Lecture 7: Introduction to the Singular Value Decomposition
p points in \( n = 2 \) dimensions

\[ U = \text{basis via SVD} \]

\[ \Rightarrow \text{if } X \in \mathbb{R}^{n \times p} \Rightarrow X = U \Sigma V^T \]

\( \hat{U} = \text{basis via GS} \)

\( U_1 \) is the 1d subspace that is closest to all of the \( X_i \)'s (i.e. best 1d subspace fit)
\[ r_i = y_i - (ax_i + b) \]

\[ \text{minimize } \sum_i r_i^2 \]

\[ \text{SVD} \]

\[ x_i, y_i \in \mathbb{R} \]

\[ r_i \]

\[ x_i \in \mathbb{R}^2 \]

\[ U_i \]

\[ X_i \]
Ex. Find a 1d-subspace (line through origin) that is closest to a set of points \( x_1, x_2, \ldots, x_n \in \mathbb{R}^p 

\Rightarrow \text{Find } a \text{ to minimize sum of squared distances}

Projection Matrices

\[
P_A^1 = A(A^*A)^{-1}A^* = P_A^2 = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}
\]

Given a subspace \( \mathcal{S} \) spanned by columns of \( A \),
the orthogonal complement of \( \mathcal{S} \) (also orthogonol complement of \( A \)) is the set of all vectors \( b \) orthogonal to all columns of \( A \) (i.e. orthogonal to every vector in \( \mathcal{S} \)).

Let \( A \in \mathbb{R}^{p \times n} \), \( B \in \mathbb{R}^{p \times (p-n)} \) be orthogonal complements

\[ AB = 0 \text{ and any } x \in \mathbb{R}^p \text{ can be written as } x = Au + Bv \text{ for some } u \in \mathbb{R}^p, v \in \mathbb{R}^{p-n} \]

\[ P_A + P_B = I \]

\[ I - P_A = P_B \quad \text{and} \quad I - P_B = P_A \]

\[ \Rightarrow P_A \text{ is also a projection matrix!} \]
distance from \( x_i \) to line \( a \):

\[
    d_i = \| x_i - P_a x_i \|_2
\]

\[
    = \| x_i - a (a^T a)^{-1} a^T x_i \|_2
\]

\[
    = \| x_i - \frac{a a^T}{a^T a} x_i \|_2
\]

\[
    = \| (I - \frac{a a^T}{a^T a}) x_i \|_2
\]

\[
    = x_i^T (I - \frac{a a^T}{a^T a}) (I - \frac{a a^T}{a^T a}) x_i
\]

\[
    = x_i^T (I - \frac{a a^T}{a^T a}) x_i
\]

\[
    = x_i^T x_i - \frac{(a^T x_i)^2}{a^T a}
\]

---

Want to minimize

\[
    \sum_{i=1}^p d_i^2 = \sum_{i=1}^p \left( x_i^T x_i - \frac{(a^T x_i)^2}{a^T a} \right)
\]

constant with respect to \( a \)

\[
\Rightarrow \arg \min_a \sum_{i=1}^p d_i^2(a) = \arg \max_a \sum_{i=1}^p \frac{a^T x_i x_i^T a}{a^T a} - \arg \max_a \sum_{i=1}^p a^T x_i x_i^T a
\]

Let \( X = [x_1, x_2, \ldots, x_p] \in \mathbb{R}^{n \times p} \)

\[
\Rightarrow \quad U_1 = \arg \max_a \sum_{i=1}^p \frac{a^T x_i x_i^T a}{a^T a} - \arg \max_a \sum_{i=1}^p \frac{a^T x_i x_i^T a}{a^T a}
\]

---

value of \( a \) that achieves maximum is 1st left singular vector of \( X \), denoted \( U_1 \).

also, \( \sigma_1 = \max_a \frac{\| X^T a \|_2}{\| a \|_2} = \| X U_1 \|_2 \cdot \| U_1 \|_2 \)

is called the 1st singular value of \( X \)

(also called the "operator norm" of \( X : \| X \|_{op} = \| X \|_2 \))

The bigger \( \sigma_1 \) is, the smaller \( \sum d_i \) is

\[
\Rightarrow \text{the better the } x_i \text{'s are aligned with a 1-D subspace}
\]
The Singular Value Decomposition

Consider a matrix $X \in \mathbb{R}^{n \times p}$. There always exists matrices $U, \Sigma, V$ such that

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

- $U \in \mathbb{R}^{n \times n}$ is orthogonal ($U^* U = U U^* = I$), called left singular vectors
- $V \in \mathbb{R}^{p \times p}$ is orthogonal ($V^T V = V V^T = I$), called right singular vectors
- $\Sigma \in \mathbb{R}^{n \times n}$ is diagonal, diagonal elements called singular values

The columns of $U$ form an orthonormal basis for the columns of $X$.

The singular values weight (scale the length) of the corresponding singular vectors.

- The number of non-zero singular vectors is the RANK of $X$.

The columns of $V^T$ (rows of $V$) are the basis coefficients (weights on the columns of $U \Sigma$) needed to represent each column of $X$. 
- \( U \) gives orthonormal for all of \( \mathbb{R}^n \).

- 1st \( r \) columns of \( U \) give basis of best \( r \)-dim subspace fit to columns of \( X \).

- \( \sigma_i \)'s indicate how important each subspace dimension is to representing/approximating data.

- 1st \( r \) columns of \( V \) give coordinates/locations of each \( x_i \) within the subspace spanned by \( U_1, \ldots, U_r \).

- \( U_i = \text{best 1d subspace fit to all data} \)

\[ \tilde{x}_i^{(1)} = x_i - P_{U_1} x_i = \text{ith residual} \quad \forall \, i \]

- \( U_2 = \text{best 1d subspace fit to all } \tilde{x}_i^{(1)} \)

\[ \tilde{\tilde{x}}_i^{(2)} = x_i - P_{U_1, U_2} x_i = \tilde{x}_i^{(1)} - P_{U_2} \tilde{x}_i^{(1)} \quad \forall \, i \]

- \( U_{k+1} = \text{best 1d subspace fit to all } \tilde{\tilde{x}}_i^{(2)} \)

- Recall Gram-Schmidt:

\[ U_1 = x_i / \| x_i \|_2 \]

\[ \tilde{x}_2 = x_2 - P_{U_1} x_2 \]

\[ U_2 = \tilde{x}_2 / \| \tilde{x}_2 \|_2 \]

Order of points matters!

Basis vectors less interpretable.

Singular values \( \sigma_1, \sigma_2, \) etc. indicate how spread out points are in the subspace.