

# Homework 1

**Exercise 1** First we prove a lemma

This one is fun note that for a variable  $x$  when we want to maximize over  $y$  take:

$$\frac{|x^*Ay|}{\|x\|_2\|y\|_2} \tag{1}$$

$$= \frac{|\langle y, A^*x \rangle|}{\|x\|_2\|y\|_2} \tag{2}$$

$$\leq \frac{\|y\|_2\|A^*x\|}{\|x\|_2\|y\|_2} \tag{3}$$

$$= \frac{\|A^*x\|}{\|x\|_2} \tag{4}$$

$$\tag{5}$$

So we know that this quantity is always upper bounded by  $\frac{\|A^*x\|}{\|x\|_2}$ . However we can actually choose a  $y$  that achieve this bound. Note that for cauchy shwartz equality is achieve if one vector is a constant multiple of the other. There are two cases to consider now.

case 1: if  $A^*x = 0$  then this bound is trivially achieved because the final bound  $\frac{\|A^*x\|}{\|x\|_2} = 0$  and the initial bound is  $\frac{|x^*Ay|}{\|x\|_2\|y\|_2} = 0$  regardless of  $y$ .

case 2: if  $A^*x \neq 0$  then choose  $y = A^*x$  and then the inequality above turns into an equality (by equality for cauchy shwartz) and we actually achieve this bound. so:

$$\sup_y \frac{|x^*Ay|}{\|x\|_2\|y\|_2} \tag{6}$$

$$= \frac{\|A^*x\|}{\|x\|_2} \tag{7}$$

$$\tag{8}$$

Now assume that  $A$  is invertible

Next we want to minimize this bound with respect to  $x$ :

$$= \frac{\|A^*x\|}{\|x\|_2} \tag{9}$$

$$= \|A^* \frac{x}{\|x\|}\| \tag{10}$$

before we move one we would like to note that in general we have that  $\|T^{-1}y\| \leq \|T^{-1}\|\|y\|$

for invertible  $T$  we have that  $y = Tx, x = T^{-1}y$  so:

$$\|x\| \leq \|T^{-1}\| \|Tx\| \quad (11)$$

$$\frac{1}{\|T^{-1}\|} \leq \|T\| \frac{\|x\|}{\|Tx\|} \quad (12)$$

$$(13)$$

Then as a result we have since  $A^*$  is invertible:

$$= \|A^* \frac{x}{\|x\|}\| \quad (14)$$

$$\geq \frac{1}{\|(A^{-1})^*\|} \quad (15)$$

so any choice of  $x$  will be lower bounded by this constant. We just need to find some choice of  $x$  that will achieve this bound. first choosing  $x = A^{-*}z$  we get:

$$= \|A^* \frac{x}{\|x\|}\| \quad (16)$$

$$= \frac{\|z\|}{\|A^{-*}z\|} \quad (17)$$

note that since  $A$  is finite and invertible  $A^{-*}$  has a finite norm and namely there is an  $z$  that achieve that bound such that  $\frac{\|A^{-*}z\|}{\|z\|} = \|A^{-*}\|$ . take that  $z$  then we have that:

$$= \frac{1}{\|A^{-*}\|} \quad (18)$$

so we found a  $x = A^{-*}z$  that achieves that bound. thus the  $\inf_x \frac{\|A^*x\|}{\|x\|_2} = \frac{1}{\|A^{-*}\|}$ . Note that we are doing this in the two norm. So the largest singular value of  $A^{-*}$  is the smallest singular value of  $A^*$  which is the smallest singular value of  $A$  (excluding zero singular values) so this just turns out to be  $\sigma_{\min}$  of  $A$ .

Now since  $A$  is invertible this this has a nonzero norm thus  $\beta > 0$

Assume that  $\beta > 0$  we can do the supremum steps just like we did before because things cancel out from there we get that:

$$0 < \beta = \inf_x \frac{\|A^*x\|}{\|x\|} \quad (19)$$

assume by way of contradiction that  $A$  was not invertible. If this is the case then  $A^*$  is also not invertible and we can choose  $x \neq 0, A^*x = 0$  thus the infum of this function would be zero:

$$\beta = \inf_x \frac{\|A^*x\|}{\|x\|} = 0 \quad (20)$$

but this contradicts  $\beta > 0$  so  $A$  must be invertible.

**Exercise 2** To prove this first note that we can obtain all of the nonzero singular values of  $A$  from eigenvalues of either  $AA^T$  or  $A^T A$ . To prove this note (taking the reduced form):

$$A^T A = V \Sigma U^T U \Sigma V \quad (21)$$

$$(22)$$

from here note that since  $U$  is  $m \times r$  then  $U^T U = I_r$  so:

$$= V \Sigma^2 V^T \quad (23)$$

We can extend this to an orthonormal completion ( $V_c$ ) for the zero singular value of  $A$ . but just notice that each of these  $\sigma^2$  is an eigenvalue because:

$$V_c \Sigma^2 V_c^T = A^T A \quad (24)$$

$$V_c \Sigma^2 = A^T A V_c \quad (25)$$

similarly for the other direction:

$$A A^T = U \Sigma V^T V \Sigma U^T \quad (26)$$

$$= U \Sigma^2 U^T \quad (27)$$

So now these same  $\sigma^2$  are also the nonzero einvalues of  $AA^T$ . (the other zero ones can be obtained by orthonormal completion and appending zero eigenvalues as stated before)

So the non zero eigenvalue of  $AA^T$ ,  $A^T A$  coincide. thus We consider  $QAA^T Q^T$ .

This is a  $m - 1$  compression of  $AA^T$ . as a result the eigenvalues of this matrix will satisfy

$$\sigma_i^2 \leq \mu_i \leq \sigma_{i+1}^2 \quad (28)$$

defining  $\hat{\sigma}_i = \sqrt{\mu_i}$  we get:

$$\sigma_i^2 \leq \hat{\sigma}_i^2 \leq \sigma_{i+1}^2 \quad (29)$$

So taking square roots of all of the sides (since everything is nonnegative as a result of semipositive definiteness or the fact that the singular values are nonnegative):

$$\sigma_i \leq \hat{\sigma}_i \leq \sigma_{i+1} \quad (30)$$

But these values are literally the square roots of eigenvalues of  $QA(QA)^T$ . which means that they are the singular values of  $QA$

**Exercise 3** first remember that all finite dimensional matrix norms are equivalent so that means for any matrix norm  $\|\cdot\|$  there is a constant  $M$  such that  $\|\cdot\| \leq M \|\cdot\|_2$ . So we will first prove density for the two norm

let  $\epsilon > 0$  remember that the two norm of a matrix is simply its largest singular value. Take the svd of a matrix of rank  $r$ .  $A = \sum_{k=1}^r \sigma_k u_k v_k^*$ . Take the fullrank completion of  $u$  and

v by just setting the remaining singular values to zero if any are zero  $A = \sum_{k=1}^{\min(n,m)} \sigma_k u_k v_k^*$ . Set  $B = \sum_{k=1}^{\min(n,m)-1} \sigma_k u_k v_k^* + (\sigma_{\min(n,m)} + \epsilon) u_{\min(n,m)} v_{\min(n,m)}^*$ . (Note if A is not full rank this is just multiplied by epsilon, but since we would have set it to zero it sthe same).

Take:

$$\|A - B\|_2 = \quad (31)$$

$$\left\| \sum_{j=1}^{\min(n,m)-1} (\sigma_j - \sigma_j) u_j v_j^* + (\sigma_{\min(n,m)} - \sigma_{\min(n,m)} + \epsilon) u_{\min(n,m)} v_{\min(n,m)} \right\|_2 \quad (32)$$

$$= \|\epsilon u_{\min(n,m)} v_{\min(n,m)}\|_2 = \epsilon \quad (33)$$

So we can arbitrarily approximate any matrix in the 2 norm.

Now take an arbitrary matrix norm note that:

$$\|A - B\| \leq M \|A - B\|_2 \quad (34)$$

choose B so that  $\|A - B\|_2 \leq \epsilon/M$  then:

$$\|A - B\| \leq \epsilon \quad (35)$$

by that B.

#### Exercise 4

a) Recall that unitary matrices satisfy  $UU^H = U^H U = I$  so they are normal and as a result are unitarily diagonalizable. Furthermore note that all of their eigenvalues have magnitude one:

$$\langle v, v \rangle = \langle Uv, Uv \rangle = \langle \lambda v, \lambda v \rangle \quad (36)$$

$$= |\lambda|^2 \langle v, v \rangle \quad (37)$$

for some eigenvector  $v$ . note that since this was nonzero that means that  $|\lambda|^2 = 1$  so they all have absolute magnitude one. as a result let each eigenvalue be  $\lambda_j$  note that we can write this in polar form as  $\lambda_j = e^{i\theta_j}$  for some theta since they are magnitude one (these thetas are real angles). then:

$$U = Q \Lambda Q^* = Q e^{i\theta} Q^* \quad (38)$$

Where  $\theta$  is the diagonal entries of our thetas we defined eralier

set  $\Theta = Q \theta Q^*$  (this matrix is clearly hermitian since all of hte thetas are real) then clearly  $e^{i\Theta} = Q e^{i\theta} Q^* = U$ . And we are done

b) essentially what we need to show is that any matrix can be decomposed as a positive definite matrix and a unitary matrix

Note that earlier we have proven that if A matrix is hermitian positive definite it has a square root. take  $\sqrt{A^* A} = S$  namely  $V \Sigma V^T$  where  $\Sigma$  is the square roots of hte eigenvalues

so that  $V\Sigma V^T V\Sigma^T V^T = V\Sigma^2 V^T$  is the eigenvalue decomposition (we can do this since  $A^*A$  is positive semidefinite and hermitian, so this square root is also positive semidefinite). Note this is also hermitian. but note that  $V$  is literally the svd right singular vectors of  $A$  so take:

$$UV^T S = UV^T V\Sigma V^T = U\Sigma V^T = A \quad (39)$$

now note that if we took the full svd of  $A$  then  $U, V$  are both unitary so  $UV^T$  is also unitary. By part one we can write this as  $e^{i\Theta} = UV^T$  so then:

$$A = (UV^T)S = e^{i\Theta}S \quad (40)$$

So we have proved one part. with Hermitian  $S$

Now similarly we can take the square root of  $AA^*$  which is  $R = U\Sigma U^T$  where  $\Sigma$  is the square roots of the eigenvalues of  $AA^*$  like before. (we can do this since  $AA^*$  is positive semidefinite, so this square root is also positive semidefinite ) Then:

$$RUV^T = U\Sigma U^T UV^T = U\Sigma V^T = A \quad (41)$$

So noticing that since like before  $UV^T = e^{i\Theta}$  is orthonormal for some hermitian  $\Theta$  then:

$$A = R(UV^T) = Re^{-i\Theta} \quad (42)$$

And from here we have proven all but the statement that  $R$  and  $S$  are orthonormally similar. This follows from the fact that  $S = V\Sigma V^T$  we can multiply on both sides with  $UV^T$  to transform this in to  $R$ . So the hermitian matrices are similar to one another.

### Exercise 5

a) Assume that  $A$  is invertible. that means that all of its singular values are nonzero and each  $U = \tilde{U}, V^T = \tilde{V}^T, \Sigma = \tilde{\Sigma}$  are full rank thus:

$$AU^+ = (\tilde{U}\tilde{\Sigma}\tilde{V}^T)\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^T \quad (43)$$

$$= U\Sigma V^T V\Sigma^{-1}U = U\Sigma\Sigma^{-1}U^T = UU^T = I \quad (44)$$

So this is the inverse.

b) from here on out just assume that the tildes are there, I don't want to type them all out. I will say if we are using the untruncated version

$$(AA^+)AA^T \quad (45)$$

$$= U\Sigma V^T V\Sigma^{-1}U^T U\Sigma V^T V\Sigma^{-1}U^T \quad (46)$$

note that  $V^T$  is an  $r \times n$  matrix with orthonormal rows so  $V^T V$  is  $r \times r$  identity:

from here note that  $U$  is an  $m \times r$  matrix with orthonormal columns so  $U^T U$  is the identity  $r \times r$ :

$$= U\Sigma\Sigma^{-1}\Sigma\Sigma^{-1}U^T \quad (47)$$

$$= U\Sigma\Sigma^{-1}U^T \quad (48)$$

$$= U\Sigma V^T V\Sigma^{-1}U^T \quad (49)$$

$$= AA^+ \quad (50)$$

$$(51)$$

So it is a projection. Going even further with simplifying we get

$$= U\Sigma V^T V\Sigma^{-1}U^T = U\Sigma\Sigma^{-1}U^T = UU^T \quad (52)$$

and this is just a projection onto the columns space of  $U$ . So this is a projection onto the range of  $A$ .

Recall that if  $P$  is a projection that  $I - P$  is also a projection so  $I - AA^+$  is a projection onto the orthogonal complement of the range of  $A$  which is the null space of  $A^T$

Now in like manner take  $A^+A$  then:

$$A^+AA^+A \quad (53)$$

$$= V\Sigma^{-1}U^T U\Sigma V^T V\Sigma^{-1}U^T U\Sigma V^T \quad (54)$$

$$= V\Sigma^{-1}U^T U\Sigma\Sigma^{-1}U^T U\Sigma V^T \quad (55)$$

$$= V\Sigma^{-1}U^T UU^T U\Sigma V^T \quad (56)$$

$$= V\Sigma^{-1}U^T U\Sigma V^T \quad (57)$$

$$= A^+A \quad (58)$$

So its a projection. Specifically if we keep reducing we get  $V\Sigma^{-1}U^T U\Sigma V^T = V\Sigma^{-1}\Sigma V^T = VV^T$  which is a projection onto the column space of  $V$  or the range of  $A^T$ .

Similarly  $I - A^+A$  must also be a projection by the rationale provided before. it is a projection onto the orthogonal complement of this which is the null space of  $A$ .

c) take:

$$AA^+A = U\Sigma V^T V\Sigma^{-1}U^T U\Sigma V^T \quad (59)$$

$$= U\Sigma\Sigma^{-1}U^T U\Sigma V^T \quad (60)$$

$$= UU^T U\Sigma V^T \quad (61)$$

$$= U\Sigma V^T \quad (62)$$

$$(63)$$

Oh so I guess it is true for all matrices! I guess that makes sense because  $AA^+$  maps things to the column space of  $A$ . but after going through  $A$  we are already in the column space so this just functions as an identity

d) take  $A^+A$  remember that we reduced this earlier to  $VV^T$ . so we have that if  $VV^T = I_n$  then this is true. however  $V$  is  $n \times r$ . So this can only be true if  $V$  is full rank or we have at least  $n$  nonzero singular value of  $A$ . In which case  $V$  would be orthonormal with  $VV^T = I_n$

Similarly for  $AA^+$  recall that we reduced this to  $UU^T$ . We want this to be the  $I_m$  identity. recall that  $u$  is  $m \times r$ . So this can only be true if  $U$  is full rank or we have at least  $m$  nonzero singular value of  $A$ . In which case  $U$  would be totally orthonormal with  $UU^T = I_m$

e) It is not a well conditioned operation take for example  $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$  and  $B = \begin{bmatrix} 0 & 0 \\ 0 & \epsilon \end{bmatrix}$ . Clearly  $\|A\|_2 = 1, \|B\| = \epsilon$ . So we have a small perturbation. additionally note that

$A^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$  so that:

$$\frac{\|(A - B)^+ - A^+\|_2}{\|A\|_2} \quad (64)$$

$$= \left\| \left( \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} \right)^+ - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\| \quad (65)$$

$$= \left\| \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\epsilon} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\| \quad (66)$$

$$= \left\| \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\epsilon} \end{bmatrix} \right\| \quad (67)$$

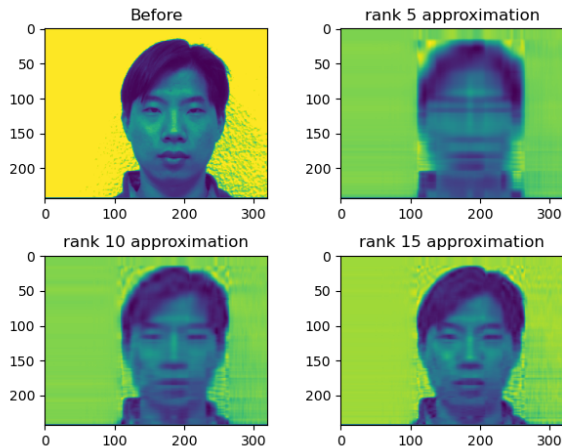
$$= \frac{1}{\epsilon} \quad (68)$$

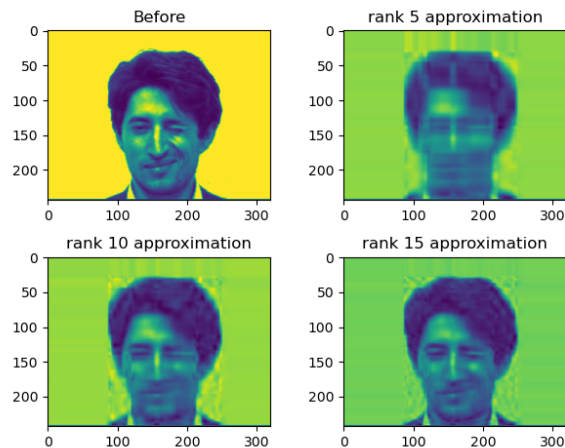
So what we find is that this operation here has condition number that grows uncontrollably as we send  $\epsilon \rightarrow 0$ . So we cannot control this growth. The idea (for me) is that when we are computing the psuedo inverse if we have something really close to zero then its sent to a very large number, but if it is zero its sent to zero. So we have this discrepancy that accounts for the gigantic condition number.

## Exercise 6

For these problems I have been using the entire dataset. I do not show all of the images.

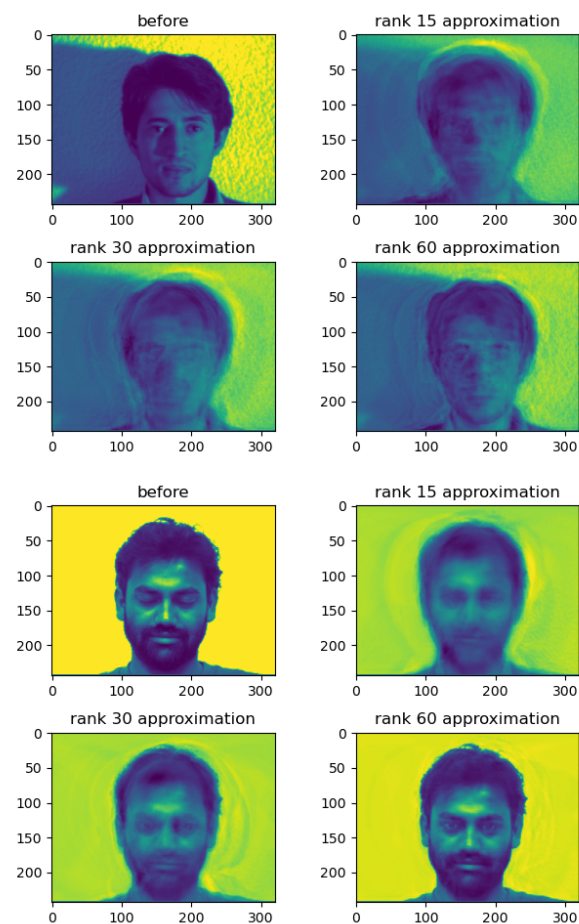
A) For this first part here are two sets of images I chose:





The accuracy mostly depended on the  $k$  I chose to approximate it, there not really variation in the faces I chose, but mostly the  $k$ . You can see that the accuracy gets a lot better and less blurry as we increase  $k$ . To do this I computed the svd using numpy of a single image (after I loaded it in with matplotlib in python). I then removed all but the first  $k$  singular values and the remultiplied  $U, \Sigma, V^T$  to get the new approximation.

B) With this problem I took two different sets of images:



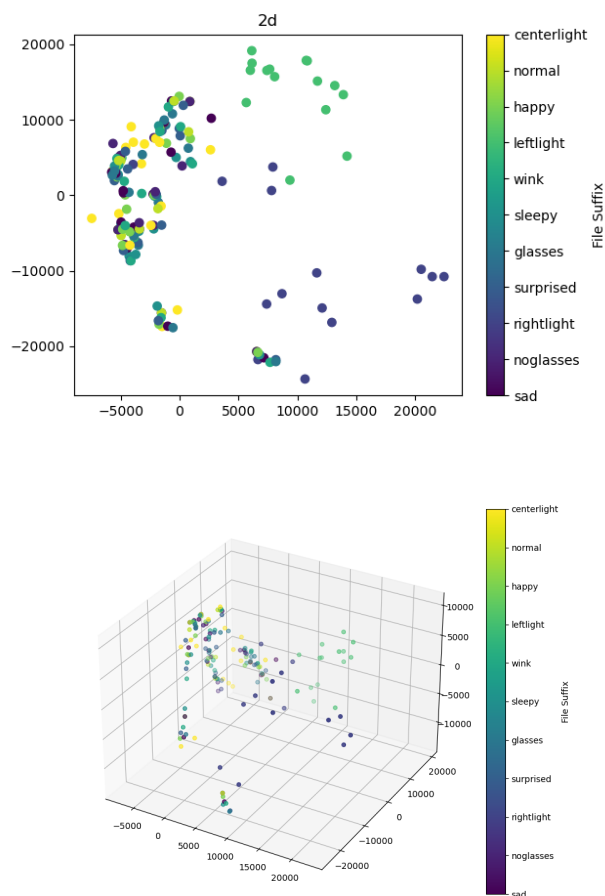
This one was more interesting, it took me more singular values to get a higher definition



(consistent with the idea its harder to compress multipl images at once). And these ones there was variation if I chose different images, for example the second image came into clarity much faster than the first (although he opens his eyes which is kind of funny, most likely a result of the compression because most people are opening their eyes).

With this problem I loaded in all of the images and stacked their flattened versions into a matrix. I then took the svd like in the previous problem and kept only the first k nonsingular values before multiplying hte matrix back together. Note that in this case it was really important that I computed the reduced svd in numpy so I didn't crash my computer.

C) For this one we first want to center the data so we center each image around the mean image. We learned that the compression matrix is literally the first k collumns of the matrix  $U$  where  $U$  is the  $U$  of the original data.. so we take the SVD and multiply our new matrix by  $U_k^*$  so we multiply  $U_k^* A_c$ . I then plot the components:



These were somewhat separated by the type of lighting they were introduced to like center light or night light. That seemed to be the main differentiator as a lot of different face expressions seemed to congregate.