Geometric means of more than two matrices

D. A. Bini[1] B. Iannazzo[2] B. Meini[1] Federico Poloni[3]

[1]University of Pisa
[2]University of Perugia
[3]Scuola Normale Superiore, Pisa

SIAM Linear Algebra conference 2009

A physical problem

Elasticity experiments [Hearmon, 1952; Moakher, 2006]:
Several experimental measures of either the stiffness tensor or its inverse (compliance tensor).

Problem

How to average them?

Requirement: Averaging inverses (compliance) should yield the same result as averaging the tensors and then inverting

$$M(A, B, C, \dots)^{-1} = M(A^{-1}, B^{-1}, C^{-1}, \dots)$$

In the scalar case, this holds true for the geometric mean

A mathematical problem

At the same time [Ando–Li–Mathias, 2003; Bhatia, 2005; + others]

Definition

$$GM(a_1, a_2, \dots, a_k) = \sqrt[k]{a_1 a_2 \dots a_k} \quad \text{for scalar } a_i > 0$$

Problem

Find a sensible generalization of the geometric mean to SPD matrices

What do we expect from a geometric mean?

[Ando–Li–Mathias, 2003]: ten properties that a *bona-fide* geometric mean should have:

- compatibility with scalars: $GM(A,B,C) = (ABC)^{1/3}$ for commuting A, B, C
- simmetry: $GM(A,B,C) = GM(B,A,C) = \dots$
- monotonicity: $A < A' \Rightarrow GM(A,B,C) < GM(A',B,C)$
- Congruence invariance: $GM(S^*AS, S^*BS, S^*CS) = S^*GM(A,B,C)S$
- Inversion invariance: $GM(A^{-1}, B^{-1}, C^{-1}) = GM(A,B,C)^{-1}$

$\dots +$ others (concavity, continuity…)

Remark

These do not define GM uniquely!

Mean of two matrices

There is already a sound definition of the geometric mean of two matrices

Definition

$$GM(A, B) = A(A^{-1}B)^{1/2}$$

(not what you would expect at first!)

Compatibility with scalars + congruence invariance determine it uniquely

The geometrical meaning of the geometric mean

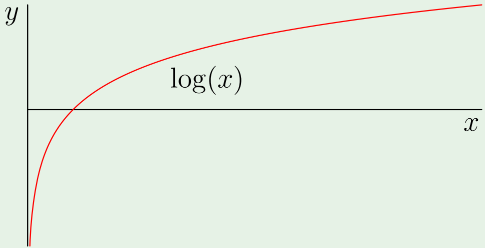
Natural Riemannian metric on SPD matrices

$$ds = \left\| A^{-1/2} dA A^{-1/2} \right\|_2$$

gets more and more "curved" when A approaches singularity

Example

In dimension 1,
logarithmic scale



The geometrical meaning of the geometric mean

Natural Riemannian metric on SPD matrices

$$ds = \left\| A^{-1/2} dA A^{-1/2} \right\|_2$$

Explicit expression for the geodesic joining A and B:

$$\gamma(t) = A(A^{-1}B)^t \qquad t \in [0, 1]$$

Geometric mean = midpoint of the geodesic!

Generalizing to $k \geq 3$ matrices

Problem

How do we define the mean of $k \geq 3$ matrices?

The ALM properties do not define it uniquely!

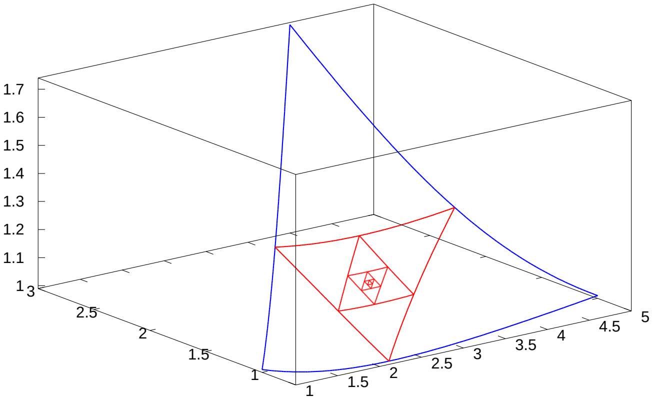
Idea [Ando–Li–Mathias, 2003]: symmetrization procedure

$$\begin{aligned} A_1 &= GM(B, C) & A_2 &= GM(B_1, C_1) \\ B_1 &= GM(C, A) & B_2 &= GM(C_1, A_1) & \dots \\ C_1 &= GM(A, B) & C_2 &= GM(A_1, B_1) \end{aligned}$$

A_i, B_i, C_i converge to the same matrix $GM_{ALM}(A, B, C)$

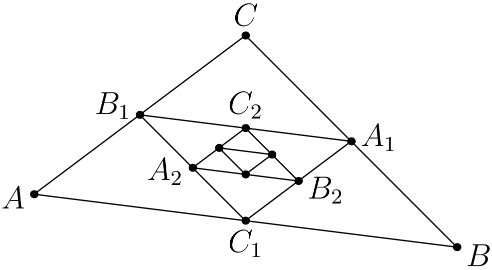
ALM construction

$$A_1 = GM(B, C) \qquad A_2 = GM(B_1, C_1)$$
$$B_1 = GM(C, A) \qquad B_2 = GM(C_1, A_1) \qquad \ldots$$
$$C_1 = GM(A, B) \qquad C_2 = GM(A_1, B_1)$$



ALM construction

$$A_1 = GM(B, C) \qquad A_2 = GM(B_1, C_1)$$
$$B_1 = GM(C, A) \qquad B_2 = GM(C_1, A_1) \qquad \dots$$
$$C_1 = GM(A, B) \qquad C_2 = GM(A_1, B_1)$$



On the plane (Euclidean metric), converges to the centroid of ABC.

ALM mean: properties

- Constructive definition
- Satisfies the ten ALM properties
- May be generalized to $k = 4$ or more: $A_1 = GM(B, C, D, \dots)$
- Slow to compute: linear convergence, cost grows as $k!$ (factorial)

Problem

Is there a faster algorithm to compute it?

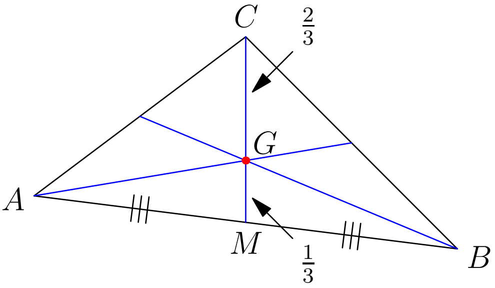
Problem

Is there a faster algorithm to compute another mean that satisfies the ten ALM properties?

Considering the medians

Theorem

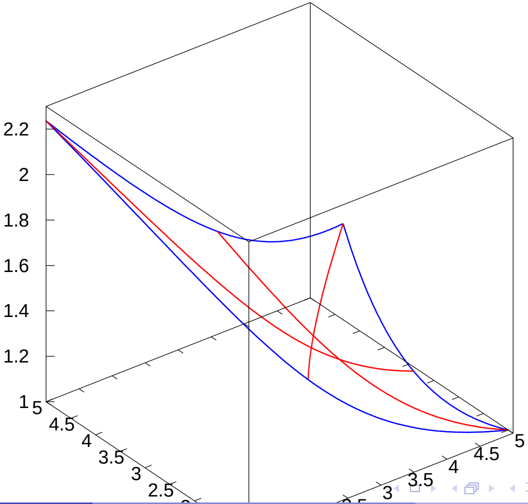
On the Euclidean plane, the three medians of a triangle meet in the centroid at 2/3 of their length.



No iteration needed! Can we do the same for matrices?

Considering the medians

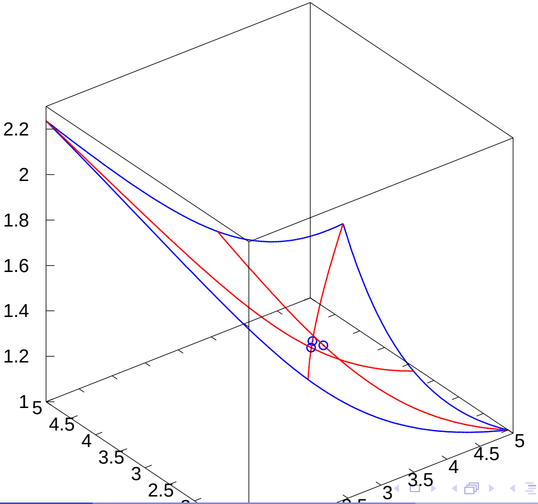
In the geometry of SPD matrices, the medians don't meet!



Considering the medians

In the geometry of SPD matrices, the medians don't meet!

...but the points at 2/3 of the corresponding geodesics are very close



Algorithm

$A_1 =$ the point at $2/3$ of the geodesic joining A and $GM(B, C)$
$B_1 =$ the point at $2/3$ of the geodesic joining A and $GM(C, A)$
$C_1 =$ the point at $2/3$ of the geodesic joining A and $GM(A, B)$

(They are not the same point!)

A_2, B_2, C_2 defined in the same way starting from $A_1 B_1 C_1$, and so on

Theorem

A_i, B_i, C_i converge to the same matrix $GM_{new}(A, B, C)$

Properties of the new mean

- Constructive definition
- Generally different from GM_{ALM}
- Satisfies the ten ALM properties
- May be generalized to $k \geq 4$:
  $A_1 = 1/k$ of the geodesic joining A and $GM(B, C, D, \dots)$
- Convergence order 3: faster than GM_{ALM}
- ...though it still grows as $k!$ (factorial)

Some numbers

5
1.92542947898189
2.90969918536362
2.35774114351751
2.61639158463414
2.48316587472793
2.54876054375880
2.51571460655576
2.53217471946628
2.52392903948587
2.52804796243998
2.52598752310721
2.52701749813482
2.52650244948321
2.52675995852183
2.52663120018107
2.52669557839604
2.52666338904971
2.52667948366316
2.52667143634151

5
2.59890269690271
2.53027293208879
2.53025171828977
2.53025171828977

Example

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 4 & 3 \\ 3 & 3 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

Shown $(A_i)_{11}$ for both iterations

Elasticity measures, data from [Hearmon, '52] (up to 6 6×6 matrices): up to $100\times$ faster

New directions

Idea: can we obtain new means from the composition of means with less arguments? E.g.

$$A, B, C, D \mapsto GM(GM(A, B), GM(C, D))$$
$$A, B, C \mapsto GM(A, GM(B, C))$$

We need a new framework to deal with non-symmetric mean-like functions.

Definition

A quasi-mean is a map that is not symmetric in its arguments, but satisfies the other Ando-Li-Mathias properties (with some technical changes).

Invariance groups

Definition

Q quasi-mean, σ permutation:

$$(Q\sigma)(A_1, A_2, A_3) := Q(A_{\sigma(1)}, A_{\sigma(2)}, A_{\sigma(3)})$$

Definition

Invariance group of a quasi-mean Q:

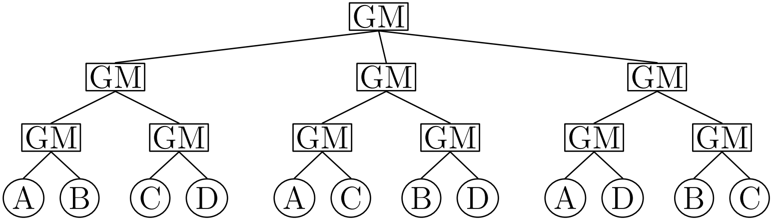
$$I(Q) := \{\text{all } \sigma \text{ s.t. } Q\sigma(\dots) = Q(\dots)\}$$

Q quasi-mean + $\{I(Q) = \text{all permutations}\} \Rightarrow Q$ is a geometric mean (all ALM properties)

A positive result

This is a geometric mean of four matrices:

$$A, B, C, D \mapsto GM(GM(GM(A,B), GM(C,D)), \\ GM(GM(A,C), GM(B,D)), \\ GM(GM(A,D), GM(B,C)))$$



A mean of 4 matrices is reduced to one of 3: computational advantage
(4× to 10× speedup on the elasticity data)

Negative results

Strong assumption

The symmetries of a quasi-mean obtained by composition (like $Q(R(\dots), S(\dots)))$ are only those deriving from the symmetries of the underlying quasi-means (like Q, R, S)

that is, no "unexpected properties" appear

Negative results

Strong assumption

The symmetries of a quasi-mean obtained by composition (like $Q(R(\dots), S(\dots)))$ are only those deriving from the symmetries of the underlying quasi-means (like Q, R, S)

Theorem

A geometric mean of $k \geq 5$ matrices cannot be built composing (quasi-)means of less matrices

Idea of the proof: the group of permutations of $k \geq 5$ elements is (nearly) a simple group

Negative results

Strong assumption

The symmetries of a quasi-mean obtained by composition (like $Q(R(\dots), S(\dots)))$ are only those deriving from the symmetries of the underlying quasi-means (like Q, R, S)

Theorem

A geometric mean of $k \geq 5$ matrices cannot be built composing (quasi-)means of less matrices

Theorem

If a geometric mean of $k \geq 5$ matrices is obtained via composition of simpler (quasi-)means + a limit process (like GM_{ALM}, GM_{new}), then the ingredients must be means of k-1 matrices

We cannot do any better than the current algorithms

Thanks for your attention!