Jordan Radke
Pset 2
MAT 714

Link to code on github

Code is available here.

Problem A

a. Since v is in the span of these vectors, we can write it as a linear combination of them, $v = \sum_i c_i w_i$. Then using the orthogonality:

$$\begin{aligned}\langle v, w_j \rangle &= \langle \sum_i c_i w_i, w_j \rangle = \sum c_i \langle w_i, w_j \rangle \\ &= c_j \langle w_j, w_j \rangle.\end{aligned}$$

This implies that $c_j = \frac{\langle v, w_j \rangle}{\|w_j\|^2}$.

b. Here, inner-product bars denote $\langle \cdot, \cdot \rangle_A$ unless stated otherwise. For the base case, we have

$$p_1 = r_1 - \frac{\langle r_1, p_0 \rangle}{\|p_0\|^2} p_0$$

So, taking the inner product of both sides with r_0, and using the fact that $r_0 = p_0$:

$$\begin{aligned}\langle p_1, r_0 \rangle &= \langle r_1 - \frac{\langle r_1, r_0 \rangle}{\|r_0\|^2} r_0, r_0 \rangle \\ &= \langle r_1, r_0 \rangle - \frac{r_1, r_0}{\|r_0\|} \langle r_0, r_0 \rangle = 0.\end{aligned}$$

For the inductive step, assume for all $m < n$ that the assertion is true. Now, let $j < n$:

$$\begin{aligned}\langle p_n, p_j \rangle &= \left\langle r_n - \sum_{l=0}^{n-1} \frac{\langle r_n, p_l \rangle}{\|p_l\|^2} p_l, p_j \right\rangle \\ &= \langle r_n, p_j \rangle - \sum_{l=0}^{n-1} \frac{\langle r_n, p_l \rangle}{\|p_l\|^2} \langle p_l, p_j \rangle\end{aligned}$$

Since our inductive hypothesis the desired orthogonality holds for all $j < n$ we can apply it to every term of this sum. This gives us:

$$\langle r_n, p_j \rangle - \sum_{l=0}^{n-1} \frac{\langle r_n, p_l \rangle}{\|p_l\|^2} \delta_{lj} \langle p_l, p_j \rangle = \langle r_n, p_j \rangle - \frac{\langle r_n, p_j \rangle}{\|p_j\|^2} \langle p_j, p_j \rangle = 0.$$

Above, δ_{lj} is the Kronecker delta.

c. Since the ϕ_n form an orthogonal basis, we can, by part a. above, write:

$$v = \sum_n \langle v, \phi_n \rangle \phi_n, \quad w = \sum_m \langle w, \phi_m \rangle \phi_m.$$

We also used the fact that $\|phi_n\| = \|phi_m\| = 1$ in the above.

To prove property i., we use the bilinearity of the inner product, the fact that ϕ_n are eigenvectors of A, and the fact they are orthonormal.

$$\begin{aligned}
\langle Av, w \rangle = \left\langle A \left(\sum_n \langle v, \phi_n \rangle \phi_n \right), \sum_m \langle w, \phi_m \rangle \phi_m \right\rangle &= \sum_{m,n} \langle v, \phi_n \rangle \langle w, \phi_m \rangle \langle A\phi_n, \phi_m \rangle \\
&= \sum_{m,n} \langle v, \phi_n \rangle \langle w, \phi_m \rangle \lambda_n \delta_{n,m} \\
&= \sum_n \lambda_n \langle v, \phi_n \rangle \langle w, \phi_n \rangle.
\end{aligned}$$

Property ii. follows since A is positive-definite. For any basis vector ϕ_n:

$$0 < \langle A\phi_n, \phi_n \rangle = \lambda_n \langle \phi_n, \phi_n \rangle = \lambda_n$$

For property iii., we use the representation formula from property i. to show:

$$\min_j \lambda_j \sum_n \langle v, \phi_n \rangle^2 \leq \sum_n \lambda_n \langle v, \phi_n \rangle^2 \leq \max_j \lambda_j \sum_n \langle v, \phi_n \rangle^2$$

Since the middle term is exactly $\langle Av, v \rangle$, this is the inequality we're after, since the minimum eigenvalue is λ_1, the maximum is λ_N and

$$\|v\|^2 = \sum_n \langle v, \phi_n \rangle^2$$

because ϕ_n is an orthonormal basis.

For the final property:

$$\begin{aligned}
\langle Av, Av \rangle &= \left\langle A \left(\sum_n \langle v, \phi_n \rangle \phi_n \right), A \left(\sum_m \langle v, \phi_m \rangle \phi_m \right) \right\rangle \\
&= \sum_{m,n} \langle v, \phi_n \rangle \langle v, \phi_m \rangle \langle A\phi_n, A\phi_m \rangle \\
&= \sum_{m,n} \langle v, \phi_n \rangle \langle v, \phi_m \rangle \lambda_n \lambda_m \delta_{mn} \\
&= \sum_n \lambda_n^2 \langle v, \phi_n \rangle^2 \leq \lambda_N^2 \|v\|^2.
\end{aligned}$$

Then, taking square roots gives us the inequality we want.

d. From the update formulas for p_{n+1}, r_n and w_n:

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

Then, using the update for p_n we can rewrite r_n so that this equals

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

e. By Cayley-Hamilton, A is a root of its own characteristic polynomial. The characteristic polynomial of A is a polynomial of degree n:

$$p(\lambda) = \det(A - \lambda I) = \lambda^n c_n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

So we have (since $\det A \neq 0$ ensures that $A^n \neq 0$):

$$p(A) = c_n A^n + c_{n-1}A^{n-1} \cdots + c_1 A + c_0 I = 0$$

Rearranging this equality and dividing by c_n gives us A^n as a linear combination of $A^{n-1}, \ldots A, I$.
f. For part i., using the Richardson iteration formula to rewrite u_{n+1} and the fact that $Au = f$ we get:

$$\begin{aligned} e_{n+1} = u_{n+1} - u &= u_n + \alpha(f - Au_n) - u \\ &= u_n - u + \alpha(Au - Au_n) \\ &= e_n - \alpha A e_n \\ &= (I - \alpha A)e_n. \end{aligned}$$

To show part ii., we can express e_n as a linear combination of the basis ϕ_n as in part c. Then since

$$A(\sum_n c_n\phi_n) = \sum_n c_n\lambda_n\phi_n,$$

we get that:

$$\|e_{n+1}\| = \|(I - \alpha A)e_n\| \leq \max_j \|(1 - \alpha\lambda_j)Ie_n\| = \max_j |1 - \alpha\lambda_j| \|e_n\|.$$

For part iii., we note first that if $\alpha > 0$, the the triangle inequality tells us that for every j

$$|1 - \alpha\lambda_j| \leq 1 + \alpha\lambda_j.$$

Then, taking maximums on both sides:

$$\max_j |1 - \alpha\lambda_j| \leq \max_j 1 + \alpha\lambda_j.$$

But the right-hand side is exactly the maximum ρ associated to $-\alpha$:

$$\max_j |1 - (-\alpha)\lambda_j| = \max 1 + \alpha\lambda_j$$

So, we can conclude that we can restrict our search to $\alpha > 0$, since every $-\alpha < 0$ has a positive α associated to it with equal or lesser ρ.

Now, since λ_1 and λ_N are the extreme values of the eigenvalues $\lambda_1 \leq \cdots \lambda_N$, we only need to consider these, since the maximum will only be achieved for one of these two indices. First, in the case $\lambda_1 = \lambda_N$, then $\alpha = \frac{2}{\lambda_1+\lambda_N}$ is certainly optimal because

$$\left|1 - \frac{2}{\lambda_1 + \lambda_N}\lambda_1\right| = \left|1 - \frac{2\lambda_1}{2\lambda_1}\right| = 0.$$

(And the desired inequality is true.)

So we can assume that $\lambda_1 \neq \lambda_N$. In the first case, say, for arbitrary $\alpha > 0$ the maximum is achieved for $j = 1$, so we have $|1 - \alpha\lambda_1| \geq |1 - \alpha\lambda_N|$. Then, we can decrease $|1 - \alpha\lambda_1|$ by increasing α until $|1 - \alpha\lambda_1| = |1 - \alpha\lambda_N|$, but not any more, since then the maximum would be achieved by λ_N instead. Thus, we can minimize ρ by choosing α such that $|1 - \alpha\lambda_1| = |1 - \alpha\lambda_N|$. Because $\lambda_1 \neq \lambda_N$, this means this occurs when $1 - \alpha\lambda_1 = -(1 - \alpha\lambda_N)$ which means that $\alpha = \frac{2}{\lambda_1+\lambda_N}$. We can make the parallel argument in the case that the maximum is achieved for $j = N$, so that this α will minimize ρ in either case.

To show the inequality, we can again treat it in the two cases that λ_1 or λ_N maximizes ρ. In the first case:

$$\rho = 1 - \frac{2\lambda_1}{\lambda_1 + \lambda_N} = \frac{\lambda_1 + \lambda_N}{\lambda_1 + \lambda_N} - \frac{2\lambda_1}{\lambda_1 + \lambda_N} = \frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1} < 1$$

In the second case,

$$\rho = -\left(1 - \frac{2\lambda_N}{\lambda_N + \lambda_1}\right) = \frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1} < 1.$$

Both hold because $\lambda_1 < \lambda_N$.

For part iv., we can just repeat the inequalities of part iii, since in the case of inequality, we know again that $c \leq \lambda_1 < \lambda_N \leq C$.

g.

For part i., use the fact that $w_0 = Ap_0$ and $p_0 = r_0$. Then the updates for r_n and w_n tell us that $r_1 = r_0 - \alpha_0 w_0 = r_0 - \alpha_0 A r_0$.

For part ii., we use the update for w_n and p_n to see:

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

But from rearranging terms in the update for r_n, $Ap_{n-1} = \frac{r_n - r_{n-1}}{-\alpha_{n-1}}$. So we get:

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

For part iii., we first show the normalized version of part i. Rearranging and dividing out by $\|r_0\|$:

$$\begin{aligned} &\frac{1}{\|r_0\|}\left(Ar_0 = \frac{1}{\alpha_0} r_0 - \frac{1}{\alpha_0} r_1\right) \\ \implies &Aq_0 = \gamma_0 r_0 - \frac{1}{\alpha\|r_0\|}\|r_1\| q_1 \end{aligned}$$

By the definition of β_0, we see that the coefficient on q_1 is:

$$\frac{1}{\alpha_0}\frac{\sqrt{r_1^t r_1}}{\sqrt{r_0^t r_0}} = \frac{\sqrt{\beta_0}}{\alpha_0} \equiv \delta_0.$$

For the normalized version of the recurrence relation in part ii., we rearrange and divide out by $\alpha_n \|r_n\|$ to get:

$$\begin{aligned} \alpha_n A r_n &= r_n - r_{n+1} + \frac{\alpha_n \beta_{n-1}}{\alpha_{n-1}}(r_n - r_{n-1}) \\ \implies A q_n &= \gamma_n q_n - \frac{\|r_{n+1}\|}{\alpha_n \|r_n\|} q_{n+1} - \frac{\beta_{n-1}}{\alpha_{n-1}} \frac{\|r_{n-1}\|}{\|r_n\|} q_{n-1} \end{aligned}$$

From the updates for β_n and β_{n-1}:

$$\frac{\|r_{n+1}\|}{\|r_n\|} = \sqrt{\beta_n}$$

and

$$\frac{\|r_n\|^2}{\|r_{n-1}^2\|}\frac{1}{\alpha_{n-1}}\frac{\|r_{n-1}\|}{\|r_n\|} q_{n-1} = \frac{\sqrt{\beta_{n-1}}}{\alpha_{n-1}} q_{n-1}$$

So, using the definition of δ_n and δ_{n-1} this gives us the recurrence relation we want.

Part iv. is just the matrix version of the n equations above: each row of the matrix equality $AQ_n = Q_n T_n - \delta_{n-1} q_n e_n^t$ is given one of the equations above for $0 \leq j \leq n-1$. Note that the last row requires the additional term $\delta_{n-1} q_n e_n^t$ since q_n is not in the span of $\{q_j\}_{j=0}^{n-1}$.

For part v., we left-multiply both sides of the matrix equality above by Q_n^t:

$$Q_n^t A Q_n = Q_n^t Q T_n - Q_n^t \delta_{n-1} q_n e_n^t.$$

But since Q_n's columns form an orthonormal basis, $Q_n^t Q_n = I$ and because $q_n \perp \text{span}\{q_0, \ldots, q_{n-1}\}$, this gives us:

$$Q_n^t A Q_n = T_n$$

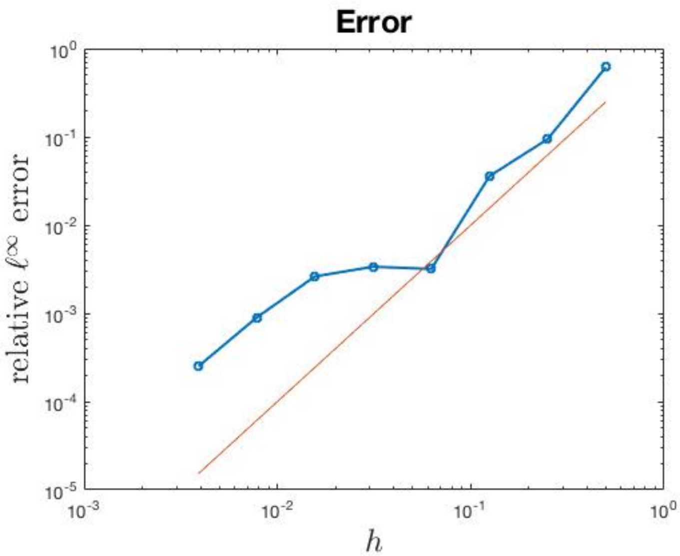
Problem B

This code used to find the minimum number of grid points necessary for an interpolation of sufficient resolution is available on github under 'gaussian_interpolation.m'. I found the minimum number of gridpoints to be $N = 100$, in order to get the interpolation error under 10^{-2}.

Problem C

a. This implementation is available on github as well, under the files 'wave_equation.m', 'wave_fourier.m' and 'error_benchmarking.m'. The numerical solution was tested against the analytic solution found via Fourier series:

$$u(x,y,t) = \sum_{m,n}^{\infty} b_{m,n} \sin(\pi\sqrt{m^2+n^2}t)\sin(m\pi x)\sin(n\pi y)$$



where the Fourier sine coefficients are

$$b_{m,n} = \frac{4}{\pi\sqrt{m^2+n^2}} \int_0^1 \int_0^1 g(x,y) \sin m\pi x \sin n\pi y \, dx \, dy$$

I kept the first 100 terms in the Fourier series (cutting off at $m = n = 10$). A plot of the log − log error and the function h^2 is above.

b. We start with the three-point discretization of the ODE $u''(t) = \lambda u(t)$:

$$u^{n-1} - 2u^n + u^{n+1} = (dt)^2 \lambda u^n.$$

Setting $z = (dt)^2\lambda$ this means the characteristic polynomial is

$$\pi(\zeta; z) = 1 - 2\zeta + \zeta^2 - z\zeta = 1 - (2+z)\zeta + \zeta^2.$$

Both of the roots of this binomial have modulus less than or equal to 1 when $-4 \le z \le 0$ (at least in the real case, which is all we need here). We can use this in conjunction with the past problem set—where we showed that the eigenvalues of the $2d$ grid Laplacian matrix are $\lambda_n + \lambda_m$, where λ_n are the eigenvalues of the $1d$ grid Laplacian matrix. The eigenvalue of the $1d$ Laplacian with largest magnitude will be approximately $-\frac{4}{(dx)^2}$ so to ensure all eigenvalues are in the domain of stability, we need:

$$-4 \le \left(-\frac{4}{(dx)^2} - \frac{4}{(dx)^2}\right)(dt)^2 \le 0 \implies \frac{(dt)^2}{(dx)^2} \le \frac{1}{2}.$$

c. For the Von Neumann stability analysis, we consider the action of the discretization on the (grid) wave function $W_{j,k} = e^{ij(dx)\xi}e^{ik(dx)\xi}$. So after some algebra:

$$\begin{aligned} W_{jk}^n + 1 &= 2W_{jk}^n - W_{jk}^{n-1} + \frac{(dt)^2}{(dx)^2}\left(W_{j,k-1}^n + W_{j-1,k}^n + W_{j+1,k}^n + W_{j,k+1}^n\right) \\ &= \left(1 + 2\frac{(dt)^2}{(dx)^2}\cos((dx)\xi)\right) W_{jk}. \end{aligned}$$

Since $\cos(y) \leq 1$ for any y, this means the amplification factor of any given mode is less than

$$1 + 2\frac{(dt)^2}{(dx)^2}$$

Letting $\alpha = \frac{2}{(dx)^2}$, we see that this satisfies the condition for stability since $(dt)^2 \leq dt$ as soon as $dt \leq 1$.

Sources

-"Finite difference methods for ordinary and partial differential equations", by Leveque
-"Introduction to partial differential equations," by Olver
-MATLAB documentation and Wikipedia
-code posted on Canvas and MAT 714 github repository