9.2 Singular value decomposition

In the diagonalisation and in the Jordan decomposition, we had three parts in the form $A = UDV$. There we had

in general not diagonal

> *U and V are inverse to each other.* $(A = UJU^{-1})$

Now, if we drop that condition, we can actually fulfil the following two properties even for rectangle matrices

> *(1) D is diagonal,* $A = UDV$
>
> *(2) U and V are unitary square matrices.* (conserve angles and lengths)

Since we allow U and V to be any square matrices, which means $U \in \mathbb{F}^{n\times n}$ and $V \in \mathbb{F}^{m\times m}$, we can consider an arbitrary rectangular matrix $A \in \mathbb{F}^{m\times n}$.

rectangular $\rightarrow A = U \cdot \underset{\Sigma}{\underset{\parallel}{D}} \cdot V^*$

Singular value decomposition of A

$$
\underset{\text{arbitrary}}{m\overset{n}{\boxed{A}}} = \underset{\text{unitary}}{m\overset{m}{\boxed{U}}} \cdot \underset{\text{diagonal}}{m\overset{n}{\boxed{\Sigma}}} \cdot \underset{\text{unitary}}{n\overset{n}{\boxed{V^*}}} . \tag{9.3}
$$

The word "diagonal" for a rectangular matrix Σ is of course not literally correct. It means the following here:

$$
\Sigma = m\underset{\text{if } m \geq n}{\overset{n}{\begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ & 0 & \end{pmatrix}}} \quad \text{or} \quad \Sigma = m\underset{\text{if } m \leq n}{\overset{n}{\left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & 0 \\ & & s_m & \end{array}\right)}} , \tag{9.4}
$$

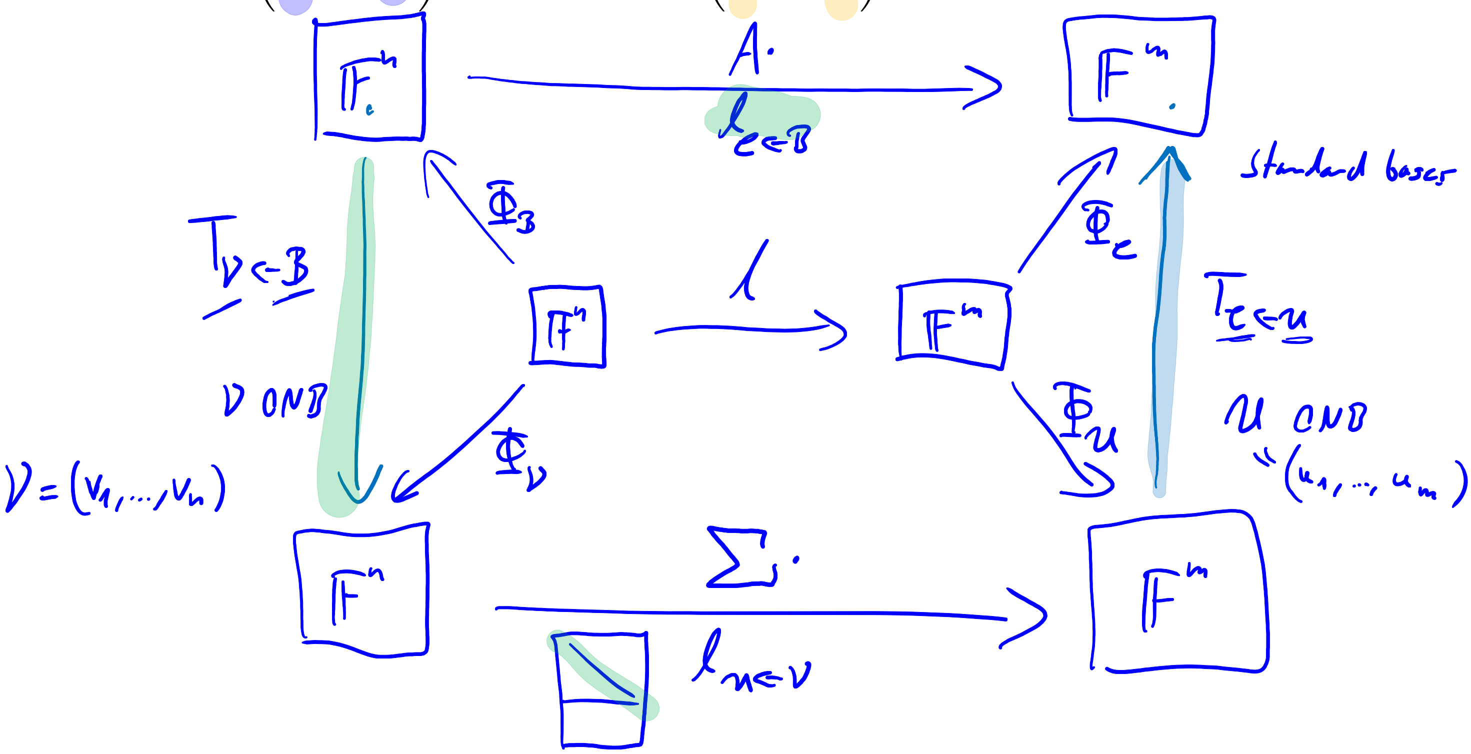
where every other entry is zero.

The equation $A = U\Sigma V^*$ tells us that A and Σ are equivalent matrices, $A \sim \Sigma$. The matrix A is the matrix representation of the linear map $\ell := f_A : \mathbb{F}^n \to \mathbb{F}^m$ given by $\mathbf{x} \mapsto A\mathbf{x}$ with respect to the standard bases $\mathcal{B}$ in $\mathbb{F}^n$ and $\mathcal{C}$ in $\mathbb{F}^m$. The change of basis to an ONB $\mathcal{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ in $\mathbb{F}^n$ and an ONB $\mathcal{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ in $\mathbb{F}^m$ gives us another matrix represenation $\ell_{\mathcal{U}\leftarrow\mathcal{V}}$ which is the "diagonal" matrix from (9.4):

$$\underbrace{\ell_{\mathcal{C}\leftarrow\mathcal{B}}}_{A} = \underbrace{T_{\mathcal{C}\leftarrow\mathcal{U}}}_{U}\ \underbrace{\ell_{\mathcal{U}\leftarrow\mathcal{V}}}_{\Sigma}\ \underbrace{T_{\mathcal{V}\leftarrow\mathcal{B}}}_{V^*}$$

with

$$U = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix} = T_{\mathcal{C}\leftarrow\mathcal{U}}, \qquad V = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} = T_{\mathcal{B}\leftarrow\mathcal{V}}, \qquad V^* = V^{-1} = T_{\mathcal{V}\leftarrow\mathcal{B}}.$$



Because of $A \sim \Sigma$ and the characterisation of equivalences given by Proposition 8.20, we know $\operatorname{rank}(\Sigma) = \operatorname{rank}(A) =: r \leq m, n$. Hence, exactly r of the entries s_i in (9.4) are non-zero. Of course, we can choose $\mathbf{u}_i$, $\mathbf{v}_i$ in such an order that we have $s_1, \dots, s_r$ as the non-zero elements.

Hence, we can see the matrix Σ from (9.4) as the following matrix representation:

$$\Sigma = \ell_{\mathcal{U}\leftarrow\mathcal{V}} = \begin{array}{c} \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_r \\ \mathbf{u}_{r+1} \\ \vdots \\ \mathbf{u}_m \end{array} \begin{array}{c} \begin{array}{cccccc} \mathbf{v}_1 & \cdots & \mathbf{v}_r & \mathbf{v}_{r+1} & \cdots & \mathbf{v}_n \end{array} \\ \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & \\ & & s_r & \\ \hline & & & \\ & & & \\ & & & \end{array}\right) \end{array} \in \mathbb{F}^{m\times n} \qquad (9.5)$$

Multiplying $A = U\Sigma V^*$ from the right with V gets us $AV = U\Sigma$.

$$A = U\Sigma V^* \iff AV = U\Sigma$$

Let us look at this in more detail:

$$\begin{pmatrix} | & & | \\ A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \\ | & & | \end{pmatrix} = A \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} = AV = U\Sigma$$

$$= \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{pmatrix} \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & \\ & & s_r & \\ \hline & & & \end{array}\right) = \begin{pmatrix} | & & | & | & & | \\ s_1\mathbf{u}_1 & \cdots & s_r\mathbf{u}_r & \mathbf{o} & \cdots & \mathbf{o} \\ | & & | & | & & | \end{pmatrix}.$$

$(AB)^* = B^*A^*$

$\begin{pmatrix} | & \\ A^*u_1 & \cdots \\ | & \end{pmatrix}$

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

$=$

$\begin{pmatrix} | & \\ \overline{s_1} v_1 & \cdots \\ | & \end{pmatrix}$

Therefore, we have:

$$A\mathbf{v}_1 = s_1\mathbf{u}_1, \quad A\mathbf{v}_2 = s_2\mathbf{u}_2, \quad \ldots, \quad A\mathbf{v}_r = s_r\mathbf{u}_r, \quad A\mathbf{v}_{r+1} = \mathbf{o}, \quad \ldots, \quad A\mathbf{v}_n = \mathbf{o}. \tag{9.6}$$

Analogously, we get from $A^* = (U\Sigma V^*)^* = V\Sigma^* U^*$ that $A^*U = V\Sigma^*$ and hence:

$$A^*\mathbf{u}_1 = \overline{s_1}\mathbf{v}_1, \quad \ldots, \quad A^*\mathbf{u}_r = \overline{s_r}\mathbf{v}_r, \quad A^*\mathbf{u}_{r+1} = \mathbf{o}, \quad \ldots, \quad A^*\mathbf{u}_m = \mathbf{o}. \tag{9.7}$$

From (9.5) or (9.6), we get:

Exercise: $\mathrm{Ker}(A^*) = \mathrm{Ran}(A)^\perp$
$\mathrm{Ran}(A^*) = \mathrm{Ker}(A)^\perp$

Proposition 9.9. Kernel and range of A

$$\begin{aligned} \mathrm{Ker}(A) &= \mathrm{Ker}(\ell) = \mathrm{Span}(\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n), & \mathrm{Ker}(A)^\perp &= \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_r) \\ \mathrm{Ran}(A) &= \mathrm{Ran}(\ell) = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_r), & \mathrm{Ran}(A)^\perp &= \mathrm{Span}(\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m) \end{aligned}$$

Please recognise the rank-nullity theorem $\dim(\mathrm{Ker}(A)) + \dim(\mathrm{Ran}(A)) = \dim(\mathbb{F}^n)$.

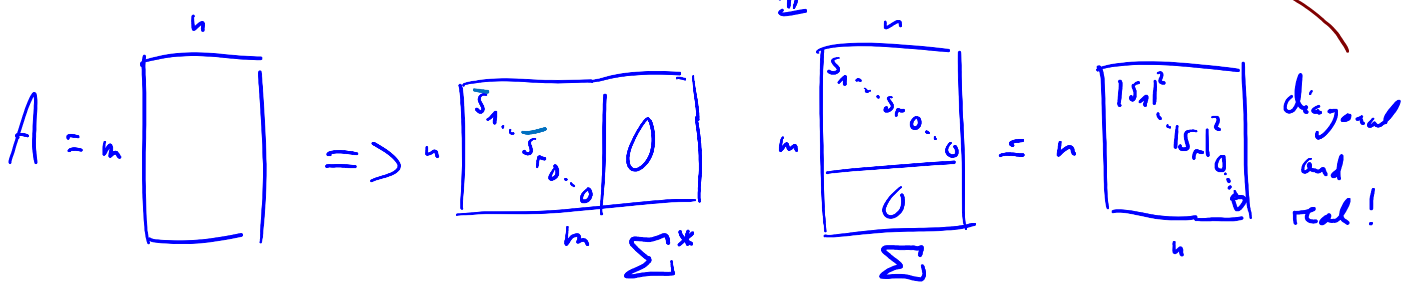
Of course, we see that the decomposition $A = U\Sigma V^*$ from (9.3) is useful as a representation of the corresponding linear map. We will later see how we can use this in applications. However, the question remains how to get U, V and Σ?

Let us go back to the result: From $A = U\Sigma V^*$, we would get:

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

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^* \overbrace{U^*U}^{\mathbb{1}} \Sigma V^* = V\Sigma^*\Sigma V^* \tag{9.8}$$

$$\text{and} \qquad AA^* = (U\Sigma V^*)(U\Sigma V^*)^* = U\Sigma \underbrace{V^*V}_{\mathbb{1}} \Sigma^* U^* = U\Sigma\Sigma^* U^*. \tag{9.9}$$



$$A^*A = V \underbrace{\Sigma^*\Sigma}_{\text{diagonal}} V^* \qquad AA^* = U \underbrace{\Sigma\Sigma^*}_{\text{diagonal}} U^*$$

selfadjoint selfadjoint

$\Rightarrow$ diagonalisable with ONB of eigenvectors and real eigenvalues

$$(A^*A)^* = A^*A^{**} = A^*A$$

Because of (9.5) and $\overline{s_i}s_i = s_i\overline{s_i} = |s_i|^2$, we have square matrices

$$\Sigma^*\Sigma = \underbrace{\begin{pmatrix} |s_1|^2 & & & & & \\ & \ddots & & & & \\ & & |s_r|^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}}_{n} n \quad \text{and} \quad \Sigma\Sigma^* = \underbrace{\begin{pmatrix} |s_1|^2 & & & & & \\ & \ddots & & & & \\ & & |s_r|^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}}_{m} m$$

that are also diagonal. Hence, (9.8) and (9.9) show us the unitary diagonalisations of the square matrices AA^* and A^*A. Recall that both matrices are self-adjoint and have by Proposition 6.44 in fact an ONB consisting of eigenvectors. These orthonormal eigenvectors (w.r.t to standard inner product!) are chosen as the columns of the matrices U and V.

Therefore, we find the eigenvalues of A^*A on the diagonal of $\Sigma^*\Sigma$ and the eigenvalues of AA^* on the diagonal of $\Sigma\Sigma^*$.

Actually, we can choose the number s_i as we want in $\mathbb{F}$ as long as $|s_i|^2$ is the ith eigenvalue of A^*A or AA^*. A simple choice is, of course, $s_1, \ldots, s_r$ being real and positive numbers.

In summary, we now have everything for U, V and Σ:

Definition 9.10. Singular values and singular vectors

Let $A \in \mathbb{F}^{m\times n}$. The (non-negative) square roots from the eigenvalues of A^*A are called the singular values of A and we order them from highest to lowest (counted with multiplicities):

$$s_1 \geq s_2 \geq \cdots \geq s_n \geq 0\,.$$

The vectors from an orthonormal family $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ consisting of eigenvectors of A^*A, with the same order as for $s_1^2, \ldots, s_n^2$, are called the right-singular vectors of A. In the same way, the vectors from an orthonormal family $(\mathbf{u}_1, \ldots, \mathbf{u}_m)$ consisting of eigenvectors of AA^*, with the same order as for $s_1^2, \ldots, s_n^2$, are called the left-singular vectors of A.

The factorisation

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

given by $U = (\mathbf{u}_1 \cdots \mathbf{u}_m)$, $V = (\mathbf{v}_1 \cdots \mathbf{v}_n)$ and Σ from (9.4), is called the singular value decomposition of A. In short: SVD.

To summarise everything, let us state the whole algorithm:

Algorithm for the SVD of A

Given: *An arbitrary matrix* $A \in \mathbb{F}^{m \times n}$.

Wanted: *Unitary matrices* $U \in \mathbb{F}^{m \times m}$, $V \in \mathbb{F}^{n \times n}$ *and diagonal matrix* $\Sigma \in \mathbb{R}^{m \times n}$ *for the singular value decomposition* $A = U\Sigma V^*$ *of* A.

Algorithm:

- *Calculate the matrix* A^*A *and all eigenvalues* $\lambda_1, \dots, \lambda_n$ *(counted with multiplicities) and use the ordering such that* $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.
- *Find an ONB* $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ *consisting of eigenvectors for* A^*A.
- *Define* r *for number of the last non-zero eigenvalue* λ_i.
- *For* $i = 1, \dots, r$:

$A\mathbf{v}_i = s_i \mathbf{u}_i$

  - *Set* $s_i := \sqrt{\lambda_i}$. > 0
  - *Set* $\mathbf{u}_i := \frac{1}{s_i} A\mathbf{v}_i$, *cf.* (9.6).
- *Set* $s_{r+1}, \dots, s_m := 0$.
- *Add to* $(\mathbf{u}_1, \dots, \mathbf{u}_r)$ *a family* $(\mathbf{u}_{r+1}, \dots, \mathbf{u}_m)$ *such that it is an ONB of* $\mathbb{F}^m$.
- *Set* $U := (\mathbf{u}_1 \cdots \mathbf{u}_m)$, $V := (\mathbf{v}_1 \cdots \mathbf{v}_n)$ *and* Σ *as in* (9.4) *or* (9.5).

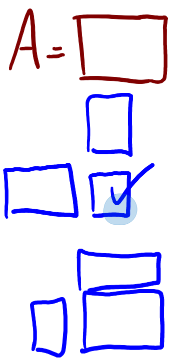
Two options: (1) Calculate SVD of A
(2) Calculate SVD of $A^* =: B$

Choose the easiest

$A = U\Sigma V^*$

Rule of thumb: Calculating the singular vectors

- *Alternatively, one finds* $\mathbf{u}_1, \dots, \mathbf{u}_m$ *as the eigenvectors of* AA^*. *However, that is more costly than using* $\mathbf{v}_i$ *if we already have them.*
- *In the case* $m < n$, AA^* *is smaller than* A^*A, *hence it is better to calculate the eigenvalues* λ_i *and eigenvectors* $\mathbf{u}_i$ *of the matrix* A^*A *and to use* (9.7) *for getting* $\mathbf{v}_i$.
- *Depending on the application, the eigenvectors* $\mathbf{u}_i$ *and* $\mathbf{v}_i$ *for* $i > r$ *might not be important (cf.* (9.11)*).*



Example 9.11. Consider the matrix $A = \begin{pmatrix} 1 & \sqrt{3} \\ -2 & 0 \end{pmatrix}$. We have $m = n = 2$ and $\mathbb{F} = \mathbb{R}$.

- Calculate

$$A^*A = \begin{pmatrix} 1 & -2 \\ \sqrt{3} & 0 \end{pmatrix} \begin{pmatrix} 1 & \sqrt{3} \\ -2 & 0 \end{pmatrix} = \begin{pmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}$$

selfadjoint / symmetric

and get $\det(A^*A - \lambda\mathbb{1}) = (5-\lambda)(3-\lambda) - 3 = \lambda^2 - 8\lambda + 12 = (\lambda - 2)(\lambda - 6)$. The eigenvalues of A^*A are (in decreasing order) $\lambda_1 = 6$ and $\lambda_2 = 2$.

$$\underset{6}{\lambda_1} \geq \underset{2}{\lambda_2} \geq 0$$

- The eigenvector $\mathbf{v}_1$ for $\lambda_1 = 6$: We solve the linear equation: $(A^*A - 6\mathbb{1})\mathbf{v}_1 = \mathbf{o}$,

$$\text{hence} \quad \begin{pmatrix} -1 & \sqrt{3} \\ \sqrt{3} & -3 \end{pmatrix} \mathbf{v}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longleftarrow \quad \mathbf{v}_1 = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \quad \text{(normalised).}$$

  Eigenvector $\mathbf{v}_2$ for $\lambda_2 = 2$: We solve the equation $(A^*A - 2\mathbb{1})\mathbf{v}_2 = \mathbf{o}$,

$$\text{and get} \quad \begin{pmatrix} 3 & \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} \mathbf{v}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longleftarrow \quad \mathbf{v}_2 = \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix} \quad \text{(normalised).}$$

  Both vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are automatically orthogonal (Proposition 6.43).
- Both eigenvalues, $\lambda_1 = 6$ and $\lambda_2 = 2$, are non-zero. Hence $r = 2$.
- Next thing, we calculate $s_1 := \sqrt{\lambda_1} = \sqrt{6}$ and

$$\mathbf{u}_1 := \frac{1}{s_1} A\mathbf{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & \sqrt{3} \\ -2 & 0 \end{pmatrix} \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} = \frac{1}{2\sqrt{6}}\begin{pmatrix} 2\sqrt{3} \\ -2\sqrt{3} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

- Then $s_2 := \sqrt{\lambda_2} = \sqrt{2}$ and

$$\mathbf{u}_2 := \frac{1}{s_2} A\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \sqrt{3} \\ -2 & 0 \end{pmatrix} \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix} = \frac{1}{2\sqrt{2}}\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- In summary, we get:

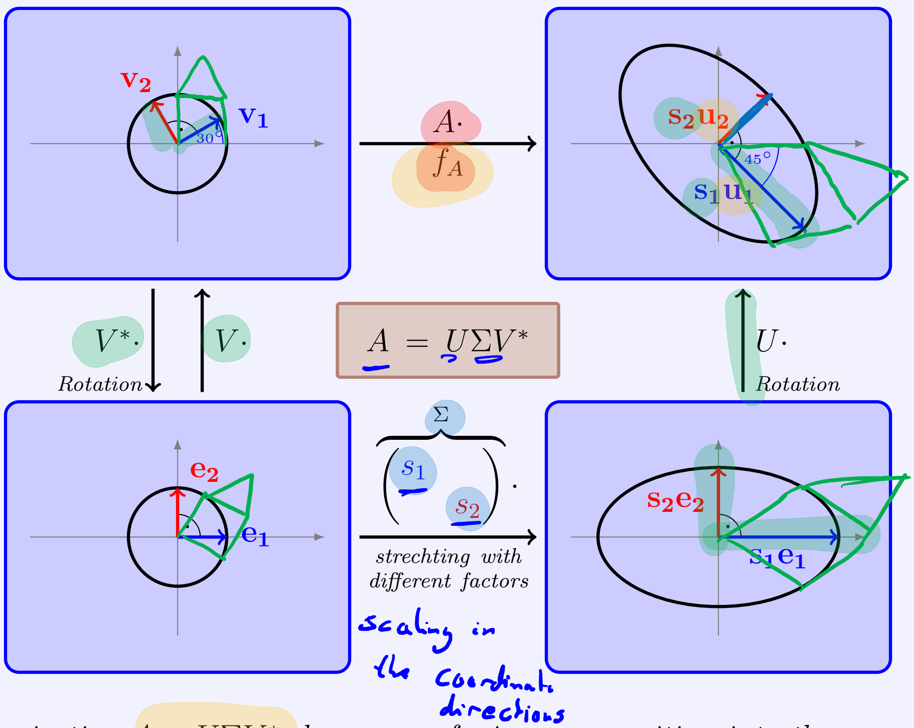
$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad V = \frac{1}{2}\begin{pmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{2} \end{pmatrix}.$$

Because we started in $\mathbb{F} = \mathbb{R}$, we could do the whole calculation inside $\mathbb{R}$.

The unitary matrices U and V are indeed orthogonal matrices if all entries are real and, by our Definition in 5.29, describe rotations (if $\det(\cdot) = 1$) or reflections (if $\det(\cdot) = -1$). In the example above, both matrices are rotations: U rotates $\mathbb{R}^2$ by $-45°$ and V rotates it by $30°$.

In Chapter 3, we have seen that linear maps $f_A : \mathbb{R}^2 \to \mathbb{R}^2$ with $\mathbf{x} \mapsto A\mathbf{x}$ can only stretch, rotate and reflect. Hence, a linear map changes the unit circle into an ellipse or it collapses into a line or point. By using the SVD, we can explain in more details what happens exactly. Let us look at our Example 9.11:

$A = U\Sigma V^*$ means: $f_A =$ rotating, stretching, rotating



The factorisation $A = U\Sigma V^*$ decompose f_A in a composition into three maps:

(1) *a rotation by $-30°$ (that is multiplication by $V^* = V^{-1}$),*

(2) *stretching separately into two directions (with factors $s_1 = \sqrt{6}$ in direction of the x_1-axis and with factor $s_2 = \sqrt{2}$ in direction of the x_2-axis) and*

(3) *a rotation by $-45°$ (multiplication by U).*

The major axis and minor axis of the ellipse, which is construed by f_A from the unit circle, are given by the eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ and the lengths are given by the singular values s_1 and s_2.

The singular values s_i give us the *stretching factors* in certain (orthogonal) directions. For the largest singular value, s_1, we have

$$s_1 = \|s_1\mathbf{u}_1\| = \|A\mathbf{v}_1\| = \max\{\|A\mathbf{x}\| : \mathbf{x} \in \mathbb{F}^n, \|\mathbf{x}\| = 1\} =: \|A\|. \qquad (9.10)$$

The here defined number $\|A\|$ is the already introduced matrix norm of A. It says how long the vector $A\mathbf{x} \in \mathbb{F}^m$ can be at most when $\mathbf{x} \in \mathbb{F}^n$ has length 1. The matrix norm fulfils the three properties of a norm.

$$A = U\Sigma V^* = U \cdot \begin{pmatrix} s_1 &&&& \\ & s_2 &&& \\ && \ddots && \\ &&& s_r & \\ &&&& \ddots \end{pmatrix} V^*, \quad r = \mathrm{rank}(A)$$

small

$$A_k = U \begin{pmatrix} s_1 &&&&& \\ & \ddots &&&& \\ && s_k &&& \\ &&& 0 && \\ &&&& \ddots & \\ &&&&& 0 \end{pmatrix} V^*$$

set some of the small ones to zero

$\mathrm{rank}(A_k) = k$

A_k low-rank approximation of A

Now we look at an important application of the SVD. We start with a calculation:

A as sum of r dyadic products

$$A = U\Sigma V^* = U \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & \\ & & s_r & \\ \hline & & & \end{array}\right) V^* = U\left(\left(\begin{array}{ccc|c} s_1 & & & \\ & & & \\ \hline & & & \end{array}\right) + \cdots + \left(\begin{array}{ccc|c} & & & \\ & & s_r & \\ \hline & & & \end{array}\right) \right) V^*$$

$$= U \left(\begin{array}{ccc|c} s_1 & & & \\ & & & \\ \hline & & & \end{array}\right) V^* + \cdots + U \left(\begin{array}{ccc|c} & & & \\ & & s_r & \\ \hline & & & \end{array}\right) V^*$$

$$= s_1 \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix} \begin{pmatrix} — \mathbf{v}_1^* — \end{pmatrix} + \cdots + s_r \begin{pmatrix} | \\ \mathbf{u}_r \\ | \end{pmatrix} \begin{pmatrix} — \mathbf{v}_r^* — \end{pmatrix} = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^* \quad (9.11)$$

As we know, A has rank r. Each of the r terms in (9.11) has rank 1. Depending on the rate of decay of the singular values

$$s_1 \geq s_2 \geq \cdots \geq s_r > 0$$

we could omit some terms in the sum (9.11) without changing the matrix so much. We call this a *low-rank matrix approximation* of A.

Example 9.12. Let us look at an 8-bit-grey picture with 537×358 pixels (which shows the only moon the planet Earth has at the moment):

0 ~> black

255 ~> white

This can be saved as a matrix $A \in \mathbb{R}^{537\times 358}$ where, in the entries, only integer values

$0, 1, \ldots, 255$ are allowed.

rank(A) = 358

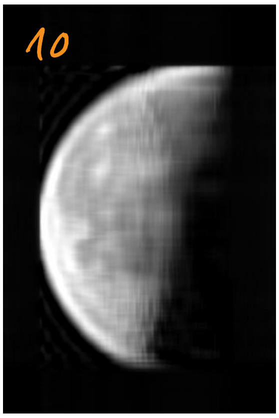
$$A = \begin{pmatrix} 1 & 3 & 7 & 5 & \ldots & 16 & 8 \\ 3 & 7 & 3 & 3 & \ldots & 11 & 12 \\ \vdots & \vdots & \vdots & \vdots & & & \\ 6 & 7 & \ldots & 248 & \ldots & 7 & 6 \\ \vdots & \vdots & \vdots & \vdots & & & \\ 4 & 8 & 8 & 5 & \ldots & 4 & 8 \\ 4 & 8 & 3 & 3 & \ldots & 6 & 6 \\ 2 & 3 & 3 & 4 & \ldots & 9 & 9 \end{pmatrix}$$

187 KB

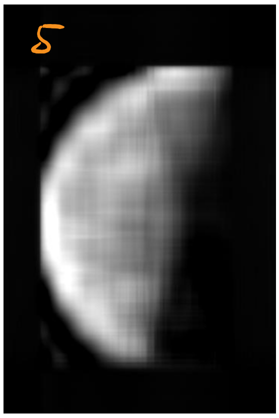
For calculations, we convert the number entries into the range $[0, 1]$ instead $[0, 255]$. Most pictures should be full-rank matrices and here we can calculate the rank and actually get $r = n$. Now let us write A in the representation given by equation (9.11). Now we stop the summation instead of after $r = 358$ steps at $k = 50$, 30, 10 or 5 terms and we get the following pictures:

43.8 KB

26.7 KB

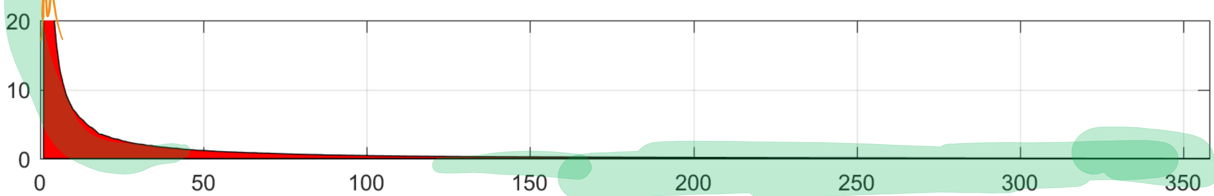


8.8 KB



4.4 KB

The decay of the singular values $s_1 \geq \cdots \geq s_{358}$ below shows us why we already have at only 30 terms in the sum a very good approximation.



The first singular values ($s_1 \approx 144$, $s_2 \approx 50$ and $s_3 \approx 35$) are not shown in the picture, for obvious reason.

In Example 9.12, we have seen that a given matrix A with rank $r = 358$ can be well approximated by matrices A_k with rank $k = 50$, 30, 10 or 5. For

$$A = \sum_{i=1}^{r} s_i \mathbf{u}_i \mathbf{v}_i^* \qquad \text{and} \quad k \in \{1, \ldots, r\} \quad \text{we set} \qquad A_k := \sum_{i=1}^{k} s_i \mathbf{u}_i \mathbf{v}_i^*.$$

A_k has rank k and is in fact the best $m \times n$-matrix with rank k for the approximation of

A. We measure the error of approximation by using the matrix norm in equation (9.10):

$$\begin{aligned}
\|A - A_k\| &= \left\| U \left(\left(\begin{array}{ccccc|c} s_1 & & & & & \\ & \ddots & & & & \\ & & s_k & & & \\ & & & \ddots & & \\ & & & & s_r & \\ \hline & & & & & \end{array}\right) - \left(\begin{array}{ccc|c} s_1 & & & \\ & \ddots & & \\ & & s_k & \\ \hline & & & \end{array}\right) \right) V^* \right\| \\
&= \left\| U \left(\begin{array}{ccc|c} s_{k+1} & & & \\ & \ddots & & \\ & & s_r & \\ \hline & & & \end{array}\right) V^* \right\| = s_{k+1} \quad \text{(largest singular value left).}
\end{aligned}$$

In short:

$$s_{k+1} \;=\; \textit{distance of } A \textit{ to the set of all matrices with rank } k \tag{9.12}$$

In particular, s_1 is the distance of A to the set of all matrices with rank 0, which consists only of the zero matrix 0.

At the end, let us take a look at the special case $m = n$, which means A and Σ are square matrices. In this case, eigenvalues and singular values are related in the following sense:

- A is invertible if and only if all the singular values are non-zero (see also Proposition 6.28 for the same claim with eigenvalues).

  The smallest singular value of A, s_n, gives the distance of A to the set of all $n \times n$-matrices with rank $n - 1$ or smaller (which are exactly the singular matrices) by equation (9.12).

  The equation $A^{-1} = (U\Sigma V^*)^{-1} = V\Sigma^{-1}U^*$ gives the SVD of A^{-1}. Therefore, the singular values of A^{-1} are $1/s_1, \ldots, 1/s_n$. The largest of these, meaning $1/s_n$, is $\|A^{-1}\|$.

- We know from Corollary 9.8 that the product of all eigenvalues of a given matrix A is exactly $\det(A)$. Since

  $$\det(A) = \det(U\Sigma V^*) = \det(U)\det(\Sigma)\det(V^*) \;\Rightarrow\; |\det(A)| = \det(\Sigma)\,,$$

  we know that the product of all singular values, which is $\det(\Sigma)$, is equal to the absolute value of $\det(A)$.

- If A is normal, which means $A^*A = AA^*$, then A can be diagonalised by using a unitary matrix: $A = XDX^*$. Then $D = \mathrm{diag}(d_1, \ldots, d_n)$ is a diagonal matrix with the eigenvalues of A as entries and $X = (\mathbf{x}_1 \cdots \mathbf{x}_n)$ consists of eigenvectors for A.

Hence $A^*A = XD^*DX^* = X\,\mathrm{diag}(|d_1|^2, \dots, |d_n|^2)\,X^*$. The eigenvalues λ_i of A^*A are, on the one hand, given by $\lambda_i = \overline{d_i}d_i = |d_i|^2$ and, on the other hand, they can be written as $\lambda_i = s_i^2$ by using the singular values $s_i \geq 0$ of A. Therefore, we get:

$$s_i = |d_i|.$$

The singular values of A are exactly the absolute values of the eigenvalues of A.

Summary

- A lot of techniques in Linear Algebra deal with suitable factorisations of a given matrix A:
- From Section 3.11.5: The Gaussian elimination are summarised by a left multiplication with a lower triangular matrix ◺ and a permutation matrix P. Hence, ◺PA is the row echelon form K of A and we have $PA = LK$ with lower triangular matrix $L :=$ ◺$^{-1}$.
- From Section 5.5: A linearly independent family of vectors $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ from $\mathbb{F}^m$ can be transformed into an ONS $(\mathbf{q}_1, \dots, \mathbf{q}_n)$ by using the Gram-Schmidt procedure. Therefore, we have for $k = 1, \dots, n$ always $\mathbf{a}_k \in \mathrm{Span}(\mathbf{q}_1, \dots, \mathbf{q}_k)$. For the matrices $A := (\mathbf{a}_1 \ \dots \ \mathbf{a}_n)$ and $Q := (\mathbf{q}_1 \ \dots \ \mathbf{q}_n)$ we find $A = QR$, where $R \in \mathbb{F}^{n\times n}$ is an invertible upper triangular matrix.
- If we decompose A into a product UDV, then we have different approaches.
- For diagonalisable matrices, we can choose $U = X$ and $V = X^{-1}$ where in X the columns are eigenvectors of A and form a basis. Then D has the eigenvalues of A on the diagonal, counted with multiplicities. See Chapter 6. We also know that selfadjoint and even normal matrices A are always diagonalisable, we can choose eigenvectors in such a way that they form an ONB, which means $X^* = X^{-1}$.
- For non-diagonalisable matrices we still can write $A = XDX^{-1}$ but now D is not diagonal. We use the Jordan normal form as a substitute. We get the important result that all (square) matrices $A \in \mathbb{C}^{n\times n}$ have such a Jordan normal form and therefore this decomposition. Note that we actually need the complex numbers here.
- For the singular value decomposition, the two matrices U and V are not connected such that we can also bring rectangular matrices A into "diagonal" structure. On the diagonal D (that is often denoted by Σ), we find the so-called *singular values* of A. The singular value decomposition is used for low rank approximation.