Geometric means of matrices: analysis and algorithms

D.A. Bini
Università di Pisa

VDM60 - Nonlinear Evolution Equations and Linear Algebra
Cagliari – September 2013

Cor:
Happy 60th

Outline

The Problem and its motivations

In certain applications we are given a set of k positive definite matrices $A_1, \ldots, A_k \in \mathcal{P}_n$ which represent measures of some physical object

Problem:

To compute an average $G = G(A_1, \ldots, A_k) \in \mathcal{P}_n$ such that

$$G(A_1, \ldots, A_k)^{-1} = G(A_1^{-1}, \ldots, A_k^{-1})$$

Elasticity tensor analysis, image processing, radar detection, subdivision schemes, [Hearmon 1952, Moakher 2006, Barbaresco 2009, Barachant et al. 2010, Itai, Sharon 2012] ;

The Problem and its motivations

In certain applications we are given a set of k positive definite matrices $A_1, \ldots, A_k \in \mathcal{P}_n$ which represent measures of some physical object

Problem:

To compute an average $G = G(A_1, \ldots, A_k) \in \mathcal{P}_n$ such that

$$G(A_1, \ldots, A_k)^{-1} = G(A_1^{-1}, \ldots, A_k^{-1})$$

Elasticity tensor analysis, image processing, radar detection, subdivision schemes, [Hearmon 1952, Moakher 2006, Barbaresco 2009, Barachant et al. 2010, Itai, Sharon 2012] ;

Scalar case: the geometric mean $(\prod_{i=1}^{k} A_i)^{1/k}$ is the ideal choice

Matrix case: things are more complicated

The Problem and its motivations

In certain applications we are given a set of k positive definite matrices $A_1, \ldots, A_k \in \mathcal{P}_n$ which represent measures of some physical object

Problem:

To compute an average $G = G(A_1, \ldots, A_k) \in \mathcal{P}_n$ such that

$$G(A_1, \ldots, A_k)^{-1} = G(A_1^{-1}, \ldots, A_k^{-1})$$

Elasticity tensor analysis, image processing, radar detection, subdivision schemes, [Hearmon 1952, Moakher 2006, Barbaresco 2009, Barachant et al. 2010, Itai, Sharon 2012] ;

Scalar case: the geometric mean $(\prod_{i=1}^{k} A_i)^{1/k}$ is the ideal choice

Matrix case: things are more complicated

An additional request: If $A_1, \ldots, A_k \in \mathcal{A} \subset \mathcal{P}_n$ then it is required that $G \in \mathcal{A}$. In the design of certain radar systems [Farina, Fortunati 2011], [Barbaresco 2009] $\mathcal{A}$ is the set of Toeplitz matrices.

Means of two matrices: an “easy” case

Many authors analyzed the problem of extending the concept of geometric mean from scalars to matrices [Anderson, Trapp, Ando, Li, Mathias, Bhatia, Holbrook, Kosaki, Lawson, Lim, Moakher, Petz, Temesi,...]

Some attempts to extend the geometric mean from scalars to matrices

- $G(A,B) := (AB)^{1/2}$: drawbacks $G(A,B) \notin \mathcal{P}_n$, $G(A,B) \neq G(B,A)$
- $G(A,B) := \exp(\frac{1}{2}(\log A + \log B))$: several drawbacks Def. of matrix function

Means of two matrices: an "easy" case

Many authors analyzed the problem of extending the concept of geometric mean from scalars to matrices [Anderson, Trapp, Ando, Li, Mathias, Bhatia, Holbrook, Kosaki, Lawson, Lim, Moakher, Petz, Temesi,...]

Some attempts to extend the geometric mean from scalars to matrices

- $G(A,B) := (AB)^{1/2}$: drawbacks $G(A,B) \notin \mathcal{P}_n$, $G(A,B) \neq G(B,A)$
- $G(A,B) := \exp(\frac{1}{2}(\log A + \log B))$: several drawbacks Def. of matrix function

A good definition

$$\boxed{G(A,B) = A(A^{-1}B)^{1/2}}$$

Means of two matrices: an "easy" case

Many authors analyzed the problem of extending the concept of geometric mean from scalars to matrices [Anderson, Trapp, Ando, Li, Mathias, Bhatia, Holbrook, Kosaki, Lawson, Lim, Moakher, Petz, Temesi,...]

Some attempts to extend the geometric mean from scalars to matrices

- $G(A,B) := (AB)^{1/2}$: drawbacks $G(A,B) \notin \mathcal{P}_n$, $G(A,B) \neq G(B,A)$
- $G(A,B) := \exp(\frac{1}{2}(\log A + \log B))$: several drawbacks Def. of matrix function

A good definition

$$\boxed{G(A,B) = A(A^{-1}B)^{1/2} = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}}$$

Means of two matrices: an "easy" case

Many authors analyzed the problem of extending the concept of geometric mean from scalars to matrices [Anderson, Trapp, Ando, Li, Mathias, Bhatia, Holbrook, Kosaki, Lawson, Lim, Moakher, Petz, Temesi,...]

Some attempts to extend the geometric mean from scalars to matrices

- $G(A,B) := (AB)^{1/2}$: drawbacks $G(A,B) \notin \mathcal{P}_n$, $G(A,B) \neq G(B,A)$
- $G(A,B) := \exp(\frac{1}{2}(\log A + \log B))$: several drawbacks ▸ Def. of matrix function

A good definition

$$\boxed{G(A,B) = A(A^{-1}B)^{1/2} = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}}$$

This mean is uniquely defined by the Ando-Li-Mathias (ALM) axioms: ten properties that a "good" mean should satisfy

P1 Consistency with scalars. If A, B commute then $G(A, B) = (AB)^{1/2}$

P2 Joint homogeneity. $G(\alpha A, \beta B) = (\alpha\beta)^{1/2} G(A, B)$, $\alpha, \beta > 0$

P3 Permutation invariance. $G(A, B) = G(B, A)$

- P1 Consistency with scalars. If A, B commute then $G(A,B) = (AB)^{1/2}$
- P2 Joint homogeneity. $G(\alpha A, \beta B) = (\alpha\beta)^{1/2} G(A,B)$, $\alpha, \beta > 0$
- P3 Permutation invariance. $G(A,B) = G(B,A)$
- P4 Monotonicity. If $A \succeq A'$, $B \succeq B'$, then $G(A,B) \succeq G(A',B')$
- P5 Continuity from above. If A_j, B_j are monotonic decreasing sequences converging to A, B, respectively, then $\lim_j G(A_j,B_j) = G(A,B)$
- P6 Joint concavity. If $A = \lambda A_1 + (1-\lambda)A_2$, $B = \lambda B_1 + (1-\lambda)B_2$, then

$$G(A,B) \succeq \lambda G(A_1,B_1) + (1-\lambda)G(A_2,B_2)$$

- P1 Consistency with scalars. If A, B commute then $G(A,B)=(AB)^{1/2}$
- P2 Joint homogeneity. $G(\alpha A, \beta B)=(\alpha\beta)^{1/2}G(A,B)$, $\alpha,\beta>0$
- P3 Permutation invariance. $G(A,B)=G(B,A)$
- P4 Monotonicity. If $A \succeq A'$, $B \succeq B'$, then $G(A,B) \succeq G(A',B')$
- P5 Continuity from above. If A_j, B_j are monotonic decreasing sequences converging to A, B, respectively, then $\lim_j G(A_j,B_j)=G(A,B)$
- P6 Joint concavity. If $A=\lambda A_1+(1-\lambda)A_2$, $B=\lambda B_1+(1-\lambda)B_2$, then

$$G(A,B) \succeq \lambda G(A_1,B_1)+(1-\lambda)G(A_2,B_2)$$

- P7 Congruence invariance. $G(S^TAS, S^TBS)=S^TG(A,B)S$
- P8 Self-duality $G(A,B)^{-1}=G(A^{-1},B^{-1})$

P1 Consistency with scalars. If A, B commute then $G(A,B) = (AB)^{1/2}$

P2 Joint homogeneity. $G(\alpha A, \beta B) = (\alpha\beta)^{1/2} G(A,B)$, $\alpha, \beta > 0$

P3 Permutation invariance. $G(A,B) = G(B,A)$

P4 Monotonicity. If $A \succeq A'$, $B \succeq B'$, then $G(A,B) \succeq G(A',B')$

P5 Continuity from above. If A_j, B_j are monotonic decreasing sequences converging to A, B, respectively, then $\lim_j G(A_j, B_j) = G(A,B)$

P6 Joint concavity. If $A = \lambda A_1 + (1-\lambda)A_2$, $B = \lambda B_1 + (1-\lambda)B_2$, then

$$G(A,B) \succeq \lambda G(A_1,B_1) + (1-\lambda)G(A_2,B_2)$$

P7 Congruence invariance. $G(S^T A S, S^T B S) = S^T G(A,B) S$

P8 Self-duality $G(A,B)^{-1} = G(A^{-1}, B^{-1})$

P9 Determinant identity $\det G(A,B) = (\det A \det B)^{1/2}$

P10 Arithmetic–geometric–harmonic mean inequality:

$$\left(\frac{A^{-1} + B^{-1}}{2}\right)^{-1} \preceq G(A,B) \preceq \frac{A+B}{2}.$$

Motivation in terms of Riemannian geometry

Several authors [Bhatia, Holbrook, Lim, Moakher, Lawson] studied the geometry of positive definite matrices endowed with the Riemannian metric with the distance defined by

$$d(A, B) = \| \log(A^{-1/2} B A^{-1/2}) \|_F$$

For scalars, $d(a, b) = | \log(a) - \log(b) |$

Motivation in terms of Riemannian geometry

Several authors [Bhatia, Holbrook, Lim, Moakher, Lawson] studied the geometry of positive definite matrices endowed with the Riemannian metric with the distance defined by

$$d(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_F$$

For scalars, $d(a,b) = |\log(a) - \log(b)|$

It holds that

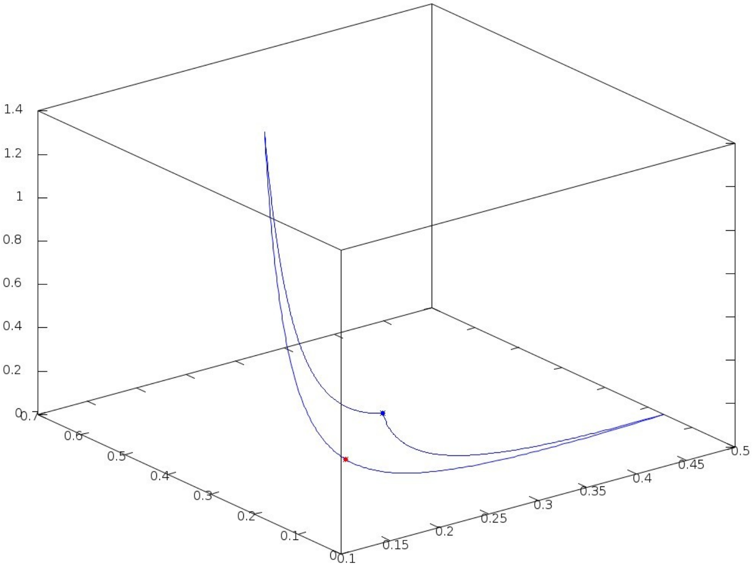
$$d(A,B) = d(A^{-1},B^{-1})$$

moreover, the geodesic joining A and B has equation

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

thus $G(A,B) = A(A^{-1}B)^{\frac{1}{2}}$ is the midpoint of the geodesic joining A and B

Explog mean and geometric mean



Explog mean and geometric mean

The explog mean does not satisfy the following ALM properties

- P4 Monotonicity
- P7 Congruence invariance

The ALM mean: The case of $k \geq 3$ matrices

Remark

The ALM-properties uniquely define the geometric mean of two matrices A and B

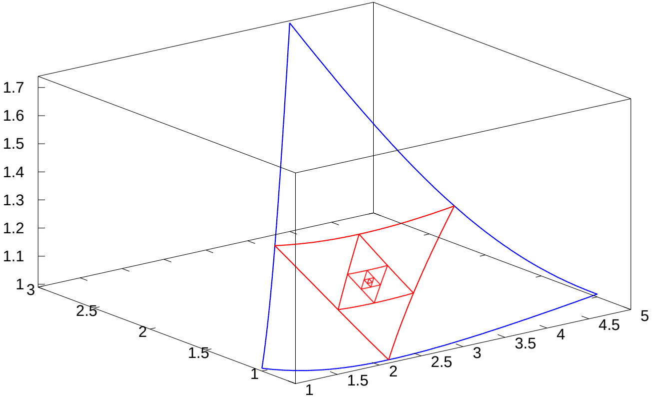
For $k > 2$ matrices there exist infinitely many matrix means satisfying the ALM-properties

One of these means is the Ando–Li–Mathias (ALM) mean
The ALM mean [Ando–Li–Mathias, 2003]:

$$\begin{array}{lll} A_1 = G(B,C) & A_2 = G(B_1,C_1) & A_3 = G(B_2,C_2) \\ B_1 = G(C,A) & B_2 = G(C_1,A_1) & B_3 = G(C_2,A_2) \quad \ldots \\ C_1 = G(A,B) & C_2 = G(A_1,B_1) & C_3 = G(A_2,B_2) \end{array}$$

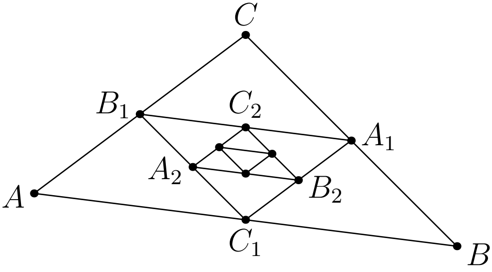
The three sequences have a common limit defined as the ALM mean $G_{ALM}(A, B, C)$

Computing the ALM mean



Remark

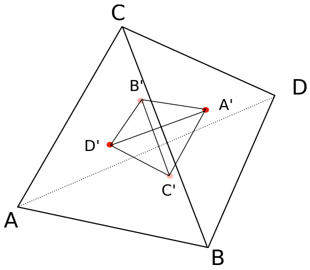
The same construction in the Euclidean geometry converges to the centroid of the triangle ABC.



Properties of the ALM mean

Recursively generalizable to $k \geq 4$ matrices $A_1, \dots, A_k$

$$A_i^{(\nu+1)} = G(A_1^{(\nu)}, \dots, A_{i-1}^{(\nu)}, A_{i+1}^{(\nu)}, \dots, A_k^{(\nu)}), \quad i = 1, \dots, k$$



- it satisfies the 10 ALM properties
- **problem**: slow convergence (linear with rate 1/2)
- **problem**: complexity $O(k! p^k n^3)$, p: number of iterations

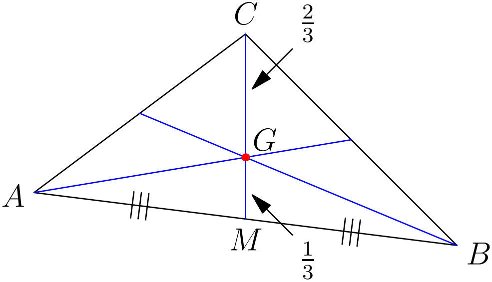
We may provide a different definition which leads to a substantial algorithmic improvement [B., Meini, Poloni, 2010], [Nakamura 2009]

In fact we overcome the first drawback about the slow convergence

It is based on the following

Remark

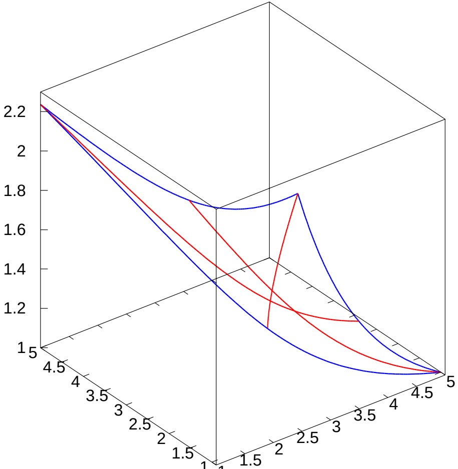
The three medians of a triangle meet at a single point, the centroid, at 2/3 of their length.



What happens with matrices?

Problem

In the Riemannian geometry the medians (geodesics) generally **do not intersect**



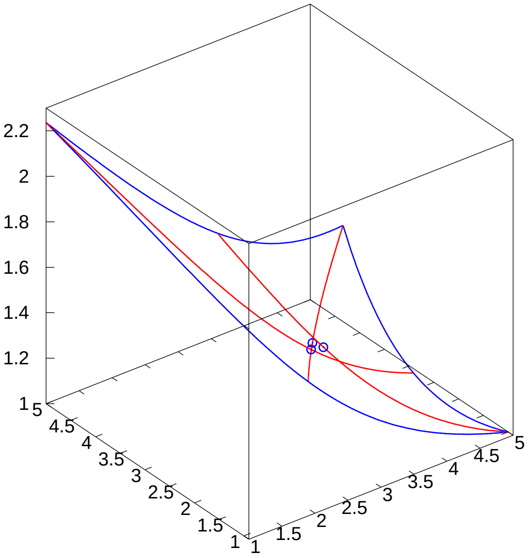
A new definition

Define A_1, B_1, C_1 the points in the medians at distance $2/3$ from the vertices A, B, C, respectively.
Generally, A_1, B_1 and C_1 are pairwise different

Similarly, define A_2, B_2, C_2 from $A_1B_1C_1$ and so forth

$$A_{\nu+1} = A_\nu \#_{\frac{2}{3}} (B_\nu \#_{\frac{1}{2}} C_\nu)$$
$$B_{\nu+1} = B_\nu \#_{\frac{2}{3}} (C_\nu \#_{\frac{1}{2}} A_\nu)$$
$$C_{\nu+1} = C_\nu \#_{\frac{2}{3}} (A_\nu \#_{\frac{1}{2}} B_\nu)$$

where $A \#_t B := \gamma(t) = A^{-1}(AB)^t$ is the matrix in the geodesic joining A and B at distance t from A



2.2
2
1.8
1.6
1.4
1.2
1
5
4.5
4
3.5
3
2.5
2
1.5
1
1
1.5
2
2.5
3
3.5
4
4.5
5

Properties and remarks

Theorem (A new mean)

For any positive definite matrices A, B, C, the sequences A_ν, B_ν, C_ν obtained this way converge to the same limit. We define this limit $G(A, B, C)$ as the geometric mean of A, B, C

Theorem (Convergence with order 3)

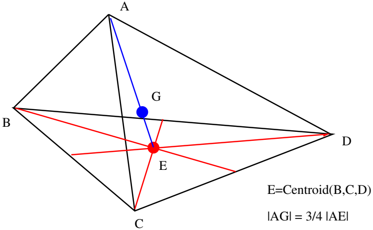
The convergence speed is of **order three***, that is the error is $O(\lambda^{3^\nu})$, for $0 < \lambda < 1$, while for the ALM mean the error is $O(2^{-\nu})$.*

Theorem (ALM properties)

In general $G(A, B, C)$ is **different** *from $G_{ALM}(A, B, C)$, but fulfills the 10 ALM properties*

Properties and remarks

The mean can be generalized to $k \geq 4$ matrices; for $k = 4$ the barycenter of a tetrahedron is in the segment joining each vertex with the centroid of the triangle (facet) formed by the remaining points, at distance $3/4$; the nice properties are preserved.



In general on has:

$$A_i^{(\nu+1)} = A_i^{(\nu)} \#_{\frac{k-1}{k}} G(A_1^{(\nu)}, \ldots, A_{i-1}^{(\nu)}, A_{i+1}^{(\nu)}, \ldots, A_k^{(\nu)}),$$

complexity $O(k! p^k n^3)$, p: number of iterations

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

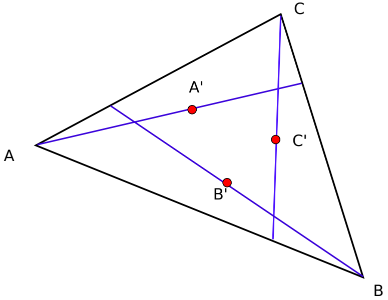
$$A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 4 & 3 \\ 3 & 3 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

The entry a_{11} is displayed at step i
Left: *ALM mean*
Right: *New mean*

Physical applications: speedup by a factor of 200 (mean of $k = 6$ matrices)

A general class of means

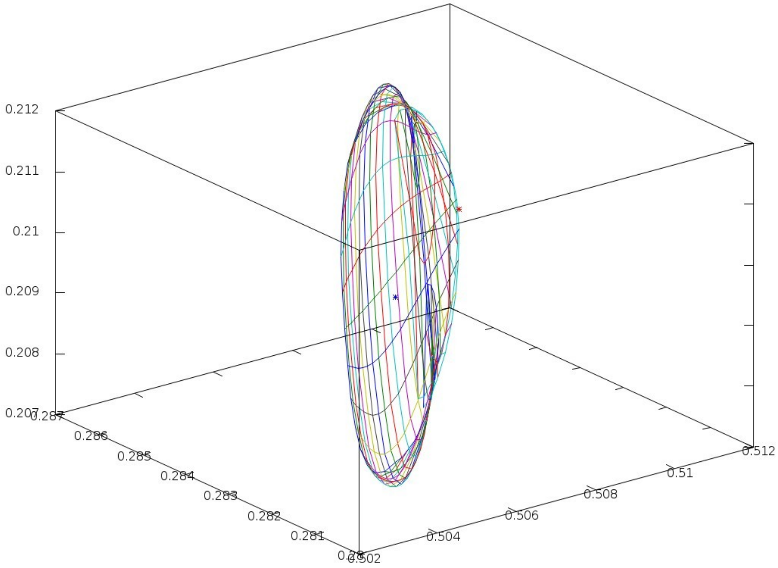
Given $0 < s, t \leq 1$ define the sequences



$$A' = A\#_s(B\#_t C)$$
$$B' = B\#_s(C\#_t A)$$
$$C' = C\#_s(A\#_t B)$$

- observe that for $s = 1$, $t = 1/2$ this gives the ALM mean. For $s = 2/3$, $t = 1/2$ this gives the mean based on medians
- the three sequences converge to the same limit $G(s,t)$ for $s,t \in [0,1]$ unless $s = 0$, or $s = 1$ and $t = 0,1$
- this limit $G(s,t)$ satisfies the 10 ALM properties for any s,t
- the set formed by $G(s,t)$ has a "small" diameter

Some experiments



Important question

Is there a way to overcome the exponential complexity?

Two solutions:

- the cheap mean
- the Karcher mean

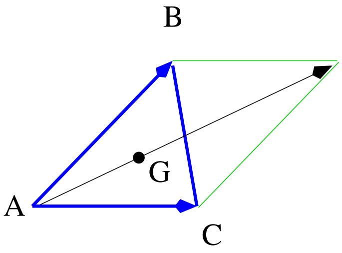
The cheap mean (overcoming the exponential complexity)

Remark

In the Euclidean geometry, given the triangle of vertices A, B, C, the centroid can be viewed as

$$G = A + \frac{1}{3}((B - A) + (C - A) + (A - A))$$

that is, **the arithmetic mean of the tangent vectors** *of the geodesics joining A with B, C and A, respectively*



That is, G lies in the geodesic passing through A and tangent to the arithmetic mean of the tangent vectors in A to the geodesics from A to B, from A to C and from A to A

We can repeat the construction in the Riemannian geometry.

In the Riemannian geometry the nonzero tangent vectors are **easily computable**. In fact differentiating the equation of the two geodesics $\gamma_{AB}(t) = A(A^{-1}B)^t$, $\gamma_{AC}(t) = A(A^{-1}C)^t$ at $t = 0$ we get the tangent vectors

$$V_B = A\log(A^{-1}B), \quad V_C = A\log(A^{-1}C)$$

The geodesic passing through A tangent to $V = \frac{1}{3}(V_B + V_C)$ is

$$\gamma(t) = A\exp(A^{-1}V)^t$$

We can repeat the construction in the Riemannian geometry.

In the Riemannian geometry the nonzero tangent vectors are **easily computable**. In fact differentiating the equation of the two geodesics $\gamma_{AB}(t) = A(A^{-1}B)^t$, $\gamma_{AC}(t) = A(A^{-1}C)^t$ at $t = 0$ we get the tangent vectors

$$V_B = A\log(A^{-1}B), \quad V_C = A\log(A^{-1}C)$$

The geodesic passing through A tangent to $V = \frac{1}{3}(V_B + V_C)$ is

$$\gamma(t) = A\exp(A^{-1}V)^t = A\exp(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)))^t$$

We can repeat the construction in the Riemannian geometry.

In the Riemannian geometry the nonzero tangent vectors are **easily computable**. In fact differentiating the equation of the two geodesics $\gamma_{AB}(t) = A(A^{-1}B)^t$, $\gamma_{AC}(t) = A(A^{-1}C)^t$ at $t = 0$ we get the tangent vectors

$$V_B = A\log(A^{-1}B), \quad V_C = A\log(A^{-1}C)$$

The geodesic passing through A tangent to $V = \frac{1}{3}(V_B + V_C)$ is

$$\gamma(t) = A\exp(A^{-1}V)^t = A\exp(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)))^t$$

For $t = 1$ we get the value

$$A' = A\exp(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)))$$

From the equation of the geodesic one deduces the iteration

$$A_{\nu+1} = A_\nu \exp[\tfrac{1}{3}(\log((A_\nu)^{-1}B_\nu) + \log((A_\nu)^{-1}C_\nu))]$$
$$B_{\nu+1} = B_\nu \exp[\tfrac{1}{3}(\log((B_\nu)^{-1}C_\nu) + \log((B_\nu)^{-1}A_\nu))]$$
$$C_{\nu+1} = C_\nu \exp[\tfrac{1}{3}(\log((C_\nu)^{-1}A_\nu) + \log((C_\nu)^{-1}B_\nu))]$$

In general, for k matrices $A_1, A_2, \ldots, A_k$ we may define the **non recursive iteration** [B., Iannazzo 2011]

$$A_i^{(\nu+1)} = A_i^{(\nu)} \exp[\frac{1}{k}\sum_{j=1}^{k} \log((A_i^{(\nu)})^{-1}A_j^{(\nu)})], \quad i = 1, 2, \ldots, k$$

Polynomial cost: $O(pk^2n^3)$, where p is the number of iterations

What we can prove

Theorem (Local convergence)

If the three sequences converge then they converge to the **same** *limit G and* **convergence is cubic**

Theorem

The matrix G satisfies the ALM properties P1,P2,P3, P7, P8, P9

Remark

We have a counterexample where **monotonicity** *P4 is not satisfied if the matrices are* **very far from each other**

Properties P5, P7 and P10 are usually proved relying on monotonicity. It is not clear if they are satisfied

For simplicity, we refer to G as the **cheap mean**

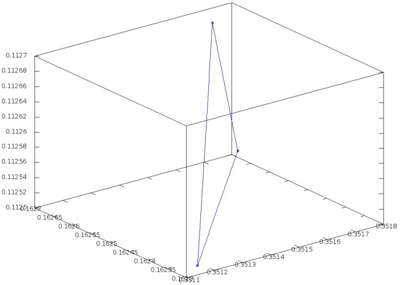
Numerical experiments

	cnd= 1.e2			cnd= 1.e4			cnd= 1.e8		
k	Cheap	NBMP	Dist.	Cheap	NBMP	Dist.	Cheap	NBMP	Dist.
3	1.e-2	1.e-2	5.e-3	1.e-2	1.e-2	3.e-2	1.e-2	1.e-2	3.e-2
4	2.e-2	2.e-1	6.e-3	2.e-2	2.e-1	2.e-2	2.e-2	2.e-2	8.e-2
5	2.e-2	1.e0	7.e-3	3.e-2	2.e0	4.e-2	3.e-1	2.e0	5.e-2
6	3.e-2	1.e+1	5.e-2	4.e-2	3.e+1	2.e-2	4.e-2	3.e+1	5.e-2
7	3.e-2	2.e+2	8.e-3	5.e-3	4.e+2	2.e-2	5.e-2	4.e+2	1.e-2
8	4.e-2	2.e+3	1.e-2	6.e-2	5.e+3	2.e-2	7.e-2	5.e+3	3.e-2
9	4.e-2	*	–	7.e-2	*	–	7.e-2	*	–
10	5.e-2	*	–	9.e-2	*	–	1.e-1	*	–

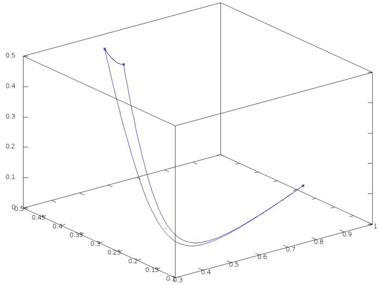
Table: CPU times in seconds, rounded to one digit, required to compute the NBMP mean G_1 and the Cheap mean G_2, together with the distances $||G_1 - G_2||_2/||G_1||_2$. A "$*$" denotes a CPU time larger than 10^4 seconds.

Numerical experiments: plotting the means

Three means



Original matrices



Numerical experiments: distances of the means

Riemannian distances of the three matrices A, B, C

	A	B	C
A		0.084	0.57
B			0.65

Distances between the means

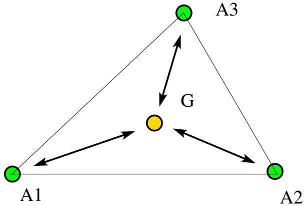
	ALM	$NBMP$	$Cheap$
ALM		3.1e-4	5.2e-4
$NBMP$			8.1e-4

The Karcher mean

Definition: The Karcher mean $G = G(A_1, \ldots, A_k)$ is the matrix G where the following function takes its minimum.

$$f(X) = \sum_{i=1}^{k} d(X, A_i)^2$$

$$d(A, B) = \| \log(A^{-1/2} B A^{-1/2}) \|_F$$

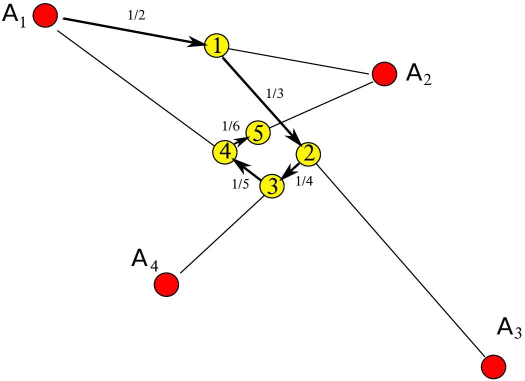


Property: The Karcher mean is unique and satisfies the 10 ALM axioms.

It is also called least squares mean, or Riemannian mean [Moakher], [Bhatia], [Holbrook], [Jeuris, Vandebril, Vandereycken].

The Karcher mean: a simple algorithm

Inductive mean (Spiral descent) [Holbrook 2012]



$$S_\nu = S_{\nu-1} \#_{\frac{1}{\nu}} A_{1+(\nu \mod k)},$$
$$S_1 = A_1$$

Theorem (good and bad news)

$\lim_\nu S_\nu = G$, *convergence is sublinear*

The Karcher mean: a more efficient algorithm

The gradient of $f(X)$ is

$$\nabla_X f = 2X^{-1} \sum_{i=1}^{k} \log(XA_i^{-1}) = 2 \sum_{i=1}^{k} \log(A_i^{-1}X)X^{-1}$$

This way, the Karcher mean can be viewed as the unique positive solution of the matrix equation $\nabla_X f = 0$, that is

$$\sum_{i=1}^{k} \log(A_i^{-1}X) = 0$$

There exist algorithms for computing G,
[Matrix means toolbox: `bezout.dm.unipi.it/software/mmtoolbox`]
[B., Iannazzo, LAA 2012]

The Karcher mean: a more efficient algorithm

Fixed-point iteration

$$X_{\nu+1} = X_\nu - \theta X_\nu \sum_{i=1}^{k} \log(A_i^{-1} X_i), \qquad \theta > 0$$

Remarks

- In the case of scalars, choosing $\theta = 1/k$ provides a quadratically convergent iteration
- If A_i commute with each other, i.e., $A_i A_j = A_j A_i$, then the convergence is quadratic if $\theta = 1/k$ as well

Convergence analysis

Theorem

Let G be the positive definite solution of the matrix equation. Define the relative error $E_\nu = G^{-1/2}(X_\nu - G)G^{-1/2}$, $e_\nu = vec(E_\nu)$. Then

$$e_{\nu+1} = (I - \theta H)e_\nu + O(\|e_\nu\|^2) \quad H = \sum_{i=1}^k H_i$$

$$H_i = \beta(W_i), \quad \beta(x) = x/(e^x - 1)$$

$$W_i = \log(M_i) \otimes I - I \otimes \log(M_i)$$

$$M_i = G^{1/2} A_i^{-1} G^{1/2}$$

Convergence analysis

Theorem

Let G be the positive definite solution of the matrix equation. Define the relative error $E_\nu = G^{-1/2}(X_\nu - G)G^{-1/2}$, $e_\nu = vec(E_\nu)$. Then

$$e_{\nu+1} = (I - \theta H)e_\nu + O(\|e_\nu\|^2) \quad H = \sum_{i=1}^k H_i$$

$$H_i = \beta(W_i), \quad \beta(x) = x/(e^x - 1)$$

$$W_i = \log(M_i) \otimes I - I \otimes \log(M_i)$$

$$M_i = G^{1/2} A_i^{-1} G^{1/2}$$

Remark

M_i are positive definite as well as H_i and H. Therefore, for $\theta > 0$ small enough, $\rho(I - \theta H) < 1$ and local convergence occurs.
The Courant-Fischer theorem enables to find the optimal value for θ in function of the condition numbers c_i of M_i

Convergence analysis

Corollary

Let c_i be the spectral condition number of M_i. Then for the spectral radius $\rho(I - \theta H)$ one has

$$\rho(I - \theta H) \le \frac{\sum_{i=1}^{k} \log c_i}{\sum_{i=1}^{k} \frac{c_i+1}{c_i-1} \log c_i} < 1, \qquad \theta = \frac{2}{\sum_{i=1}^{k} \frac{c_i+1}{c_i-1} \log c_i}$$

Remarks

- If matrices A_i, $i = 1, \ldots, k$, are "close" to each other and to G, then $M_i \approx I$ so that $c_i \approx 1$ and $\rho(I - \theta H) \approx 0$ independently of $\mathrm{cond}(G)$. In practice, the convergence is fast;
- If matrices A_i, $i = 1, \ldots, k$, commute then $\rho(I - \theta H) = 0$ with $\theta = 1/k$
- If matrices A_i, $i = 1, \ldots, k$, almost commute but are "far" from each other, the choice $\theta = 1/k$ leads to convergence failure

Some numerical experiments

Number of iterations for k random matrices with condition number 10^2 and 10^4, respectively

k	3	4	5	6	7	8	9	10
cond=10^2	17	17	16	16	15	15	14	14
cond=10^4	41	37	35	31	29	29	29	28

Number of iterations for 10 random matrices with condition number 20 and 10^5, respectively, lying in a neighborhood of radius ϵ.

ϵ	0.5	10^{-1}	10^{-2}	10^{-3}	10^{-4}
cond= 20	5	4	2	1	1
cond= 10^5	22	19	14	12	6

Structured mean: definition and algorithms

Fact: Unfortunately, the Riemannian mean does not preserve structure.
$A_1, \ldots, A_k \in \mathcal{A} \subset \mathcal{P}_n \;\not\Rightarrow\; G(A_1, \ldots, A_k) \in \mathcal{A}$

Example: $A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$, $B = I$, $G = \begin{bmatrix} 1.3590 & 0.3860 & -0.0611 & 0.0173 \\ 0.3860 & 1.2978 & 0.4034 & -0.0611 \\ -0.0611 & 0.4034 & 1.2978 & 0.3860 \\ 0.0173 & -0.0611 & 0.3860 & 1.3590 \end{bmatrix}$

A different definition is needed [B., Iannazzo, Jeuris, Vandebril 2013]:

Let $\sigma(t) : \mathbb{R}^q \to \mathbb{R}^{n\times n}$ be a differentiable map and define $\mathcal{A} = \sigma(\mathbb{R}^q) \cap \mathcal{P}_n$, where $\mathcal{P}_n$ is the set of positive definite matrices.

Definition (Structured mean)

The **structured mean with respect to $\mathcal{A}$** of the matrices

$$A_i = \sigma(a_i) \in \mathcal{A}, \quad a_i \in \mathbb{R}^q, \quad i = 1, \ldots, k$$

is the set

$$G_{\mathcal{A}} = \{X = \sigma(t) \in \mathcal{A} : \quad f(X) = \inf_{Y \in \mathcal{A}} f(Y)\}$$

Properties that we can prove

- $G_{\mathcal{A}}$ is not empty if $\mathcal{A}$ is a linear space

Properties that we can prove

- $G_{\mathcal{A}}$ is not empty if $\mathcal{A}$ is a linear space
- If $\mathcal{A}$ is geodesically convex, then the structured mean coincides with the Riemannian mean.

Properties that we can prove

- $G_{\mathcal{A}}$ is not empty if $\mathcal{A}$ is a linear space
- If $\mathcal{A}$ is geodesically convex, then the structured mean coincides with the Riemannian mean.
- If $\sigma(\mathbb{R}^q)$ is a matrix algebra then it is geodesically convex

Properties that we can prove

- $G_{\mathcal{A}}$ is not empty if $\mathcal{A}$ is a linear space
- If $\mathcal{A}$ is geodesically convex, then the structured mean coincides with the Riemannian mean.
- If $\sigma(\mathbb{R}^q)$ is a matrix algebra then it is geodesically convex
- the structured mean is not generally unique

Properties that we can prove

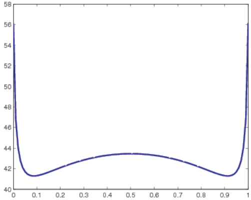
- $G_{\mathcal{A}}$ is not empty if $\mathcal{A}$ is a linear space
- If $\mathcal{A}$ is geodesically convex, then the structured mean coincides with the Riemannian mean.
- If $\sigma(\mathbb{R}^q)$ is a matrix algebra then it is geodesically convex
- the structured mean is not generally unique

Example of non uniqueness

Let $A = I$, $B = \mathrm{diag}(\alpha, \alpha^{-1})$, $\sigma(t) = A + t(B - A)$, $0 \le t \le 1$.

the function $f(t) = \delta^2(\sigma(t), A) + \delta^2(\sigma(t), B)$

is symmetric w.r.t. $t = 1/2$ and for $\alpha = 100$ is such that



Properties that we can prove

[P8] Self duality $G(A_1, \ldots, A_k)^{-1} = G(A_1^{-1}, \ldots, A_k^{-1})$ is satisfied in the following form

$$G_{\mathcal{A}}(A_1, \ldots, A_k)^{-1} = G_{\mathcal{A}^{-1}}(A_1^{-1}, \ldots, A_k^{-1}), \quad \mathcal{A}^{-1} = \{A^{-1} : \quad A \in \mathcal{A}\}$$

The inverse of the structured mean w.r.t. $\mathcal{A}$ coincides with the structured mean of the inverses w.r.t. $\mathcal{A}^{-1}$

[P7] congruence invariance $G(S^T A_1 S, \ldots, S^T A_k S) = S^T G(A_1, \ldots, A_k) S$ is satisfied in the following form

$$G_{S^T \mathcal{A} S}(S^T A_1 S, \ldots, S^T A_k S) = G_{\mathcal{A}}(A_1, \ldots, A_k)$$

Properties that we can prove

[P2] joint homogeneity $G_{\mathcal{A}}(\alpha_1 A_1, \ldots, \alpha_k A_k) = (\prod_i \alpha_i)^{1/k} G(A_1, \ldots, A_k)$ holds if $\mathcal{A}$ is a linear space

[P3] permutation invariance

[P11] repetition invariance: $G(A_1, \ldots, A_k, A_1, \ldots, A_k) = G(A_1, \ldots, A_k)$

Computing the structured mean: A vector equation

The set of structured means $G = \sigma(g)$ is a subset of the set of stationary points for $f(\sigma(t))$, i.e., $\nabla_t f(\sigma(t)) = 0$

The vectors g are the solutions of the vector equation $\nabla_t f(t; a_1, \ldots, a_k) = 0$ such that $\sigma(g)$ is positive definite

From the chain rule of derivatives, this leads to a vector equation which, for $\mathcal{A}$ linear space, takes the form

$$U^T \mathrm{vec}(\sigma(t)^{-1} \sum_{i=1}^{k} \log(\sigma(t) A_i^{-1})) = 0$$

where $\sigma(t) : \mathbb{R}^q \to \mathbb{R}^{n\times n}$ is linear and $\mathrm{vec}(\sigma(t)) = Ut$ for $U \in \mathbb{R}^{q\times n^2}$

Algorithms for solving the matrix/vector equation

In the **structured case** we consider iterations of the kind

$$t^{(\nu+1)} = \varphi(t^{(\nu)}), \quad \varphi(t) = t - \theta V(t)^{-1} \nabla_t f(t)$$

where $V(t)$ is a suitable invertible matrix.

Remark:
If $V(t)$ is the Jacobian matrix of $\nabla_t f(t)$ then the algorithm is Newton's iteration

Newton's iteration is very expensive: the Jacobian depends on all the matrices $A_1, \ldots, A_k$

Cheaper iterations can be obtained by looking for a matrix V which depends only on the current approximation.

Here are some possibilities:

1. $V(t) = D, \quad D = U^T U$
2. $V(t) = \left[U^T(\sigma(t)^{-1} \otimes \sigma(t)^{-1})U\right]$
3. $V^{-1}(t) = D^{-1}U^T(\sigma(t) \otimes \sigma(t))UD^{-1}$

Motivation:

1. projected gradient descent method w.r.t. the Euclidean inner product $\langle A, B\rangle = \text{trace}(AB)$
2. projected gradient descent method w.r.t. the "natural" scalar product $\langle A, B\rangle_X = \text{trace}(AX^{-1}BX^{-1})$
3. "projected version" of the transformation performed in the unstructured case: $\nabla_X f \quad \to \quad X(\nabla_X f)X$

Convergence analysis

Convergence analysis can be performed by computing the Jacobian of $\varphi(\sigma(t))$ through the Fréchet derivative

By performing a convergence analysis similar to that of the unstructured case one finds that the Jacobian of $\varphi(t)$ in G is

$$J = I - \theta V^{-1} U(I \otimes G) H (I \otimes G^{-1}) U$$

where H is the same as in the unstructured case, that is

$$H = \sum_{i=1}^{k} H_i \qquad H_i = \beta(\log(M_i) \otimes I - I \otimes \log(M_i))$$

$$\beta(t) = t/(e^t - 1) \qquad M_i = G^{1/2} A_i^{-1} G^{1/2}$$

In the case of the second algorithm we find exactly the same bounds of the unstructured case:

$$\rho(J) \le \frac{\sum_{i=1}^{k} \log c_i}{\sum_{i=1}^{k} \frac{c_i+1}{c_i-1} \log c_i}, \quad \text{for} \quad \theta = \frac{2}{\sum_{i=1}^{k} \frac{c_i+1}{c_i-1} \log c_i}$$

where

$$c_i = \frac{\lambda_{\max}(M_i)}{\lambda_{\min}(M_i)}, \quad M_i = G^{\frac{1}{2}} A_i^{-1} G^{\frac{1}{2}}$$

Some experiments

Case 1: $n = 10$, $k = 5$, $\epsilon = 10^{-3}$

$$A_i = H + \epsilon \texttt{Toep}(\texttt{rand}(n,1)), \quad i = 1, \ldots, p, \qquad H = \texttt{Toep}([5\,1 \ldots 1])$$

$\text{cond}(H) = 2.5$

Case 2: $n = 10$, $k = 5$, $\epsilon = 10^{-3}$

$$A_i = H + \epsilon \texttt{Toep}(\texttt{rand}(n,1)), \quad i = 1, \ldots, p, \qquad H = \texttt{Toep}([n : -1 : 1])$$

$\text{cond}(H) = 132.36$

Case 3: $n = 10$, $k = 5$, $\epsilon = 10^{-3}$ $A_i = \texttt{Toep}(t)$,

```
t= rand(n,1); t(1)=t(1)-min(eig(Toep(t)))+1.e-3;
```

Number of iterations

Case	Iter. 1	Iter 2.	Iter. 3
1	33	3	11
2	> 1000	3	47
3	> 1000	32	183

Conclusion

- Matrix geometric means (MGM) are needed in the applications
- For two postive definite matrices there is a unique definition of MGM
- For $k > 2$ matrices there are many MGMs
- The NBMP can be computed faster than the ALM mean, however, the cost grows exponentially with k
- The cheap mean has a polynomial cost but does not satisfy the monotonicity property
- The Karcher mean has the good properties: it requires the solution of a matrix equation.
- Effective algorithms exist for solving this matrix equation
- A structured mean has been introduced with the property of preserving structures
- An effective algorithm for its computation has been provided

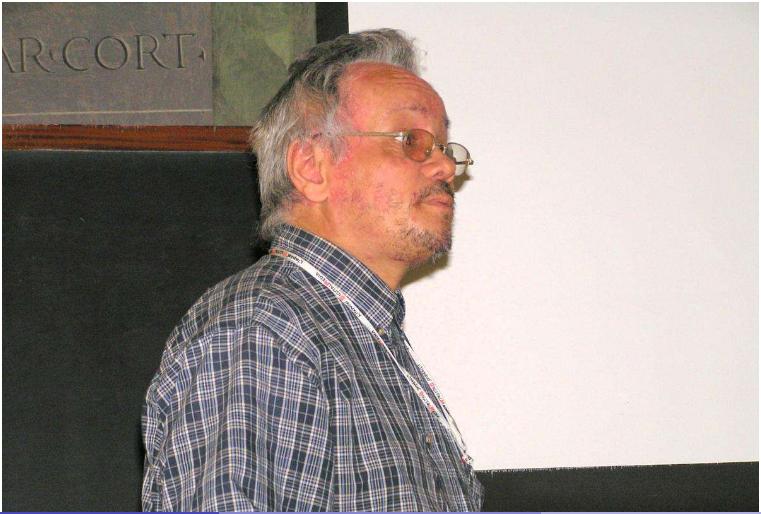
Open problems and things to do

- Global convergence of the Cheap mean
- Monotonicity of the Cheap mean for close input matrices
- Global convergence of the Richardson iteration for the Karcher mean
- Analysis of the distances of the different geometric means
- Conditions for the uniqueness of the structured mean
- Structured means through positive parametrizations

Thank you for your attention

and

Happy birthday to Cor!



Bibliography

- T. Ando, C. Li, and R. Mathias. Geometric means. *Linear Algebra and its Applications*, 385:305–334, 2004.
- R. Bhatia. The Riemannian mean of positive definite matrices. Matrix Information Geometry, (Eds.) F. Nielsen, R. Bhatia, Springer 2013.
- R. Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, 2007.
- R. Bhatia and J. Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra and its Applications*, 413(2-3):594–618, 2006.
- D. Bini, B. Meini, and F. Poloni. An effective matrix geometric mean satisfying the ando–li–mathias properties. *Mathematics of Computation*, 79(269):437–452, 2010.
- D. A. Bini and B. Iannazzo. A note on computing matrix geometric means. *Advances in Computational Mathematics*, 35(2-4):175–192, 2011.
- D. A. Bini, B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. Linear Algebra and Appl., 438(4):1700–1710, 2013.
- D. A. Bini, B. Iannazzo, B. Jeuris, R. Vandebril, Geometric means of structured matrices, BIT, in press.

Bibliography

- N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
- J. Holbrook, No dice: a deterministic approach to the Cartan centroid, J. Ramanujan Math. Soc., Vol. 27 No. 4 – Dec 2012.
- B. Iannazzo. The geometric mean of two matrices from a computational viewpoint. Preprint, available on arXiv, 2011.
- B. Iannazzo, B. Meini. Palindromic matrix polyomials, matrix functions and integral representations, Linear Algebra Appl., 434-1 (2011), pp. 174-184.
- B. Jeuris, R. Vandebril, and B. Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39:379–402, 2012.
- J. Lapuyade-Lahorgue and F. Barbaresco. Radar detection using Siegel distance between autoregressive processes, application to HF and X-band radar. In *Radar Conference, 2008. RADAR '08, IEEE*, Rome, May 2008.
- J. Lawson and Y. Lim. Monotonic properties of the least squares mean. *Mathematische Annalen*, 351(2):267–279, 2011.

Bibliography

- Y. Lim, Y., M. Pálfia, The matrix power means and the Karcher mean. J. Funct. Anal. 262, 1498–1514 (2012)
- M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.
- M. Moakher. On the averaging of symmetric positive-definite tensors. *Journal of Elasticity*, 82(3):273–296, 2006.
- K. Nakamura. Geometric means of positive operators. *Kyunpook Mathematical Journal*, 49:167–181, 2009.
- M. Pálfia. The Riemann barycenter computation and means of several matrices. *Int. J. Comput. Math. Sci*, 3(3):128–133, 2009.

Matrix Function

Given $A = SDS^{-1}$, $D = \mathrm{diag}(d_1, \ldots, d_n)$, and a function F defined on $\{d_1, \ldots, d_n\}$, define

$$F(A) = SF(D)S^{-1}, \quad F(D) = \mathrm{diag}(F(d_1), \ldots, F(d_n))$$

back