

CS 6210: HOMEWORK 4

Instructor: Anil Damle

Due: October 30, 2023

POLICIES

You may discuss the homework problems freely with other students, but please refrain from looking at their code or writeups (or sharing your own). Ultimately, you must implement your own code and write up your own solution to be turned in. Your solution, including plots and requested output from your code should be typeset and submitted via the Gradescope as a pdf file. Additionally, please submit any code written for the assignment. This can be done by either including it in your solution as an appendix, or uploading it as a zip file to the separate Gradescope assignment.

QUESTION 1:

Assume that we are using a stationary iterative method to solve $Ax = b$ with the splitting $A = M - N$ and initial guess $x^{(0)}$, and that $x^{(1)}, \dots, x^{(k)}$ have been computed using the iteration $Mx^{(j+1)} = Nx^{(j)} + b$. Normally we would consider $x^{(k)}$ as our current approximation of the solution. However, maybe there is some process that allows us to accelerate the convergence of our method and draw an interesting connection to Krylov subspace methods.

Let's consider one approach to accomplishing this goal. Specifically, we would like to construct coefficients $\{v_j^{(k)}\}_{j=1}^k$ for each iteration k such that

$$y^{(k)} = \sum_{j=0}^k v_j^{(k)} x^{(j)}$$

gives a better approximation to our true solution, denoted x . Let us define $G = M^{-1}N$ and

$$p_k(z) = \sum_{j=0}^k v_j^{(k)} z^j$$

- (a) Assuming $x^{(0)} = 0$, is $y^{(k)}$ consistently part of a Krylov subspace? If so, which Krylov subspace? (By consistently, I mean can you define a matrix H and vector w that do not depend on k such that $y^{(k)} \in \mathcal{K}_k(H, w)$.)
- (b) Now, let us further assume that $p_k(1) = 1$. Prove that

$$y^{(k)} - x = p_k(G) e^{(0)}$$

- (c) Prove that if B is similar to a Hermitian matrix then

$$\rho(p_k(B)) = \max_{\lambda_i \in \lambda(B)} |p_k(\lambda_i)|$$

where $\rho(p_k(B))$ is the spectral radius of $p_k(B)$.

We now assume that the iteration matrix G is similar to a Hermitian matrix and has real eigenvalues $\lambda_1, \dots, \lambda_n$. Recall that for convergence (from any initial guess) we require that

$$-1 < \lambda_n \leq \lambda_1 < 1.$$

Moreover, let α and β be such that

$$-1 < \alpha \leq \lambda_n \leq \dots \leq \lambda_1 \leq \beta < 1.$$

- (d) Since we may write $\|y^{(k)} - x\| = \|p_k(G)e^{(0)}\|$, what choice of p_k would give us zero error? Is such a choice feasible?

Since we may write

$$\max_{\lambda_i \in \lambda(B)} |p_k(\lambda_i)| \leq \max_{\alpha \leq \lambda \leq \beta} |p_k(\lambda)|,$$

it seems reasonable to pick p_k to be small on the interval $[\alpha, \beta]$. The ideal choice, given the constraint on $p_k(1)$, is a scaled and shifted version of the k^{th} Chebyshev polynomial. These polynomials may be defined by the recursion $c_j(z) = 2zc_{j-1}(z) - c_{j-2}(z)$ where $c_0(z) = 1$ and $c_1(z) = z$. Alternatively, we may write $c_j(z) = \cos(j\theta)$ where $\theta = \arccos(z)$. Specifically, we may choose our polynomial to be

$$p_k(z) = \frac{c_k\left(-1 + 2\frac{z-\alpha}{\beta-\alpha}\right)}{c_k(\mu)},$$

where $\mu = 1 + 2\frac{1-\beta}{\beta-\alpha}$. You can verify that $p_k(1) = 1$. Notably, $c_k(z)$ has the property that it is bounded between -1 and 1 in the interval $[-1, 1]$, but then grows rapidly outside of this interval. So, in the formula above, $c_k(\mu)$ becomes large as $k \rightarrow \infty$. With the chosen scaling we ensure that $p(z)$ is small in the interval $[\alpha, \beta]$ while satisfying $p(1) = 1$.

- (e) Given the above choice for $p_k(z)$ prove that there exists a constant C such that

$$\|y^{(k)} - x\|_2 \leq C \left(\frac{1}{c_k(\mu)} \right) \|x - x^{(0)}\|_2,$$

where C may depend on the matrix G .

- (f) Let $\alpha = -0.9$ and $\beta = 0.9$. Plot $c_k(\mu)$, on a logarithmic scale, for $k = 0, 1, \dots, 100$.

We will now consider using this acceleration method in conjunction with the Jacobi iteration. For the remainder of this problem assume that A is a real symmetric matrix that is strictly diagonally dominant and has positive diagonal entries.

- (g) Under the aforementioned assumptions, prove that the iteration matrix associated with A is similar to a Hermitian matrix.
- (h) Implement the Jacobi method both with and without Chebyshev acceleration. You can find pseudo-code in Golub and Van Loan 4th edition, section 11.2.8 (3rd edition, section 10.1.5) that leverages the three-term recurrence for Chebyshev polynomials for an efficient implementation.

Build some non-singular test problems via matrices A , and vectors $x^{(0)}$ and b where the eigenvalues of G approach ± 1 . You may use a built in routine to compute eigenvalues and set the eigenvalue bounds as $\alpha = (-1 + \lambda_n)/2$ and $\beta = (1 + \lambda_1)/2$. Use your algorithm both

with and without the acceleration to solve $Ax = b$. You may stop your algorithm when the 2 norm of the residual is less than 10^{-6} or you have run 1000 iterations. Provide error plots, on a logarithmic scale, of the 2 norm of the residual vs iteration both with and without the acceleration. Comment on your observations.

- (i) How does the scheme here relate to our discussion of Krylov methods?

QUESTION 2 (UNGRADED BUT INTERESTING; WE USED THIS RESULT IN CLASS):

Here we will characterize the behavior of A^k in terms of its spectral radius.

- (a) Prove that if $\rho(A) < 1$ then there exists a sub-multiplicative matrix norm $\|\cdot\|$ such that $\|A\| < 1$. Hint: start with the Schur decomposition and note that for any non-singular matrix S and norm $\|\cdot\|$, $\|A\|_S = \|S^{-1}AS\|$ is also a norm.
- (b) Given $A \in \mathbb{R}^{n \times n}$, a matrix norm $\|\cdot\|$, and some $\epsilon > 0$ prove that there exist constants α (depending on the norm) and $\beta_{A,\epsilon}$ (depending on A , the norm, and ϵ) such that

$$\alpha\rho(A)^k \leq \|A^k\| \leq \beta_{A,\epsilon}(\rho(A) + \epsilon)^k.$$

QUESTION 3:

We are now going to consider some details about how Krylov methods behave in specific situations.

- (a) Given a symmetric positive definite matrix A and vector b , prove that if the Lanczos process breaks down at some point (*i.e.* $\beta_k = 0$ using the notation from class and Trefethen and Bau) then the subspace $\mathcal{K}_k(A, b)$ contains a solution to the linear system $Ax = b$. In principle we might be worried that if $\beta_k = 0$ things have gone horribly wrong since we cannot construct the next vector in our orthonormal basis. However, this result shows that in this context everything has actually gone remarkably well.
- (b) Given a non-singular diagonalizable matrix A with at most p distinct eigenvalues and a vector b , show that a solution to $Ax = b$ exists in $\mathcal{K}_k(A, b)$ for some $k \leq p$. In other words, we certainly have a solution in the p^{th} Krylov subspace, though we may find one sooner in some special circumstances.

QUESTION 4:

Here, we will consider the applicability of Krylov methods to solving a set of closely related linear systems. Specifically, we are given a real symmetric $n \times n$ matrix A and a set of M real numbers $\{\sigma_i\}_{i=1}^M$ (you may assume none of the σ_i are eigenvalues of A), and we want to solve the set of M linear systems

$$(A - \sigma_i I) x_i = b$$

for $\{x_i\}_{i=1}^M$. You may assume we do not have any reason to use an initial guess besides $\vec{0}$ for all of the given linear systems.

Devise a Krylov subspace based iterative method to “simultaneously” solve this collection of linear systems in the sense that you construct M sequences of iterates each with the property that $x_i^{(k)} \rightarrow (A - \sigma_i)^{-1} b$ as $k \rightarrow \infty$. In addition, your algorithm must satisfy the following properties:

- Use no more than one matrix vector product with A at each iteration. A single iteration constitutes computing $x_i^{(k)}$ for $i = 1, \dots, M$.
- Converge (in exact arithmetic) for any σ_i that is not an eigenvalue of A and converge in at most ℓ iterations for every i if A has ℓ distinct eigenvalues.
- Have a storage cost that is naïvely $\mathcal{O}(Mnk)$ and computational complexity per iteration that is naïvely $T_{mult}(A) + \mathcal{O}(Mnk) + \mathcal{O}(Mk^3)$, but can be improved to $\mathcal{O}(Mn) + \mathcal{O}(Mk)$ and $T_{mult}(A) + \mathcal{O}(Mn)$ respectively. You do not have to work out all the details on the improvement, but you do have to make a convincing argument that such an improvement is possible.

Given this problem and set of requirements address the following:

- (a) State your algorithm for addressing the above problem and prove why it satisfies the desired criteria. Be sure to both clearly articulate your final algorithm and provide the desired proofs.
- (b) Let's say we are given a set of non-singular real symmetric preconditioners $M_i^{-1} \approx (A - \sigma_i)^{-1}$ and their Cholesky factorizations $M_i^{-1} = L_i L_i^T$. Do you think that these preconditioners can be incorporated into your algorithm without adversely impacting its computational benefits (*i.e.* could you devise an algorithm that is faster than solving the M problems independently)? Why or why not?