Distances between subspaces

Jackson Van Dyke

October 12 and 14, 2020

I originally gave this talk in Professor Yen-Hsi Tsai's course "Mathematics in Deep Learning" (M393) at UT Austin in Fall 2020.
It is based off of *this talk*, by Professor Lek-Heng Lim.

Motivation

- Start with k objects (images, text, etc.) with N features.
- I.e. a collection of k vectors of dimension N.

Example

If we start with k images, we can split it into p squares and take the grayscale values to get k vectors in $\mathbb{R}^p$.

- Then we turn these vectors into some kind of subspace. The three types we will consider are:
  - linear subspaces (vector subspaces),
  - affine subspaces (shifted vector subspaces),
  - ellipsoids (higher-dimensional ellipses).
- Before doing anything else with these subspaces, we want to develop some notion of distance between them.

Table of contents

Review: linear subspaces

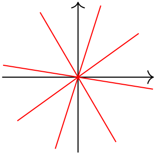
- Consider the real vector space $\mathbb{R}^N$.
- A **linear subspace of** $\mathbb{R}^N$ is a subset which is also a vector space.
- In particular, it **contains** 0**.**

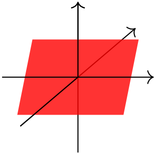
Example

Linear subspaces of $\mathbb{R}^2$ are lines **through the origin**.

Example

The 2-dimensional linear subspaces of $\mathbb{R}^3$ are planes **through the origin**.





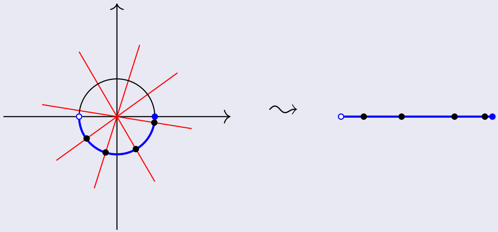
Distance

Question

What is the distance between two linear subspaces?

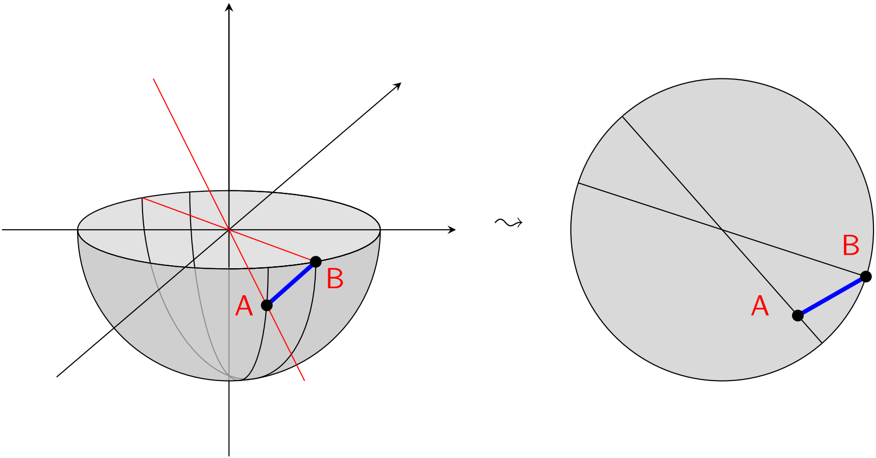
Example

For lines in $\mathbb{R}^2$, we just need to take the angle.



So now we want to formalize this in high dimensions.

Higher-dimensional picture



distance (A,B) = blue.

Higher-dimensional setup

Let $a_1, \ldots, a_k \in \mathbb{R}^N$ and $b_1, \ldots, b_k \in \mathbb{R}^N$ be (separately) linearly independent sets of vectors. Write their spans as:

$$A \coloneqq \operatorname{Span}\{a_1, \ldots, a_k\} \subset \mathbb{R}^N \qquad B \coloneqq \operatorname{Span}\{b_1, \ldots, b_k\} \subset \mathbb{R}^N .$$

Since the vectors were linearly independent, A and B are both k-dimensional linear subspaces of $\mathbb{R}^N$.

Therefore A and B are points of the **Grassmannian**.

$$A, B \in \operatorname{Gr}(k, N) \coloneqq \left\{ k - \text{dim'l linear subspaces of } \mathbb{R}^N \right\} .$$

Principal vectors and angles

- Write $\widehat{a}_1 \in A$ and $\widehat{b}_1 \in B$ for the vectors which

$$\begin{aligned} \text{maximize} \quad & a^T b \\ \text{such that} \quad & \|a\| = \|b\| = 1 \end{aligned}$$

  for $a \in A$, $b \in B$.
- Write $\widehat{a}_2 \in A$ and $\widehat{b}_2 \in B$ for the vectors which

$$\begin{aligned} \text{maximize} \quad & a^T b \\ \text{such that} \quad & \|a\| = \|b\| = 1 \\ & a^T \widehat{a}_1 = 0, \quad b^T \widehat{b}_1 = 0 \end{aligned}$$

  for $a \in A$ and $b \in B$.
- In general we ask for $\widehat{a}_j$ (resp. $\widehat{b}_j$) to be orthogonal to $\widehat{a}_i$ (resp. $\widehat{b}_i$) for all $i < j$.

Grassmann distance

- **Summary of principal vectors:** $\widehat{a}_1$ and $\widehat{b}_1$ are unit vectors which have minimal angle between them. The vectors $\widehat{a}_i$ and $\widehat{b}_i$ are defined the same way, except you insist that they are orthogonal to the previously chosen vectors.
- We can think of the principal vectors as forming a basis which is convenient for measuring angles.
- Define the **principal angles** θ_j by

$$\cos\theta_j = \widehat{a}_j^T \widehat{b}_j \ .$$

  Note that $\theta_1 \leq \ldots \leq \theta_k$.
- The **Grassmann distance** between the linear subspaces A and B is given by:

$$d_k(A, B) = \left(\sum_{i=1}^{k} \theta_i^2\right)^{1/2} \quad .$$

Metric?

We have been using the word "distance" a bit loosely.

Technically, d defines a **metric** on $\mathrm{Gr}(k, N)$ because it satisfies:

1. $d(A, B) = 0$ if and only if $A = B$,
2. $d(A, B) = d(B, A)$, and
3. $d(A, C) \leq d(A, B) + d(B, C)$

for all A, B, and $C \in \mathrm{Gr}(k, N)$.

Computing principal angles

- For any orthonormal basis of A (resp. B) we can store the vectors as columns, to represent A as a matrix M_A (resp. M_B).
- Then we can compute the singular value decomposition (SVD):

$$M_A^T M_B = U \Sigma V^T$$

where

$$\Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{pmatrix} .$$

- The principal angles then satisfy

$$\cos \theta_i = \sigma_i \ .$$

- The principal vectors are the columns of:

$$M_A U \qquad\qquad M_B V \ .$$

An example

- By separating images into three regions:

$$\text{2 images of someone's face } \rightsquigarrow v_1, v_2 \in \mathbb{R}^3$$

- If v_1 and v_2 are linearly independent, we get a plane:

$$F := \operatorname{Span}(v_1, v_2) = \{m_1 v_1 + m_2 v_2 \mid m_1, m_2 \in \mathbb{R}\} \subset \mathbb{R}^3 \ .$$

- 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?

Question

How do we compare subspaces of different dimensions?

Schubert varieties

- For $k \leq \ell$, we would like a notion of distance between
$$A \in \mathrm{Gr}(k, N) \qquad\qquad B \in \mathrm{Gr}(\ell, N) \ .$$
- Consider the set of ℓ-planes containing A:
$$\Omega_+(A) := \{P \in \mathrm{Gr}(\ell, N) \mid A \subseteq P\}$$
and the set of all k-planes containing B:
$$\Omega_-(B) := \{P \in \mathrm{Gr}(k, N) \mid P \subseteq B\} \ .$$
These are called **Schubert varieties**. E.g.
$$\begin{aligned} \Omega_+(\text{the line}) &= \{\text{planes containing the line}\} \\ \Omega_-(\text{plane}) &= \{\text{lines contained in the plane}\} \ . \end{aligned}$$
- **Strategy:** measure distance from A to $\Omega_-(B)$, and B to $\Omega_+(A)$ and compare.

Distance between linear subspaces of different dimensions

The distance from A to $\Omega_-(B)$ is given by:

$$\delta_- = \min\left\{d_k(P, A) \,|\, P \in \Omega_-(B)\right\} \ .$$

and the distance from B to $\Omega_+(A)$ is given by

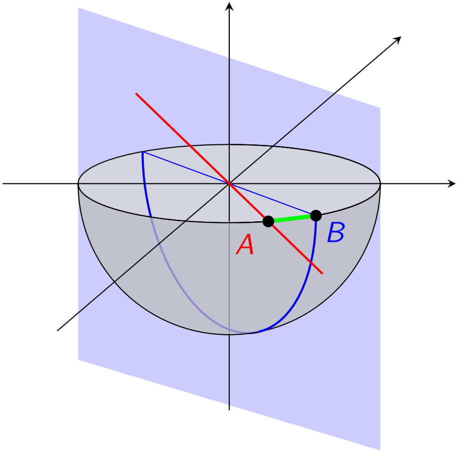
$$\delta_+ = \min\left\{d_\ell(P, B) \,|\, P \in \Omega_+(A)\right\} \ .$$

Theorem 1 (Ye-Lim 2016 [YL16])

$\delta_+ = \delta_-$, *and the common value is:*

$$\delta(A, B) = \left(\sum_{i=1}^{\min(k,\ell)} \theta_i^2\right)^{1/2} \ .$$

Now A is still a line, but B is a plane, both still in $\mathbb{R}^3$.



The distance is the only principal angle that can be defined: the first one. So

$$\delta(\mathrm{A}, \mathrm{B}) = \text{green} \ .$$

Metric?

- Recall d was a metric on $\mathrm{Gr}(k, N)$.
- The space of all linear subspaces in all dimensions is the **doubly infinite Grassmannian**: $\mathrm{Gr}(\infty, \infty) = \sqcup_{k=1}^{\infty} \mathrm{Gr}(k, \infty)$.

Question

Does δ define a metric on $\mathrm{Gr}(\infty, \infty)$?

No: it only satisfies symmetry.

$$\delta(A, B) = 0 \quad \iff \quad A \subseteq B \text{ or } B \subseteq A$$

Counterexample

Let $L_1, L_2 \in \mathrm{Gr}(1, N)$, $P \in \mathrm{Gr}(2, N)$ such that $L_1, L_2 \subset P$.
Triangle inequality $\implies \delta(L_1, L_2) = \delta(L_1, P) = 0$. **Contradiction.**

Premetric.

Instead, δ is what is called a **premetric** (or **distance**) on $\mathrm{Gr}(\infty, \infty)$, since it satisfies:

1. $d(A, B) \geq 0$,
2. $d(A, A) = 0$, and
3. $d(A, B) = d(B, A)$

for all $A, B \in \mathrm{Gr}(\infty, \infty)$.

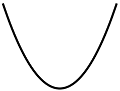
This can be thought of more as a way to measure *separation*, in the sense of the distance between a point and a set.

Metric after all?

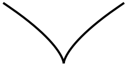
- Recall we can express $\delta(A,B) = \left(\sum_{i=1}^{\min(k,\ell)=k} \theta_i^2\right)^{1/2}$.
- Instead of stopping at the small dimension (k), we can artificially keep going, by defining $\theta_i = \pi/2$ for $i \gneq k$.
- Then
$$d_\infty(A,B) = \left(\sum_{i=1}^{\max(k,\ell)=\ell} \theta_i^2\right)^{1/2} \tag{1}$$
is a **metric** on $\mathrm{Gr}(\infty,\infty)$.
- When restricted to $\mathrm{Gr}(k,\infty)$, this agrees with d_k.
- Note that the topology we got from δ is not metrizable (in fact it is not even Hausdorff).

Digression: Schubert varieties

- In algebraic geometry, Schubert varieties primarily act as one of the most important (and well-studied) **singular varieties**.
- Classically, a **variety** is a subspace (of e.g. $\mathbb{R}^N$) defined as the points where some polynomials vanish.
- These can be nice and smooth: e.g. $y - x^2 = 0$ in $\mathbb{R}^2$.



- Or not nice and **singular**: e.g. $y^3 - x^2 = 0$ in $\mathbb{R}^2$.



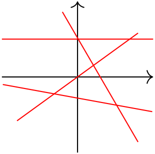
- So these Schubert varieties are actually the subset where some polynomials vanish inside of some huge $\mathbb{R}^D$.

Affine subspaces

- Let $A \in \mathrm{Gr}(k, N)$ be a k-dimensional linear subspace and $b \in \mathbb{R}^N$ to be thought of as the "shift away" from the origin.
- Write $\{a_1, \ldots, a_k\}$ for some basis of A.
- The associated **affine subspace** is:

$$A + b := \left\{ m_1 a_1 + \ldots + m_k a_k + b \in \mathbb{R}^N \;\middle|\; \lambda_i \in \mathbb{R} \right\} \subset \mathbb{R}^N .$$

In particular, they don't have to contain the origin.

E.g. $\mathrm{Graff}(0, N) = \mathbb{R}^N$, and $\mathrm{Graff}(1, N) =$ 

Together, the affine subspaces form the **affine Grassmannian**:

$$\mathrm{Graff}(k, N) = \left\{ k\text{-dim'l affine subspaces of } \mathbb{R}^N \right\} .$$

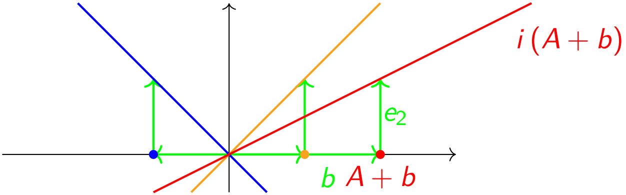
Embedding Graff in (a bigger) Gr

- **Strategy:** view affine subspaces as linear subspaces of a higher-dimensional space, and take d_{Gr}:

$$\mathrm{Graff}\,(k, N) \overset{i}{\hookrightarrow} \mathrm{Gr}\,(k+1, N+1)$$

$$A + b \longmapsto \mathrm{Span}\,(A \cup \{b + e_{n+1}\})$$

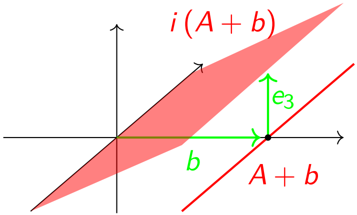
- When $k = 0$ and $N = 1$, i sends points of $\mathbb{R}$ to lines of $\mathbb{R}^2$.
- Given a point $\bullet$, taking this span is the same as drawing a line from the point a unit distance above $\bullet$ through the origin.



Embedding Graff in (a bigger) Gr

$$\mathrm{Graff}(1,2) \xhookrightarrow{i} \mathrm{Gr}(2,3)$$

$$A + b \longmapsto \mathrm{Span}\left(A \cup \{b + e_3\}\right)$$



A metric on Graff

We use this embedding to define the distance between two affine subspaces:

$$d_{\mathrm{Graff}(k,N)}\left(A+b, B+c\right) := d_{\mathrm{Gr}(k+1,N+1)}\left(i\left(A+b\right), i\left(B+c\right)\right) \ .$$

- d_{Graff} is a metric because d_{Gr} is.
- If $b = c = 0$, this is just the usual Grassmannian distance.
- Just as the distance between linear subspaces was calculated using the principal angles, there are **affine principal angles** such that this distance is written as before.
- These angles are also computationally manageable.

An example

- By separating two images into three regions we get $v_1, v_2 \in \mathbb{R}^3$.
- If they are linearly independent, we get a line L which contains those points:

$$L := \{m_1 v_1 + m_2 v_2 \mid m_1 + m_2 = 1, m_1, m_2 \in \mathbb{R}\} \subset \mathbb{R}^3 .$$

  This is the **affine span/hull of v_1 and v_2**, following e.g. [SR20].
- The affine hull is the smallest affine subspace containing the data. In particular, it is contained in the linear subspace F from before.
- For two new photos of someone, again we get a line and we can take the distance to L to compare to the originals.

Question

How do we compare subspaces of different dimensions?

Distance for inequidimensional affine subspaces

For $k \leq \ell$, we would like a notion of distance between

$$A + b \in \mathrm{Graff}\,(k, N) \qquad\qquad B + c \in \mathrm{Graff}\,(\ell, N)\ .$$

As in the linear case, define

$$\begin{aligned}\Omega_+\,(A + b) &:= \{P + q \in \mathrm{Graff}\,(\ell, N) \mid A + b \subseteq P + q\}\\ \Omega_-\,(B + c) &:= \{P + q \in \mathrm{Graff}\,(k, N) \mid P + q \subseteq B + c\}\ .\end{aligned}$$

Theorem 2 (Lim-Wong-Ye 2018 [LWY18])

$d_{\mathrm{Graff}(k,N)}\,(A + b, \Omega_-\,(B + c)) = d_{\mathrm{Graff}(\ell,N)}\,(B + c, \Omega_+\,(A + b))$, *and it is explicitly given via the affine principle angles.*

d_{Graff} is a metric because d_{Gr} is.

Ellipsoids

- $M \in \mathbb{R}^{k \times k}$ is a real symmetric positive definite matrix
  $\iff$ all eigenvalues of M are positive,
  $\iff$ $\forall$ non-zero column vectors z we have $z^T M z > 0$.
- Such a matrix M determines an **ellipsoid**:

$$E_M := \left\{ x \in \mathbb{R}^k \,\middle|\, x^T M x \leq 1 \right\} \ .$$

Example

If M is the identity matrix, then this is just the closed ball of dimension N.

- We will define a distance between E_A and E_B by finding one between the matrices A and B.

PDS cone and distance between ellipsoids

$\mathbb{S}^k_{++}$ = the cone of real symmetric positive definite matrices.

Metric on $\mathbb{S}^k_{++}$ given by:

$$\mathbb{S}^k_{++} \times \mathbb{S}^k_{++} \xrightarrow{\quad \delta_2 \quad} \mathbb{R}_+$$

$$(A, B) \xmapsto{\delta_2} \left(\sum_{j=1}^n \log^2\left(\lambda_j\left(A^{-1}B\right)\right)\right)^{1/2} .$$

This is a convenient metric because it is very **invariant**. I.e. it satisfies:

$$\delta_2\left(XAX^T, XBX^T\right) = \delta_2(A, B)$$
$$\delta_2\left(XAX^{-1}, XBX^{-1}\right) = \delta_2(A, B)$$
$$\delta_2\left(A^{-1}, B^{-1}\right) = \delta_2(A, B) .$$

δ_2 has applications to computer vision, medical imaging, radar signal processing, statistical inference, and other areas.

An example

- Assume we are given k articles written about Halloween, and we count the occurrences of the terms
  - pumpkins,
  - skeletons, and
  - trick-or-treating

  to yield k vectors in $\mathbb{R}^3$.
- Write E for the smallest ellipsoid in $\mathbb{R}^3$ containing these vectors. If they are linearly independent, it is k-dimensional.
- For some other collection of k articles, we can count the same three words and form a second ellipsoid. Then we can measure the distance to E.
- The inverse of the distance gives the likelihood that the new articles are about Halloween.
- If we wanted to compare fewer than k articles to the originals, we would have needed to compare E to an ellipsoid of dim $\lneq k$.

Sub-ellipsoids

- There is a partial order on $\mathbb{S}^k_{++}$ given by:

$$A \preccurlyeq B \qquad \Longleftrightarrow \qquad B - A \in \mathbb{S}^k_+ \ ,$$

  where $\mathbb{S}^k_+$ consists of real symmetric positive semi-definite matrices.
- $A \preccurlyeq B$ iff $E_B \subseteq E_A$.
- If I want to compare $A \in \mathbb{S}^k_{++}$ to $M \in \mathbb{S}^\ell_{++}$ (for $k \leq \ell$) then I can write

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \ , \tag{2}$$

  where M_{11} is the upper left $k \times k$ block of M, and compare A to M_{11}.
- We will use this notion of containment to define the analogues of Schubert varieties.

Analogue of Schubert varieties

For $k \leq \ell$, we would like a notion of distance between

$$A \in \mathbb{S}^k_{++} \qquad\qquad B \in \mathbb{S}^\ell_{++} .$$

Define the convex set of ellipsoids containing/contained in E_A / E_B:

$$\Omega_+ (A) \coloneqq \left\{ M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \in \mathbb{S}^\ell_{++} \,\middle|\, M_{11} \preccurlyeq A \right\}$$

$$\Omega_- (B) \coloneqq \left\{ M \in \mathbb{S}^k_{++} \,\middle|\, B_{11} \preccurlyeq M \right\}$$

where B_{11} is the upper left $k \times k$ block of B, M_{11} is the upper left $k \times k$ block of M.

Distance between inequidimensional ellipsoids

Theorem 3 (Lim-Sepulchre-Ye 2019 [LSY19])

$\delta_2(A, \Omega_-(B)) = \delta_2(B, \Omega_+(A))$. *The common value is*

$$\delta_2^+(A, B) = \left(\sum_{j=1}^{k} \log^2 \lambda_j \left(A^{-1} B_{11}\right)\right)^{1/2}$$

where k is such that

$$\lambda_j \left(A^{-1} B_{11}\right) \leq 1$$

for $j = k+1, \ldots, m$.

- We are implicitly putting the smaller-dimensional matrix in the first argument of δ_2^+,
- So asking for it to be symmetric does not make sense, i.e. it is not a metric on $\sqcup_{k=1}^{\infty} \mathbb{S}_{++}^{k}$.

Future directions

- A **category** is (roughly) a collection of objects and arrows between the objects which satisfy some conditions.
- In [DHKK13], the authors define a notion of distance between any two objects of a category.

Example

The collection of half-dimensional subspaces of a given even-dimensional manifold[a] fit naturally into a category called the **Fukaya category**. Roughly, we have an object for every subspace, and an arrow whenever they intersect.

[a] Technically they're Lagrangians in a symplectic manifold.

Question

Is this a useful distance for our purposes? Is it computable?

Summary:

Assume we have a way to pass from raw data to a subspace:

$$\text{raw data} \quad \rightsquigarrow \quad \{v_i\} \in \mathbb{R}^N \quad \rightsquigarrow \quad \text{subspace} \subseteq \mathbb{R}^N$$

When the subspace is linear, affine, or an ellipsoid, there is a metric (or premetric) which on the space of such subspaces (of any dimension!) which is realistic to calculate.

So we can distinguish data by measuring the distance between the associated subspaces.

References

George Dimitrov, Fabian Haiden, Ludmil Katzarkov, and Maxim Kontsevich, *Dynamical systems and categories*, 2013.

L. Lim, R. Sepulchre, and K. Ye, *Geometric distance between positive definite matrices of different dimensions*, IEEE Transactions on Information Theory **65** (2019), no. 9, 5401–5405.

Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye, *The grassmannian of affine subspaces*, 2018.

K. Sharma and R. Rameshan, *Image set classification using a distance-based kernel over affine grassmann manifold*, IEEE Transactions on Neural Networks and Learning Systems (2020), 1–14.

Ke Ye and Lek-Heng Lim, *Schubert varieties and distances between subspaces of different dimensions*, SIAM J. Matrix Anal. Appl. **37** (2016), no. 3, 1176–1197. MR 3543155