Averaging positive definite matrices

E. Massart, J. Hendrickx, P.-A. Absil

ICTEAM, UCL, Belgium

March 2016
The set of Positive Definite (PD) matrices

\[ P_n = \{ A \in \mathbb{R}^{n \times n} \mid A = A^T \text{ and } \lambda_i > 0 \ \forall i = 1, \ldots, n \} \]
The set of Positive Definite (PD) matrices

\[ \mathbb{P}_n = \{ A \in \mathbb{R}^{n\times n} \mid A = A^T \text{ and } \lambda_i > 0 \ \forall i = 1, \ldots, n \} \]

For \( n = 2 \):

\[ A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{P}_2 \]

\[ \begin{cases} \tau(A) = a + c > 0 \\ \det(A) = ac - b^2 > 0 \end{cases} \]
The set of Positive Definite (PD) matrices

\[ \mathbb{P}_n = \{ A \in \mathbb{R}^{n \times n} \mid A = A^T \text{ and } \lambda_i > 0 \ \forall i = 1, \ldots, n \} \]

For \( n = 2 \):

\[ \delta(A, B) = \| A - B \|_F \]

\[ \downarrow \]

\[ \delta(A, B) = \| \log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}) \|_F \]
Properties expected from a mean on $\mathbb{P}_n$

ALM list (10 criteria):

- Joint homogeneity: $M(\alpha A_1, A_2) = \sqrt{\alpha} M(A_1, A_2)$
- Invariance under permutation: $M(A_1, A_2) = M(A_2, A_1)$
- Invariance under inversion: $M(A_1^{-1}, A_2^{-1}) = M(A_1, A_2)^{-1}$
- ...

$\Rightarrow$ Unique 2-variable ALM-compliant mean
2-variable ALM-compliant mean

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

⇒ Unique 2-variable ALM-compliant mean
Multivariate ALM-compliant mean

Several possibilities, the most widespread = Karcher mean:

\[ K(A_1, \ldots, A_N) = \arg\min_{X \in \mathbb{P}_n} \sum_{i=1}^{N} \delta^2(X, A_i) \]

where

\[ \delta(A, B) = ||\log(A^{-\frac{1}{2}} BA^{-\frac{1}{2}})||_F \]
Several possibilities, the most widespread = Karcher mean:

\[ K(A_1,\ldots,A_N) = \arg\min_{X \in \mathbb{P}_n} \sum_{i=1}^{N} \delta^2(X, A_i) \]

Optimization problem on manifold

⇒ solved using iterative algorithm, e.g., Steepest Descent (SD)
⇒ expensive to compute

⇒ Approximate it instead!
Our goal

Produce a multivariate PD mean, which should be:

1. cheap
2. as close to the Karcher mean as possible
3. as ALM-compliant as possible

⇒ Tradeoff between (1) and \((2), (3)\)

Tool : start from the 2-variable ALM-compliant PD mean, for which there exists a closed-form expression...
Progressive merging approach

Progressive merging approach combined with shuffling

Performances
Progressive merging approach

Progressive merging approach combined with shuffling

Performances
Progressive Merging algorithm

Idea: make successively one step towards each input matrix. Example for $N = 5$: 
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Trick to compute the multivariate arithmetic mean in $\mathbb{R}^2$
Application of this trick on the sphere
Application of this trick on the sphere
Application of this trick on the sphere
Application of this trick on the sphere
Application of this trick on the sphere
Application of this trick on the sphere
Application of this trick on the sphere

⇒ doesn’t work anymore

⇒ last points under-emphasized
Application of this trick on $\mathbb{P}_2$
Application of this trick on $\mathbb{P}_2$
Application of this trick on $\mathbb{P}_2$
Application of this trick on $\mathbb{P}_2$
Application of this trick on $\mathbb{P}_2$

$\Rightarrow$ doesn’t work anymore

$\Rightarrow$ last points over-emphasized
The result depends on the order of the inputs

\[ \delta_{rel}(i) = \frac{\delta(M, A_i)}{\delta(K, A_i)} \] with \[ \delta(A, B) = \| \log(A^{-\frac{1}{2}} BA^{-\frac{1}{2}}) \|_F \]
Progressive merging approach

Progressive merging approach combined with shuffling

Performances
Progressive Merging algorithm

Idea: make successively one step towards each input matrix.

Example for $N = 5$:

Key observation: result $X_{N-1}$ depends on the initial order of the matrices.
Progressive Merging algorithm with shuffling

- Choose a set of \( k \) permutations
- Run the PM algorithm with each permutation
  \( \Rightarrow \) estimates \( X_{N-1,1}, \ldots, X_{N-1,k} \)
- Average estimates \( X_{N-1,1}, \ldots, X_{N-1,k} \).

Example: \( N = 5 \)

\[
\begin{align*}
A_1 & \quad X_{4,1} \\
A_2 & \quad X_{4,2} \\
A_3 & \quad X_{4,3} \\
A_4 & \\
A_5 &
\end{align*}
\]
Choice of the permutations: in-shuffling
Choice of the permutations: in-shuffling

![Diagram of card arrangement](image)
Choice of the permutations: in-shuffling
Choice of the permutations: in-shuffling
Choice of the mean for the last averaging step

- Arithmetic mean
- PM algorithm
- Crude mean

\[ M_{Cr}(A_1, \ldots, A_N) = M(M_A(A_1, \ldots, A_N), M_H(A_1, \ldots, A_N)) \]

with

\[ M_A(A_1, \ldots, A_N) = \frac{1}{N} \sum_{i=1}^{N} A_i \quad \text{and} \quad M_H(A_1, \ldots, A_N) = \left( \frac{1}{N} \sum_{i=1}^{N} A_i^{-1} \right)^{-1} \]
Progressive merging approach

Progressive merging approach combined with shuffling

Performances
Performances: ALM-compliant means?

7-9 of the 10 ALM properties fulfilled, depending on the variant

Lost properties:

- invariance under permutation:
  \[ M(A_1, \ldots, A_N) = M(A_{p(1)}, \ldots, A_{p(N)}) \]

- invariance under inversion:
  \[ M(A_1^{-1}, \ldots, A_N^{-1}) = M(A_1, \ldots, A_N)^{-1} \]

- determinant equality:
  \[ \det(M(A_1, \ldots, A_N)) = (\det A_1 \cdot \ldots \cdot \det A_N)^{1/N} \]
Performances: estimate of $K$?

Remember:

$$K(A_1, \ldots, A_N) = \arg\min_{X \in \mathcal{P}_n} \sum_{i=1}^{N} \delta^2(X, A_i)$$

Proximity of $M$ to $K$:

$$E_{rel} = \frac{\delta(M, K)}{\frac{1}{N} \sum_{i=1}^{N} \delta(A_i, K)}$$
Performances: estimate of $K$?

$$E_{rel} = \frac{\delta(M, K)}{\sum_{i=1}^{N} \delta(A_i, K)}$$

$N = 25, n = 20$

- IS−PM−Ar
- IS−PM−Cr
- IS−PM−PM
- Cheap
- SD−Auto−10
- SD−Auto−15
Performances: estimate of $K$?

\[ E_{rel} = \frac{\delta(M, K)}{N} \sum_{i=1}^{N} \delta(A_i, K) \]

\[ N = 10, n = 50 \]

![Graph showing error $E_{rel}$ vs. CPU time for different algorithms.](image)
Performances: estimate of $K$?

$$E_{rel} = \frac{1}{N} \sum_{i=1}^{N} \delta(A_i, K)$$

$N = 10, n = 20, c = 10^4$
Conclusion

PD matrices

Averaging algorithm
Performances:

- 7-9 ALM criteria satisfied
- Non-dominated by state-of-the-art
- Advantage \( \uparrow \) for big / ill-conditioned matrices