MATH 5110 – Applied Linear Algebra and Matrix Analysis

❖ More Metric Spaces

--Distances between objects of different dimensions

Instructor: He Wang
Department of Mathematics
Northeastern University

https://hewang.sites.northeastern.edu/

Motivation:

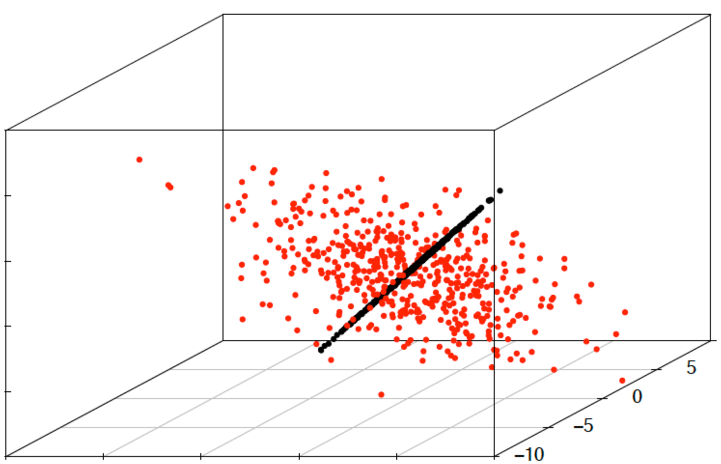
k objects: genes, tweets, images, etc.

n features: expression levels, term frequencies, frames, etc

j-th object described by feature vector $\vec{a}_1 \in \mathbb{R}^n$

Data set is described as $A = [\vec{a}_1 \ldots \vec{a}_k] \in \mathbb{R}^{n \times k}$
(transposed from machine learning)

Often what is important is subspace defined by A, e.g., $im\ A$ or *principal subspaces of A defined by eigenvectors of covariance matrix.*



Problem:

Suppose $\vec{a}_1 \ldots \vec{a}_k \in \mathbb{R}^n$ and $\vec{b}_1 \ldots \vec{b}_k \in \mathbb{R}^n$ two collections of k linearly independent vectors.

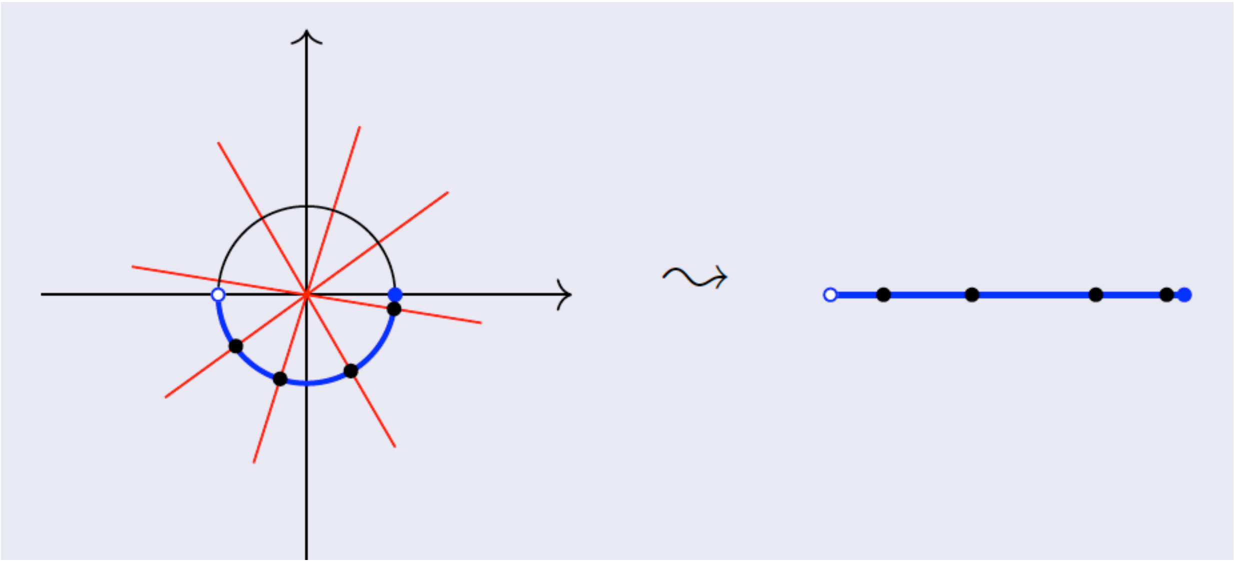
We want measure of separation of subspace $span(\vec{a}_1 \ldots \vec{a}_k)$ and $span(\vec{b}_1 \ldots \vec{b}_k)$

Solution: distances or angles between subspaces of the same dimension. (equivalent)

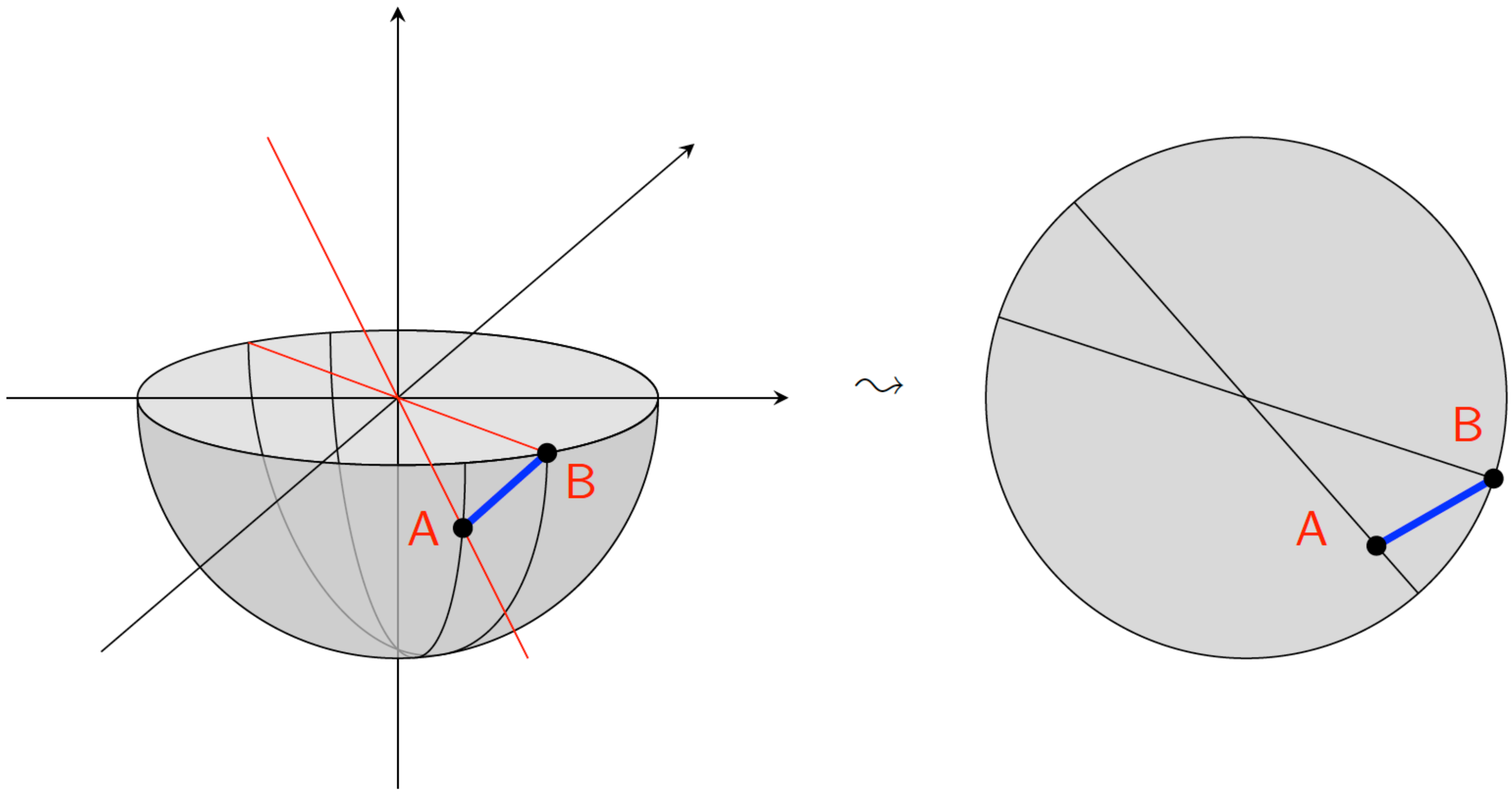
Question: What is the distance between two linear subspaces?

Example: Dimension 1 Subspaces in $\mathbb{R}^2$.

We can take the angle.



Higher-dimensional



Grassmann manifold

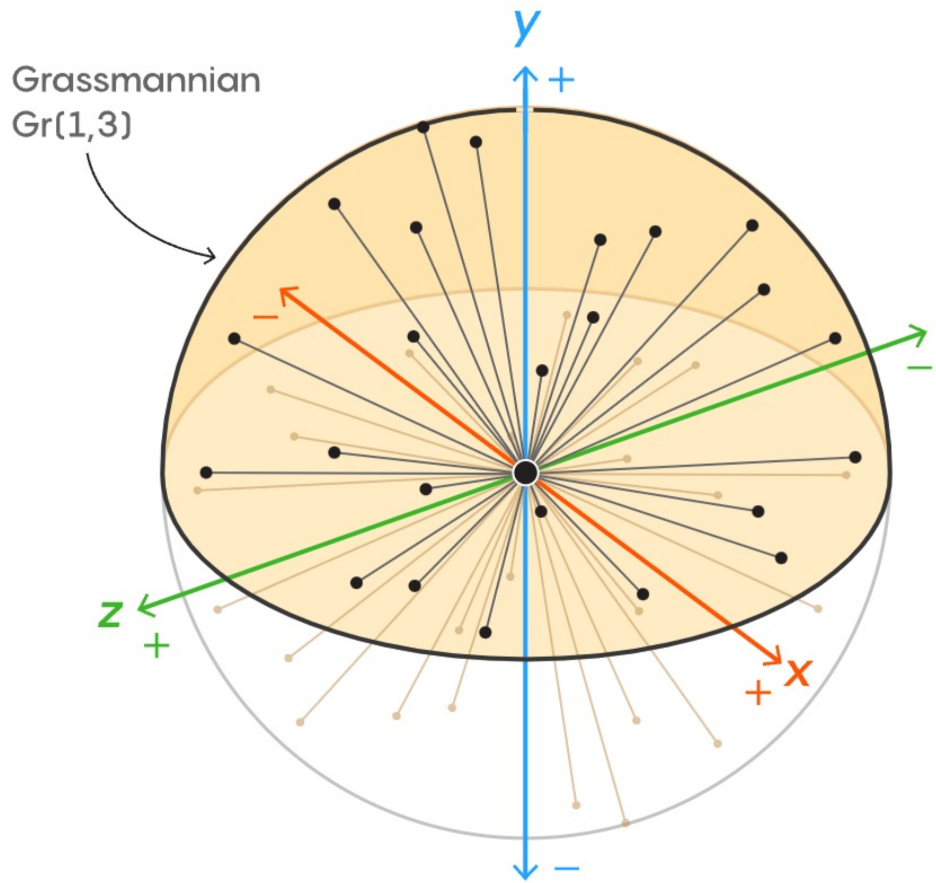
As a set, the **Grassmannian** is defined as the set of all k-dimensional subspaces of $\mathbb{R}^n$,

$$Gr(k,n) \coloneqq \{k - \text{dim} \;\; \text{subspaces of } \mathbb{R}^n\}$$

Suppose $\vec{a}_1 \ldots \vec{a}_k \in \mathbb{R}^n$ and $\vec{b}_1 \ldots \vec{b}_k \in \mathbb{R}^n$ two collections of k linearly independent vectors.

Write subspace $\mathrm{U} = span(\vec{a}_1 \ldots \vec{a}_k)$ and $W = span\left(\vec{b}_1 \ldots \vec{b}_k\right)$. Both V and W are points of the **Grassmannian**.

https://en.wikipedia.org/wiki/Grassmannian


Grassmannian
Gr(1,3)
y
+
−
x
+
−
z
+
−

Grassmann manifold

Stiefel manifold:

$$V(k,n) \coloneqq \{n\times k \text{ orthonormal matrices}\} \subset \mathbb{R}^{n\times k}$$

The set of $\mathrm{k}\times k$ orthogonal matrices (i.e., $AA^T = A^T A = I$) **general orthogonal group** and is usually denoted by $O(k)$.

Grassmann manifold is the quotient

$$Gr(k,n) = V(k,n)/O(k).$$

The natural right action of $O(k)$ on $V(k,n)$ is by rotates a k-frame in the space it spans.

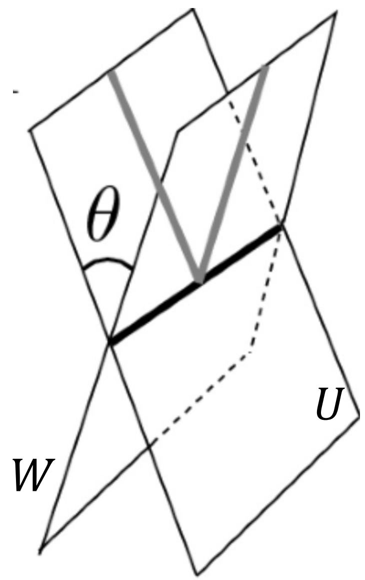
Rich geometry on $Gr(k,n)$: smooth Riemannian manifold, algebraic variety, homogeneous space, geodesic orbit space

➢ Principal vectors and angles

Recall angle θ between two **unit** vectors $\vec{u}$ and $\vec{v}$ is defined by

$$\cos\theta = \frac{\vec{u}^T\vec{v}}{\|\vec{u}\|\|\vec{v}\|} = \vec{u}^T\vec{v}$$

Standard way to measure deviation between two subspaces U and W is using **Principal vectors**.



$\mathrm{U} = span(\vec{u}_1 \ldots \vec{u}_k)$ and $W = span(\vec{w}_1 \ldots \vec{w}_k)$

Principal angles between subspaces

Define **principal vectors** $(\vec{a}_j^*, \vec{b}_j^*)$ recursively as the solutions to the optimization problem:

Maximize: $\vec{a}^T\vec{b}$ (minimize angle between $\vec{a}$ and $\vec{b}$)

subject to

Unit vectors $\vec{a} \in U$ and $\vec{b} \in W$

$$\vec{a}^{\mathrm{T}}\vec{a}_1 = \vec{a}^{\mathrm{T}}\vec{a}_2 = \cdots = \vec{a}^{\mathrm{T}}\vec{a}_{j-1} = 0$$

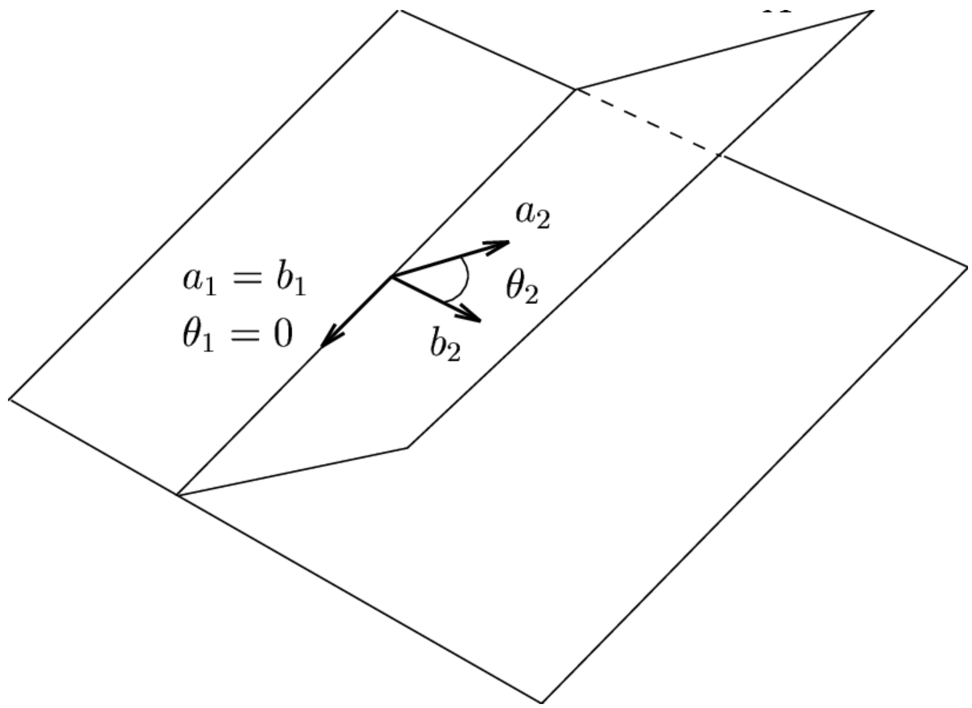
$$\vec{a}^{\mathrm{T}}\vec{a}_1 = \vec{a}^{\mathrm{T}}\vec{a}_2 = \cdots = \vec{a}^{\mathrm{T}}\vec{a}_{j-1} = 0$$

(orthogonal)

Principle angles θ_j defined by

$$\cos\theta_j \coloneqq \vec{a}_j^{*T}\vec{b}_j^{*} \qquad \text{for } j = 1,2,\ldots,k$$

Property: $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$



More generally, principle angles θ_j can be computed using QR and SVD

Åke Björck and Gene H. Golub https://www.jstor.org/stable/2005662

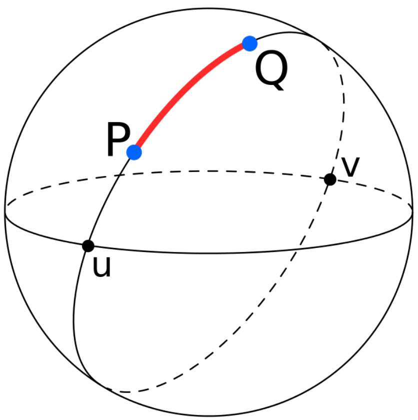
Grassmann distance

The **Grassmann distance** between the linear subspaces U and W is given by:

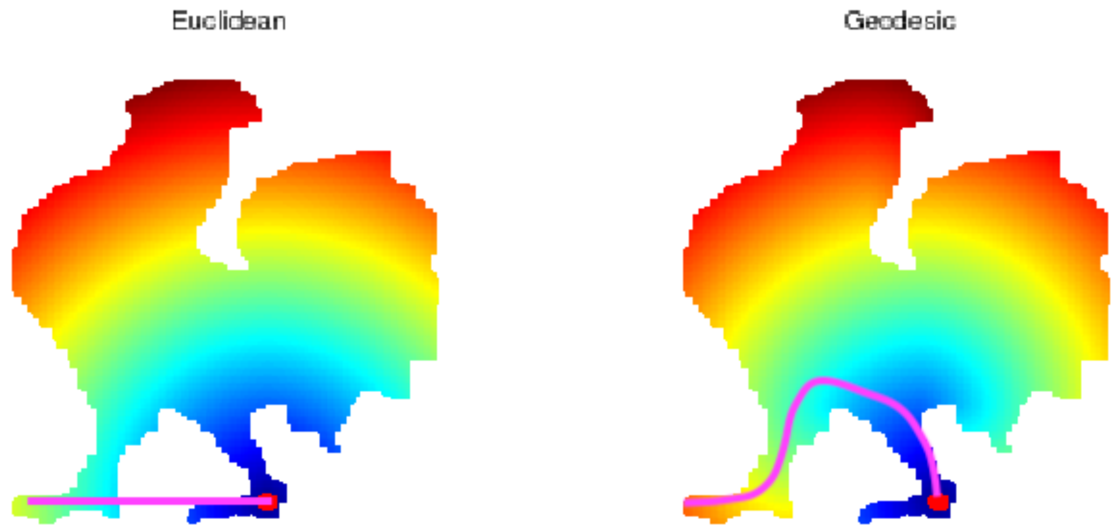
$$dist(U,W) \coloneqq \left[\sum_{i=1}^{k} \theta_i^2\right]^{1/2}$$

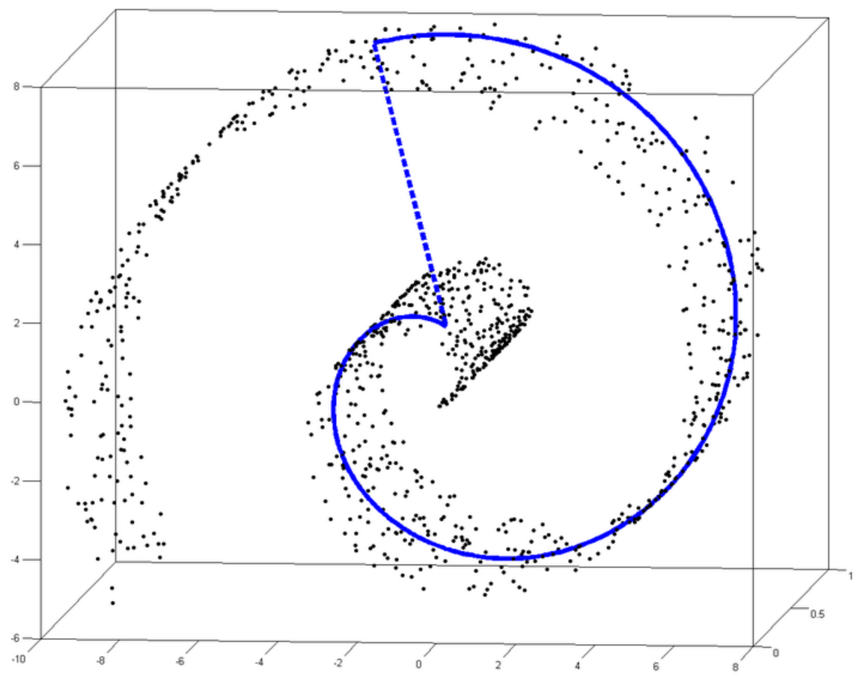
This defines the *geodesic distance* on the Grassmann manifold:

It is *intrinsic*, i.e., does not depend on any embedding



Euclidean v.s. Geodesic





Recall the definition of metric

Definition (Metric). Let S be a **set**. A **metric**(distance) on S is a binary function

$$d\colon S \times S \longrightarrow \mathbb{R}$$

such that for vectors $\vec{u}, \vec{v}, \vec{w} \in S$ and a scalar $c \in \mathbb{R}$, the following hold:

(1.) $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$

(2.) $d(\vec{u}, \vec{v}) = 0$ if and only if $\vec{u} = \vec{v}$

(3.) $d(\vec{u}, \vec{w}) \leq d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w})$

We call S a metric space metric function d.

Grassmann distance	$d_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \left(\sum_{i=1}^{k} \theta_i^2\right)^{1/2}$
Asimov distance	$d^{\alpha}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \theta_k$
Binet–Cauchy distance	$d^{\beta}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \left(1 - \prod_{i=1}^{k} \cos^2\theta_i\right)^{1/2}$
Chordal distance	$d^{\kappa}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \left(\sum_{i=1}^{k} \sin^2\theta_i\right)^{1/2}$
Fubini–Study distance	$d^{\phi}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \cos^{-1}\left(\prod_{i=1}^{k} \cos\theta_i\right)$
Martin distance	$d^{\mu}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \left(\log \prod_{i=1}^{k} 1/\cos^2\theta_i\right)^{1/2}$
Procrustes distance	$d^{\rho}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = 2\left(\sum_{i=1}^{k} \sin^2(\theta_i/2)\right)^{1/2}$
Projection distance	$d^{\pi}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \sin\theta_k$
Spectral distance	$d^{\sigma}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = 2\sin(\theta_k/2)$

Computing principal angles

- Take **orthonormal** bases for subspaces and store them as columns of matrices $A, B \in \mathbb{R}^{n\times k}$ (e.g., QR)
- Then, compute the singular value decomposition (SVD):

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

Note that singular values $0 \le \sigma_j \le 1$ by orthonormality of columns of A and B

- The **principal angles** then satisfy
  - $\cos\theta_j = \sigma_j$
- **Principal vectors** given by the columns

$$AU = [p_1 \dots p_k] \qquad BV = [q_1 \dots q_k]$$

Åke Björck and Gene H. Golub

https://www.jstor.org/stable/2005662

Application:

By separating images into three regions:

2 images of someone's face → $\vec{u}, \vec{v} \in \mathbb{R}^3$

If $\vec{u}, \vec{v}$ are linearly independent, we get a plane $F = span(\vec{u}, \vec{v})$:

For two new photos of someone, again we get a plane and we can take the distance to *F* as a way to compare to the original photos.

But what if I only have one picture of someone, and I want to compare it to the two I started with?

Questions: Distances between non-equi-dimensional subsapces?

Wong, 1967; Differential geometry of Grassmann manifolds. Proc. Nat. Acad. Sci., 57, no. 3, pp. 589– 594

Ye-LHL, *Schubert varieties and distances between subspaces of different dimensions," SIAM J. Matrix Anal. Appl., 37 (2016), no. 3, pp. 1176-1197.*

L.-H. Lim, R. Sepulchre, and K. Ye, Geometric distance between positive defnite matrices of different dimensions," IEEE Trans. Inform. Theory, 2019

The grassmannian of affine subspaces
https://arxiv.org/abs/1807.10883
Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye, (2018)

The distance can be generalized to **Affine** subspaces (shifted vector subspaces), or more generally, Ellipsoids (higher-dimensional ellipses) in the same dimension or different dimensions. .

Reference:

Book: Matrix Computations. by Gene H. Golub and Charles F. Van Loan (fourth edition)

Lecture notes:

https://helper.ipam.ucla.edu/publications/glws1/glws1_15465.pdf

https://web.ma.utexas.edu/users/vandyke/notes/deep_learning_presentation/presentation.pdf

Papers:

Schubert varieties and distances between subspaces of different dimensions
https://www.stat.uchicago.edu/~lekheng/work/schubert.pdf

Grassmannian Learning: Embedding Geometry Awareness in Shallow and Deep Learning
https://arxiv.org/pdf/1808.02229.pdf

Analysis of Temporal Tensor Datasets on Product Grassmann Manifold
https://ieeexplore.ieee.org/document/9856982

Quantum Algorithm for Computing Distances Between Subspaces
https://arxiv.org/abs/2308.15432