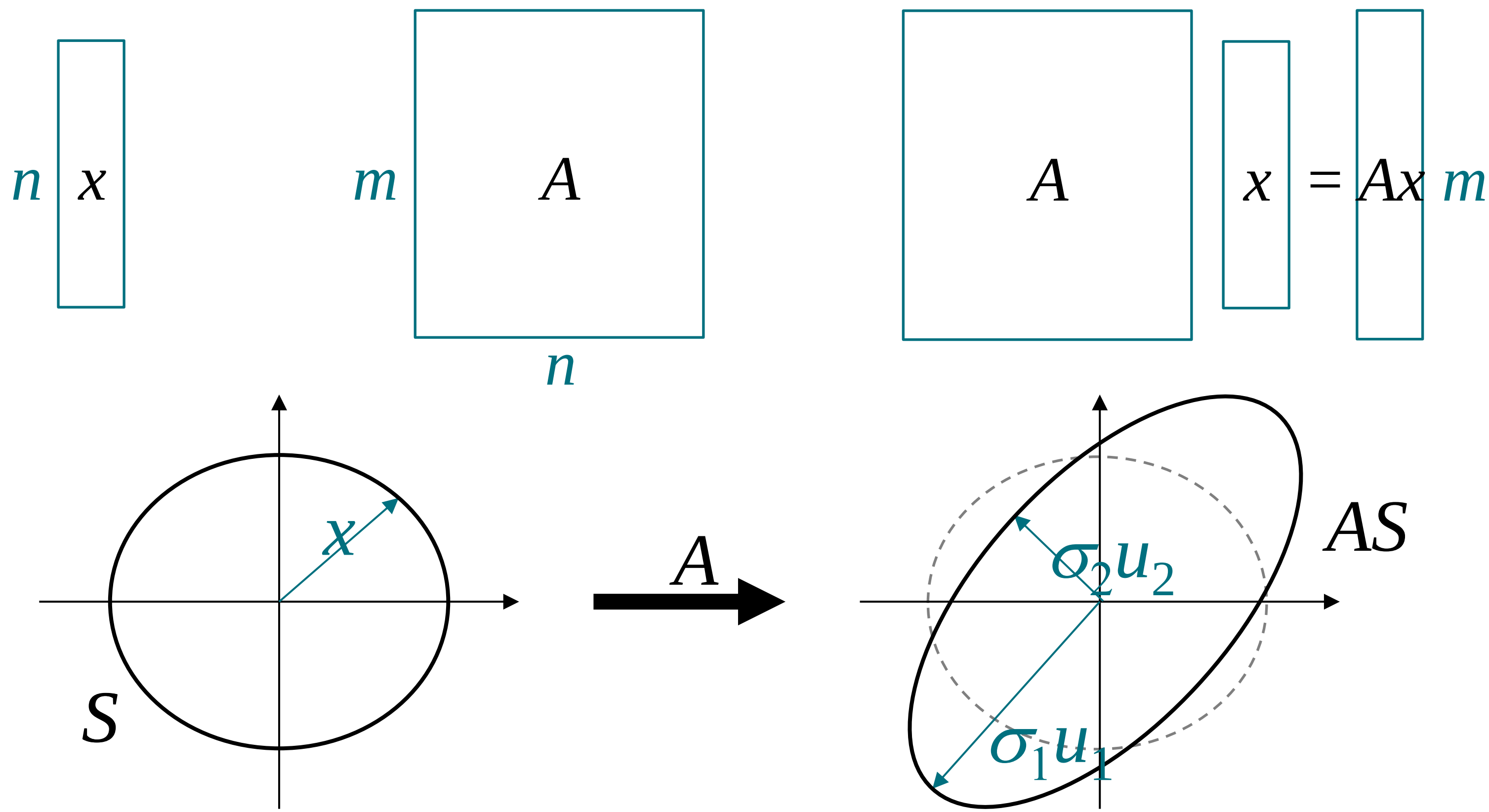
Lecture 12 More on the SVD

Mathématiques appliquées (MATH0504-1)
B. Dewals, Ch. Geuzaine

Reminder

The image of the unit sphere S under any $m \times n$ matrix is a hyperellipse.

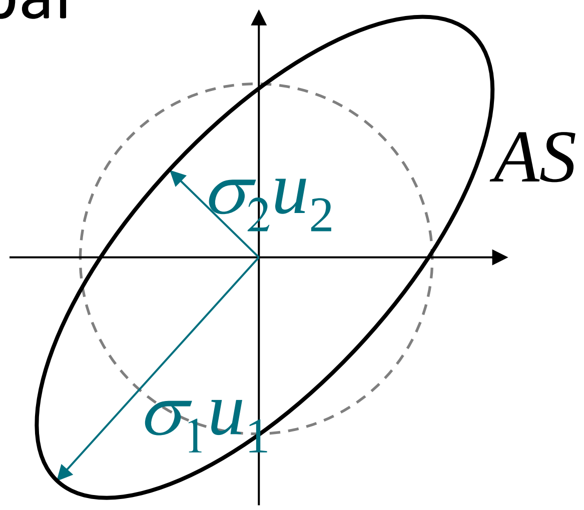


Reminder

In $\mathbb{R}^m$, a hyperellipse is a surface obtained by

- stretching the unit sphere in $\mathbb{R}^m$
- by some factors $\sigma_1, \ldots, \sigma_m$
- in some orthonormal directions $u_1, \ldots, u_m \in \mathbb{R}^m$

The vectors $\{\sigma_i u_i\}$ are the principal semiaxes of the hyperellipse:



$\sigma_1, \ldots, \sigma_n$ are the singular values

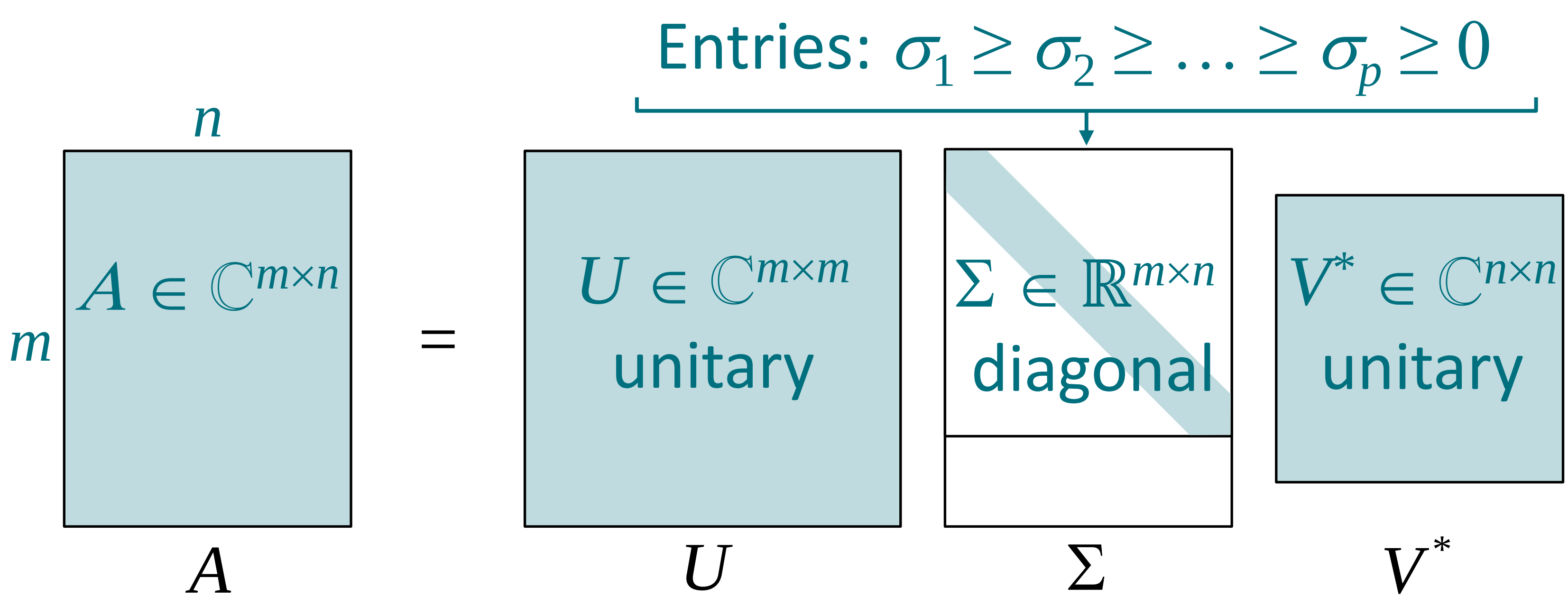
$u_1, \ldots, u_m$ are the (left) singular vectors

Reminder

Given $A \in \mathbb{C}^{m\times n}$, not necessarily of full rank, the SVD decomposition of A is a factorization

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

where

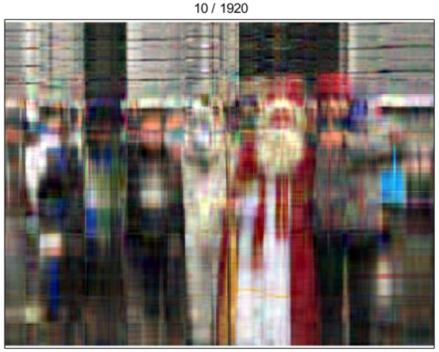


Learning objectives

Understand the SVD as a change of bases, and relate it to the eigenvalue decomposition

Study matrix properties via the SVD

Understand the principle of low-rank approximations



~ 1 %

~ 14 %

Full HD

1 – Change of bases

(Lecture 5 in Trefethen & Bau, 1997)

Change of bases

The SVD makes it possible to view any matrix A as a diagonal matrix… provided that we use proper bases for the domain and range spaces

Consider $b = Ax$

Let us

- expand b in the basis of the left singular vectors of A (the columns of U)
- Expand x in the basis of the right singular vectors of A (the columns of V)

Change of bases

In these new bases, we have

$$b = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} b' = Ub'$$

$$x = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} x' = Vx'$$

and thus $b' = U^* b$ and $x' = V^* x$.

Thus:

$$b = Ax \iff U^* b = U^* Ax = U^* U \Sigma V^* x = \Sigma V^* x$$

i.e. $b' = \Sigma x'$

Change of bases

Thus any matrix A reduces to the diagonal matrix Σ when

- the range is expressed in the basis of columns of U
- the domain is expressed in the basis of the columns of V

2 – SVD vs. eigenvalue decomposition

(Lecture 5 in Trefethen & Bau, 1997)

Eigenvalue decomposition

If a square matrix $A \in \mathbb{C}^{m \times m}$ **possesses m linearly independent eigenvectors**, the eigenvalue decomposition of A is

$$A = X \Lambda X^{-1}$$

where

- the columns of X are the eigenvectors
- Λ is an $m \times m$ diagonal matrix whose entries are the eigenvalues of A

Eigenvalue decomposition

If similarly as before we now expand b and x (in $b = Ax$) in the basis of the eigenvectors, then the new vectors

$$b' = X^{-1}b, \qquad x' = X^{-1}x$$

satisfy $b' = \Lambda x'$.

What are the differences with the SVD?

- SVD uses two bases instead of one (the eigenvectors)
- SVD uses orthonormal bases (while eigenvectors are in general not orthogonal)
- All matrices (even rectangular) have an SVD!

LIÈGE université Sciences Appliquées

3 – Matrix properties via the SVD

(Lecture 5 in Trefethen & Bau, 1997)

Rank

Let r be the number of nonzero singular values. Then $\operatorname{rank}(A) = r$.

Indeed:

- the rank of a diagonal matrix is the number of its nonzero entries;
- since U and V have full rank, we have

$$\operatorname{rank}(A) = \underbrace{\operatorname{rank}(U\Sigma V^*) = \operatorname{rank}(\Sigma)}_{\textit{This is shown in the next slide}} = r$$

Rank: demonstration

Reminder: $\operatorname{rank}(A\,B) \leq \min\left[\operatorname{rank}(A), \operatorname{rank}(B)\right]$

$$\operatorname{rank}\left(U\Sigma V^{*}\right) \leq \min\left[\operatorname{rank}(U), \operatorname{rank}(\Sigma), \operatorname{rank}\left(V^{*}\right)\right]$$

$$= \min\left[m, \operatorname{rank}(\Sigma), n\right] = \operatorname{rank}(\Sigma) \quad ①$$

$$\operatorname{rank}(\Sigma) = \operatorname{rank}(I\,\Sigma\,I) = \operatorname{rank}\left(U^{*}U\,\Sigma\,V^{*}V\right)$$

$$\leq \min\left[\operatorname{rank}\left(U^{*}\right), \operatorname{rank}\left(U\,\Sigma\,V^{*}\right), \operatorname{rank}(V)\right]$$

$$= \min\left[m, \operatorname{rank}\left(U\,\Sigma\,V^{*}\right), n\right] = \operatorname{rank}\left(U\,\Sigma\,V^{*}\right) \quad ②$$

① ② $\rightarrow \operatorname{rank}\left(U\,\Sigma\,V^{*}\right) = \operatorname{rank}(\Sigma)$

Norm

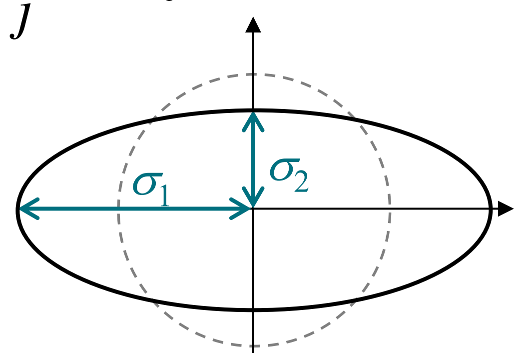
$$\|A\|_{(m,n)} = \max_{\substack{x\in\mathbb{C}^n \\ \|x\|_{(n)}\neq 0}} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \max_{\substack{x\in\mathbb{C}^n \\ \|x\|_{(n)}=1}} \|Ax\|_{(m)}$$

The 2-norm of the matrix is equal to the first (largest) singular value: $||A||_2 = \sigma_1$

We established this in the existence proof (from the geometric interpretation of SVD)!

More quickly, since $A = U\Sigma V^*$ with unitary U and V, we have

$$||A||_2 = \underbrace{||U\Sigma V^*||_2 = ||\Sigma||_2}_{\textit{This is shown in the next four slides}} = \max_j |\sigma_j| = \sigma_1$$



Norm: demonstration

Consider a unitary matrix U.

We proceed in four steps.

Step 1

First, let's show that $\|Ux\|_2 = \|x\|_2$:

$$\|Ux\|_2^2 = (Ux)^*(Ux) = x^*\underbrace{U^*U}_{I}x = x^*x = \|x\|_2^2$$

Norm: demonstration (cont'd)

Reminder: $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$

Step 2 We demonstrate that $\|UM\|_2 \leq \|M\|_2$

a) Consider a vector x with $\|x\|_2 = 1$, such that:

$$\|UMx\|_2 = \|UM\|_2$$

b) Let's evaluate the norm of UM:

$$\|UM\|_2 = \|UMx\|_2 = \|U(Mx)\|_2 = \|Mx\|_2$$

$$\leq \|M\|_2 \|x\|_2 = \|M\|_2$$

$\|AB\|_2 \leq \|A\|_2 \|B\|_2$

Step 1

Norm: demonstration (cont'd)

Reminder: $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$

Step 3 We demonstrate that $\|M\|_2 \leq \|UM\|_2$

a) Consider a vector y with $\|y\|_2 = 1$, such that:

$$\|My\|_2 = \|M\|_2$$

b) Let's evaluate the norm of M:

Step 1

$$\|M\|_2 = \|My\|_2 = \|U(My)\|_2 = \|(UM)y\|_2$$

$$\leq \|UM\|_2 \|y\|_2 = \|UM\|_2$$

$\|AB\|_2 \leq \|A\|_2 \|B\|_2$

Norm: demonstration (cont'd)

Step 4

From Step 2, we have:

$$\|UM\|_2 \leq \|M\|_2$$

From Step 3, we have:

$$\|M\|_2 \leq \|UM\|_2$$

Hence,

$$\|UM\|_2 = \|M\|_2$$

Eigenvalues

The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*

Proof: from

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

we see that A^*A and $\Sigma^*\Sigma$ are similar.

Hence, they have the same eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ with $n - p$ additional zero eigenvalues if $n > p$.

Note that the right (resp. left) singular vectors are eigenvectors of A^*A (resp. AA^*).

Two matrices A and B are *similar* if, for some invertible matrix P, we have $B = P^{-1} A P$

Similar matrices have the same eigenvalues.

Indeed, if v is an eigenvector of A with eigenvalue λ, $P^{-1} v$ is an eigenvector of B with the same eigenvalue λ:

$$A v = \lambda v$$

$$PBP^{-1} v = \lambda v$$

$$B P^{-1} v = \lambda P^{-1} v$$

So, every eigenvalue of A is an eigenvalue of B, and conversely since one can interchange A and B.

Eigenvalues

If $A^* = A$, then the singular values of A are the absolute values of the eigenvalues of A.

Proof *Reminder:* a Hermitian matrix has

- a full set of orthogonal eigenvectors,
- and all its eigenvalues are real,

so that its eigenvalue decomposition can be written $A = Q \Lambda Q^*$, with Q unitary and Λ diagonal and real.

Let $|\Lambda|$ and $\mathrm{sign}(\Lambda)$ denote the diagonal matrices with entries $|\lambda_i|$ and $\mathrm{sign}(\lambda_i)$, respectively.

Eigenvalues

We can then write

$$A = Q\Lambda Q^* = Q|\Lambda|\mathrm{sign}(\Lambda)Q^* = Q|\Lambda|W^*$$

since $\mathrm{sign}(\Lambda)Q^*$ is unitary if Q^* is unitary.

Inserting permutation matrices (i.e., square matrices that have exactly one entry of 1 in each row and each column and 0s elsewhere) as factors of Q and W^* to reorder the entries of $|\Lambda|$ in non-increasing order, this is an SVD of A, with the singular values equal to the diagonal entries of $|\Lambda|$, i.e., the absolute values of the eigenvalues.

LIÈGE université Sciences Appliquées

4 – Low-rank approximations

(Lecture 5 in Trefethen & Bau, 1997)

Sum of rank-one matrices

Thanks to the SVD we can express A as the sum of r rank-one matrices:

$r = \text{rank}(A) = \text{rank}(\Sigma)$

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*$$

Proof: write Σ as a sum of r matrices Σ_j

$$A = U\Sigma V^* = U\left(\sum_{j=1}^{r} \Sigma_j\right)V^* = \sum_{j=1}^{r}\left(U\Sigma_j V^*\right) = \sum_{j=1}^{r}\left(u_j \sigma_j v_j^*\right)$$

with $\Sigma_j = \text{diag}\left(0,\ldots,0,\sigma_j,0,\ldots,0\right)$

Sum of rank-one matrices

Thanks to the SVD we can express A as the sum of r rank-one matrices:

$r = \mathrm{rank}(A) = \mathrm{rank}(\Sigma)$

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*$$

There are many other ways to express a matrix as a sum of rank-one matrices (e.g. simply as the sum of its m rows, or its n columns, etc.).

But using the SVD leads to an **approximation** with a remarkable property: the k-th *partial* sum captures as much "energy" of A as possible.

Low-rank approximation

For any $k \leq r$, define the partial sum

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^*$$

Then $||A - A_k||_2 = \sigma_{k+1} = \min_{\text{rank}(B)=k} ||A - B||_2$

This tells us that the “best” rank-k approximation of a matrix is obtained by the k-th partial sum A_k !

This has numerous applications, from image compression to the approximation of PDEs.

Low-rank approximation: proof

① Since $A_k = U\text{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)V^*$, $\text{rank}(A_k) = k$ and we have $A - A_k = U\text{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_r)V^*$.

Thus $||A - A_k||_2 = \sigma_{k+1}$.

② a Suppose that there exists $B \in \mathbb{C}^{m\times n}$, such that $\text{rank}(B) = k$ and $||A - B||_2 < \sigma_{k+1}$.

Then we can find orthonormal vectors $w_1, \ldots, w_{n-k}$ in $\mathbb{C}^n$ such that $\text{null}(B) = \langle w_1, \ldots, w_{n-k}\rangle$.

For all $w \in \langle w_1, \ldots, w_{n-k}\rangle$ we then have $Bw = 0$, $(A - B)w = Aw$, and

$$||Aw||_2 = ||(A-B)w||_2 \leq ||A-B||_2||w||_2 < \sigma_{k+1}||w||_2$$

Low-rank approximation: proof

② b However there is another subspace of $\mathbb{C}^n$ for which $||Aw||_2 \geq \sigma_{k+1} ||w||_2$: the one spanned by the $k+1$ first right singular vectors: $\langle v_1, \ldots, v_{k+1} \rangle$.

Indeed: for $w \in \langle v_1, \ldots, v_{k+1} \rangle$, i.e. $w = \sum_{i=1}^{k+1} c_i v_i$ and $Aw = \sum_{i=1}^{k+1} \sigma_i c_i u_i$, we have

$$||w||_2^2 = \sum_{i=1}^{k+1} c_i^2$$

$$||Aw||_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 c_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} c_i^2$$

Low-rank approximation: proof

②c We thus have:

$$||Aw||_2 < \sigma_{k+1}||w||_2, \ \forall w \in \langle w_1, \ldots, w_{n-k} \rangle$$

$$||Aw||_2 \geq \sigma_{k+1}||w||_2, \ \forall w \in \langle v_1, \ldots, v_{k+1} \rangle$$

But these two subspaces of $\mathbb{C}^n$ must have a nonzero intersection, as the sum of their dimensions is: $(n\text{-}k) + (k{+}1) = n{+}1 > n$.

This a contradiction and, therefore, there cannot exist a matrix B such that $||A-B||_2 < \sigma_{k+1}$.

The best low-rank approximation in 2-norm is thus

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^* \quad !$$

Take-home messages

Every matrix is diagonal if the range is expressed in the basis of columns of U and the domain is expressed in the basis of the columns of V

Compared to the eigenvalue decomposition:

- two orthonormal bases (instead of one: the eigenvectors)
- applicable to non-square matrices
- non-zero singular values of A are the square roots of the eigenvalues of A^*A
- if $A^* = A$, the singular values of A are the absolute values of the eigenvalues of A

The SVD allows finding the best low-rank approximation of A.

Thank you!