

# 18.335 Take-Home Midterm Exam: Spring 2023

Posted Friday 12:30pm April 14, due **11:59pm Monday April 17.**

## Problem 0: Honor code

Copy and sign the following in your solutions:

*I have not used any resources to complete this exam other than my own 18.335 notes, the textbook, running my own Julia code, and posted 18.335 course materials.*

---

your signature

## Problem 1: (32 points)

Given two real vectors  $u = (u_1, u_2, \dots, u_n)^T$  and  $v = (v_1, v_2, \dots, v_n)^T$ , computing the dot product  $f(u, v) = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = u^T v$  in floating point arithmetic with left to right summation is backward stable. The computed dot product  $\hat{f}(u, v)$  satisfies the *component-wise* backward error criteria

$$\hat{f}(u, v) = (u + \delta u)^T v, \quad \text{where} \quad |\delta u| \leq n \epsilon_{\text{mach}} |u| + \mathcal{O}(\epsilon_{\text{mach}}^2).$$

The notation  $|w|$  indicates the vector  $|w| = (|w_1|, |w_2|, \dots, |w_n|)^T$ , i.e., the vector obtained by taking the absolute value of each entry of  $w$ .

- (a) Using the dot product algorithm  $\hat{f}(u, v)$ , derive an algorithm  $\hat{g}(A, b)$  for computing the matrix-vector product  $g(A, b) = Ab$  in floating point arithmetic, and show that it satisfies the component-wise backward stability criteria

$$\hat{g}(A, b) = (A + \delta A)b, \quad \text{where} \quad |\delta A| \leq n \epsilon_{\text{mach}} |A| + \mathcal{O}(\epsilon_{\text{mach}}^2),$$

where the notation  $|B|$  indicates the matrix obtained by taking the absolute value of each entry of  $B$ .

- (b) Suppose the algorithm  $\hat{g}(A, b)$  is used to compute matrix-matrix products  $C = AB$  by computing one column of the matrix  $C$  at a time. Is the resulting floating-point algorithm  $\hat{h}(A, B)$  component-wise backward stable in the sense that there is a matrix  $\delta A$  such that

$$\hat{h}(A, B) = (A + \delta A)B, \quad \text{where} \quad |\delta A| \leq n \epsilon_{\text{mach}} |A| + \mathcal{O}(\epsilon_{\text{mach}}^2)?$$

Explain why or why not.

## Problem 2: (32 points)

Given an  $n$ -dimensional subspace  $\mathcal{V}$ , the standard Rayleigh–Ritz projection approximates a few ( $n \ll m$ ) eigenvalues of an  $m \times m$  matrix  $A$  by finding a scalar  $\lambda$  and  $x \in \mathcal{V}$  such that  $Ax - \lambda x \perp \mathcal{V}$ , i.e., the residual is perpendicular to the subspace. A *two-sided* Rayleigh–Ritz projection uses a second subspace  $\mathcal{W}$  (not orthogonal to  $\mathcal{V}$ ) and searches for a scalar  $\lambda$  and  $x \in \mathcal{V}$  such that

$$Ax - \lambda x \perp \mathcal{W}, \quad \text{and} \quad x \in \mathcal{V}, \quad (1)$$

i.e., the residual is perpendicular to the *second* subspace. In this problem,  $A$  is diagonalizable.

- (a) Let  $V$  and  $W$  be a pair of bases for  $\mathcal{V}$  and  $\mathcal{W}$ , and let  $\lambda$  (finite) and  $w$  solve the eigenvalue problem  $Bw = \lambda Mw$ , where  $B = W^T A V$  and  $M = W^T V$ . Show that  $\lambda$  and  $x = Vw$  satisfy the criteria in (1).

- (b) Suppose that  $\mathcal{V} = \text{span}\{x_1, \dots, x_n\}$  and  $\mathcal{W} = \text{span}\{y_1, \dots, y_n\}$ , where  $Ax_i = \lambda_i x_i$  and  $y_i^T A = \lambda_i y_i^T$  for  $i = 1, \dots, n$ , are a pair of  $n$ -dimensional *right and left invariant subspaces* of  $A$ . If the bases  $V$  and  $W$  are chosen to be *bi-orthonormal*, meaning that  $W^T V = I$ , show that  $\lambda$  and  $x$  from part (a) are an eigenpair of the full  $m \times m$  matrix  $A$ .

**Hint 1:** In part (b), consider the similarity transform  $[W \ W_\perp]^T A [V \ V_\perp]$ , where  $V_\perp$  and  $W_\perp$  are a biorthonormal basis for the orthogonal complements,  $\mathcal{V}^\perp$  and  $\mathcal{W}^\perp$ , respectively. **Hint 2:** The right and left eigenvectors of a diagonalizable matrix can be made biorthonormal (why?).

## Problem 2: (36 points)

The method of Generalized Minimal RESiduals (GMRES) uses  $n$  iterations of the Arnoldi method to construct a sequence of approximate solutions  $x_1, x_2, \dots, x_n$  to the  $m \times m$  linear system  $Ax = b$ . At the  $n^{\text{th}}$  iteration, the approximate solution  $x_n = Q_n y_n$  is constructed by solving the least-squares problem,

$$y_n = \underset{y}{\text{argmin}} \|\tilde{H}_n y - \|b\|e_1\|,$$

where  $\tilde{H}_n$  is an  $(n+1) \times n$  upper Hessenberg matrix and  $Q_n$  is the usual orthonormal basis for the Krylov subspace  $\mathcal{K}_n(A, b) = \text{span}\{b, Ab, A^2b, \dots, A^{n-1}b\}$ .

- Describe an algorithm based on Givens rotations that exploits the upper Hessenberg structure of  $\tilde{H}_n$  to solve the  $(n+1) \times n$  least-squares problem in  $\mathcal{O}(n^2)$  flops.
- If the QR factorization  $\tilde{H}_{n-1} = \Omega_{n-1} R_{n-1}$  is known from the previous iteration, explain how to update the QR factorization to  $\tilde{H}_n = \Omega_n R_n$  cheaply using a single Givens rotation.
- Using your result from part (b), explain how the solution to the least-squares problem can also be updated cheaply from the solution at the previous iteration.
- What is the approximate flop count for updating the least-squares solution at the  $n^{\text{th}}$  step of GMRES? You may use big- $\mathcal{O}$  notation to express the asymptotic scaling in  $n$ .