distances between objects of different dimensions

Lek-Heng Lim

University of Chicago

joint work with: Ke Ye (CAS) and Rodolphe Sepulchre (Cambridge)

thanks: DARPA D15AP00109, NSF DMS-1209136, NSF IIS-1546413,
Frank Sottile (Texas A&M)

overview

- fundamental problem: before we set out to do anything with objects, we usually need a notion of separation between them
- usually well-known for same type of objects of the *same dimension*
- what about same type of objects of different dimensions?

problem 1 distance between two linear subspaces of different dimensions?

problem 2 metric between two linear subspaces of different dimensions?

problem 3 distance between two affine subspaces of the same dimension?

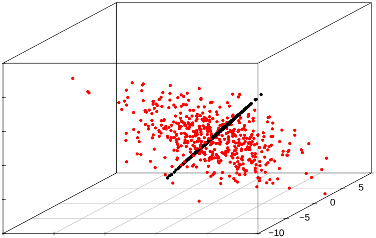
problem 4 distance between two affine subspaces of different dimensions?

problem 5 distance between two covariance matrices of different dimensions?

linear subspaces

why subspaces

- k objects: genes, tweets, images, etc
- n features: expression levels, term frequencies, frames, etc
- jth object described by feature vector $a_j = [a_{1j}, \dots, a_{nj}]^{\mathsf{T}} \in \mathbb{R}^n$
- data set described by $A = [a_1, \dots, a_k] \in \mathbb{R}^{n \times k}$
  - massive: n large
  - high-dimensional: k large
- often what is important is not A but subspace defined by A
  - $\operatorname{span}\{a_1, \dots, a_k\}$ or $\mu + \operatorname{span}\{a_1 - \mu, \dots, a_k - \mu\}$
  - principal subspaces of A defined by eigenvectors of covariance matrix



classical problem

Problem

$a_1, \dots, a_k \in \mathbb{R}^n$ and $b_1, \dots, b_k \in \mathbb{R}^n$ two collections of k linearly independent vectors; want measure of separation of subspace spanned by $a_1, \dots, a_k$ and the subspace spanned by $b_1, \dots, b_k$

- two possible solutions: distances or angles between subspaces *of the same dimension*
- turns out to be equivalent
- classical problem in matrix computations [Golub–Van Loan, 2013]
- notations:
  - write $\langle a_1, \dots, a_k \rangle \coloneqq \operatorname{span}\{a_1, \dots, a_k\}$
  - subspace $A \subseteq \mathbb{R}^n$, write $P_A \in \mathbb{R}^{n \times n}$ for orthogonal projection onto A

principal angles between subspaces

- standard way to measure deviation between two subspaces
- measure principal angles $\theta_1, \dots, \theta_k \in [0, \pi/2]$ between them
- define principal vectors (a_j^*, b_j^*) recursively as the solutions to the optimization problem

$$\begin{array}{rl} \text{maximize} & a^\mathsf{T} b \\ \text{subject to} & a \in \langle a_1, \dots, a_k \rangle, \; b \in \langle b_1, \dots, b_k \rangle, \\ & a^\mathsf{T} a_1 = \cdots = a^\mathsf{T} a_{j-1} = 0, \; \|a\| = 1, \\ & b^\mathsf{T} b_1 = \cdots = b^\mathsf{T} b_{j-1} = 0, \; \|b\| = 1, \end{array}$$

  for $j = 1, \dots, k$
- principle angles given by

$$\cos\theta_j = {a_j^*}^\mathsf{T} b_j^*, \quad j = 1, \dots, k$$

- clearly $\theta_1 \le \cdots \le \theta_k$

readily computable

- may be computed using QR and SVD [Björck–Golub, 1973]
- take orthonormal bases for subspaces and store them as columns of matrices $A, B \in \mathbb{R}^{n \times k}$ (e.g., Householder QR)
- let SVD of $A^\mathsf{T} B \in \mathbb{R}^{k \times k}$ be

$$A^\mathsf{T} B = U \Sigma V^\mathsf{T}$$

  where $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_k)$ and $\sigma_1 \geq \cdots \geq \sigma_k$ are the singular values
- note $0 \leq \sigma_i \leq 1$ by orthonormality of columns of A and B
- principal angles given by

$$\cos \theta_i = \sigma_i, \quad i = 1, \dots, k$$

- principal vectors given by

$$AU = [p_1, \dots, p_k], \quad BV = [q_1, \dots, q_l]$$

basic geometry of subspaces

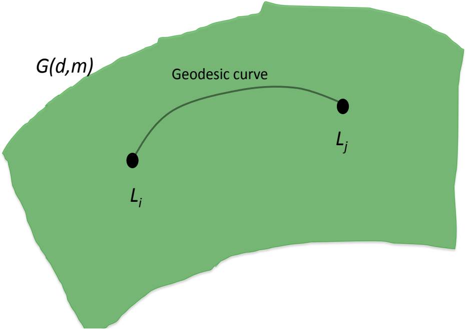
- k-dimensional linear subspace $\mathbb{A}$ in $\mathbb{R}^n$ is an element of the Grassmann manifold $\mathrm{Gr}(k, n)$
- Stiefel manifold: $\mathrm{V}(k, n)$ set of $n \times k$ orthonormal matrices $A \in \mathbb{R}^{n \times k}$
- Grassmann manifold: $\mathbb{A} = \mathrm{span}(A) \in \mathrm{Gr}(k, n)$ represents an equivalence class
$$\mathrm{Gr}(k, n) = \mathrm{V}(k, n)/\,\mathrm{O}(k)$$
- rich geometry: smooth Riemannian manifold, algebraic variety, homogeneous space, geodesic orbit space

Grassmann distance

- geodesic distance: along geodesic between $\mathbb{A}$ and $\mathbb{B}$ on $\mathrm{Gr}(k, n)$

$$d_{\mathrm{Gr}(k,n)}(\mathbb{A}, \mathbb{B}) = \left[\sum_{i=1}^{k} \theta_i^2\right]^{1/2}$$

- $d_{\mathrm{Gr}(k,n)}$ is *intrinsic*, i.e., does not depend on any embedding
- but $d_{\mathrm{Gr}(k,n)}(\mathbb{A}, \mathbb{B})$ undefined for $\mathbb{A} \in \mathrm{Gr}(k, n)$, $\mathbb{B} \in \mathrm{Gr}(l, n)$, $k \neq l$



distances between equidimensional subspaces

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	$d^{\sigma}_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = 2\sin(\theta_k/2)$

distances between nonequidimensional subapces?

- dependence on principal angles not a coincidence

Theorem (Wong, 1967; Ye–LHL, 2016)

any valid distance function $d(\mathbb{A}, \mathbb{B})$ on subspaces must be a function of only their principal angles

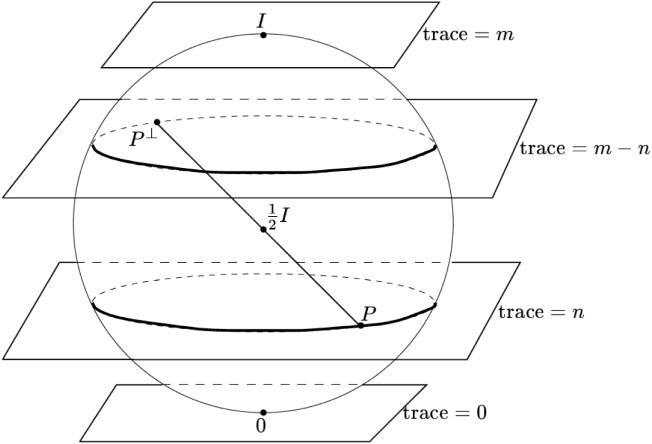
$$A^{\mathsf{T}} B = U(\cos\Theta)V^{\mathsf{T}}, \quad \Theta = \operatorname{diag}(\theta_1, \dots, \theta_k)$$

- however none works for subspaces of different dimensions
- one solution: embed $\mathrm{Gr}(n, 0), \mathrm{Gr}(n, 1), \dots, \mathrm{Gr}(n, n)$ in some bigger space and measure distance in that space

J. Conway, R. Hardin, N. Sloane. "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Exp. Math.*, **5** (1996), no. 2, pp. 139–159

example

- simultaneous embedding of $\mathrm{Gr}(n,0), \mathrm{Gr}(n,1), \dots, \mathrm{Gr}(n,n)$ into sphere in $\mathbb{R}^{(n-1)(n+2)/2}$ as orthogonal projectors
- chordal distance $\|AA^{\mathsf{T}} - BB^{\mathsf{T}}\|_F = \sqrt{2} d^{\kappa}_{\mathrm{Gr}(k,n)}(\mathbb{A}, \mathbb{B})$



instrinsic distance?

- want an intrinsic distance for subspaces of different dimensions
- must agree with the geodesic distance on $\mathrm{Gr}(k,n)$ when both subspaces are of the same dimension

$$d_{\mathrm{Gr}(k,n)}(\mathbb{A},\mathbb{B}) = \Bigl[\sum\nolimits_{i=1}^{k} \theta_i^2\Bigr]^{1/2}$$

- solution: inspired by Schubert calculus

what we propose

- given subspaces $\mathbb{A}$ of dimension k and $\mathbb{B}$ of dimension l
- WLOG assume $k < l$
- define

$$\Omega_+(\mathbb{A}) := \{\mathbb{Y} \in \mathrm{Gr}(l, n) : \mathbb{A} \subseteq \mathbb{Y}\}$$
$$\Omega_-(\mathbb{B}) := \{\mathbb{X} \in \mathrm{Gr}(k, n) : \mathbb{X} \subseteq \mathbb{B}\}$$

- $\Omega_+(\mathbb{A})$ and $\Omega_-(\mathbb{B})$ are Schubert varieties in $\mathrm{Gr}(l, n)$ and $\mathrm{Gr}(k, n)$ respectively
- two possibilities for our distance:

$$\delta_+(\mathbb{A}, \mathbb{B}) = \min\{d_{\mathrm{Gr}(l,n)}(\mathbb{X}, \mathbb{B}) : \mathbb{X} \in \Omega_+(\mathbb{A})\}$$
$$\delta_-(\mathbb{A}, \mathbb{B}) = \min\{d_{\mathrm{Gr}(k,n)}(\mathbb{Y}, \mathbb{A}) : \mathbb{Y} \in \Omega_-(\mathbb{B})\}$$

intrinsic distance for inequidimensional subspaces

Theorem (Ye–LHL, 2016)

for any two subspaces $\mathbb{A}$ *of dimension* k *and* $\mathbb{B}$ *of dimension* l,

$$\delta_+(\mathbb{A},\mathbb{B}) = \delta_-(\mathbb{A},\mathbb{B})$$

- denote common value by $\delta(\mathbb{A},\mathbb{B})$

Theorem (Ye–LHL, 2016)

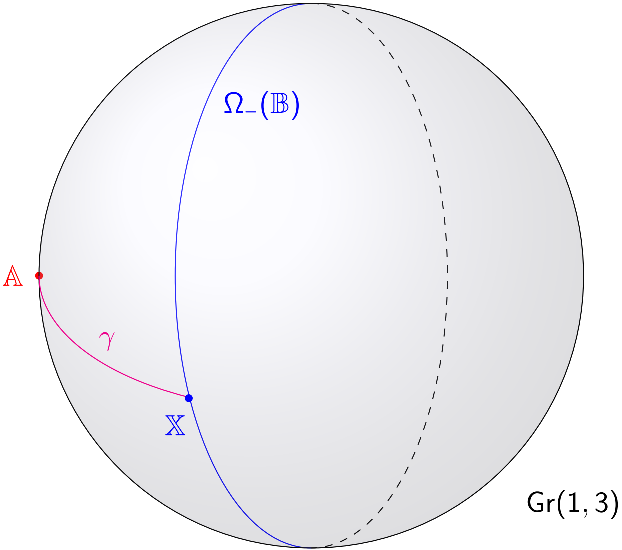
for any two subspaces $\mathbb{A}$ *of dimension* k *and* $\mathbb{B}$ *of dimension* l,

$$\delta(\mathbb{A},\mathbb{B}) = \Bigl[\sum_{i=1}^{\min(k,l)} \theta_i^2\Bigr]^{1/2}$$

where

$$A^{\mathsf{T}}B = U(\cos\Theta)V^{\mathsf{T}}, \quad \Theta = \operatorname{diag}(\theta_1,\dots,\theta_{\min(k,l)},1,\dots,1)$$

pictorial view



properties

- agrees with $d_{\mathrm{Gr}(k,n)}$ when $k = l$
- easily computable via singular value decomposition
- does not depend on n: inclusion

$$i : \mathrm{Gr}(k, n) \to \mathrm{Gr}(k, n+1)$$

  is isometric and $d_{\mathrm{Gr}(k,\infty)}$ defines metric on $\mathrm{Gr}(k, \infty)$
- distance in the sense of distance of a point to a set

extends to all other distances

Theorem (Ye–LHL, 2016)

does not matter if $d_{\mathrm{Gr}(k,n)}$ is replaced by any other distances, always have $\delta^*_+(\mathbb{A},\mathbb{B}) = \delta^*_-(\mathbb{A},\mathbb{B})$, $* = \alpha,\beta,\kappa,\mu,\pi,\rho,\sigma,\phi$

Asimov distance	$\delta^\alpha(\mathbb{A},\mathbb{B}) = \theta_{\min(k,l)}$
Binet–Cauchy distance	$\delta^\beta(\mathbb{A},\mathbb{B}) = \left(1 - \prod_{i=1}^{\min(k,l)} \cos^2\theta_i\right)^{1/2}$
Chordal distance	$\delta^\kappa(\mathbb{A},\mathbb{B}) = \left(\sum_{i=1}^{\min(k,l)} \sin^2\theta_i\right)^{1/2}$
Fubini–Study distance	$\delta^\phi(\mathbb{A},\mathbb{B}) = \cos^{-1}\left(\prod_{i=1}^{\min(k,l)} \cos\theta_i\right)$
Martin distance	$\delta^\mu(\mathbb{A},\mathbb{B}) = \left(\log\prod_{i=1}^{\min(k,l)} 1/\cos^2\theta_i\right)^{1/2}$
Procrustes distance	$\delta^\rho(\mathbb{A},\mathbb{B}) = 2\left(\sum_{i=1}^{\min(k,l)} \sin^2(\theta_i/2)\right)^{1/2}$
Projection distance	$\delta^\pi(\mathbb{A},\mathbb{B}) = \sin\theta_{\min(k,l)}$
Spectral distance	$\delta^\sigma(\mathbb{A},\mathbb{B}) = 2\sin(\theta_{\min(k,l)}/2)$

metric?

- δ is a premetric but not a metric on doubly infinite Grassmannian

$$\mathrm{Gr}(\infty, \infty) := \coprod_{k=1}^{\infty} \mathrm{Gr}(k, \infty)$$

  which parameterizes subspaces of all dimensions
- e.g., $\delta(\mathbb{A}, \mathbb{B}) = 0$ if $\mathbb{A} \subset \mathbb{B}$, triangle inequality not satisfied
- no mathematically natural way to make $\mathrm{Gr}(\infty, \infty)$ into a metric space: category of metric space does not admit coproduct

our proposal

- given two subspaces in $\mathbb{R}^n$, $\mathbb{A}$ of dimension k and $\mathbb{B}$ of dimension l
- WLOG assume $k < l$, principal angles $\theta_1, \dots, \theta_k$, now *define*

$$\theta_{k+1} = \cdots = \theta_l = \pi/2$$

- get metrics on $\mathrm{Gr}(\infty, \infty)$ [Ye–LHL, 2016]

$$d_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A}, \mathbb{B}) = \Bigl(\sum\nolimits_{i=1}^{l} \theta_i^2\Bigr)^{1/2} = \Bigl((l-k)\pi^2/4 + \sum\nolimits_{i=1}^{k} \theta_i^2\Bigr)^{1/2}$$

$$d^{\kappa}_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A}, \mathbb{B}) = \Bigl(\sum\nolimits_{i=1}^{l} \sin^2\theta_i\Bigr)^{1/2} = \Bigl(l-k + \sum\nolimits_{i=1}^{k} \sin^2\theta_i\Bigr)^{1/2}$$

$$d^{\rho}_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A}, \mathbb{B}) = \Bigl(2\sum\nolimits_{i=1}^{l} \sin^2(\theta_i/2)\Bigr)^{1/2} = \Bigl(l-k + 2\sum\nolimits_{i=1}^{k} \sin^2(\theta_i/2)\Bigr)^{1/2}$$

- essentially root mean square of two pieces of information: $\delta^*(\mathbb{A}, \mathbb{B})$ and $\epsilon(\mathbb{A}, \mathbb{B}) := |\dim\mathbb{A} - \dim\mathbb{B}|^{1/2}$

$$d^*_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A}, \mathbb{B}) = \sqrt{\delta^*(\mathbb{A}, \mathbb{B})^2 + c_*^2\epsilon(\mathbb{A}, \mathbb{B})^2}$$

moreover

- what about $* = \alpha, \beta, \phi, \mu, \pi, \sigma$?
- not very interesting:

$$d^*_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A},\mathbb{B}) = \begin{cases} d^*_{\mathrm{Gr}(k,\infty)}(\mathbb{A},\mathbb{B}) & \text{if } \dim \mathbb{A} = \dim \mathbb{B} = k \\ c_* & \text{if } \dim \mathbb{A} \neq \dim \mathbb{B} \end{cases}$$

- constants $c_* > 0$ given by

$$c = c_\alpha = \pi/2, \quad c_\beta = c_\phi = c_\pi = c_\kappa = c_\rho = 1, \quad c_\sigma = \sqrt{2}, \quad c_\mu = \infty$$

- how to interpret?

$$\max_{\mathbb{X}\in\Omega_+(\mathbb{A})} d^*_{\mathrm{Gr}(l,n)}(\mathbb{X},\mathbb{B}) = d^*_{\mathrm{Gr}(\infty,\infty)}(\mathbb{A},\mathbb{B}) = \max_{\mathbb{Y}\in\Omega_-(\mathbb{B})} d^*_{\mathrm{Gr}(k,n)}(\mathbb{Y},\mathbb{A})$$

provided $n > 2l$

summary in english

- given two subspaces in $\mathbb{R}^n$, $\mathbb{A}$ of dimension k and $\mathbb{B}$ of dimension l
- distance of $\mathbb{A}$ to nearest k-dimensional subspace contained in $\mathbb{B}$ *equals* distance of $\mathbb{B}$ to nearest l-dimensional subspace containing $\mathbb{A}$
- common value gives distance between $\mathbb{A}$ and $\mathbb{B}$
- distance of $\mathbb{A}$ to furthest k-dimensional subspace contained in $\mathbb{B}$ *equals* distance of $\mathbb{B}$ to furthest l-dimensional subspace containing $\mathbb{A}$
- common value gives metric between $\mathbb{A}$ and $\mathbb{B}$

volumetric analogue

- $\mu_{k,n}$ natural probability density on $\mathrm{Gr}(k, n)$
- what we showed [Ye–LHL, 2016],

$$\mu_{l,n}\big(\Omega_+(\mathbb{A})\big) = \mu_{k,n}\big(\Omega_-(\mathbb{B})\big)$$

- probability a random l-dimensional subspace contains $\mathbb{A}$ *equals* probability a random k-dimensional subspace is contained in $\mathbb{B}$
- common value does not depend on the choices of $\mathbb{A}$ and $\mathbb{B}$ but only on k, l, n and is given by

$$\frac{l!(n-k)!\prod_{j=l-k+1}^{l}\omega_j}{n!(l-k)!\prod_{j=n-k+1}^{n}\omega_j}$$

$\omega_m := \pi^{m/2}/\Gamma(1+m/2)$ is volume of unit 2-norm ball in $\mathbb{R}^m$

affine subspaces

affine subspaces

- $\mathbb{A} \in \mathrm{Gr}(k, n)$ k-dimensional linear subspace, $b \in \mathbb{R}^n$ displacement of $\mathbb{A}$ from the origin
- $A = [a_1, \dots, a_k] \in \mathbb{R}^{n \times k}$ basis of $\mathbb{A}$, then a k-dimensional affine subspace is

$$\mathbb{A} + b \coloneqq \{\lambda_1 a_1 + \cdots + \lambda_k a_k + b \in \mathbb{R}^n : \lambda_1, \dots, \lambda_k \in \mathbb{R}\}$$

- $[A, b_0]$ orthogonal affine coordinates if

$$[A, b_0] \in \mathbb{R}^{n \times (k+1)}, \quad A^{\mathsf{T}} A = I, \quad A^{\mathsf{T}} b_0 = 0$$

- $\mathrm{Graff}(k, n)$ Grassmannian of affine subspaces in $\mathbb{R}^n$ is set of all k-dimensional affine subspaces of $\mathbb{R}^n$

Grassmannian of affine subspaces

- $\mathrm{Graff}(k, n)$ is smooth manifold
- $\mathrm{Graff}(k, n)$ is universal quotient bundle of $\mathrm{Gr}(k, n)$

$$0 \to S \to \mathrm{Gr}(k, n) \times \mathbb{R}^n \to \mathrm{Graff}(k, n) \to 0$$

- $\mathrm{Graff}(k, n)$ is homogeneous space

$$\mathrm{Graff}(k, n) \cong \mathrm{E}(n)/(\mathrm{O}(n-k) \times \mathrm{E}(k))$$

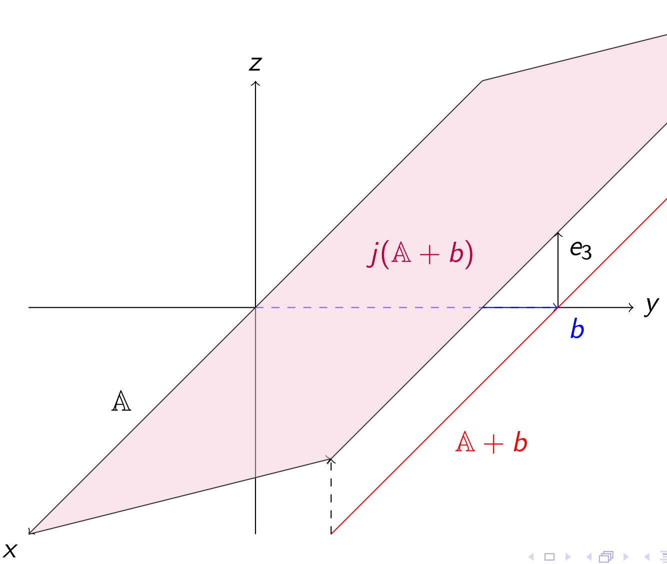
  where $\mathrm{E}(n)$ is group of orthogonal affine transformations
- $\mathrm{Graff}(k, n)$ is Zariski open subset of $\mathrm{Gr}(k+1, n+1)$

$$\mathrm{Gr}(k+1, n+1) = X \cup X^c, \quad X \cong \mathrm{Graff}(k, n), \quad X^c \cong \mathrm{Gr}(k+1, n)$$

- first three do not give useful distance on $\mathrm{Graff}(k, n)$, last one does

embedding Graff(k, n) into Gr$(k+1, n+1)$

$$j : \mathrm{Graff}(k,n) \to \mathrm{Gr}(k+1,n+1), \quad \mathbb{A} + b \mapsto \mathrm{span}(\mathbb{A} \cup \{b + e_{n+1}\})$$



distance between affine subspaces

- define distance between two k-dimensional affine subspaces as

$$d_{\mathrm{Graff}(k,n)}(\mathbb{A}+b,\mathbb{B}+c) := d_{\mathrm{Gr}(k+1,n+1)}\bigl(j(\mathbb{A}+b), j(\mathbb{B}+c)\bigr)$$

- reduces to Grassmann distance when $b = c = 0$
- if $[A, b_0]$ and $[B, c_0] \in \mathbb{R}^{n\times(k+1)}$ are orthogonal affine coordinates, then

$$d_{\mathrm{Graff}(k,n)}(\mathbb{A}+b,\mathbb{B}+c) = \Bigl(\sum_{i=1}^{k+1} \phi_i^2\Bigr)^{1/2}$$

- affine principal angles defined by

$$\phi_i = \cos^{-1}\tau_i, \quad i = 1,\dots,k+1,$$

where $\tau_1 \geq \cdots \geq \tau_{k+1}$ are singular values of

$$\begin{bmatrix} A & b_0/\sqrt{1+\|b_0\|^2} \\ 0 & 1/\sqrt{1+\|b_0\|^2}\end{bmatrix}^{\mathsf{T}} \begin{bmatrix} B & c_0/\sqrt{1+\|c_0\|^2} \\ 0 & 1/\sqrt{1+\|c_0\|^2}\end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$$

affine subspaces of different dimensions?

$$\Omega_+(\mathbb{A}+b) := \{\mathbb{X}+y \in \operatorname{Graff}(l,n) : \mathbb{A}+b \subseteq \mathbb{X}+y\}$$
$$\Omega_-(\mathbb{B}+c) := \{\mathbb{Y}+z \in \operatorname{Graff}(k,n) : \mathbb{Y}+z \subseteq \mathbb{B}+c\}$$

Theorem (LHL–Wong–Ye, 2018)

$k \le l \le n$, $\mathbb{A}+b \in \operatorname{Graff}(k,n)$, $\mathbb{B}+c \in \operatorname{Graff}(l,n)$, then

$$d_{\operatorname{Graff}(k,n)}\bigl(\mathbb{A}+b, \Omega_-(\mathbb{B}+c)\bigr) = d_{\operatorname{Graff}(l,n)}\bigl(\mathbb{B}+c, \Omega_+(\mathbb{A}+b)\bigr),$$

and their common value is

$$\delta(\mathbb{A}+b, \mathbb{B}+c) = \Bigl(\sum_{i=1}^{\min(k,l)+1} \phi_i^2\Bigr)^{1/2},$$

where $\phi_1, \dots, \phi_{\min(k,l)+1}$ are affine principal angles corresponding to

$$\begin{bmatrix} A & b_0/\sqrt{1+\|b_0\|^2} \\ 0 & 1/\sqrt{1+\|b_0\|^2} \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} B & c_0/\sqrt{1+\|c_0\|^2} \\ 0 & 1/\sqrt{1+\|c_0\|^2} \end{bmatrix} \in \mathbb{R}^{(k+1)\times(l+1)}$$

works with other distances too

common value $\delta^*(\mathbb{A} + b, \mathbb{B} + c)$ given by:

$$\delta^\alpha(\mathbb{A} + b, \mathbb{B} + c) = \phi_{k+1}, \qquad \delta^\beta(\mathbb{A} + b, \mathbb{B} + c) = \Bigl(1 - \prod_{i=1}^{k+1} \cos^2 \phi_i\Bigr)^{1/2},$$

$$\delta^\pi(\mathbb{A} + b, \mathbb{B} + c) = \sin \phi_{k+1}, \qquad \delta^\mu(\mathbb{A} + b, \mathbb{B} + c) = \Bigl(\log \prod_{i=1}^{k+1} \frac{1}{\cos^2 \phi_i}\Bigr)^{1/2},$$

$$\delta^\sigma(\mathbb{A} + b, \mathbb{B} + c) = 2 \sin(\phi_{k+1}/2), \qquad \delta^\phi(\mathbb{A} + b, \mathbb{B} + c) = \cos^{-1}\Bigl(\prod_{i=1}^{k+1} \cos \phi_i\Bigr),$$

$$\delta^\kappa(\mathbb{A} + b, \mathbb{B} + c) = \Bigl(\sum_{i=1}^{k+1} \sin^2 \phi_i\Bigr)^{1/2}, \quad \delta^\rho(\mathbb{A} + b, \mathbb{B} + c) = \Bigl(2 \sum_{i=1}^{k+1} \sin^2(\phi_i/2)\Bigr)^{1/2}$$

ellipsoids

same thing different names

- real symmetric positive definite matrices $A \in \mathbb{R}^{n \times n}$
- ellipsoids centered at the origin in $\mathbb{R}^n$,
$$\mathcal{E}_A := \{x \in \mathbb{R}^n : x^{\mathsf{T}} A x \le 1\}$$
- inner products on $\mathbb{R}^n$,
$$\langle \cdot, \cdot \rangle_A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x, y) \mapsto x^{\mathsf{T}} A y$$
- covariances of nondegenerate random variables $X : \Omega \to \mathbb{R}^n$,
$$A = \operatorname{Cov}(X) = E[(X - \mu)(X - \mu)^{\mathsf{T}}]$$
- many more: diffusion tensors, sums-of-squares polynomials, mean-centered Gaussians, etc

PSD cone

- $\mathbb{S}^n$ vector space of real symmetric or complex Hermitian matrices
- $\mathbb{S}^n_{++}$ cone of real symmetric positive definite or complex Hermitian positive definite matrices
- rich geometric structures
  - Riemannian manifold
  - symmetric space
  - Bruhat–Tits space
  - CAT(0) space
  - metric space of nonpositive curvature
- Riemannian metric

$$ds^2 = \operatorname{tr}(A^{-1} dA)^2$$

induced by the trace inner product $\operatorname{tr}(A^{\mathsf{T}} B)$ on tangent space $\mathbb{S}^n$

Riemannian distance

- most awesome distance on $\mathbb{S}^n_{++}$:

$$\delta_2 : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \to \mathbb{R}_+, \quad \delta_2(A, B) = \Bigl[\sum\nolimits_{j=1}^{n} \log^2(\lambda_j(A^{-1}B))\Bigr]^{1/2}$$

- invariant under
  - congruence:
$$\delta_2(XAX^{\mathsf{T}}, XBX^{\mathsf{T}}) = \delta_2(A, B)$$
  - similarity:
$$\delta_2(XAX^{-1}, XBX^{-1}) = \delta_2(A, B)$$
  - inversion:
$$\delta_2(A^{-1}, B^{-1}) = \delta_2(A, B)$$
- for comparison, all matrix norms are at best invariant under
  - unitary transformations: Frobenius, spectral, nuclear, Schatten, Ky Fan
  - permutations and scaling: operator p-norms, Hölder p-norms, $p \neq 2$

important in applications

optimization δ_2 equivalent to the metric defined by the self-concordant log barrier in semidefinite programming, i.e., $\log\det : \mathbb{S}^n_{++} \to \mathbb{R}$

statistics δ_2 equivalent to the Fisher information metric for Gaussian covariance matrix estimation problems

linear algebra δ_2 gives rise to the matrix geometric mean

other areas computer vision, medical imaging, radar signal processing, pattern recognition

- for $A \in \mathbb{S}^m_{++}$, $B \in \mathbb{S}^n_{++}$, $m \neq n$, can we define $\delta_2(A, B)$?

analogues of our Schubert varieties

- assume $m \le n$, $A \in \mathbb{S}^m_{++}$, $B \in \mathbb{S}^n_{++}$
- convex set of n-dimensional ellipsoids containing $\mathcal{E}_A$

$$\Omega_+(A) := \left\{ G = \begin{bmatrix} G_{11} & G_{12} \\ G_{12}^* & G_{22} \end{bmatrix} \in \mathbb{S}^n_{++} : G_{11} \preceq A \right\}.$$

- convex set of m-dimensional ellipsoids contained in $\mathcal{E}_B$

$$\Omega_-(B) := \{H \in \mathbb{S}^m_{++} : B_{11} \preceq H\},$$

  where B_{11} is upper left $m \times m$ principal submatrix of B
- recall partial order on $\mathbb{S}^n_{++}$

$$A \preceq B \qquad \text{if and only if} \qquad B - A \in \mathbb{S}^n_{+}$$

Riemannian distance for inequidimensional ellipsoids

Theorem (LHL–Sepulchre–Ye, 2018)

for any $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$,

$$\delta_2\big(A, \Omega_-(B)\big) = \delta_2\big(B, \Omega_+(A)\big)$$

- denote common value by $\delta_2^+(A, B)$

Theorem (LHL–Sepulchre–Ye, 2018)

if B_{11} upper left $m \times m$ principal submatrix of B, then

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

where k is such that $\lambda_j(A^{-1}B_{11}) \le 1$ for $j = k+1, \dots, m$

summary in english

- given two ellipsoids, $\mathcal{E}_A$ of dimension m and $\mathcal{E}_B$ of dimension n
- distance from $\mathcal{E}_A$ to the set of m-dimensional ellipsoids contained in $\mathcal{E}_B$ equals the distance from $\mathcal{E}_B$ to the set of n-dimensional ellipsoids containing $\mathcal{E}_A$
- common value gives distance between $\mathcal{E}_A$ and $\mathcal{E}_B$

references

- L.-H. Lim, R. Sepulchre, and K. Ye, "Geometric distance between positive definite matrices of different dimensions," *IEEE Trans. Inform. Theory*, to appear.
- L.-H. Lim, K. Wong, and K. Ye, "The Grassmannian of affine subspaces," `http://arxiv.org/abs/1807.10883`.
- K. Ye and L.-H. Lim, "Schubert varieties and distances between subspaces of different dimensions," *SIAM J. Matrix Anal. Appl.*, **37** (2016), no. 3, pp. 1176–1197.