The Geometric Matrix Mean: an Adaptation for Structured Matrices

D.A. Bini, B. Iannazzo, B. Jeuris, R. Vandebril
University of Pisa, University of Perugia, and KU Leuven

KU LEUVEN

Geometric matrix mean

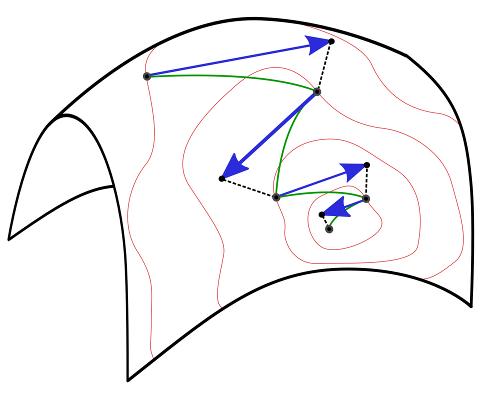
Barycenter for the following settings

- Domain: set of (symmetric) positive definite (PD) matrices $\mathcal{P}_n$;
- Distance measure: based on natural PD metric:

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

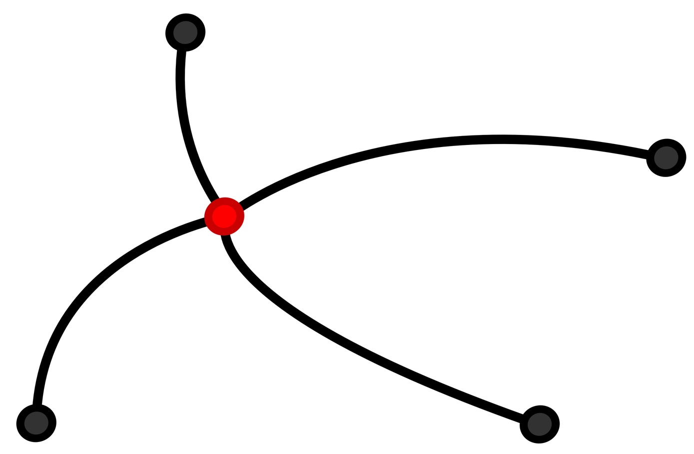
Computed using manifold optimization

- Gradient, Hessian, ... incorporate the Riemannian geometry of the set;
- Iterative steps are taken along the manifold.



Barycenter

Minimizer of the sum of squared distances to given elements A_i



Depends on

- Search space/Domain S: restriction to the desired matrix structure;
- Distance measure d: influences additional properties of the barycenter.

$$\Rightarrow B_S(A_1,\ldots,A_k) = \arg\min_{X\in S} \sum_{i=1}^{k} d^2(X,A_i)$$

Kähler metric mean

Barycenter for the following settings

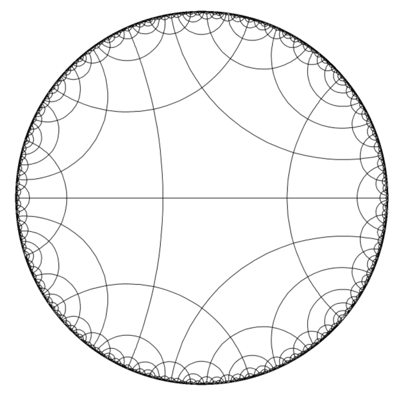
- Domain: set of (Hermitian) PD Toeplitz matrices $\mathcal{T}_n$;
- Distance measure: Kähler metric.

The metric is defined using an application-inspired transformation of the elements of $\mathcal{T}_n$:

$$T \in \mathcal{T}_n \to (P, \mu_1, \ldots, \mu_{n-1}) \in R_*^+ \times D^{n-1},$$

with D the open complex unit disk.

- R_*^+: the geometry of positive numbers;
- D: hyperbolic geometry (Poincaré disk).



Structured matrix adaptations

Why?

- Many applications use a matrix structure which is an intersection of $\mathcal{P}_n$ and an additional (vectorspace) structure;
- The geometric mean generally destroys additional structure.

How?

- Domain restriction: minimize over $S \subset \mathcal{P}_n$ instead of $\mathcal{P}_n$ itself.

Properties?

Unstructured

- *Permutation invariance:* A permutation of the elements A_i does not change the result;
- *Joint homogeneity:* for $\alpha_1,\ldots,\alpha_k > 0$,

$$A_i \to \alpha_i A_i$$
$$B_{\mathcal{P}_n} \to (\alpha_1 \cdots \alpha_k)^{1/k} B_{\mathcal{P}_n};$$

- *Inversion invariance:*

$$B_{\mathcal{P}_n}(A_1^{-1},\ldots,A_k^{-1}) = B_{\mathcal{P}_n}(A_1,\ldots,A_k)^{-1}.$$

Structured

- *Permutation invariance:* Still holds because of the definition of the cost function;
- *Joint homogeneity:* Remains valid if S is the intersection of $\mathcal{P}_n$ with a vectorspace;
- *Inversion invariance:* Define S^{-1} as the set of inverses of the nonsingular elements in S, then

$$B_{S^{-1}}(A_1^{-1},\ldots,A_k^{-1}) = B_S(A_1,\ldots,A_k)^{-1}.$$

Computation and properties

Transformation to product space $R_*^+ \times D^{n-1}$, optimization can be decoupled into n separate averaging operations.

- Barycenter of the coefficients P_i in R_*^+: exactly the (scalar) geometric mean $(P_1 \cdots P_k)^{1/k}$;
- Barycenter of the coefficients $\mu_{j,i}$ (separate for each j) in D:
  - Real matrices: $\mathcal{C}\left((\mathcal{C}(\mu_{j,1}) \cdots \mathcal{C}(\mu_{j,k}))^{1/k} \right)$, where $\mathcal{C}$ is the Cayley transform, $\mathcal{C}(z) = (1-z)/(1+z)$;
  - Complex matrices: no explicit formula, but fast scalar optimization algorithms can be constructed.

Properties?

- Barycenter properties (permutation invariance, repetition invariance, ...) hold by construction;
- Properties of the geometric mean which translate well through the transformation are preserved, e.g., joint homogeneity:

$$T \to \alpha T$$
$$(P, \mu_1, \ldots, \mu_{n-1}) \to (\alpha P, \mu_1, \ldots, \mu_{n-1})$$

- Many other geometric mean properties do not hold or are not well-defined because of the transformation, showing the clear distinction between the means;
- Useful because of its decoupling property and its close relation to the application via the transformation.

Two preconditioners

Suppose S is the intersection of a vectorspace $\mathcal{V}$ and $\mathcal{P}_n$, then parametrize $\mathcal{V}$ using $\sigma : \mathbb{R}^q \to \mathbb{R}^{n\times n}$, $q = \dim(\mathcal{V})$, such that $\mathrm{vec}(\sigma(t)) = Ut$. Denote by $\Gamma(X)$ the Euclidean gradient of the unstructured barycenter cost function $f_{B_{\mathcal{P}_n}}$ at $X \in \mathcal{P}_n$.
From this, the gradient for the structured barycenter cost function f_{B_S} can be obtained in both the Euclidean and the Riemannian geometry.

- Euclidean geometry:

$$\left(U^T U\right)^{-1} U^T \,\mathrm{vec}(\Gamma(\sigma(t)));$$

- Riemannian geometry:

$$\left(U^T \left(\sigma(t)^{-1} \otimes \sigma(t)^{-1}\right) U\right)^{-1} U^T \,\mathrm{vec}(\Gamma(\sigma(t))).$$

The pre-multiplied terms in the expressions can be interpreted as preconditioners to the Euclidean gradient. Using the geometric information of $\mathcal{P}_n$, a more effective gradient and more efficient algorithms are obtained.

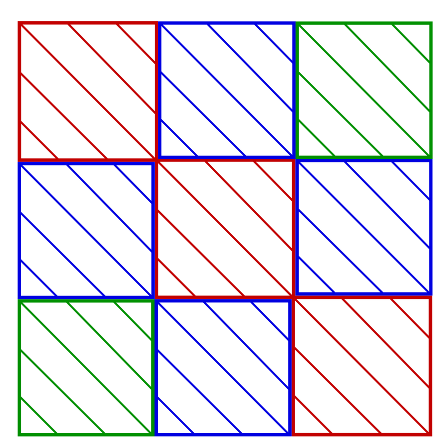
TBBT matrices

Generalize the theory of the Kähler mean to the set of PD Toeplitz-block Block-Toeplitz (TBBT) matrices $\mathcal{B}_{p,n}$ (an $n \times n$ block structure of $p \times p$ matrix blocks):

- Adaptation of the transformation to the block structure:

$$T \in \mathcal{B}_{p,n} \to (P, U_1, \ldots, U_{n-1}) \in \mathcal{P}_p \times \mathcal{D}_p^{n-1},$$

where $\mathcal{D}_p$ is a generalization of D;
- The computational advantage of decoupling the separate coefficients can again be exploited to average large matrices.



Contact

Email: `ben.jeuris@cs.kuleuven.be`; Tel: +32 16 327087.