NLA 2021-2022
The singular value decomposition

Martin Sombra

27 October 2021

Singular values and singular vectors

The *singular value decomposition (SVD)* extends this factorization to any matrix, even outside the square case.

The *key ingredient* is

$$\boxed{\text{decoupling } Q \text{ and } Q^T}$$

Let A be an $m \times n$ matrix with $m \geqslant n$. There are two sets of *singular vectors*

$$u_1, \ldots, u_m \in \mathbb{R}^m \ (\textit{left}) \quad \text{and} \quad v_1, \ldots, v_n \in \mathbb{R}^n \ (\textit{right})$$

Both sets form orthogonal bases, and are connected by the relation

$$A\, v_i = \sigma_i\, u_i, \quad i = 1, \ldots, n,$$

for the *singular values* $\sigma_1, \ldots, \sigma_n \geqslant 0$

The SVD

In matrix form: both

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m \times m} \quad \text{and} \quad V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

are *orthogonal* square matrices:

$$U^{-1} = U^T \quad \text{and} \quad V^{-1} = V^T.$$

Set

$$\Sigma = \begin{array}{cl} n & \\ \left[\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma \\ & \mathbb{0} & \end{array}\right] & \begin{array}{l} \\ n \\ \\ m-n \end{array} \end{array}$$

The *(full) SVD* is the factorization $A\,V = U\,\Sigma$ or equivalently

$$A = U\,\Sigma\,V^T$$

Example

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

$$= U\Sigma V^T = \begin{bmatrix} 0.32 & -0.95 \\ 0.95 & 0.32 \end{bmatrix} \begin{bmatrix} 6.71 & 0 \\ 0 & 2.24 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \\ -0.71 & 0.71 \end{bmatrix}$$

Decomposition into matrices of rank 1

Set $r = \text{rank}(A)$. Then

- $u_1, \dots, u_r$ is a basis of $\text{Im}(A)$
- $v_{r+1}, \dots, v_n$ is a basis of $\text{Ker}(A)$

The column-row multiplication of $U\Sigma$ and V^T separates A into r pieces of rank 1:

$$A = \sum_{i=1}^{r} \sigma_i \, u_i \, v_i^T$$

In the example:

$$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = 6.71 \begin{bmatrix} 0.32 \\ 0.95 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \end{bmatrix} + 2.24 \begin{bmatrix} -0.95 \\ 0.32 \end{bmatrix} \begin{bmatrix} -0.71 & 0.71 \end{bmatrix}$$

We have that

$$\sigma_1 = 6.71 > \sigma_2 = 2.24$$

$\rightsquigarrow$ the first piece is “more representative” of A

The thin and the reduced SVD's

The *thin SVD* avoids the 0's in the lower part of Σ and uses a diagonal matrix for the singular values:

$$A = U_n \Sigma_n V^T$$

with

$$\begin{aligned} U_n &= \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} & m \times n\text{-orthogonal} \\ \Sigma_n &= \operatorname{diag}(\sigma_1, \ldots, \sigma_n) & n \times n\text{-diagonal} \end{aligned}$$

The *reduced SVD* keeps only the nonzero singular values to remove the parts that are going to produce zeros for sure:

$$A = U_r \Sigma_r V_r^T$$

with

$$\begin{aligned} U_r &= \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} & m \times r\text{-orthogonal} \\ \Sigma_r &= \operatorname{diag}(\sigma_1, \ldots, \sigma_r) & r \times r\text{-diagonal} \\ V_r^T &= \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} & n \times r\text{-orthogonal} \end{aligned}$$

Reducing to SPD matrices

To identify the singular values and vectors, we can consider the SPD matrices

$$A^T A = (V \Sigma^T U^T)(U \Sigma V^T) = V \Sigma^T \Sigma V^T \in \mathbb{R}^{n \times n}$$
$$A A^T = (U \Sigma V^T)(V \Sigma^T U^T) = U \Sigma \Sigma^T U^T \in \mathbb{R}^{m \times m}$$

Then

- V contains the orthonormal eigenvectors of $A^T A$
- U contains the orthonormal eigenvectors of $A A^T$
- $\sigma_1, \ldots, \sigma_r$ are the nonzero eigenvalues of both $A^T A$ and $A A^T$

Existence and computation of the SVD

Consider the diagonalization

$$A^T A = Q \Lambda Q^T$$

and let $v_1, \ldots, v_n$ be the columns of Q, ordered so that $v_1, \ldots, v_r$ correspond to the nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0$

Set

$$\sigma_k = \lambda_k^{1/2} \quad \text{and} \quad u_k = \sigma_k^{-1} A\, v_k, \qquad k = 1, \ldots, r$$

The u_k's are orthonormal:

$$\begin{aligned} \langle u_j, u_k \rangle &= u_j^T u_k = (\sigma_j^{-1} A\, v_j)^T (\sigma_k^{-1} A\, v_k) \\ &= \sigma_j^{-1} \sigma_k^{-1} v_j^T A^T A\, v_k = v_j^T v_k = \langle v_j, v_k \rangle = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \end{aligned}$$

$\rightsquigarrow$ the reduced SVD

$$A = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

Existence and computation of the SVD (cont.)

For the full SVD, take

$$v_{r+1}, \ldots, v_n \in \mathbb{R}^n \quad \text{and} \quad u_{r+1}, \ldots, u_m \in \mathbb{R}^m$$

completing $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$ to orthonormal bases. Then

$$A = U \Sigma V^T$$

with $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}$, $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & \mathbb{0} & & \end{bmatrix}$$

Example (cont.)

Set as before $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$. Then

$$A^T A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \quad \text{and} \quad A\,A^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

The eigenvalues and eigenvectors of $A^T A$ are

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Hence the right singular vectors are

$$v_1 = \frac{1}{2^{1/2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} \quad \text{and} \quad v_2 = \frac{1}{2^{1/2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix}$$

and the singular values are

$$\sigma_1 = 45^{1/2} = 6.71 \quad \text{and} \quad \sigma_2 = 5^{1/2} = 2.24$$

Moreover, the left singular vectors are

$$u_1 = \sigma_1^{-1} A\, v_1 = \frac{1}{10^{1/2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0.32 \\ 0.95 \end{bmatrix}$$

$$u_2 = \sigma_2^{-1} A\, v_2 = \frac{1}{10^{1/2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.95 \\ 0.32 \end{bmatrix}$$

Hence $A = U \Sigma V^T$ with

$$U = \begin{bmatrix} 0.32 & -0.95 \\ 0.95 & 0.32 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 6.71 & \\ & 2.24 \end{bmatrix}, \quad V = \begin{bmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{bmatrix}$$

Some particular cases

- Let $S = Q \Lambda Q^T$ be the diagonalization of a SPD matrix. Then
$$U = V = Q \quad \text{and} \quad \Sigma = \Lambda$$
- The singular values of an orthonormal $n \times n$ matrix Q are all equal to 1.
- Let $A = x\, y^T$ be an $m \times n$ matrix of rank 1, with $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$
Its reduced SVD is $A = U_1 \Sigma_1 V_1^T$
$$U_1 = \frac{x}{\|x\|_2}, \quad \Sigma_1 = \left[\|x\|_2 \, \|y\|_2 \right], \quad V_1 = \frac{y}{\|y\|_2}$$

The geometry of the SVD

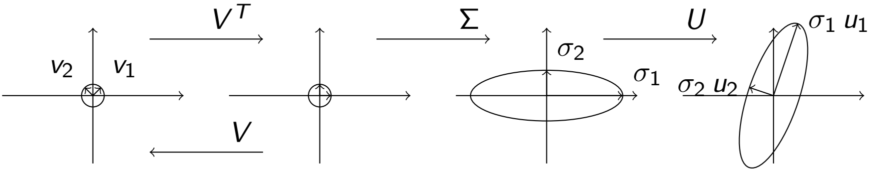
The SVD decomposes the matrix as

$$\text{orthogonal} \times \text{diagonal} \times \text{orthogonal}\,.$$

The unit sphere $\mathbb{S}_n$ of $\mathbb{R}^n$ is send to the ellipsoid $A\,\mathbb{S}_n$ of $\mathbb{R}^m$ centered at the origin and with axes

$$\sigma_i\,u_i, \quad i = 1, \ldots, r.$$

In dimension 2 we can draw the process: for $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ it gives



Computing the 2-norm and the Frobenius norm

The 2-norm of A can be computed in terms of its SVD: this norm is invariant with respect to multiplication by orthogonal matrices and so

$$\|A\|_2 = \|U^T A V\|_2 = \|\Sigma\|_2 = \sigma_1$$

The Frobenius norm is also invariant with respect to multiplication by orthogonal matrices, and so

$$\|A\|_{\mathrm{F}} = \|U^T A V\|_{\mathrm{F}} = \|\Sigma\|_{\mathrm{F}} = \left(\sum_{i=1}^{r} \sigma_i^2\right)^{1/2}$$

The best low rank approximation

Eckart-Young theorem: for $k = 1, \ldots, r$, the matrix

$$A_k = \sum_{i=1}^{k} \sigma_i \, u_i \, v_i^T = U_k \, \Sigma_k \, V_k^T$$

is the best rank k approximation of A with respect to both the 2-norm and the Frobenius norm

Hence for $\| \cdot \| = \| \cdot \|_2$ or $\| \cdot \| = \| \cdot \|_{\mathrm{F}}$ we have that

$$\|A - B\| \geqslant \|A - A_k\|$$

for any other $m \times n$ matrix of rank $\leqslant k$

The best low rank approximation (cont.)

For both norms, by the orthogonal invariance we have that

$$\|A - A_k\| = \|\Sigma - \Sigma_k\| = \left\| \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_r \\ & & & \mathbb{0} & & \end{bmatrix} \right\|$$

Hence

$$\|A - A_k\|_2 = \sigma_{k+1} \quad \text{and} \quad \|A - A_k\|_{\mathrm{F}} = \left(\sum_{i=k+1}^{r} \sigma_i \right)^{1/2}$$

Image compression

A B/W image of $m \times n$ pixels can be coded by an $m \times n$ matrix A with entries $a_{i,j} \in [0,1]$, indicating the brightness of the (i,j)-pixel:

$$0 \text{ (black)} \cdots \text{ gray } \cdots 1 \text{ (white)}$$

Instead of transmitting/storing A, we can replace it by its k-th rank approximation

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

which has size $k\,(m+n+1)$ (or $k\,(m+n)$ if we store $\sigma_i \, u_i$)

The *relative error of the approximation* is

$$\frac{\|A - A_k\|_2}{\|A\|_2} = \frac{\sigma_{k+1}}{\sigma_1}$$

and its *compression ratio* is

$$\frac{k\,(m+n)}{m\,n}.$$

Image compression (cont.)

Here is a 320×200-picture of a clown and its approximations:

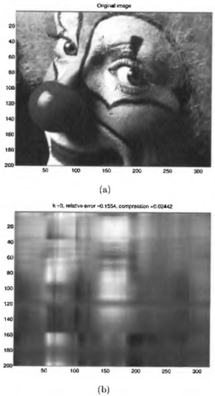


Fig. 3.3. *Image compression using the SVD.* (a) *Original image.* (b) *Rank* $k = 3$ *approximation.*

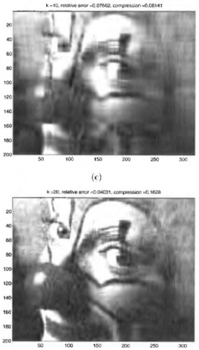


Fig. 3.3. *Continued.* (c) *Rank* $k = 10$ *approximation.* (d) *Rank* $k = 20$ *approximation.*