Numerical Linear Algebra Project 3

Peter Weber

January 22, 2018

1 Graphics Compression

1.1 Task 1: Proof of low rank approximations

1.1.1 Spectral Norm

Let $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$ and $A = U\Sigma V^T$ be the singular value decomposition of A where Σ is a diagonal matrix with $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n \geq 0$, and U, V have orthogonal columns. Then, the best rank k $(k < n)$ approximation of A with respect to the spectral norm is given by

$$A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T.$$

Proof. First, note that

$$||A - A_k||_2 = ||U \cdot diag(0, ..., \sigma_{k+1}, ..., \sigma_n) V^T||_2 = ||diag(0, ..., \sigma_{k+1}, ..., \sigma_n)||_2 = \sigma_{k+1},$$

since U and V are orthogonal.
We still have to show that A_k is the best rank k approximation. Let $B = XY^T \in \mathbb{R}^{m \times n}$ be a matrix with rank k and X, Y have k columns. Then there is a linear combination of the $k+1$ linearly independent vectors $(v_1, ..., v_{k+1})$ of matrix V, that maps B to its null space $Bw = 0$, where $w = \gamma_1 v_1 + ... + \gamma_{k+1} v_{k+1}$ is chosen to have norm 1. With $\gamma = (\gamma_1, ..., \gamma_k)^T$ it follows

$$\begin{aligned} ||(A-B)w||_2^2 = ||Aw||_2^2 = ||U\Sigma V^T w||_2^2 = ||\Sigma V^T w||_2^2 = ||\Sigma\gamma||_2^2 = \\ \gamma_1^2 \sigma_1^2 + ... + \gamma_{k+1}^2 \sigma_{k+1}^2 \geq \sigma_{k+1}^2 (\gamma_1^2 + ... + \gamma_{k+1}^2) = \sigma_{k+1}^2, \end{aligned} \tag{1.1}$$

since the singular values σ_i are ordered. Hence, A_k is the best rank k approximation.

1.1.2 Frobenius Norm

Let $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ and $A = U\Sigma V^T$ be the singular value decomposition of A where Σ is a diagonal matrix with $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n \geq 0$, and U, V have orthogonal columns. Then, the best rank k $(k < n)$ approximation of A with respect to the Frobenius norm is given by

$$A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T.$$

Proof. First, note that

$$||A - A_k||_F^2 = ||U \cdot diag(0,...,\sigma_{k+1},...,\sigma_n)V^T||_F^2 = ||diag(0,...,\sigma_{k+1},...,\sigma_n)||_F^2 = \sum_{i=k+1}^{n} \sigma_i^2,$$

since U and V are orthogonal.
We still have to show that A_k is the best rank k approximation. Let $B = XY^T \in \mathbb{R}^{m\times n}$ be a matrix with rank k and X, Y have k columns. We denote σ_i as the i-th singular value of its respective matrix. Then we have by the triangle inequality, if $A = A' + A''$, $\sigma_1(A) \leq \sigma_1(A') + \sigma_1(A'')$. If A'_k and A''_k are the closest rank k approximations of A' and A'', respectively, then for $i, j \geq 1$

$$\begin{aligned} \sigma_i(A') + \sigma_j(A'') &= \sigma_1(A' - A'_{i-1}) + \sigma_1(A'' - A''_{j-1}) \\ &\geq \sigma_1(A - A'_{i-1} - A''_{j-1}) \\ &\geq \sigma_{i+j-1}(A). \end{aligned} \tag{1.2}$$

Setting $A' = A - B$, $A'' = B$, $i = j + 1$ we obtain

$$\sigma_i(A - B) + \sigma_{k+1}(B) = \sigma_i(A - B) \geq \sigma_{i+k}(A)$$

where $\sigma_{k+1}(B) = 0$.
Hence, we have

$$||A - B||_F^2 = \sum_{i=1}^{n} \sigma_i(A - B)^2 \geq \sum_{i=k+1}^{n} \sigma_i(A)^2 = ||A - A_k||_F^2. \quad \square$$

Both proofs are taken from the Low-rank approximation article in Wikipedia.

1.2 Task 2: Lossy jpeg image

The next figures show two examples of image compression from the jpeg format. In the first figure one sees four subfigures for every image, where the relative number of components is 0.001, 0.01, 0.1, and 1. In particular, the details of the compression are given in the printout in Fig. 1.2. It is interesting to note, that in order to display something useful, the image that has letters needs a significantly higher rank approximation than the image without letters.
In Fig. 1.3 I plot the relativ Frobenius Norm as a function of the rank of the approximation. Surprisingly, only a rank one approximation already captures around 90% of the Frobenius Norm.

2 Principal Components Analysis

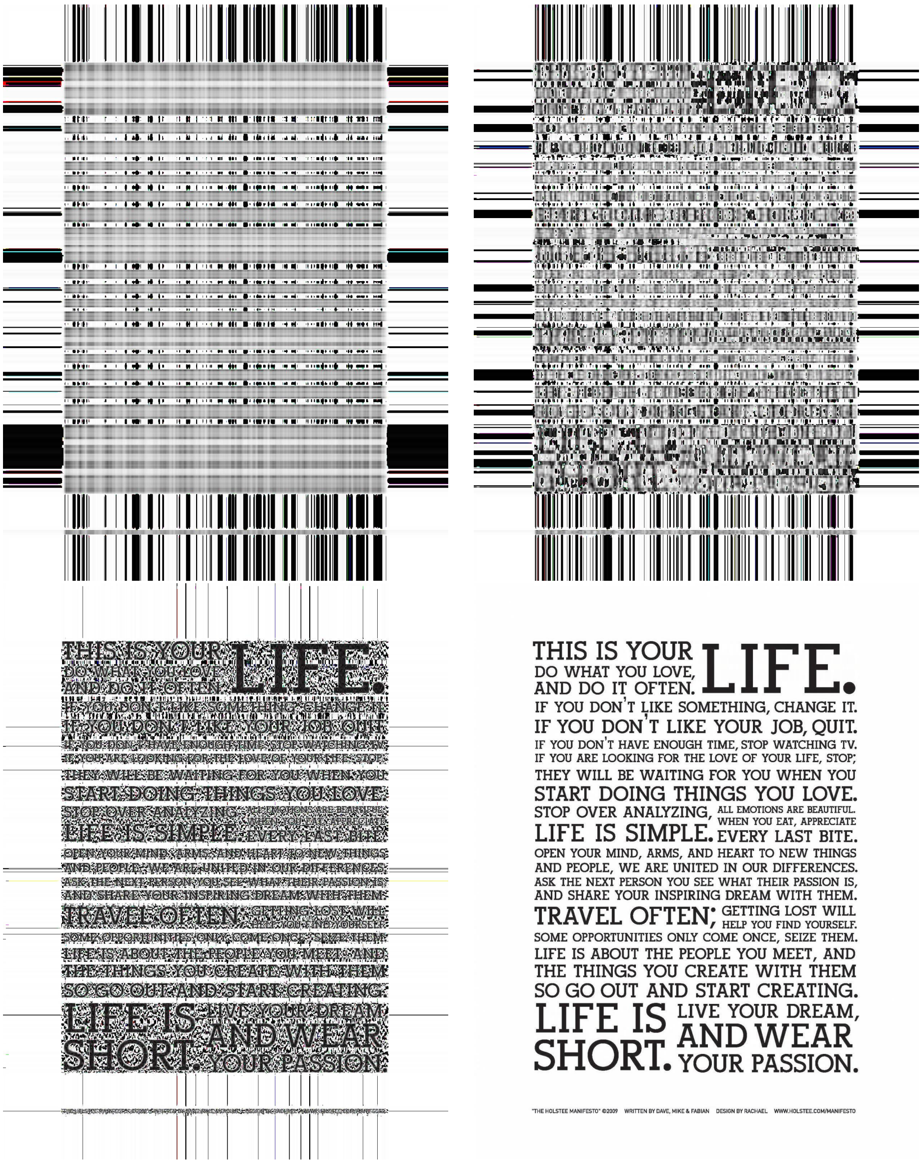
2.1 Task 1: Toy Example

For this subproblem, all the relevant information is printed out by my code. Here I will only include the plots of the results of the PCA, and a short discussion.
In figur 2.1, I plot the standard deviation and the cumulative variance of the principal components of the covariance matrix for the toy problem. The same plots, but for the correlation matrix are displayed in Fig. 2.2. When, for instance, choosing the Kaiser-rule as a rule for the stopping criterion, which states that one should include only the eigenvalues larger 1 one would take the first three principal components when working with the covariance matrix. One would choose the first two principal components when working with the correlation matrix. In any case, for such a small dataset, with only 4 features, a PCA is not really necessary.

2.2 Task 2: Gene Expressions

For this subproblem, all the relevant information is printed out by my code. Here I will only include the plots of the results of the PCA, and a short discussion.
First, I reproduce the plot from the assignment sheet to verify that my results are correct. This can be seen in Fig. 2.3.
In figure 2.4, I plot the standard deviation and the cumulative variance of each of the principal components. Here it is interesting that the first two principal comonents already capture almost 90% of the variance and the first three principal components almost 95%. So choosing just a few principal components is sufficient to describe the data appropriately. Additionally, the dimensionality of this problem is huge since the number of features correpsponds to 58581. Doing a principal component analysys without using the SVD would therefore result in a computationally and memory intense problem, because one has to deal with a 58581x58581 covariance matrix. Hence, using the SVD reduces immensely the computational burden, as, even when using all principal components, the new matrix dimension is 20x20.

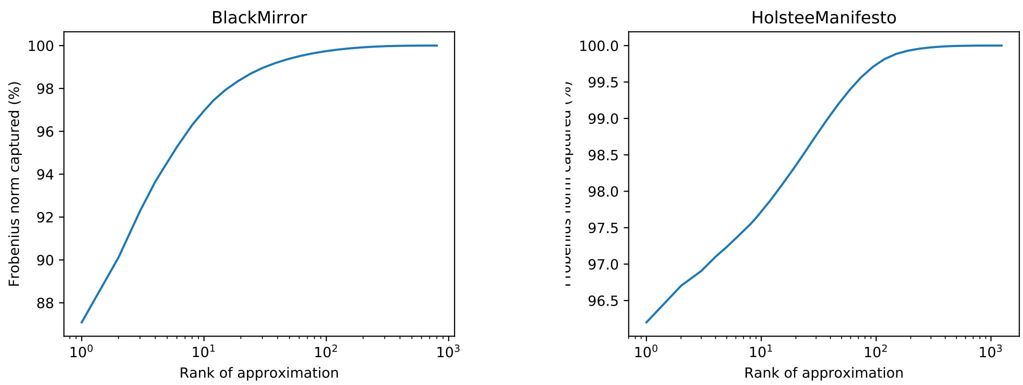
(a) Black Mirror Season 4 image compression. See next figure for details of compression.



(b) Holstee Manifesto image compression. See next figure for details of compression.

```
------------------------------------------------------------

--------------------  BlackMirror --------------------

Number of components: [1, 8, 80, 800]

Relative number of components: [0.001, 0.01, 0.1, 1]

Averaged Frobenius Norms: [  95905.  106048.  109747.  110129.]

Relative Frobenius Norms in (%): ['87.085', '96.295', '99.653', '100.000']
------------------------------------------------------------

--------------------  HolsteeManifesto --------------------

Number of components: [2, 13, 124, 1236]

Relative number of components: [0.001, 0.01, 0.1, 1]

Averaged Frobenius Norms: [ 318113.  322211.  328410.  328963.]

Relative Frobenius Norms in (%): ['96.702', '97.947', '99.832', '100.000']
Images saved
```

Figure 1.2: Image compression information for the images on the previous page. The Black-Mirror part corresponds to the first four images and the HolsteeManifesto to the last four, respectively. Every row vector characterizes all four images, where the name of the row is the property that is characterized.



(a) Black Mirror.

(b) Holstee Manifesto.

Figure 1.3: Relative Frobenius Norm as a function of the rank of the approximation.

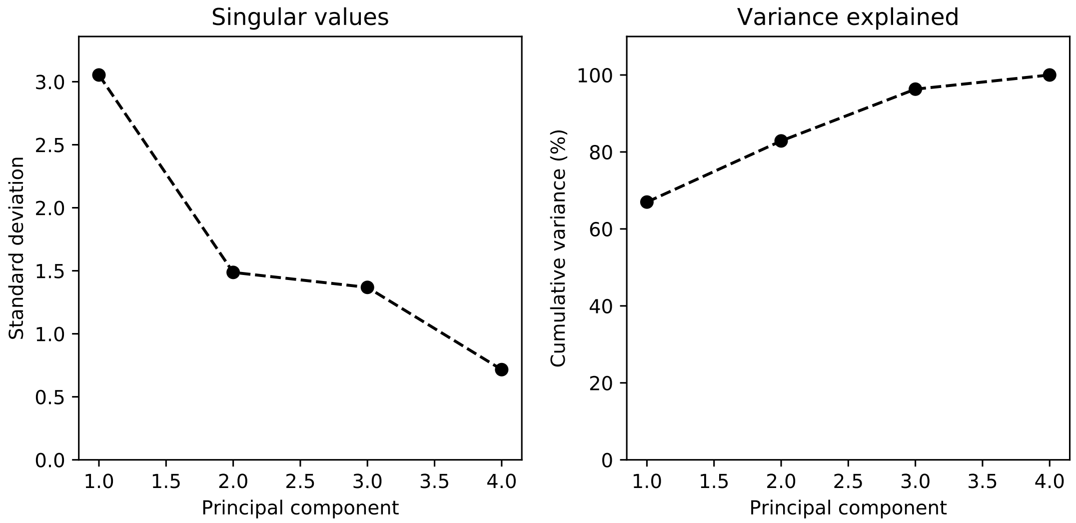


Figure 2.1: Toy example. Singular values and cumulative variance for covariance matrix.

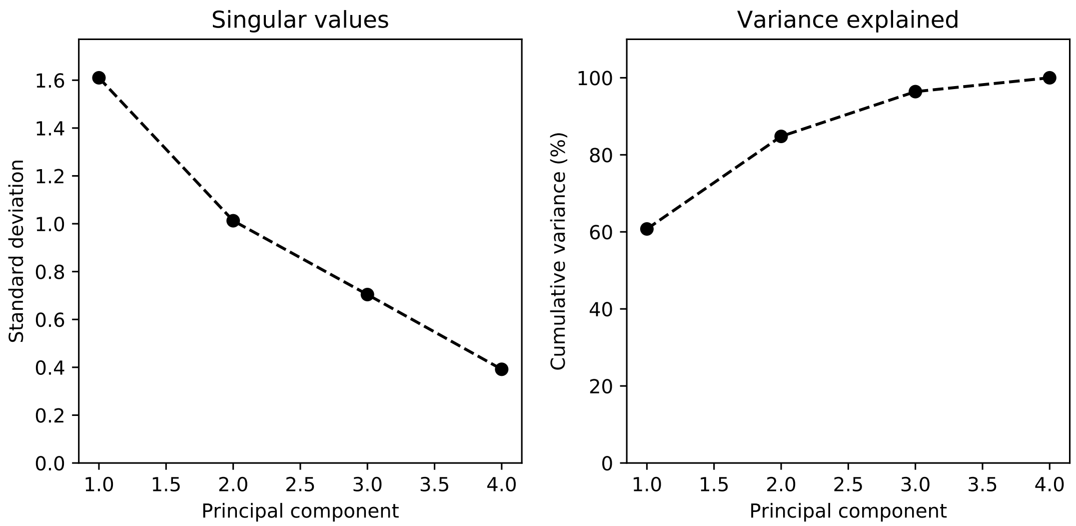


Figure 2.2: Toy example. Singular values and cumulative variance for correlation matrix.

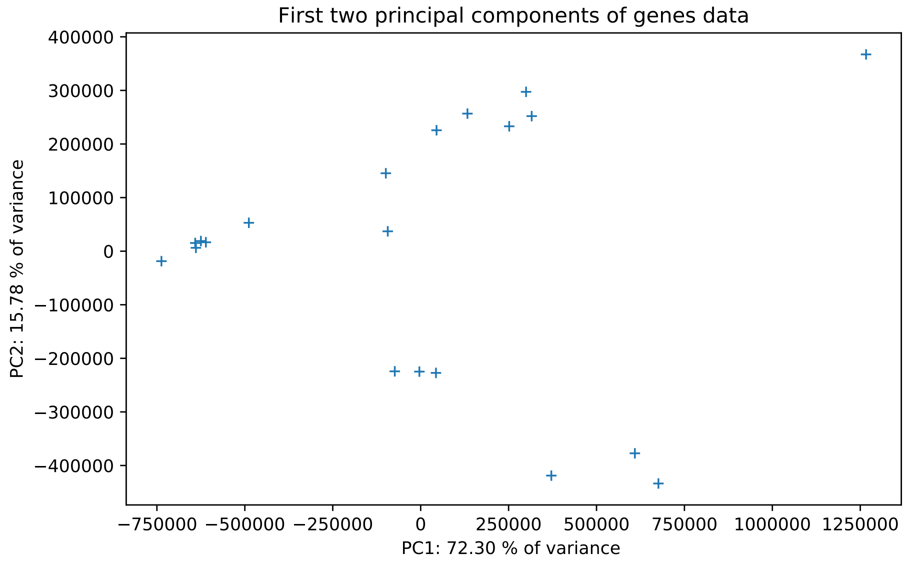


Figure 2.3: Genes data set. First vs. second principal component of covariance matrix.

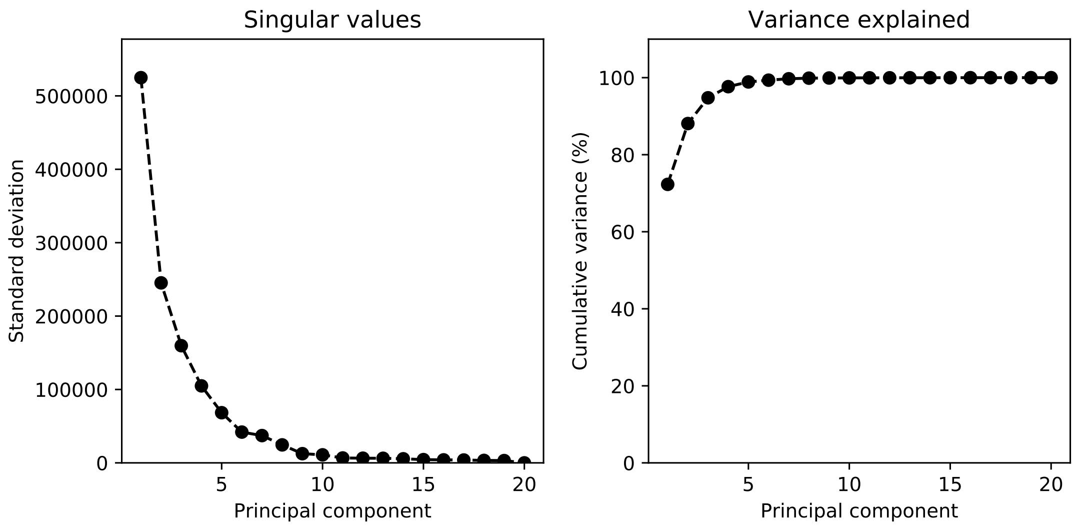


Figure 2.4: Genes data set. Singular values and cumulative variance of covariance matrix.