THE UNIVERSITY OF CHICAGO

MSCA 37016
Advanced Linear Algebra for Machine Learning

Lecture 4

Danny Ng

Principal Component Analysis PCA

- **Principal component analysis** PCA is:
  1. Dimension reduction technique
  2. Exploit correlation / **covariance** among a set of data points
  3. **Project** the data points into a suitable lower-dimensional space to reveal any low-dimensional structure of the data
  4. Retain as much of the **variation** in the data as possible
- It is not about predicting one variable in the data by others as in regression

- **Step 1:** Start with matrix of n observed data points in $\mathbb{R}^d$

one data point

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \cdots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix} \in \mathbb{R}^{n \times d}$$

- **Step 2:** Compute the j^{th} column mean

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

- **Step 3:** Centering the data per column

$$X_c = X - \begin{pmatrix} \mu_1 & \cdots & \mu_d \\ \vdots & \cdots & \vdots \\ \mu_1 & \cdots & \mu_d \end{pmatrix}$$

Variance = Norm²

- **Variance** of an explanatory variables i.e. a column of X_c

$$\vec{x}_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

$$\begin{aligned} \mathrm{Var}(X_j) &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-\mu_j\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-0\right)^2 = \frac{1}{n}\left\|\vec{x}_j\right\|^2 \end{aligned}$$

- Standard deviation is

$$SD(X_j) = \frac{1}{\sqrt{n}}\left\|\vec{x}_j\right\|$$

Covariance = Dot Product

- **Covariance** of two explanatory variables i.e. two columns of X_c

$$\vec{x}_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix} \quad \text{and} \quad \vec{x}_k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix}$$

$$\begin{aligned} \mathrm{Cov}(X_j, X_k) &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-\mu_j\right)(x_{ik}-\mu_k) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-0\right)(x_{ik}-0) \\ &= \frac{1}{n}\left(\vec{x}_j \cdot \vec{x}_k\right) \end{aligned}$$

Covariance Properties

- Variance and covariance

$$\operatorname{Var}(X_j) = \frac{1}{n}\|\vec{x}_j\|^2$$
$$= \frac{1}{n}(\vec{x}_j \cdot \vec{x}_j) = \operatorname{Cov}(X_j, X_j)$$

- Symmetry of covariance

$$\operatorname{Cov}(X_j, X_k) = \frac{1}{n}(\vec{x}_j \cdot \vec{x}_k)$$
$$= \frac{1}{n}(\vec{x}_k \cdot \vec{x}_j) = \operatorname{Cov}(X_k, X_j)$$

Uncorrelated = Orthogonal

- If two centered i.e. mean 0 explanatory variables are **uncorrelated**, then

$$0 = \operatorname{Cov}(X_j, X_k) = \frac{1}{n}(\vec{x}_j \cdot \vec{x}_k)$$

i.e. the two column vectors $\vec{x}_j, \vec{x}_k \in \mathbb{R}^n$ are orthogonal

$$\vec{x}_j \perp \vec{x}_k$$

- Two perspectives of data matrix $X \in \mathbb{R}^{n \times d}$:
  1. n rows of data points in $\mathbb{R}^d$
  2. d column vectors of explanatory variables in $\mathbb{R}^n$

Covariance Matrix

- **Covariance matrix** of d explanatory variables $X = (X_1, \dots, X_d)$ is

$$\mathrm{Cov}(X) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_2, X_1) & \ddots & & \vdots \\ \vdots & & \ddots & \mathrm{Cov}(X_{d-1}, X_d) \\ \mathrm{Cov}(X_d, X_1) & \cdots & \mathrm{Cov}(X_d, X_{d-1}) & \mathrm{Var}(X_d) \end{pmatrix} \in \mathbb{R}^{d \times d}$$

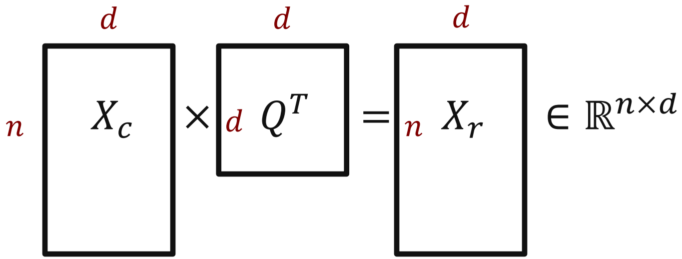
- Computationally, each entry $\mathrm{Cov}(X)_{jk} = \mathrm{Cov}\left(X_j, X_k\right) = \frac{1}{n}\left(\vec{x}_j \cdot \vec{x}_k\right)$. Therefore,

$$\mathrm{Cov}(X) = \frac{1}{n} \begin{pmatrix} - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_d & - \end{pmatrix} \begin{pmatrix} | & & | \\ \vec{x}_1 & \dots & \vec{x}_d \\ | & & | \end{pmatrix} = \frac{1}{n} X_c^T X_c$$

- $\boxed{\text{Covariance matrix is always symmetric}}$

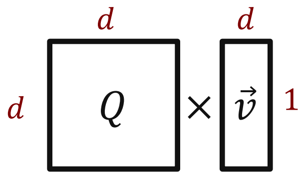
Principal Component Analysis PCA (cont'd)

- **Step 4:** Find orthogonal matrix $Q \in \mathbb{R}^{d\times d}$ to rotate the centered data X_c such that the rotated data X_r look "axis-aligned"

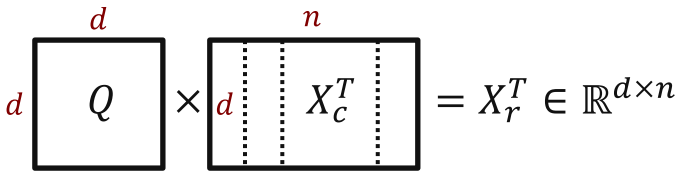


- **Question:** Why is X_c multiplied on the right side by Q^T instead of on the left side by Q?

- Linear transformation (e.g. rotation) is usually a matrix-vector multiplication



- Here each data point "vector" is a row of X_c so we need to do



$$X_c Q^T = X_r \in \mathbb{R}^{n\times d}$$

Principal Component Analysis PCA (cont’d)

- **Step 5:** Further decompose the rotated data X_r into unit vectors U and norms D

$$X_c Q^T = \overbrace{X_r}^{\text{look “axis-aligned”}} = \overbrace{\begin{pmatrix} | & & | \\ \vec{u}_1 & \dots & \vec{u}_d \\ | & & | \end{pmatrix}}^{\substack{\text{all } d \text{ column vectors are statistically} \\ \text{uncorrelated i.e. } \vec{u}_j \perp \vec{u}_k \in \mathbb{R}^n}} = \underbrace{\begin{pmatrix} | & & | \\ \hat{u}_1 & \dots & \hat{u}_d \\ | & & | \end{pmatrix}}_{\substack{U \in \mathbb{R}^{n \times d} \\ d \text{ orthonormal column vectors in } \mathbb{R}^n}} \underbrace{\begin{pmatrix} \|\vec{u}_1\| & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \|\vec{u}_d\| \end{pmatrix}}_{\substack{D \in \mathbb{R}^{d \times d} \\ \text{diagonal matrix of the norms}}} \in \mathbb{R}^{n \times d}$$

- In other words, PCA is a decomposition of the data matrix $X_c = X_r Q = UDQ$

Singular Value Decomposition SVD

- **Fact:** Any matrix $A \in \mathbb{R}^{n\times d}$ can be decomposed into a product of 3 "nice" matrices

singular values of A

$$A = UDV^T = \underbrace{\begin{pmatrix} | & & | \\ \hat{u}_1 & \dots & \hat{u}_d \\ | & & | \end{pmatrix}}_{\substack{U \in \mathbb{R}^{n\times d} \\ d \text{ orthonormal column vectors in } \mathbb{R}^n}} \underbrace{\begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d \end{pmatrix}}_{\substack{D \in \mathbb{R}^{d\times d} \\ \text{diagonal matrix}}} \underbrace{\begin{pmatrix} - & \hat{v}_1 & - \\ & \vdots & \\ - & \hat{v}_d & - \end{pmatrix}}_{\substack{V^T \in \mathbb{R}^{d\times d} \\ \text{orthogonal matrix}}}$$

- This is the most powerful matrix decomposition because it is applicable to any matrix and reveals a lot of the structure of a matrix!

Singular Value vs Rank

- The diagonal entries of

$$D = \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d \end{pmatrix} \in \mathbb{R}^{d \times d}$$

are the **singular values** of matrix A

- **Fact:**

$$\mathrm{rank}(A) = \begin{pmatrix} \text{number of non−zero} \\ \text{singular values} \end{pmatrix}$$

SVD and Invertibility

- **Fact:** For square matrix $A \in \mathbb{R}^{d \times d}$
  1. U is also an orthogonal matrix
  2. A is invertible if and only if all d singular values $\sigma_i \neq 0$
  3. In this case, the matrix inverse A^{-1} is simply

$$\begin{aligned} A^{-1} &= (UDV^T)^{-1} \\ &= (V^T)^{-1} D^{-1} U^{-1} \\ &= V \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 1/\sigma_d \end{pmatrix} U^T \end{aligned}$$

Non-Uniqueness of Singular Value Decomposition

- Different computational softwares might yield some variation e.g.
  1. Negate $\hat{u}_i$ to $-\hat{u}_i$ as long as the corresponding σ_i becomes $-\sigma_i$
  2. Permute the columns of U and V and the diagonal entries of D together

  without affecting the overall matrix product

$$A = UDV^T = \begin{pmatrix} | & & | \\ \hat{u}_1 & \dots & \hat{u}_d \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d \end{pmatrix} \begin{pmatrix} - & \hat{v}_1 & - \\ & \vdots & \\ - & \hat{v}_d & - \end{pmatrix}$$

- Convention:
  1. If $\sigma_i < 0$, then change $\hat{u}_i$ to $-\hat{u}_i$ so the singular value $-\sigma_i > 0$ is always positive
  2. Re-order the columns of U and V so that $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$ are sorted descending

Principal Component Analysis via SVD

- Perform PCA by computing SVD of the centered data matrix

$$X_c = X_r Q = UDQ = UDV^T$$

$$\sigma_1 \geq \cdots \geq \sigma_d \geq 0$$

- **Principal components:** Orthonormal basis $\{\hat{v}_1, \ldots, \hat{v}_d\}$ of the rotation matrix

$$Q^T = V = \begin{pmatrix} | & & | \\ \hat{v}_1 & \cdots & \hat{v}_d \\ | & & | \end{pmatrix} \in \mathbb{R}^{d \times d}$$

are the major / minor axes of the data point ellipsoid

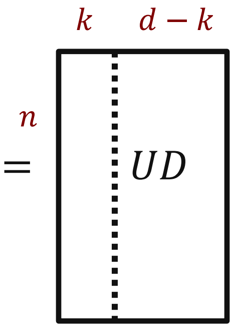
- **Variance:** Singular value

$$\sigma_i = \|\vec{u}_i\|$$

is (proportional to) the data variation along the i^{th} principal component axis

- **Dimension reduction:** Keep only the first k columns of X_r (with highest variances σ_i's)

$X_r =$



$\in \mathbb{R}^{n \times d}$

Eigenvalue and Eigenvector

- Let $A \in \mathbb{R}^{d\times d}$ be a square matrix. If a non-zero vector $\vec{v} \in \mathbb{R}^d$ satisfies

  $$A\vec{v} = \lambda\vec{v}$$

  for some $\lambda \in \mathbb{R}$, then $\vec{v}$ is an **eigenvector** for A with corresponding **eigenvalue** λ

- Geometrically, A transforms $\vec{v}$ by simply scaling it by λ

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \lambda = -4$$

e.g.

$$A = \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix} \quad \vec{v} = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \quad \lambda = 0$$

e.g.

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \quad \vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \lambda_1 = 4$$

$$\vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \lambda_2 = -2$$

Finding Eigenvalue

- **Question:** How to find all eigenvalues (if any) of A?

- The following statements are equivalent:
  1. λ is an eigenvalue of A
  2. $A\vec{v} = \lambda\vec{v}$ for some $\vec{v} \neq \vec{0}$
  3. Homogeneous system
  $$(A - \lambda I_d)\vec{v} = \vec{0}$$
  has a non-trivial $\vec{v} \neq \vec{0}$ solution
  4. Matrix $A - \lambda I_d$ is non-invertible
  5. $\det(A - \lambda I_d) = 0$

Example of $A - \lambda I_d$

e.g.

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \qquad \vec{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \lambda = 4$$

$$\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 4 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} 1 - 4 & 3 \\ 3 & 1 - 4 \end{pmatrix}}_{A - \lambda I_d} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Characteristic Polynomial

- The function

$$f(\lambda) = \det(A - \lambda I_d)$$

  is called the **characteristic polynomial** of matrix $A \in \mathbb{R}^{d \times d}$

- **Facts:**
  1. It is a d^{th} degree polynomial
  2. Roots / zeros are eigenvalues of A
  3. There can be at most d distinct roots i.e. eigenvalues
  4. No simple formula to solve for roots of 5^{th}+ degree polynomial

Eigenvalue Examples

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix}$$

$$f(\lambda) = \det(A - \lambda I_2)$$

$$= \det\begin{pmatrix} 2-\lambda & 3 \\ 4 & -2-\lambda \end{pmatrix} = \lambda^2 - 16$$

It has 2 distinct eigenvalues

$$\lambda_1 = 4 \qquad \lambda_2 = -4$$

Eigenvalue Examples

e.g.

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

$$f(\lambda) = \det \begin{pmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{pmatrix} = (2-\lambda)^2$$

Only one eigenvalue $\lambda = 2$

e.g.

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

$$f(\lambda) = \det \begin{pmatrix} 0-\lambda & 1 \\ -1 & 0-\lambda \end{pmatrix} = \lambda^2 + 1$$

No (real number) eigenvalue

Triangular Matrix Eigenvalues

- For upper or lower triangular (including diagonal) matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

- Characteristic polynomial is

$$\begin{aligned} f(\lambda) &= \det(A - \lambda I_d) \\ &= (a_{11} - \lambda) \times \cdots \times (a_{dd} - \lambda) \end{aligned}$$

- Eigenvalues are the diagonal entries of A

$$\lambda_1 = a_{11}, \ldots, \lambda_d = a_{dd}$$

Finding Eigenvector

- **Question:** If $A \in \mathbb{R}^{d\times d}$ has eigenvalue λ, how to find all corresponding eigenvectors $\vec{v}$?

- **Answer:** Just non-zero solutions of the homogeneous system

$$(A - \lambda I_d)\vec{v} = \vec{0}$$

- The null space

$$\begin{aligned} E_\lambda &= \mathrm{null}(A - \lambda I_d) \\ &= \{\vec{v} : (A - \lambda I_d)\vec{v} = \vec{0}\} \end{aligned}$$

is called the **eigenspace** for λ

Eigenspace Examples

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix}$$

$$\lambda_1 = 4 \qquad A - \lambda_1 I_2 = \begin{pmatrix} -2 & 3 \\ 4 & -6 \end{pmatrix}$$

$$E_4 = \mathrm{span}\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix}\right)$$

$$\lambda_2 = -4 \qquad A - \lambda_2 I_2 = \begin{pmatrix} 6 & 3 \\ 4 & 2 \end{pmatrix}$$

$$E_{-4} = \mathrm{span}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)$$

Eigenspace Examples

e.g.

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

Only one eigenvalue $\lambda = 2$

$$A - \lambda I_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$E_2 = \mathbb{R}^2 = \operatorname{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

still can find 2 linearly independent eigenvectors

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$$

Only one eigenvalue $\lambda = 2$

$$A - \lambda I_2 = \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}$$

$$E_2 = \operatorname{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$$

only 1 linearly independent eigenvector

Diagonalization

- A matrix $A \in \mathbb{R}^{d \times d}$ is **diagonalizable** if:
  1. It has d linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_d$
  2. So enough eigenvectors to form a basis for $\mathbb{R}^d$

- Let $\lambda_1, \ldots, \lambda_d$ be the corresponding (not necessarily distinct) eigenvalues

Diagonalizable vs Not Examples

e.g.

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

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \;\&\; \lambda_1 = 2 \qquad \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;\&\; \lambda_2 = 2$$

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$$

Only has one (1-dimensional) eigenspace E_2 so A cannot have two linearly independent eigenvectors

$$E_2 = \text{span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$$

Eigen-Decomposition / Spectral Decomposition

- Suppose $A \in \mathbb{R}^{d\times d}$ is diagonalizable with d linearly independent eigenvectors

$$\vec{v}_1, \dots, \vec{v}_d$$

and corresponding eigenvalues $\lambda_1, \dots, \lambda_d$

- Let

$$V = \begin{pmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_d \\ | & & | \end{pmatrix} \in \mathbb{R}^{d\times d}$$

$$D = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{pmatrix} \in \mathbb{R}^{d\times d}$$

- Since $A\vec{v}_i = \lambda_i \vec{v}_i$ for each $i = 1, \dots, d$, we can write

$$A\begin{pmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_d \\ | & & | \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_d \\ | & & | \end{pmatrix}\begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{pmatrix}$$

In other words,

$$AV = VD$$

Eigen-Decomposition / Spectral Decomposition

- Since V has d linearly independent column vectors, it is invertible. We can rewrite

$$AV = VD$$

as

$$\boxed{A = VDV^{-1}}$$

- We decompose A into a product of (structurally simpler) invertible and diagonal matrices

e.g.

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

e.g.

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

Non-Uniqueness of Eigen-Decomposition

- Can re-order the eigenvectors

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

- **Convention 1:** Sort the eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_d$$

descending

- Can scale each eigenvector

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix}$$
$$= \begin{pmatrix} 3 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & -2 \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} 15 & -1 \\ 10 & 2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 15 & -1 \\ 10 & 2 \end{pmatrix}^{-1}$$

- **Convention 2:** Use normalized unit eigenvectors

$$\begin{pmatrix} 3/\sqrt{13} & 1/\sqrt{5} \\ 2/\sqrt{13} & -2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 3/\sqrt{13} & 1/\sqrt{5} \\ 2/\sqrt{13} & -2/\sqrt{5} \end{pmatrix}^{-1}$$

Diagonalizable Matrix Properties

- If $A \in \mathbb{R}^{d \times d}$ is diagonalizable with

$$A = VDV^{-1}$$

- **Facts:**
  1. Determinant

$$\det(A) = \det(D) = \lambda_1 \times \cdots \times \lambda_d$$

  2. Trace

$$\operatorname{tr}(A) = \operatorname{tr}(D) = \lambda_1 + \cdots + \lambda_d$$

  3. Rank

$$\begin{aligned} \operatorname{rank}(A) &= \operatorname{rank}(D) \\ &= \text{number of non-zero } \lambda_i\text{'s} \end{aligned}$$

- **Facts (cont'd):**
  4. Compute matrix power easily

$$\begin{aligned} A^k &= (VDV^{-1})(VDV^{-1}) \ldots (VDV^{-1}) \\ &= VD^kV^{-1} \end{aligned}$$

e.g. Markov chain k^{th} step transition probability matrix

- Set of eigenvalues is called **spectrum** of A. It captures a lot of characteristics of the matrix.

Diagonalizability

- **Question:** How do we know if a matrix $A \in \mathbb{R}^{d \times d}$ is diagonalizable i.e. its eigenspaces have enough dimensions together to yield d linearly independent eigenvectors?

- **Answer:** In general, it is difficult to tell unless we do the computation to obtain the eigenspaces and their dimensions

Different Eigenvalues

- **Fact:** Eigenvectors $\vec{v}_1, \dots, \vec{v}_k$ that correspond to different eigenvalues $\lambda_1 \neq \cdots \neq \lambda_k$ are always linearly independent

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix}$$

$$\lambda_1 = 4 \qquad E_4 = \operatorname{span}\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix}\right)$$

$$\lambda_2 = -4 \qquad E_{-4} = \operatorname{span}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)$$

Not coincident that eigenvectors from E_4 and E_{-4} are linearly independent

Totally Distinct Eigenvalues

- **Fact:** If a matrix $A \in \mathbb{R}^{d \times d}$ has d distinct eigenvalues

$$\lambda_1 \neq \cdots \neq \lambda_d$$

then

1. Eigenspaces $E_{\lambda_1}, \dots, E_{\lambda_d}$ are all 1-dimensional
2. Choose an eigenvector $\vec{v}_i \in E_{\lambda_i}$ then $\{\vec{v}_1, \dots, \vec{v}_d\}$ is linearly independent
3. A is diagonalizable

- But the converse is not true i.e. this is a sufficient but not necessary condition

- A matrix with repeated eigenvalues can still be diagonalizable

e.g.

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

$$= \begin{pmatrix} 0 & 1/\sqrt{2} \\ -1 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1/\sqrt{2} \\ -1 & 1/\sqrt{2} \end{pmatrix}^{-1}$$

Symmetric Matrix is Orthogonally Diagonalizable

- **Fact:** If $A \in \mathbb{R}^{d\times d}$ is symmetric, then

  1. It has eigenvalue(s)
  2. It has d linearly independent eigenvectors i.e. enough
  3. It is diagonalizable
  4. Eigenvectors $\vec{v}_i$ and $\vec{v}_j$ for distinct eigenvalues $\lambda_i \neq \lambda_j$ are orthogonal

$$\vec{v}_i \perp \vec{v}_j$$

e.g.

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

$$\lambda_1 = 4 \qquad E_4 = \operatorname{span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$$

$$\lambda_2 = -2 \qquad E_{-2} = \operatorname{span}\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$$

e.g.

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -2 \end{pmatrix}$$

$$\lambda_1 = 4 \qquad E_4 = \operatorname{span}\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix}\right)$$

$$\lambda_2 = -4 \qquad E_{-4} = \operatorname{span}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)$$

Symmetric Matrix is Orthogonally Diagonalizable

- **Fact:** If $A \in \mathbb{R}^{d\times d}$ is symmetric, then

  4. Eigenvectors from different eigenspaces $E_{\lambda_i} \neq E_{\lambda_j}$ are already orthogonal
  5. Basis within each eigenspace can made orthogonal using **Gram-Schmidt orthogonalization process**
  6. Overall, A has an orthonormal basis of eigenvectors $\hat{v}_1, \dots, \hat{v}_d$

e.g.

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

$$\begin{aligned} E_2 = \mathbb{R}^2 &= \operatorname{span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}\right) \\ &= \operatorname{span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}\right) \\ &= \operatorname{span}\underbrace{\left(\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right)}_{\text{orthonormal eigenvectors}} \end{aligned}$$

by Gram-Schmidt

Spectral Theorem

- **Fact:** If $A \in \mathbb{R}^{d \times d}$ is symmetric, then

  6. It has an orthonormal basis of eigenvectors $\hat{v}_1, \dots, \hat{v}_d$
  7. Can be decomposed into

$$A = QDQ^T$$

$$Q = \begin{pmatrix} | & & | \\ \hat{v}_1 & \dots & \hat{v}_d \\ | & & | \end{pmatrix} \in \mathbb{R}^{d \times d}$$

$$D = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{pmatrix} \in \mathbb{R}^{d \times d}$$

Singular Value and Singular Vector

- For general $A \in \mathbb{R}^{n \times d}$, its SVD is

$$AV = UD$$

$$A \begin{pmatrix} | & & | \\ \hat{v}_1 & \dots & \hat{v}_d \\ | & & | \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ \hat{u}_1 & \dots & \hat{u}_d \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d \end{pmatrix}$$

$$A\hat{v}_i = \sigma_i \hat{u}_i \quad \text{for } i = 1, \dots, d$$

left singular vector

singular value

right singular vector

Principal Component Analysis via Cov(X)

- **Question:** Is PCA related to spectral decomposition of the (symmetric) covariance matrix $\mathrm{Cov}(X)$ of data X?

- **Method 1:** Do spectral decomposition

$$\mathrm{Cov}(X) = C = QDQ^T$$

$$D = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{pmatrix}$$

- **Answer:** Yes. Centered data matrix X_c and its covariance matrix C are related:

  1. Eigenvectors of C are $Q = V$ are left singular vectors of X_c
  2. Eigenvalues of C and singular values of X_c are related

$$\lambda_i = \sigma_i^2/n$$

- **Method 2:** Do SVD on data $X_c = U\Sigma V^T$

$$C = \frac{1}{n}X_c^T X_c = \frac{1}{n}(V\Sigma^T U^T)(U\Sigma V^T) = \frac{1}{n}V\Sigma^2 V^T = V\begin{pmatrix} \sigma_1^2/n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d^2/n \end{pmatrix}V^T$$

Total Variance and Variance Explained

$$\mathrm{Cov}(X) = \begin{pmatrix} \mathrm{Var}(X_1) & \dots & \mathrm{Cov}(X_1, X_d) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \dots & \mathrm{Var}(X_d) \end{pmatrix} = Q \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{pmatrix} Q^T$$

- **Total variance** in the data X is

$$\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_d) = \mathrm{tr}(\mathrm{Cov}(X)) = \lambda_1 + \cdots + \lambda_d$$

$$\lambda_i = \begin{pmatrix} \text{variance along } i^{\text{th}} \text{ principal} \\ \text{component direction } \hat{v}_i \end{pmatrix} \propto i^{\text{th}} \text{ axis length of the data ellipsoid}$$

- Percent of **variance explained** by the first k principal components (out of all d) is

$$\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_d} = \frac{\sigma_1^2 + \cdots + \sigma_k^2}{\sigma_1^2 + \cdots + \sigma_d^2} \in (0,1)$$

Matrix Invertibility Criteria (cont'd)

- The following statements are equivalent:
  1. The matrix $A \in \mathbb{R}^{d \times d}$ is invertible
  2. The matrix A^T is invertible
  3. The determinant $\det(A) \neq 0$
  4. The d row (or column) vectors of A are linearly independent / span $\mathbb{R}^d$ / is a basis for $\mathbb{R}^d$
  5. The matrix A has full rank d
  6. The system $A\vec{x} = \vec{0}$ only has trivial solution $\vec{x} = \vec{0}$
  7. $\text{null}(A) = \{\vec{0}\}$
  8. $\text{nullity}(A) = 0$
  9. The system $A\vec{x} = \vec{b}$ has unique solution $\vec{x} = A^{-1}\vec{b}$
  10. $\lambda = 0$ is not an eigenvalue of A
  11. All singular values $\sigma_i > 0$ of A are positive