Math 501 **Name:** Qilin Ye
Spring 2021
Final
May 7, 2021
Time: 11:00 am -1:00 pm

This exam contains 8 pages (including this cover page) and 5 questions. Total of points is 100.

- **Show your work.**
- an A4 cheatsheet (two sides) is allowed.
- No calculator, cell/smart phone or other electronic device.

Grade Table (for teacher use only)

Question	Points	Score
1	16	
2	24	
3	20	
4	20	
5	20	
Total:	100	

MATH 501 Final Exam

Qilin Ye

May 7, 2021

Problem 1

Prove that if the eigenvalues of A satisfy $|\lambda_1| > |\lambda_i|$ for all other i's then

$$\lambda_1 = \lim_{m\to\infty} \operatorname{tr}(A^{m+1})/\operatorname{tr}(A^m).$$

Proof. Notice that if λ is an eigenvalue of A, then λ^k is one of A^k, and they correspond to the same eigenvector x:

$$A^k x = A^{k-1}(Ax) = \lambda A^{k-1}x = \lambda A^{k-2}(Ax) = \ldots = \lambda^k x.$$

Therefore $\operatorname{tr}(A^k) = \lambda_1^k + \lambda_2^k + \ldots + \lambda_n^k$. It follows that

$$\begin{aligned}\lim_{m\to\infty} \frac{\operatorname{tr}(A^{m+1})}{\operatorname{tr}(A^m)} &= \lim_{m\to\infty} \frac{\lambda_1^{m+1} + \ldots + \lambda_n^{m+1}}{\lambda_1^m + \ldots + \lambda_n^m} \\ &= \lim_{m\to\infty}\left[\lambda_1 + \frac{(\lambda_1-\lambda_2)\lambda_2^m + \ldots + (\lambda_1-\lambda_n)\lambda_n^m}{\lambda_1^m + \ldots + \lambda_n^m}\right]\end{aligned}$$

where the large fraction tends to 0 as $m \to \infty$ because $(\lambda_i/\lambda_1)^m \to 0$ for all $i \geqslant 2$. Therefore the limit is λ_1. ☐

Problem 2

(a) Prove that if $A_{m\times n}$ is of rank n then A^*A is nonsingular.

(b) Prove that if $A_{m\times n}$ is of rank n then A^*A is Hermitian and positive definite.

(c) Prove that if $A_{m\times n}$ is of rank n then $A^+ = (A^*A)^{-1}A^*$. *In the proof please clarify the size of matrices at each step.*

Proof.

(a) If $A^*Ax = 0$ then in particular $x^*A^*Ax = 0$, and so

$$0 = x^*A^*Ax = \langle Ax, Ax\rangle = \|Ax\|^2$$

which implies $Ax = 0$. But since A has full column rank, this implies $x = 0$. Thus A^*A is invertible.

(b) That A^*A is Hermitian is trivial: $(A^*A)^* = A^*A^{**} = A^*A$. For positive definiteness:

$$x^*A^*Ax = \langle Ax, Ax\rangle = \|Ax\|^2 \geqslant 0,$$

and since it is of full column rank, 0 can be obtained if and only if $x = 0$.

(c) For this one, it suffices to check that $(A^*A)^{-1}A^*$ satisfies all Penrose properties:

$$A_{m\times n}[(A^*A)^{-1}A^*]_{n\times m}A_{m\times n} = A_{m\times n}\underbrace{(A^*A)^{-1}(A^*A)}_{=I_{n\times n}} = A \tag{1}$$

$$[(A^*A)^{-1}A^*]_{n\times m}A_{m\times n}[(A^*A)^{-1}A^*]_{n\times m} = (A^*A)^{-1}_{n\times n}\underbrace{(A^*A)(A^*A)^{-1}}_{I_{n\times n}}A^*_{n\times m} = (A^*A)^{-1}A^* \tag{2}$$

$$\begin{aligned}(A_{m\times n}[(A^*A)^{-1}A^*]_{n\times m})^* &= A^{**}_{m\times n}((A^*A)^{-1})^*_{n\times n}A^* \\ &= A_{m\times n}((A^*A)^*)^{-1}_{n\times n}A^* \\ [2\text{(b)}: A^*A \text{ Hermitian}] &= A((A^*A)^*)^{-1}A = A_{m\times n}[(A^*A)^{-1}A^*]_{n\times m}.\end{aligned} \tag{3}$$

$$I^*_{n\times n} = ([(A^*A)^{-1}A^*]_{n\times m}A_{m\times n})^* = A^*_{n\times m}A^{**}_{m\times n}((A^*A)^{-1})^*_{n\times n} = A^*A(A^*A)^{-1} = I_{n\times n}. \tag{4}$$

□

Problem 3

Let $\|A\|_2$ denote the matrix norm subordinate to the Euclidean 2-norm. Let the singular values of A be $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_n$. Prove that $\|A\|_2 = \sigma_1$.

Proof. First notice that if U is any unitary matrix then for x, a vector of corresponding size, $\|Ux\| = \|x\|$. Indeed,

$$\|Ux\|^2 = \langle Ux, Ux\rangle = \langle x, U^*Ux\rangle = \langle x, x\rangle = \|x\|^2$$

(so it holds in particular for $\|\cdot\|_2$). Now let $A = U\Sigma V^T$ be the singular value decomposition of A. It follows that

$$\begin{aligned}\|A\|_2 = \sup_{\|x\|_2=1}\|Ax\|_2 &= \sup_{\|x\|_2=1}\|U\Sigma V^Tx\|_2 \\ &= \sup_{\|x\|_2=1}\|DQx\|_2 && (U \text{ unitary}) \\ &= \sup_{\|y\|_2=1}\|Dy\|_2 && (y := V^Tx;\ V^T \text{ unitary}) \\ &= \|De^{(1)}\|_2 = \sqrt{\sigma_1^2} = |\sigma_1| = \sigma_1.\end{aligned}$$

□

Problem 4

(a) Consider $f(x) := x^m$, $m \in \mathbb{N}$. Show that

$$f[x_0, \ldots, x_n] = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n > m. \end{cases}$$

Proof. (a) If we can prove that when $n = m$ the divided difference is 1 then we are done, as

$$f[x_0, \ldots, x_{m+1}] = \frac{f[x_1, \ldots, x_{m+1}] - f[x_0, \ldots, x_m]}{x_{m+1} - x_0} = \frac{1-1}{x_{m+1} - x_0} = 0,$$

and inductively one can show that the same holds for any $n > m$. But indeed this is true, since by the MVT for divided difference, if $x_0, ..., x_m \in [a,b]$ then

$$f[x_0, ..., x_m] = \frac{f^{(m)}(\xi)}{m!} \qquad \text{for some } \xi \in [a,b].$$

On the other hand, regardless of the value of ξ, the m^{th} order derivative of ξ^m is $m(m-1)... = m!$, so the fraction simplifies nicely to 1, proving our claim.

(b) We approach this identity by induction. The base case is obviously true:

$$f[x_0] = f(x_0) = \sum_{i=0}^{0}\left[f(x_i)\prod_{j\neq i}(x_i - x_j)^{-1}\right].$$

Now for the inductive step, we suppose that the equation holds for any divided difference with k indeterminate, that is, we have

$$f[x_0, ..., x_k] = \sum_{i=0}^{k}\left[f(x_i)\prod_{j\neq i}(x_i - x_j)^{-1}\right] \textbf{ and } f[x_1, ..., x_{k+1}] = \sum_{m=1}^{k+1}\left[f(x_m)\prod_{\ell\neq m}(x_m - x_\ell)^{-1}\right].$$

Recall the recursive formula that defines the divided difference:

$$f[x_0, ..., x_{k+1}] = \frac{f[x_1, ..., x_{k+1}] - f[x_0, ..., x_k]}{x_{k+1} - x_0}.$$

Substituting our induction hypothesis into this, we recover precisely the formula we want for $f[x_0, ..., x_{k+1}]$. To avoid cumbersome notation, I will not use a chain of equations but instead directly compute the coefficient of $f(x_i)$. For $f(x_0)$, since it is only determined by $-f[x_0, ..., x_k]$, we have coefficient

$$-\frac{\prod_{j=1}^{k}(x_0 - x_j)^{-1}}{x_{k+1} - x_0} = \prod_{j\neq 0}(x_0 - x_j)^{-1}.$$

Similarly, the coefficient of $f(x_{k+1})$ is only determined by $f[x_1, ..., x_{k+1}]$ so it is

$$\frac{\prod_{\ell=0}^{k}(x_{k+1} - x_\ell)^{-1}}{x_{k+1} - x_0} = \prod_{\ell\neq k+1}(x_{k+1} - x_\ell)^{-1}.$$

For other terms, i.e., $f(x_i)$ with $1 \leqslant i \leqslant k$, the coefficient is determined by both $f[x_0, ..., x_k]$ and $f[x_1, ..., x_{k+1}]$:

$$\frac{\prod_{\ell=1,\ell\neq i}^{k+1}(x_i - x_\ell)^{-1} - \prod_{j=0,j\neq i}^{k}(x_i - x_j)^{-1}}{x_{k+1} - x_0}$$

and writing the difference in terms of common denominator gives

$$((x_i - x_0) - (x_i - x_{k+1}))\prod_{\substack{j=0\\ j\neq i}}^{k+1}(x_i - x_j)^{-1} \cdot (x_{k+1} - x_0)^{-1} = \prod_{\substack{j=0\\ j\neq i}}(x_i - x_j)^{-1}.$$

This shows that all coefficients match, and we are done with our induction hypothesis. The claim follows. □

Problem 5

(a) Suppose A is an $m \times n$ matrix. Show that for any $x \in \mathbb{C}^m$, AA^+x is the best approximation of x (w.r.t. $\|\cdot\|_2$) in the column space of A.

(b) Show that if $\{u_1, u_2, \ldots\}$ is an orthonormal basis for an inner-product space E, then

$$P_n f := \sum_{i=1}^{n} \langle f, u_i \rangle u_i$$

is the best approximation of f in $U_n := \operatorname{span}\{u_1, \ldots, u_n\}$.

Proof.

(1) Recall (from a theorem proven in class) that the minimal solution to $Ay = x$ is given by $y = A^+x$. Therefore $x - Ay$ attains minimal 2-norm when $y = A^+x$ i.e., $Ay = AA^+x$ is the best approximation of x in the column space of A.

(2) We first show that $f - P_n f$ is orthogonal to $P_n f$. In particular, we show that $f - P_n f$ is orthogonal to all u_i for $1 \leqslant i \leqslant n$ (so that it must be orthogonal to everything in $\operatorname{span}\{u_1, \ldots, u_n\}$, in which there lies $P_n f$).

$$\langle f - P_n f, u_i \rangle = \langle f, u_i \rangle - \langle P_n f, u_i \rangle = \langle f, u_i \rangle - \langle \sum_{i=1}^{n} \langle f, u_i \rangle u_i, u_i \rangle = \langle f, u_i \rangle - \langle f, u_i \rangle = 0$$

(since if $i \neq j$ then $\langle \langle f, u_j \rangle u_j, u_i \rangle = \langle f, u_j \rangle \langle u_j, u_i \rangle = 0$). Now, if we have $P_n f' \in \operatorname{span}\{u_1, \ldots, u_n\}$, then

$$\underbrace{(f - P_n f)}_{\in \text{span}^\perp} \perp \underbrace{(P_n f - P_n f')}_{\in \text{span}} \implies \|f - P_n f'\|^2 = \|f - P_n f\|^2 + \|P_n f - P_n f'\|^2 \geqslant \|f - P_n f\|^2,$$

proving the minimality of $\|f - P_n f\|$. □