PRINCIPLE COMPONENT ANALYSIS AND SINGULAR VALUE DECOMPOSITION

吳沛遠 **Pei-Yuan Wu**
Dept. Electrical Engineering
National Taiwan University

1

2022/10/7

AGENDA

- Linear Algebra Preliminaries

- Spectral theorem for Symmetric matrices

- Principle Component Analyais (PCA)

- Singular Value Decomposition (SVD)

LINEAR ALGEBRA PRELIMINARIES

3

VECTOR SPACE

Vector space $(V, F, +, \cdot)$
- V: A set of objects (called vectors)
- F: Field (e.g., $\mathbb{R}$ or $\mathbb{C}$)
- Addition of vectors $+: V \times V \to V$ (i.e., addition is closed in V)
  - ➢ Commutativity: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$
  - ➢ Associativity: $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$
  - ➢ $\exists \boldsymbol{0} \in V$ such that $\boldsymbol{0} + \boldsymbol{x} = \boldsymbol{x} + \boldsymbol{0} = \boldsymbol{x}, \forall \boldsymbol{x} \in V$
    - ✓ Proposition: $\boldsymbol{0}$ is unique, called the *zero vector*. (Exercise)
  - ➢ $\forall \boldsymbol{x} \in V, \exists (-\boldsymbol{x}) \in V$ such that $\boldsymbol{x} + (-\boldsymbol{x}) = \boldsymbol{0}$
    - ✓ Proposition: $\forall \boldsymbol{x} \in V, (-\boldsymbol{x})$ is unique, called the *additive inverse* of $\boldsymbol{x}$. (Exercise)
- Scalar multiplication: $\cdot: F \times V \to V$ (i.e., scalar multiplication is closed in V)
  - ➢ Compatibility: $a(b\boldsymbol{x}) = (ab)\boldsymbol{x}, \forall \boldsymbol{x} \in V, a, b \in F$
  - ➢ Distributivity: $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}, \ (a + b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}, \forall \boldsymbol{x}, \boldsymbol{y} \in V, a, b \in F$
  - ➢ $1\boldsymbol{x} = \boldsymbol{x}$
- Example: Euclidean space $\mathbb{R}^n$.

INNER PRODUCT SPACE

- An ***inner product space*** is a vector space V over the field F ($\mathbb{R}$ or $\mathbb{C}$) together with an inner product

$$\langle \cdot, \cdot \rangle : V \times V \to F$$

That satisfies the following three axioms for all vectors $x, y, z \in V$ and all scalars $a \in F$:

➢ Conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$

➢ Linearity in the first argument

$$\langle ax, y \rangle = a\langle x, y \rangle$$
$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

➢ Positive definiteness

$$\langle x, x \rangle \geq 0$$
$$\langle x, x \rangle = 0 \Leftrightarrow x = 0$$

EUCLIDEAN SPACE

- Euclidean space $\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \dots, x_n \in \mathbb{R} \right\}$

- $\mathbb{R}^n$ is a vector space over field $\mathbb{R}$ by defining elementwise addition and scalar multiplication: (Exercise: Prove that vector space axioms hold)

  ➢ Addition: $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$

  ➢ Scalar multiplication: $\alpha \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix}, \alpha \in \mathbb{R}$

- $\mathbb{R}^n$ is an inner product space by defining inner (dot) product:

  ➢ Inner (dot) product: $\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle_{\mathbb{R}^n} = x_1 y_1 + \dots + x_n y_n \in \mathbb{R}$

COMPLEX EUCLIDEAN SPACE

- Complex Euclidean space $\mathbb{C}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ldots, x_n \in \mathbb{C} \right\}$

- $\mathbb{C}^n$ is a vector space over field $\mathbb{C}$ by defining elementwise addition and scalar multiplication: (Exercise: Prove that vector space axioms hold)

  ➤ Addition: $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$

  ➤ Scalar multiplication: $\alpha \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix}, \alpha \in \mathbb{C}$

- $\mathbb{C}^n$ is an inner product space by defining inner (dot) product:

  ➤ Inner (dot) product: $\left\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\rangle_{\mathbb{C}^n} = x_1 \overline{y_1} + \cdots + x_n \overline{y_n} \in \mathbb{C}$

LINEAR MAPPING

- Let V and W be vector spaces over the same field F. A function $T: V \to W$ is said to be a ***linear mapping*** if it preserves the addition and scalar multiplication operations. Namely,
$$T(a\boldsymbol{x} + b\boldsymbol{y}) = aT(\boldsymbol{x}) + bT(\boldsymbol{y}), \forall \boldsymbol{x}, \boldsymbol{y} \in M, a, b \in F$$

- Example: Let F be a field ($\mathbb{R}$ or $\mathbb{C}$). Each matrix $A \in F^{m \times n}$ can be considered as a linear mapping $T: F^n \to F^m$ defined as
$$T(\boldsymbol{x}) = A\boldsymbol{x}$$

- Proposition: Let F be a field ($\mathbb{R}$ or $\mathbb{C}$). Each linear mapping $T: F^n \to F^m$ can be represented by a matrix $A \in F^{m \times n}$ such that
$$T(\boldsymbol{x}) = A\boldsymbol{x}, \forall \boldsymbol{x} \in F^n$$

EIGEN-VECTOR AND EIGEN-VALUE

- Let V be vector space over the same field F, $T: V \to V$ be a linear mapping. We call $v \in V$ an **_eigenvector_** of T if $v \neq 0$ and
$$T(v) = \lambda v$$

  where $\lambda \in F$. We call λ the **_eigenvalue_** associated with v.

- Let F be a field ($\mathbb{R}$ or $\mathbb{C}$). A matrix $A \in F^{n \times n}$ can be viewed as a linear mapping $T: F^n \to F^n$, $T(x) = Ax$. Hence we call λ the eigenvalue (of A) associated with eigenvector (of A) $v \in F^n$ if $v \neq 0$ and
$$Av = \lambda v$$

- Exercise: Eigenvector and eigenvalue of
$$A = \begin{bmatrix} -1 & 6 \\ 3 & 2 \end{bmatrix}$$

SPECTRAL THEOREM FOR SYMMETRIC MATRICES

10

SYMMETRIC MATRIX IS DIAGONALIZABLE

▪ **Theorem 5:** If $A \in \mathbb{R}^{n \times n}$ is *symmetric*, then there exists diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and *orthogonal* matrix $U \in \mathbb{R}^{n \times n}$ such that $A = U\Lambda U^T$

$$U = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n]$$

eigenvectors

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \lambda_n \end{bmatrix}$$

Eigen Values

$$A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k$$

Symmetric:
$A = [a_{ij}] \in \mathbb{C}^{n \times n}$ is symmetric if $a_{ij} = a_{ji}$

Orthogonal matrix:
$U = [\boldsymbol{v}_1 \dots \boldsymbol{v}_n] \in \mathbb{R}^{n \times n}$ is an orthogonal matrix if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ are orthogonal and have unit length

$$\boldsymbol{v}_k^T \boldsymbol{v}_\ell = \begin{cases} 1 & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}$$

That is, $U^T U = I$, namely, $U^{-1} = U^T$.

HERMITIAN MATRIX IS DIAGONALIZABLE

- **Theorem 4:** If $A \in \mathbb{C}^{n \times n}$ is *Hermitian*, then there exists diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and *unitary* matrix $U \in \mathbb{C}^{n \times n}$ such that $A = U\Lambda U^{\mathrm{H}}$

$$U = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$$

eigenvectors

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \lambda_n \end{bmatrix}$$

Eigen Values

$$A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k$$

Conjugate transpose:
The conjugate transpose of $A \in \mathbb{C}^{m \times n}$ *is* denoted as $A^H = \overline{A^T} \in \mathbb{C}^{n \times m}$

Hermitian:
$A = [a_{ij}] \in \mathbb{C}^{n \times n}$ is Hermitian if $a_{ij} = \overline{a_{ji}}$

Unitary Matrix:
$U = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_n] \in \mathbb{C}^{n \times n}$ is an unitary matrix if $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal and have unit length

$$\boldsymbol{v}_k^H \boldsymbol{v}_\ell = \begin{cases} 1 & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}$$

That is, $U^H U = I$

INVARIANT SUBSPACE

- M is called a **_subspace_** of vector space (V, F), if it is closed under vector addition and multiplication. Namely,
$$a\boldsymbol{x} + b\boldsymbol{y} \in M, \forall \boldsymbol{x}, \boldsymbol{y} \in M, a, b \in F$$

- Let V be a vector space, $T: V \to V$ be a linear mapping. M is called an **_invariant subspace_** of T if $T(\boldsymbol{x}) \in M, \forall \boldsymbol{x} \in M$.

- Let F be a field ($\mathbb{R}$ or $\mathbb{C}$). A matrix $A \in F^{m \times n}$ can be viewed as a linear mapping $T: F^n \to F^m, T(\boldsymbol{x}) = A\boldsymbol{x}$. We call M an invariant subspace of A if $A\boldsymbol{x} \in M, \forall \boldsymbol{x} \in M$.

ADJOINT OPERATOR

- Let H_1, H_2 be inner product spaces, $T: H_1 \to H_2$ be a linear mapping. If a linear mapping $T^*: H_2 \to H_1$ satisfies

$$\langle T\mathbf{x}, \mathbf{y} \rangle_{H_2} = \langle \mathbf{x}, T^*\mathbf{y} \rangle_{H_1}, \forall \mathbf{x} \in H_1, \mathbf{y} \in H_2$$

  Then we call T^* *the **adjoint*** of T.

- **Proposition:** T^* is unique (if it does exist) (Exercise)

- **Proposition:** $(T^*)^* = T$

Proof: $\langle T^*\boldsymbol{y}, \boldsymbol{x} \rangle_{H_1} = \overline{\langle \boldsymbol{x}, T^*\boldsymbol{y} \rangle_{H_1}} = \overline{\langle T\boldsymbol{x}, \boldsymbol{y} \rangle_{H_2}} = \langle \boldsymbol{y}, T\boldsymbol{x} \rangle_{H_2}, \forall \boldsymbol{x} \in H_1, \boldsymbol{y} \in H_2$

- Example: Suppose linear mapping $T: \mathbb{C}^n \to \mathbb{C}^m$ is represented by $A \in \mathbb{C}^{m \times n}$. Since

$$\langle T\boldsymbol{x}, \boldsymbol{y} \rangle = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^H(A\boldsymbol{x}) = (A^H\boldsymbol{y})^H \boldsymbol{x} = \langle \boldsymbol{x}, A^H\boldsymbol{y} \rangle,$$

  therefore T has adjoint $T^*: \mathbb{C}^m \to \mathbb{C}^n$, which is represented by $A^H \in \mathbb{C}^{m \times n}$.

SELF ADJOINT

- Let H be inner product space, $T\colon H \to H$ be a linear mapping. We call T **_self-adjoint_** if T^* exists and

$$T = T^*$$

- Example: If linear mapping $T\colon \mathbb{C}^n \to \mathbb{C}^n$ is self-adjoint and represented by $A \in \mathbb{C}^{n\times n}$, then $A = A^H$, i.e., A is Hermitian.

- **Proposition:** If λ is an eigenvalue of self-adjoint mapping T, then $\lambda \in \mathbb{R}$.

  Proof: Let $\boldsymbol{x}$ be an eigenvector associated with λ, then
$$\lambda\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \langle T\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{x}, T^*\boldsymbol{x}\rangle = \langle \boldsymbol{x}, T\boldsymbol{x}\rangle = \bar{\lambda}\langle \boldsymbol{x}, \boldsymbol{x}\rangle$$

  Therefore $\lambda = \bar{\lambda}$.

NORMAL OPERATOR

- Let H be inner product space, $T: H \to H$ be a linear mapping. We call T **_normal_** if T^* exists and $TT^* = T^*T$.

- Example: Suppose linear mapping $T: \mathbb{C}^n \to \mathbb{C}^n$ is represented by $A \in \mathbb{C}^{n \times n}$, then T is normal iff $AA^H = A^H A$.

- **Proposition:** If $A \in \mathbb{C}^{n \times n}$ is normal, then
$$\|A\boldsymbol{x} - \lambda\boldsymbol{x}\|^2 = \left\|A^H \boldsymbol{x} - \bar{\lambda}\boldsymbol{x}\right\|^2, \forall \boldsymbol{x} \in \mathbb{C}^n, \lambda \in \mathbb{C}$$

Proof:
$$\|A\boldsymbol{x} - \lambda\boldsymbol{x}\|^2 = \boldsymbol{x}^H\left(A^H A - \lambda A^H - \bar{\lambda}A + |\lambda|^2 I\right)\boldsymbol{x} = \boldsymbol{x}^H\left(AA^H - \bar{\lambda}A - \lambda A^H + |\lambda|^2 I\right)\boldsymbol{x} = \left\|A^H \boldsymbol{x} - \bar{\lambda}\boldsymbol{x}\right\|^2$$

- **Proposition:** Let $A = U\Lambda U^H$ for which $\Lambda \in \mathbb{C}^{n \times n}$ is diagonal and $U \in \mathbb{C}^{n \times n}$ is unitary, then A is normal

Proof: $AA^H = U\Lambda\Lambda^H U^H = U\Lambda^H \Lambda U^H = A^H A$

NORMAL MATRIX IS DIAGONALIZABLE

- **Lemma 1:** If $v_1, v_2, \ldots, v_k$ are eigenvectors of an operator T, then $M_k = \{x | \langle x, v_i \rangle = 0, \forall i = 1, 2, \ldots, k\}$ is an invariant subspace of T^*.

Proof:
$$\langle T^* x, v_i \rangle = \langle x, T v_i \rangle = \overline{\lambda_i} \langle x, v_i \rangle = 0, \forall x \in M_k, i = 1, 2, \ldots, k$$

NORMAL MATRIX IS DIAGONALIZABLE

- **Lemma 2:** If M is a non-trivial invariant subspace of linear mapping $T: \mathbb{C}^n \to \mathbb{C}^n$, then there exists $\lambda \in \mathbb{C}, \boldsymbol{v} \in M, \boldsymbol{v} \neq \boldsymbol{0}$ such that $T\boldsymbol{v} = \lambda\boldsymbol{v}$

Proof: Take $A \in \mathbb{C}^{n \times n}$ which represents T. Take $U \in \mathbb{C}^{n \times m}$ whose columns $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ form an orthonormal basis of M. Define $\psi: \mathbb{C}^m \to M$ as

$$\psi\left(\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}\right) = \sum_{i=1}^{m} c_i \boldsymbol{u}_i$$

Then ψ is a one-to-one onto linear mapping. Consider linear mapping $\psi^{-1} \circ T \circ \psi: \mathbb{C}^m \to \mathbb{C}^m$ which can be represented by $U^H A U \in \mathbb{C}^{m \times m}$, it is evident that there exists some $\lambda \in \mathbb{C}, \boldsymbol{c} \in \mathbb{C}^m, \boldsymbol{c} \neq \boldsymbol{0}$ such that $U^H A U \boldsymbol{c} = \lambda \boldsymbol{c}$. (Hint: Take λ sastisfying $\det(U^H A U - \lambda I) = 0$ and nonzero vector $\boldsymbol{c} \in Null(U^H A U - \lambda I)$). Note that

$$\psi^{-1} \circ T \circ \psi(\boldsymbol{c}) = U^H A U \boldsymbol{c} = \lambda \boldsymbol{c} \Rightarrow T(\psi(\boldsymbol{c})) = \psi(\lambda \boldsymbol{c}) = \lambda \psi(\boldsymbol{c})$$

Therefore $\boldsymbol{v} = \psi(\boldsymbol{c}) \in M$ is a nonzero vector satisfying $T\boldsymbol{v} = \lambda\boldsymbol{v}$. Q.E.D.

NORMAL MATRIX IS DIAGONALIZABLE

- **Theorem 3:** If $A \in \mathbb{C}^{n \times n}$ is normal, then there exists diagonal matrix $\Lambda \in \mathbb{C}^{n \times n}$ and unitary matrix $U \in \mathbb{C}^{n \times n}$ such that $A = U\Lambda U^H$.

Proof: We will construct pairwise-orthogonal eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ with associating eigenvalues $\lambda_1, \dots, \lambda_n$ of A as follows:

For $k = 1, 2, \dots, n$, suppose $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1}$ are pairwise-orthogonal eigenvectors of A. By Lemma 1,

$$M_{k-1} = \{\boldsymbol{x} \in \mathbb{C}^n | \langle \boldsymbol{x}, \boldsymbol{v}_i \rangle = 0, \forall i = 1, \dots, k-1\}$$

is a non-trivial invariant subspace of A. Therefore, by Lemma 2, there exists $\lambda_k \in \mathbb{C}$, $\boldsymbol{v}_k \in M_{k-1}, \boldsymbol{v}_k \neq \boldsymbol{0}$ such that $A^H \boldsymbol{v}_k = \overline{\lambda_k} \boldsymbol{v}_k$, which implies $A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k$. In other words, $\boldsymbol{v}_k$ is an eigenvector of A with associating eigenvalue λ_k which is orthogonal to $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1}$. Hence $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are pairwise-orthogonal eigenvectors of A.

The above procedure thus constructs pairwise-orthogonal eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ with associating eigenvalues $\lambda_1, \dots, \lambda_n$ of A.

NORMAL MATRIX IS DIAGONALIZABLE

- **Theorem 3:** If $A \in \mathbb{C}^{n \times n}$ is normal, then there exists diagonal matrix $\Lambda \in \mathbb{C}^{n \times n}$ and unitary matrix $U \in \mathbb{C}^{n \times n}$ such that $A = U\Lambda U^H$.

Proof: (cont'd) We have shown the existence of pairwise-orthogonal eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, normalized to unit length $\|\boldsymbol{v}_i\| = 1$, with associating eigenvalues $\lambda_1, \ldots, \lambda_n$. Set $U = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_n] \in \mathbb{C}^{n \times n}$, then $U^H U = I$, i.e., U is unitary. Since for every $\boldsymbol{c} \in \mathbb{C}^n$, one has

$$AU\boldsymbol{c} = A\sum_{i=1}^{m} c_i \boldsymbol{v}_i = \sum_{i=1}^{m} c_i A\boldsymbol{v}_i = \sum_{i=1}^{m} c_i \lambda_i \boldsymbol{v}_i = U\Lambda\boldsymbol{c}$$

Where $\Lambda = diag(c_1, \ldots, c_n)$. Therefore $AU = U\Lambda$ which implies $A = U\Lambda U^{-1} = U\Lambda U^H$.
Q.E.D.

That is, a normal matrix A can always be written as $A = U\Lambda U^H$, where U is an unitary matrix of eigenvectors, and Λ is a diagonal matrix of the associated eigenvalues.

HERMITIAN MATRIX IS DIAGONALIZABLE

Since every Hermitian matrix is normal (so is diagonizable) and self-adjoint (so eigenvalues are real), one immediately obtains Theorem 4, as restated below:

- **Theorem 4:** If $A \in \mathbb{C}^{n \times n}$ is *Hermitian*, then there exists diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and *unitary* matrix $U \in \mathbb{C}^{n \times n}$ such that $A = U \Lambda U^{\mathrm{H}}$.

REAL SYMMETRIC MATRIX IS DIAGONALIZABLE WITH REAL EIGENVECTORS

- **Lemma 3:** If M is a non-trivial invariant subspace of self-adjoint operator $T: \mathbb{R}^n \to \mathbb{R}^n$, then there exists $\lambda \in \mathbb{R}, \boldsymbol{v} \in M, \boldsymbol{v} \neq \boldsymbol{0}$ such that $T\boldsymbol{v} = \lambda\boldsymbol{v}$

Proof: Take symmetric matrix $A \in \mathbb{R}^{n \times n}$ which represents T. Take $U \in \mathbb{R}^{n \times m}$ whose columns $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ form an orthonormal basis of M. Define $\psi: \mathbb{R}^m \to M$ as

$$\psi\left(\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}\right) = \sum_{i=1}^{m} c_i \boldsymbol{u}_i$$

Then ψ is a one-to-one onto linear mapping. Consider linear mapping $\psi^{-1} \circ T \circ \psi: \mathbb{R}^m \to \mathbb{R}^m$ which can be represented by $U^T A U \in \mathbb{R}^{m \times m}$. Note that $U^T A U$ is self-adjoint, so it has eigenvalue $\lambda \in \mathbb{R}$ and there is non-zero $\boldsymbol{c} \in \mathbb{R}^m$ for which $U^T A U \boldsymbol{c} = \lambda\boldsymbol{c}$. Note that

$$\psi^{-1} \circ T \circ \psi(\boldsymbol{c}) = U^T A U \boldsymbol{c} = \lambda\boldsymbol{c} \Rightarrow T\big(\psi(\boldsymbol{c})\big) = \psi(\lambda\boldsymbol{c}) = \lambda\psi(\boldsymbol{c})$$

Therefore $\boldsymbol{v} = \psi(\boldsymbol{c}) \in M$ is a nonzero vector satisfying $T\boldsymbol{v} = \lambda\boldsymbol{v}$. Q.E.D.

REAL SYMMETRIC MATRIX IS DIAGONALIZABLE WITH REAL EIGENVECTORS

- **Theorem 5:** If $A \in \mathbb{R}^{n \times n}$ is symmetric, then there exists diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $A = U \Lambda U^T$.

Proof: We will construct pairwise-orthogonal eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ with associating eigenvalues $\lambda_1, \dots, \lambda_n$ of A as follows:

For $k = 1, 2, \dots, n$, suppose $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1}$ are pairwise-orthogonal eigenvectors of A. By Lemma 1,

$$M_{k-1} = \{\mathbf{x} \in \mathbb{R}^n | \langle \mathbf{x}, \boldsymbol{v}_i \rangle = 0, \forall i = 1, \dots, k-1\}$$

is a non-trivial invariant subspace of $A^T = A$. Therefore, by Lemma 3, there exists $\lambda \in \mathbb{R}$, $\boldsymbol{v}_k \in M_{k-1}$, $\boldsymbol{v}_k \neq \boldsymbol{0}$ such that $A\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k$. In other words, $\boldsymbol{v}_k$ is an eigenvector of A with associating eigenvalue λ_k which is orthogonal to $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1}$. Hence $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are pairwise-orthogonal eigenvectors of A.

The above procedure thus constructs pairwise-orthogonal eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ with associating eigenvalues $\lambda_1, \dots, \lambda_n$ of A.

REAL SYMMETRIC MATRIX IS DIAGONALIZABLE WITH REAL EIGENVECTORS

- **Theorem 5:** If $A \in \mathbb{R}^{n \times n}$ is symmetric, then there exists diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $A = U \Lambda U^T$.

Proof: (cont'd): We have shown the existence of pairwise-orthogonal eigenvectors $v_1, \dots, v_n$, normalized to unit length $\|v_i\| = 1$, with associating eigenvalues $\lambda_1, \dots, \lambda_n$. Set $U = [v_1 \dots v_n] \in \mathbb{R}^{n \times n}$, then $U^T U = I$, i.e., U is orthogonal. Since for every $c \in \mathbb{R}^n$, one has

$$AUc = A \sum_{i=1}^{m} c_i v_i = \sum_{i=1}^{m} c_i A v_i = \sum_{i=1}^{m} c_i \lambda_i v_i = U \Lambda c$$

Where $\Lambda = diag(c_1, \dots, c_n)$. Therefore $AU = U\Lambda$ which implies $A = U\Lambda U^{-1} = U\Lambda U^T$.
**Q.E.D.**

# PRINCIPLE COMPONENT ANALYAIS (PCA)

25

DIMENSIONALITY REDUCTION

- Mapping the original high-dimensional data onto a lower-dimensional subspace.
- Pros for dimensionality reduction
  - ➢ Clarity of representation, ease of understanding (e.g., visualization in 2D or 3D)
  - ➢ Data compression, computational cost reduction.
  - ➢ Noise reduction, prevention of overfitting
- Cons for dimensionality reduction
  - ➢ oversimplification: loss of important or relevant information

M

N X

→

m

N $\widetilde{X}$
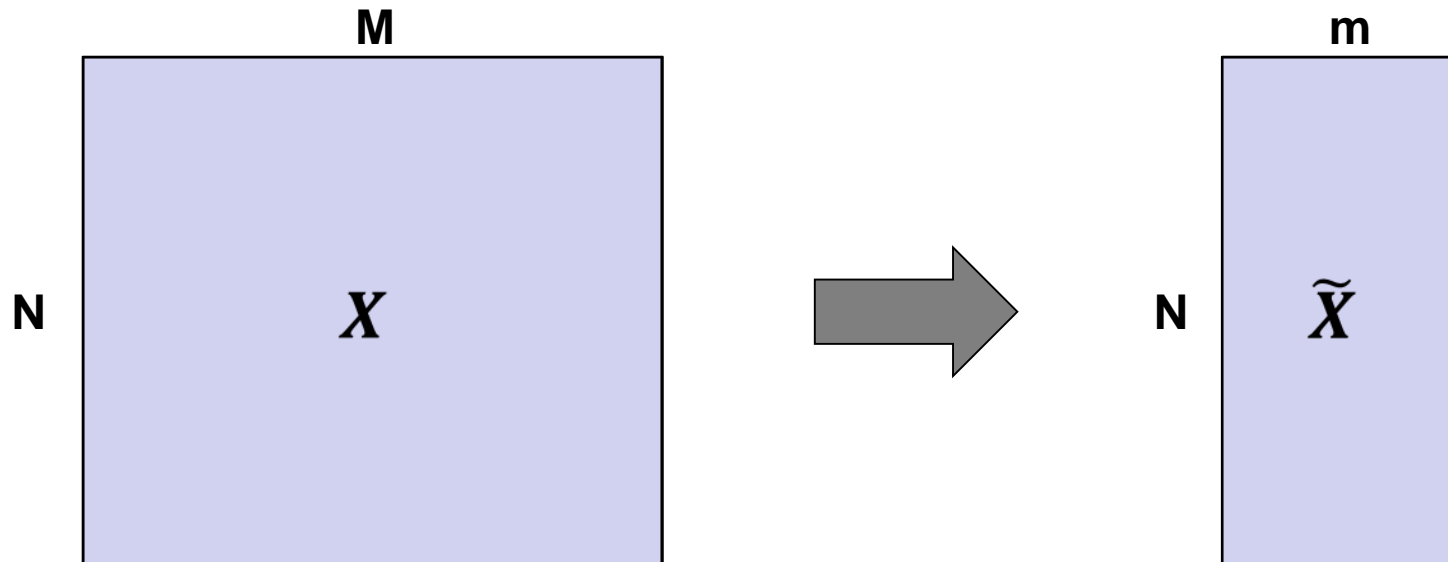
PRINCIPAL COMPONENT ANALYSIS (PCA)

- PCA's target: finding the best lower dimensional sub-space that conveys most of the variance in the original data

- Example: If we were to compress 2-D data to 1-D subspace, then PCA prefers projecting to the **black** line, since it preserves more variance comparing to *blue* line.
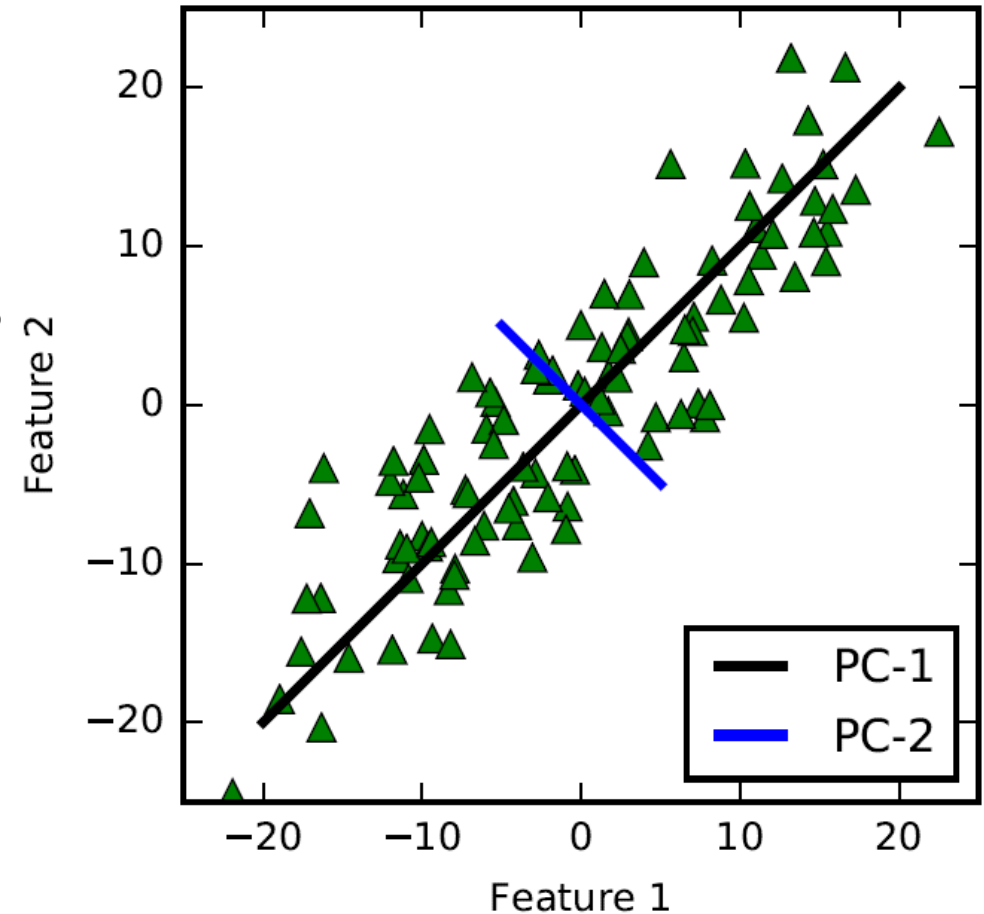
PRINCIPLE AXES

- Objective of PCA: Given data in $\mathbb{R}^M$, want to *rigidly rotate* the axes to new positions (principle axes) with the following properties:

  ➢ *Ordered such that principle axis 1 has the highest variance, axis 2 has the next highest variance, …, and axis M has the lowest variance.*

  ➢ *Covariance among each pair of the principal axes is zero.*

- The k'th **principle component** is the projection to the k'th principle axis.

- Keep the first $m < M$ principle components for dimensionality reduction.
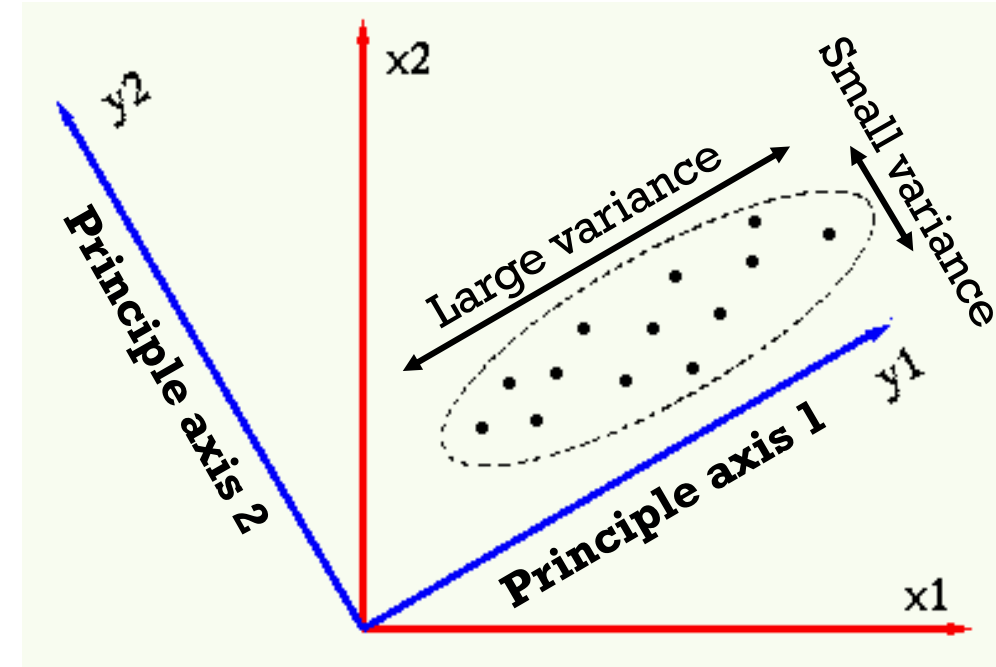
PRINCIPLE COMPONENT COMPUTATION

- Given N data $x_1, \ldots, x_N \in \mathbb{R}^M$, PCA first compute the covariance matrix for the data

$$\Sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T = U \Lambda U^T$$

  where $\mu \in \mathbb{R}^M$ is the data mean.

- Since Σ is symmetric, Σ can be written as $\Sigma = U \Lambda U^T$, where $U = [v_1 \ldots v_n]$ is unitary matrix of eigenvectors (of Σ), $\Lambda = diag(\lambda_1, \ldots, \lambda_M)$ is diagonal matrix of the associated eigenvalues arranged in non-ascending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0$. (Note that all eigenvalues are non-negative real scalars since Σ is ***semi-positive definite***.)

- For data $x \in \mathbb{R}^M$, compute its 1st principle component as $v_1^T x$, 2nd principle component as $v_2^T x, \ldots,$ M'th principle component as $v_M^T x$

> *Positive definite:*
> $A \in \mathbb{R}^{n \times n}$ is semi-positive definite if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$. If the equality holds only when $x = 0$, then A is positive definite.

PRINCIPLE COMPONENTS ARE UNCORRELATED

- The covariance of the k'th and ℓ'th principle components of data $x_1, \ldots, x_N$ is

$$\frac{1}{N}\sum_{i=1}^{N}\left[v_k^T(x_i - \mu)\right]\left[v_\ell^T(x_i - \mu)\right] = \frac{1}{N}\sum_{i=1}^{N}v_k^T(x_i - \mu)(x_i - \mu)^T v_\ell$$

$$= v_k^T \Sigma v_\ell = v_k^T U\Lambda U^T v_\ell = e_k^T \Lambda e_\ell = \begin{cases} \lambda_k & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}$$

Therefore

➢ The variance of the k'th principle components is λ_k.

$\Rightarrow$ *principle axis 1 has the highest variance, axis 2 has the next highest variance, ..., and axis M has the lowest variance.*

➢ The covariance of different principle components is zero.

➢ $\Rightarrow$ *Covariance among each pair of the principal axes is zero.*

SINGULAR VALUE DECOMPOSITION (SVD)

31

THEOREM OF SVD

- **Theorem:** Suppose $A \in \mathbb{R}^{m \times n}$, then there exists a factorization, called **singular value decomposition** of A, of the form

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

  where $U \in \mathbb{R}^{m \times m}$ and $\mathrm{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices, $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix with non-negative numbers on the diagonal.

  <u>Proof:</u> Since $A^T A \in \mathbb{R}^{n \times n}$ is symmetric and semi-positive definite, it can be factorized in the form

$$A^T A = V \Lambda V^T$$

  where $\mathrm{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_n] \in \mathbb{R}^{n \times n}$ is orthogonal matrix, $\Lambda = diag(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n \times n}$ is a diagonal matrix of eigenvalues arranged in non-ascending order

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

THEOREM OF SVD

Proof (cont'd): Let $p = \max\{i : \lambda_i > 0\}$ be the rank of A. Define $\boldsymbol{u}_i = \frac{A\boldsymbol{v}_i}{\sqrt{\lambda_i}} \in \mathbb{R}^m$ for $i = 1, \dots, p$. Then

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \langle \frac{A\boldsymbol{v}_i}{\sqrt{\lambda_i}}, \frac{A\boldsymbol{v}_j}{\sqrt{\lambda_j}} \rangle = \frac{\langle \boldsymbol{v}_i, A^T A \boldsymbol{v}_j \rangle}{\sqrt{\lambda_i}\sqrt{\lambda_j}} = \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_i}} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = \delta_{ij} = \begin{cases} 1, \text{if i} = \text{j} \\ 0, \text{if i} \neq \text{j} \end{cases}$$

That is, $\boldsymbol{u}_1, \dots, \boldsymbol{u}_p$ are pairwise orthogonal unit vectors in $\mathbb{R}^m$. By Gram-Schmidt process we can construct $\boldsymbol{u}_{p+1}, \dots, \boldsymbol{u}_m \in \mathbb{R}^m$ (if $p < m$) such that $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ are pairwise orthogonal unit vectors. This gives orthogonal matrix $U = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_m] \in \mathbb{R}^{m \times m}$.

We have the following claim:

Claim: $\boldsymbol{u}_i^T A \boldsymbol{v}_j = \sqrt{\lambda_j} \delta_{ij}$ for $i = 1, \dots, m, \text{j} = 1, \dots, n$

Proof: If $\text{j} \leq p$, then $A\boldsymbol{v}_j = \sqrt{\lambda_j} \boldsymbol{u}_j$, therefore $\boldsymbol{u}_i^T A \boldsymbol{v}_j = \sqrt{\lambda_j} \boldsymbol{u}_i^T \boldsymbol{u}_j = \sqrt{\lambda_j} \delta_{ij}$.

If $\text{j} > p$, since $\boldsymbol{v}_j$ is an eigenvector of $A^T A$ with associating eigenvalue $\lambda_j = 0$, therefore

$$\left(A\boldsymbol{v}_j\right)^T \left(A\boldsymbol{v}_j\right) = \boldsymbol{v}_j^T A^T A \boldsymbol{v}_j = \boldsymbol{v}_j^T \lambda_j \boldsymbol{v}_j = 0$$

Hence $A\boldsymbol{v}_j = \boldsymbol{0}$ and $\boldsymbol{u}_i^T A \boldsymbol{v}_j = 0 = \sqrt{\lambda_j} \delta_{ij}$. **Q.E.D.**

Therefore $\Sigma = U^T A V = diag(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p}, 0, \dots 0) \in \mathbb{R}^{m \times n}$ is a diagonal matrix.

Therefore $A = U\Sigma V^T$. **Q.E.D.**