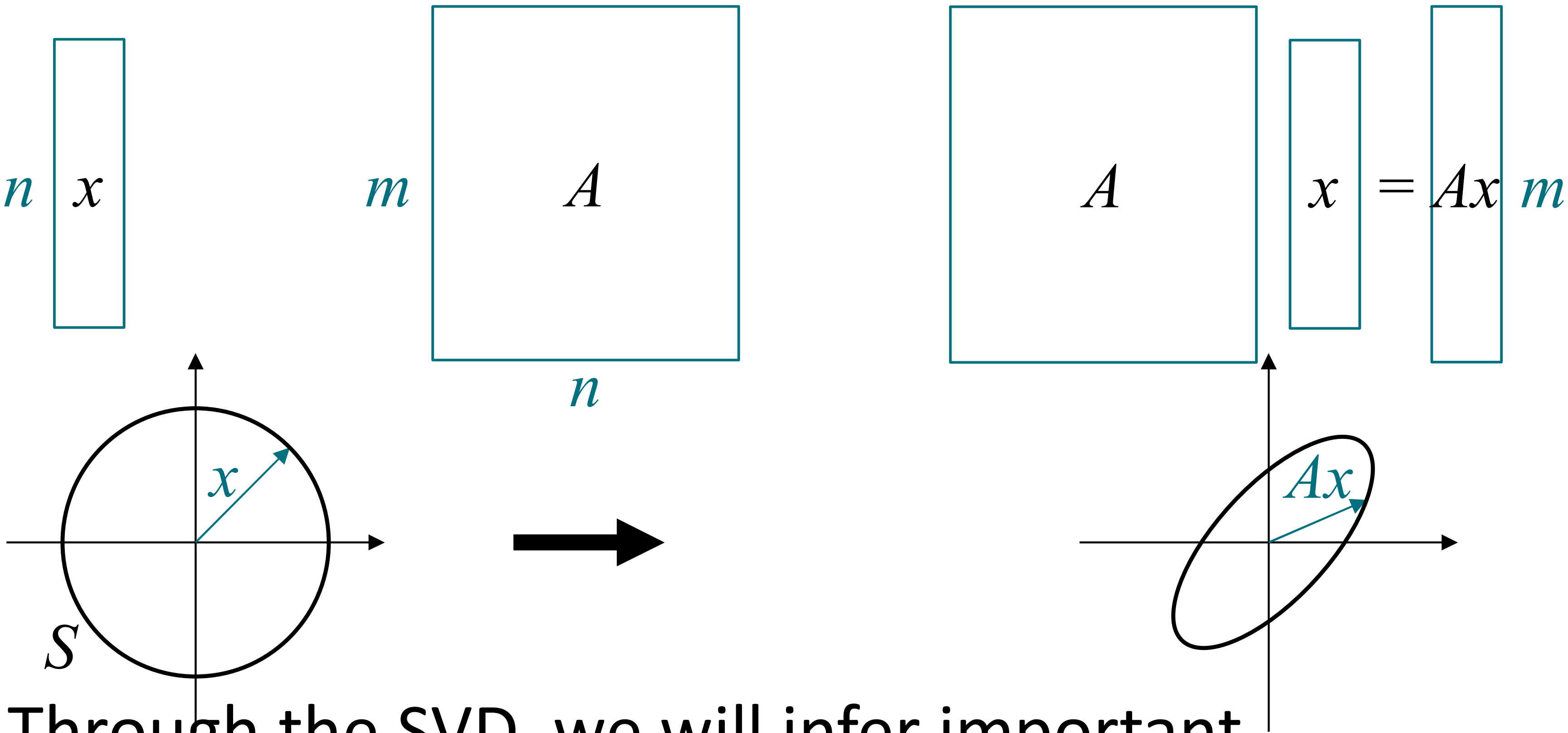
Lecture 11 Singular value decomposition

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

Singular value decomposition (SVD) at a glance …

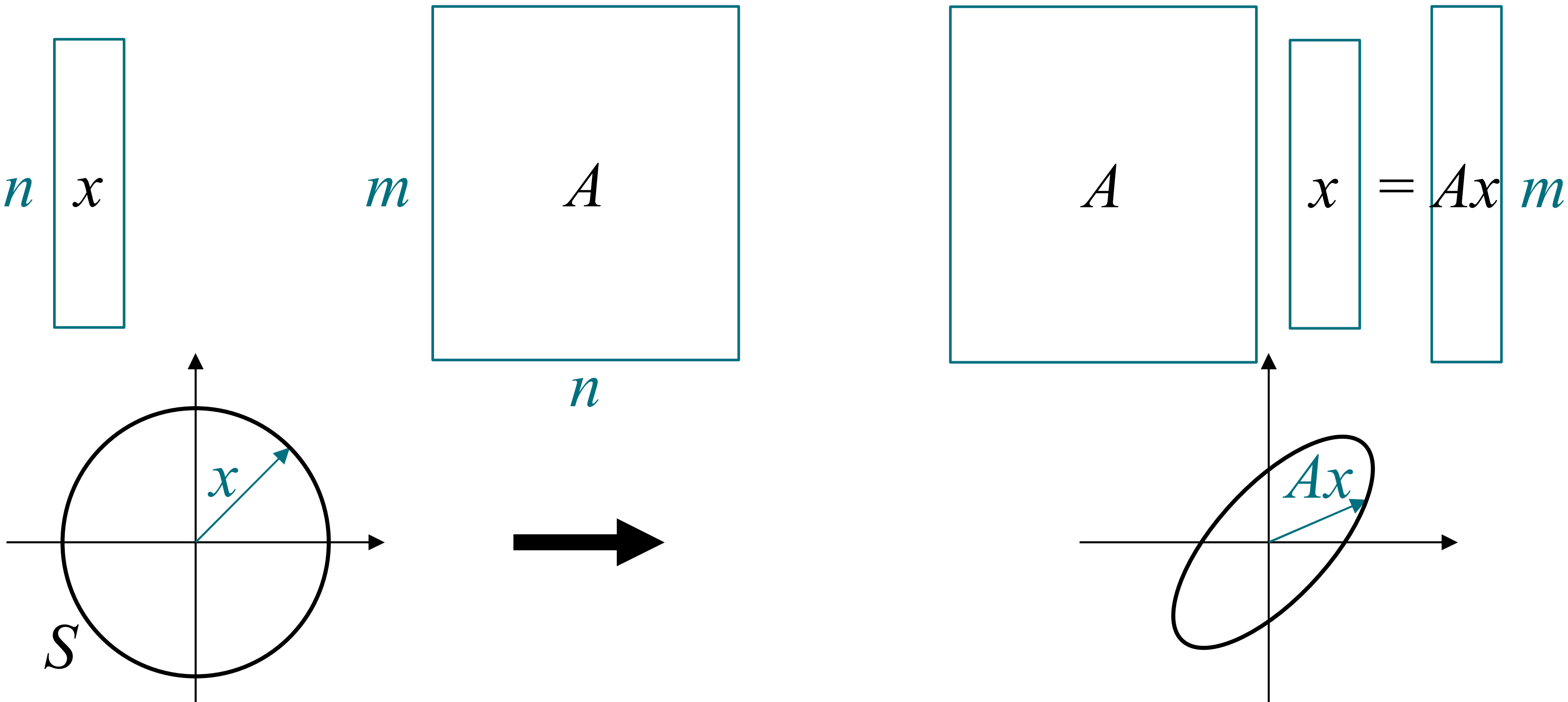
Motivation: the image of the unit sphere S under any $m \times n$ matrix transformation is a hyperellipse.



Through the SVD, we will infer important properties of matrix A from the shape of AS!

Singular value decomposition (SVD) at a glance …

The singular value decomposition (SVD)
is a particular **matrix factorization**.



Through the SVD, we will infer important
properties of matrix A from the shape of AS!

Why is the singular value decomposition of particular importance?

The reasons for looking at SVD are twofold:

1. The computation of the SVD is used as an intermediate step in many algorithms of practical interest.

2. From a conceptual point of view, the SVD enables a deeper understanding of many aspects of linear algebra.

Learning objectives & outline

Become familiar with the SVD and its geometric interpretation, and get aware of its significance

1. Reminder of some fundamentals in linear algebra
2. Geometric interpretation
3. From “reduced SVD” to “full SVD”, and formal definition
4. Existence and uniqueness

1 - Reminder: fundamentals in linear algebra

In this section, we briefly review the concepts of adjoint matrix, matrix rank, unitary matrix as well as matrix norms (Chapters 2 and 3 in Trefethen & Bau, 1997).

Adjoint of a matrix

The *adjoint* (or Hermitian conjugate) of an $m \times n$ matrix A, written A^*, is the $n \times m$ matrix

- whose i, j entry
- is the complex conjugate of the j, i entry of A.

If $A = A^*$, A is *Hermitian* (or *self-adjoint*).

For a real matrix A,

- the adjoint is the *transpose*: $A^* = A^T$,
- if the matrix is Hermitian, that is $A = A^T$, then it is *symmetric*.

Matrix rank

The *rank* of a matrix is the number of linearly independent columns (or rows) of a matrix.

The numbers of linearly independent columns and rows of a matrix are equal.

An $m \times n$ matrix of *full rank* is one that has the maximal possible rank (the lesser of m and n).

If $m \geq n$, such a matrix is characterized by the property that it maps no two distinct vectors to the same vector.

Matrix rank

Image of the unit sphere S by a *full-rank matrix*: no distinct vectors are mapped to the same vector.

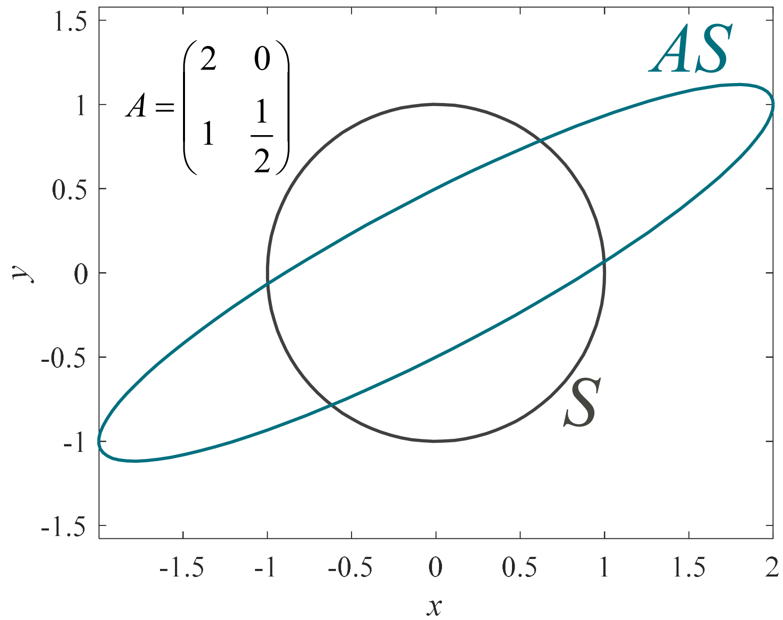
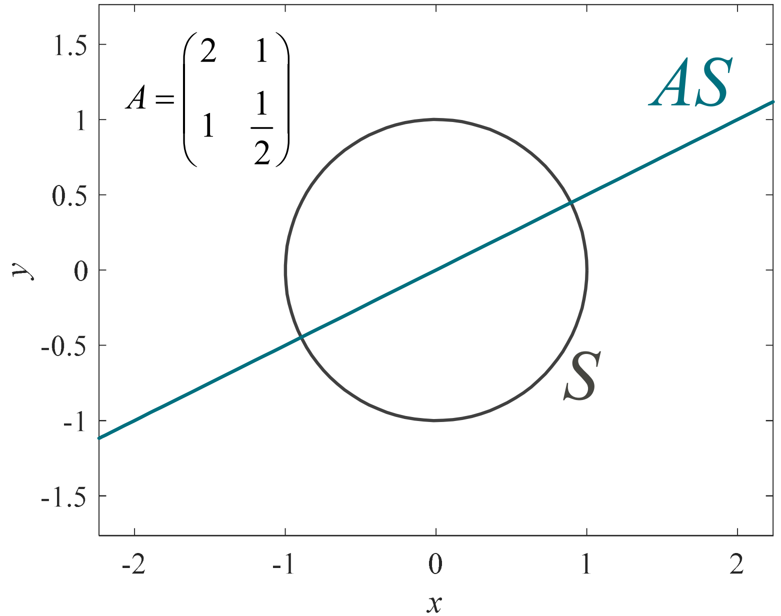


Image of the unit sphere S by a *rank-deficient* matrix: distinct vectors are mapped to the same vector.



Unitary matrix

A square matrix $Q \in \mathbb{C}^{m \times m}$, is unitary (or *orthogonal*, in the real case), if

$$Q^* = Q^{-1},$$

i.e.

$$Q^* Q = I.$$

The columns q_i of a unitary matrix form an othonormal basis of $\mathbb{C}^m$: $(q_i)^* q_j = \delta_{ij}$, with δ_{ij} the Kronecker delta.

A rotation matrix is a typical example of a unitary matrix

A rotation matrix R may write:

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

The image of a vector is the same vector, rotated counter clockwise by an angle θ.

Matrix R

- is orthogonal
- and $R^* R = R^T R = I$.

(Induced) matrix norms are defined from the action of the matrix on vectors

For a matrix $A \in \mathbb{C}^{m\times n}$, and given vector norms

- $\|\cdot\|_{(n)}$ on the domain of A
- $\|\cdot\|_{(m)}$ on the range of A

the *induced matrix norm* $\|\cdot\|_{(n)}$ is the smallest number C for which the following inequality holds for all $x \in \mathbb{C}^n$:

$$\|Ax\|_{(m)} \leq C\|x\|_{(n)}$$

It is the maximum factor by which A can “stretch” a vector x.

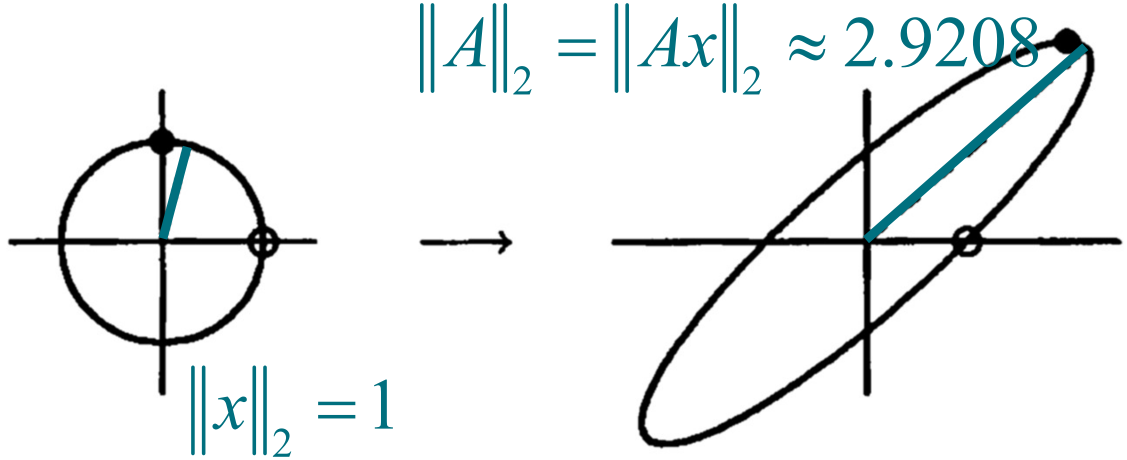
(Induced) matrix norms are defined from the action of the matrix on vectors

The matrix norm can be defined equivalently in terms of the images of the unit vectors under A:

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

This form is convenient for visualizing induced matrix norms, as in this example.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$



2 – Geometric interpretation

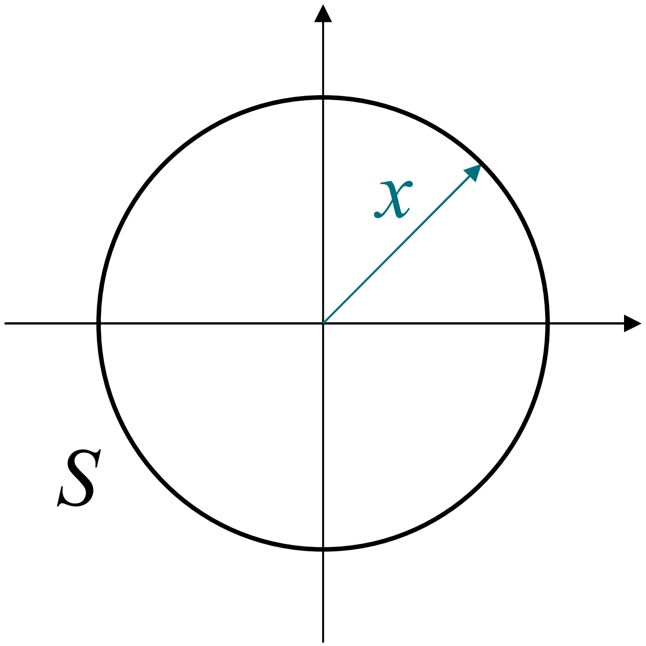
In this section, we introduce conceptually the SVD, by means of a simple geometric interpretation (Chapter 4 in Trefethen & Bau, 1997).

Geometric interpretation

Let S be the unit sphere in $\mathbb{R}^n$.

Consider any matrix $A \in \mathbb{R}^{m\times n}$, with $m \geq n$.

Assume for the moment that A has full rank n.



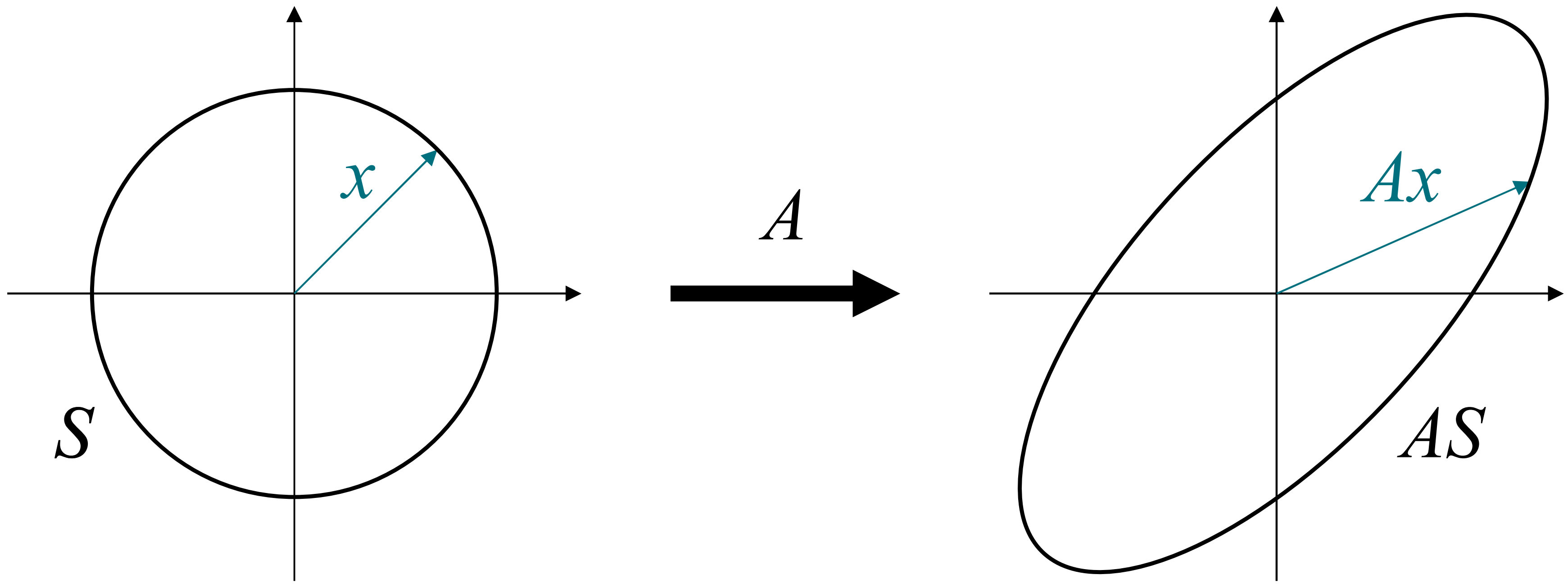
$$x \in \mathbb{R}^n : \|x\| = 1$$

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

Geometric interpretation

The image AS is a *hyperellipse* in $\mathbb{R}^m$.

This fact is not obvious; but let us assume for now that it is true. It will be proved later.



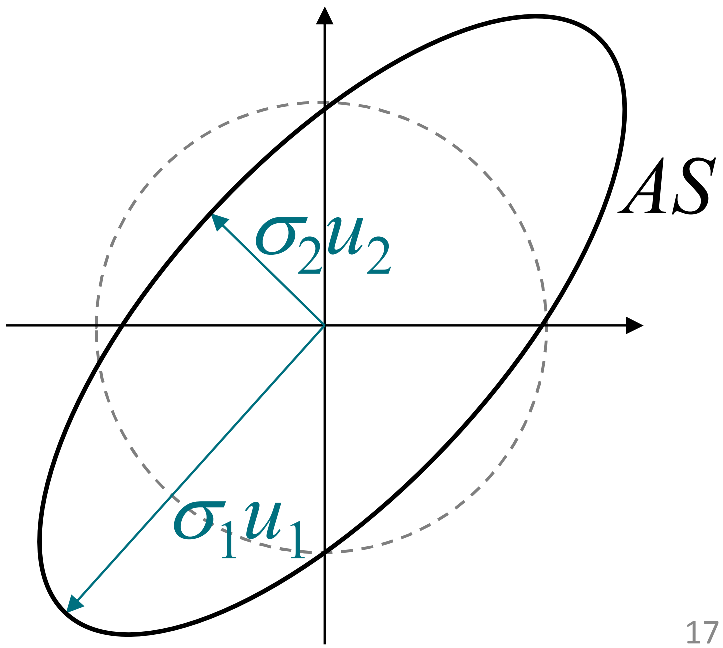
A "hyperellipse" is the m-dimensional generalization of an ellipse in 2D

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

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

For convenience, let us take the u_i to be unit vectors, i.e. $\|u_i\|_2 = 1$.

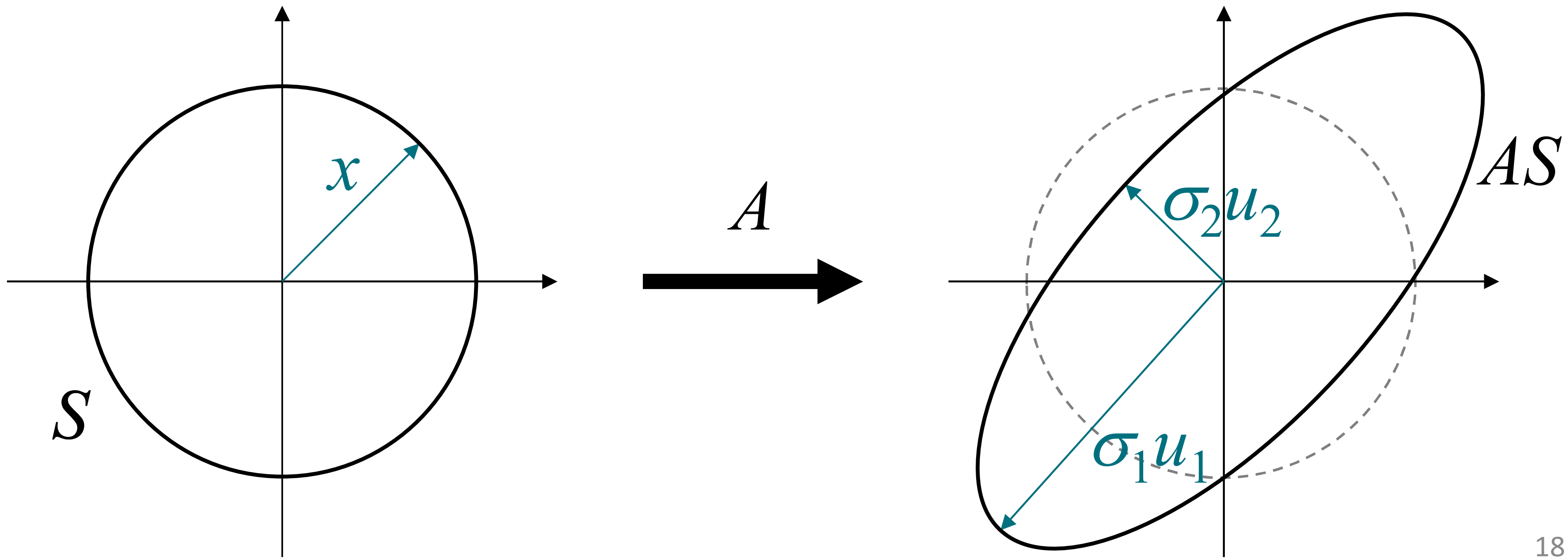
The vectors $\{\sigma_i u_i\}$ are the *principal semiaxes* of the hyperellipse.



A “hyperellipse” is the m-dimensional generalization of an ellipse in 2D

If A has rank r,
exactly r of the lengths σ_i will be nonzero.

In particular, if $m \geq n$,
at most n of them will be nonzero.



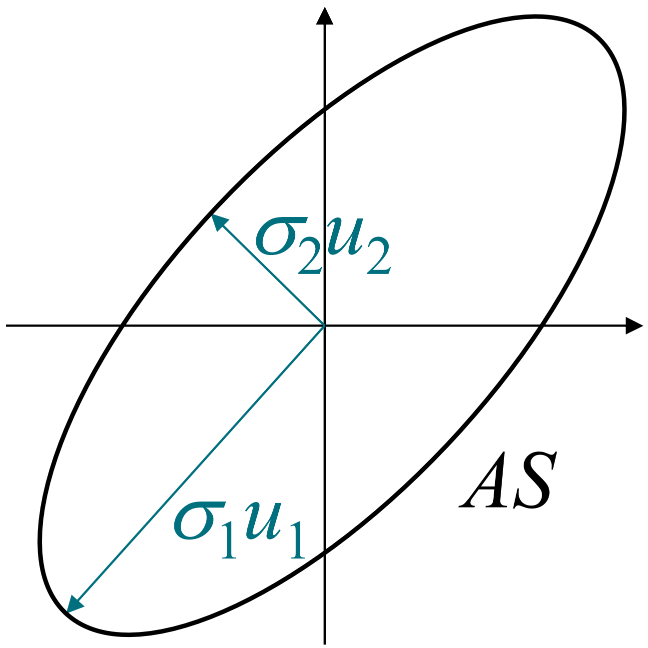
Singular values

We stated at the beginning that the SVD enables characterizing properties of matrix A from the shape of AS. Here we go for three definitions …

We **define** the n ***singular values*** of matrix A as the lengths of the n principal semiaxes of AS, noted $\sigma_1, \ldots, \sigma_n$.

It is conventional to number the singular values in descending order:

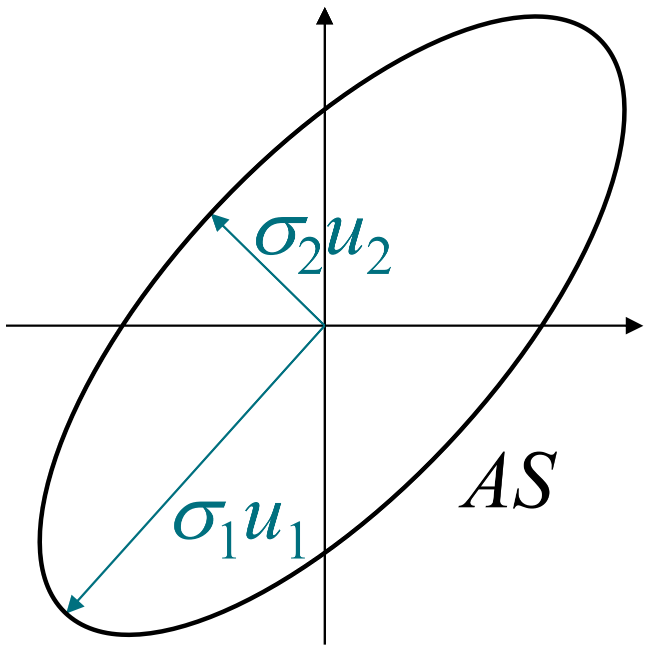
$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$.



Left singular vectors

We also **define** the n ***left singular vectors*** of matrix A as

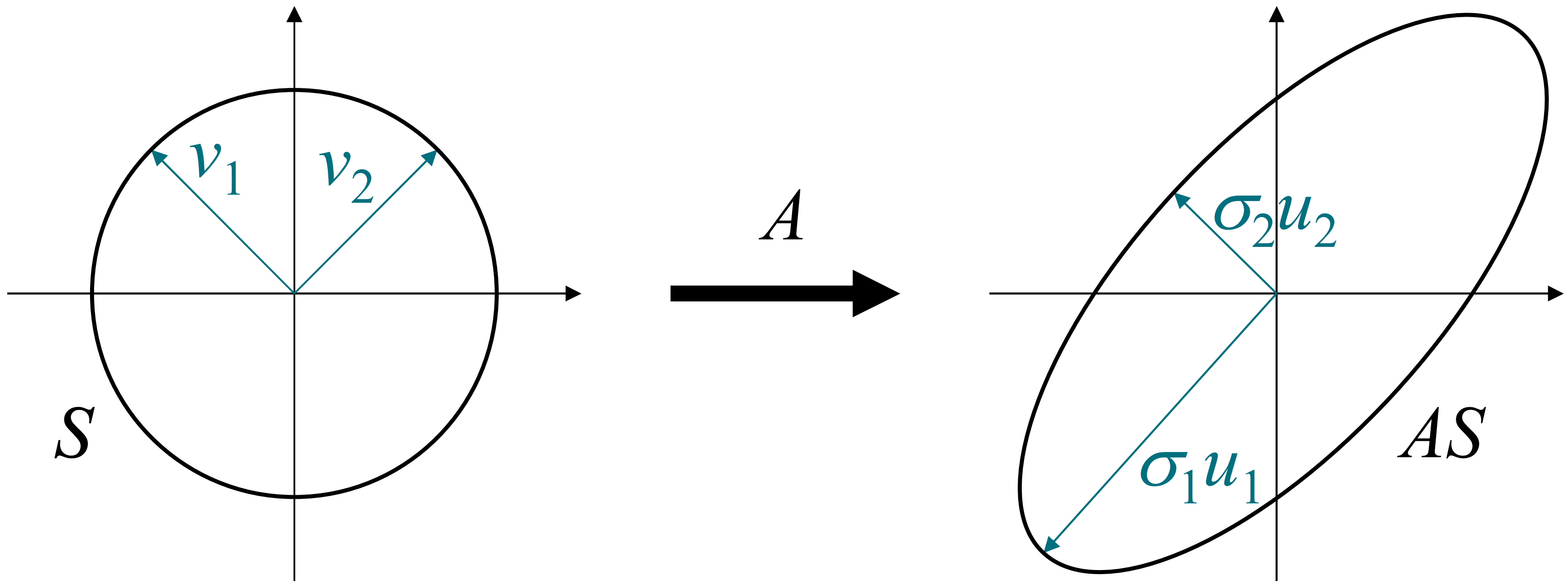
- the unit vectors $\{u_1, \ldots, u_n\}$
- oriented in the directions of the principal semiaxes of AS,
- numbered to correspond with the singular values.



Thus, the vector $\sigma_i\, u_i$ is the i^{th} largest principal semiaxis.

Right singular vectors

We also **define** the n ***right singular vectors*** of matrix A as the unit vectors $\{v_1, \ldots, v_n\} \in S$ that are the preimages of the principal semiaxes of AS, numbered so that $A\,v_j = \sigma_j\,u_j$.



Important remarks

The terms "left" and "right" singular vectors will be understood later as we move forward with a more formal description of the SVD.

In the geometric interpretation presented so far, we assumed that matrix A is real and $m = n = 2$.

Actually, the SVD applies

- to both real and complex matrices,
- whatever the number of dimensions.

3 – From reduced to full SVD, and formal definition

In this section, we distinguish between the so-called “reduced SVD”, often used in practice, and the “full SVD”. We also introduce the formal definition of SVD (Chapter 4 in Trefethen & Bau, 1997).

The equations relating right and left singular vectors can be expressed in matrix form

We just mentioned that the equations relating right singular vectors $\{v_j\}$ and left singular vectors $\{u_j\}$ can be written

$$A\, v_j = \sigma_j\, u_j \qquad\qquad 1 \le j \le n$$

This collection of vector equations can be expressed as a matrix equation.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} v_1\, v_2 \dots v_n \end{bmatrix} = \begin{bmatrix} u_1\, u_2 \dots u_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

The equations relating right and left singular vectors can be expressed in matrix form

This matrix equation can be written in a more compact form:

to distinguish from U, Σ in the "full SVD"

$$AV = \hat{U}\hat{\Sigma}$$

with

- $\hat{\Sigma}$ an $n \times n$ diagonal matrix with real entries (as A was assumed to have full rank n)
- $\hat{U}$ an $m \times n$ matrix with orthonormal columns
- V an $n \times n$ matrix with orthonormal columns

Thus, V is unitary (i.e. $V^* = V^{-1}$), and we obtain:

$$A = \hat{U}\hat{\Sigma}V^*$$

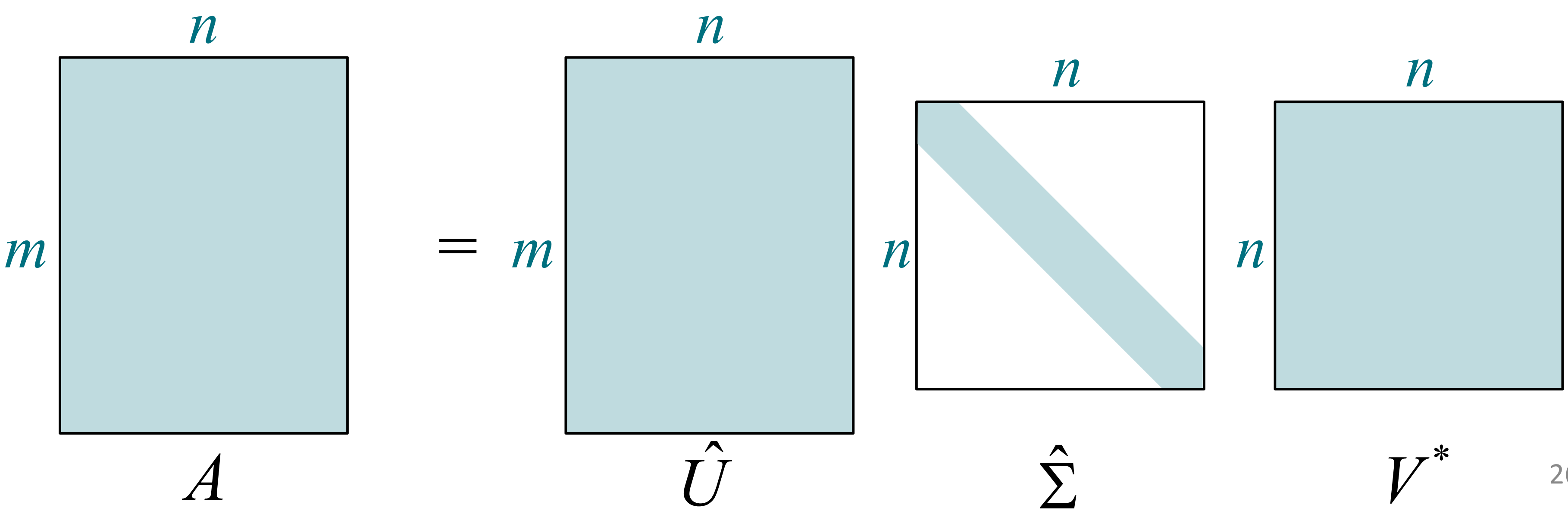
Reduced SVD

The factorization of matrix A in the form

$$A = \hat{U}\hat{\Sigma}V^*$$

is called a ***reduced singular values decomposition***, or reduced SVD, of matrix A.

Schematically, it looks like this ($m \geq n$):

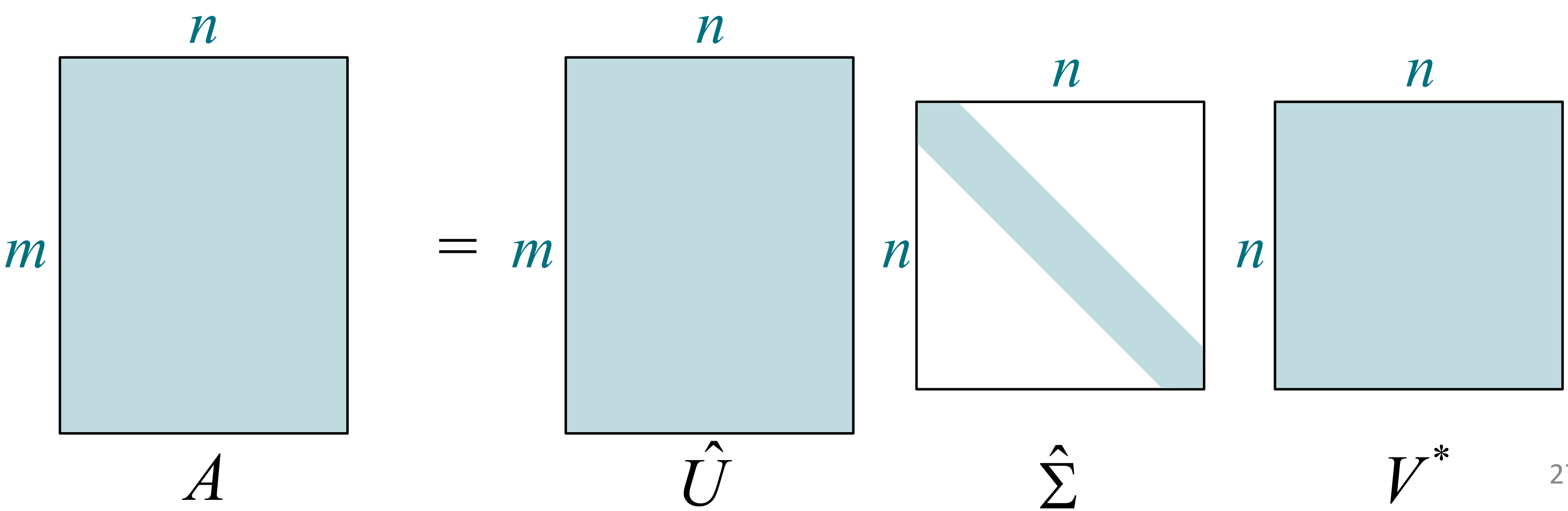


From reduced SVD to … full SVD

The columns of $\hat{U}$ are n orthonormal vectors in the m-dimensional space $\mathbb{C}^m$.

Unless $m = n$, they do not form a basis of $\mathbb{C}^m$, nor is $\hat{U}$ a unitary matrix.

However, we may "upgrade" $\hat{U}$ to a unitary matrix!

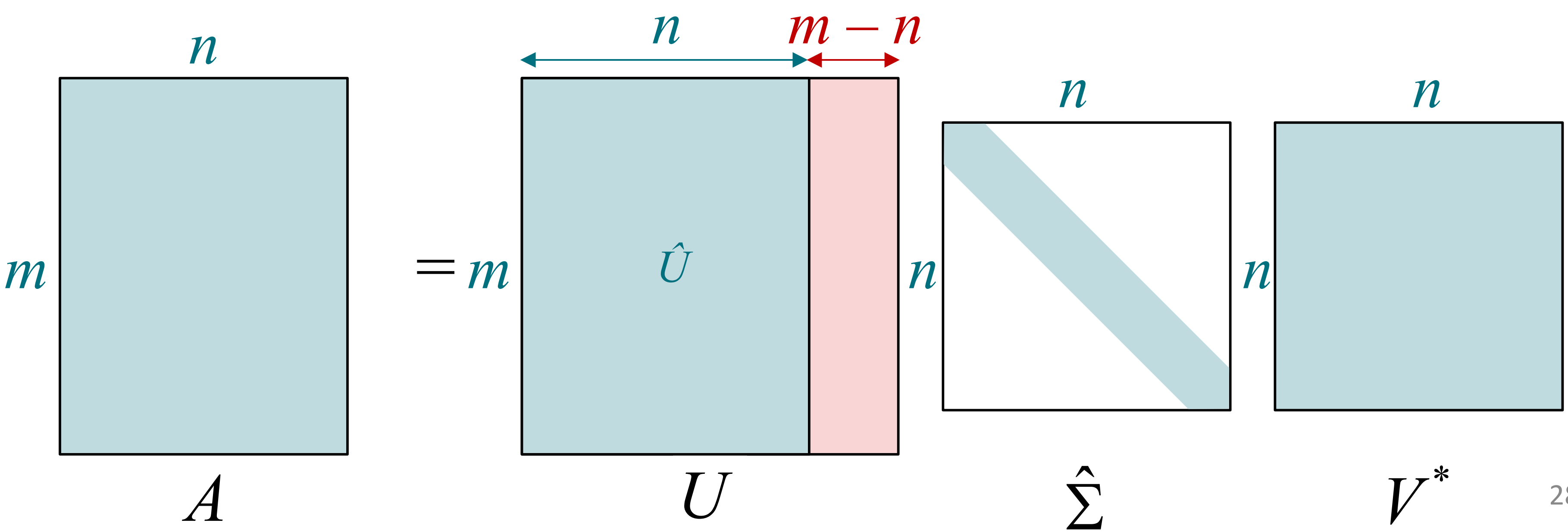


From reduced SVD to … full SVD

Let us adjoin an additional $m - n$ orthonormal columns to matrix $\hat{U}$, so that it becomes unitary.

The $m - n$ additional orthonormal columns are chosen arbitrarily and the result is noted U.

However, Σ must change too …

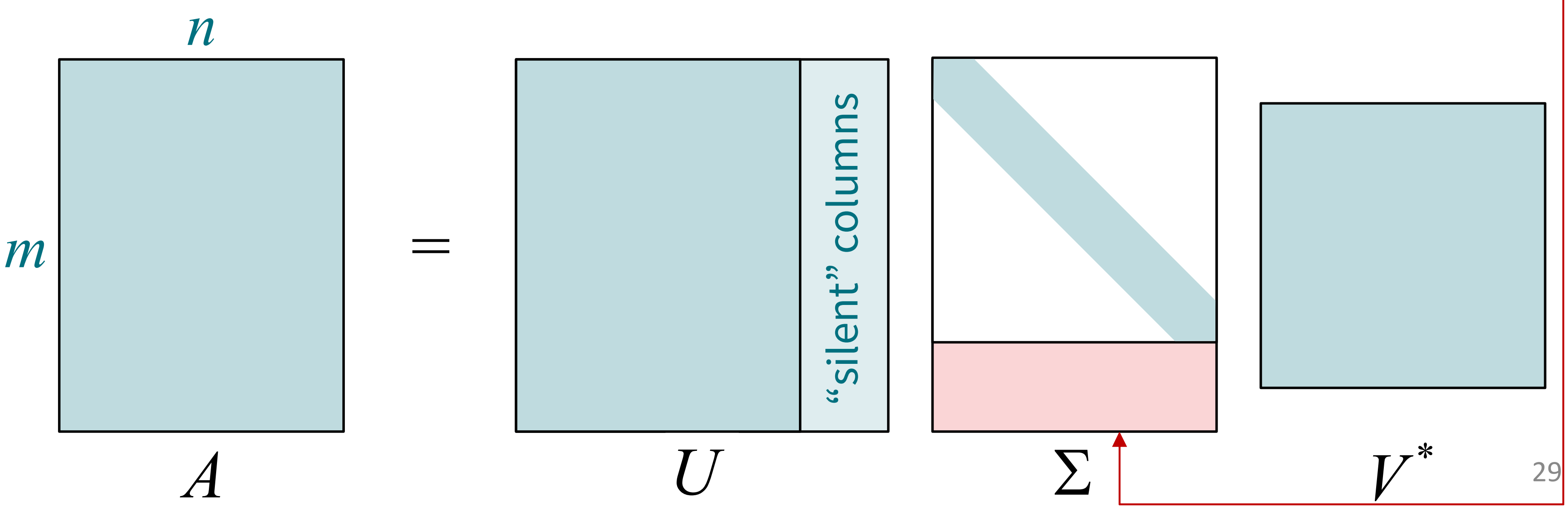


From reduced SVD to … full SVD

For the product to remain unaltered, the last $m - n$ columns of U should be multiplied by zero.

Accordingly, let Σ be the $m \times n$ matrix consisting of

- $\hat{\Sigma}$ in the upper $n \times n$ block
- together with $m - n$ rows of zeros below.

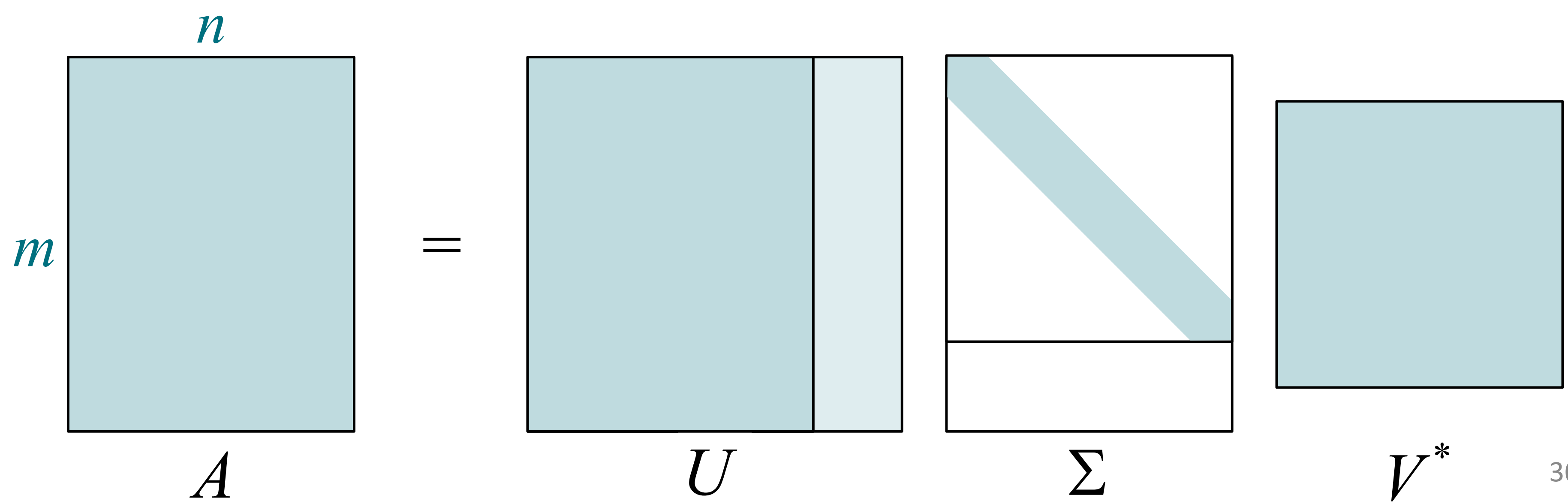


From reduced SVD to … full SVD

We get a new factorization of A, called ***full SVD***:

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

- U is an $m \times m$ unitary matrix,
- V is an $n \times n$ unitary matrix,
- Σ is an $m \times n$ diagonal matrix with real entries



Generalization to the case of a matrix A which does not have full rank

If matrix A is rank-deficient (i.e. of rank $r < n$), only r (instead of n) of the left singular vectors are deduced from the size of the hyperellipse

BUT the full SVD still applies,

- by introducing $m - r$ (instead of $m - n$) additional arbitrary orthonormal columns to construct the unitary matrix U;
- the matrix V also needs $n - r$ arbitrary orthonormal columns to extend the r columns determined from the hyperellipse geometry
- matrix Σ has only r non-zero diagonal entries.

Formal definition of the SVD

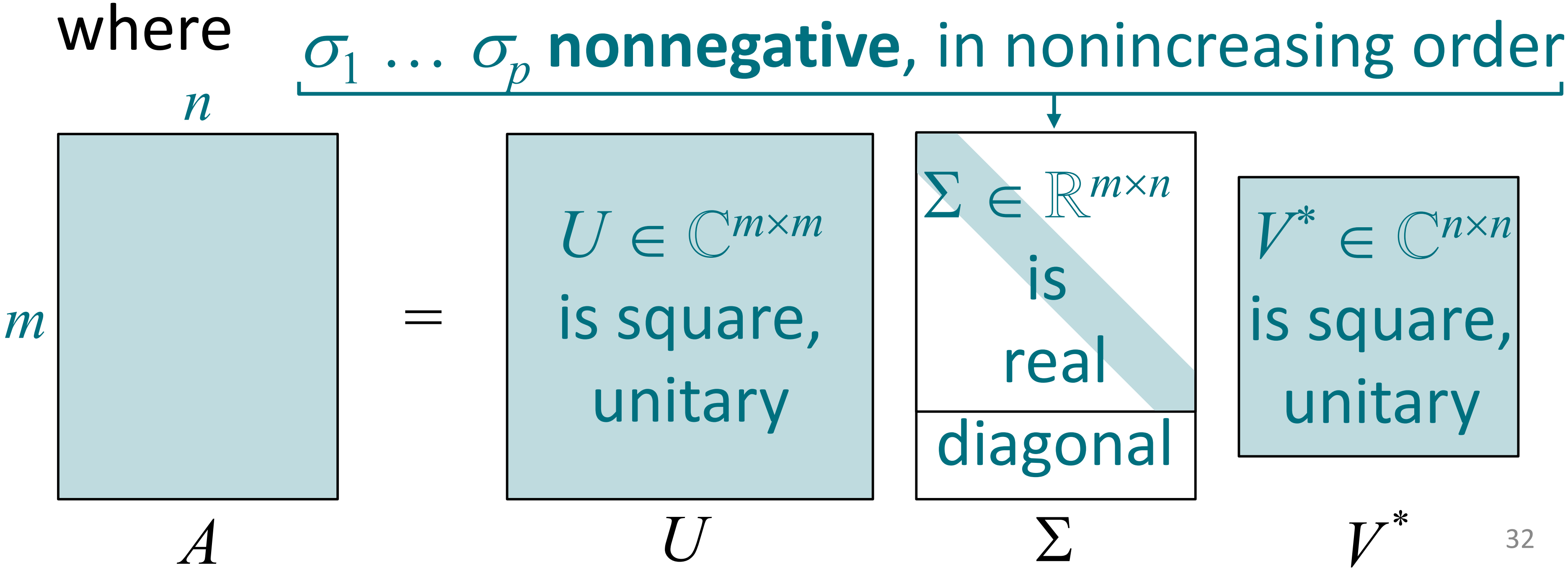
Let m and n be arbitrary (we do not require $m \geq n$).

Given $A \in \mathbb{C}^{m \times n}$, not necessarily of full rank,

a singular value decomposition of A is a factorization

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

where



Consequently, the image of the unit sphere in $\mathbb{R}^n$ under a map $A = U \Sigma V^*$ is a hyperellipse in $\mathbb{R}^m$

1. The unitary map V^* preserves the sphere
2. The diagonal matrix Σ stretches the sphere into a hyperellipse
3. The final unitary map U rotates, or reflects, the hyperellipse without changing its shape.

Thus,

- if we can prove that every matrix has an SVD,
- we will have proved that the image of the unit sphere under any linear map is indeed a hyperellipse.

4 – Existence and uniqueness

In this section, we demonstrate the existence of the SVD, the uniqueness of the singular values, as well as under some specific conditions, the uniqueness of the singular vectors (Chapter 4 in Trefethen & Bau, 1997).

Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition $A = U\Sigma V^*$

To prove the existence of the SVD,

- we first isolate the direction of the largest action of A,
- then we proceed by *induction* on the dimension of A.

The proof takes 5 steps.

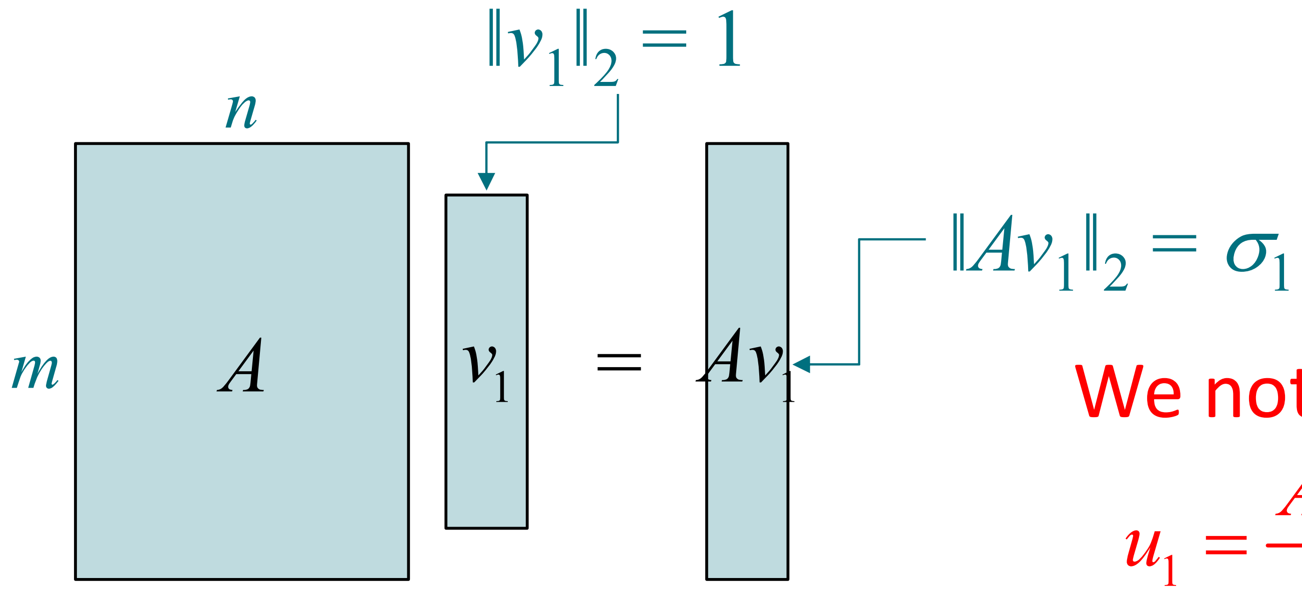
Every matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition $A = U \Sigma V^*$

$$\|A\| = \max_{\substack{x \in \mathbb{C}^n \\ \|x\|=1}} \|Ax\|$$

① Set $\sigma_1 = \|A\|_2$.

From the definition of the matrix norm, there must be a vector $v_1 \in \mathbb{C}^n$

with $\|v_1\|_2 = 1$ and $\|Av_1\|_2 = \sigma_1$



We note:

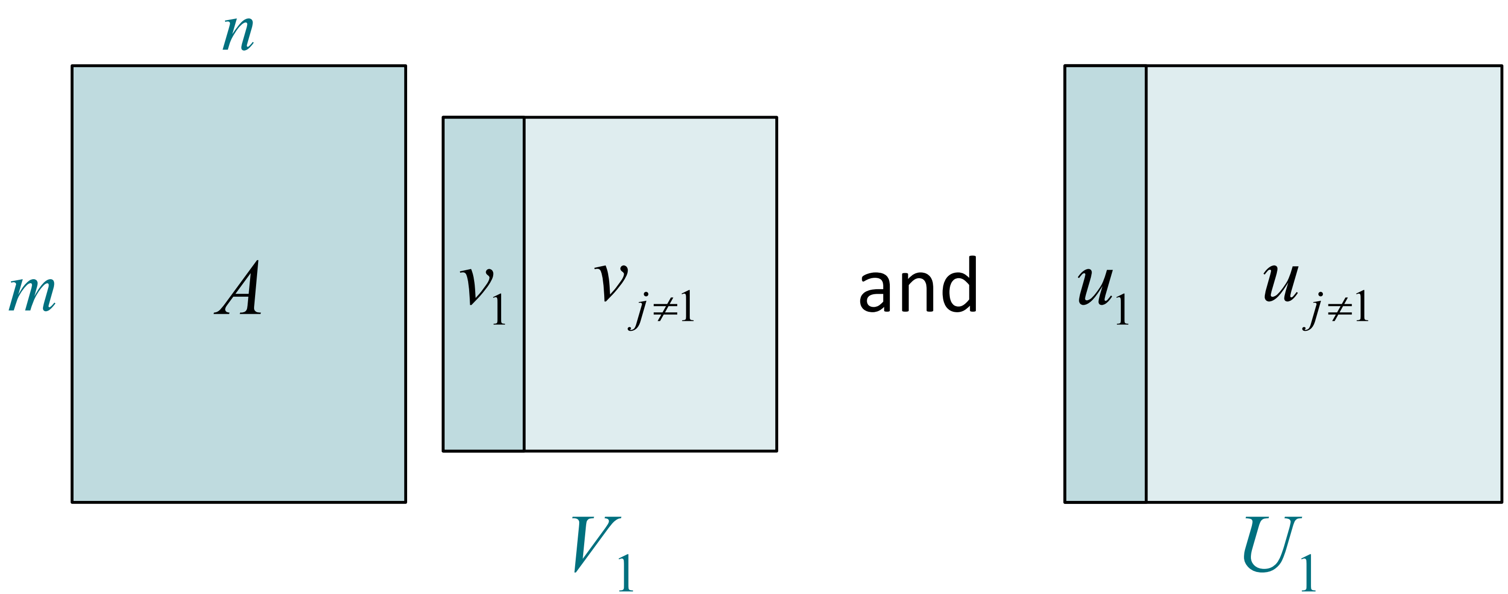
$$u_1 = \frac{Av_1}{\sigma_1}$$

Every matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition $A = U \Sigma V^*$

② Consider any extensions

- of v_1 to an orthonormal basis $\{v_j\}$ of $\mathbb{C}^n$
- and of u_1 to an orthonormal basis $\{u_j\}$ of $\mathbb{C}^m$

Let U_1 and V_1 denote the unitary matrices with columns u_j and v_j, respectively.

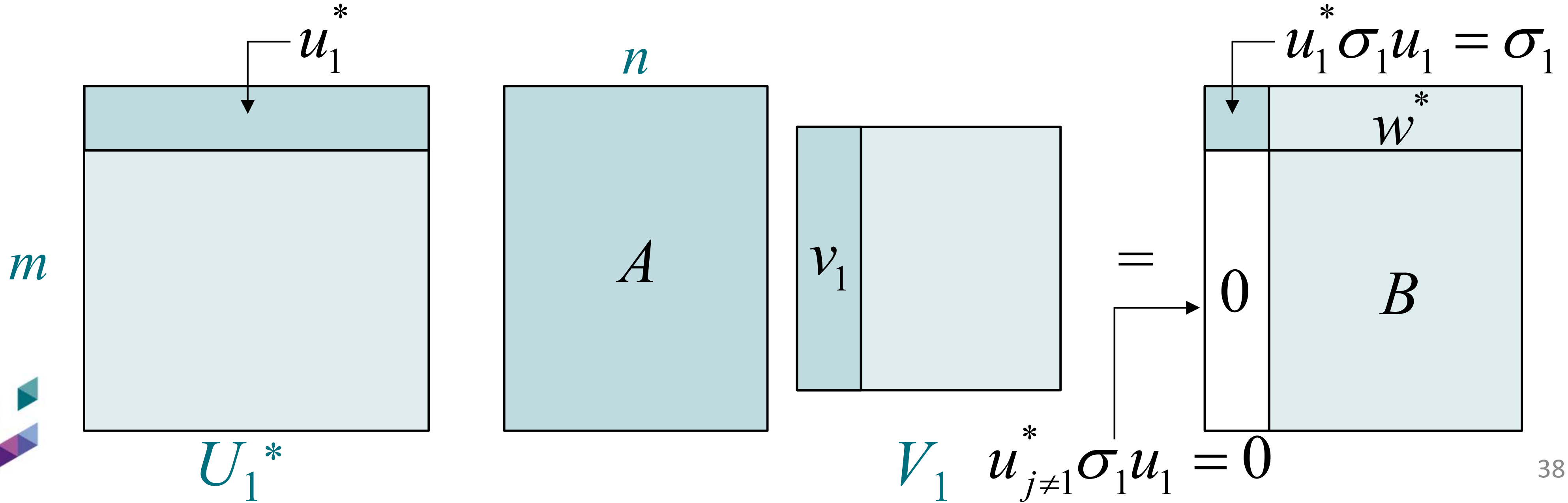


Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition $A = U\,\Sigma\, V^*$

③ Then we have

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix}$$

where 0 is a column vector of dimension $m-1$,
w^* is a row vector of dimension $n-1$, and
B has dimensions $(m-1)\times(n-1)$.



Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition $A = U\Sigma V^*$

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

④ Furthermore,

$$\left\| \underbrace{\begin{bmatrix} \sigma_1 & w^* \\ 0 & B \end{bmatrix}}_{S} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 \geq \sigma_1^2 + w^* w = \left(\sigma_1^2 + w^* w\right)^{1/2} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2$$

Implying (from the definition of matrix norms)

$$\|S\|_2 \geq \left(\sigma_1^2 + w^* w\right)^{1/2}$$

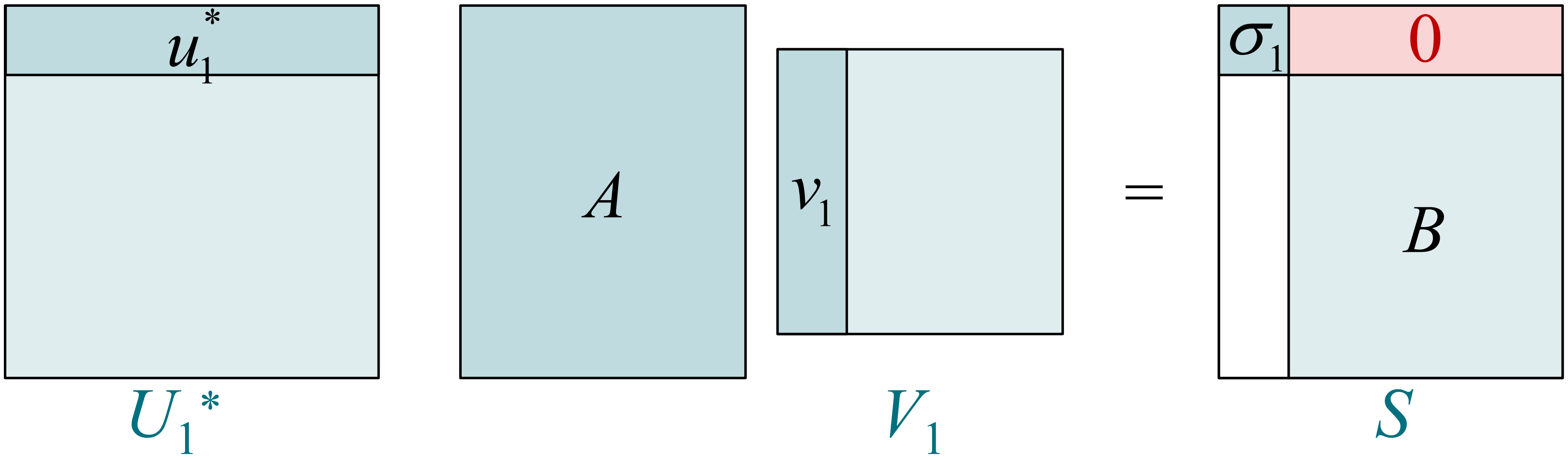
BUT, since U_1 and V_1 are unitary, we know that

$$\|S\|_2 = \|U_1^* A V_1\|_2 = \|A\|_2 = \sigma_1$$

This implies $w = 0$.

Every matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition $A = U \Sigma V^*$

④ To sum up, this is what we know at this stage:



Hence,

$$A = U_1 \, S \, V_1^*$$

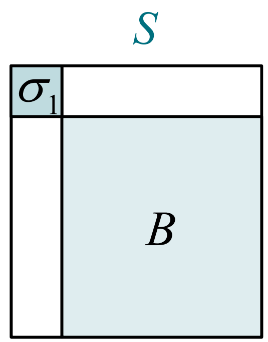
Every matrix $A \in \mathbb{C}^{m \times n}$ has a singular value decomposition $A = U \Sigma V^*$

⑤ If $n = 1$ or $m = 1$, we are done!

Otherwise, the submatrix B describes the action of A on the subspace orthogonal to v_1.

By the induction hypothesis, B has an SVD

$$B = U_2 \Sigma_2 V_2^*.$$



Now it is easily verified that

$$A = U_1 \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_2 \end{bmatrix}^*}_{S} V_1^*$$

is an SVD of A, completing the proof of existence.

Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition $A = U\,\Sigma\,V^*$

⑤ Written out in full:

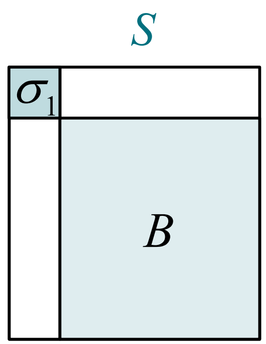
$$\underbrace{U_1\begin{bmatrix}1 & 0\\ 0 & U_2\end{bmatrix}}_{\text{Unitary matrix}}\begin{bmatrix}\sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\underbrace{\begin{bmatrix}1 & 0\\ 0 & V_2\end{bmatrix}^* V_1^*}_{\text{Unitarymatrix}}$$

$$= U_1\begin{bmatrix}\sigma_1 & 0\\ 0 & U_2\Sigma_2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & V_2^*\end{bmatrix}V_1^*$$

$$= U_1\begin{bmatrix}\sigma_1 & 0\\ 0 & U_2\Sigma_2 V_2^*\end{bmatrix}V_1^*$$

$$= U_1\begin{bmatrix}\sigma_1 & 0\\ 0 & B\end{bmatrix}V_1^* = U_1\,S\,V_1^* = A$$

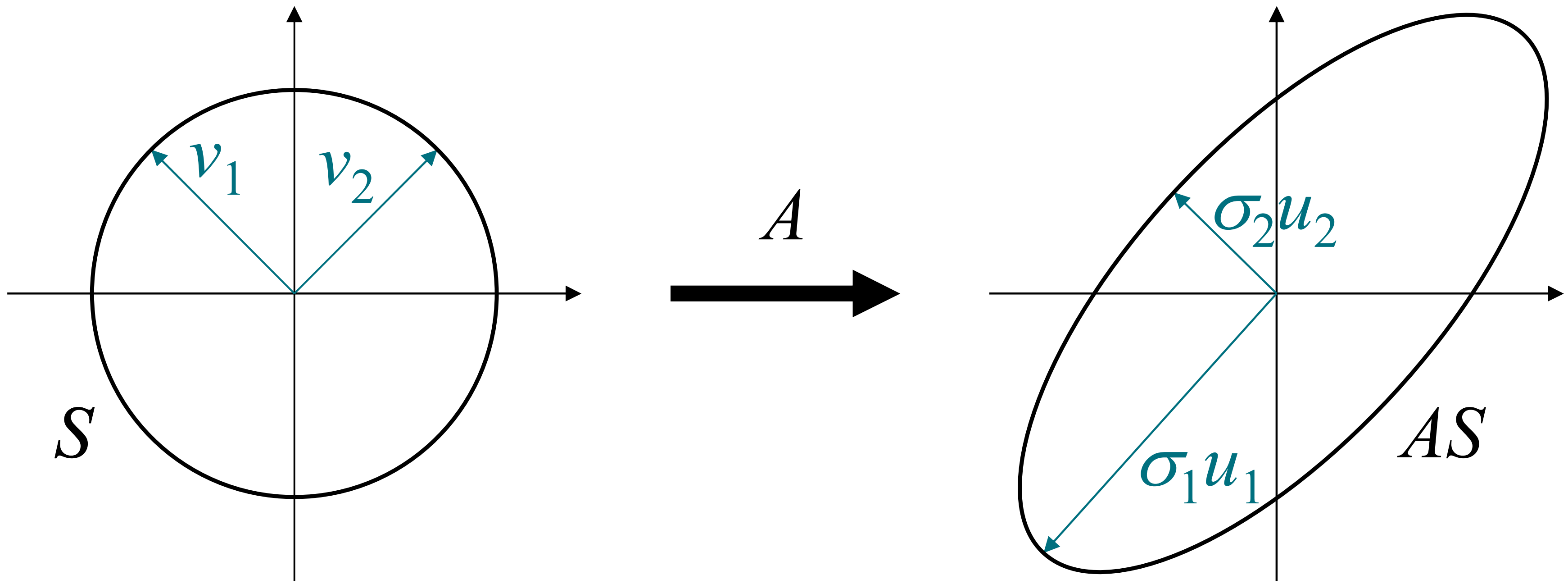
The product of two unitary matrices is another unitary matrix.



Uniqueness

The singular values $\{\sigma_j\}$ are uniquely determined.

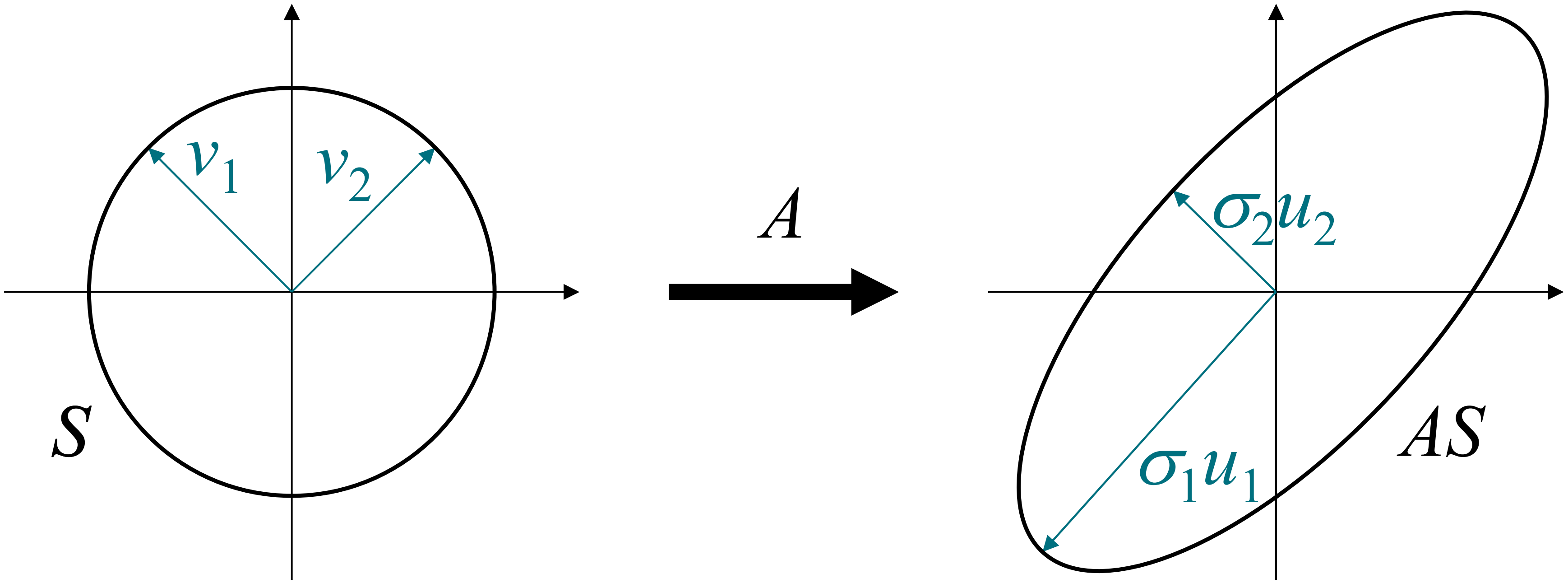
If A is square and the σ_j are distinct, the left and right singular vectors $\{u_j\}$ and $\{v_j\}$ are uniquely determined up to complex signs.



Uniqueness

Geometrically, the proof is straightforward:

- if the semiaxis lengths of a hyperellipse are distinct,
- then the semiaxes themselves are determined by the geometry, up to signs.



Take-home messages

SVD is an important factorization method, which applies for all rectangular, real or complex matrices

It decomposes the matrix into three factors

- a unitary matrix
- a real diagonal matrix, with nonnegative entries
- another unitary matrix

It has a broad range of implications and applications!

What’s next?

Every matrix is diagonal if only one uses the proper bases for the domain and range spaces.

SVD vs. eigenvalue decomposition

- existence
- rectangular vs. square matrices
- orthonormal bases in the SVD, not eigenvectors

Link with matrix rank, range, null space, norm …

Low-rank approximations

Low-rank approximations of a matrix

~ 1 %

10 / 1920

~ 4 %

30 / 1920

~ 14 %

100 / 1920

100 %

1920 / 1920

Full HD