Sheet 1 Section B

Question 4.
From the definition of the vector 1-norm, show that

$$\|A\|_1 = \max_j \sum_i \left|a_{ij}\right|.$$

Attempted Solution: WLOG, assume $A \in \mathbb{R}^{m\times n}$. We need to show 2 things:

(i) $\|A\|_1 \le \max_{1\le j\le n} \sum_{i=1}^m |a_{ij}|$;

(ii) This upper bound is attainable.

For (i), from the definition of vector 1-norm, we have

$$\begin{aligned}
\|A\|_1 &= \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_1}{\|x\|_1} = \max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \|Ax\|_1 && (1)\\
&= \max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \sum_{i=1}^m |(Ax)_i| && (2)\\
&= \max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \sum_{i=1}^n \left|\sum_{j=1}^n a_{ij}x_j\right| && (3)\\
&\le \max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \sum_{i=1}^n \sum_{j=1}^n \left|a_{ij}\right|\left|x_j\right| && (4)\\
&= \max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \sum_{j=1}^n \left(\left|x_j\right| \sum_{i=1}^m \left|a_{ij}\right|\right) && (5)\\
&\le \left(\max_{\substack{x\in\mathbb{R}^n\\ \|x\|_1=1}} \sum_{j=1}^n \left|x_j\right|\right)\left(\max_{1\le j\le n} \sum_{i=1}^m \left|a_{ij}\right|\right) && (6)\\
&= \max_{1\le j\le n} \sum_{i=1}^m \left|a_{ij}\right|. && (7)
\end{aligned}$$

(1) (2) are by the definition of vector induced norm and vector norm; (3) is by the definition of matrix product; (4) is by triangle inequality; (5) is due to finite sums; (6) is simple calculation; (7) is by $\|x\|_1 = 1$.

For (ii), let $\hat{j} = \operatorname{argmax}_{1\le j\le n} \sum_{i=1}^m |a_{ij}|$. Define x by $x_j = 0$ if $j \ne \hat{j}$ and $x_{\hat{j}} = 1$. Then we have $\|Ax\|_1 = \max_{1\le j\le n} \sum_{i=1}^m \left|a_{ij}\right|$.

Question 5.

Full SVD. Prove the existence of $A = U \begin{bmatrix} \Sigma \\ 0_{(m-n)\times n} \end{bmatrix} V^*$, where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary matrices i.e., $U^*U = I_m$ and $V^*V = I_n$, and $\Sigma \in \mathbb{R}^{n\times n}$ is diagonal.

Attempted Solution: There are essentially 2 (maybe more) ways of proving the existence of the full SVD of any matrix.

The *first* way is using induction. Assume WLOG that $A \in \mathbb{C}^{m\times n}$ with $m \geq n$. Since we are in finite dimensional [1] vector spaces, $\exists v_1 \in \mathbb{C}^n$ s.t. $\|v_1\|_2 = 1$ and $\|Av_1\|_2 = \|A\|_2$. Denote $\|A\|_2 = \sigma_1$. Then $\exists u_1 \in \mathbb{C}^m$ s.t. $\|u\|_2 = 1$ and $Av_1 = \sigma_1 u_1$ as we can take $u_1 = \frac{Av_1}{\|Av_1\|_2}$. Now construct orthogonal matrices U_1 and V_1 as follows:

$$V_1 = \left[v_1, \hat{V}_1\right] \in \mathbb{C}^{n\times n}, \quad U_1 = \left[u_1, \hat{U}_1\right] \in \mathbb{C}^{m\times m}$$

where $\hat{V}_1 \in \mathbb{C}^{n\times(n-1)}$ and $\hat{U}_1 \in \mathbb{C}^{m\times(m-1)}$ are matrices that complete v_1 and u_1 to orthonormal bases. This can be done via the Gram-Schmidt orthogonalisation. Consider the matrix S given as follows:

$$S := U_1^\top A V_1 = \begin{bmatrix} u_1^\top \\ \hat{U}_1 \end{bmatrix} A \begin{bmatrix} v_1 & \hat{V}_1 \end{bmatrix} = \begin{bmatrix} u_1^\top \\ \hat{U}_1 \end{bmatrix} \begin{bmatrix} \sigma_1 u_1 & A\hat{V}_1 \end{bmatrix} = \begin{bmatrix} \sigma_1 & w^\top \\ 0 & B \end{bmatrix},$$

where $w^\top = u_1^\top A \hat{V}_1 \in mathbbC^{n-1}$ and $B = \hat{U}_1 A \hat{V}_1 \in \mathbb{C}^{(m-1)\times(n-1)}$. Now, I claim that $w = 0$. Indeed, observe that

$$\|S\|_2^2 \geq \frac{\left\| S \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2^2}{\left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2^2} = \frac{\left\| \begin{bmatrix} \sigma_1^2 + w^\top w \\ Bw \end{bmatrix} \right\|_2^2}{\sigma_1^2 + w^\top w} \geq \frac{\left(\sigma_1^2 + w^\top w\right)^2}{\sigma_1^2 + w^\top w} = \sigma_1^2 + \|w\|_2^2.$$

On the other hand,

$$\|S\|_2^2 = \left\|U_1^\top A V_1\right\|_2^2 = \|A\|_2^2 = \sigma_1^2.$$

These two statements imply that $\|w\|_2 = 0$ and hence $w = 0$ which proves the claim. By the inductive hypothesis, assume the SVD of B exists s.t. $B = \hat{U}\hat{\Sigma}\hat{V}^\top$. Then

$$A = U_1 S V_1^\top = U_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{U} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{\Sigma} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \hat{V}^\top \end{bmatrix} V_1^\top.$$

Define $U = U_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{U} \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 0 & \hat{V}^\top \end{bmatrix} V_1^\top$. The induction will reach a column vector (since $m \geq n$) for which the existence of the SVD trivially holds.

The *second* method is by exploiting the eigendecomposition for symmetric matrices A^*A which is very similar to the one given in the lecture notes except now we are not assuming that A is full rank.

[1] In a finite-dimensional vector space, such as $\mathbb{C}^n$, the unit sphere $S^{n-1} = \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$ is compact. The function $f(v) = \|Av\|_2$ is continuous in v on the compact set S^{n-1}. By the extreme value theorem, $f(v)$ achieves a maximum on the unit sphere. That is, there exists some unit vector $v_1 \in S^{n-1}$ s.t. $f(v_1) = \max_{v\in S^{n-1}} \|Av\|_2$ This means that v_1 maximizes $\|Av\|_2$, and hence $\sigma_1 = \|Av_1\|_2 = \max_{v\neq 0} \frac{\|Av\|_2}{\|v\|_2} = \|A\|_2$.

Question 6.
What is the SVD of a normal matrix A, with respect to the eigenvalues and eigenvectors? What if A is (real) symmetric? And unitary?

Attempted Solution:

- *Normal matrices.* It is well-knonw that a matrix $A \in \mathbb{C}^{n\times n}$ is normal iff A is unitarily diagonalisable i.e. $A = U\Lambda U^*$ where Λ is diagonal and U is unitary (check this!). However, Λ can have complex or negative entries on its diagonal which forbids Λ to be the Σ we would like. The resolution to this problem is the following technique. Let $\Lambda = \text{diag}(\lambda_i)$ where $\lambda_i = r_i e^{i\theta}$ with $r_i \geq 0$. Define E by $E = \text{diag}(e^{i\theta})$, then observe that $\Lambda = |\Lambda|E$ and E is obviously an orthogonal matrix. Thus, the SVD of a nomral matrix A is $A = U|\Lambda|(EU^*)$.

- *Symmetric matrices.* It is well-known that a symmetric matrix A always have an eigendecomposition with real eigenvalues and orthogonal eigenvectors i.e. $A = U\Lambda U^*$ where Λ is real and diagonal. Then A can be written as $A = U|\Lambda|FU^*$ where $F = \text{diag}\left(\text{sign}(\lambda_i(A))\right)$. Now it remains to use a permutation matrix on both sides of $|\Lambda|$ to reorder the singular values.

- *Unitary matrices.* We can take the SVD of a unitary matrix A to be $A = AII$ and the singular values of A are all 1. It is also easy to note that the eigenvalues of unitary matrices have modulus 1 and hence $\sigma_i = |\lambda_i|$. However, it is in general not the case that the columns of A are its eigenvectors. For instance, an eigenvector of $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is $[1, 1]^\top$.

Question 7.
If $A \in \mathbb{R}^{n\times n}$ is nonsingular, what is the SVD of A^{-1} in terms of that of A ?

Attempted Solution: Let the SVD of A be $A = U\Sigma V^\top$ where $U, \Sigma, V \in \mathbb{R}^{n\times n}$. Since A is nonsingular, all singular values of A are strictly postive i.e. $\sigma_i > 0, \forall 1 \leq i \leq n$. Note that $(U\Sigma V^\top)\left(V\Sigma^{-1}U^\top\right) = I$ as U and V are orthogonal. Now the "SVD" of A^{-1} is given by $A^{-1} = V\Sigma^{-1}U^\top$ except for the fact that the singular values of A^{-1} are ordered in increasing order. By reordering the diagonal entries of Σ^{-1} using a permutation matrix of the form

$$P = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \iddots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

we obtain the SVD of A to be $A = \tilde{V}\tilde{\Sigma}\tilde{U}$, where $\tilde{V} = VP$, $\tilde{\Sigma} = P\Sigma^{-1}P$, and $\tilde{U} = PU^\top$.

NB. Let A be an n by n matrix. Multiplying the above permutation matrix P from the left reorder the columns of A and multiplying P from the right reorder the rows of A.

Question 8.
Let B be a square $n \times n$ matrix. Bound the i th singular values of AB using $\sigma_i(A)$ and $\sigma_i(B)$: Specifically, prove that for each i,

$$\sigma_i(A)\sigma_n(B) \leq \sigma_i(AB) \leq \sigma_i(A)\sigma_1(B)$$

Attempted Solution: WLOG assume that $A \in \mathbb{R}^{m\times n}$ with $m \leq n$. We have

$$\begin{aligned}
\sigma_i(AB) &= \max_{\dim(S)=i} \min_{\substack{x\in S\\ \|x\|_2=1}} \|ABx\|_2 && (8)\\
&\leq \max_{\dim(S)=i} \min_{\substack{x\in S\\ \|x\|_2=1}} \|A\|_2\|Bx\|_2 && (9)\\
&= \sigma_1(A) \max_{\dim(S)=i} \min_{\substack{x\in S\\ \|x\|_2=1}} \|Bx\|_2 && (10)\\
&= \sigma_1(A)\sigma_i(B). && (11)
\end{aligned}$$

(8) (11) are by Courant-Fisher minmax theorem; (9): any vector-induced matrix norm is subordinate; (10): $\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}$.

Then, it follows that $\sigma_i(AB) = \sigma_i((AB)^\top) = \sigma_i(B^\top A^\top) \leq \sigma_1(B^\top)\sigma_i(A^\top) = \sigma_i(B)\sigma_i(A)$ since $\sigma_i(A) = \sigma_i(A^\top)$ for any matrix A.

For the lower bound, we use the following fact which can be proven easily: $\sigma_n(A) = \min_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_2}{\|x\|_2}$. This fact implies that $\|Ax\|_2 \geq \sigma_n(A)\|x\|_2 \ \forall x \in \mathbb{R}^n$ and hence $\|ABx\|_2 \geq \sigma_n(A)\|Bx\|_2$. We have

$$\begin{aligned}
\sigma_i(AB) &= \max_{\dim(S)=i} \min_{\substack{x\in S\\ \|x\|_2=1}} \|ABx\|_2\\
&\geq \sigma_n(A) \max_{\dim(S)=i} \min_{\substack{x\in S\\ \|x\|_2=1}} \|Bx\|_2\\
&= \sigma_n(A)\sigma_i(B).
\end{aligned}$$

Similarly, it follows that $\sigma_i(AB) = \sigma_i((AB)^\top) = \sigma_i(B^\top A^\top) \geq \sigma_n(B^\top)\sigma_i(A^\top) = \sigma_n(B)\sigma_i(A)$. And both the upper bound and lower bound hold for all $1 \leq i \leq n$.