Worksheet: Singular Value Decomposition

Joaquin Rapela

February 12, 2024

1 Ken's worksheet

1. Solve Ken's worksheet on the SVD.

   For the pseudocolor image I recommend to use a blue-white-red colorscale, as in Figure 1, instead of the blue-green colorscale used in Ken's worksheet.

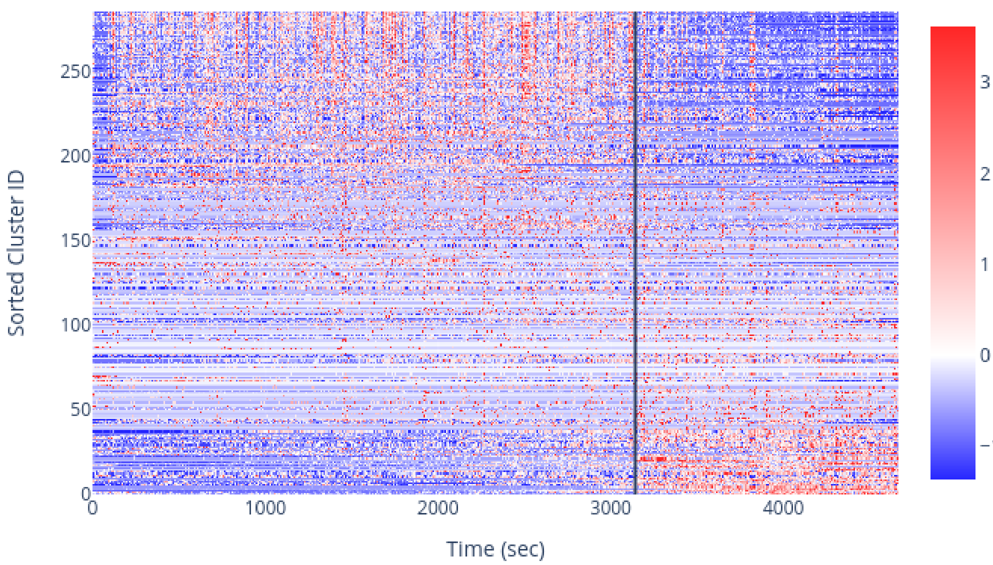

Figure 1: z-scores of binned spikes times of all neurons (bin size=1 sec, unsorted neurons). Neurons have been sorted according to their weight along the first left singular vector.

2. Explore the relation between neurons' activities and subject' responses by plotting the first left singular vector (i.e., first column of matrix U of the SVD $U\Sigma V^\intercal$) superimposed with vertical lines at the times of the subject's responses.

2 A few informative notes on the SVD (no exercises in this section)

Definition 1 (The SVD). *Given $M \in \mathbb{C}^{m\times n}$, a singular value decomposition (SVD) of M is a factorisation:*

$$M = USV^*$$

where

$$\begin{aligned} U &\in \mathbb{C}^{m\times m} \quad \textit{is unitary,} \\ V &\in \mathbb{C}^{n\times n} \quad \textit{is unitary,} \\ S &\in \mathbb{C}^{m\times n} \quad \textit{is diagonal.} \end{aligned}$$

In addition, it is assumed that the diagonal entries s_k of S are nonnegative and in nonincreasing order; that is, $s_1 \geq s_2 \geq \ldots \geq s_p \geq 0$, where $p = \min(m, n)$.

Definition 2 (Rank of a matrix). *The column rank of a matrix is the dimension of the space spanned by its columns. Similarly, the row rank of a matrix is the dimension of the space spanned by its rows. The column rank of a matrix is always equal to its row rank. This is a corollary of the SVD. So we refer to this number simply as the rank of a matrix.*

The rank of a matrix can be interpreted as a measure of the complexity of the matrix. Matrices with lower rank are simpler than those with larger rank.

The SVD decomposes a matrix as a sum of rank-one (i.e., very simple) matrices.

$$M = \sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^*$$

There are multiple other decompositions as sums of rank-one matrices. If $M \in \mathbb{C}^{m\times n}$, then it can be decomposed as a sum of m rank-one matrices given by its rows (i.e., $M = \sum_{i=1}^{m} \mathbf{e}_i \mathbf{m}_{i,\cdot}^*$, where $\mathbf{e}_i$ is the m-dimensional canonical unit vector, and $\mathbf{m}_{i,\cdot}$ is the ith row of M), or as a sum of n rank-one matrices given by its columns (i.e., $M = \sum_{j=1}^{n} \mathbf{m}_{\cdot,j} \mathbf{e}_j^*$, where $\mathbf{e}_j$ is the n-dimensional canonical unit vector, and $\mathbf{m}_{\cdot,j}$ is the jth column of M), or a sum of mn rank-one matrices each containing only one non-zero element (i.e., $M = \sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij} E_{ij}$, where E_{ij} is the matrix with all entries equal to zero, except the ij entry that is one, and m_{ij} is the entry of M at position ij).

A unique characteristic of the SVD compared to these other decompositions is that, if the rank of a matrix is r, then its SVD yields optimal approximations of lower rank ν, for $\nu = 1, \ldots, r$, as shown by Theorem 1.

Definition 3 (Frobenius norm). *The Frobenius norm of matrix $M \in \mathbb{C}^{m\times n}$ is*

$$\|M\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij}^2 \right)^{1/2}$$

Note that

$$\|M\|_F = \sqrt{tr(M^*M)} = \sqrt{tr(MM^*)} \tag{1}$$

Lemma 1 (Orthogonal matrices preserve the Frobenius norm)**.** *Let $M \in \mathbb{C}^{m\times n}$ and let $P \in \mathbb{C}^{m\times m}$ and $Q \in \mathbb{C}^{n\times n}$ be orthogonal matrices. Then*

$$\|PMQ\|_F = \|M\|_F$$

Proof.

$$\|PMQ\|_F = \sqrt{tr((PMQ)(PMQ)^*)} = \sqrt{tr(PMQQ^*M^*P^*)} = \sqrt{tr(PMM^*P^*)} \tag{2}$$
$$= \sqrt{tr(P^*PMM^*)} = \sqrt{tr(MM^*)} = \|M\|_F \tag{3}$$

Notes:

1. The first equality in Eq. 2 follows Eq. 1.
2. The second equality in Eq. 2 uses the fact that $(AB)^* = B^*A^*$.
3. The third equality in Eq. 2 holds because Q is orthogonal (i.e., $QQ^* = I$).
4. The first equality in Eq. 3 uses the cyclic property of the trace (i.e., tr(ABC)=tr(CAB)).
5. The first equality in Eq. 3 holds by the orthogonality of P.
6. The last equality in Eq. 3 again applies Eq. 1.

□

A direct consequence of Lemma 1 is that the Frobenius norm of any matrix $M = USV^*$ is

$$\|M\|_F = \|USV^*\|_F = \|S\|_F = \sqrt{\sum_{k=1}^{r} s_k^2}$$

Another consequence of Lemma 1 is the error in approximating a matrix M of rank r with its truncated SVD of rank ν (i.e., $M_\nu = \sum_{k=1}^{\nu} s_k \mathbf{u}_k \mathbf{v}_k^*$) is

$$\|M - M_\nu\|_F = \|\sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^* - \sum_{k=1}^{\nu} s_k \mathbf{u}_k \mathbf{v}_k^*\|_F = \|\sum_{k=\nu+1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^*\|_F = \sqrt{\sum_{k=\nu+1}^{r} s_k^2} \tag{4}$$

Theorem 1 (Eckart-Young-Mirsky)**.** *Let $M \in \mathbb{C}^{m\times n}$ be of rank r with singular value decomposition $M = \sum_{k=1}^{r} s_k \mathbf{u}_k \mathbf{v}_k^*$. For any ν with $0 \le \nu \le r$, define the* ***truncated SVD*** *as*

$$M_\nu = \sum_{k=1}^{\nu} s_k \mathbf{u}_k \mathbf{v}_k^* \tag{5}$$

Then

$$\|M - M_\nu\|_F = \inf_{\substack{\tilde{M} \in \mathbb{C}^{m\times n} \\ rank(\tilde{M}) \leq \nu}} \|M - \tilde{M}\|_F = \sqrt{\sum_{k=\nu+1}^{r} s_k^2} \tag{6}$$

Proof. We use the Weyl's inequality that relates the singular values of a sum of two matrices to the singular values of each of these matrices. Precisely, if $X, Y \in \mathbb{C}^{m\times n}$ and $s_i(X)$ is the ith singular value of X, then

$$s_{i+j-1}(X+Y) \leq s_i(X) + s_j(Y) \tag{7}$$

Let $\tilde{M}$ be a matrix of rank at most ν. Applying Eq. 7 to $X = M - \tilde{M}$, $Y = \tilde{M}$ and $j - 1 = \nu$ we obtain

$$s_{i+\nu}(M) \leq s_i(M - \tilde{M}) + s_{\nu+1}(\tilde{M}) = s_i(M - \tilde{M}) \tag{8}$$

The last equality in Eq. 8 holds because $\tilde{M}$ has rank less or equal to ν, and therefore its $\nu + 1$ singular value is zero.

$$\|M - M_\nu\|_F^2 = \sum_{j=\nu+1}^{r} s_j^2(M) = \sum_{i=1}^{r-\nu} s_{i+\nu}^2(M) \leq \sum_{i=1}^{r-\nu} s_i^2(M - \tilde{M}) \leq \sum_{i=1}^{\min(m,n)} s_i^2(M - \tilde{M}) \tag{9}$$

$$= \|M - \tilde{M}\|_F^2 \tag{10}$$

Notes:

1. The first equality in Eq. 9 holds by Eq. 4.
2. The second equality in Eq. 9 used the change of variables $i = j - \nu$.
3. The first inequality in Eq. 9 used Eq. 8
4. The last inequality in Eq. 9 is true because $r - \nu \leq \min(m, n)$ and adding squared singular values to the sum in the left hand side only increases this sum.
5. The equality in Eq. 10 again holds by Eq. 6 and by the fact that singular values of index larger than the rank of a matrix are zero.

The last equality in Eq. 6 follows from Eq. 4.

□

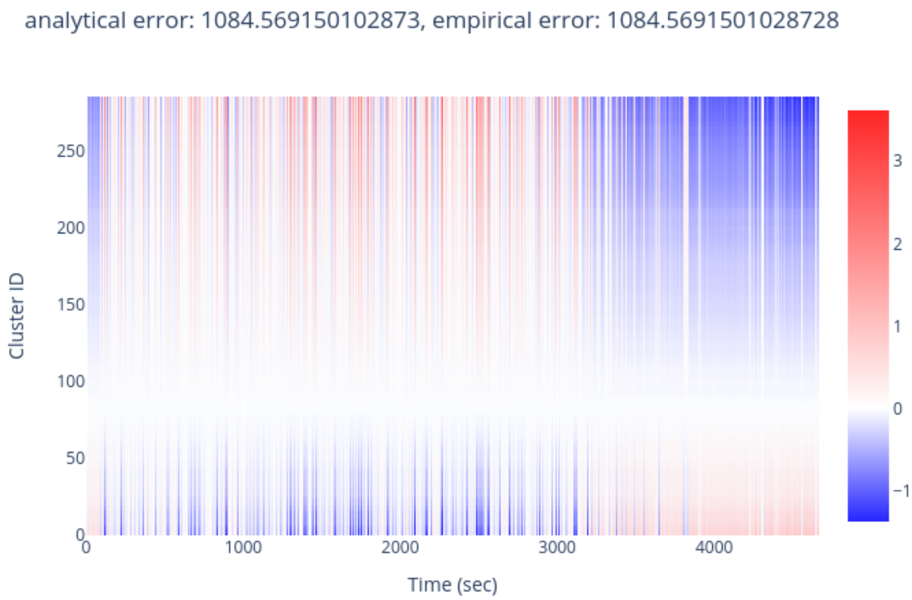

Figure 2: Low-rank approximation of the image in Figure 1 using a truncated SVD or rank 1. The title reports the empirical and analytical errors of the reconstruction. The empirical error is the Frobenius norm of the difference between the low-rank approximation and the image in Figure 1. The analytical error is computed from the singular values of the image in Figure 1 using Eq. 6. Click on the image to access its interactive version.

3 Verification of the Eckart-Young-Mirsky theorem

It is remarkable that Theorem 1 allows us to compute a lower bound on the error that will be achieved by any low-rank approximation by just using the singular values of the matrix to be approximated.

Verify Theorem 1, for different ranks $\nu \in \{1, \dots, 10\}$, by computing truncated SVD approximations of rank ν (Eq. 5) to the matrix shown in Figure 1. For each of these approximations, plot the approximation and insert in the title the emprical and analytical error. The empirical error is the Frobenius norm of the difference between the matrix if Figure 1 and its approximation of rank ν. The analytical error is the sum of singular values in Eq. 6. Figure 2 shows such a plot for a rank $\nu = 1$.

You should also observe that as the rank of the approximation increases, the approximation becomes more similar to the original matrix in Figure 1.