The Abdus Salam International Centre for Theoretical Physics

2470-4

Advanced School and Workshop on Matrix Geometries and Applications

1 - 12 July 2013

Computational issues of matrix geomtric means (1)

Dario Bini

Università di Pisa
Italy

Computational issues of matrix geometric means

D.A. Bini
Università di Pisa

Trieste – July 2013

Aims and contents

We wish to present and discuss some **computational issues** related to matrix geometric means including

- theory and formal properties
- design and analysis of algorithms for their computation
- implementation and use of the algorithms in Octave/Matlab

Four parts:

1. Means of two matrices, motivation and preliminaries
2. Means of $k > 2$ matrices, complexity analysis
3. Overcoming the exponential complexity: the Cheap mean
4. Overcoming the exponential complexity: the (structured) Karcher mean

Contents of the first part

- motivation of the problem and applications
- matrix functions: square root, exponential and logarithm
- mean of two matrices: properties and algorithms
- implementation in Octave/Matlab
- plotting geodesics in the Riemannian geometry
- plotting matrices and their geometric mean
- analysis of the numerical conditioning:
  The Fréchet derivative and the Kronecker product

The Problem and its motivations

In certain applications we are given matrices $A_1, \ldots, A_k \in \mathcal{P}_n$ which represent measures of some physical object. Here, $\mathcal{P}_n$ is the set of $n \times n$ real symmetric positive definite matrices

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

Something about matrix functions

In the following we need the concept of function of matrices

We restrict ourselves to diagonalizable matrices

For the treatment of the general case, we refer to the book by Nick Higham "Function of Matrices: Theory and Computation," SIAM 2008

Definition

Let A be an $n \times n$ matrix such that $A = VDV^{-1}$, where $D = \mathrm{diag}(d_1, \ldots, d_n)$ and $\det V \neq 0$.
Let $f(x) : \Omega \to \mathbb{R}$ be such that $d_1, \ldots, d_n \in \Omega$.
We define

$$f(A) = Vf(D)V^{-1}, \quad \text{where} \quad f(D) = \mathrm{diag}(f(d_1), \ldots, f(d_n))$$

Something about matrix functions

Some properties

- if A is real symmetric then $f(A)$ is real symmetric
- $f(SAS^{-1}) = Sf(A)S^{-1}$ for any nonsingular matrix S
- $f(A)^T = f(A^T)$
- $f(A)$ is a polynomial in A
- $f(A)$ commutes with A

Functions of interest:

- $F = \exp(A)$: A is any matrix `F=expm(A)`
- $F = \log(A)$: A has positive eigenvalues `F=logm(A)`
- $F = \mathrm{sqrt}(A)$: A has nonnegative eigenvalues `F=sqrtm(A)`
- $F = A^t$, $t \in \mathbb{R}$: A has nonnegative eigenvalues `F=expm(t*logm(A))`

Means of two matrices

Many authors analyzed the problem of extending the concept of geometric mean from scalars to matrices. For a survey and historical notes see
R. Bhatia, "The Riemannian mean of positive definite matrices", Springer 2013

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

A good definition

$$G(A,B) = A(A^{-1}B)^{\frac{1}{2}} = A^{\frac{1}{2}}\left[A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right]^{\frac{1}{2}}A^{\frac{1}{2}}$$

Proof of the second equality

$$A(A^{-1}B)^{\frac{1}{2}} = A(A^{-\frac{1}{2}}A^{-\frac{1}{2}}BA^{-\frac{1}{2}}A^{\frac{1}{2}})^{\frac{1}{2}} = A^{\frac{1}{2}}\left[A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right]^{\frac{1}{2}}A^{\frac{1}{2}}$$

The Ando-Li-Mathias properties

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

We refer to $G(A, B)$ as the **matrix geometric mean**

Throughout, for $A, B \in \mathcal{P}_n$, we write $A \succeq B$ if $A - B$ is positive semidefinite

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

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

P3 Symmetry. $G(A,B)=G(B,A)$

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

P5 Congruence invariance. $G(S^TAS,S^TBS)=S^TG(A,B)S$

P6 Continuity from above. If A_j, B_j are monotonic decreasing sequences converging to A, B, respectively, then $\lim_j G(A_j,B_j)=G(A,B)$

P7 Joint concavity. If $A=\lambda A_1+(1-\lambda)A_2$, $B=\lambda B_1+(1-\lambda)B_2$, then

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

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

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

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

$$d(A,B) = \| \log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) \|_F, \quad ||X||_F = (\mathrm{trace}(X^T X))^{\frac{1}{2}}$$

For scalars, $d(a,b) = |\log(a) - \log(b)|$: the Riemannian distance coincides with the Euclidean distance of the logarithms

	Euclidean metric	Riemannian metric
equation of the segment	$(1-t)a + tb$	$a^{1-t} b^t$
midpoint of the segment	$(a+b)/2$	$(ab)^{1/2}$

Motivation in terms of Riemannian geometry

Some properties of $d(A,B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|_F$

- $d(A,B) = d(B,A)$

$$d(A,B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|_F = \|\log(SS^T)\|_F, \quad S = A^{-\frac{1}{2}}B^{\frac{1}{2}}$$

$$d(B,A) = \|\log(B^{-\frac{1}{2}}AB^{-\frac{1}{2}})\|_F = \|\log(RR^T)\|_F, \quad R = B^{-\frac{1}{2}}A^{\frac{1}{2}}$$

$$R = S^{-1} \quad \Rightarrow \quad \log(SS^T) = -\log(R^TR) \quad \Rightarrow \quad d(A,B) = d(B,A)$$

since R^TR and RR^T are orthogonally similar

(if $R = U^T\Sigma V$ is the SVD of R then $R^TR = V^T\Sigma^2V$, $RR^T = U^T\Sigma^2U$)

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

$$\begin{aligned} d(A,B) =& \|\log(SS^T)\|_F = \|\log(S^{-T}S^{-1})\|_F \\ =& \|\log(A^{\frac{1}{2}}B^{-1}A^{\frac{1}{2}})\|_F = d(A^{-1},B^{-1}) \end{aligned}$$

The geodesic joining A and B has equation

$$\gamma(t) = A(A^{-1}B)^t = A^{\frac{1}{2}} \left[A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right]^t A^{\frac{1}{2}}, \quad t \in [0, 1],$$

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

Remarks:

- The midpoint of a segment $[A, B]$ in the Euclidean geometry corresponds to the arithmetic mean $M = (A + B)/2$
- The midpoint of a segment in the Riemannian geometry corresponds to the geometric mean
- Changing the metric in “logarithmic way” has had the effect to move from the arithmetic to the geometric mean without the need of applying the exp-log transformation
- the geometric interpretation has the advantage to provide hints and suggestions on designing algorithms

Computing the geometric mean of two matrices

The mean

$$G = A(A^{-1}B)^{\frac{1}{2}} = A^{\frac{1}{2}} \left[A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right]^{\frac{1}{2}} A^{\frac{1}{2}}$$

can be computed by performing 2 matrix square roots, one matrix inversion and 4 matrix multiplications

The same mean can be viewed as the solution of the matrix equation

$$XA^{-1}X = B$$

which is a particular instance of the algebraic Riccati equation

$$XFX + XG + HX + L = 0$$

Any algorithm for solving this equation is a tool for computing the geometric mean G [D. Bini, B. Iannazzo, B. Meini, Numerical solution of algebraic Riccati equations, SIAM 2012]

Computing the geometric mean of two matrices

Using the formula $G = A(A^{-1}B)^{1/2}$, even though possible, is deprecable from the numerical point of view

Better algorithms can be obtained by relying on the formulae:

$$A(A^{-1}B)^t = R\left[R^{-1}BR^{-1}\right]^t R, \qquad R = A^{1/2}$$
$$A(A^{-1}B)^t = B(B^{-1}A)^{1-t} = S\left[S^{-1}AS^{-1}\right]^{1-t} S, \quad S = B^{1/2}$$

Observe that we have factored $A = A^{\frac{1}{2}}A^{\frac{1}{2}}$ and $B = B^{\frac{1}{2}}B^{\frac{1}{2}}$

In the following we denote $A\#_t B := A(A^{-1}B)^t$

We design a matlab function that computes $A\#_t B$ relying on the above formulae

Computing the geometric mean of two matrices

Observe that the first formula requires the inversion of $A^{\frac{1}{2}}$ while the second formula requires the inversion of $B^{\frac{1}{2}}$

Recall that for a given operator norm $\|\cdot\|$, the **condition number** of A is $\mathrm{cond}(A) = \|A\|\,\|A^{-1}\|$. It measures the sensibility of the variation of the entries of A^{-1} under perturbations of the entries of A.

The larger the condition number, the larger the numerical errors expected in the computed values of A^{-1}.

For this reason it is more convenient to choose the formula where the condition number of the matrix to invert is smaller

The matrix H^t for a given $H \in \mathcal{P}_n$ is computed by means of its Schur decomposition as follows

$$H^t = QD^tQ^T, \quad H = QDQ^T, \quad Q^TQ = I, \quad D \ \text{ diagonal}$$

```
[Q,D]=schur(H);      H=Q*diag(diag(D).^t)*Q'
```

Some matlab functions

```
function G =sharp(A,B,t)
% G = sharp(A,B,t) computes the point G=g(t) of the geodesic
%     g joining A and B, g(0)=A and g(1)=B
%   A,B: positive definite matrices, t real

mA=cond(A);  mB=cond(B);
if  mA<=mB
    RA=sqrtm(A);
    IRA=inv(RA);
    [U V]=schur(IRA'*B*IRA);
    G=RA'*U*diag(diag(V).^t)*U'*RA;
else
    RB=sqrtm(B);
    IRB=inv(RB);
    [U V]=schur(IRB'*A*IRB);
    G=RB'*U*diag(diag(V).^(1-t))*U'*RB;
end
```

A more efficient computation

A more efficient matlab function can be designed by replacing the factorization

$$A = A^{\frac{1}{2}} A^{\frac{1}{2}}$$

with the Cholesky factorization

$$A = LL^T, \quad L \text{ lower triangular}$$

That is,

$$A(A^{-1}B)^t = L\left[L^{-1}BL^{-1}\right]^t L, \qquad A = LL^T$$

$$A(A^{-1}B)^t = B(B^{-1}A)^{1-t} = L\left[L^{-1}AL^{-1}\right]^{1-t}, \quad B = LL^T$$

Computing the Cholesky factor L is much less expensive than computing the matrix square root. The Octave/Matlab command

```
S = chol(A);
```

provides the matrix $S = L^T$ such that $A = LL^T$

Some matlab functions

```
function G =sharp_cy(A,B,t)
% G = sharp_cy(A,B,t) computes the point G=g(t) of the geodes:
%      g joining A and B, g(0)=A and g(1)=B
%    A,B: positive definite matrices, t real

mA=cond(A);  mB=cond(B);
if  mA<=mB
    L=chol(A);
    IL=inv(L);
    [U V]=schur(IL'*B*IL);
    G=L'*U*diag(diag(V).^t)*U'*L;
else
    L=chol(B);
    IL=inv(L);
    [U V]=schur(IL'*A*IL);
    G=L'*U*diag(diag(V).^(1-t))*U'*L;
end
```

Other algorithms for computing $A\#_{\frac{1}{2}}B$

- Averaging techniques [Anderson Trapp, 1980]
- Matrix sign function [Higham et al.]
- Cyclic reduction [Iannazzo, Meini]
- Continued fractions [Raïssouli, Leazizi, 2003]
- Polar decomposition [Iannazzo 2011]
- Gaussian quadrature applied to $A\#_t B = \frac{1}{\pi}\int_0^1 \frac{(tB^{-1}+(1-t)A^{-1})^{-1}}{\sqrt{t(1-t)}}dt$ [Iannazzo 2011]

For more details see B. Iannazzo, The geometric mean of two matrices from a computational point of view. arXiv 1201.0101v, 2011

Plotting matrices

A symmetric 2×2 matrix $A = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix}$ can be uniquely associated with $(x_1, x_2, x_3) \in \mathbb{R}^3$. Therefore it can be easily plotted

Function for plotting a matrix A

```
function plot_matrix(a,style)
% plot_matrix(a) plots the position of the 2x2
% matrix a with a given style
% INPUT: a:  2x2 matrix
%        style: string with the plot style. Examples of style
%        'r*', 'b+', 'gx'

plot3(a(1,1),a(1,2),a(2,2),style)
```

Plotting geodesics

Function for plotting the geodesic joining A, B

```
function plot_geodesic(a,b)
% plot_geodesic(a,b) plot the geodesic joining a and b
% a,b: 2x2 positive definite matrices
n=200;
t=0:1/(n-1):1;
for i=1:n
   g=sharp(a,b,t(i));
   x(i)=g(1,1);   y(i)=g(1,2);   z(i)=g(2,2);
end
plot3(x,y,z);
```

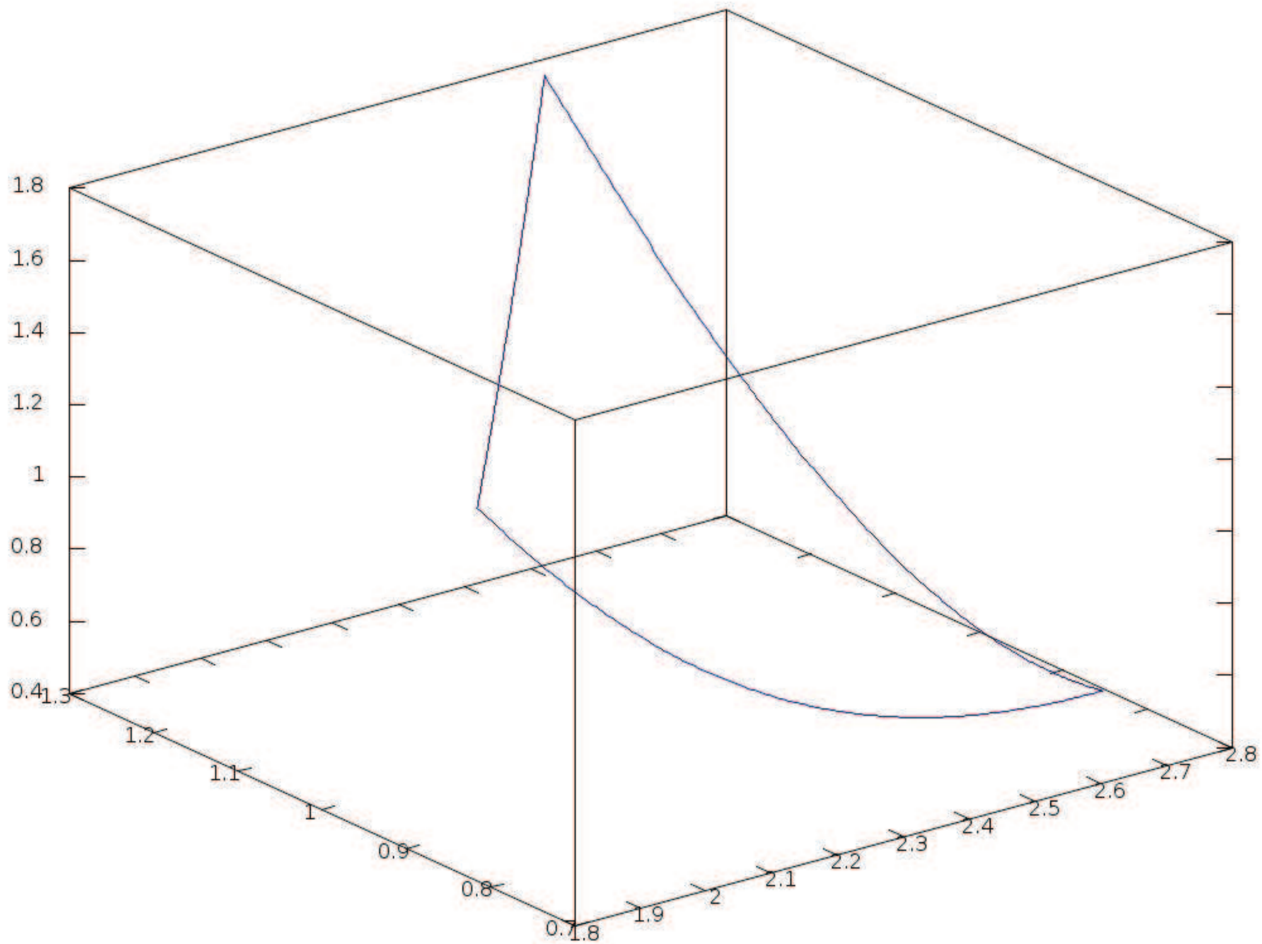
Plotting geodesics

Function for plotting a Riemannian triangle with vertices A, B and C

```
function plot_triangle(a,b,c)
% plot_triangle(a,b,c) plot the triangle of vertices a,b,c
% in the riemannian geometry
% a,b,c: 2x2 positive definite matrices
n=200;
t=0:1/(n-1):1;
for i=1:n
   g=sharp(a,b,t(i));
     x(i)=g(1,1); y(i)=g(1,2); z(i)=g(2,2);
   g=sharp(b,c,t(i));
     x(i+n)=g(1,1); y(i+n)=g(1,2); z(i+n)=g(2,2);
   g=sharp(c,a,t(i));
     x(i+2*n)=g(1,1); y(i+2*n)=g(1,2); z(i+2*n)=g(2,2);
end
plot3(x,y,z);
```

Some pictures

Well conditioned matrices, close to each other



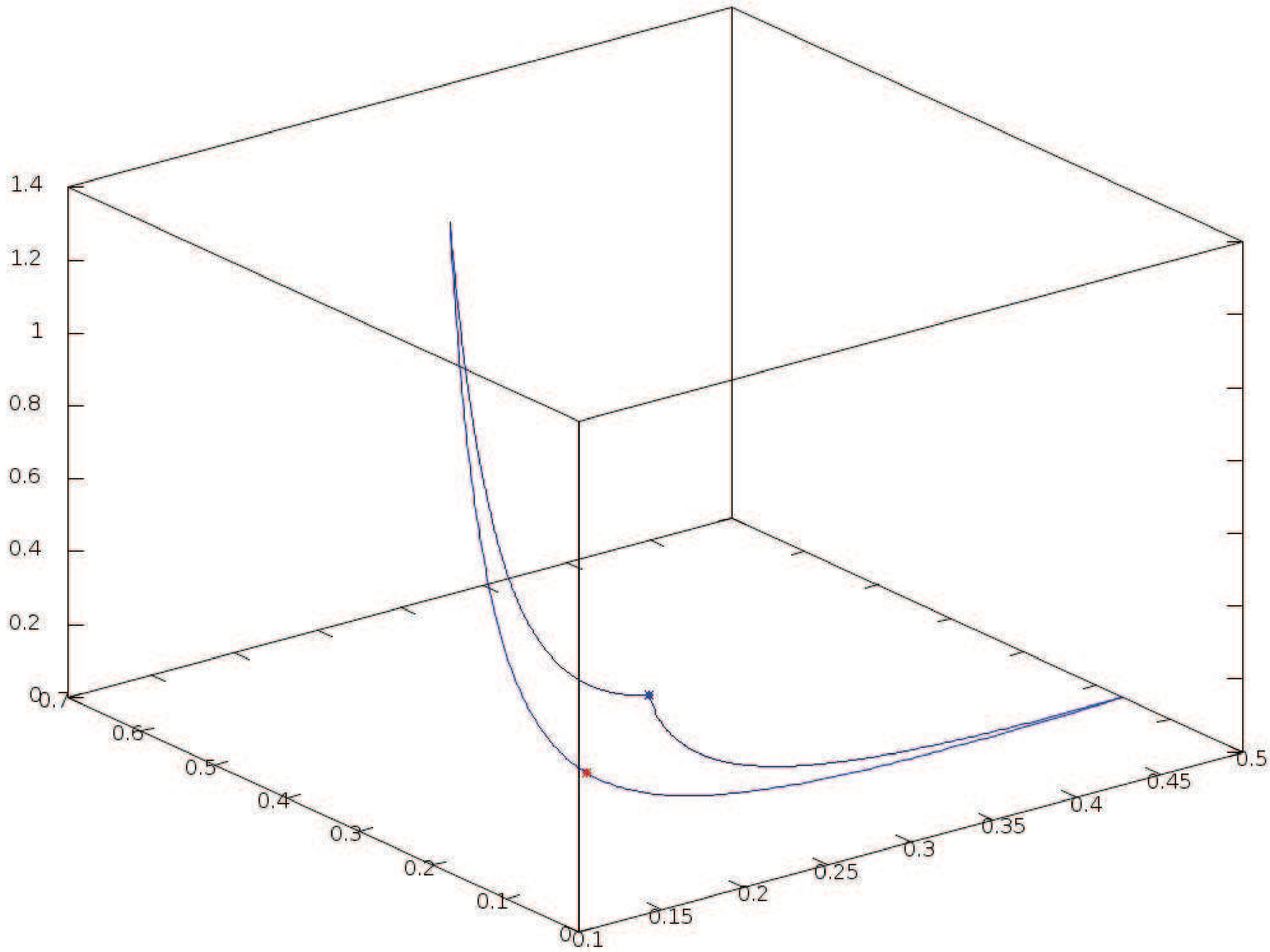
Explog mean and geometric mean

Let us try to compare the geometric mean and the explog mean

Given 2×2 matrices A, B, the following code plots the Riemannian triangle of vertices A, B and H where $H = \exp(\frac{1}{2}(\log A + \log B))$ is the explog mean. Moreover it marks the geometric mean with a red $*$ and the explog mean with a blue $*$.

```
gxpl=expm(0.5*logm(a)+0.5*logm(b)); % explog mean
g=sharp(a,b,0.5);                   % geometric mean
plot_triangle(a,b,gxpl);
hold on
plot_matrix(g,'r*');
plot_matrix(gxpl,'b*');
```

Explog mean and geometric mean



Explog mean and geometric mean

The explog mean does not satisfy the following ALM properties

- P4 Congruence invariance
- P5 Monotonicity

Analysis of the conditioning: tools

We analyze the sensitivity of the matrix geometric mean

$$G : \mathcal{P}_n \times \mathcal{P}_n \to \mathcal{P}_n, \quad G(X, Y) = X \#_{\frac{1}{2}} Y$$

to perturbations in both arguments X and Y.
This analysis can be carried out by relying on the following tools:

The Kronecker product

The Fréchet derivative

These tools will be fundamental later on in the convergence analysis of some matrix iterations

The Kronecker product

Given matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, define the **Kronecker product**

$$A \otimes B = [a_{i,j}B] \in \mathbb{R}^{mp\times nq}$$

In particular

$$I \otimes B = \begin{bmatrix} B & & \\ & \ddots & \\ & & B \end{bmatrix}, \quad A \otimes I = \begin{bmatrix} a_{1,1}I & \dots & a_{1,n}I \\ \vdots & \ddots & \vdots \\ a_{m,1}I & \dots & a_{m,n}I \end{bmatrix}$$

The Kronecker product is related to the operator "vec", where $\mathrm{vec}(A) \in \mathbb{R}^{mn}$ is the vector obtained by stacking the columns of A.

$$\mathrm{vec}([u_1, u_2, \dots, u_n]) = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Kronecker product

$$\text{vec}(AU) = \text{vec}([Au_1, Au_2, \ldots, Au_n]) = \begin{bmatrix} Au_1 \\ Au_2 \\ \vdots \\ Au_n \end{bmatrix}$$

$$= \begin{bmatrix} A & & \\ & \ddots & \\ & & A \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = (I \otimes A)\text{vec}(U)$$

Similarly, $\text{vec}(UA) = (A^T \otimes I)\text{vec}(U)$

Kronecker product

In the "vec" notation we can represent linear operators on matrices in terms of an $n^2 \times n^2$ matrix:

$$U \to AU \qquad I \otimes A$$

$$U \to UA \qquad A^T \otimes I$$

$$U \to AUB \qquad B^T \otimes A$$

$$U \to AU + UA \qquad I \otimes A + A^T \otimes I$$

Kronecker product

Properties

- $\mathrm{vec}(AB) = (I \otimes A)\mathrm{vec}(B) = (B^T \otimes I)\mathrm{vec}(A)$
- $(A \otimes B)(C \otimes D) = (A \otimes C)(B \otimes D)$
- $\det(A_n \otimes B_m) = (\det A_n)^m (\det B_m)^n$, for $A_n \in \mathbb{R}^{n\times n}$, $B_m \in \mathbb{R}^{m\times m}$.
- $(A_n \otimes B_m)^{-1} = A_n^{-1} \otimes B_m^{-1}$
- $(A_n \otimes B_m)^T = A_n^T \otimes B_m^T$
- $\lambda_{i,j}(A_n \otimes B_m) = \lambda_i(A_n)\lambda_j(B_m)$, $i = 1 : n$, $j = 1 : m$.
- $A_n x = \lambda x$, $B_m y = \lambda y \Rightarrow (A_n \otimes B_m)(x \otimes y) = \lambda\mu(x \otimes y)$
- $\lambda_{i,j}(I_n \otimes B_m + A_n \otimes B_m) = \lambda_i(A_n) + \lambda_j(B_m)$, $i = 1 : n$, $j = 1 : m$.

The Fréchet derivative

Given the matrix function $F(X)$, the Fréchet derivative of $F(X)$ at the point X is a linear mapping $E \in \mathbb{R}^{n\times n} \to L_F(X,E) \in \mathbb{R}^{n\times n}$ such that

$$F(X+E) - F(X) = L_F(X,E) + o(\|E\|).$$

We denote $K_F(X)$ the $n^2 \times n^2$ matrix which represents $L_F(X,E)$ in the "vec" notation

$$\text{vec}(L_F(X,E) = K_F(X)\text{vec}(E)$$

Properties:

$L_{g*h}(X,E) = L_g(X,E)h(X) + g(X)L_h(X,E)$	product rule
$L_{g\circ h}(X,E) = L_g(h(X), L_h(X,E))$	chain rule
$L_{f^{-1}}(f(X),E) = L_f^{-1}(X,E)$	derivative of inverse functior

The same properties in terms of the matrix K

$$K_{g*h}(X,E) = (h(X)^T \otimes I)K_g(X) + (I \otimes g(X))K_h(X)$$

$$K_{g\circ h}(X,E) = K_g(h(X))K_h(X)$$

$$K_{f^{-1}}(f(X),E) = K_f(X)^{-1}$$

Examples:

$F(X)$	$L_F(X,E)$	$K_F(X)$
X	$E \to E$	I
AXB	$E \to AEB$	$B^T \otimes A$
X^2	$E \to XE + EX$	$I \otimes X + X^T \otimes I$
$X^{1/2}$	$E \to F: \quad X^{\frac{1}{2}}F + FX^{\frac{1}{2}} = E$	$(I \otimes X^{\frac{1}{2}} + X^{\frac{1}{2}\,T} \otimes I)^{-1}$
X^{-1}	$E \to -X^{-1}EX^{-1}$	$X^{-T} \otimes X^{-1}$

Numerical conditioning

Let

$$G : \mathcal{P}_n \times \mathcal{P}_n \to \mathcal{P}_n, \quad G(X,Y) = X \#_{\frac{1}{2}} Y = Y(Y^{-1}X)^{\frac{1}{2}}$$

We analyze the condition number of the function $G(X,Y)$ w.r.t. X given by

$$\kappa = \|G(X+E,Y) - G(X,Y)\|_F / \|E\|_F$$

Set $F(X) = G(X,Y)$ and recall that

$$G(X+E,Y) - G(X,Y) = L_F(X,E) + o(\|E\|_F)$$

so that

$$\kappa = \|K_F\|_2 := \max_{\|u\|_2=1} \|K_F u\|_2 = \rho(K_F^T K_F)^{\frac{1}{2}}$$

It is sufficient to evaluate K_F and perform a spectral analysis of $K_F^T K_F$

Since $F(X) = Y(Y^{-1}X)^{\frac{1}{2}}$, from the properties of the Fréchet derivative one has

$$K_F = (I \otimes Y)(I \otimes Z^T + Z \otimes I)^{-1}(I \otimes Y^{-1}), \qquad Z = (XY^{-1})^{\frac{1}{2}}$$

that is

$$K_F = (I \otimes Z + Z \otimes I)^{-1}$$

Since $Z = Y^{\frac{1}{2}} \left[Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}} \right]^{\frac{1}{2}} Y^{-\frac{1}{2}}$ and the central term is symmetric, one finds that

$$\kappa = \|K_F\|_2 \le \frac{\mathrm{cond}(Y)}{2 \min \lambda(Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}})}$$

A similar analysis holds for perturbations on Y

Remarks: If X and Y are "close" to each other so that $XY^{-1} \approx I$ then κ_2 is bounded by the condition numbers of X and Y

If X and Y are "far", then the spectrum of the matrix $Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}}$ plays a role.