HW5:

(1a) The SVD of matrix X is given by

$X = U \Sigma V^T$ — orthogonal, orthogonal, Diagonal — and the solution for PCA assuming zero mean is $XX^T = U \Lambda U^T$

$\therefore \; XX^T = (U\Sigma V^T)(U \Sigma V^T)^T$

$= U \Sigma V^T U^T \Sigma^T V \quad \& \; \Sigma^T = \Sigma$ as Σ is Diaganol

$= U \Sigma^2 U^T V^T V$ and $V^T V = I$ denity by matrix properties

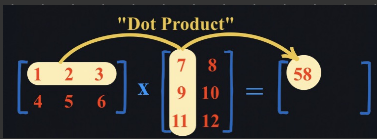
$\therefore \; = U \Sigma^2 U^T = XX^T = U \Lambda U^T$ hence showing the relationship btwn (SVD) of X & XX^T with $\Lambda = \Sigma^2$ as U matrices are equivalent $\therefore$ the eigenvalues of XX^T in Λ equal the singular values of X in Σ.

Now, as each column of U is an Eigenvector with the direction of greatest variance and, as seen in lecture, our eigen values λ along the diagonal $\Lambda = \Sigma^2$ show the total variance of our dataset along it's colum basis (i.e. $(XX^T)u_d = \lambda_d u_d$ slide 44 lec 9) with the largest λ_d values indicating the greatest variance $\therefore$ the principal components (PC) of X are along the columns of U from PCA and as U from PCA equals U from SVD as shown above $\therefore$ the columns of U are the PC of X.

(1b) In part a we showed that λ (the eigenvalues) of XX^T along the diagonal Λ are the square of the singular values of X hence the eigen values of the PCA of X cannot be negative hence they are non-negative as a result of squaring. Also as seen of slide 50 lec 9 we know each eigen value λ_d is the total variance along u_d $\therefore$ we can see intuitively that λ_d is non-negative as total variance cannot be negative as by definition variance is non-negative. Lastly, from lecture 9 slide 50, we know $\lambda_d = \sum_{i=1}^{N} (x_i'^{(d)})^2$ $\therefore$ λ_d cannot be negative as it is the sum of squares.

(1c)

Base: Using the hint we will prove first that $Tr(AB) = Tr(BC)$ for square matrices A & B of the same size (i.e. $N \times N$) $\therefore$ we have $\underbrace{\sum_{i=1}^{N} (AB)_{ii}}_{\text{(by hint)}} = \sum_{i=1}^{N} \sum_{j=1}^{N} (A_{ij})(B_{ji}) = Tr(AB)$ by the definition of matrix multiplication:



Note: to expand $(AB)_{ii}$ in terms of another sum we can write $(AB)_{ii} = \sum_{j=1}^{N} A_{ij} B_{ji}$ by the definition of matrix multiplication as seen in the photo stating that (i,j)th entry of (AB) is given by the dot product of the i-th row of A and j-th column of B.

Now looking at $Tr(BA)$ we observe that $Tr(BA) = \sum_{i=1}^{N} (BA)_{ii} = \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} A_{ji}$ and with simple change in variable names we see that $Tr(BA) = \sum_{j=1}^{N} \sum_{i=1}^{N} B_{ji} A_{ij}$

$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ji} A_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} B_{ji} = Tr(BA) = Tr(AB)$ hence proven.

(by transitive properties of matrix multiplication and summation properties.)

Inductive Step: Assume the idenity holds for $K \geq 2$ matrices.

- Now, we need to generalize this for $K+1$ many $N \times N$ square matrices $X_1, X_2, X_3, \ldots, X_K, X_{K+1}$. Given $K+1$ many of these square matrices we need to show $Tr(X_1 X_2 \cdots X_K) = Tr(X_2 X_3 \cdots X_K X_{K+1} X_1) = \cdots = Tr(X_K X_{K-1} \cdots X_1)$.

First, apply the idenity for K matrices $X_2, X_3 \ldots X_{K+1}$ $\therefore$

$$Tr(X_2 X_3 \cdots X_{K+1} X_1) = Tr(X_3 X_4 \cdots X_{K+1} X_1 X_2) = \cdots = Tr(X_{K+1} X_1 X_2 \cdots X_K)$$

Then as $Tr(AB) = T(BA)$ by our base case then we can interchange X_1 with X_{K+1} $\therefore$ $Tr(X_{K+1} X_1 X_2 \cdots X_K) = Tr(X_1 X_{K+1} X_2 \cdots X_K)$ and repeat this cycle until X_1 & X_{K+1} appear in their correct original position hence proven for $K+1$ matrices $\therefore$ for any number of matrices.

(1d) We need $(2N+1)K$ values to store the truncated SVD with K singular values as we have N rows for each point and K columns for each new basis axis for matrices U & V hence $2NK$ values. Then, for our diagonal matrix Σ, we have only our K singular values along the diagonal hence we have to store only those K-values $\therefore$ yielding $2NK + K$ total values and $2NK + K = (2N+1)K < N^2$ $\therefore$ making storing the truncated SVD more efficient for $K < \frac{N^2}{(2N+1)}$ bcz $(2N+1)K < N^2$.

(1E) We want to show that $U\Sigma = U'\Sigma'$ → $N\times N$ matrix consisting of the first N rows of Σ
↳ $D\times N$ matrix consisting of the first N columns of U.

and similar to part C where
$(AB)_{ii} = \sum_{j=1}^{N} A_{ij}B_{ji}$ ∴ we can extend this to $(AB)_{ij} = \sum_{k=1}^{M} A_{ik}B_{kj}$ for the element of the product of matrices A of size $(l\times M)$ and B of size $(M\times P)$. Now apply this generalization to $U\Sigma$ s.t. U has size $(D\times D) = A$ & Σ with size $(D\times N) = B$ ∴

$(U\Sigma)_{ij} = \sum_{k=1}^{D} U_{ik}\Sigma_{kj} = \sum_{k=1}^{N} U_{ik}\Sigma_{kj} + \sum_{k=N+1}^{D} U_{ik}\Sigma_{kj}$ as Σ has non-zeroes $\forall\, 1\le i\le N$ with rank $x=N$ on Σ_{ii} ∴ $\Sigma_{kj} = 0$ $\forall\, k > N$ ∴ $\sum_{k=N+1}^{D} U_{ik}\Sigma_{kj} = 0$ as $k > N \ge i$.
∴ $(U\Sigma)_{ij} = \sum_{k=1}^{N} U_{ik}\Sigma_{kj}$ hence only the first N rows and N colums of Σ and U respectfully are present in computing $(U\Sigma)_{ij}$ which is equal to $(U'\Sigma)_{ij}$. ∴ $(U\Sigma)_{ij} = (U'\Sigma')_{ij}$ ∴ $U'\Sigma'$ & $U\Sigma$ share all the same elements with each Σ_{ii} element $\forall\, i > N$ removing the N+1 to D columns and rows of U & D respectfully from the product calculation ∴ $U\Sigma = U'\Sigma'$ as they share the same elements for the resulting diversions.

(1F) U' is an $D\times N$ matrix ∴ $(U')^T$ is an $N\times D$ matrix ∴ using matrix multiplication $(U')(U')^T$ is a $D\times D$ matrix and $(U')^T(U')$ is an $N\times N$ matrix hence the two products do not have the same dimensions meaning they're not equal hence $(U')(U')^T \neq (U')^T(U')$ ∴ U' is not orthogonal.

(1g) From the problem statement we know that the columns of U' are orthogonal ∴ linearly independent and we know that the dot product of ortho ~~onal~~ vectors dotted by itself is equal to 1 as they are parallal i.i the same direction and the dot-product of two linearly independent vectors is always zero. With that in mind, we know that the elements of $(U')^T(U')$ are the results of the dot-product of the rows of $(U')^T$ and columns of U' and if a column in U' is multiplied by it'self in $(U')^T$ it will result in a one, else, if it is multiplied by any other in $(U')^T$ it will result in zero as all columns in U' are linearly independent ∴ $(U')^T(U')$ will have 1's along it's diagonal and zeroes everywhere else resulting in the idenity matrix $I_{N\times N}$. Now, for $(U')(U')^T$ is the dot product of the rows of U' by themselves ∴ as we assumed in E that $D > N$ ∴ we have more rows than colums hence we have linear overlap meaning that not all rows are necessarily linearly independent hence that $(U')(U')^T$ will have elements outside the diagonal that are not zeroes ∴ $(U')(U')^T$ doesn't result in the idenity matrix $I_{D\times D}$.

(1h) From lecture we know that the pseudoinverse for SVD $A = U\Sigma V^T$ is $A^+ = V\Sigma^+ U^T$ where Σ^+ is a diagonal matrix including only positive elements $K > 0$ that are inverses of themselves

$\therefore$ we can show this as $\Sigma^+ = \Sigma^{-1}$

$\therefore$ $A^+ = V\Sigma^+ U^T = V\Sigma^{-1} V^T$

(1I)

We know that both U & V are orthogonal by definition $\therefore$ $U^{-1} = U^T$ s.t. $UU^T = U^TU = I$ and the same for V as it is orthogonal. Also, we know that for a diagonal matrix it's transpose is equal to itself which is obvious

$\therefore$ $X^{+'} = (X^TX)^{-1}X^T = X^{-1}(X^T)^{-1}X^T$ and plugin in for $X = U\Sigma V^T$ we get

$= (U\Sigma V^T)^{-1}((U\Sigma V^T)^T)^{-1}(U\Sigma V^T)^T = U^{-1}\Sigma^{-1}(V^T)^{-1}(U^T)^{-1}(\Sigma^T)^{-1}(V)^{-1}U^T\Sigma^T V$

and as Σ is a diagonal matrix & U & V are orthogonal then by the above properties we mentioned we have:

$= U^{-1}\Sigma^{-1}(V^{-1})^{-1}(U^T)^{-1}\Sigma^{-1}V^TU^T\Sigma V$ and moving things around we get:

$$= \underbrace{U^{-1}(U^T)^{-1}}_{I}\underbrace{(V^{-1})^{-1}}_{V}\Sigma^{-1}\underbrace{\Sigma^{-1}\Sigma}_{I}\underbrace{V^TV}_{I}U^T$$

$= (UU^T)^{-1}V\Sigma^{-1}U^T$ → Identity matrix

$= V\Sigma^{-1}U^T = X^+$ hence proven.

(4) The conditional number (CN) of a given ~~matrix~~ is the result of it's max singular value (σ_{max}) divided by it's min singular value (σ_{min}) $\therefore$ as all are singular values are positive as our singular values are the positive squareroots $\therefore$ our $CN \geq 1$ as we have more than one singular value. Now the CN of $\Sigma = \frac{\sigma_{max}}{\sigma_{min}}$ $\therefore X^TX = (U\Sigma V^T)^T(U\Sigma V^T) = \underbrace{U^TU}_{\to I} \cancel{\Sigma^T} \Sigma V V^T$ $\to \Sigma$ as Σ is a diagonal matrix

$= \Sigma \Sigma V V^T = V \Sigma^2 V^T$ $\therefore$ the singular values of X^TX are in Σ^2 $\therefore$

CN of $X^TX = \frac{(\sigma_{max})^2}{(\sigma_{min})^2}$ $\therefore$ as $\frac{\sigma_{max}}{\sigma_{min}} > 1$ $\therefore \frac{(\sigma_{max})^2}{(\sigma_{min})^2} > \frac{\sigma_{max}}{\sigma_{min}}$ $\therefore$ CN of X^TX is is greater hence the expression in problem I is more prone to numerical error.