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
- $j$th object described by feature vector $a_j = [a_{1j}, \ldots, a_{nj}]^T \in \mathbb{R}^n$
- data set described by $A = [a_1, \ldots, a_k] \in \mathbb{R}^{n \times k}$
  - massive: $n$ large
  - high-dimensional: $k$ large
- often what is important is not $A$ but subspace defined by $A$
  - $\text{span}\{a_1, \ldots, a_k\}$ or $\mu + \text{span}\{a_1 - \mu, \ldots, a_k - \mu\}$
  - principal subspaces of $A$ defined by eigenvectors of covariance matrix
classical problem

Problem

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

- 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 \( \langle a_1, \ldots, a_k \rangle := \text{span}\{a_1, \ldots, a_k\} \)
  - 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 $\theta_1, \ldots, \theta_k \in [0, \pi/2]$ between them
- define principal vectors $(a^*_j, b^*_j)$ recursively as the solutions to the optimization problem

$$\begin{align*}
\text{maximize} & \quad a^T b \\
\text{subject to} & \quad a \in \langle a_1, \ldots, a_k \rangle, \ b \in \langle b_1, \ldots, b_k \rangle, \\
& \quad a^T a_1 = \cdots = a^T a_{j-1} = 0, \ |a| = 1, \\
& \quad b^T b_1 = \cdots = b^T b_{j-1} = 0, \ |b| = 1,
\end{align*}$$

for $j = 1, \ldots, k$

- principle angles given by

$$\cos \theta_j = a^*_j \cdot b^*_j, \quad j = 1, \ldots, k$$

- clearly $\theta_1 \leq \cdots \leq \theta_k$
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 $A^T B \in \mathbb{R}^{k \times k}$ be

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

where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_k)$ 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

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

- principal vectors given by

$$AU = [p_1, \ldots, p_k], \quad BV = [q_1, \ldots, q_l]$$
basic geometry of subspaces

- $k$-dimensional linear subspace $\mathbb{A}$ in $\mathbb{R}^n$ is an element of the Grassmann manifold $\text{Gr}(k, n)$
- **Stiefel manifold**: $V(k, n)$ set of $n \times k$ orthonormal matrices $A \in \mathbb{R}^{n \times k}$
- **Grassmann manifold**: $\mathbb{A} = \text{span}(A) \in \text{Gr}(k, n)$ represents an equivalence class

$$\text{Gr}(k, n) = V(k, n)/O(k)$$

- rich geometry: smooth Riemannian manifold, algebraic variety, homogeneous space, geodesic orbit space

L.-H. Lim (Chicago)
Grassmann distance

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

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

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

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

Asimov distance
\[ d_{\text{Gr}(k,n)}^{\alpha}(A, B) = \theta_k \]

Binet–Cauchy distance
\[ d_{\text{Gr}(k,n)}^{\beta}(A, B) = \left( 1 - \prod_{i=1}^{k} \cos^2 \theta_i \right)^{1/2} \]

Chordal distance
\[ d_{\text{Gr}(k,n)}^{\kappa}(A, B) = \left( \sum_{i=1}^{k} \sin^2 \theta_i \right)^{1/2} \]

Fubini–Study distance
\[ d_{\text{Gr}(k,n)}^{\phi}(A, B) = \cos^{-1} \left( \prod_{i=1}^{k} \cos \theta_i \right) \]

Martin distance
\[ d_{\text{Gr}(k,n)}^{\mu}(A, B) = \left( \log \prod_{i=1}^{k} 1/ \cos^2 \theta_i \right)^{1/2} \]

Procrustes distance
\[ d_{\text{Gr}(k,n)}^{\rho}(A, B) = 2 \left( \sum_{i=1}^{k} \sin^2(\theta_i/2) \right)^{1/2} \]

Projection distance
\[ d_{\text{Gr}(k,n)}^{\pi}(A, B) = \sin \theta_k \]

Spectral distance
\[ d_{\text{Gr}(k,n)}^{\sigma}(A, B) = 2 \sin(\theta_k/2) \]
distances between nonequidimensional subspaces?

- 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*

\[
A^T B = U(\cos \Theta)V^T, \quad \Theta = \text{diag}(\theta_1, \ldots, \theta_k)
\]

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

example

- simultaneous embedding of \(\text{Gr}(n, 0), \text{Gr}(n, 1), \ldots, \text{Gr}(n, n)\) into sphere in \(\mathbb{R}^{(n-1)(n+2)/2}\) as orthogonal projectors
- chordal distance \(\|AA^T - BB^T\|_F = \sqrt{2}d_{\text{Gr}(k,n)}^\kappa(A, B)\)
intrinsinc distance?

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

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

- solution: inspired by Schubert calculus
what we propose

- given subspaces $A$ of dimension $k$ and $B$ of dimension $l$
- WLOG assume $k < l$
- define

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

- $\Omega_+(A)$ and $\Omega_-(B)$ are Schubert varieties in $\text{Gr}(l, n)$ and $\text{Gr}(k, n)$ respectively

- two possibilities for our distance:

$$\delta_+(A, B) = \min\{d_{\text{Gr}(l, n)}(X, B) : X \in \Omega_+(A)\}$$
$$\delta_-(A, B) = \min\{d_{\text{Gr}(k, n)}(Y, A) : Y \in \Omega_-(B)\}$$
intrinsic distance for inequidimensional subspaces

Theorem (Ye–LHL, 2016)

for any two subspaces $A$ of dimension $k$ and $B$ of dimension $l$, 

$$
\delta_+(A, B) = \delta_-(A, B)
$$

denote common value by $\delta(A, B)$

Theorem (Ye–LHL, 2016)

for any two subspaces $A$ of dimension $k$ and $B$ of dimension $l$, 

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

where 

$$
A^T B = U(\cos \Theta)V^T, \quad \Theta = \text{diag}(\theta_1, \ldots, \theta_{\min(k,l)}, 1, \ldots, 1)
$$
pictorial view

\[ \Omega_-(\mathbb{B}) \]

\[ \text{Gr}(1, 3) \]
properties

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

$$i : Gr(k, n) \rightarrow Gr(k, n + 1)$$

is isometric and $d_{Gr(k,\infty)}$ defines metric on $Gr(k, \infty)$

- distance in the sense of distance of a point to a set
Theorem (Ye–LHL, 2016)

**does not matter if** $d_{\text{Gr}(k,n)}$ **is replaced by any other distances, always have**

$$\delta^*_+(A,B) = \delta^*_-(A,B), \; * = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi$$

<table>
<thead>
<tr>
<th>Distance</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asimov distance</td>
<td>$\delta^\alpha(A,B) = \theta_{\text{min}(k,l)}$</td>
</tr>
<tr>
<td>Binet–Cauchy distance</td>
<td>$\delta^\beta(A,B) = \left(1 - \prod_{i=1}^{\text{min}(k,l)} \cos^2 \theta_i\right)^{1/2}$</td>
</tr>
<tr>
<td>Chordal distance</td>
<td>$\delta^\kappa(A,B) = \left(\sum_{i=1}^{\text{min}(k,l)} \sin^2 \theta_i\right)^{1/2}$</td>
</tr>
<tr>
<td>Fubini–Study distance</td>
<td>$\delta^\phi(A,B) = \cos^{-1} \left(\prod_{i=1}^{\text{min}(k,l)} \cos \theta_i\right)$</td>
</tr>
<tr>
<td>Martin distance</td>
<td>$\delta^\mu(A,B) = \left(\log \prod_{i=1}^{\text{min}(k,l)} 1/ \cos^2 \theta_i\right)^{1/2}$</td>
</tr>
<tr>
<td>Procrustes distance</td>
<td>$\delta^\rho(A,B) = 2 \left(\sum_{i=1}^{\text{min}(k,l)} \sin^2(\theta_i/2)\right)^{1/2}$</td>
</tr>
<tr>
<td>Projection distance</td>
<td>$\delta^\pi(A,B) = \sin \theta_{\text{min}(k,l)}$</td>
</tr>
<tr>
<td>Spectral distance</td>
<td>$\delta^\sigma(A,B) = 2 \sin(\theta_{\text{min}(k,l)}/2)$</td>
</tr>
</tbody>
</table>
\( \delta \) is a premetric but not a metric on doubly infinite Grassmannian

\[
\text{Gr}(\infty, \infty) := \bigsqcup_{k=1}^{\infty} \text{Gr}(k, \infty)
\]

which parameterizes subspaces of all dimensions

- e.g., \( \delta(A, B) = 0 \) if \( A \subset B \), triangle inequality not satisfied
- no mathematically natural way to make \( \text{Gr}(\infty, \infty) \) into a metric space: category of metric space does not admit coproduct
our proposal

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

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

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

$$d_{\text{Gr}(\infty, \infty)}(A, B) = \left(\sum_{i=1}^{l} \theta_i^2\right)^{1/2} = \left((l - k)\pi^2/4 + \sum_{i=1}^{k} \theta_i^2\right)^{1/2}$$

$$d_{\text{Gr}(\infty, \infty)}^\kappa(A, B) = \left(\sum_{i=1}^{l} \sin^2 \theta_i\right)^{1/2} = \left(l - k + \sum_{i=1}^{k} \sin^2 \theta_i\right)^{1/2}$$

$$d_{\text{Gr}(\infty, \infty)}^\rho(A, B) = \left(2 \sum_{i=1}^{l} \sin^2(\theta_i/2)\right)^{1/2} = \left(l - k + 2 \sum_{i=1}^{k} \sin^2(\theta_i/2)\right)^{1/2}$$

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

$$d_{\text{Gr}(\infty, \infty)}^*(A, B) = \sqrt{\delta^*(A, B)^2 + c^2_\ast \epsilon(A, B)^2}$$
moreover

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

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

- 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?

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

provided $n > 2l$
summary in english

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

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

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

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

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

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

- $A \in \text{Gr}(k, n)$ $k$-dimensional linear subspace, $b \in \mathbb{R}^n$ displacement of $A$ from the origin

- $A = [a_1, \ldots, a_k] \in \mathbb{R}^{n \times k}$ basis of $A$, then a $k$-dimensional affine subspace is

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

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

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

- 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

- \( \text{Graff}(k, n) \) is smooth manifold
- \( \text{Graff}(k, n) \) is universal quotient bundle of \( \text{Gr}(k, n) \)
  
  \[
  0 \rightarrow S \rightarrow \text{Gr}(k, n) \times \mathbb{R}^n \rightarrow \text{Graff}(k, n) \rightarrow 0
  \]

- \( \text{Graff}(k, n) \) is homogeneous space
  
  \[
  \text{Graff}(k, n) \cong E(n)/(O(n-k) \times E(k))
  \]

  where \( E(n) \) is group of orthogonal affine transformations
- \( \text{Graff}(k, n) \) is Zariski open subset of \( \text{Gr}(k+1, n+1) \)
  
  \[
  \text{Gr}(k+1, n+1) = X \cup X^c, \quad X \cong \text{Graff}(k, n), \quad X^c \cong \text{Gr}(k+1, n)
  \]

- first three do not give useful distance on \( \text{Graff}(k, n) \), last one does
embedding $\text{Graff}(k, n)$ into $\text{Gr}(k + 1, n + 1)$

\[ j : \text{Graff}(k, n) \to \text{Gr}(k + 1, n + 1), \quad A + b \mapsto \text{span}(A \cup \{b + e_{n+1}\}) \]
distance between affine subspaces

- define distance between two $k$-dimensional affine subspaces as

\[
d_{\text{Graff}}(k,n)(A + b, B + c) := d_{\text{Gr}}(k+1,n+1)(j(A + b), j(B + c))
\]

- 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

\[
d_{\text{Graff}}(k,n)(A + b, B + c) = \left(\sum_{i=1}^{k+1} \phi_i^2\right)^{1/2}
\]

- affine principal angles defined by

\[
\phi_i = \cos^{-1} \tau_i, \quad i = 1, \ldots, k + 1,
\]

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

\[
\begin{bmatrix}
A & b_0 / \sqrt{1 + \|b_0\|^2} \\
0 & 1 / \sqrt{1 + \|b_0\|^2}
\end{bmatrix}^T \begin{bmatrix}
B & c_0 / \sqrt{1 + \|c_0\|^2} \\
0 & 1 / \sqrt{1 + \|c_0\|^2}
\end{bmatrix} \in \mathbb{R}^{(k+1) \times (k+1)}
\]
affine subspaces of different dimensions?

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

**Theorem (LHL–Wong–Ye, 2018)**

If \( k \leq l \leq n \), \( A + b \in \text{Graff}(k, n) \), \( B + c \in \text{Graff}(l, n) \), then

\[ d_{\text{Graff}(k,n)}(A + b, \Omega_-(B + c)) = d_{\text{Graff}(l,n)}(B + c, \Omega_+(A + b)), \]

and their common value is

\[ \delta(A + b, B + c) = \left( \sum_{i=1}^{\min(k,l)+1} \phi_i^2 \right)^{1/2}, \]

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

\[
\begin{bmatrix}
A & b_0 / \sqrt{1 + \|b_0\|^2} \\
0 & 1 / \sqrt{1 + \|b_0\|^2}
\end{bmatrix}^T
\begin{bmatrix}
B & c_0 / \sqrt{1 + \|c_0\|^2} \\
0 & 1 / \sqrt{1 + \|c_0\|^2}
\end{bmatrix} \in \mathbb{R}^{(k+1) \times (l+1)}
\]
works with other distances too

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

$$
\delta^\alpha (\mathbb{A} + b, \mathbb{B} + c) = \phi_{k+1},
\delta^\pi (\mathbb{A} + b, \mathbb{B} + c) = \sin \phi_{k+1},
\delta^\sigma (\mathbb{A} + b, \mathbb{B} + c) = 2 \sin(\phi_{k+1}/2),
\delta^\kappa (\mathbb{A} + b, \mathbb{B} + c) = \left( \sum_{i=1}^{k+1} \sin^2 \phi_i \right)^{1/2},
\delta^\rho (\mathbb{A} + b, \mathbb{B} + c) = 2 \sin(\phi_{k+1}/2),
\delta^\phi (\mathbb{A} + b, \mathbb{B} + c) = \cos^{-1} \left( \prod_{i=1}^{k+1} \cos \phi_i \right),
\delta^\mu (\mathbb{A} + b, \mathbb{B} + c) = \left( 1 - \prod_{i=1}^{k+1} \cos^2 \phi_i \right)^{1/2},
\delta^\nu (\mathbb{A} + b, \mathbb{B} + c) = \left( \log \prod_{i=1}^{k+1} \frac{1}{\cos^2 \phi_i} \right)^{1/2}.
$$
ellipsoids
same thing different names

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

- $S^n$ vector space of real symmetric or complex Hermitian matrices
- $S^n_{++}$ 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

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

induced by the trace inner product $\text{tr}(A^T B)$ on tangent space $S^n$
Riemannian distance

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

$$
\delta_2 : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \rightarrow \mathbb{R}_+, \quad \delta_2(A, B) = \left[ \sum_{j=1}^{n} \log^2(\lambda_j(A^{-1}B)) \right]^{1/2}
$$

- invariant under
  - congruence:
    $$
    \delta_2(XAX^T, XBX^T) = \delta_2(A, B)
    $$
  - 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** $\delta_2$ 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** $\delta_2$ equivalent to the Fisher information metric for Gaussian covariance matrix estimation problems

**linear algebra** $\delta_2$ 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 $\delta_2(A, B)$?
analogue of our Schubert varieties

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

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

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

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

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

$$A \preceq B \quad \text{if and only if} \quad B - A \in S^n_+$$
Riemannian distance for inequidimensional ellipsoids

**Theorem (LHL–Sepulchre–Ye, 2018)**

for any $A \in \mathbb{S}_+^m$ and $B \in \mathbb{S}_+^n$,

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

- 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

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

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

L.-H. Lim (Chicago)  
cross-dimensional distances  
37 / 39
given two ellipsoids, $E_A$ of dimension $m$ and $E_B$ of dimension $n$

distance from $E_A$ to the set of $m$-dimensional ellipsoids contained in $E_B$ equals the distance from $E_B$ to the set of $n$-dimensional ellipsoids containing $E_A$

common value gives distance between $E_A$ and $E_B$
