Solution to problem 1

a) Let $v_1 = 1$, $v_2 = x$, and $v_3 = x^2$. Use the Gram-Schmidt orthogonalization method to determine an ON-basis.

$$
\begin{aligned}
e_1 &= \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{1^2 + 1^2 + 1^2 + 1^2}} = \frac{1}{2}, \\
w_2 &= v_2 - \langle v_2, e_1 \rangle e_1 = x - \frac{1}{2}(-1 - 1/3 + 1/3 + 1)e_1 = x, \\
e_2 &= \frac{w_2}{\|w_2\|} = \frac{x}{\sqrt{(-1)^2 + (-1/3)^2 + (1/3)^2 + 1^2}} = \frac{3x}{\sqrt{20}}, \\
w_3 &= v_3 - \langle v_3, e_1 \rangle e_1 - \langle v_3, e_2 \rangle e_2 = x^2 - \langle x^2, 1/2 \rangle 1/2 - \langle x^2, 3x/\sqrt{20} \rangle \frac{3x}{\sqrt{20}} \\
&= x^2 - \frac{5}{9}, \\
e_3 &= \frac{w_3}{\|w_3\|} = \frac{x^2 - 5/9}{\sqrt{(4/9)^2 + (4/9)^2 + (4/9)^2 + (4/9)^2}} = \frac{x^2 - 5/9}{8/9} = \frac{9}{8}x^2 - \frac{5}{8}.
\end{aligned}
$$

b) The least squares projection of the data unto M is given by $g^* = \langle y, e_1 \rangle e_1 + \langle y, e_2 \rangle e_2 + \langle y, e_3 \rangle e_3$.

$$
\begin{aligned}
\langle y, e_1 \rangle e_1 &= \frac{1}{2}\frac{1}{2}\langle y, 1 \rangle 1 = \frac{7}{4}, \\
\langle y, e_2 \rangle e_2 &= \frac{3}{\sqrt{20}}\frac{3}{\sqrt{20}}\langle y, x \rangle x = \frac{33}{20}x, \\
\langle y, e_3 \rangle e_3 &= \left\langle y, \frac{9}{8}x^2 - \frac{5}{8} \right\rangle \left(\frac{9}{8}x^2 - \frac{5}{8} \right) = \left[\frac{9}{8}\langle y, x^2 \rangle - \frac{5}{8}\langle y, 1 \rangle \right] \left(\frac{9}{8}x^2 - \frac{5}{8} \right) \\
&= -\frac{1}{2} \left(\frac{9}{8}x^2 - \frac{5}{8} \right) = -\frac{9}{16}x^2 + \frac{5}{16}.
\end{aligned}
$$

The least squares approximation is $g^*(x) = -\frac{9}{16}x^2 + \frac{33}{20}x + \frac{33}{16}$.

Solution to problem 2

a) The matrix A is symmetric. Therefore the eigenvalues are real (not necessary in order to solve the problem, but useful to know). The Gersgorin circles are the same for rows and columns. For the original matrix, we have

$$
\begin{aligned}
|\lambda_1 - 0.5| &\leq 0.2 + 0.3 = 0.5 &\Rightarrow& \quad 0 \leq \lambda_1 \leq 1 \\
|\lambda_2 + 2.0| &\leq 0.2 &\Rightarrow& \quad -2.2 \leq \lambda_2 \leq -1.8 \\
|\lambda_3 + 0.5| &\leq 0.3 &\Rightarrow& \quad -0.8 \leq \lambda_3 \leq -0.2.
\end{aligned}
$$

The largest eigenvalue (the one to compute) is λ_1.

For the first shift $A + 1.25I$, we get, for $\mu_j = \lambda_j + 1.25$,

$$1.25 \leq \mu_1 \leq 2.25$$
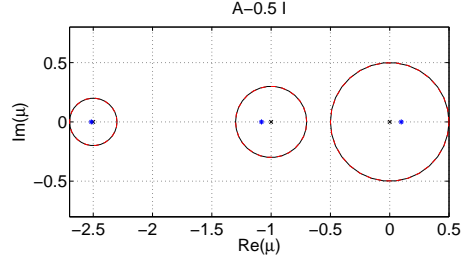$$-0.95 \leq \mu_2 \leq 0.55$$
$$0.45 \leq \mu_3 \leq 1.05$$
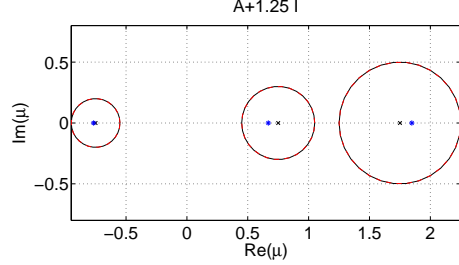
The power method will converge to μ_1 as desired. The error is given by

$$\left(\frac{\max(|\mu_2|, |\mu_3|)}{|\mu_1|}\right)^k \leq \left(\frac{1.05}{1.25}\right)^k = 0.84^k,$$

where the worst case scenario was chosen as estimate. (The best case is $(0.55/2.25)^k \approx 0.24^k$.)

For the second shift $A - 0.5I$, we get, for $\mu_j = \lambda_j - 0.5$,

$$-0.5 \leq \mu_1 \leq 0.5$$
$$-2.7 \leq \mu_2 \leq -2.3$$
$$-1.3 \leq \mu_3 \leq -0.7$$

Inverse iteration converges to the smallest eigenvalue μ_1 as desired. However, now the convergence rate is determined by the inverse of the eigenvalues, since we are actually using the power method on A^{-1}. The error is given by

$$\left(\frac{\max(|\mu_2|^{-1}, |\mu_3|^{-1})}{|\mu_1|^{-1}}\right)^k = \left(\frac{|\mu_1|}{\min(|\mu_2|, |\mu_3|)}\right)^k \leq \left(\frac{0.5}{0.7}\right)^k \approx 0.71^k,$$

where the worst case scenario has again been chosen as estimate. (The best case is convergence in zero iterations if the shift is exactly the eigenvalue.)

Answer: Inverse iteration is likely to converge in fewer iterations than the power method, since the worst case convergence rate is faster than for the power method. (Furthermore, the best case convergence is also better than for the power method.)

b) The power method does one normalization, one matrix-vector multiplication and one eigenvalue estimate in each iteration, leading to

$$C_p = k_p(3N + 2N^2 + 2N) = k_p N(2N + 5).$$

With inverse iteration, the matrix is first factorized and then, for each iteration, one normalization, one solve, and one eigenvalue estimate is

2

performed, leading to

$$C_i = 2N^3/3 + k_i(3N + 2N^2 + 2N) = 2N^3/3 + k_i N(2N + 5).$$

The two methods use the same number of operations when

$$
\begin{aligned}
k_p N(2N+5) &= 2N^3/3 + k_i N(2N+5) \\
&\Leftrightarrow \\
(k_p - k_i)N(2N+5) &= 2N^3/3 \\
&\Leftrightarrow \\
(k_p - k_i) &= \frac{2N^3/3}{N(2N+5)} \approx \frac{N}{3}.
\end{aligned}
$$

Conclusion: We can do a lot of extra iterations with the power method for a large full matrix before it pays off to change to inverse iteration. However, for a sparse matrix, the extra cost for the factorization is much less notable.

Solution to problem 3

a)

$$
\begin{aligned}
\|B\|_1 &= \max(0.8, 0.85, 0.6, 0.95) = 0.95 \\
\|B\|_\infty &= \max(1.1, 0.95, 0.55, 0.6) = 1.1
\end{aligned}
$$

b)

$$
x^1 = \begin{bmatrix} -0.1 & 0.2 & -0.15 & 0.65 \\ 0.25 & 0.5 & 0.1 & -0.1 \\ -0.15 & -0.1 & 0.2 & -0.1 \\ 0.3 & 0.05 & 0.15 & 0.1 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \\ -0.1 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \\ 0.3 \\ -0.2 \end{bmatrix} = \begin{bmatrix} 0.325 \\ 0.175 \\ 0.205 \\ -0.045 \end{bmatrix}.
$$

c) The iterations converge if $\rho(B) < 1$. We have $\rho(B) \le \|B\|$ all norms. Since $\|B\|_1 = 0.95$, we have $\rho(B) \le 0.95 < 1$. Hence, the iterative method converges.

d)

$$Bu = (D + L + U)u = b$$

Jacobi's method:

$$x^{k+1} = -D^{-1}(L + U)x^k + D^{-1}b.$$

In this case:

$$
x^{k+1} = \begin{bmatrix} 0 & 2 & -1.5 & 6.5 \\ -0.5 & 0 & -0.2 & 0.2 \\ 0.75 & 0.5 & 0 & 0.5 \\ -3 & -0.5 & -1.5 & 0 \end{bmatrix} x^k + \begin{bmatrix} -2 \\ 2 \\ -6 \\ -10 \end{bmatrix}.
$$

Remark: Both x and u were used in the question by accident and either one can be used in the answer.

Solution to problem 4

a) $0 - 0 \cdot \lambda = 0 \Rightarrow$ the PDE is parabolic.

b)
$$u_j^{n+1} = u_j^n + \frac{\lambda \Delta t}{h^2} \left(u_{j+1}^n - 2u_j^n + u_{j-1}^n \right), \quad j = 0, \ldots, M-1,$$

$$\Rightarrow$$

$$u^{n+1} = \begin{pmatrix} a & b & & & & c \\ c & a & b & & & \\ & \ddots & \ddots & \ddots & & \\ & & c & a & b \\ b & & & c & a \end{pmatrix} u^n,$$

where
$$a = 1 - 2\frac{\lambda \Delta t}{h^2},$$
$$b = c = \frac{\lambda \Delta t}{h^2}.$$

c) Taylor-expansion around (x_j, t_n), denote $u(x_j, t_n)$ by u:

$$u_j^{n+1} = u + \Delta t u_t + \frac{\Delta t^2}{2} u_{tt} + \frac{\Delta t^3}{6} u_{ttt} + \mathcal{O}(\Delta t^4),$$
$$u_{j+1}^n = u + h u_x + \frac{h^2}{2} u_{xx} + \frac{h^3}{6} u_{xxx} + \frac{h^4}{24} u_{xxxx} + \mathcal{O}(h^5),$$
$$u_{j-1}^n = u - h u_x + \frac{h^2}{2} u_{xx} - \frac{h^3}{6} u_{xxx} + \frac{h^4}{24} u_{xxxx} + \mathcal{O}(h^5).$$

Move all terms in the approximation of the PDE to the left-hand side:

$$\frac{u + \Delta t u_t + \frac{\Delta t^2}{2} u_{tt} + \frac{\Delta t^3}{6} u_{ttt} + \mathcal{O}(\Delta t^4) - u}{\Delta t} -$$
$$\lambda \frac{u + h u_x + \frac{h^2}{2} u_{xx} + \frac{h^3}{6} u_{xxx} + \frac{h^4}{24} u_{xxxx} + \mathcal{O}(h^5) - 2u + u - h u_x + \frac{h^2}{2} u_{xx} - \frac{h^3}{6} u_{xxx} + \frac{h^4}{24} u_{xxxx} + \mathcal{O}(h^5)}{h^2} =$$
$$u_t + \frac{\Delta t}{2} u_{tt} + \mathcal{O}(\Delta t^2) - \lambda u_{xx} + \mathcal{O}(h^2) = [u_t = \lambda u_{xx}] =$$
$$\mathcal{O}(\Delta t) + \mathcal{O}(h^2).$$

Hence the local truncation error is $\mathcal{O}(\Delta t) + \mathcal{O}(h^2)$.

d) A Fourier expansion and orthogonality of $e^{i \omega x_j}$ yields

$$\begin{aligned} \hat{u}_\omega^{n+1} &= \hat{u}_\omega^n \left(1 + \lambda \frac{\Delta t}{h^2} \left(e^{i\omega h} - 2 + e^{-i\omega h} \right) \right) = \\ &= \hat{u}_\omega^n \left(1 + 2\lambda \frac{\Delta t}{h^2} \left(\cos(\omega h) - 1 \right) \right) = \hat{Q}_\omega \hat{u}_\omega^n. \end{aligned}$$

Since $\cos(\omega h) - 1 \in [-2, 0]$ $\hat{Q}_\omega \leq 1$. Remains to derive a condition such that $\hat{Q}_\omega \geq -1$. A condition for this is that $2\lambda \frac{\Delta t}{h^2} \left(\cos(\omega h) - 1 \right) \geq -2$. Since $\cos(\omega h) - 1 \geq -2$ this means that $\frac{\lambda \Delta t}{h^2} \leq \frac{1}{2}$ for stability, i.e. $\Delta t \leq \frac{h^2}{2\lambda}$.

Solution to problem 5

a) Let v be an arbitrary element in $\mathcal{V}$ multiplying the equation with v and integrating over $[0, 1]$ results in

$$\int_0^1 -u''v \, dx = \int_0^1 fv \, dx$$

integrating the left hand side by parts results in

$$\int_0^1 u'v' \, dx - [u'v]_0^1 = \int_0^1 fv \, dx. \tag{1}$$

The boundary conditions can be written as

$$u'(0) = 1 - au(0), \quad u'(1) = 1 - au(1),$$

inserting this into equation (1) yields

$$\int_0^1 u'v' \, dx - v(1)(1 - au(1) + v(0)(1 - a(0)) = \int_0^1 fv \, dx$$

Since v is arbitrary this hold for all $v \in \mathcal{V}$. We can now state our variation problem as

Find $u \in \mathcal{V}$ such that

$$\int_0^1 u'v' \, dx - v(1)(1 - au(1) + v(0)(1 - a(0)) = \int_0^1 fv \, dx \quad \forall v \in \mathcal{V}.$$

b) Let $\mathcal{V}_0 = \{v \in \mathcal{V} \mid v(0) = v(1) = 0\}$, the variational formulation reads

Find $u \in \mathcal{V}_0$ such that

$$\int_0^1 u'v' \, dx = \int_0^1 fv \, dx \quad \forall v \in \mathcal{V}_0.$$

Let $N > 0$ be an integer, and define $h = 1/(N+1)$ and let $x_i = ih$ for $i = 0, 1, 2, \ldots, N, N+1$. Now define the linear hat functions ϕ_i as

$$\phi_i = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{for } x \in (x_{i-1}, x_i], \\ 1 - \frac{x - x_i}{x_{i+1} - x_i} & \text{for } x \in (x_i, x_{i+1}], \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, N$. Define $\mathcal{V}_h = \text{span}\{\phi_1, \phi_2, \ldots, \phi_N\}$. The finite element method for equation is given by

Find $u_h \in \mathcal{V}_h$ such that

$$\int_0^1 u_h'v_h' \, dx = \int_0^1 fv_h \, dx \quad \forall v_h \in \mathcal{V}_h.$$

5

Let $v = (v_1, v_2, \ldots, v_N)^T$ and define $v_h(x) = \sum_{i=1}^N v_i \phi_i(x)$, then

$$v^T A v = \sum_{i=1}^N \sum_{j=1}^N v_i \int_0^1 \phi_i' \phi_j' \, dx \, v_j$$

$$= \int_0^1 \sum_{i=1}^N (v_i \phi_i') \sum_{j=1}^N (v_j \phi_j') \, dx = \int_0^1 |v_h'|^2 \, dx \geq 0,$$

with equality iff $v_h' \equiv 0$ or equivalently $v_h = Const$. But since $v_h(0) = v_h(1) = 0$ we have equality iff $v_h \equiv 0$. Thus $v^T A v > 0$ for each $v \neq 0$, that is, A is positive definite.

c) For $x \in [0, 1]$ we have that

$$|u(x)| = \left| \int_0^x u'(y) \, dy \right| = \left| \int_0^x 1 u'(y) \, dy \right| \leq \left| \int_0^x 1^2 \, dy \right|^{1/2} \left| \int_0^x u'(y)^2 \, dy \right|^{1/2}$$

$$\leq 1 \left| \int_0^1 u'(x)^2 \, dy \right|^{1/2} = \|u'\|_{L^2(0,1)},$$

and thus

$$\|u\|_{L^2(0,1)} = \left| \int_0^1 (u(x))^2 \, dx \right|^{1/2} \leq \left| \int_0^1 \|u'\|_{L^2(0,1)}^2 \, dx \right|^{1/2} = \|u'\|_{L^2(0,1)}.$$

Moreover, we have that

$$\|u'\|_{L^2(0,1)}^2 = \left| \int_0^1 u'u' \, dx \right| = \left| \int_0^1 fu \, dx \right| \leq \left| \int_0^1 f^2 \, dx \right|^{1/2} \left| \int_0^1 u(x)^2 \, dx \right|^{1/2}$$

$$= \|f\|_{L^2(0,1)} \|u\|_{L^2(0,1)} \leq \|f\|_{L^2(0,1)} \|u'\|_{L^2(0,1)}$$

that is, $\|u'\|_{L^2(0,1)} \leq |f\|_{L^2(0,1)}$. Putting it all together yields

$$\|u\|_{L^2(0,1)} + \|u'\|_{L^2(0,1)} \leq 2\|u'\|_{L^2(0,1)} \leq 2|f\|_{L^2(0,1)}. \qquad \square$$