Singular Value Decomposition

my favorite among all factorizations

yes, it is worth studying yet another factorization

SVD is <u>general</u>: can find SVD for any matrix A, even if A is:

- Nonsymmetric
- Noninvertible
- Invertible, but with nearly linearly dependent columns
- Nonsquare (we will focus on the square case for today)
- Complex-valued

SVD is <u>meaningful</u>:

- Rich geometric & algebraic information
- Enables low-rank approximations
- Exposes bases for nullspace, for image of A, and for subspaces perpendicular to these

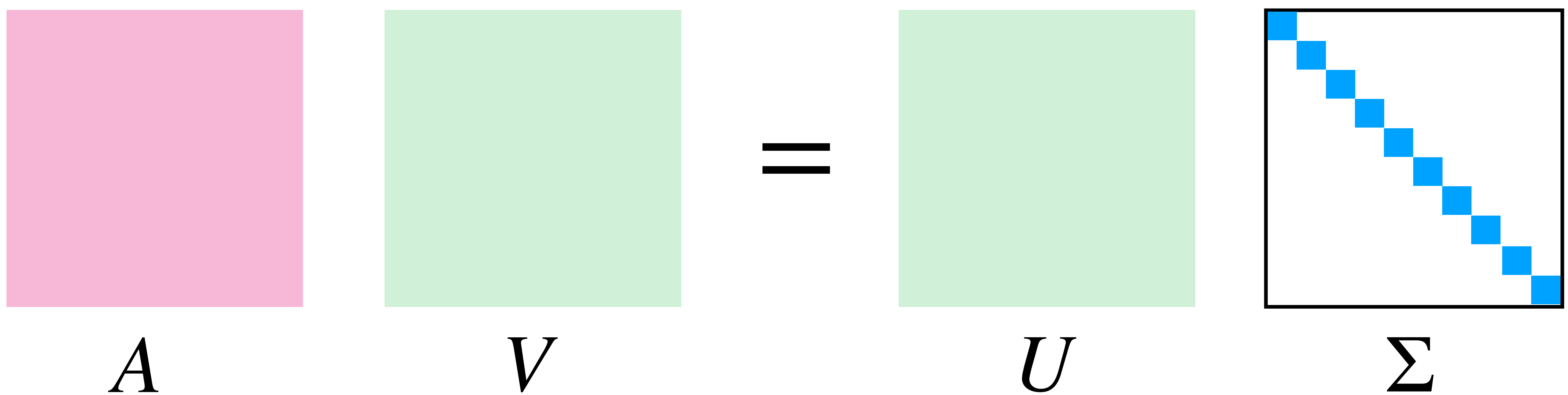
The SVD is a factorization:

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

U and V are orthogonal

Σ is diagonal with nonnegative entries, arranged from largest to smallest

Equivalent to $AV = U\Sigma$

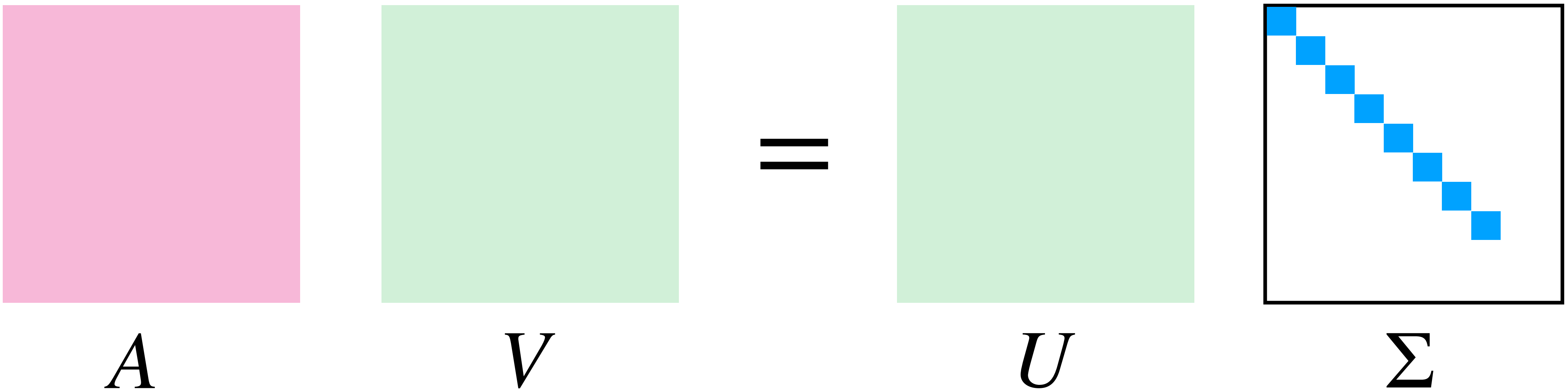


$AV = U\Sigma$ implies $AV_1 = \sigma_1 U_1$ and $AV_2 = \sigma_2 U_2$ and so on

Columns of U are the *left singular vectors* of A

Columns of V are the *right singular vectors* of A

Entries on the diagonal of Σ are the *singular values* of A

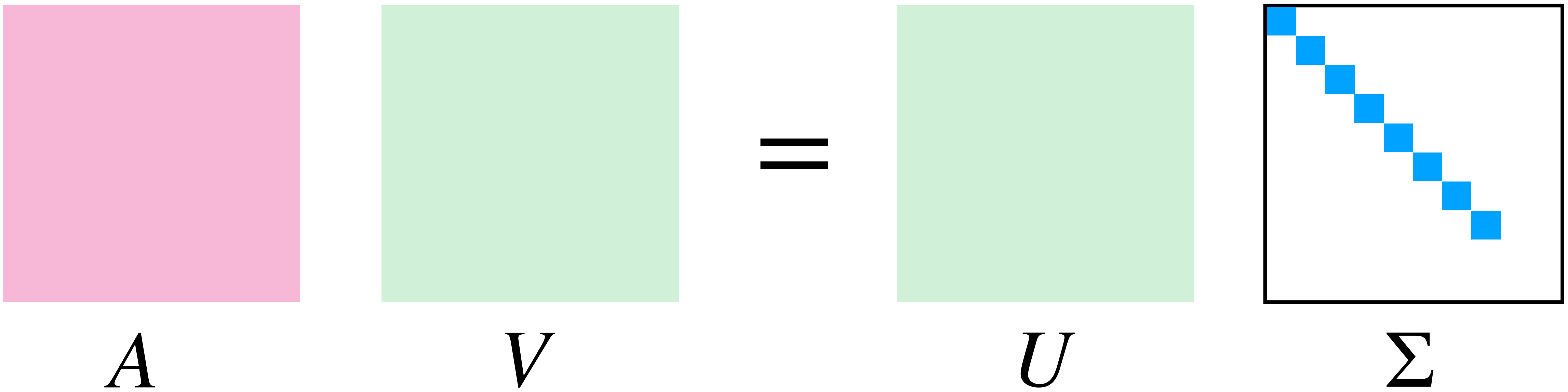


Example: you take the SVD of a ten-by-ten matrix A.

You notice two zero singular values; the first eight are positive.

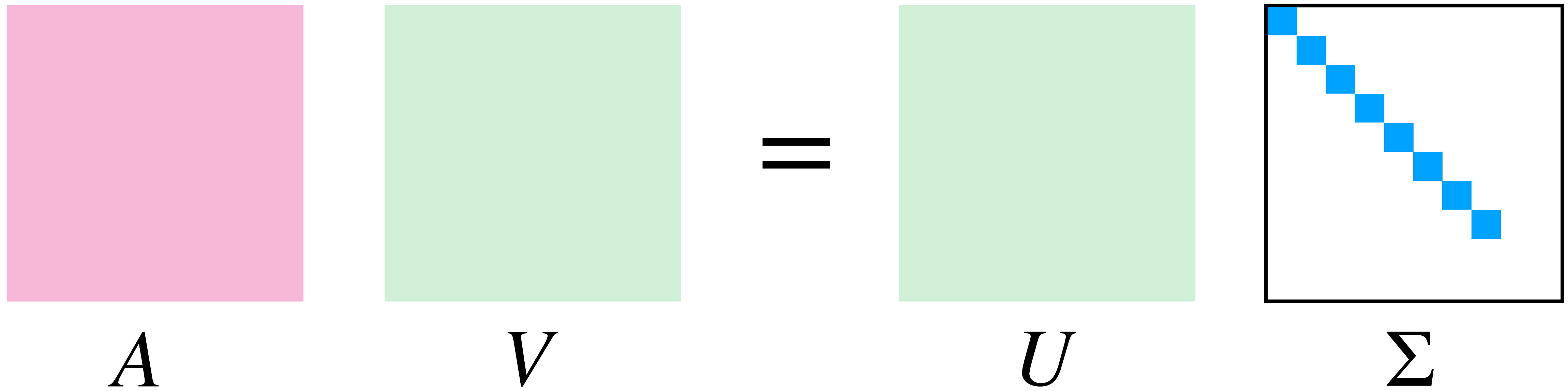
That is, $\sigma_9 = 0 = \sigma_{10}$ while $\sigma_8 > 0$.

What does this mean?



Example: you take the SVD of a ten-by-ten matrix A.
You notice two zero singular values; the first eight are positive.
That is, $\sigma_9 = 0 = \sigma_{10}$ while $\sigma_8 > 0$.
What does this mean?

- We know $AV_9 = \overrightarrow{0}$ and $AV_{10} = \overrightarrow{0}$



Example: you take the SVD of a ten-by-ten matrix A.
You notice two zero singular values; the first eight are positive.
That is, $\sigma_9 = 0 = \sigma_{10}$ while $\sigma_8 > 0$.
What does this mean?

- We know $AV_9 = \overrightarrow{0}$ and $AV_{10} = \overrightarrow{0}$
- A is noninvertible
- $\{V_9, V_{10}\}$ is a basis for the nullspace of A
- $\{U_1, \cdots, U_8\}$ is a basis for the range of A (span of cols of A)
- The equation $Ax = b$ has a solution only if $b \perp U_9$ and $b \perp U_{10}$
- If x is one solution to $Ax = b$, another is $(x + 2.2V_9 - 4.7V_{10})$

Low rank approximation

What if the first three singular values are much larger than the others?
We can get a good approximation by dropping all of the small ones:

$$A = \sum_{j=1}^{n} \sigma_j U_j V_j^T \qquad\qquad A \approx \sum_{j=1}^{3} \sigma_j U_j V_j^T$$

This is the **best** possible rank-3 approximation of A.

Definition: the *Frobenius norm* of a matrix M is the square root of the sum of the squares of the entries of M, written $\|M\|_F$

Theorem: If A has the SVD $AV = U\Sigma$, and we want to find a rank k matrix B to make $\|A - B\|_F$ as small as possible, the answer is $B = \sum_{j=1}^{k} \sigma_j U_j V_j^T$.

Application 5: low-rank approximation

Full (rank 600) image

$$A = \sum_{j=1}^{600} \sigma_j U_j V_j^T$$

store 360000 values

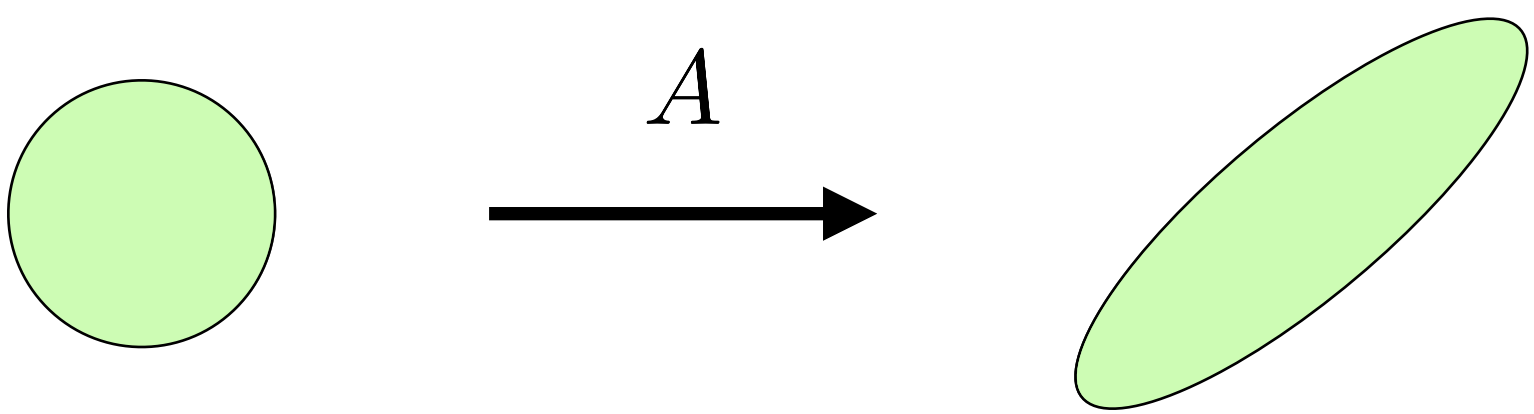
Rank 15 approximation

$$A \approx \sum_{j=1}^{15} \sigma_j U_j V_j^T$$

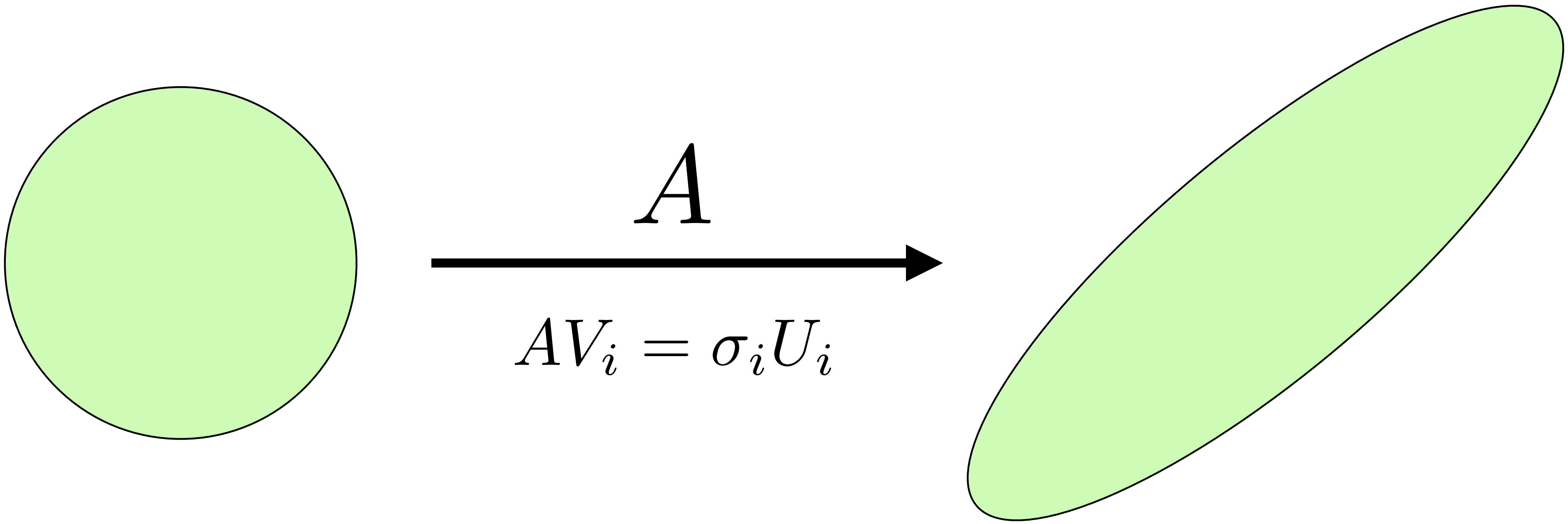
store 18015 values

Geometric introduction of SVD

Let A be n-by-n.
Idea: a linear transformation carries the n-dimensional unit ball to an n-dimensional hyper-ellipse.



$$AV_i = \sigma_i U_i$$



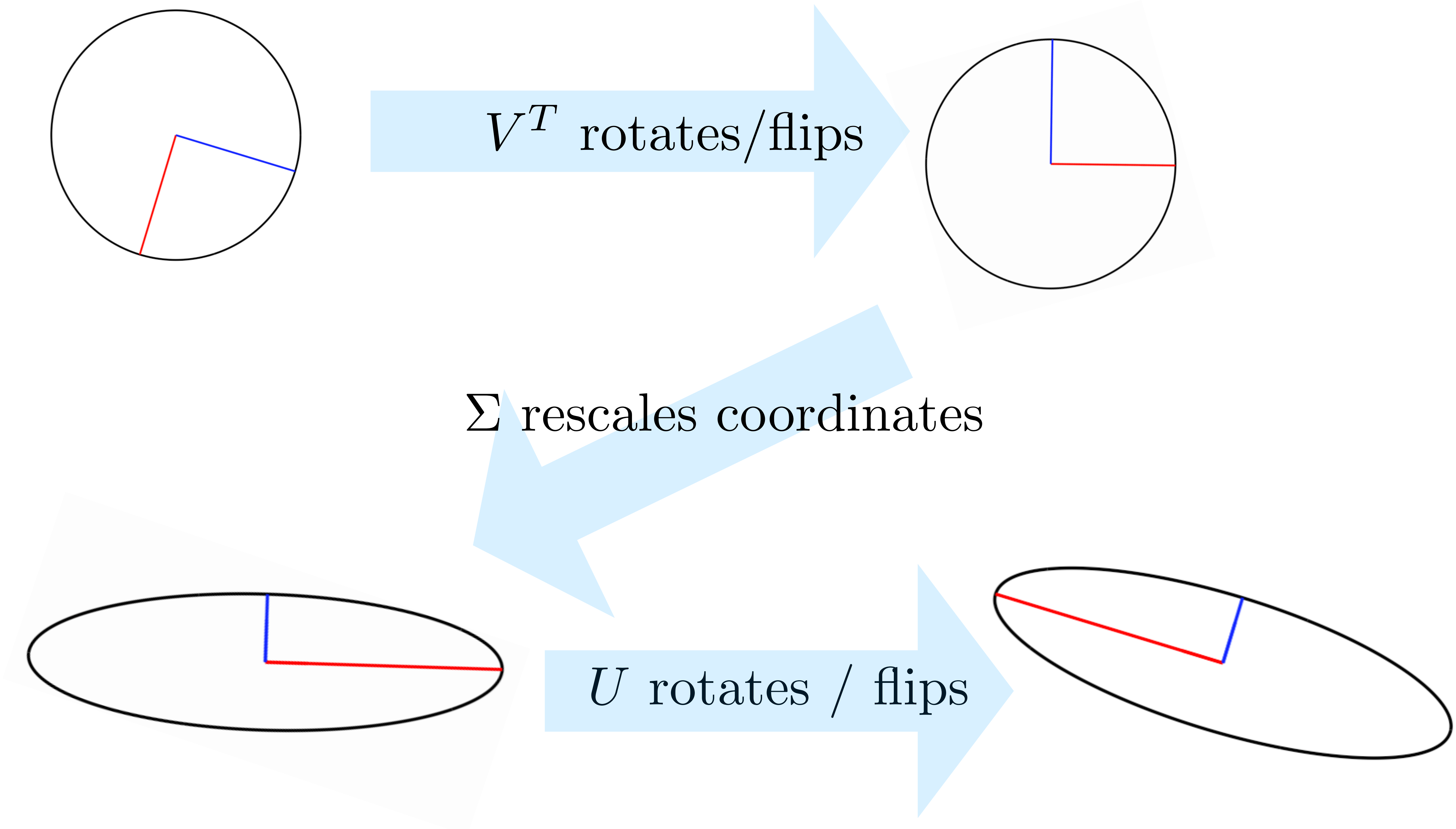
Pictorial, 2D singular value decomposition algorithm!

1. find unit vector V_1 so that $\|AV_1\|$ is as large as possible
2. set $\sigma_1 = \|AV_1\|$ and $U_1 = \dfrac{1}{\sigma_1} AV_1$
3. find unit vector V_2 so that $\|AV_2\|$ is as small as possible

(alternative if $n > 2$: find unit V_2 with $V_2 \perp V_1$ so AV_2 is as large as possible)

4. set $\sigma_2 = \|AV_2\|$ and $U_2 = \dfrac{1}{\sigma_2} AV_2$

$$Aw = U\Sigma V^T w$$



Algebraic construction / definition of SVD

(in practice, avoid actually forming $A^T A$ – we'll see how later)

Theorem: *a symmetric matrix in* $\mathbb{R}^{n\times n}$ *has a full set of* n *mutually perpendicular eigenvectors.*

Even if A is nonsymmetric, $A^T A$ is symmetric.

Let $V_1, V_2, \cdots, V_n$ denote the (unit, $\perp$) eigenvectors of $A^T A$.
Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ denote the corresponding eigenvalues of $A^T A$.
Exercise 1: show that the eigenvalues are non-negative, $\lambda_i \geq 0$.
Define the singular values of A as $\sigma_i = \sqrt{\lambda_i}$.
Exercise 2: show that $\|AV_i\| = \sigma_i$ for each $i = 1 \cdots n$.
Define $U_i = \frac{1}{\sigma_i} AV_i$ for $\sigma_i > 0$.
Exercise 3: show that the U_i are also perpendicular unit vectors.

Let $V_1, V_2, \cdots, V_n$ denote the (unit, $\perp$) eigenvectors of $A^T A$.
Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ denote the corresponding eigenvalues of $A^T A$.
Exercise 1: show that the eigenvalues are non-negative, $\lambda_i \geq 0$.
Define the singular values of A as $\sigma_i = \sqrt{\lambda_i}$.
Exercise 2: show that $\|AV_i\| = \sigma_i$ for each $i = 1 \cdots n$.
Define $U_i = \frac{1}{\sigma_i} A V_i$ for $\sigma_i > 0$.
Exercise 3: show that the U_i are also perpendicular unit vectors.

E1 hint: $0 \leq \|AV_i\|^2 = \cdots$

E2 hint: modify E1 solution.

Let $V_1, V_2, \cdots, V_n$ denote the (unit, $\perp$) eigenvectors of $A^T A$.
Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ denote the corresponding eigenvalues of $A^T A$.
Exercise 1: show that the eigenvalues are non-negative, $\lambda_i \geq 0$.
Define the singular values of A as $\sigma_i = \sqrt{\lambda_i}$.
Exercise 2: show that $\|AV_i\| = \sigma_i$ for each $i = 1 \cdots n$.
Define $U_i = \frac{1}{\sigma_i} AV_i$ for $\sigma_i > 0$.
Exercise 3: show that the U_i are also perpendicular unit vectors.

E1 hint: $0 \leq \|AV_i\|^2 = \cdots$
E1 solution: $0 \leq \|AV_i\|^2 = V_i^T A^T A V_i = V_i^T \lambda_i V_i = \lambda_i \|V_i\|^2 = \lambda_i$.

E2 hint: modify E1 solution.
E2 solution: $\|AV_i\|^2 = \cdots = \lambda_i$, so $\|AV_i\| = \sigma_i$.

E3 solution: $\|U_i\| = \frac{1}{\sigma}\|AV_i\| = \frac{\sigma}{\sigma} = 1$, and also
$U_i^T U_j = \frac{1}{\sigma_i \sigma_j} V_i^T A^T A V_j = \frac{1}{\sigma_i \sigma_j} V_i^T \lambda_j V_j = 0.$

Let $V_1, V_2, \cdots, V_n$ denote the (unit, $\perp$) eigenvectors of $A^T A$.
Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ denote the corresponding eigenvalues of $A^T A$.
Exercise 1: show that the eigenvalues are non-negative, $\lambda_i \geq 0$.
Define the singular values of A as $\sigma_i = \sqrt{\lambda_i}$.
Exercise 2: show that $\|AV_i\| = \sigma_i$ for each $i = 1 \cdots n$.
Define $U_i = \frac{1}{\sigma_i} AV_i$ for $\sigma_i > 0$.
Exercise 3: show that the U_i are also perpendicular unit vectors.

For any i with $\sigma_i = 0$, we can't define $U_i = \frac{1}{\sigma_i} AV_i$.
However, if $\sigma_i = 0$, we know $AV_i = 0$ so $AV_i = \sigma_i U_i$ for any choice of U_i.
Therefore, we can choose any U_i we like for these i.
Choose these U_i so that U is orthogonal.

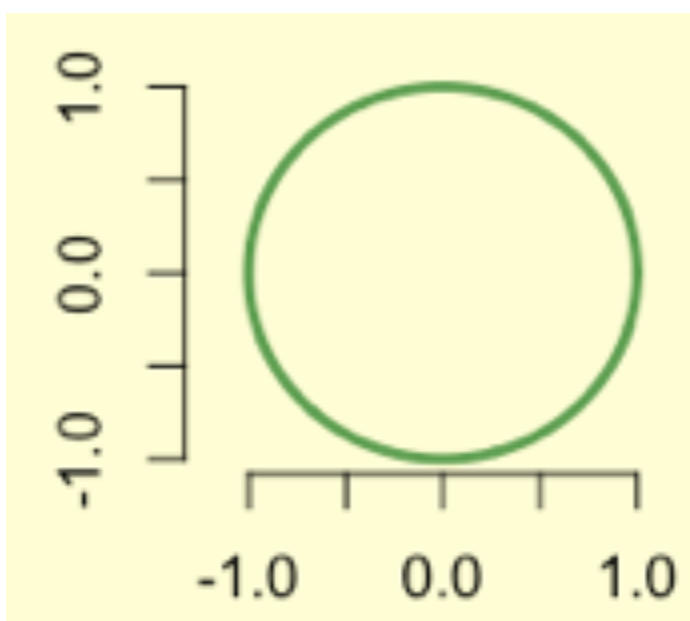
Put it together:
We have $AV_i = \sigma_i U_i$ for each i.
Collect the U_i and V_i into matrices.
Now $AV = U\Sigma$ or $A = U\Sigma V^T$

Application 1: matrix 2-norm

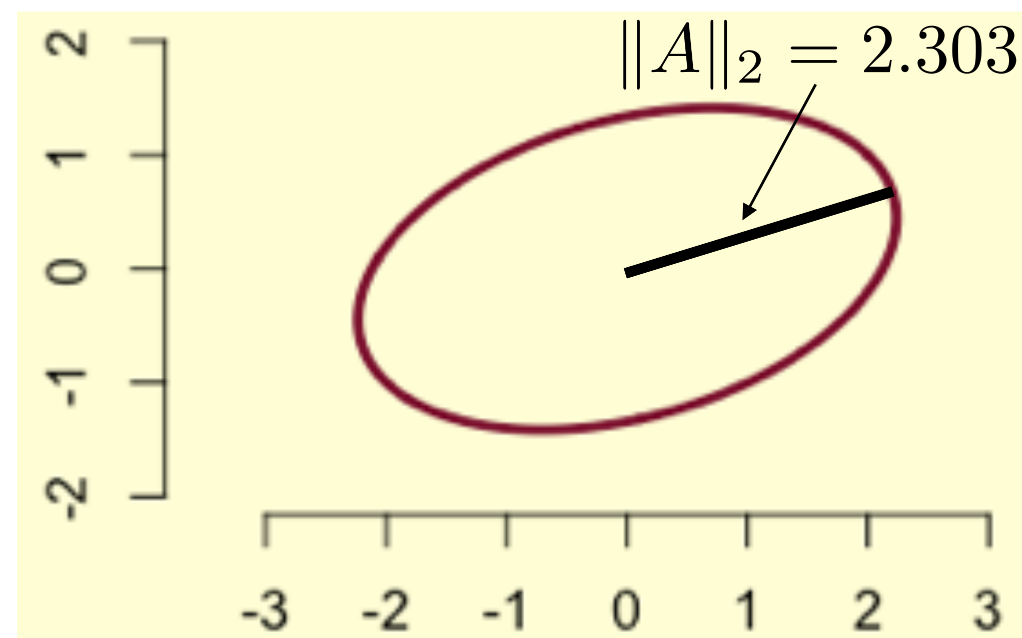
Definition: the 2-norm of a matrix A is the largest possible value of $\|Ax\| = \|Ax\|_2$ given that $\|x\| = 1$.

$$\|A\| = \|A\|_2 = \sigma_1$$

$\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}:$



$\rightarrow$



Application 2: matrix rank

rank(A) = # nonzero singular values of A

numerical rank: the # of singular values greater than a tolerance (like 10^{-14})

Application 3: condition number

$$\kappa(A) = \sigma_1 / \sigma_m$$

In solving Ax=b, if you know b to an accuracy of k digits, you will find x accurate to only $k - \log_{10}(\kappa(A))$ digits

Small condition numbers are good (1 is optimal).
(Orthogonal matrices have condition number 1)

Vandermonde matrices are very poorly conditioned.
(we'll meet these soon)

Application 4: solve Ax=b for square A

1) get SVD $A = U\Sigma V^T$
2) compute $U^T b$
3) Solve diagonal system $\Sigma w = U^T b$
4) compute $x = Vw$

```
A <- matrix(rnorm(16),nrow=4) # random 4x4 matrix
b <- rnorm(4) # random rhs vector
sing <- svd(A)
u <- sing$u           # extract left singular vectors
sigma <- sing$d       # extract vector of singular values
v <- sing$v           # extract right singular vectors
x <- v %*% (1/sigma * (t(u) %*% b)) # get x = A^{-1}b
print(max(abs(A %*% x - b))) # should be close to 0
```

```
## [1] 1.776357e-15
```