The Abdus Salam
International Centre
for Theoretical Physics

2470-6

Advanced School and Workshop on Matrix Geometries and Applications

1 - 12 July 2013

Computational issues of matrix geometric means (3)

Dario Bini
Università di Pisa
Italy

Strada Costiera, 11 - 34151 - Trieste - Italy • Tel. +39 0402240111 • Fax. +39 040224163 • sci_info@ictp.it • www.ictp.it

Computational issues of matrix geometric means
Part 3

D.A. Bini
Università di Pisa

Trieste – July 2013

Contents

We have seen that computing the ALM mean of k matrices of size n has a cost of $O(k!p^k n^3)$ where p is the number of iterations required at each step.

The mean based on medians allows one to reduce substantially the number p of iterations due to its cubic convergence

Is there a way to overcome the exponential complexity?

Two solutions:

- the Cheap mean
- the Karcher mean

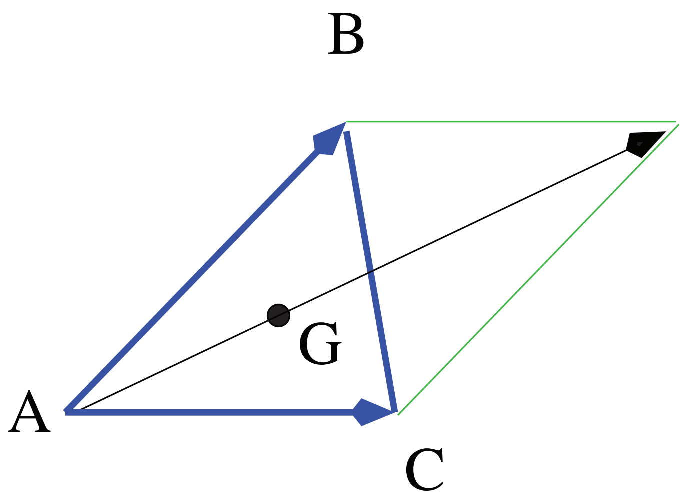
The cheap mean (overcoming the exponential complexity)

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 **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

No examples are known where the sequences do not converge.

An equivalent formulation

The iteration can be reformulated in a more convenient way. For simplicity, we omit the iteration index ν

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

From this expression it follows that A_i' is symmetric and positive definite, so the sequences are well defined

What we can prove

Theorem (Local convergence)

If the three sequences converge then 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

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

Properties P6, P7 and P10 are usually proved relying on monotonicity. It is not clear if they are satisfied

We refer to G as the **cheap mean**

Convergence

We prove that the sequences cannot converge to different limits

If by absurd, $\lim_\nu A_\nu = \overline{A}$, $\lim_\nu B_\nu = \overline{B}$, $\lim_\nu C_\nu = \overline{C}$, then from the fixed point iteration one obtains

$$\overline{A} = \overline{A} \exp\left(\frac{1}{3}\left(\log(\bar{A}^{-1}\overline{B}) + \log(\bar{A}^{-1}\overline{C})\right)\right)$$

that is,

$$\log(\bar{A}^{-1}\overline{B}) + \log(\bar{A}^{-1}\overline{C}) = 0$$

This equation implies that

- $\overline{A}$ is the ALM mean of $\overline{B}$ and $\overline{C}$
- $\overline{B}$ is the ALM mean of $\overline{A}$ and $\overline{C}$
- $\overline{C}$ is the ALM mean of $\overline{A}$ and $\overline{B}$

This is the iteration step for computing the ALM mean. This step is contractive, therefore $\overline{A}$ cannot coincide with $G(\overline{B}, \overline{C})$

Order of convergence

We prove that the order of convergence is 3. Consider the iteration

$$A_i' = A_i^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1}^{k} \log(A_i^{-\frac{1}{2}} A_j A_i^{-\frac{1}{2}})\right) A_i^{\frac{1}{2}}$$

Introduce the following notation

$$A_i^{-\frac{1}{2}} A_j A_i^{-\frac{1}{2}} = I + A_i^{-\frac{1}{2}}(A_j - A_i) A_i^{-\frac{1}{2}} =: I + X_{i,j}$$

$$E_{i,j} = A_j - A_i, \quad X_{i,j} = A_i^{-\frac{1}{2}} E_{i,j} A_i^{-\frac{1}{2}}$$

Rewrite the iteration as

$$A_i' = A_i^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1}^{k} \log(I + X_{i,j})\right) A_i^{\frac{1}{2}}$$

Recall that if $\|X\|, \|W\| \le \epsilon < 1$ then

$$\log(I+X) = X - \frac{1}{2}X^2 + O(\epsilon^3)$$
$$\exp(W) = I + W + \frac{1}{2}W^2 + O(\epsilon^3)$$

Use the above equations for $X = X_{ij}$ in,

$$A_i' = A_i^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1}^{k} \log(I + X_{i,j})\right) A_i^{\frac{1}{2}}$$

and obtain

$$A_j' = A_j + \frac{1}{k}\sum_{i=1}^{k} E_{i,j} - \frac{1}{2k}\sum_{i=1}^{k} E_{i,j}A_j^{-1}E_{i,j} + \frac{1}{2k^2}\sum_{r,s=1}^{k} E_{r,j}A_j^{-1}E_{s,j}$$

Subtracting two copies of equation

$$A_j' = A_j + \frac{1}{k}\sum_{i=1}^{k} E_{i,j} - \frac{1}{2k}\sum_{i=1}^{k} E_{i,j}A_j^{-1}E_{i,j} + \frac{1}{2k^2}\sum_{r,s=1}^{k} E_{r,j}A_j^{-1}E_{s,j}$$

for indices h and j yields

$$E_{h,j}' = E_{h,j} + \frac{1}{k}\sum_{i=1}^{k}(E_{i,h} - E_{i,j}) - \frac{1}{2k}\sum_{i=1}^{k}(E_{i,h}A_h^{-1}E_{i,h} - E_{i,j}A_j^{-1}E_{i,j})$$
$$+ \frac{1}{2k^2}\sum_{r,s=1}^{k}(E_{r,h}A_h^{-1}E_{s,h} - E_{r,j}A_j^{-1}E_{s,j})$$

By definition of $E_{i,j}$, the blue term is zero. Moreover, from the identities

$$XYZ - UVW = (X-U)YZ + U(Y-V)Z + UV(Z-W),$$
$$Y^{-1} - W^{-1} = Y^{-1}(W-Y)W^{-1}$$

it follows that the black term is $O(\epsilon^3)$

Proving the ALM properties

We prove them in the case of three matrices A, B, C

Consistency with scalars

$$
\begin{aligned}
A' &= A \exp\left(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)\right) \\
&= A \exp\left(\frac{1}{3}(\log(A^{-1}B)A^{-1}C)\right) \\
&= A(A^{-2}BC)^{\frac{1}{3}} = (ABC)^{\frac{1}{3}}
\end{aligned}
$$

The same holds for B' and C'

Observe that one step of the iteration provides the geometric mean

Proving the ALM properties

Joint homogeneity

Let $\widehat{A} = \alpha A$, $\widehat{B} = \beta B$, $\widehat{C} = \gamma C$

$$\begin{aligned}
\widehat{A}' &= \alpha A \exp\left(\frac{1}{3}\left((\log(A^{-1}B\frac{\beta}{\alpha}) + \log(A^{-1}C\frac{\gamma}{\alpha})\right)\right) \\
&= \alpha A \exp\left(\frac{1}{3}\left((\log(A^{-1}B) + \log(A^{-1}C) + \log(\frac{\beta\gamma}{\alpha^2}I)\right)\right) \\
&= \alpha A \exp\left(\frac{1}{3}\left(\log(A^{-1}B) + \log(A^{-1}C)\right)\right) \exp\log(\left(\frac{\beta\gamma}{\alpha^2}\right)^{\frac{1}{3}}) \\
&= (\alpha\beta\gamma)^{\frac{1}{3}} A'
\end{aligned}$$

Proving the ALM properties

Congruence invariance

Let $\widehat{A} = S^T AS$, $\widehat{B} = S^T BS$, $\widehat{C} = S^T CS$,

$$\begin{aligned}\widehat{A}' &= \widehat{A} \exp\left(\frac{1}{3}(\log(\widehat{A}^{-1}\widehat{B}) + \log(\widehat{A}^{-1}\widehat{C})\right) \\ &= S^T AS \exp\left(\frac{1}{3}(\log(s^{-1}A^{-1}BS) + \log(s^{-1}A^{-1}CS)\right) \\ &= S^T ASS^{-1} \exp\left(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)\right) S = S^T A' S\end{aligned}$$

The same holds for $\widehat{B}'$ and $\widehat{C}'$

Proving the ALM properties

Self duality

Recall that

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

so that

$$\begin{aligned}(A')^{-1} &= \exp\left(-\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C)\right) A^{-1} \\ &= \exp\left(\log(B^{-1}A)^{\frac{1}{3}} + \log(C^{-1}A)^{\frac{1}{3}}\right) A^{-1} \\ &= A^{-1} \exp\left(\log(AB^{-1})^{\frac{1}{3}} + \log(AC^{-1})^{\frac{1}{3}}\right) = \widehat{A}'\end{aligned}$$

Proving the ALM properties

Determinant identity

It follows from the property

$$\det(\exp(A+B)) = \det(\exp(A)\exp(B))$$

In fact, one has

$$\begin{aligned}\det A' &= \det A \det(\exp(\log(A^{-1}B)^{\frac{1}{3}})) \det(\exp(\log(A^{-1}C)^{\frac{1}{3}})) \\ &= \det A \det(A^{-1}B)^{\frac{1}{3}} \det(A^{-1}C)^{\frac{1}{3}} = (\det A \det B \det C)^{\frac{1}{3}}\end{aligned}$$

The lack of monotonicity

Take $\epsilon = 10^{-4}$, $0 < h \leq 3$ and define the triples A, B, C and $\widetilde{A}, B, C$ as

$$A = I, \quad \widehat{A} = I + hee^T, \quad B = \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Clearly, $\widehat{A} \geq A$, while $Q = G(\widehat{A}, B, C) - G(A, B, C)$ has a negative eigenvalue. For instance, for $h = 1$ the eigenvalues of Q are

```
-2.4131e-3,  2.2853e-2,  1.0826e-1
```

Implementation

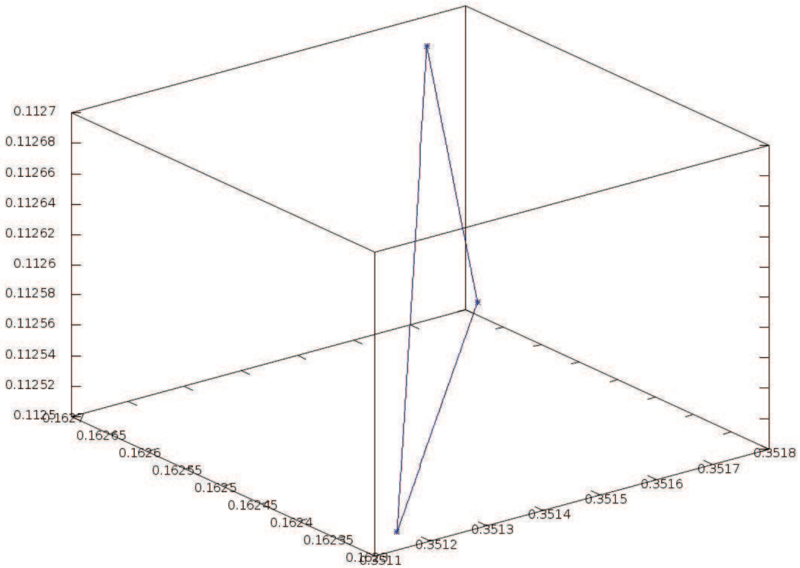
```
function [C,iter]=cheap(varargin)
% [C,iter]=CHEAP(A1,A2,...,AP) computes the Cheap mean
A=varargin; n=length(A{1}); k=length(A); tol=1d-14; maxiter=200; ct=0;
for iter=1:maxiter
    for i=1:k
        R{i}=chol(A{i});  RI{i}=inv(R{i});  S=zeros(n);
        for h=1:k
            if (h~=i);  S=S+logm(RI{i}'*A{h}*RI{i});   end
        end
        A1{i}=R{i}'*expm(1/k*S)*R{i};
    end
    ni=norm(A1{1}-A{1});
    if ni<tol  break  end
    A=A1;
end
C=A{1};
if ct==maxiter disp('Max number of iterations reached')  end
```

Numerical experiments: plotting the means

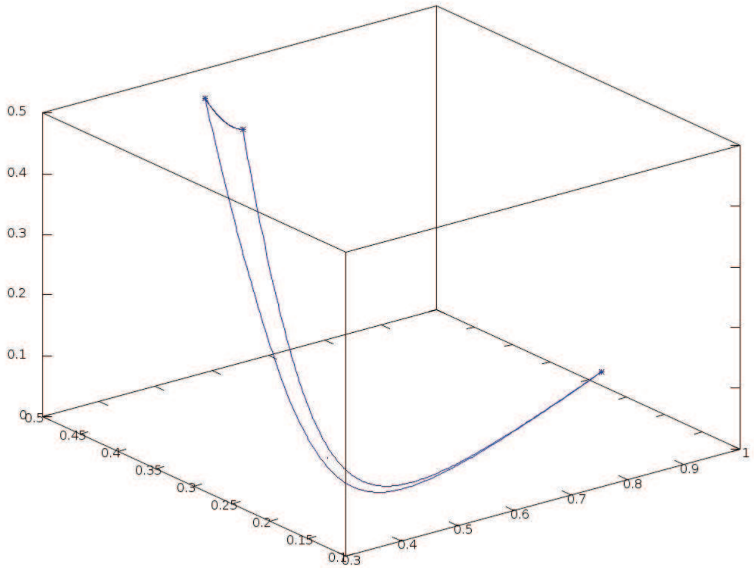
```
x=rand(2); a=x'*x;  x=rand(2); b=x'*x; x=rand(2); c=x'*x;
galm=alm(a,b,c);
gnbmp=nbmp(a,b,c);
gcheap=cheap(a,b,c);
plot_matrix(galm,'r*');
hold on;
plot_matrix(gnbmp,'b*');
plot_matrix(gcheap,'g*');
plot_triangle(galm,gnbmp,gcheap);
hold off;
```

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

Numerical experiments: timings

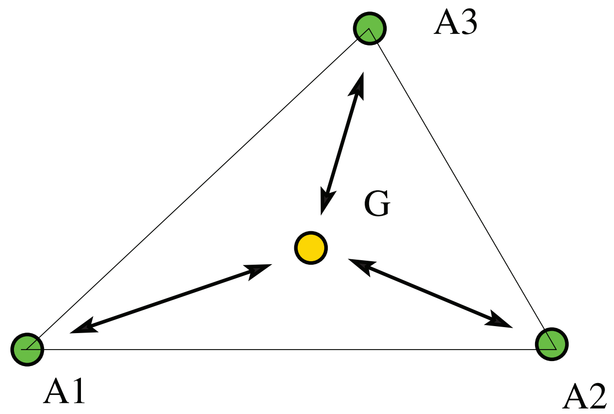
k	cnd= 1.e2 Cheap	NBMP	Dist.	cnd= 1.e4 Cheap	NBMP	Dist.	cnd= 1.e8 Cheap	NBMP	Dist.
3	1.e-2	1.e-2	5.e-3	1.e-2	1.e-2	3.e-2	1.e-2	1.e-2	3.e-2
4	2.e-2	2.e-1	6.e-3	2.e-2	2.e-1	2.e-2	2.e-2	2.e-2	8.e-2
5	2.e-2	1.e0	7.e-3	3.e-2	2.e0	4.e-2	3.e-1	2.e0	5.e-2
6	3.e-2	1.e+1	5.e-2	4.e-2	3.e+1	2.e-2	4.e-2	3.e+1	5.e-2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

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.

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



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].

Existence and uniqueness follow from the fact that $f(X)$ is *geodesically convex*, i.e., it is convex along any geodesic

The Karcher mean

The point where the function takes its minimum is the unique positive definite matrix which solves the equation $\nabla f(X) = 0$, where $\nabla f(X)$ is the gradient of $f(x)$.

It holds that the gradient of $f(X)$ is

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

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

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

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