

Homework 2

October 15, 2020

Deadline: October 26th 2020, the answer to the questions will be submitted via Canvas, and the code will be posted to github, which should include instructions to run the code. Only questions B and C will be required to be submitted in latex. No late homework will be accepted.

Rules: You are strongly encouraged to discuss the homework with your peers, in particular, piazza is a very good environment for discussing the homework. However, you need to write your own homework and you need disclose your sources.

A In this exercise you will uncover most of the details about the conjugate gradient iterations that were not covered in class. You will start from the notion of A -induced inner product, then you will explore several relations between the different

- (a) (2 points) Suppose that w_1, w_2, \dots, w_n are orthogonal, i.e., $\langle w_i, w_j \rangle = 0$ whenever $i \neq j$. Show that if v belongs to the span of w_1, w_2, \dots, w_n then

$$v = \sum_{j=1}^n \frac{\langle v, w_j \rangle}{\|w_j\|^2} w_j$$

Hint: Write the definition of $v \in \text{span}\{w_1, w_2, \dots, w_n\}$ with unknowns coefficients, then use the properties of the inner product to find each of those coefficients.

- (b) (4 points) Conjugate gradient aims to build new search directions p_n , which are A -conjugate to $\{p_j\}_{j=0}^{n-1}$. One way to achieve that is through a Gram-Schmidt procedure with a modified inner product. Suppose that r_0, r_1, \dots, r_{n^*} are the residuals of the CG iteration. Then we aim to create a A -orthogonal basis for $\text{span}\{r_0, r_1, \dots, r_n\}$ for $n \leq n^* - 1$.

Recall that Gram-Schmidt orthogonalization procedure starts with $p_0 = r_0$ and

$$p_n = r_n - \sum_{j=0}^{n-1} \frac{\langle r_n, p_j \rangle_A}{\|p_j\|_A^2} p_j \quad \text{for } 1 \leq n \leq n^* - 1, \quad (1)$$

where $n^* \leq N$ is the number of iterations to convergence, and $A \in \mathbb{R}^{N \times N}$ is symmetric.

- Explain why the number of iterations to convergence, n^* , may be strictly smaller than N .
- Prove by induction on n that

$$\langle p_n, p_j \rangle_A = 0 \quad \text{for } 0 \leq j < n \leq n^* - 1.$$

You will see in the sequel that the update rules of CG allows to have a only-three term recursion in (1).

- (c) (4 points) Given that $A \in \mathbb{R}^{N \times N}$ is a symmetric positive definite matrix then \mathbb{R}^N has an orthonormal basis of eigenvectors $\phi_1, \phi_2, \dots, \phi_N$:

$$A\phi_n = \lambda_n \phi_n \quad \text{and} \quad \langle \phi_n, \phi_j \rangle = \delta_{nj}.$$

Assuming we order the eigenvalues so that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, prove the following, for all $v, w \in \mathbb{R}^N$.

- $\langle Av, w \rangle = \sum_{n=1}^N \lambda_n \langle v, \phi_n \rangle \langle \phi_n, w \rangle$
- $\lambda_n > 0$ for $1 \leq n \leq N$.

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

iv. $\|Av\| \leq \lambda_N \|v\|$.

(d) (6 points) Deduce from the update formulas for p_n , w_n , and r_n that

$$p_{n+1} = (1 + \beta_n)p_n - \alpha_n A p_n - \beta_{n-1} p_{n-1} \quad \text{for } 1 \leq n \leq n^* - 2.$$

Hint: You may want to start with the update formula for p_{n+1} , and then replace the rest of expressions appropriately.

(e) (6 points) Prove that if $A \in \mathbb{R}^{N \times N}$ is non-singular, then A^N is a linear combination of $I, A, A^2, \dots, A^{N-1}$.
Hint: use the Cayley-Hamilton theorem.

(f) (8 points) For any $\alpha \neq 0$, the linear equation $Au = f$ is equivalent to

$$u = u + \alpha(f - Au).$$

In this context, the *Richardson iteration* is defined by

$$u_{n+1} = u_n + \alpha(f - Au_n).$$

i. Show that the error $e_n = u_n - u$ satisfies $e_{n+1} = (I - \alpha A)e_n$.

ii. Deduce that $\|e_{n+1}\| \leq \rho \|e_n\|$, where the error reduction factor is

$$\rho = \max_{1 \leq j \leq N} |1 - \alpha \lambda_j|.$$

It follows that $\|e_n\| \leq \rho^n \|e_0\|$, so if $\rho < 1$ then $e_n \rightarrow 0$ and hence the Richardson iterates u_n converge to the solution u .

iii. Prove that ρ is minimized by choosing

$$\alpha = \frac{2}{\lambda_1 + \lambda_N}$$

in which case

$$\rho = \frac{\lambda_N - \lambda_1}{\lambda_N + \lambda_1} = \frac{\kappa - 1}{\kappa + 1} < 1, \quad \text{where} \quad \kappa = \frac{\lambda_N}{\lambda_1}.$$

iv. Suppose that we do not know the exact values of the extremal eigenvalues, but only some one-sided bounds

$$0 < c \leq \lambda_1 \leq \lambda_N \leq C < \infty$$

For the (sub-optimal) choice $\alpha = 2/(c + C)$, show

$$\rho \leq \frac{C - c}{C + c} = \frac{\kappa' - 1}{\kappa' + 1} < 1, \quad \text{where} \quad \kappa' = \frac{C}{c}$$

(g) (10 points) Define the normalized CG residuals,

$$q_n = \frac{r_n}{\|r_n\|} \quad \text{for } 0 \leq n \leq n^* - 1$$

so that $\{q_0, \dots, q_{n-1}\}$ is an orthonormal basis for the Krylov space \mathcal{K}_n

i. Show that $r_1 = r_0 - \alpha_0 A r_0$

ii. Show that

$$r_{n+1} = r_n - \alpha_n A r_n + \frac{\alpha_n \beta_{n-1}}{\alpha_{n-1}} (r_n - r_{n-1}) \quad \text{for } 1 \leq n \leq n^* - 1$$

iii. Define

$$\gamma_0 = \frac{1}{\alpha_0} \quad \text{and} \quad \gamma_n = \frac{1}{\alpha_n} + \frac{\beta_{n-1}}{\alpha_{n-1}} \quad \text{for } 1 \leq n \leq n^* - 1$$

and

$$\delta_n = \frac{\sqrt{\beta_n}}{\alpha_n}$$

Deduce from parts i and ii that

$$\begin{aligned} A q_0 &= \gamma_0 q_0 - \delta_0 q_1 \\ A q_n &= -\delta_{n-1} q_{n-1} + \gamma_n q_n - \delta_n q_{n+1} \quad \text{for } 1 \leq n \leq n^* - 1 \end{aligned}$$

iv. Show that

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

where $Q = \begin{bmatrix} q_0 & q_1 & \cdots & q_{n-1} \end{bmatrix} \in \mathbb{R}^{N \times n}$ is orthogonal,

$$T_n = \begin{bmatrix} \gamma_0 & -\delta_0 & & \\ -\delta_0 & \gamma_1 & -\delta_1 & \\ & \ddots & \ddots & \ddots \\ & & -\delta_{n-3} & \gamma_{n-2} & -\delta_{n-2} \\ & & & -\delta_{n-2} & \gamma_{n-1} \end{bmatrix} \in \mathbb{R}^{n \times n} \quad (2)$$

is tridiagonal and $e_n = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^n$

v. Deduce that $Q_n^T A Q_n = T_n$

B (10 pts) Consider the function

$$f(x) = e^{-400(x-0.5)^2}$$

for $x \in [0, 1]$. Sample it on a grid $x_j = jh$ with $h = 1/N$ and $0 \leq j \leq N$, for some N to be determined. Consider the linear interpolant of f computed from the $N + 1$ samples $f(x_j)$. Numerically, find the smallest value of N such that f differs from its linear interpolant by at most 10^{-2} in the uniform norm. [Hint: Matlab has interp1.m for 1 D linear interpolation.]

C Consider the 2D wave equation

$$u_{tt} = \Delta u, \quad 0 \leq x, y \leq 1,$$

with homogeneous Dirichlet boundary conditions. Fix the initial conditions to be

$$u(x, y, 0) = 0, \quad u_t(x, y, 0) = f(x)f(y)$$

where f was defined in problem B. Consider a spatial grid $\mathbf{x}_j = (x_{j_1}, y_{j_2}) = (j_1 \Delta x, j_2 \Delta x)$ with Δx small enough to resolve the initial condition, in the sense of problem B. (I.e., take Δx less than $h = 1/N$, where the critical N was found in problem B.)

- (30 pts) Implement and test the "simplest" numerical method, which uses the 3-point formula for the second derivative in time, and the 5-point Laplacian at time t_n . It results in a two-step method. Explain how you initialize your scheme. Show a log-log plot of the error vs. the grid spacing Δx , and check from this plot that your method is second-order accurate.
- (5 pts) Consider the ODE $y''(t) = \lambda y$, and the 3-point rule for y'' as a two-step explicit time integrator. Find the region of stability of this ODE solver in terms of $\lambda(\Delta t)^2$, and plot it in the complex plane.
- (5 pts) From your answer to (b), and your knowledge of the spectrum of the discrete Laplacian, perform the "method of lines" stability analysis for the method in (a). What CFL condition does this analysis result in?
- (5 pts) Perform the von Neumann stability analysis for the method in (a), and check if the resulting CFL condition agrees with what you found in the previous question. [Hint: since this is a 2D problem, a plane wave is $\exp(ik_1 j_1 \Delta x) \exp(ik_2 j_2 \Delta x)$.]
- (5 pts) Find the modified equation that corresponds to the numerical method in (a). Solve it via Fourier series, and comment on the physics of the extra terms. Are they dissipative, dispersive, or something else?

D (Bonus 20 pts) Formulate and prove one extension of the Lax equivalence theorem to the case of linear ODE with two time derivatives.

In the next homework, you will be required to compare the finite difference approach you will develop here with spectral methods, thus you may want to thoroughly test your codes.