Project 2 NLA: SVD applications

Theodoros Lambrou

1. **Least Squares problem**

I develop two distinct functions for solving the Least Squares problem: the first utilizing Singular Value Decomposition, and the second utilizing QR factorization.

This QR factorization method considers the matrix's rank (rank(A)), which determines the solution strategy. For full-rank matrices, I apply standard QR factorization and resolve the system using back substitution. When confronting a rank-deficient matrix, I switch to QR factorization with pivoting. This approach addresses the matrix's rank limitations and also improves the computational solution's numerical stability.

When analyzing the results, the optimal degree identified as the best performer is 9. This shows that a polynomial degree of 9 gives a balance between complexity and accuracy in fitting the data. It's important to note that the considered degrees are from 3 to 10 — the selection criterion prioritizes the degree with the smallest error. This criterion is met at about 10.8455 for both the SVD and QR methods.

Below I analyse the outcomes for these 2 datasets. For the 1st dataset ("dades.csv"), both methods produce an error of 10.845499004346927 and a solution with a norm of 137.20330857368776.

For the 2nd dataset ("dades_regressio.csv") it can be seen that the outcomes differ between the 2 methods. For SVR, the error is. 4.4542020954046295 and the solution norm is 3581482594641324.0. For QR, the error is 1.149597896007536 and the solution norm 4774736.28989233.

The solutions returned by the SVD method look similar for both datasets. For the 2nd dataset however the norm is very big maybe attributing to numerical instability.

The solutions returned by the QR method are again similar for both datasets. For the 2nd dataset however, the norm is a lot smaller than the one obtained using the SVD method — this might be due to improved numerical stability.

QR factorization seems to be preferable when the matrix dimensions are known thereby enabling us to decide what is the complexity of the solution.

2. Graphics compression

2.1

Singular Value Decomposition (SVD) factorization: for any real or complex matrix A of size $m x n$, there are matrices U (orthogonal matrix of size $m x n$), Σ (diagonal matrix of size $m x n$), and V^T(conjugate transpose of V , where V is an orthogonal matrix of size $m x n$) such that: $A = U\Sigma V^T$

The singular values in Σ are ordered in decreasing order:

$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, where $\mathrm{p} = \min(\mathrm{m}, \mathrm{n})$.

The best low-rank approximation of A of rank k is given by:

$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, where u_i and v_i are the i-th columns of U and V respectively.

Using the steps of the proof in reference [1], it can be shown that the matrix A_k is the best rank k approximation to A in both the Frobenius and the 2-norm.

I will firstly show that the matrix $A_k$is the best rank k approximation to A in the Frobenius norm. For this, the following 2 lemmas are needed:

> ***Lemma 4.6:*** *The rows of A_k are the projections of the rows of A onto the subspace V_k spanned by the first k singular vectors of A*
>
> ***Lemma 4.8:*** $\left\|\mathrm{A} - \mathrm{A_k}\right\|_2^2 = \sigma_{\mathrm{k+1}}^2$

Proof for the above lemmas are found in reference [1].

Theorem 4.7: For any matrix B of rank at most k, $\left\|\mathrm{A} - \mathrm{A_k}\right\|_{\mathrm{F}} \leq \left\|\mathrm{A} - \mathrm{B}\right\|_{\mathrm{F}}$

Proof: Let B minimize $\left\|\mathrm{A} - \mathrm{A_k}\right\|_{\mathrm{F}}^2$ among all rank k or less matrices.

Let V be the space spanned by the rows of B.

The dimension of V is at most k.

Since B minimizes $\|A - A_k\|_F^2$, it must be that each row of B is the projection of the corresponding row of A onto V , otherwise replacing the row of B with the projection of the corresponding row of A onto V does not change V and hence the rank of B but would reduce $\|A - A_k\|_F^2$.

Since each row of B is the projection of the corresponding row of A, it follows that $\|A - A_k\|_F^2$ is the sum of squared distances of rows of A to V.

Since A_k minimizes the sum of squared distance of rows of A to any k -dimensional subspace, it follows that $\|A - A_k\|_F \leq \|A - B\|_F$.

Theorem 4.9: Let A be an $n \times d$ matrix. For any matrix B of rank at most k,

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$. Thus assume that A is of rank greater than k. By Lemma 4.8, $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.

Now assuming there is some matrix B of rank at most k such that B is a better 2-norm approximation to A than Ak. That is, $\|A - B\|_2 < \sigma_{k+1}$.

The null space of B, Null (B), (the set of vectors v such that Bv = 0) has dimension at least d − k. Let v1, v2,..., vk+1 be the first k + 1 singular vectors of A.

By a dimension argument, it follows that there exists a $z \neq 0$ in

$\mathrm{Null}\,(B) \cap \mathrm{Span}\,\{v1, v2, \ldots, vk + 1\}$

Scale z so that |z| = 1. We now show that for this vector z, which lies in the space of the first k + 1 singular vectors of A, that$(A - B)\,z \geq \sigma_{k+1}$.

Hence the 2-norm of A − B is at least σ_{k+1} contradicting the assumption that $\|A - B\|_2 < \sigma_{k+1}$

First $\|A - B\|_2^2 \geq |(A - B)z|^2$

Since $Bz = 0$, $\|A - B\|_2^2 \geq |Az|^2$

Since z is in the Span {v1, v2,..., vk+1}

$$\left|\mathrm{Az}\right|^2 = \left|\sum_{i=1}^{n} \sigma_i u_i v_i^T z\right|^2 = \sum_{i=1}^{n} \sigma_i^2 (v_i^T z)^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \left(v_i^T z\right)^2 = \sigma_{k+1}^2$$

.

It follows that $\|\mathrm{A} - \mathrm{B}\|_2^2 \geq \sigma_{k+1}^2$,

contradicting the assumption that $\|\mathrm{A} - \mathrm{B}\|_2 < \sigma_{k+1}$, thereby proving the theorem.

2.2

For this exercise, I chose 2 different images in greyscale so three matrices were obtained from the three components (RGB) of the .jpg files.

Four different ranks were used (1, 5, 20, 80) to get different approximations. The function implemented includes the Frobenius norm percentage (%) captured by the approximation, the relative error of the norm and the image.

Below are the results for each of the images.

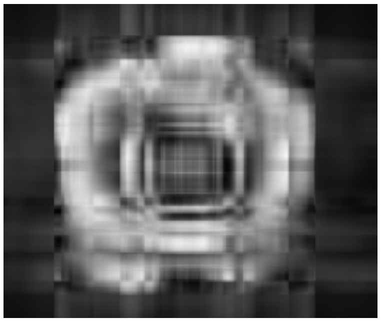
<ins>Image 1</ins>

Original

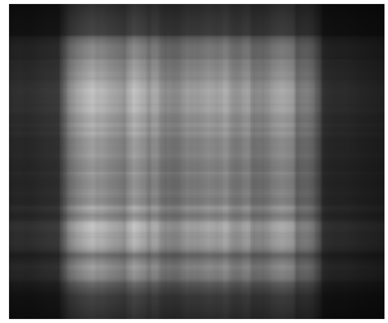
Rank 80

Rank 20

Rank 5



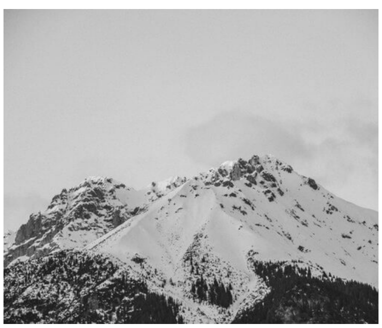
Rank 1



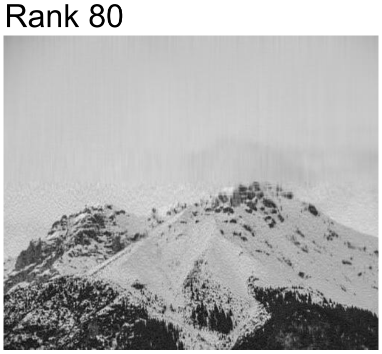
It is obvious that with higher rank the image has more clarity, and in lower ranks it is the opposite - which reflects the fact that the percentage of Frobenius norm is also closer to 100% when the image is more clear.

Image 2

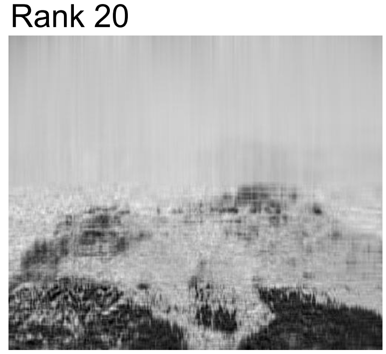
Original



Rank 80

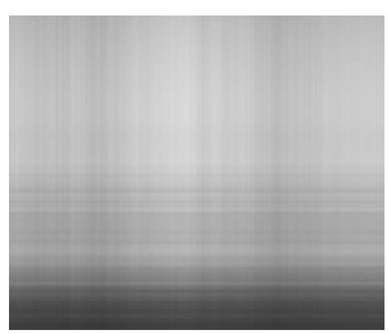


Rank 20



Rank 5

Rank 1



Similarly to the other image, higher rank improves the clarity of the image.

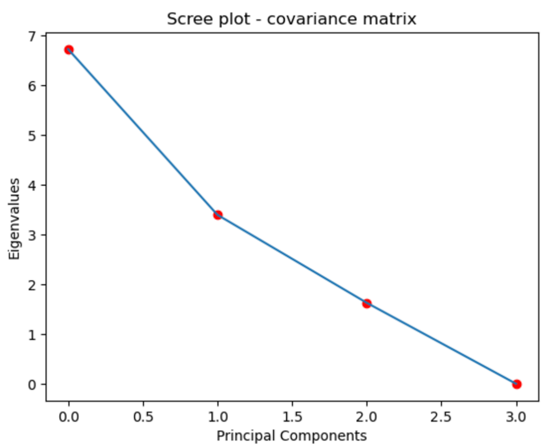
So it's important to be aware of what images are used when we proceed with image compression.

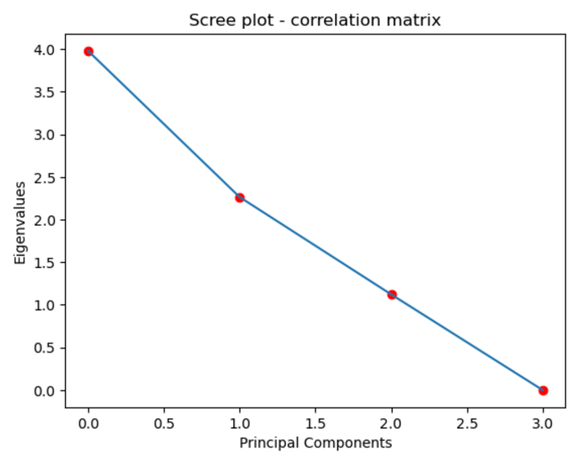
3. Principal component analysis (PCA)

The analysis below is based on the Python code provided.

These methods were used to get the principal components: Scree plots, the Kaiser rule and the 3/4 rule.

First dataset: *example.dat*



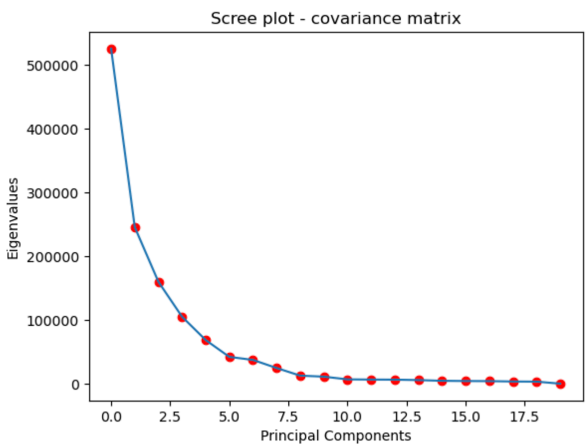


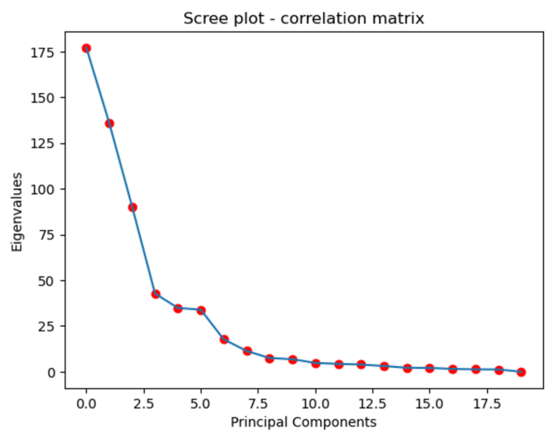
Despite the limited number of eigenvectors and eigenvalues in the dataset, the analysis of the covariance and correlation matrices suggests we can effectively eliminate two components from each. As a result, our approach would concentrate on preserving just two principal components.

According to the Kaiser rule, our matrices would be characterized by three principal components. It is interesting that this approach tends to suggest retaining more components than what would be indicated by the "elbow" method when examining the Scree plot.

The covariance matrix contains a principal component that independently accounts for more than 75% of the total variance. Applying the 3/4 rule to this matrix takes us to recommend preserving just 1 Principal Component. In contrast, the correlation matrix warrants the retention of 2 Principal Components under the same rule, a conclusion that has been confirmed from computational verification.

Repeating the same steps for the *RCsGoff.csv* dataset:





While the Scree Plot suggests preserving 5 Principal Components for the Covariance Matrix, the Correlation Matrix displays a more complex profile with two subtle ‘elbow’ points. Hence, when focusing on the second inflection point, the recommendation is to maintain either 6 or 7 Principal Components for the correlation matrix.

The application of the Kaiser Rule to the covariance matrix and correlation matrix identifies 19 Principal Components associated with eigenvalues exceeding 1.

Lastly, the 3/4 rule guides us to select 2 Principal Components for both matrices. This choice is substantiated by the variance distribution: in the covariance matrix, the top components capture 72% and 15% of variance, while in the correlation matrix, they account for 50% and 30% of the total variance.

References:

[1] https://www.cs.cmu.edu/~venkatg/teaching/CStheory-infoage/book-chapter-4.pdf