Ch5 Lecture 2

Eigenvalues of Symmetric Matrices

Positive Definite Matrices

A matrix A is called **positive definite** if $x^T Ax > 0$ for all nonzero vectors x.

. . .

A symmetric matrix $K = K^T$ is positive definite if and only if all of its eigenvalues are strictly positive.

. . .

Proof: If $\mathbf{x} = \mathbf{v} \neq \mathbf{0}$ is an eigenvector with (necessarily real) eigenvalue λ, then

$$0 < \mathbf{v}^T K\mathbf{v} = \mathbf{v}^T(\lambda\mathbf{v}) = \lambda\mathbf{v}^T\mathbf{v} = \lambda\|\mathbf{v}\|^2$$

So $\lambda > 0$

. . .

Conversely, suppose K has all positive eigenvalues.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be the orthonormal eigenvector basis with $K\mathbf{u}_j = \lambda_j\mathbf{u}_j$ with $\lambda_j > 0$.

$$\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n, \quad \text{we obtain} \quad K\mathbf{x} = c_1\lambda_1\mathbf{u}_1 + \cdots + c_n\lambda_n\mathbf{u}_n$$

. . .

Therefore,

$$\mathbf{x}^T K\mathbf{x} = \left(c_1\mathbf{u}_1^T + \cdots + c_n\mathbf{u}_n^T\right)\left(c_1\lambda_1\mathbf{u}_1 + \cdots + c_n\lambda_n\mathbf{u}_n\right) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2 > 0$$

Let $A = A^T$ be a real symmetric $n \times n$ matrix. Then

(a) All the eigenvalues of A are real.

. . .

(b) Eigenvectors corresponding to distinct eigenvalues are orthogonal.

. . .

(c) There is an orthonormal basis of $\mathbb{R}^n$ consisting of n eigenvectors of A. In particular, all real symmetric matrices are non-defective and real diagonalizable.

Example

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

. . .

We compute the determinant in the characteristic equation

$$\det(A - \lambda \mathrm{I}) = \det \begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix} = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8$$

. . .

$$\lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2) = 0$$

. . .

Eigenvectors:

For the first eigenvalue, the eigenvector equation is

$$(A - 4\mathrm{I})\mathbf{v} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \text{or} \qquad \begin{aligned} -x + y &= 0 \\ x - y &= 0 \end{aligned}$$

General solution:

$$x = y = a, \quad \text{so} \quad \mathbf{v} = \begin{pmatrix} a \\ a \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

. . .

$$\lambda_1 = 4, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \lambda_2 = 2, \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

. . .

The eigenvectors are orthogonal: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$

Proof of part (a)

Let $A = A^T$ be a real symmetric $n \times n$ matrix. Then

(a) All the eigenvalues of A are real.

. . .

Suppose λ is a complex eigenvalue with complex eigenvector $\mathbf{v} \in \mathbb{C}^n$.

$$(A\mathbf{v}) \cdot \mathbf{v} = (\lambda \mathbf{v}) \cdot \mathbf{v} = \lambda \|\mathbf{v}\|^2$$

. . .

Now, if A is real and symmetric,

$$(A\mathbf{v}) \cdot \mathbf{w} = \left(\mathbf{v}^T A^T\right) \mathbf{w} = \mathbf{v} \cdot (A\mathbf{w}) \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{C}^n$$

Therefore

$$(A\mathbf{v}) \cdot \mathbf{v} = \mathbf{v} \cdot (A\mathbf{v}) = \mathbf{v} \cdot (\lambda \mathbf{v}) = \mathbf{v}^T \overline{\lambda \mathbf{v}} = \bar{\lambda} \|\mathbf{v}\|^2$$

. . .

$\Rightarrow$, $\lambda \|\mathbf{v}\|^2 = \bar{\lambda} \|\mathbf{v}\|^2 \Rightarrow \lambda = \bar{\lambda}$, so λ is real.

Proof of part (b)

Part b: Eigenvectors corresponding to distinct eigenvalues are orthogonal.

Suppose $A\mathbf{v} = \lambda\mathbf{v}, A\mathbf{w} = \mu\mathbf{w}$, where $\lambda \neq \mu$ are distinct real eigenvalues.

. . .

$\lambda\mathbf{v}\cdot\mathbf{w} = (A\mathbf{v})\cdot\mathbf{w} = \mathbf{v}\cdot(A\mathbf{w}) = \mathbf{v}\cdot(\mu\mathbf{w}) = \mu\mathbf{v}\cdot\mathbf{w}, \quad \text{and hence} \quad (\lambda-\mu)\mathbf{v}\cdot\mathbf{w} = 0.$

. . .

Since $\lambda \neq \mu$, this implies that $\mathbf{v}\cdot\mathbf{w} = 0$, so the eigenvectors $\mathbf{v}, \mathbf{w}$ are orthogonal.

Proof of part (c)

Part c: There is an orthonormal basis of $\mathbb{R}^n$ consisting of n eigenvectors of A.

. . .

If the eigenvalues are distinct, then the eigenvectors are orthogonal by part (b).

. . .

If the eigenvalues are repeated, then we can use the Gram-Schmidt process to orthogonalize the eigenvectors.

Diagonalization of Symmetric Matrices

Diagonalizability of Symmetric Matrices: the Spectral Theorem

- Every real, symmetric matrix admits an eigenvector basis, and hence is diagonalizable.
- Moreover, since we can choose eigenvectors that form an orthonormal basis, the diagonalizing matrix takes a particularly simple form.
- Recall that an $n \times n$ matrix Q is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$.

. . .

Writing our diagonalization from the previous lecture, specifically for the case of a real symmetric matrix A:

If $A = A^T$ is a real symmetric $n \times n$ matrix, then there exists an orthogonal matrix Q and a real diagonal matrix Λ such that

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

The eigenvalues of A appear on the diagonal of Λ, while the columns of Q are the corresponding orthonormal eigenvectors.

One example of a useful symmetric matrix: Quadratic Forms

A **quadratic form** is a homogeneous polynomial of degree 2 in n variables $x_1, \dots, x_n$. For example, in x, y, z: $Q(x, y, z) = ax^2 + by^2 + cz^2 + 2dxy + 2eyz + 2fzx$.

. . .

Every quadratic form can be written in matrix form as $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

. . .

Example:

$$Q(x, y, z) = x^2 + 2y^2 + z^2 + 2xy + yz + 3xz.$$

. . .

$$\begin{aligned} x(x+2y+3z) + y(2y+z) + z^2 &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} x+2y+3z \\ 2y+z \\ z \end{bmatrix} \\ &= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{x}^T A\mathbf{x}, \end{aligned}$$

Now, if we have a quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, we always write this in terms of an equivalent symmetric matrix B as $Q(\mathbf{x}) = \mathbf{x}^T B\mathbf{x}$ where $B = \frac{1}{2}(A + A^T)$.

(See Exercise 2.4.34 in your texbook.)

So in this case, we can write $Q(\mathbf{x}) = \mathbf{x}^T B\mathbf{x}$ where

$$B = \frac{1}{2} \begin{bmatrix} 2 & 2 & 3 \\ 2 & 4 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

Check:

```
import sympy as sp
B = sp.Matrix([[2,2,3],[2,4,1],[3,1,2]])/2
x = sp.Matrix(sp.symbols(['x','y','z']))
(x.T*B*x).expand()
```

$\left[x^2 + 2xy + 3xz + 2y^2 + yz + z^2\right]$

Yes, this is the same as the original quadratic form.

Diagonalizing Quadratic Forms

Symmetric matrices are diagonalizable ⇒ we can always find a basis in which the quadratic form takes a particularly simple form. Just diagonalize:

. . .

$Q(\mathbf{x}) = \mathbf{x}^T B\mathbf{x} = x^T PDP^T x$ where P is the matrix of eigenvectors of B and D is the diagonal matrix of eigenvalues of B.

. . .

Then if we define new variables $\mathbf{y} = P^T\mathbf{x}$, we have $Q(\mathbf{x}) = \mathbf{y}^T D\mathbf{y}$

. . .

which just becomes a sum of squares:

$$q(\mathbf{x}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

Example:

Suppose we have the quadratic form $3x^2 + 2xy + 3y^2$. We can write this in matrix form as $Q(\mathbf{x}) = \mathbf{x}^T B\mathbf{x}$ where $\mathbf{x} = \binom{x_1}{x_2}$ and

$$B = \frac{1}{2}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

. . .

We diagonalize B:

$$\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = A = PDP^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

. . .

Now, if we define $\mathbf{y} = P^T\mathbf{x} = \frac{1}{\sqrt{2}}\binom{x_1+x_2}{-x_1+x_2}$, we have $Q(\mathbf{x}) = \mathbf{y}^T D\mathbf{y}$, or

$$q(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + 3x_2^2 = 4y_1^2 + 2y_2^2$$

The numbers aren't always clean, though!

```
Q,Lambda = B.diagonalize()
Q
```

$$\begin{bmatrix} \frac{-13248\cdot(1+\sqrt{3}i)(53+9\sqrt{1167}i)^{\frac{2}{3}} - \sqrt[3]{53+9\sqrt{1167}i}\left(184+\left(-16+(1+\sqrt{3}i)\sqrt[3]{53+9\sqrt{1167}i}\right)(1+\sqrt{3}i)\sqrt[3]{53+9\sqrt{1167}i}\right)^2 + 72(1+\sqrt{3}i)^2\cdot\left(11-(1+\sqrt{3}i) \cdots\right.}{792(1+\sqrt{3}i)^2\cdot(53+9\sqrt{1167}i)} & \cdots \\ \frac{28\left(-22+(1+\sqrt{3}i)\sqrt[3]{53+9\sqrt{1167}i}\right)(1+\sqrt{3}i)^2\cdot(53+9\sqrt{1167}i)+\sqrt[3]{53+9\sqrt{1167}i}\left(184+\left(-16+(1+\sqrt{3}i)\sqrt[3]{53+9\sqrt{1167}i}\right)(1+\sqrt{3}i)\sqrt[3]{53+9\sqrt{1167}i}\right)^2+5\cdots}{264(1+\sqrt{3}i)^2\cdot(53+9\sqrt{1167}i)} & \cdots \\ 1 & \cdots \end{bmatrix}$$

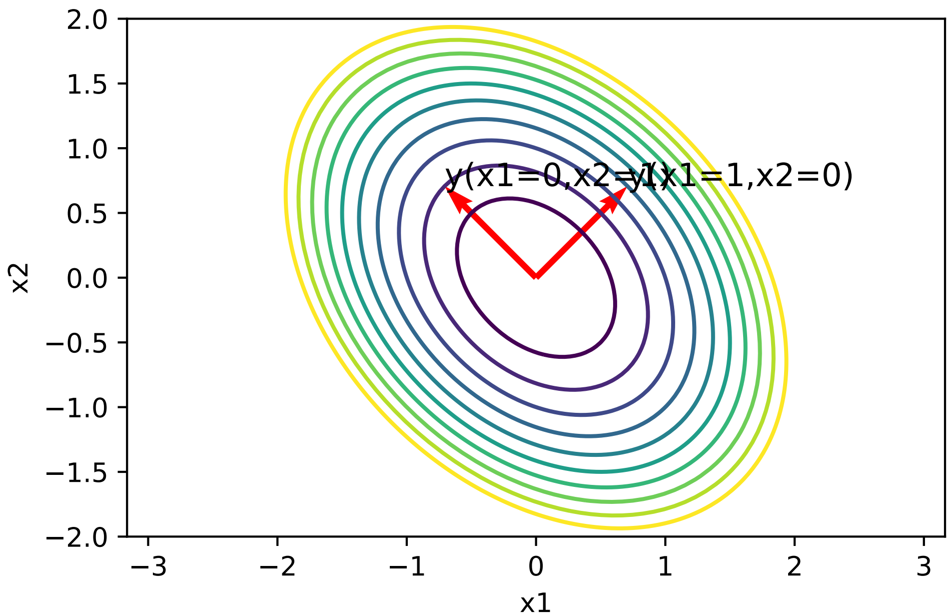
```
Lambda
```

$$\begin{bmatrix} \frac{4}{3} + \left(-\frac{1}{2} - \frac{\sqrt{3}i}{2}\right)\sqrt[3]{\frac{53}{216} + \frac{\sqrt{1167}i}{24}} + \frac{23}{18\left(-\frac{1}{2}-\frac{\sqrt{3}i}{2}\right)\sqrt[3]{\frac{53}{216}+\frac{\sqrt{1167}i}{24}}} & 0 & \cdots \\ 0 & \frac{4}{3} + \frac{23}{18\left(-\frac{1}{2}+\frac{\sqrt{3}i}{2}\right)\sqrt[3]{\frac{53}{216}+\frac{\sqrt{1167}i}{24}}} + \left(-\frac{1}{2} + \frac{\sqrt{3}i}{2}\right)\sqrt[3]{\frac{53}{216} + \frac{\sqrt{1\cdots}}{}} & \cdots \\ 0 & 0 & \cdots \end{bmatrix}$$

Yuck!

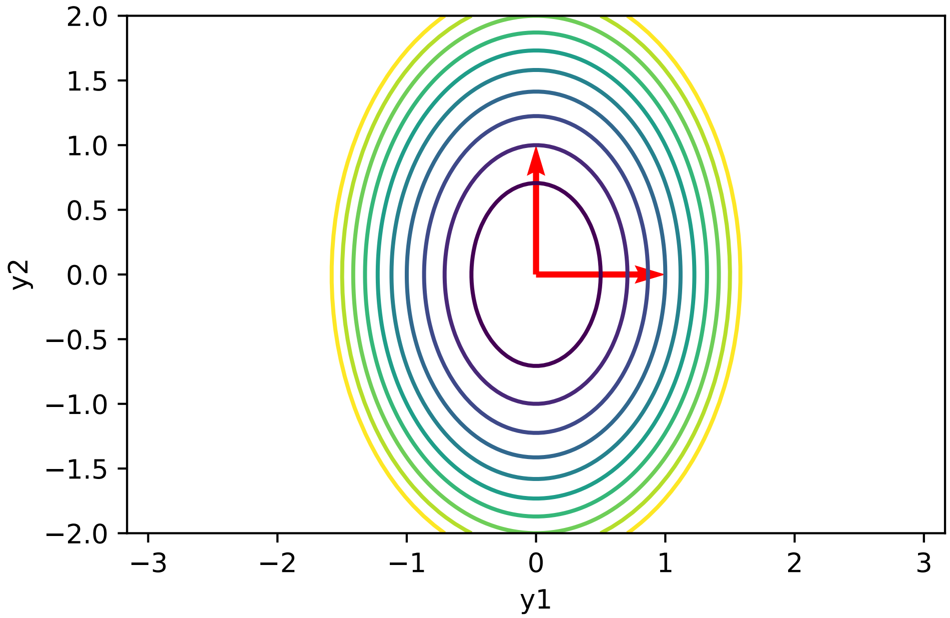
We can visualize the previous example as a *rotation* of the axes (a change of basis) to a new coordinate system where the quadratic form is just a sum of squares.

```
import numpy as np
import matplotlib.pyplot as plt
x = np.linspace(-2,2,100)
y = np.linspace(-2,2,100)
X,Y = np.meshgrid(x,y)
Z = 3*X**2+2*X*Y+3*Y**2
plt.contour(X,Y,Z,levels=[1,2,3,4,5,6,7,8,9,10])
plt.xlabel('x1')
plt.ylabel('x2')
plt.axis('equal')
# plot the vector P.T times (1,0) and (0,1)
P = np.array([[1/np.sqrt(2),1/np.sqrt(2)],[-1/np.sqrt(2),1/np.sqrt(2)]])
v1 = P.T @ np.array([1,0])
v2 = P.T @ np.array([0,1])
plt.quiver(0,0,v1[0],v1[1],angles='xy',scale_units='xy',scale=1,color='r')
plt.quiver(0,0,v2[0],v2[1],angles='xy',scale_units='xy',scale=1,color='r')
# label the two quivers ("x1=1, x2=0" and "x1=0, x2=1")
plt.text(v1[0],v1[1],'y(x1=1,x2=0)',fontsize=12)
plt.text(v2[0],v2[1],'y(x1=0,x2=1)',fontsize=12)
plt.show()
```



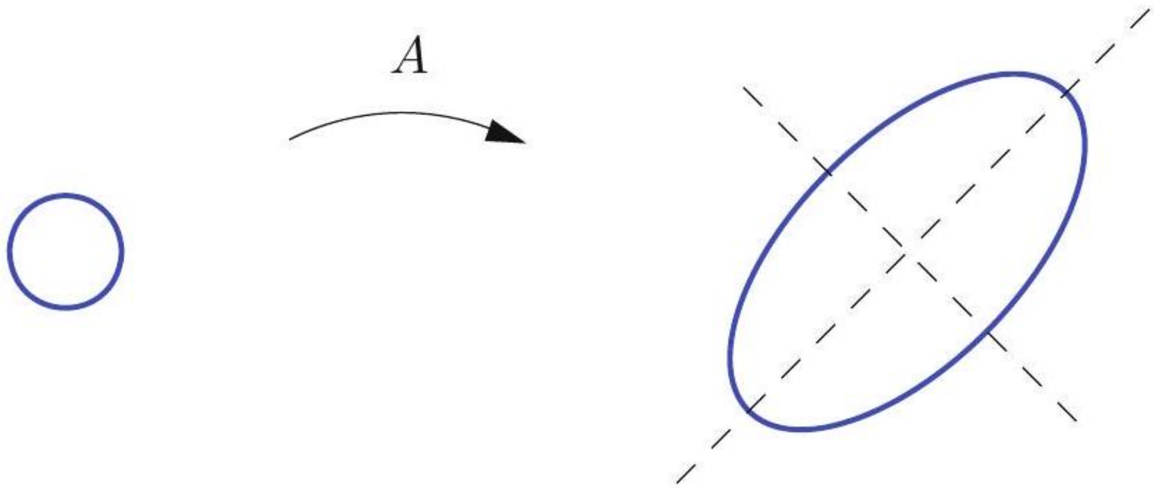
Now make the same plot but in y coordinates:

```
x = np.linspace(-2,2,100)
y = np.linspace(-2,2,100)
X,Y = np.meshgrid(x,y)
Z = 4*X**2+2*Y**2
plt.contour(X,Y,Z,levels=[1,2,3,4,5,6,7,8,9,10])
plt.xlabel('y1')
plt.ylabel('y2')
plt.axis('equal')
v1 = P.T @ np.array([1,0])
v2 = P.T @ np.array([0,1])
plt.quiver(0,0,1,0,angles='xy',scale_units='xy',scale=1,color='r')
plt.quiver(0,0,0,1,angles='xy',scale_units='xy',scale=1,color='r')
# label the two quivers ("x1=1, x2=0" and "x1=0, x2=1")
plt.show()
```



In general, we can think of the diagonalization as a *rotation* of the axes followed by a *scaling* of the axes.

We often visualize this by plotting the effects of the transformations on the unit circle.



The SVD

Singular Values

We've talked a lot about eigenvalues and eigenvectors, but these only make any sense for square matrices. What can we do for a general $m \times n$ matrix A?

. . .

It turns out we can learn a lot from the matrix $A^T A$ (or AA^T), which is always square and symmetric.

. . .

The **singular values** $\sigma_1, \dots, \sigma_r$ of an $m \times n$ matrix A are the positive square roots, $\sigma_i = \sqrt{\lambda_i} > 0$, of the nonzero eigenvalues of the associated "Gram matrix" $K = A^T A$.

. . .

The corresponding eigenvectors of K are known as the **singular vectors** of A.

. . .

All of the eigenvalues of K are real and nonnegative – but some may be zero.

. . .

If $K = A^T A$ has repeated eigenvalues, the singular values of A are repeated with the same multiplicities.

. . .

The number r of singular values is equal to the rank of the matrices A and K.

Example

Let $A = \begin{pmatrix} 3 & 5 \\ 4 & 0 \end{pmatrix}$.

$$K = A^T A = \begin{pmatrix} 3 & 4 \\ 5 & 0 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} 25 & 15 \\ 15 & 25 \end{pmatrix}$$

. . .

This has eigenvalues $\lambda_1 = 40, \lambda_2 = 10$, and corresponding eigenvectors $\mathbf{v}_1 = \binom{1}{1}, \mathbf{v}_2 = \binom{1}{-1}$.

. . .

Therefore, the singular values of A are $\sigma_1 = \sqrt{40} = 2\sqrt{10}, \sigma_2 = \sqrt{10}$.

pause

Singular Values of a Symmetric Matrix

Singular Value Decomposition

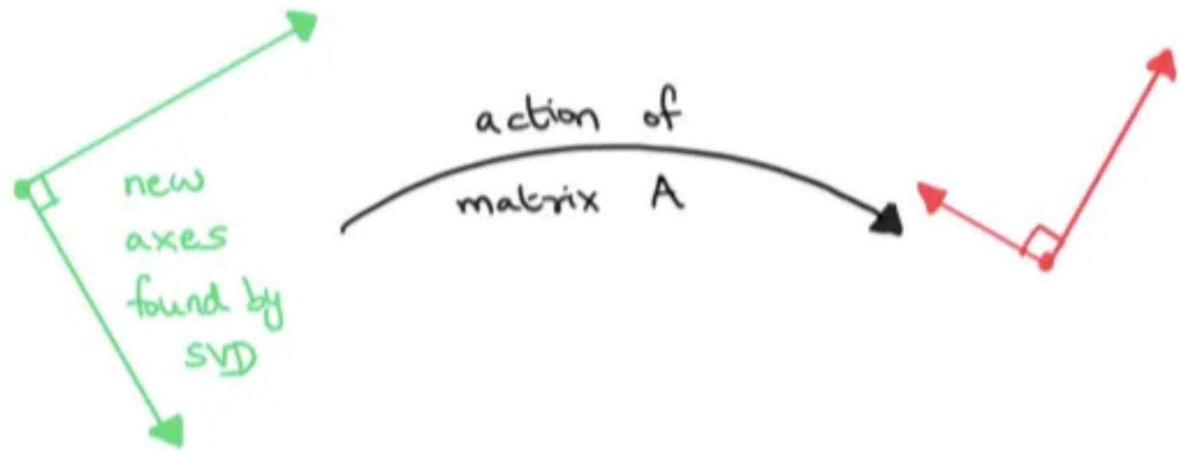
Let A be an $m \times n$ real matrix. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ diagonal matrix Σ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, with $p = \min\{m, n\}$, such that $U^T A V = \Sigma$. Moreover, the numbers $\sigma_1, \sigma_2, \ldots, \sigma_p$ are uniquely determined by A.

Proof:

Geometric interpretation of the SVD

(following closely this blog post)

Goal: to understand the SVD as finding perpendicular axes that remain perpendicular after a transformation.



Take a very simple matrix:

$$A = \begin{pmatrix} 6 & 2 \\ -7 & 6 \end{pmatrix}$$

. . .

Represents a linear map $\mathrm{T} : \mathrm{R}^2 \to \mathbf{R}^2$ with respect to the standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$.

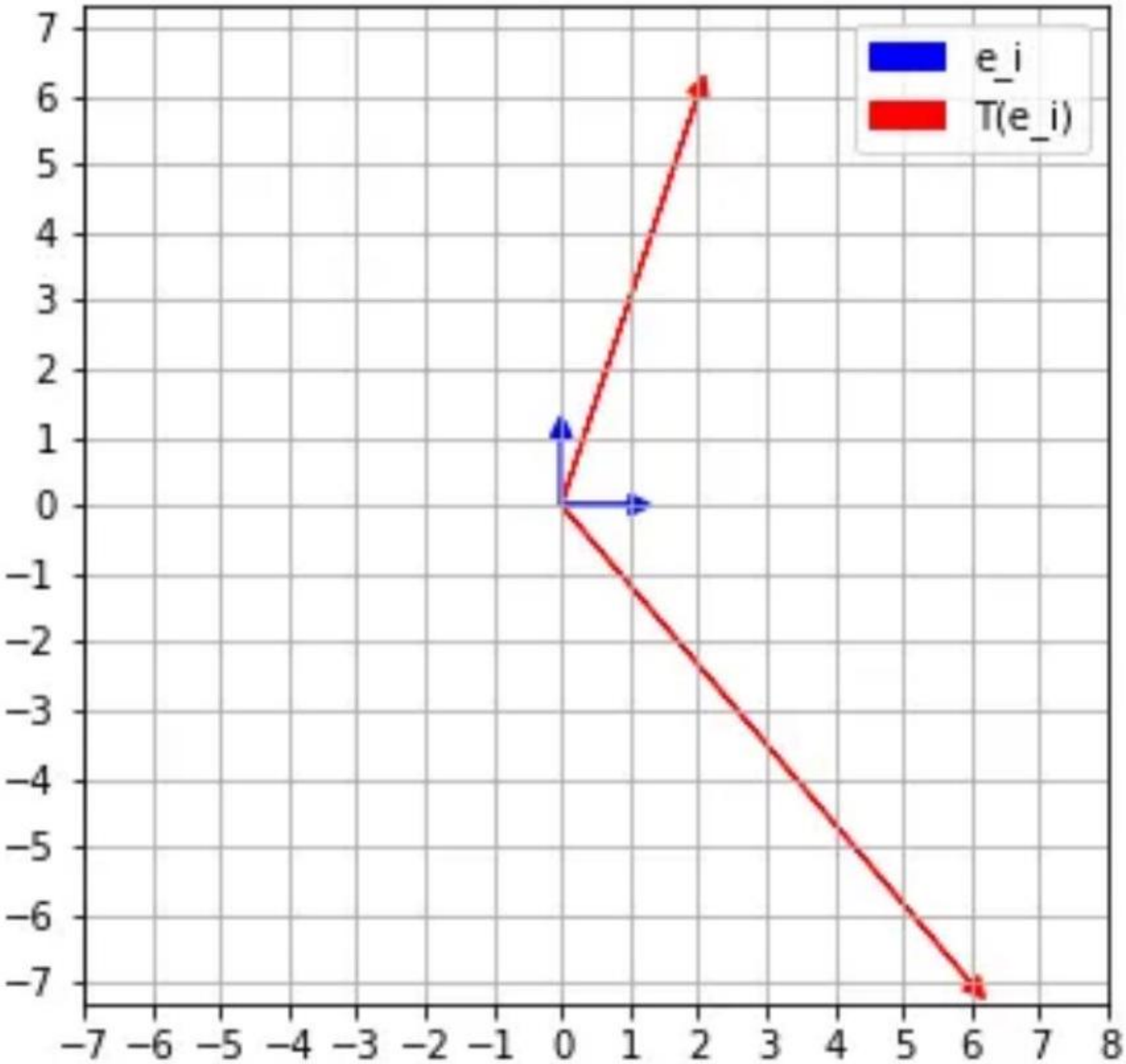
. . .

Sends the usual basis elements $e_1 \rightsquigarrow (6, -7)$ and $e_2 \rightsquigarrow (2, 6)$.

We can see what T does to any vector:

$$
\begin{aligned}
T(2,3) &= T(2,0) + T(0,3) = T\left(2e_1\right) + T\left(3e_2\right) \\
&= 2T\left(e_1\right) + 3T\left(e_2\right) = 2(6,-7) + 3(2,6) \\
&= (18,4)
\end{aligned}
$$

The usual basis elements are no longer perpendicular after the transformation:



Given a linear map T:Rm $\rightarrow R^n$ (which is represented by a $n \times m$ matrix A with respect to the standard basis),

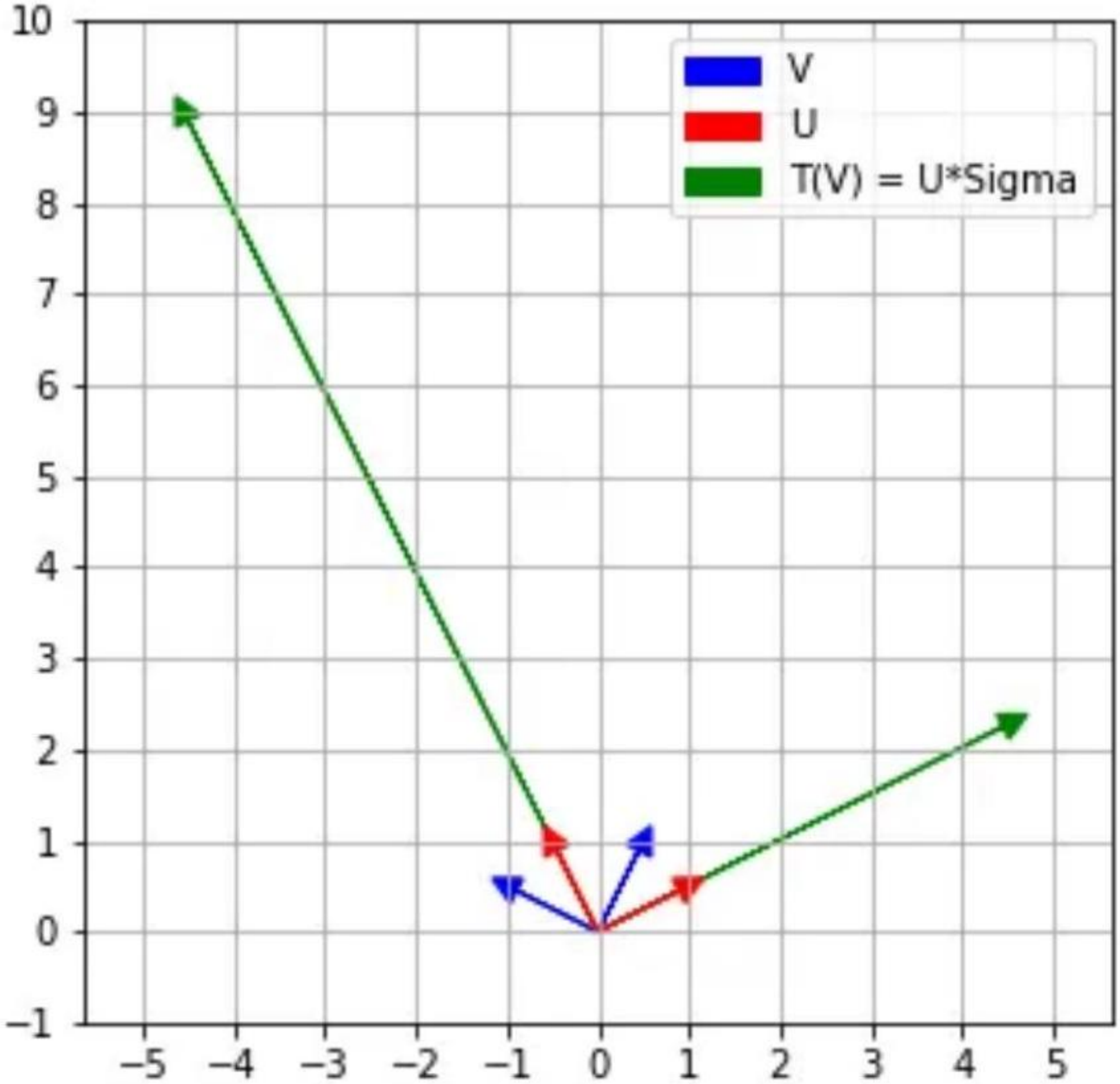
- can we find an orthonormal basis $\{v\_1, v\_2, \ldots, v\_m\}$ of R^m such that
- $\{T(v\_1), T(v\_2), \ldots$ T (v_m)} are still mutually perpendicular to each other?

. . .

This is what the SVD does!

Let the SVD of A be $A = U\Sigma V^T$

. . .



How are U and V related?