Machine Learning

CMPT 726

Mo Chen
SFU School of Computing Science
2021-09-15

Linear Algebra and Calculus Review (cont’d)

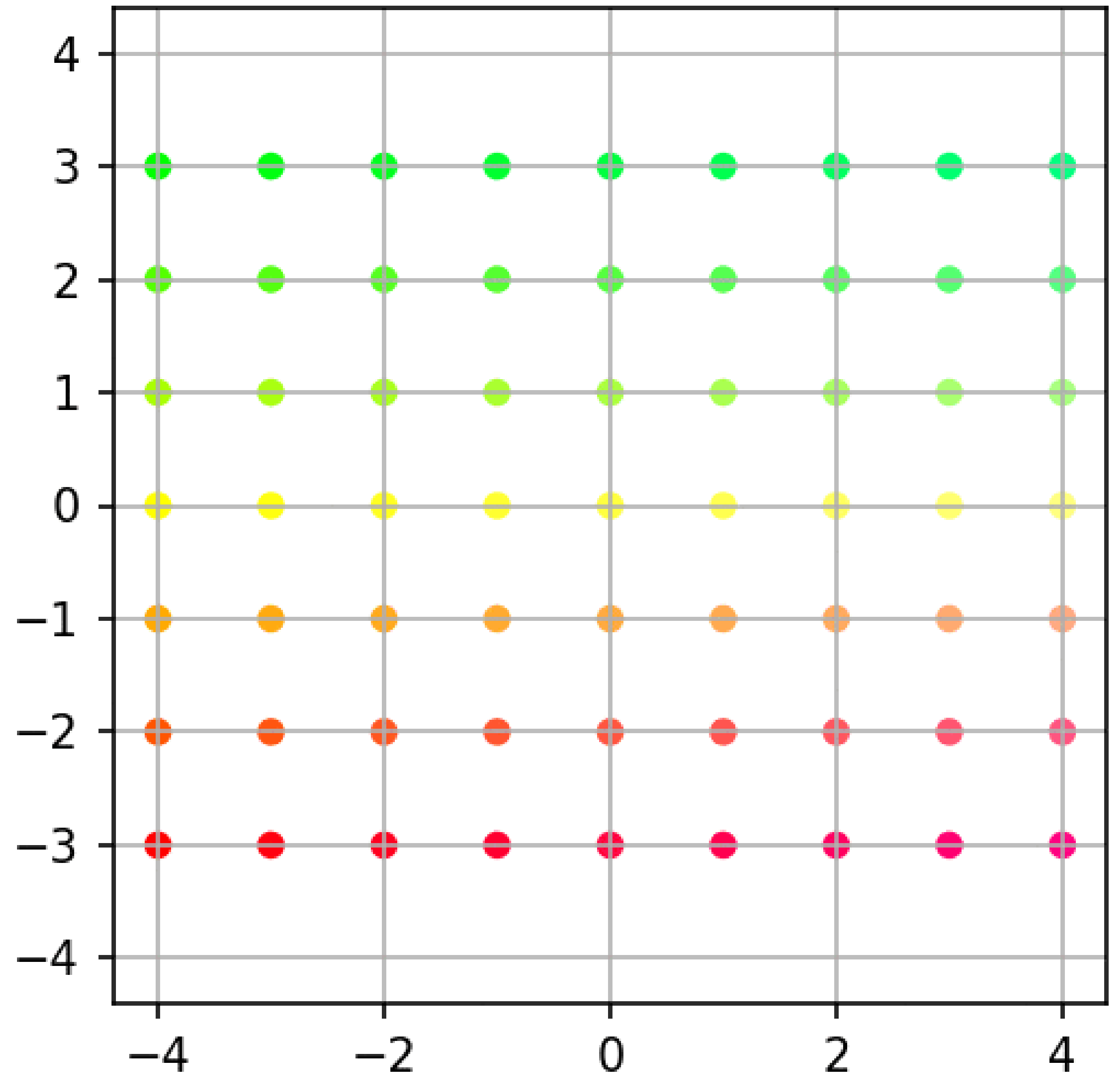
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Full-rank: $A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$



0.9	0.0
0.0	0.5

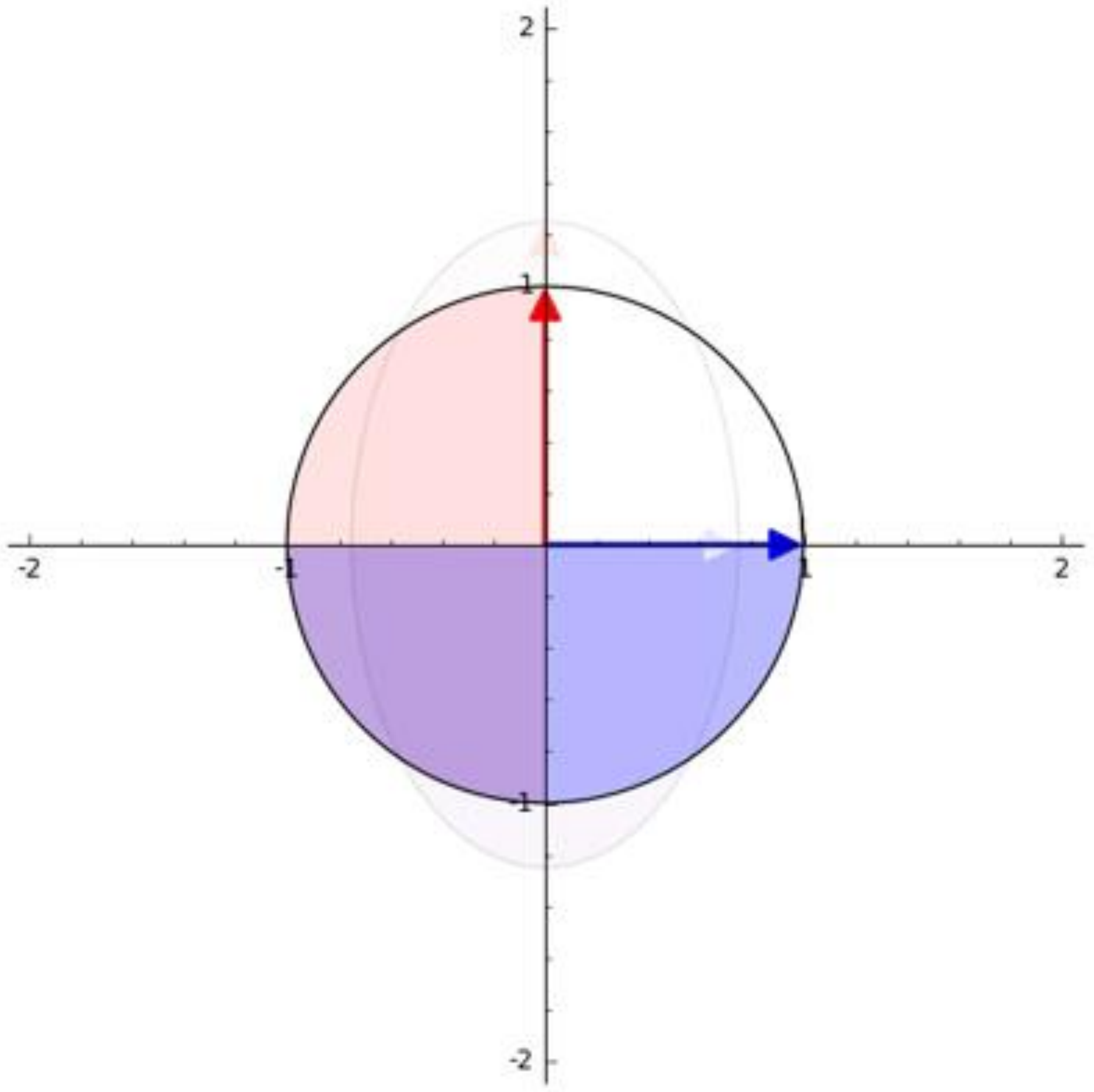
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Full-rank: $A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$



Credit: Ryan Holbrook

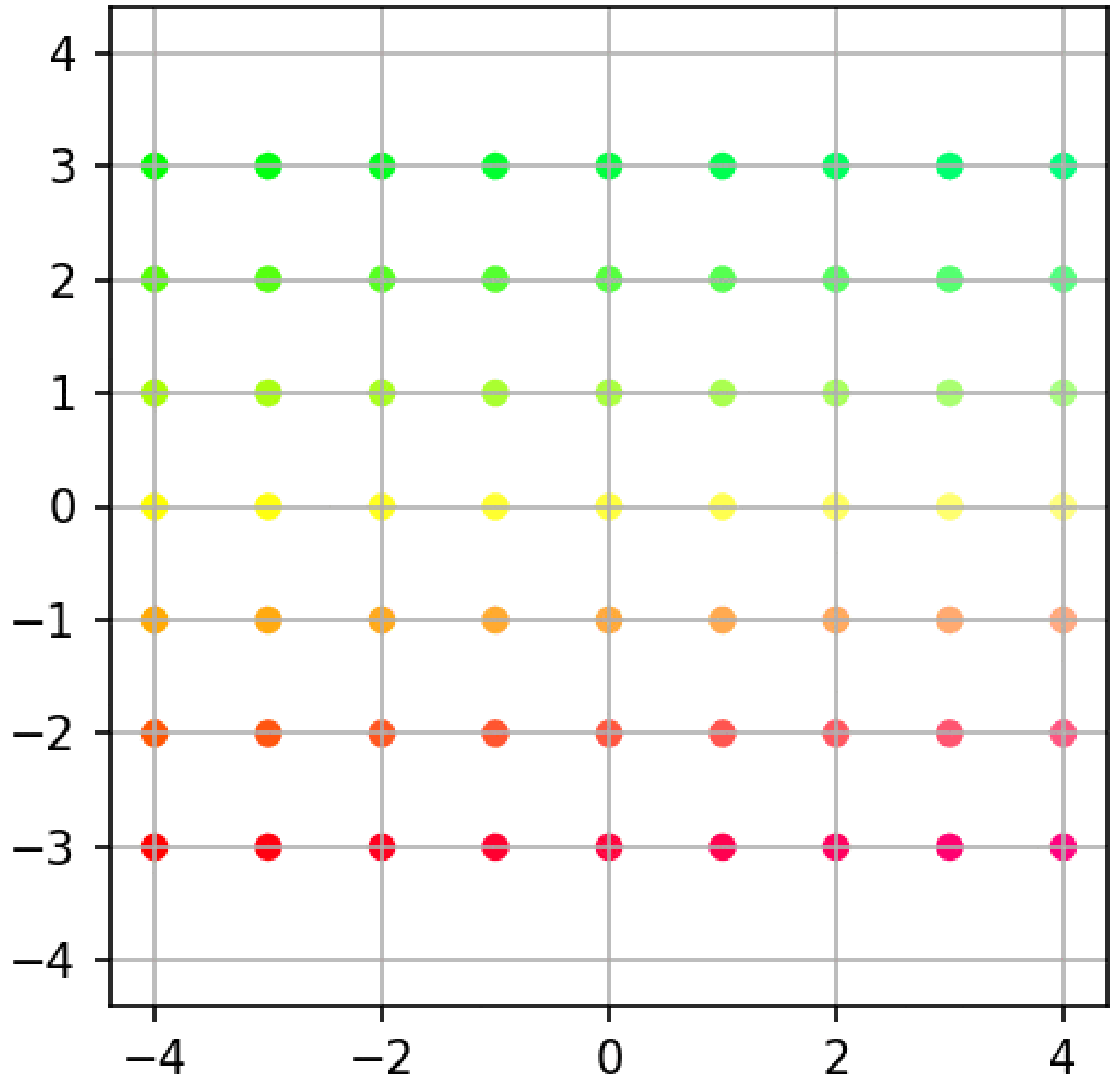
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the numbe of non-zero entries.

Rank-deficient: $A = \begin{pmatrix} 0 & 0 \\ 0 & a_{2,2} \end{pmatrix}$



0.0	0.0
0.0	0.5

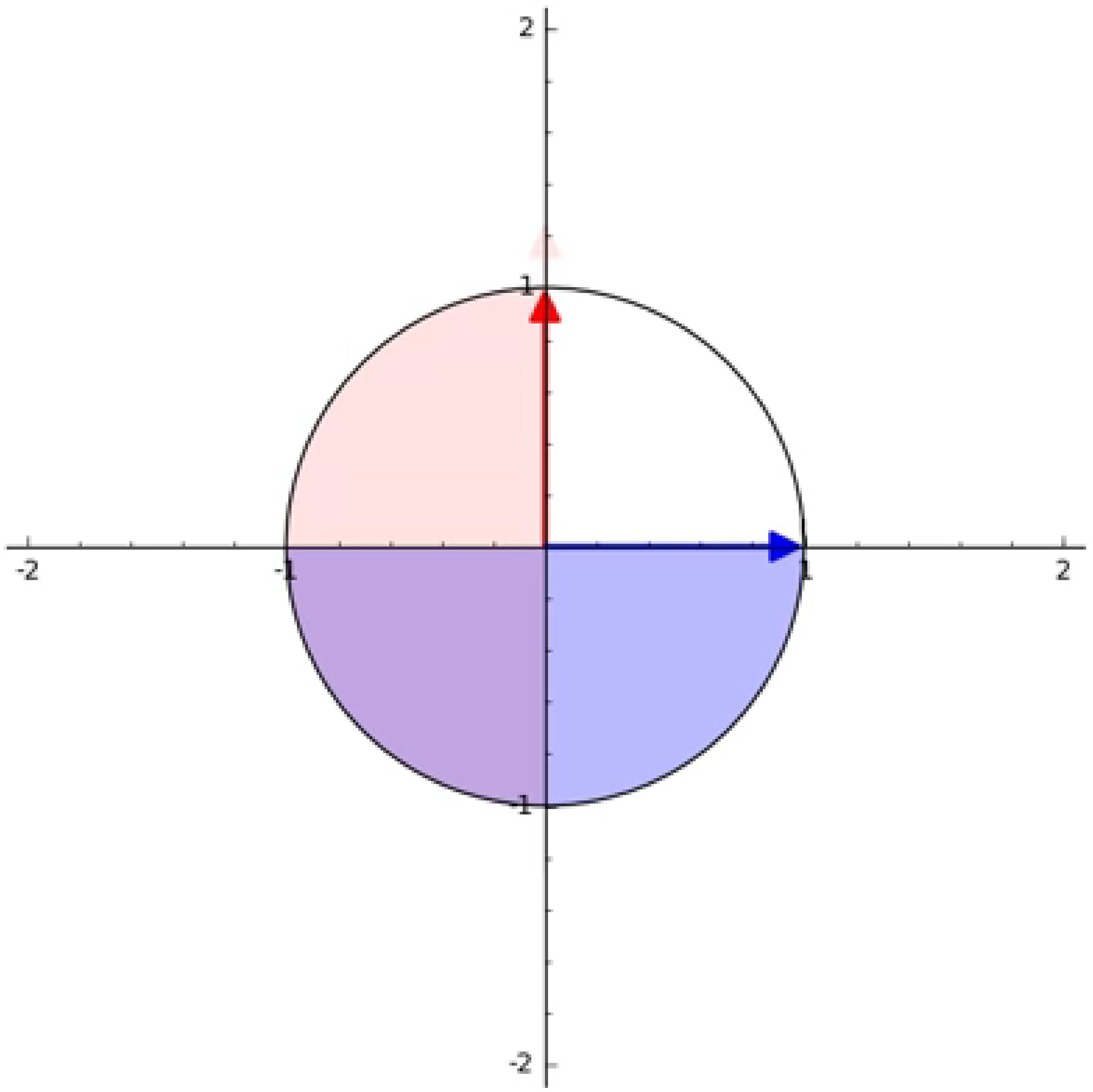
Diagonal Matrices

$$A = \begin{pmatrix} a_{1,1} & 0 \\ 0 & a_{2,2} \end{pmatrix}$$

$$A\vec{x} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 \\ a_{2,2}x_2 \end{pmatrix}$$

In the case of diagonal matrices, rank is the number of non-zero entries.

Rank-deficient: $A = \begin{pmatrix} 0 & 0 \\ 0 & a_{2,2} \end{pmatrix}$



Credit: Ryan Holbrook

Orthogonal Matrices

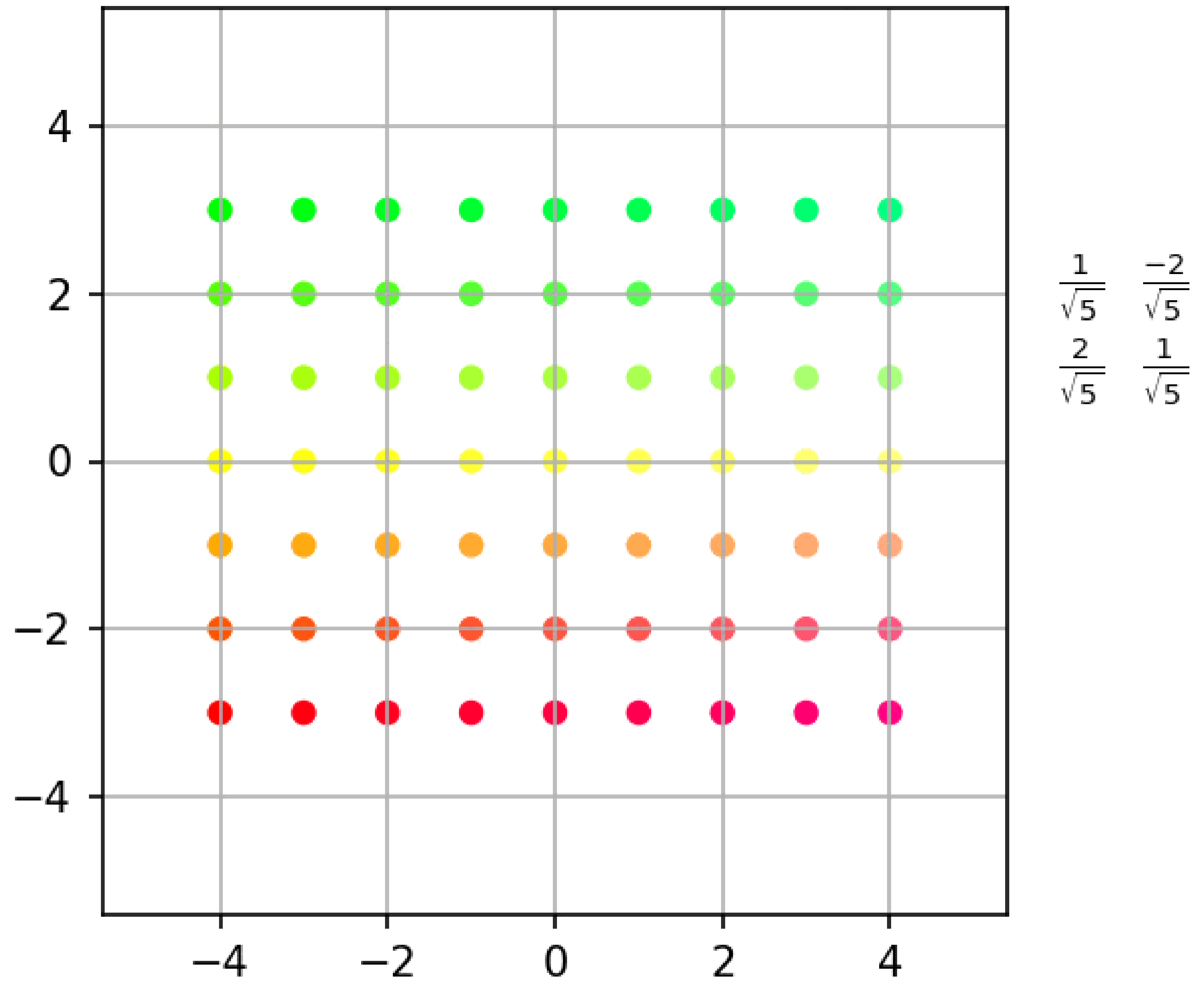
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of 1: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$



Orthogonal Matrices

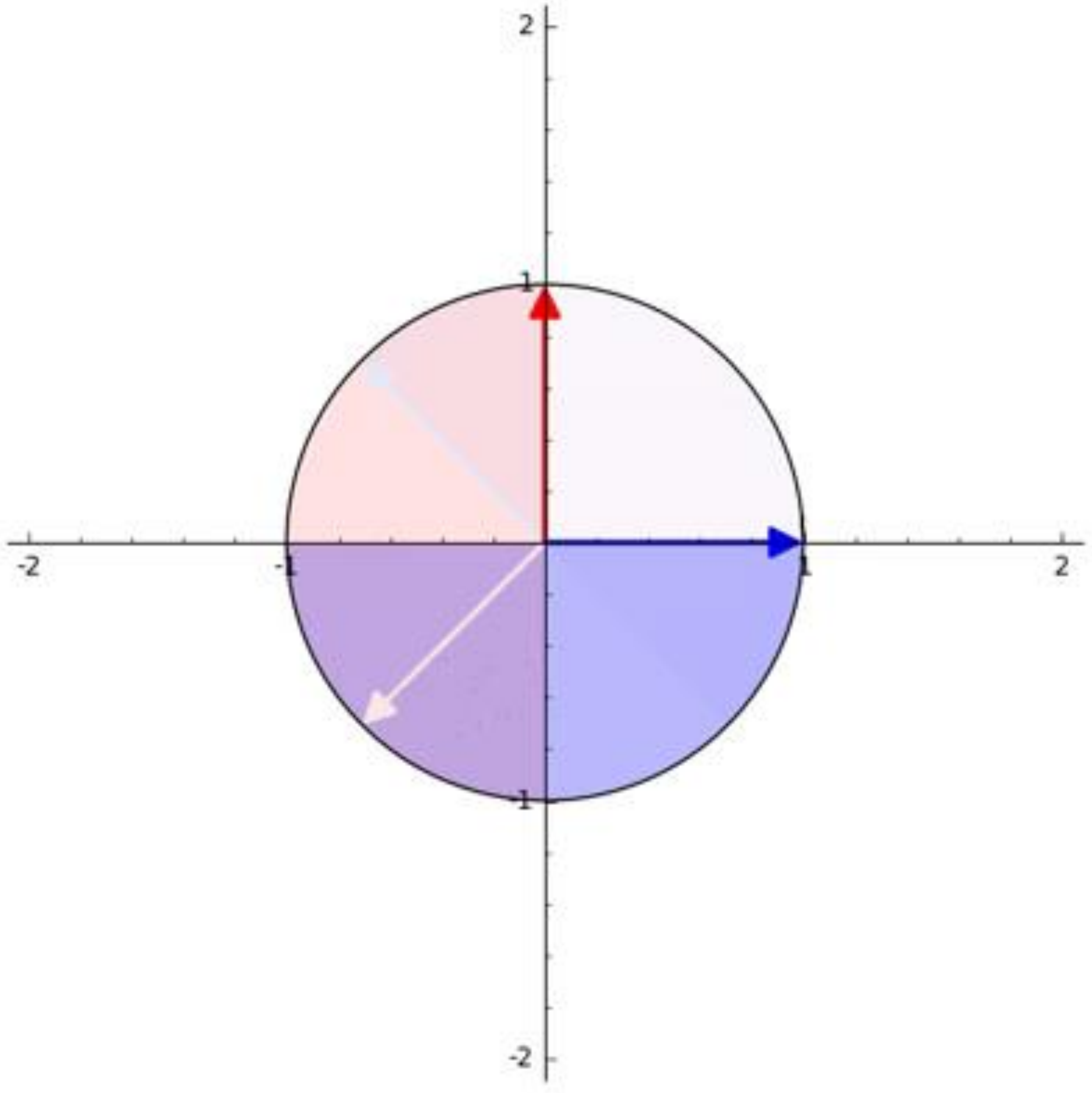
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of 1: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$



Credit: Ryan Holbrook

Orthogonal Matrices

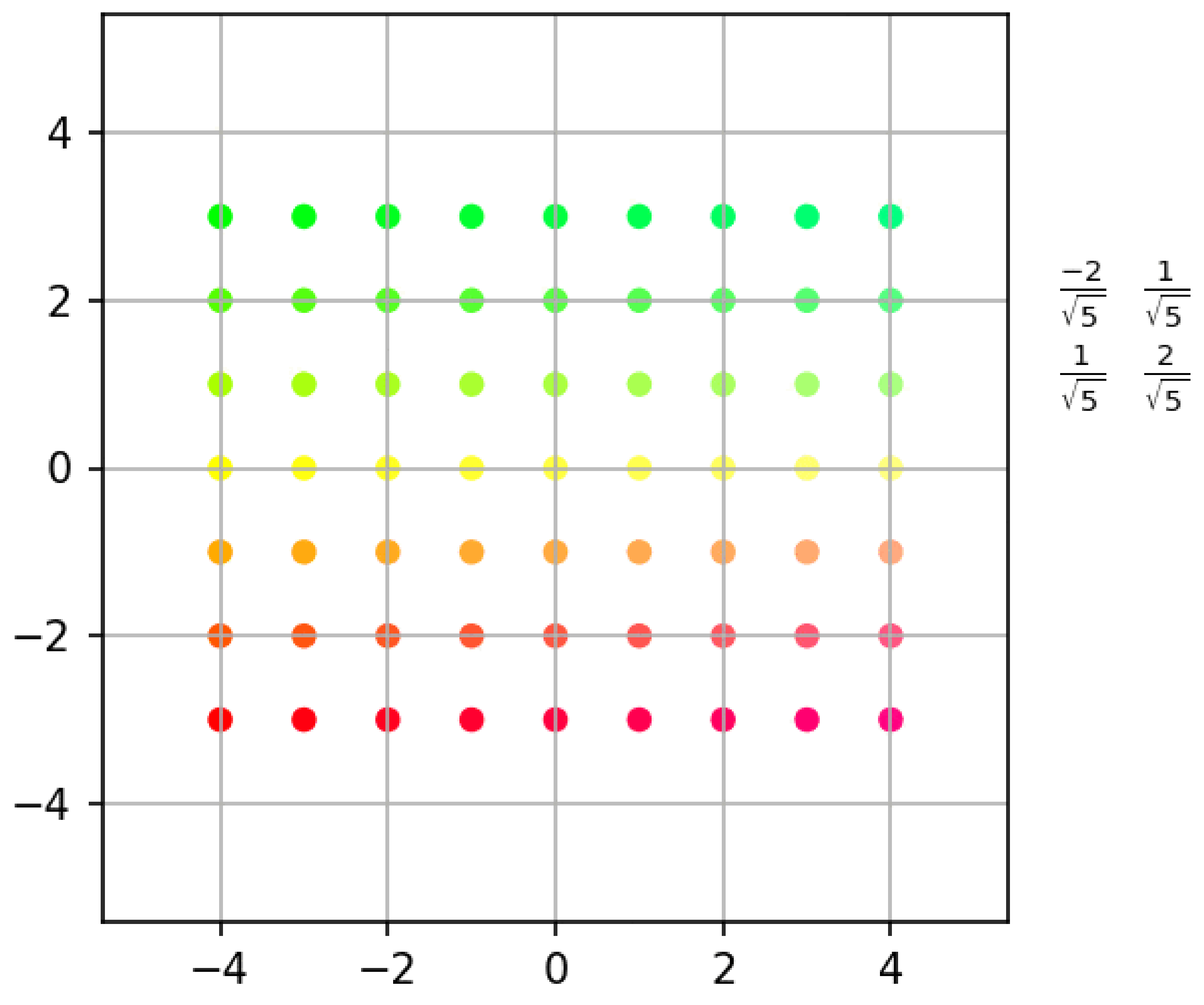
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of -1: $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$



Orthogonal Matrices

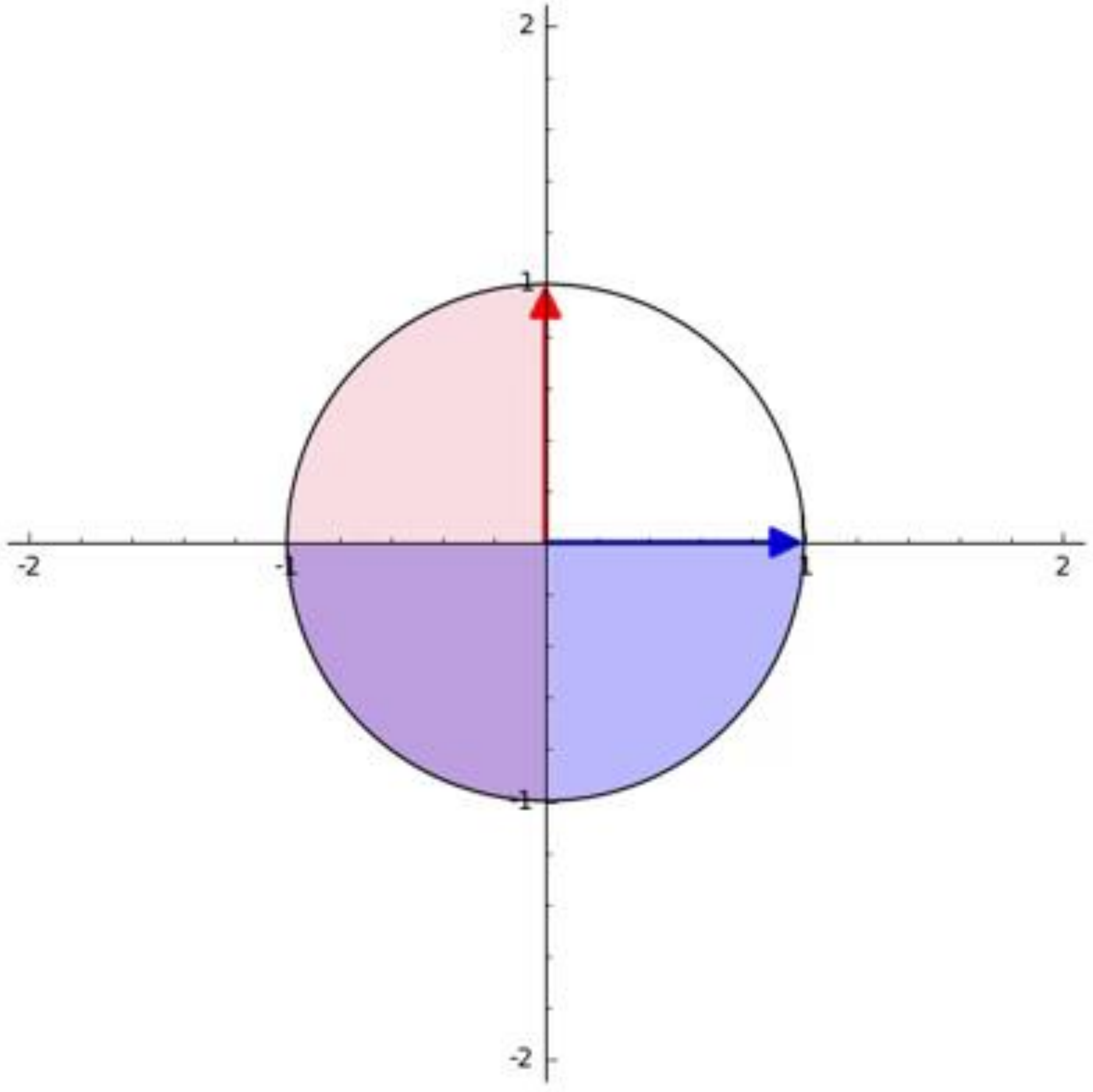
A square matrix A such that $AA^\top = I$ and $A^\top A = I$ (implies $A^{-1} = A^\top$)

$$A^\top A = \begin{pmatrix} \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 1}^\top \vec{a}_{\cdot n} \\ \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot 2}^\top \vec{a}_{\cdot n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 1} & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot 2} & \cdots & \vec{a}_{\cdot n}^\top \vec{a}_{\cdot n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The columns of A form an orthonormal basis

Determinant of -1: $A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$



Credit: Ryan Holbrook

Singular Value Decomposition (SVD)

All matrices can be decomposed into a sequence of:

1. Orthogonal matrix (rotation/reflection)

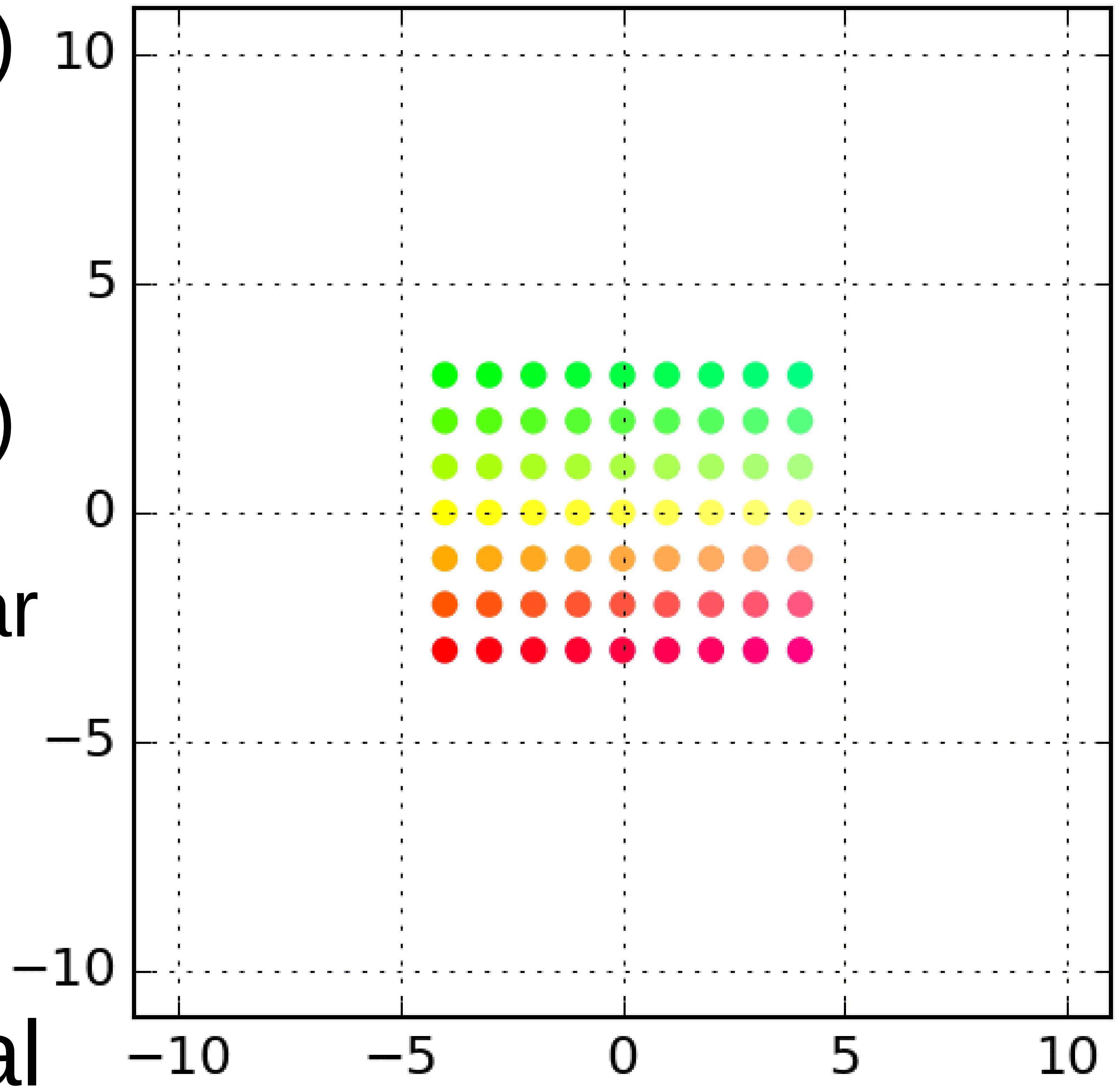
2. Diagonal matrix (scaling along axes)

3. Orthogonal matrix (rotation/reflection)

This decomposition is known as singular value decomposition (SVD):

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

U, V are orthogonal, Σ is diagonal with real non-negative entries



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

Singular Value Decomposition (SVD)

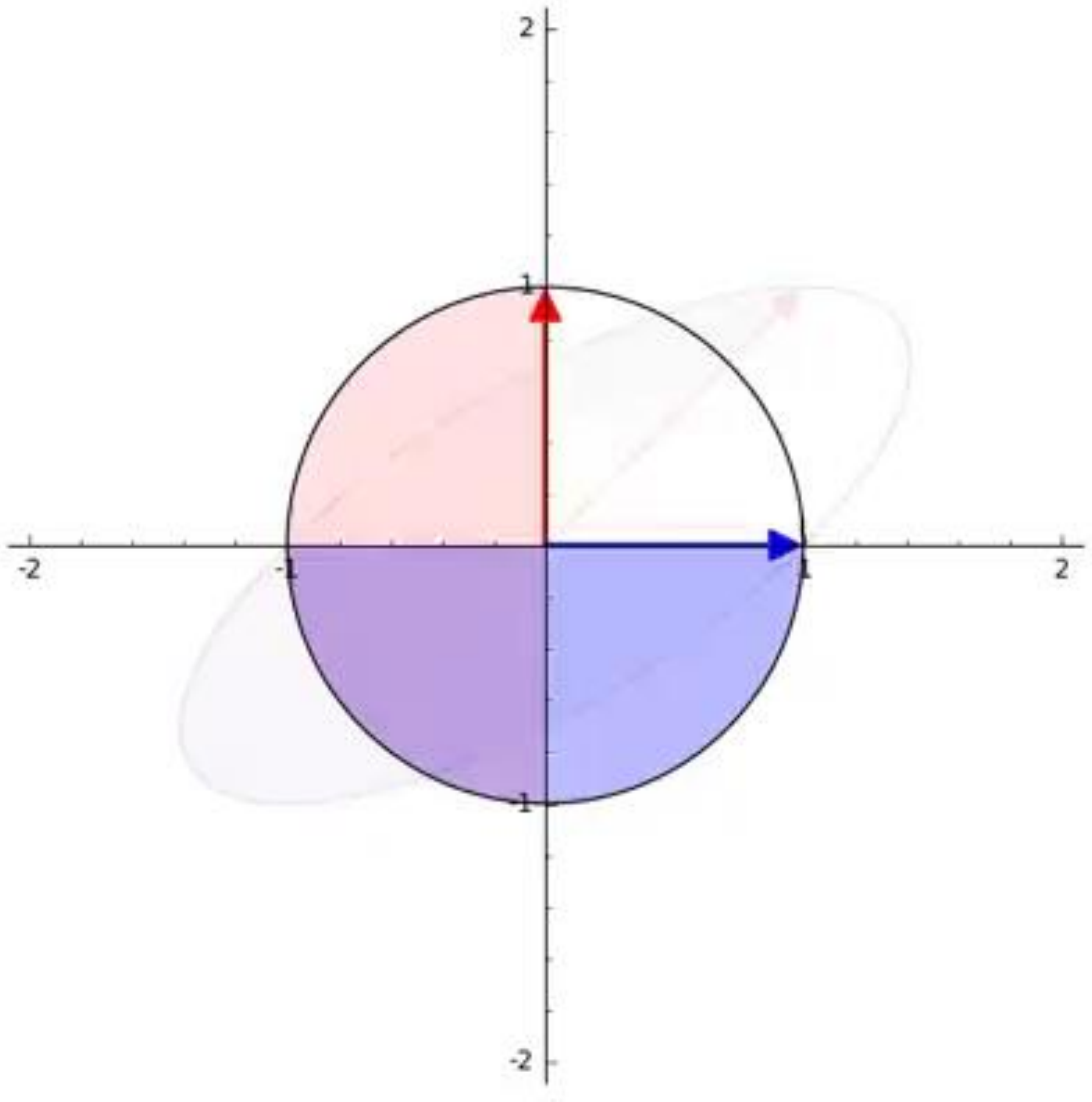
All matrices can be decomposed into a sequence of:

1. Orthogonal matrix (rotation/reflection)
2. Diagonal matrix (scaling along axes)
3. Orthogonal matrix (rotation/reflection)

This decomposition is known as singular value decomposition (SVD):

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

U, V are orthogonal, Σ is diagonal with real non-negative entries



Credit: Ryan Holbrook

Singular Value Decomposition (SVD)

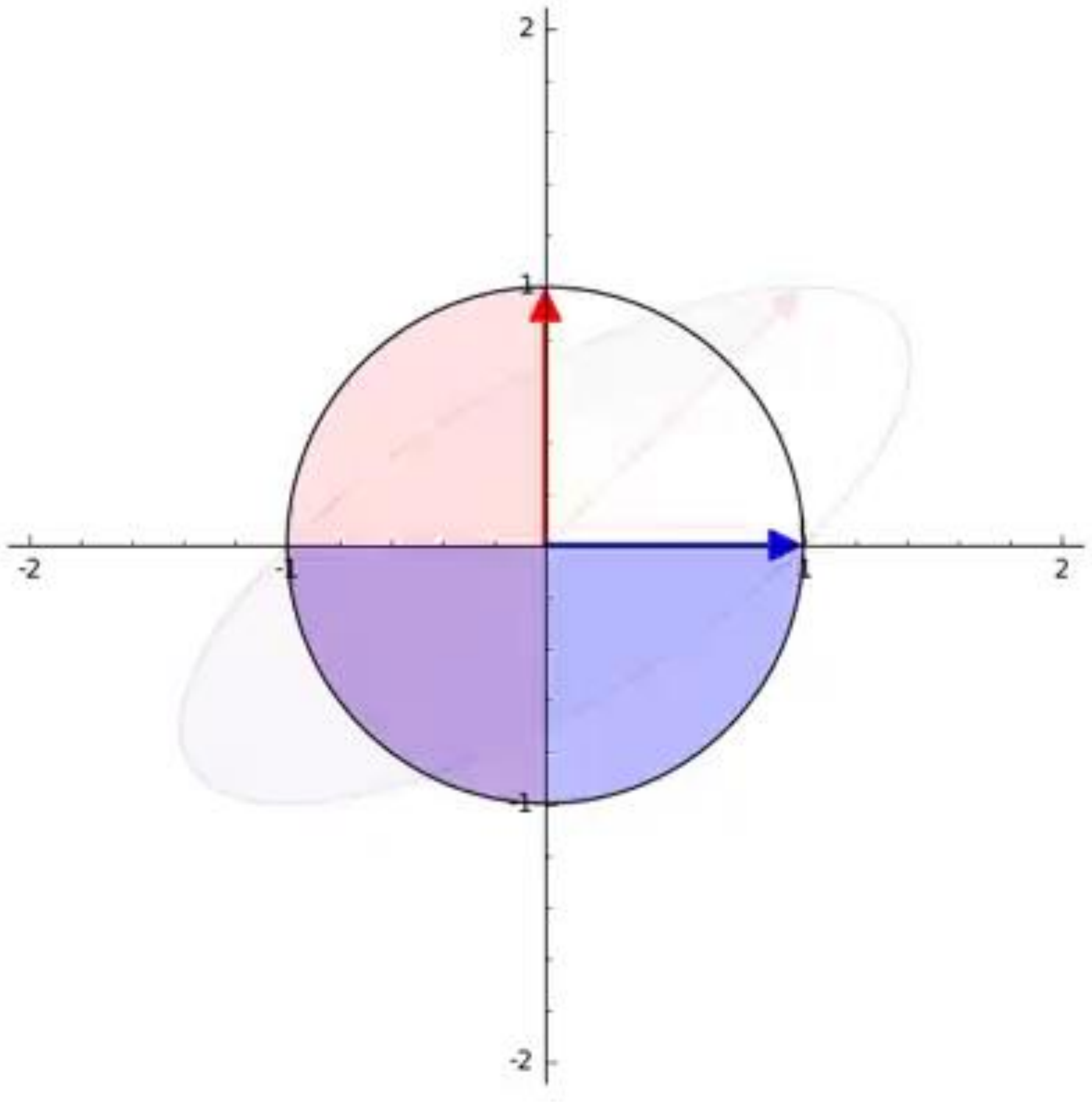
$$A = U\Sigma V^\top$$

U, V are orthogonal, $\sum$ is diagonal (possibly non-square) with real non-negative entries

The columns of V are known as the right-singular vectors

The columns of U are known as the left-singular vectors

The diagonal entries of Σ are known as the singular values



Credit: Ryan Holbrook

Singular Value Decomposition (SVD)

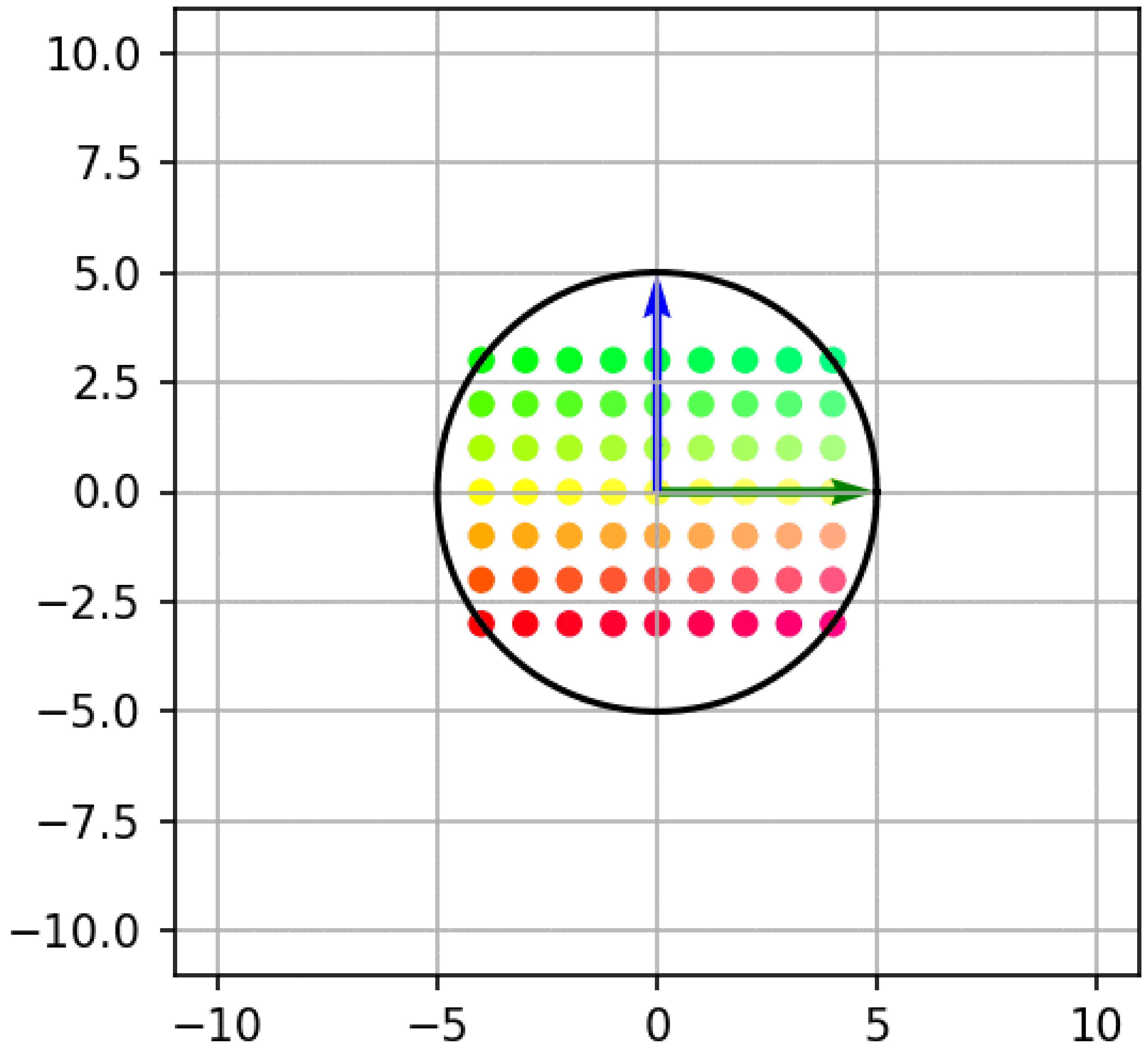
Example:

$$A = \begin{pmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ -\frac{2}{3} & \frac{1}{\sqrt{5}} & -\frac{4}{3\sqrt{5}} \\ -\frac{2}{3} & 0 & \frac{\sqrt{5}}{3} \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \underbrace{\begin{pmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}}_{V^\top}$$

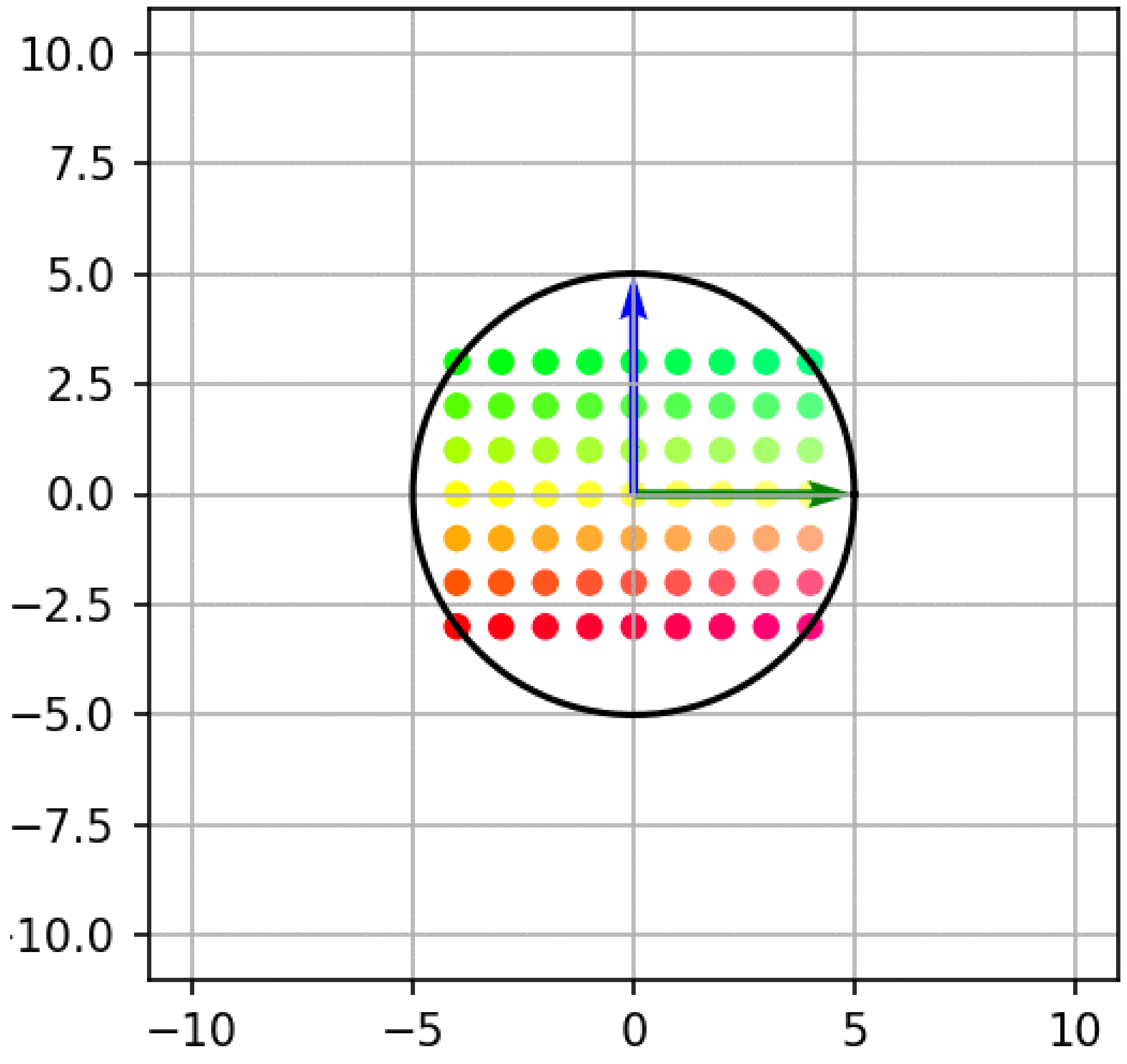
Number of non-zero diagonal entries in Σ corresponds to the rank of A.

SVD and Rank

Full-Rank (Rank = 2)



Rank-Deficient (Rank = 1)



Singular Value Decomposition (SVD)

Compact/Reduced SVD: Eliminates all rows or columns in that are all zeros

$$A = \begin{pmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}}_{\tilde{U}} \underbrace{\begin{pmatrix} 3\sqrt{10} \end{pmatrix}}_{\tilde{\Sigma}} \underbrace{\begin{pmatrix} -\frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{pmatrix}}_{\tilde{V}^{\top}}$$

Here, $\tilde{\Sigma}$ is an $r \times r$ square matrix, where r is the rank of A, $\tilde{U}$ and $\tilde{V}$ are semi-orthogonal, i.e.: possibly non-square matrices whose column vectors are orthonormal

Eigendecomposition

SVD: $A = U\Sigma V^\top$ for orthogonal U, V, diagonal Σ

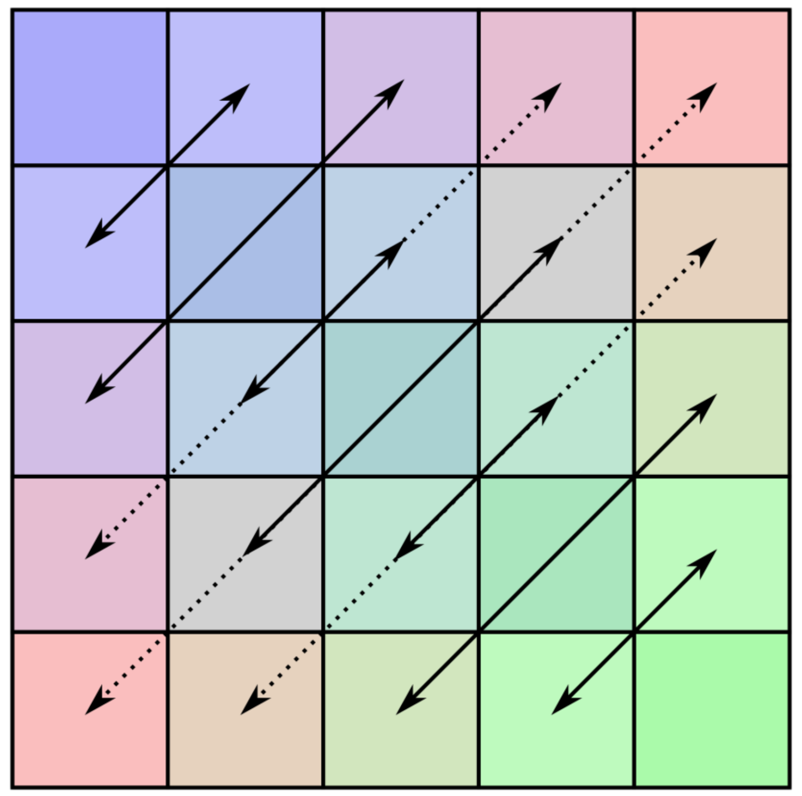
In the special case of (square) symmetric matrices:

The column vectors of U and V are the same up to sign, i.e.: there is a diagonal matrix D with ± 1 on the diagonal such that $V = UD$.

This is called eigendecomposition.

$$A = U\Sigma(UD)^\top = U\Sigma D^\top U^\top = U\Sigma D U^\top$$

$$= U(\Sigma D)U^\top := U\Lambda U^\top \text{ ,where } \Lambda = \Sigma D$$



$$A = A^\top$$

Like Σ, Λ is diagonal. Unlike Σ, Λ can be negative.

The columns of U are known as the eigenvectors
The diagonal entries of Λ are known as the eigenvalues.

For symmetric matrices, the singular values are the absolute values of eigenvalues and the left- and right-singular vectors are $\pm$ the eigenvectors.

Eigendecomposition

$$A = U\Sigma U^{\top}$$

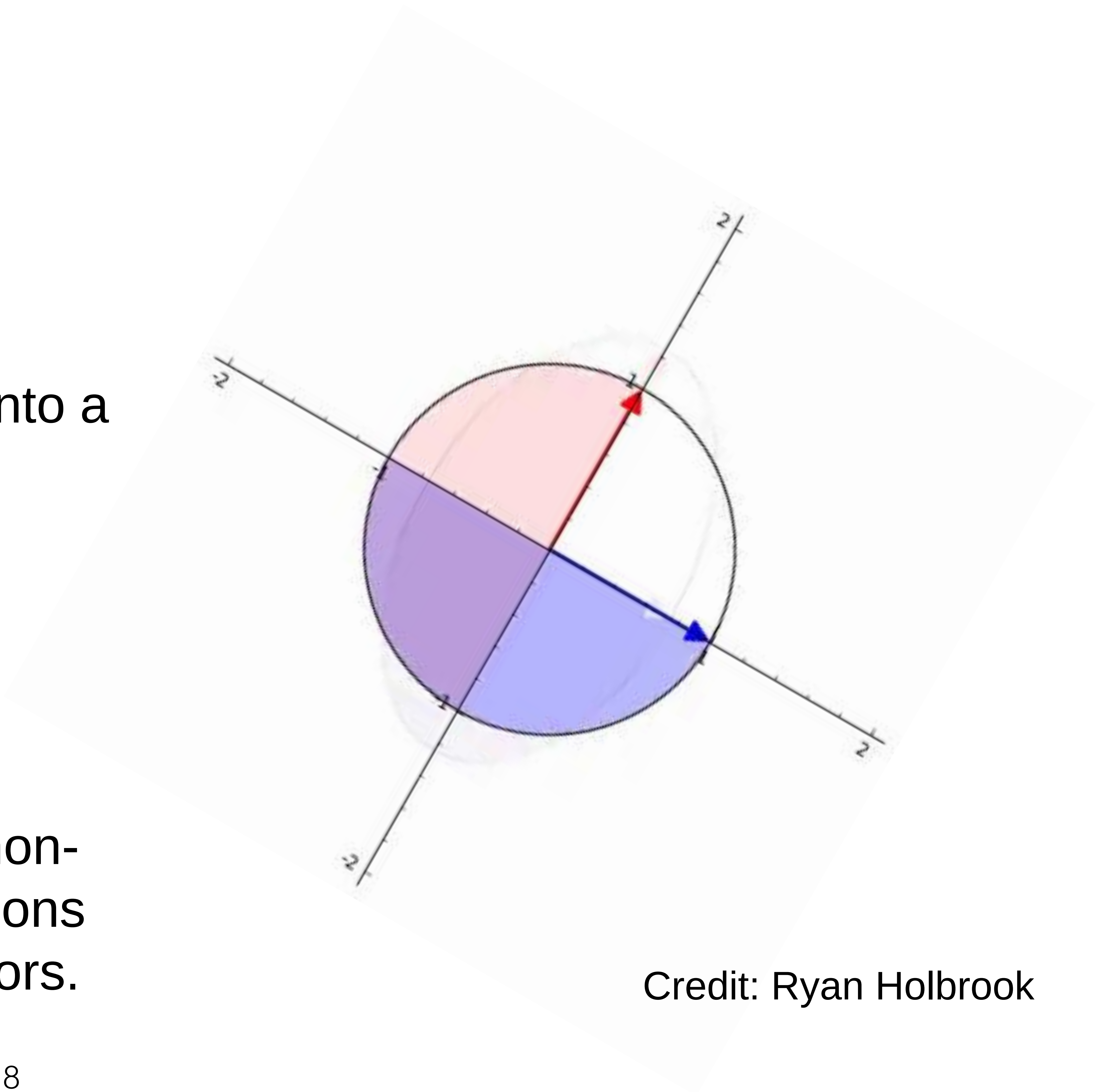
Since U is orthogonal, $U^{\top} = U^{-1}$

$$A = U\Sigma U^{-1}$$

All symmetric matrices can be decomposed into a sequence of:

1. Rotation
2. Scaling/reflection along axes
3. Reverse rotation

So any symmetric matrix essentially performs non-axis aligned scaling/reflection, where the directions along which scaling happens are the eigenvectors.

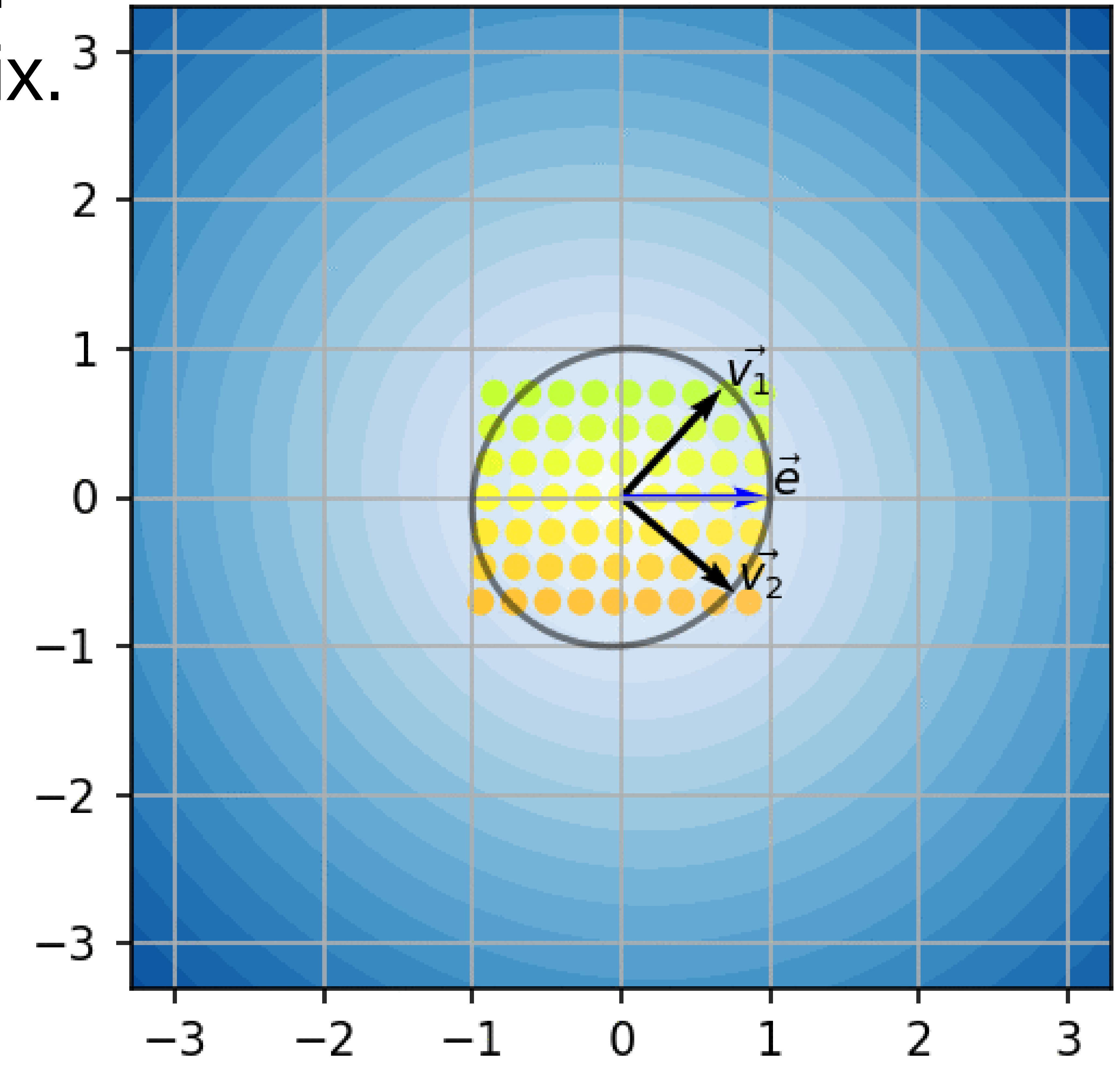


Credit: Ryan Holbrook

Eigenvectors vs. Right-Singular Vectors

Eigenvectors are the directions along which the vector retains its direction after being transformed by the matrix.

$$A\vec{u}_{.i} = \lambda_{ii}\vec{u}_{.i}$$



$\vec{v}_1$ - right-singular vector

$\vec{v}_2$ - second right-singular vector

$\vec{e}$ - eigenvector

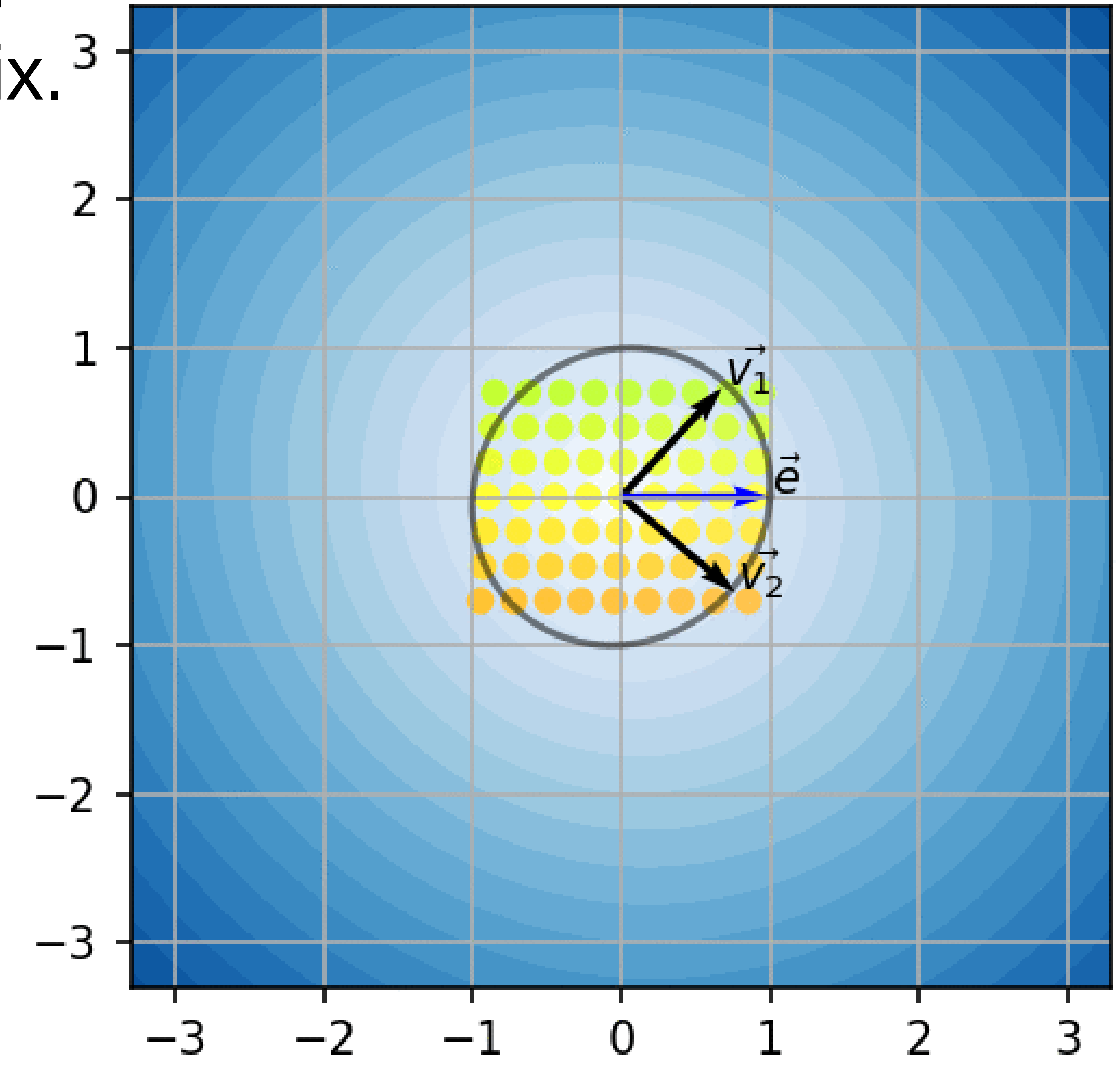
$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Eigenvectors vs. Right-Singular Vectors

Eigenvectors are the directions along which the vector retains its direction after being transformed by the matrix.

$$A\vec{u}_{.i} = \lambda_{ii}\vec{u}_{.i}$$

For asymmetric matrices, eigenvectors are not necessarily orthogonal; in this case, they are coincident



$\vec{v}_1$ - right-singular vector

$\vec{v}_2$ - second right-singular vector

$\vec{e}$ - eigenvector

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

Eigenvectors vs. Right-Singular Vectors

Eigenvectors are the directions along which the vector retains its direction after being transformed by the matrix.

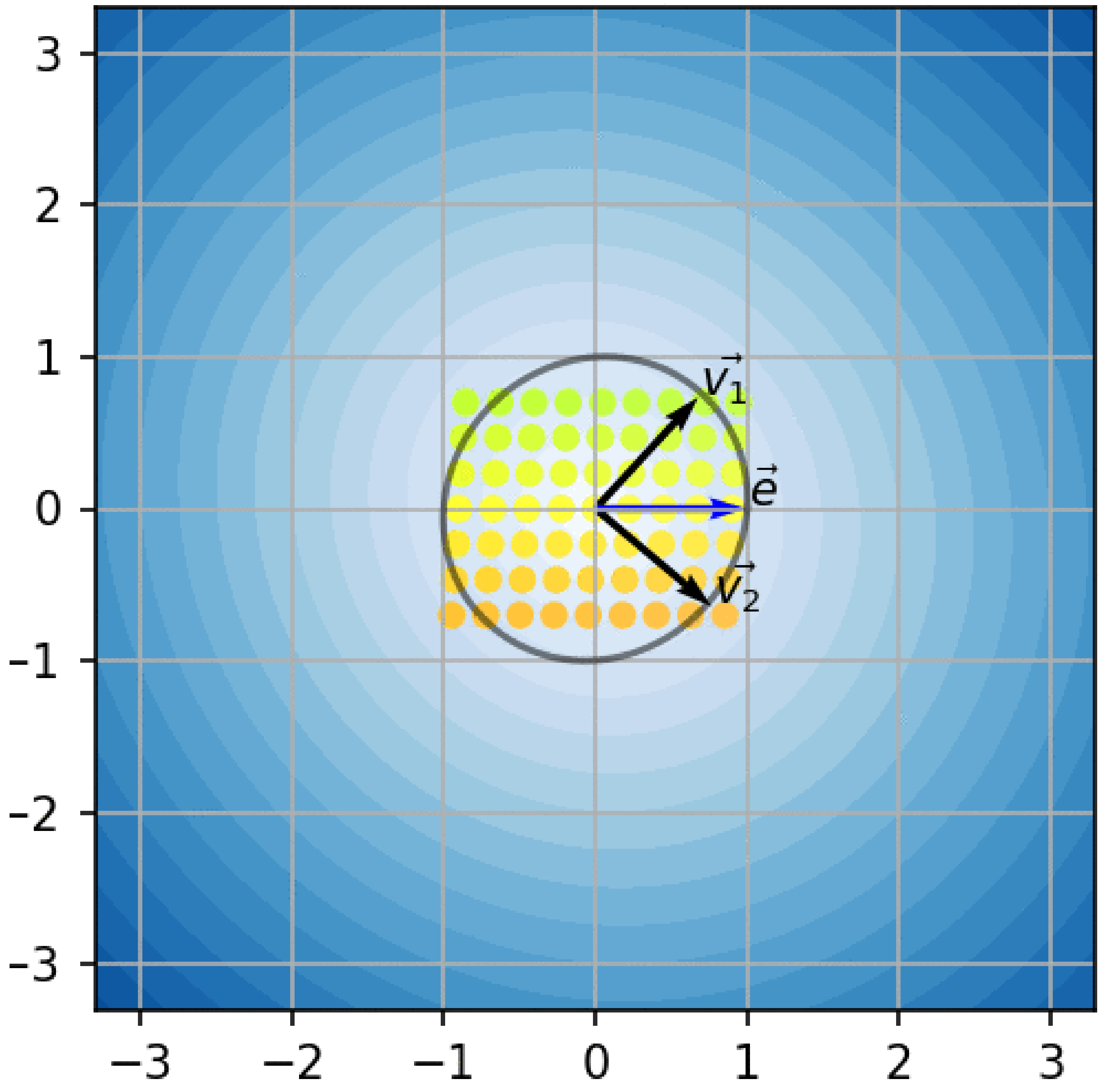
$$A\vec{u}_{\cdot i} = \lambda_{ii}\vec{u}_{\cdot i}$$

The right-singular vector with the largest singular value is the direction along which a unit vector becomes the longest after being transformed by the matrix.

$$\sigma_{1,1} = \max_{\vec{x}:\|\vec{x}\|_2=1} \| A\vec{x} \|_2$$

$$\vec{v}_{\cdot 1} = \arg\max_{\vec{x}:\|\vec{x}\|_2=1} \| A\vec{x} \|_2$$

For asymmetric matrices, eigenvectors are not necessarily orthogonal; in this case, they are coincident



$\vec{v_1}$ - right-singular vector

$\vec{v_2}$ - second right-singular vector

$\vec{e}$ - eigenvector

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

Eigendecomposition More Generally

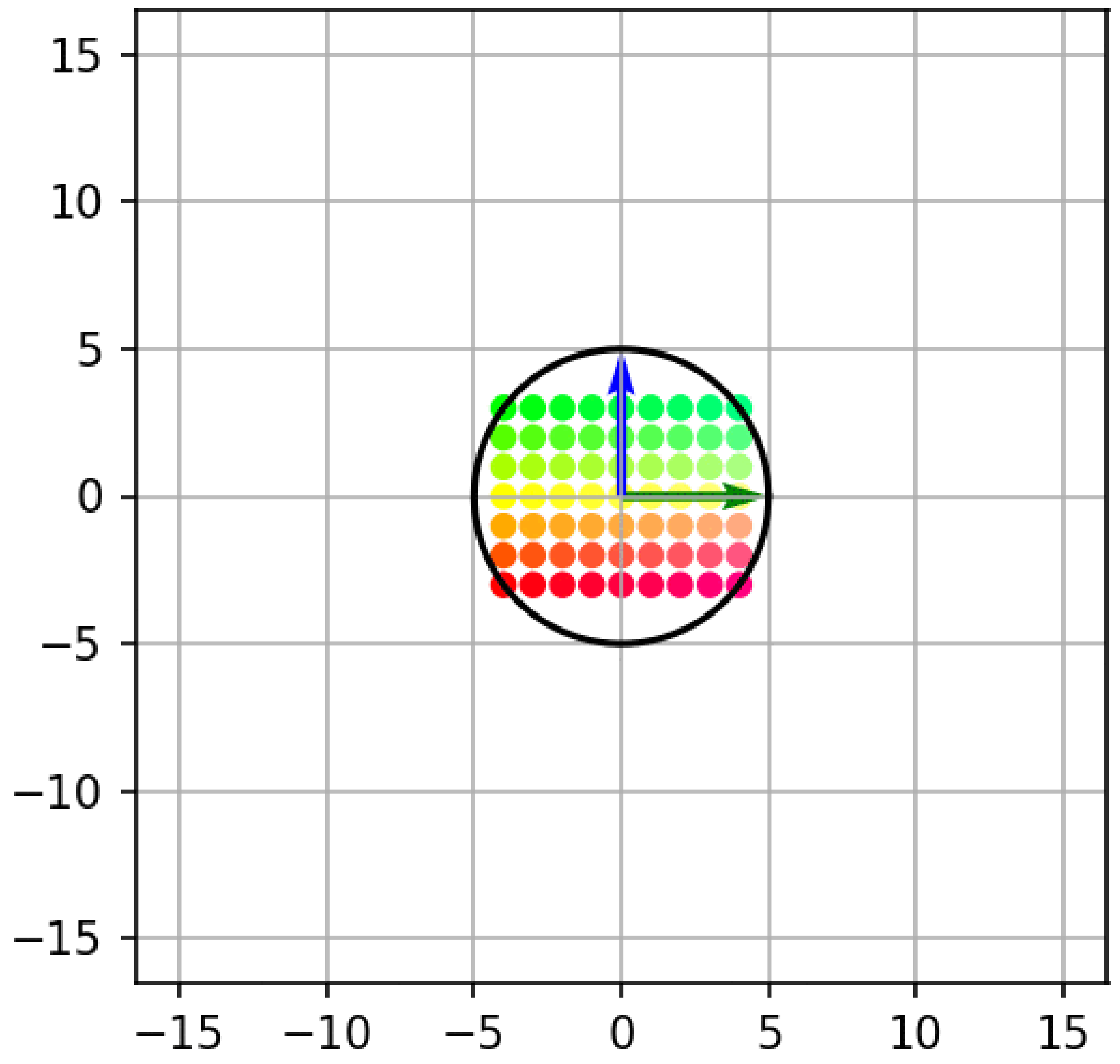
For asymmetric matrices, sometimes eigendecomposition is possible

- Only possible when the matrix is diagonalizable

In such cases:

Eigenvectors are not necessarily orthogonal
Eigenvalues and eigenvectors are not necessarily real
No straightforward geometric interpretation

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



$$A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$$

Eigendecomposition More Generally

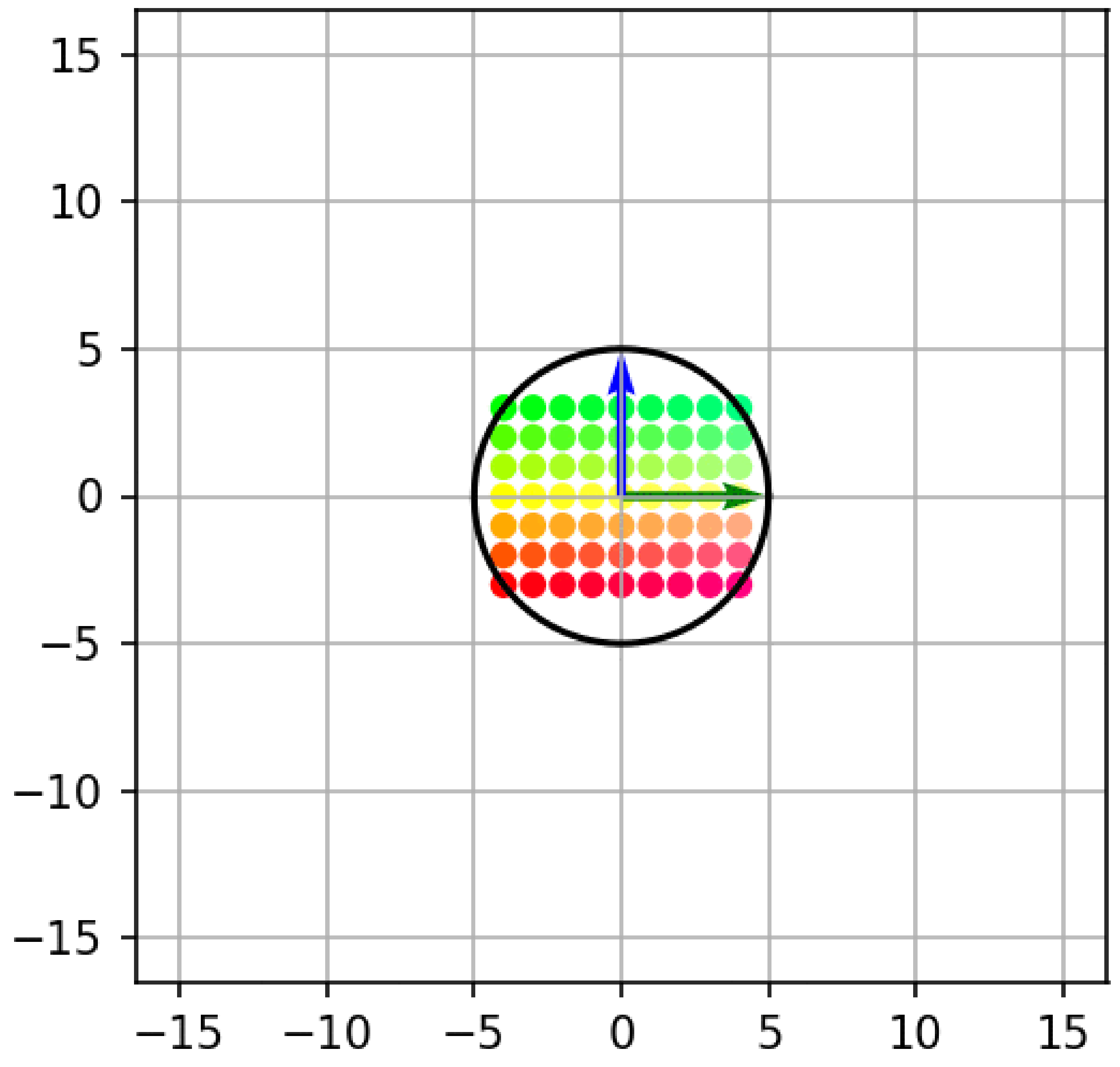
For asymmetric matrices, sometimes eigendecomposition is possible

- Only possible when the matrix is diagonalizable

In such cases:

Eigenvectors are not necessarily orthogonal
Eigenvalues and eigenvectors are not necessarily real
No straightforward geometric interpretation

$$A = U\Lambda U^{-1} \neq U\Lambda U^{\top}$$



$$A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$$

SVD and Eigendecomposition

SVD and eigendecomposition are closely related:

The right-singular vectors are eigenvectors of $A^\top A$.

The left-singular vectors are eigenvectors of $AA^\top$.

The non-zero singular values are the square roots of non-zero eigenvalues of $A^\top A$ (or equivalently the square roots of non-zero eigenvalues of $AA^\top$)

Application of Eigendecomposition

Finding the inverse of a symmetric matrix:

$$A = U\Lambda U^\top$$

$$A^{-1} = (U\Lambda U^\top)^{-1} = (U^\top)^{-1}\Lambda^{-1}U^{-1} = (U^\top)^\top\Lambda^{-1}U^\top = U\Lambda^{-1}U^\top$$

Since Λ is diagonal

$$\Lambda^{-1} = \begin{pmatrix} \frac{1}{\lambda_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_{nn}} \end{pmatrix}$$

Why?