where A is the 1D Laplacian matrix $A = tridiag(1, -2, 1)$ which is the $N-1 \times N-1$ matrix corresponding to Dirichlet boundary conditions. And absorbing the $2u_{i,j}^n$ term we can write the update in vector form

$$u^{n+1} = (rB + I_{N-1})u^n - Iu^{n-1}$$

where we are only solving for the inner $N-1 \times N-1$ points. Here is error plot:

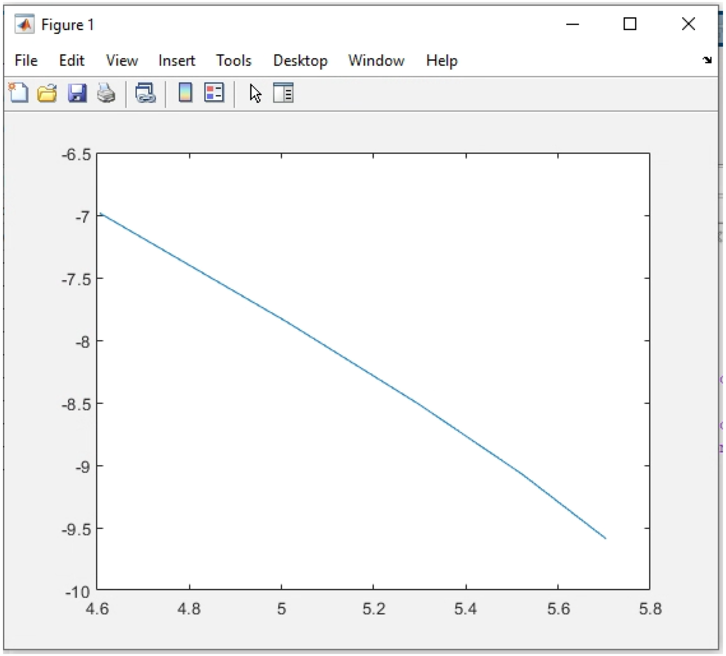

Figure 1: Error Plot

This was using N values of 100, 150, 200, 250, and 300. We get a slope of 2.35 which is around 2, so we get the right convergence.

b) We approximate the derivative with

$$y^{n+1} - 2y^n + y^{n-1} = \lambda \Delta t^2 y^n$$

And so rearranging we have

$$y^{n+1} - (2 + \lambda \Delta t^2) y^n + y^{n-1} = 0$$

From this we get the characteristic equation

$$\rho^2 - (2 + \lambda \Delta t^2)\rho + 1 = 0$$

Solving for ρ we get

$$\rho = \frac{(2+\lambda\Delta t^2) \pm \sqrt{(2+\lambda\Delta t^2)^2 - 4}}{2} = \left(1 + \frac{\lambda\Delta t^2}{2}\right) \pm \sqrt{\left(1 + \frac{\lambda\Delta t^2}{2}\right)^2 - 1}$$

Let $\alpha = \lambda\Delta t^2$ Then since we need $|\rho| \leq 1$ then we must impose

$$\left|(1+\alpha/2) \pm \sqrt{(1+\alpha/2)^2 - 1}\right| \leq 1$$

We can plot both of these in MATLAB. I have plotted here the absolute values of the positive and negative roots and also the plane $z = 1$.

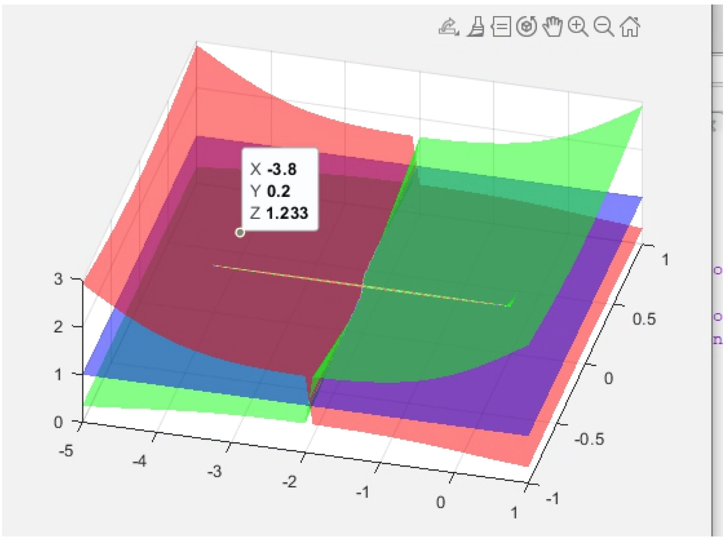

Figure 2: Error Plot

We see here that the condition is satisfied only on the real line with $-4 \leq \alpha \leq 0$. So we see then that we have to have $-4 \leq \lambda\Delta t^2 \leq 0$.

c) Using the method of lines, our semidiscrete scheme (discretized only in space but continuous in time) is

$$y''_{i,j}(t) = \frac{1}{h^2}\left(y_{i+1,j}(t) + y_{i-1,j}(t) + y_{i,j+1}(t) + y_{i,j-1}(t) - 4y_{i,j}(t)\right)$$

and in vector form we can write it as

$$y''(t) = By(t)$$

where B is the same matrix we had before, $B = kron(I, A) + kron(A, I)$ where A is the 1D Laplacian matrix $A = tridiag(1, -2, 1)$ which is the $N-1 \times N-1$ matrix corresponding to Dirichlet boundary conditions. So now discretizing the time derivative as before we have

$$\left(\frac{y^{n+1} - 2y^n + y^{n-1}}{\Delta t^2}\right) = By^n$$

We know that

$$\lambda_k(A) = -\frac{4}{h^2}\sin^2\left(\frac{k\pi h}{2}\right)$$

And since we know that

$$\lambda(B) = \lambda_i(A) + \lambda_j(A) = -\frac{4}{h^2}\left(\sin^2\left(\frac{i\pi h}{2}\right) + \sin^2\left(\frac{j\pi h}{2}\right)\right)$$

by properties of the Kronecker product then using the previous part we must have that

$$-4 \leq -\frac{4\Delta t^2}{h^2}\left(\sin^2\left(\frac{i\pi h}{2}\right) + \sin^2\left(\frac{j\pi h}{2}\right)\right) \leq 0$$

Which we can rearrange to

$$0 \leq \frac{4\Delta t^2}{h^2}\left(\sin^2\left(\frac{i\pi h}{2}\right) + \sin^2\left(\frac{j\pi h}{2}\right)\right) \leq 4$$

Noting that $\sin^2\left(\frac{i\pi h}{2}\right) + \sin^2\left(\frac{j\pi h}{2}\right) \leq 2$ then we can ask that

$$\frac{4\Delta t^2}{h^2}\cdot 2 \leq 4 \implies \frac{\Delta t^2}{h^2} \leq 2$$

this is our CFL condition. ⊠

d) A plane wave is $e^{ik_1x_j}e^{ik_2y_i}$. So if we put in $u_{i,j}^{n-1} = e^{ik_1x_j}e^{ik_2y_i}$ then considering an amplification factor $g(k)$ we will have

$$u^n = g(k_1,k_2)u^{n-1} \quad u^{n+1} = g(k_1,k_2)^2u^{n-1}$$

So with this we see that

$$\begin{aligned} g^2e^{ik_1x_j}e^{ik_2y_i} &= g\frac{\Delta t^2}{h^2}\left(e^{ik_1(x_j+h)}e^{ik_2y_i} + e^{ik_1(x_j-h)}e^{ik_2y_i} + e^{ik_1x_j}e^{ik_2(y_i+h)} + e^{ik_1x_j}e^{ik_2(y_i-h)}\right) \\ &+ 2e^{ik_1x_j}e^{ik_2y_i} - e^{ik_1x_j}e^{ik_2y_i} \end{aligned}$$

Dividing out by $e^{ik_1x_j}e^{ik_2y_i}$ gives

$$g^2 = g\frac{\Delta t^2}{h^2}\left(e^{ik_1h} + e^{-ik_1h} + e^{ik_2h} + e^{-ik_2h}\right) + 2g - 1$$

Using Euler's formula and expanding we can expand this out

$$g^2 = g\frac{\Delta t^2}{h^2}\left(2\cos(k_1h) + 2\cos(k_2h) + 2\right) + 1$$

Notice that $\cos(k_ih) = 1 - \sin^2(k_ih/2)$ And then putting this in gives

$$g^2 + g\frac{\Delta t^2}{h^2}\left(4\sin^2(k_1h/2) + 4\sin^2(k_2h/2) - 2\right) + 1 = 0$$

where we have rearranged this expression and then solving the quadratic

$$g = \frac{-\left(\frac{\Delta t^2}{h^2}\left(4\sin^2(k_1h/2)+4\sin^2(k_2h/2)-2\right)\right) \pm \sqrt{\left(\frac{\Delta t^2}{h^2}\left(4\sin^2(k_1h/2)+4\sin^2(k_2h/2)-2\right)\right)^2-4}}{2}$$

We can factor out a 2 and get

$$g = 1-\frac{\Delta t^2}{h^2}\left(2\sin^2(k_1h/2)+2\sin^2(k_2h/2)\right) \pm \sqrt{\left(1-\frac{\Delta t^2}{h^2}\left(2\sin^2(k_1h/2)+2\sin^2(k_2h/2)\right)\right)^2-1}$$

Let $\alpha = -2\left(\frac{\Delta t^2}{h^2}\left(2\sin^2(k_1h/2)+2\sin^2(k_2h/2)\right)\right)$ then we have

$$g = 1+\alpha/2 \pm \sqrt{(1+\alpha/2)^2-1}$$

which is the exact same quadratic as before so we see that since we want $|g|<1$ we have $-4<\alpha<0$ so we need

$$-4 < -2\left(\frac{\Delta t^2}{h^2}\left(2\sin^2(k_1h/2)+2\sin^2(k_2h/2)\right)\right) < 0$$

rearranged we then need

$$0 \le 2\frac{\Delta t^2}{h^2}\left(2\sin^2(k_1h/2)+2\sin^2(k_2h/2)\right) \le 4$$

dividing by 4

$$0 \le \frac{\Delta t^2}{h^2}\left(\sin^2(k_1h/2)+\sin^2(k_2h/2)\right) \le 1$$

Since $\sin^2(k_1h/2)+\sin^2(k_2h/2) \le 2$ then

$$\frac{\Delta t^2}{h^2}\cdot \le 1$$

which gives the same CFL condition. ⊠

Problem D

We can do a reduction of order using $v = y'$ then we get the system

$$y' = v \quad v' = y'' = \lambda y$$

Which gives us a first order system which is linear and by Lax-Equivalence for first order ODE, we are done.