Norwegian University of Science and Technology
Department of Mathematical Sciences

Contact during the exam:
Elena Celledoni, tlf. 735 93541

Solution of the exam in TMA4212

Monday 23rd May 2013
Time: 9:00–13:00

Allowed aids: Approved simple pocket calculator. All written and handwritten material from the course. Learning outcome **L6** and **L3** have been tested through the project work[1].

Problem 1 *Learning outcome* **L2, L7**
Consider the differential equation

$$-(p(x)u')' + r(x)u = f(x), \quad \text{for } x \in (a,b),$$

with $p \in C^1[a,b]$, $p(x) \geq c_0 > 0$ for all $x \in [a,b]$, $r \in C^0[a,b]$, $r(x) \geq 0$ for all $x \in [a,b]$ and $f \in L_2[a,b]$, subject to the boundary conditions

$$-p(a)u'(a) + \alpha\, u(a) = A, \qquad p(b)u'(b) + \beta\, u(b) = B,$$

where α and β are positive real numbers and A and B are real numbers.

a) Show that the weak formulation of the boundary value problem is

$$\text{find} \quad u \in H^1(a,b), \text{ such that } \mathcal{A}(u,v) = \ell(v) \text{ for all } v \in H^1(a,b),$$

where

$$\mathcal{A}(u,v) = \int_a^b [p(x)u'(x)v'(x) + r(x)u(x)v(x)]\,\mathrm{d}x + \alpha u(a)v(a) + \beta u(b)v(b),$$

and

$$\ell(v) = \langle f, v\rangle + Av(a) + Bv(b).$$

Solution

We multiply the boundary value problem on both sides by a test function $v \in H^1(a,b)$ and integrate between a and b to obtain

$$\int_a^b [-(p(x)u')'v + r(x)uv]\,dx = \int_a^b fv\,dx,$$

[1]The learning outcome is published on the course webpage and on the official description of the course.

and integrating by parts we get

$$(-pu')v|_a^b - \int_a^b [-(p(x)u')v' - r(x)uv]\,dx = \int_a^b fv\,dx,$$

$$\int_a^b [(p(x)u')v' + r(x)uv]\,dx + [p(a)u'(a)v(a) - p(b)u'(b)v(b)] = \int_a^b fv\,dx,$$

using the boundary conditions and rearranging the terms we get

$$\int_a^b [(p(x)u')v' + r(x)uv]\,dx + \alpha u(a)v(a) + \beta u(b)v(b) = \int_a^b fv\,dx + Av(a) + Bv(b),$$

leading to the proposed definition of $\mathcal{A}(u,v)$ (left hand side) and $\ell(v)$ (right hand side).

b) Construct a finite element approximation of the boundary value problem based on this weak formulation using piecewise linear finite element basis functions on the subdivision

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

on the interval $[a,b]$. Show that the finite element method gives rise to a set of $n+1$ simultaneous linear equations with $n+1$ unknowns $u_i = u^h(x_i)$, $i = 0,1,\ldots,n$. Show that this linear system has a unique solution. Comment on the structure of the matrix $M \in \mathbf{R}^{(n+1)\times(n+1)}$of the linear system: Is M symmetric? Is M positive deinite? Is M tridiagonal? Explain your answers.

Solution

We consider $n+1$ piecewise linear basis functions $\phi_0,\ldots,\phi_n$ such that $\phi_j(x_j) = 1$, $j = 0,\ldots,n$:

$$\phi_j(x) = \begin{cases} \frac{x-x_{j-1}}{h}, & x \in [x_{j-1}, x_j), \\ \frac{x_{j+1}-x}{h}, & x \in [x_j, x_{j+1}], \\ 0 & \text{otherwise}, \end{cases} \qquad j = 1,\ldots,n-1$$

and

$$\phi_0 = \begin{cases} \frac{x_1-x}{h}, & x \in [x_0, x_1), \\ 0 & \text{otherwise}, \end{cases} \qquad \phi_n = \begin{cases} \frac{x-x_{n-1}}{h}, & x \in [x_{n-1}, x_n], \\ 0 & \text{otherwise}. \end{cases}$$

The finite element space is defined by

$$S_h := \{w_h \in H^1 \,|\, w_h = \sum_{j=0}^{n} w_j\phi_j\},$$

and the Galerkin method is

$$\text{Find}\, u_h \in S_h\, \text{s.t.}\, \mathcal{A}(u_h, v_h) = \ell(v_h), \quad \forall v_h \in S_h.$$

Therefore u_h can be written as $u_h = \sum_{j=0}^n u_j\phi_j$, and noticing that it is necessary and sufficient to impose

$$\mathcal{A}(u_h, \phi_i) = \ell(\phi_i), \quad i = 0, \ldots, n$$

to determine u_h, using the bilinearity of $\mathcal{A}$ and the linearity of ℓ we get

$$\sum_{j=0}^{n} \mathcal{A}(\phi_j, \phi_i)u_j = \ell(\phi_i), \quad i = 0, \ldots, n$$

and letting

$$M_{i,j} := \mathcal{A}(\phi_j, \phi_i), \quad b_i := \ell(\phi_i),$$

we get the linear system

$$M\mathbf{u} = \mathbf{b},$$

with M the finite element matrix $(n+1) \times (n+1)$.

M is symmetric because

$$M_{i,j} = \mathcal{A}(\phi_j, \phi_i) = \mathcal{A}(\phi_i, \phi_j) = M_{j,i}.$$

M is positive definite, in fact consider the vector $\mathbf{v} \in \mathbf{R}^{n+1}$ and $\mathbf{v} \neq 0$ then

$$\mathbf{v}^T M\mathbf{v} = \sum_i \sum_j v_i M_{i,j} v_i = \sum_i \sum_j v_i \mathcal{A}(\phi_j, \phi_i) v_i$$

and using the bilinearity of the map we get

$$\mathbf{v}^T M\mathbf{v} = \mathcal{A}(v, v),$$

where $v = \sum_{j=0}^n v_j\phi_j \in H^1$. Further we have

$$\mathbf{v}^T M\mathbf{v} = \mathcal{A}(v, v) \geq \int_a^b pv'v'\,dx > c_0 \int_a^b (v')^2\,dx + \alpha v(a)^2 + \beta v(b)^2,$$

and since $v \neq 0$ (almost everywhere) and $v \in H^1$, and $\int_a^x v'(x)\,dx = v(x) - v(a)$ then v' is not zero (almost everywhere) unless $v(x) = v(a) \neq 0$ for almost every x so either

$$\int_a^b (v')^2\,dx > 0,$$

or $v(x) = v(a)$ almost everywhere in $[a, b]$ and

$$\alpha\, v(a)^2 > 0,$$

we can conclude that

$$\mathbf{v}^T M\mathbf{v} > 0$$

and M is positive definite.

To understand if M is tridiagonal we must consider the entries

$$M_{i,j} = \mathcal{A}(\phi_j, \phi_i), \quad |i-j| > 1,$$

the result follows after showing that ϕ_i or ϕ_j are not simultaneously different from zero in any of the subintervals $[x_l, x_{l+1}]$ for all l when $|i-j| > 1$, and the same is true for their derivatives.

Problem 2 *Learning outcome* **L5, L7**

Consider Laplace equation in 2D on a square domain $\Omega = [0,1] \times [0,1]$,

$$u_{xx} + u_{yy} = 0,$$

with Dirichlet boundary conditions,

$$u = g, \quad \text{on}\, \partial\Omega.$$

Use the 5-points formula to discretize the equation on the grid $h = \frac{1}{M+1}$, $x_i = i\,h$, $i = 1, \dots, M$, and $y_j = j\,h$, $j = 1, \dots, M$ and obtain a linear system of algebraic equations

$$A\,\mathbf{u} = \mathbf{b},$$

where $\mathbf{u}$ is a vector whose components are the numerical approximations of u on the grid of the discretization $u_{i,j} \approx u(x_i, y_j)$, ordered as follows:

$$\mathbf{u} := [u_{1,1}, u_{2,1}, \dots, u_{M,1}, u_{1,2}, \dots, u_{M,2}, \dots, u_{1,M}, \dots, u_{M,M}]^T.$$

Consider the use of the conjugate gradient method to solve this linear system.

Recall the following formula for the eigenvalues of the discrete Laplacian:

$$\mu_{j,l} = \frac{2}{h^2}\left(\cos\left(\pi j h\right) - 1\right) + \frac{2}{h^2}\left(\cos\left(\pi l h\right) - 1\right), \quad j = 1, \dots, M,\, l = 1, \dots, M.$$

a) Write out A explicitly. Show that $-A$ is a positive definite matrix whose eigenvalues lie in an interval $[a, b]$ with $\sqrt{\frac{a}{b}}$ approximately $\frac{\pi h}{2}$.

Solution

A is a $M^2 \times M^2$ with M blocks B $M \times M$ along the diagonal each of which is of the form

$$B = \frac{1}{h^2}\begin{bmatrix} -4 & 1 & & & \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & -4 \end{bmatrix}$$

along the sub and super diagonal there are $M-1$, $M \times M$ identity blocks scaled with $\frac{1}{h^2}$.

The eigenvalues of $-A$ are $-\mu_{j,l}$ one finds

$$\min_{j,l}(-\mu_{j,l}) = \frac{4}{h^2}(1-\cos(\pi h)) = a \approx \frac{4}{h^2}\frac{\pi^2 h^2}{2} = 2\pi^2,$$

$$\min_{j,l}(-\mu_{j,l}) = \max_j \frac{4}{h^2}(1-\cos(\pi j h)) = \max_j \frac{8}{h^2}\sin(\pi/2jh)^2 = \frac{8}{h^2}\sin(\pi/2Mh)^2 = b \approx \frac{8}{h^2},$$

and

$$\sqrt{\frac{a}{b}} \approx \frac{\pi h}{2}.$$

b) Assume you are given an arbitrary initial guess x_0. Use the Theorem for the convergence of the conjugate gradient method to estimate the number of iterations K necessary to obtain a relative error below the tolerance ε, that is

$$\frac{\|e_K\|_A}{\|e_0\|_A} \leq \varepsilon.$$

Recall that for a vector $\mathbf{v}$,

$$\|\mathbf{v}\|_A := \sqrt{\mathbf{v}^T A \mathbf{v}}.$$

Find an estimate for K assuming $M+1 = 100$ and $\varepsilon = 10^{-7}$.

Solution

We use the theorem about the convergence of the conjugate gradient method in chapter 14.5 in the book by Strikwerda and we have

$$\frac{\|e_K\|_A}{\|e_0\|_A} \leq 2\frac{(\sqrt{b}-\sqrt{a})^K}{(\sqrt{b}+\sqrt{a})^K} = 2\frac{(1-\sqrt{\frac{a}{b}})^K}{(1+\sqrt{\frac{a}{b}})^K} = 2\frac{(1-\frac{\pi h}{2})^K}{(1+\frac{\pi h}{2})^K}$$

and to find K we impose the condition

$$2\frac{(1-\frac{\pi h}{2})^K}{(1+\frac{\pi h}{2})^K} = \varepsilon.$$

Taking logarithms we get

$$K = \frac{\log\frac{\varepsilon}{2}}{(\log(1-\frac{\pi h}{2}) - \log(1+\frac{\pi h}{2}))}.$$

Inserting the prescribed values for ε and $M+1$ we obtain the estimate

$$K = 536.$$

Problem 3 *Learning outcome* **L1, L4, L7**
Consider the linear PDE

$$u_t + u_{xxx} = 0, \quad u(-\infty, t) = u(\infty, t) = 0, \quad x \in \mathbf{R}, \quad t \geq 0.$$

Consider the interval $[-L, L]$ with $L > 0$ sufficiently large, consider the grid $x_m = -L+hm$, $h = 2L/M$, $m = 0, \ldots, M$. Discretise with central finite differences in space and the Backward Euler method in time (backward differences in time), let $u(x_0, t) = u(x_M, t) = 0$.

Use the following central differences approximation of the third derivative

$$u_{xxx}|_{(x_m,t)} = \frac{u(x_{m+3}, t) - 3u(x_{m+1}, t) + 3u(x_{m-1}, t) - u(x_{m-3}, t)}{8h^3} + \mathcal{O}(h^2).$$

Show that the obtained method is Von Neumann stable.

Solution

Using central differences in space and backward Euler in time we obtain

$$U_m^{n+1} = U_m^n + \frac{k}{8h^3}\left(U_{m-3}^{n+1} - 3U_{m-1}^{n+1} + 3U_{m+1}^{n+1} - U_{m+3}^{n+1}\right),$$

leading to

$$U_m^{n+1} + \frac{k}{8h^3}\left(-U_{m-3}^{n+1} + 3U_{m-1}^{n+1} - 3U_{m+1}^{n+1} + U_{m+3}^{n+1}\right) = U_m^n,$$

by assuming $U_m^n = \xi^n e^{i\beta x_m}$, and $i = \sqrt{-1}$, and substituting in the method we obtain

$$\xi = \frac{1}{1+i\alpha},$$

with

$$\alpha = \frac{k}{8h^3}(\sin(3\beta h) - 3\sin(\beta h)).$$

Finally we obtain

$$\xi^*\xi = |\xi|^2 = \frac{1}{1+i\alpha}\frac{1}{1-i\alpha} = \frac{1}{1+\alpha^2} \leq 1,$$

implying Von Neumann stability.

Problem 4 *Learning outcome* **L1, L4, L7**
Let r and α be positive real numbers. Show that if $[w_1, \ldots, w_M]^T$ satisfies

$$-rw_{m-1} + (1 + 2r + \alpha)w_m - rw_{m+1} = v_m, \quad 1 \leq m \leq M-1, \quad w_0 = w_M = 0,$$

then

$$\max_{0\leq m\leq M} |w_m| \leq \max_{1\leq m\leq M-1} |v_m|.$$

Use this to show that Backward Euler converges for arbitrary $r = k/h^2$ on the problem

$$u_t = u_{xx} - u, \quad u(x, 0) = f(x), \quad u(0, t) = g_0(t), \quad u(1, t) = g_1(t).$$

Hint: You will have to take $w_m = e_m^{n+1}$ where $e_m^n := U_m^n - u(x_m, t_m)$ is the error, and U_m^n is the numerical approximation given by the Backward Euler method.

Solution

We have

$$(1 + 2r + \alpha)w_m = rw_{m-1} + rw_{m+1} + v_m$$

and then

$$(1 + 2r + \alpha) \max_{1\leq m\leq M} |w_m| \leq 2r \max_{1\leq m\leq M} |w_m| + \max_{1\leq m\leq M-1} |v_m|,$$

and then

$$\max_{1\leq m\leq M} |w_m| \leq \max_{1\leq m\leq M-1} |v_m|.$$

Considering Backward Euler we have

$$-rU_{m-1}^{n+1} + (1 + 2r + k)U_m^{n+1} - rU_{m+1}^{n+1} = U_m^n$$

and the error $e_m^n := U_m^n - u(x_m, t_n)$ satisfies the equation

$$-re_{m-1}^{n+1} + (1 + 2r + k)e_m^{n+1} - re_{m+1}^{n+1} = e_m^n - k\tau_m^n$$

where $\tau_m^n = (\mathcal{O}(k) + \mathcal{O}(h^2))$. Using the property proven at the beginning of the exercise, with $w_m = e_m^{n+1}$ and $v_m = e_m^n - k\tau_m^n$

$$\max_{1\leq m\leq M} |e_m^{n+1}| \leq \max_{1\leq m\leq M} |e_m^n| + k \max_m |\tau_m^n| \leq k \sum_{j=1}^{n} \max_m |\tau_m^j| \leq k\, n \max_{m,n} |\tau_m^n| \leq T \max_{m,n} |\tau_m^n|$$

and since the method is consistent (in fact for Backward Euler $\max_{m,n} |\tau_m^n| = \mathcal{O}(k) + \mathcal{O}(h^2)$) so we get

$$\max_m |e_m^n| \leq T(\mathcal{O}(k) + \mathcal{O}(h^2)), \quad \forall m \,\text{and}\, n$$

the last inequality implies convergence.