INF554

Ali MAMMADOV

October 2020

1 Question 1

From linear algebra we know that if the Principal Components are uncorrelated the theirs **Covariance matrix** must be a diagonal matrix.

$$Cov(CU) = \frac{(CU)^T CU}{m} = \frac{U^T C^T CU}{m} = \frac{U^T * m * SU}{m} = U^T SU$$

– U : eigenvectors of $S = Cov(C) = \frac{C^T C}{m}; C^T C = m * S$
– Since U is eigenvectors of S. $SU_j = \lambda_j U_j$, for j in {1...k}

$$U_j^T SU_j = U_j^T \lambda_j U_j = \lambda_j U_j^T U_j = \lambda_j$$

–$U_j^T Uj = 1$, cause eigenvectors are orthogonals

Thus, Cov(CU) = diag($\lambda_1, ..., \lambda_k$). This proves that Principal components are uncorrelated.

Also we know that diagonal values of a covariance matrix are equal to the variances. In our case, we can see that variances of Principal Components are equal to the eigenvalues ($\lambda_1, ..., \lambda_k$).

2 Question 2

According to the image obtained in task 2, the image quality improves with increasing **k**. From the implementation of algorithms, it is obvious that **NMF** function is compressing better than others. Cause it replaces the original data with a lower-dimensional representation matrices obtained via approximation (A[n*k], B[k*m]). The optimal one (good quality and good compression) is **SVD**. It stores one additional matrix (E[k*k]), which takes a bit more than **NMF** but the quality is the best.

3 Question 3

$$X^h = \frac{X^T X}{m}$$

$$X^{h'} = \frac{X'^T X'}{m} = \frac{(XQ)^T XQ}{m} = \frac{Q^T X^T XQ}{m} = Q^T X^h Q; X^{h'} = Q^T X^h Q$$

$$Cov(X^{h'}) = \frac{X^{h' T} X^{h'}}{m} = \frac{Q^T X^{hT} Q Q^T X^h Q}{m} = \frac{Q^T X^{hT} X^h Q}{m} = Q^T Cov(X^h) Q = Q^{-1} Cov(X^h) Q$$

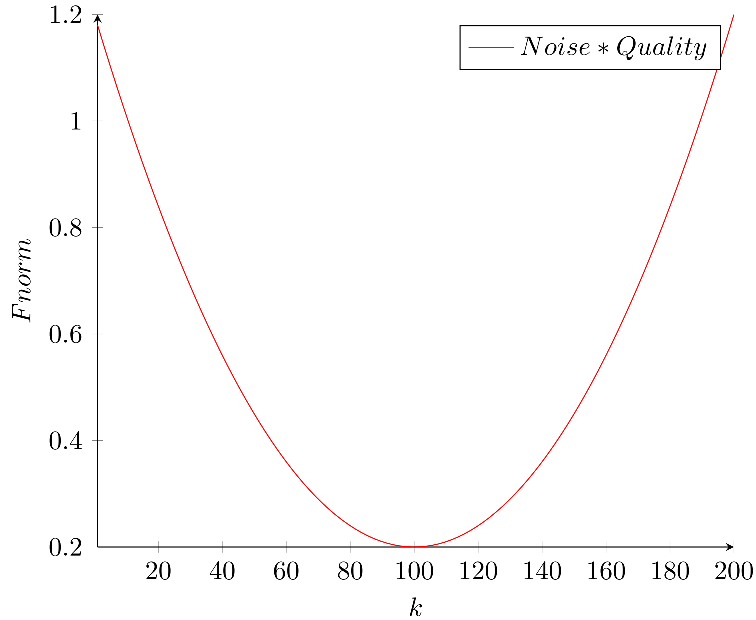
* $Q^T = Q^{-1}$, Q is orthogonal

The total amounts of the variance of X^h is the sum of diagonal elements of Cov(X^h) or the trace tr(Cov(X^h)). $Cov(X^{h'}) = Q^{-1} Cov(X^h) Q$, from the similarity transformation we know that it will always preserves the trace of matrix. Thus, tr(Cov(X^h)) = tr(Cov($X^{h'}$)). Which means that the total amounts of the variance of are equal cause they are located on the diagonal of covariance matrix.

We have the same thing with X and X': $Cov(X^{'}) = Q^{-1} Cov(X) Q$. This implies that principal components of them are the same $CU = C'U'$.

4 Question 4

Low-rank approximation decreases the noise of the image when k is not so high. As we said in the Q2 quality improves as k increases but with noise it is quite the opposite, it increases when it gets bigger and it ruins the photo. Thus, it is necessary to find the optimal value for k at which there will be little noise and good quality (see the graph). **denoising()** function finds that value of k.



5 Question 5

If the inferred structure of the noise captures the important parts of information related to the low-rank approximation, the performance of the low-rank representation possibly decreases. In our case, we see that image 2 is very pretty approximately without noise, means that noise does not captures important information. So, it is easy to denoise. But we can't say the same thing to the first image, in which noise probably captures the core of the image. It causes big changes in the data during the approximation and consequently Frobenius norm difference(between original and denoised images) increases very fast. So we should stop on small amount of the k. Which is not enough for obtaining a good approximation of the image.