A NOTE ON COMPUTING MATRIX GEOMETRIC MEANS

DARIO A. BINI AND BRUNO IANNAZZO

Abstract. A new definition is introduced for the matrix geometric mean of a set of k positive definite $n \times n$ matrices together with an iterative method for its computation. The iterative method is locally convergent with cubic convergence and requires $O(n^3k^2)$ arithmetic operations per step whereas the methods based on the symmetrization technique of Ando, Li and Mathias [*Linear Algebra Appl.*, 385 (2004), pp. 305–334] have complexity $O(n^3k!2^k)$. The new mean is obtained from the properties of the centroid of a triangle rephrased in terms of geodesics in a suitable Riemannian geometry on the set of positive definite matrices. It satisfies most part of the 10 properties stated by Ando, Li and Mathias; a counterexample shows that monotonicity is not fulfilled.

Key words. Matrix geometric mean, matrix function, Riemannian centroid, geodesic.

AMS subject classifications. 65F30, 15A15

1. Introduction.

In certain physical applications one has to represent through a *single average* matrix G the results of several experiments made up by a set of *many* positive definite $n \times n$ matrices $A_1, A_2, \ldots, A_k$. The arithmetic mean $\frac{1}{k}\sum_{i=1}^{k} A_i$ is not well-suited to represent the needed quantity since for physical reasons one of the required properties is that the average of $A_1^{-1}, \ldots, A_k^{-1}$ as well, must coincide with G^{-1} (see [10, 11]). Among the classical means of positive real numbers $a_1, \ldots, a_k$, this property is satisfied by the geometric mean $g = (\prod_{i=1}^{k} a_i)^{1/k}$.

1.1. Means of two matrices.

There is large agreement on what is the *right* definition of the geometric mean $G = A\#B$ of two positive definite matrices A and B, namely $G := A(A^{-1}B)^{1/2}$ (see [2, Chapter 4] for a concise treatment of the topic), where given a square matrix M having no nonpositive real eigenvalues, $M^{1/2}$ denotes the unique solution of the equation $X^2 = M$ whose eigenvalues lie in the right half plane. That definition was given in the seventies by Pusz and Woronowicz [13], but there are many other equivalent characterizations, the most notable of which has been provided recently in [8, 10] and is related to the Riemannian geometry obtained endowing the set $\mathbb{P}_n$ of positive definite matrices of size n with the scalar product $g(M,N) = \mathrm{tr}(A^{-1}MA^{-1}N)$ in the tangent space $T_A\mathbb{P}_n$ at A.

The link to the geometric mean is through geodesics, in fact it can be proved that there exists a unique geodesic joining two positive definite matrices A and B whose parameterization is

$$A\#_t B := A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}, \quad t \in [0,1],$$

and $A\#B = A\#_{1/2}B$ is its midpoint.

Here and hereafter, we will use the symbols $\log(A)$, $\exp(A)$, $A^t := \exp(t\log(A))$ to denote the usual functions of a square matrix. If A is diagonalizable, namely if there exists an invertible matrix M and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ such that $A = MDM^{-1}$, then $f(A) := Mf(D)M^{-1}$, where $f(D) := \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. The above definition of $A^{1/2}$ coincides with this one (see [7]).

We briefly recall some properties of the matrix exponential and logarithm which will be useful in the following, the proofs can be found in [7].

THEOREM 1.1. *The following properties hold:*

1. $\log(\alpha I) = \log(\alpha)I$ *for any positive constant* α*, in particular* $\log I = 0$*;*
2. *if* M *and* N *commute and have real positive eigenvalues then* $\log(MN) = \log(M) + \log(N)$*;*

3. *for any invertible matrix* M, $f(MAM^{-1}) = Mf(A)M^{-1}$, *in particular* $\exp(MAM^{-1}) = M\exp(A)M^{-1}$ *and* $\log(MAM^{-1}) = M\log(A)M^{-1}$;
4. $\det(\exp(A+B)) = \det(\exp(A))\det(\exp(B))$;
5. $\exp(-X) = \exp(X)^{-1}$.

In the setting of matrix functions, it is often easy to prove general results in an elegant way. For example the following result holds.

THEOREM 1.2. *Let* A *and* B *be positive definite matrices and let* f *be a function defined on the eigenvalues of* $A^{-1}B$*, then* $Af(A^{-1}B) = A^{1/2}f(A^{-1/2}BA^{-1/2})A^{1/2}$

Proof. First, observe that the matrix $A^{-1}B$ is diagonalizable. From the above definition of matrix function it follows that for any diagonalizable matrix A one has $f(N^{-1}AN) = N^{-1}f(A)N$, thus

$$Af(A^{-1}B) = Af(A^{-1/2}A^{-1/2}BA^{-1/2}A^{1/2}) = A^{1/2}f(A^{-1/2}BA^{-1/2})A^{1/2}.$$

□

Theorem 1.2 explains why $A\#_{1/2}B = A(A^{-1}B)^{1/2}$.

1.2. Means of more than 2 matrices. The generalization of the definition of geometric mean to more than two positive definite matrices seems to be considerably more difficult.

Ando, Li and Mathias [1] proposed a list of ten properties (the ALM properties) that a "good" geometric mean $G(\cdot)$ of k matrices should satisfy. Here, for simplicity we report this list in the case $k=3$ where we write $A > B$ if $A-B$ is positive definite and $A \geqslant B$ if $A-B$ is positive semi-definite.

P1 *Consistency with scalars.* If A, B, C commute then $G(A,B,C) = (ABC)^{1/3}$.
P2 *Joint homogeneity.* $G(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{1/3}G(A,B,C)$, for $\alpha,\beta,\gamma > 0$.
P3 *Permutation invariance.* For any permutation $\pi(A,B,C)$ of A, B, C, it holds that $G(A,B,C) = G(\pi(A,B,C))$.
P4 *Monotonicity.* If $A \geqslant A'$, $B \geqslant B'$, $C \geqslant C'$, then $G(A,B,C) \geqslant G(A',B',C')$.
P5 *Continuity from above.* If A_n, B_n, C_n are monotonic decreasing sequences converging to A, B, C, respectively, then $G(A_n,B_n,C_n)$ converges to $G(A,B,C)$.
P6 *Congruence invariance.* For any nonsingular S, it holds that $S^*G(A,B,C)S = G(S^*AS, S^*BS, S^*CS)$.
P7 *Joint concavity.* If $A = \lambda A_1 + (1-\lambda)A_2$, $B = \lambda B_1 + (1-\lambda)B_2$, $C = \lambda C_1 + (1-\lambda)C_2$, then $G(A,B,C) \geqslant \lambda G(A_1,B_1,C_1) + (1-\lambda)G(A_2,B_2,C_2)$ for $0 < \lambda < 1$.
P8 *Self-duality.* $G(A,B,C)^{-1} = G(A^{-1},B^{-1},C^{-1})$.
P9 *Determinant identity.* $\det G(A,B,C) = (\det A \det B \det C)^{1/3}$.
P10 *Arithmetic–geometric–harmonic mean inequality.*

$$\frac{A+B+C}{3} \geqslant G(A,B,C) \geqslant \left(\frac{A^{-1}+B^{-1}+C^{-1}}{3}\right)^{-1}.$$

It has been proved in [1] that P5 and P10 are consequences of the others. Notice that all these properties can be easily generalized to the mean of any number of matrices.

For $k=2$ this list uniquely defines $G = A\#B = A(A^{-1}B)^{1/2}$. In the case $k>2$ there are infinitely many means satisfying the ALM properties.

In [1] Ando, Li and Mathias propose a numerical scheme for computing a mean of k matrices which satisfies the ALM properties. For $k=3$ they show that the sequences

$$\begin{aligned} A^{(\nu+1)} &= B^{(\nu)}\#C^{(\nu)}, \\ B^{(\nu+1)} &= C^{(\nu)}\#A^{(\nu)}, \qquad \nu = 0,1,\dots, \\ C^{(\nu+1)} &= A^{(\nu)}\#B^{(\nu)}, \end{aligned} \tag{1.1}$$

obtained with $A^{(0)} = A$, $B^{(0)} = B$, $C^{(0)} = C$, converge to a common limit G satisfying the ALM properties. For a set $A_1, \ldots, A_k$ of $k > 3$ matrices these sequences can be defined as

$$A_i^{(\nu+1)} = G_{k-1}(A_1^{(\nu)}, \ldots, A_{i-1}^{(\nu)}, A_{i+1}^{(\nu)}, \ldots, A_k^{(\nu)}), \quad i = 1, \ldots, k, \tag{1.2}$$

where G_{k-1} denotes the mean of $k-1$ matrices recursively defined by means of the same relations. Indeed, also these sequences converge to a common limit which satisfies the ALM properties. We refer to this limit as the ALM mean. It is proved that convergence is linear with convergence factor 1/2. It is easy to find out that the computational cost of this scheme for general k is $O(n^3 k! \prod_{i=3}^{k} p_i)$ where n is the matrix size and p_i is the number of iterations needed in the computation of the means of i matrices.

A substantial improvement has been achieved in [6] relying on these observations: in the sequences (1.1) converging to the ALM mean, $A^{(\nu+1)}$ is the midpoint of the geodesics joining the matrix $B^{(\nu)}$ with $C^{(\nu)}$; in the Euclidean geometry the limit of this sequence is the centroid of the triangle ABC; the centroid is also located in the median which connects A with the midpoint of the edge BC at distance 2/3 from A, that is $A\#_{2/3}(B\#_{1/2}C)$; the three medians have the centroid as common point. Due to the negative curvature of $\mathbb{P}_n$ the three points $A\#_{2/3}(B\#_{1/2}C)$, $B\#_{2/3}(C\#_{1/2}A)$, $C\#_{2/3}(A\#_{1/2}B)$ do not coincide, but are closer to each other than the original matrices.

Therefore the iteration is given by

$$\begin{array}{ll} A^{(\nu+1)} = A^{(\nu)}\#_{2/3}(B^{(\nu)}\#C^{(\nu)}), & \\ B^{(\nu+1)} = B^{(\nu)}\#_{2/3}(C^{(\nu)}\#A^{(\nu)}), & \qquad \nu = 0, 1, \ldots \\ C^{(\nu+1)} = C^{(\nu)}\#_{2/3}(A^{(\nu)}\#B^{(\nu)}). & \end{array}$$

It is proved that the three matrix sequences have a common limit, different from the ALM mean, which satisfies the ALM properties, and the convergence is cubic. We will refer to this mean as the BMP mean. The same iteration can be generalized to the case of $k > 3$ matrices.

The computational cost is the same as the ALM scheme, however, the number p_i of iterations is reduced by relying on a numerical scheme having cubic convergence so the acceleration in certain applications is dramatic. Unfortunately, the growth of the computational cost with k is still exponential; therefore, for moderate values of k also this iteration is infeasible.

The idea of [1, 6] can be generalized by considering new means obtained by assembling existing ones through a recursive procedure. Unfortunately, it has been proved that no such definition could produce a mean whose computational cost is polynomial with respect to k [12]. In the next section we follow a different direction.

1.3. New results. In this paper, by relying on the geometric interpretation given in terms of geodesics in the Riemannian geometry on the variety $\mathbb{P}_n$, we introduce a new iteration for computing a geometric mean of k matrices with the following features: unlike the known methods, the computation of the mean of k matrices does not require computing the mean of $k-1$ matrices and no recursive process is needed; the convergence speed of the new iteration is cubic; its computational cost is polynomial, namely $O(n^3 k^2 p_k)$, where p_k is the number of iterations needed by the method (typically just a few); for $k = 2$ the limit is $A\#B$, so the proposed mean generalizes the geometric mean of two matrices; the limit of k sequences satisfies the

ALM properties P1–P3, P6, P8 and P9; we provide a counterexample where P4 is not satisfied. The counterexample requires that the matrices be very far from each other; counterexamples where the matrices A_i, $i = 1, \dots, k$ are in a relatively small neighborhood of their mean are not known. We refer to this new mean as the *Cheap Mean*.

The idea on which this iteration is based relies once again on the geometric interpretation of the centroid G of a triangle ABC. In the Euclidean geometry the centroid G satisfies the equations

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

that is, it lies in the geodesic passing through A and tangent in A to the arithmetic mean of the tangent vectors in A of the three geodesics connecting A with B, C and A, respectively. Obviously, the third vector is zero. Similar expressions are obtained starting from B and C, respectively.

In the Riemannian manifold $\mathbb{P}_n$ this procedure gives three different points A', B' and C', and can be viewed as a step of an iterative procedure converging to a possible mean. Observe that the mean of the tangent vectors is done in the tangent space at a point which is Euclidean, where it is natural to choose the arithmetic mean.

This procedure can be easily generalized to $k \geqslant 3$. Given $A_1, \dots, A_k$, it is enough to consider, for each i, the geodesic starting at A_i and whose tangent vector is the arithmetic mean of the k tangent vectors at A_i to the geodesic joining A_i with A_j (where if $i = j$ the vector is 0). Then A'_i will be the point of that geodesic for $t = 1$.

Since the tangent vector at A to the geodesic joining A and B is the symmetric matrix $A \log(A^{-1}B)$, one obtains the following iteration

$$A_i^{(\nu+1)} = A_i^{(\nu)} \exp\left(\frac{1}{k} \sum_{j=1, j\neq i}^{k} \log((A_i^{(\nu)})^{-1} A_j^{(\nu)}) \right), \quad i = 1, \dots, k, \tag{1.3}$$

with $A_i^{(0)} = A_i$, $i = 1, \dots, k$. Observe that, by Theorem 1.2, (1.3) can be equivalently rewritten as

$$A_i^{(\nu+1)} = (A_i^{(\nu)})^{1/2} \exp\left(\frac{1}{k} \sum_{j=1, j\neq i}^{k} \log((A_i^{(\nu)})^{-1/2} A_j^{(\nu)} (A_i^{(\nu)})^{-1/2}) \right) (A_i^{(\nu)})^{1/2}. \tag{1.4}$$

This equation shows that the sequences $\{A_i^{(\nu)}\}_\nu$ are formed by symmetric positive definite matrices.

For $k = 2$ the first step of the iteration yields $A_1^{(1)} = A \exp(\frac{1}{2}(\log(A^{-1}B))) = A\#B$ and similarly $B_1^{(1)} = A\#B$. Thus, in the case of two matrices the iteration yields the geometric mean since the first step.

We prove that if the sequences $\{A_i^{(\nu)}\}_\nu$ converge to the same limit for $i = 1, \dots, k$, then the convergence is cubic. Moreover we give conditions under which convergence occurs. Even though the local convergence condition may appear rather restrictive, from the many numerical experiments that we have performed we never encountered failure of convergence.

We have implemented the computation of the Cheap Mean in the Matrix Mean Toolbox, available for Matlab and Octave [5], and performed some numerical tests.

In particular, we have compared the Cheap mean with the "least square geometric mean" [3], also called "Riemannian geometric mean" [10], or Karcher mean [9], that is, the unique positive definite solution of the matrix equation

$$\sum_{i=1}^{k} X^{1/2} A_i^{-1} X^{1/2} = 0. \tag{1.5}$$

It is known that this mean satisfies all the ALM properties of a geometric mean except perhaps the monotonicity property, for which no counterexample is known so far. By means of numerical experiments we show that the Cheap mean is much faster to compute than the Karcher mean (if for the latter, the algorithms of [9, 10] or a gradient algorithm applied to (1.5) are used). In fact, in all the experiments performed so far, iteration (1.3) converges to the Cheap mean in at most 5 iterations with a relative error of the order of 10^{-15} independently of the condition number, whereas for the Karcher mean, the iterations of [9, 10] do not converge in certain cases and in the other cases require a larger computational cost. The gradient methods require always a larger computational cost.

We wish to point out that the iteration of [9] is given by

$$X^{(\nu+1)} = X^{(\nu)} \exp\left(\frac{1}{k}\sum_{i=1}^{k} \log((X^{(\nu)})^{-1}A_i)\right), \quad X_0 = A_1, \tag{1.6}$$

which is very similar to our iteration (1.3). In fact, each step of (1.3) can be viewed as k first steps of iteration (1.6) with $A_i = A_i^{(\nu)}$, $i = 1, \ldots, k$ and $X_0 = A_i^{(\nu)}$. In [9] the convergence of (1.6) has been proved in the special orthogonal group provided that the matrices A_i are sufficiently close to each other. The numerical tests show that iteration (1.6) does not converge if the matrices A_i are positive definite and not close each other and that when convergence occurs it is linear.

Another comparison that we have performed concerns the definition of geometric mean by

$$G = \exp\left(\frac{1}{k}\sum_{i=1}^{k} \log A_i\right).$$

This mean, referred to as ExpLog mean, is studied in [1], and can be computed with a cost of $O(n^3k)$ ops. However, its properties are poorer than the properties of the Cheap mean. First, the ExpLog mean of two matrices is different from $A\#B$. Second, it is not congruence invariant as shown in the numerical experiments. Third, the ExpLog mean looses the monotonicity property in a very large part of cases. In fact, from a wide numerical experimentation it turns out that even though the matrices A_i are tightly close to each other and have a moderate condition number, this mean fails to be monotone. Whereas, the Cheap mean fails to be monotone only when the matrices are severely ill conditioned and they are not tightly close to each other. Finally, for modeling reasons, any practical definitions of geometric mean should lie in a small neighborhood; from numerical experiments it turns out that the ALM, BMP, and Cheap means form a very tight cluster while the ExpLog mean lies very far from this cluster.

The paper is organized as follows: in Section 2 we prove the local cubic convergence of iteration (1.4), in Section 3 we prove most of the ALM properties for the Cheap mean and provide a counterexample for the monotonicity. In Section 4 we discuss the results of the numerical experiments.

2. Convergence analysis. Let us consider a single step of iteration (1.4) and for notational simplicity we write

$$A_i' = A_i^{1/2}\exp\left(\frac{1}{k}\sum_{j=1}^{k}\log(A_i^{-1/2}A_jA_i^{-1/2})\right)A_i^{1/2}. \tag{2.1}$$

Observe that the condition $i \neq j$ is not needed in (2.1) since the term obtained for $i = j$ is zero.

Let us introduce the following notation

$$\begin{aligned} &A_j^{-1/2}A_iA_j^{-1/2} = I + A_j^{-1/2}(A_i - A_j)A_j^{-1/2} = I + X_{i,j},\\ &X_{i,j} = A_j^{-1/2}E_{i,j}A_j^{-1/2}, \qquad E_{i,j} = A_i - A_j, \end{aligned}$$

so that equation (2.1) can be rewritten as

$$A_i' = A_i^{1/2}\exp\left(\frac{1}{k}\sum_{j=1}^{k}\log(I + X_{i,j})\right)A_i^{1/2}. \tag{2.2}$$

We recall that if $\rho(X) < 1$ then

$$\begin{aligned} &\log(I+X) = X - \tfrac{1}{2}X^2 + \tfrac{1}{3}X^3 - \cdots \doteq X - \tfrac{1}{2}X^2\\ &\exp(W) = I + W + \tfrac{1}{2}W^2 + \tfrac{1}{3!}W^3 + \cdots \doteq I + W + \tfrac{1}{2}W^2 \end{aligned} \tag{2.3}$$

where $\doteq$ denotes equality up to terms of the third order in X or in W.

Here we assume that the matrices are close enough to each other, more precisely, we assume that

$$\|A_j^{-1/2}E_{i,j}A_j^{-1/2}\| \leqslant \varepsilon < 1, \quad i,j = 1,\ldots,k,$$

for $\varepsilon > 0$ small enough, where $\|\cdot\|$ denotes the spectral norm.

Since $\|X_{i,j}\| < 1$, applying (2.3) with $X = X_{i,j}$ in (2.2) yields

$$\begin{aligned} &A_j' \doteq A_j^{1/2}\left[I + Z_j + \tfrac{1}{2}Z_j^2\right]A_j^{1/2},\\ &Z_j \doteq \tfrac{1}{k}\textstyle\sum_{i=1}^{k}(X_{i,j} - \tfrac{1}{2}X_{i,j}^2), \end{aligned}$$

whence

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

Writing down the same equation for A_h' and subtracting the two expressions yields the equation which relates $E_{h,j}' = A_h' - A_j'$ to $E_{i,j}$:

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

Now, since $E_{h,j} + \frac{1}{k}\sum_{i=1}^{k}(E_{i,h} - E_{i,j}) = A_h - A_j + \frac{1}{k}\sum_{i=1}^{k}(A_j - A_h) = 0$, one has

$$\begin{aligned} E'_{h,j} \doteq & -\frac{1}{2k}\sum_{i=1}^{k}(E_{i,h}A_h^{-1}E_{i,h} - E_{i,j}A_j^{-1}E_{i,j}) \\ & + \frac{1}{2k^2}\sum_{r,s=1}^{k}(E_{r,h}A_h^{-1}E_{s,h} - E_{r,j}A_j^{-1}E_{s,j}). \end{aligned} \tag{2.4}$$

This implies that there exists a constant σ, depending only on the matrices $A_1, \ldots, A_k$ such that $\max_{i,j}\|E'_{i,j}\| < \sigma \max_{i,j}\|E_{i,j}\|^2$, so that if the sequence $\{E_{h,j}^{(\nu)}\}_\nu$ converges to zero the convergence is at least quadratic.

We can prove more by observing that

$$\begin{aligned} \frac{1}{k}\sum_{i=1}^{k} E_{i,h}A_h^{-1}E_{i,h} &= \frac{1}{k}\sum_{i=1}^{k} A_iA_h^{-1}A_i - 2M + A_h, \\ \frac{1}{k^2}\sum_{i,j=1}^{k} E_{i,h}A_h^{-1}E_{j,h} &= MA_h^{-1}M - 2M + A_h, \end{aligned}$$

where we set $M = \frac{1}{k}\sum_{i=1}^{k} A_i$. Replacing the latter equations in (2.4) one obtains

$$E'_{h,j} = -\frac{1}{2}\left(M(A_h^{-1} - A_j^{-1})M - \frac{1}{k}\sum_{i=1}^{k} A_i(A_h^{-1} - A_j^{-1})A_i\right).$$

Since $M = \frac{1}{k}\sum_{i=1}^{k} A_i$, formally the latter expression is a quadratic form in $A_1, \ldots, A_k$, namely,

$$E'_{h,j} = \sum_{r,s=1}^{k} \eta_{r,s}A_r(A_h^{-1} - A_j^{-1})A_s, \qquad \frac{1}{2k^2}(kI - ee^T) = (\eta_{r,s}),$$

where $e = (1, \ldots, 1)^T$, that is, the matrix associated with this quadratic form is

$$Q_{h,j} = \frac{1}{2k^2}(kI - ee^T) \otimes (A_h^{-1} - A_j^{-1}),$$

where $\otimes$ denotes the Kronecker product.

Now, the matrix $kI - ee^T$ can be rewritten as

$$kI - ee^T = kUT^{-1}U^T, \quad T = U^TU \tag{2.5}$$

where $U \in \mathbb{R}^{k\times(k-1)}$, $Ue_i = e_i - e_{(i-1) \bmod k}$, for $i = 1, \ldots, k-1$, and $T = U^TU$ is the $(n-1)\times(n-1)$ symmetric tridiagonal matrix having diagonal entries equal to 2 and super-diagonal entries equal to -1. In fact, the two matrices in the left-hand and in the right-hand side of (2.5) have the vector e in their kernels and thus coincide in the linear space orthogonal to e spanned by the columns of U. Therefore

$$Q_{h,j} = \frac{1}{2k}[U \otimes I][T^{-1} \otimes (A_h^{-1} - A_j^{-1})][U^T \otimes I]$$

so that we may write

$$
\begin{aligned}
E'_{h,j} &= \begin{bmatrix} A_1 & \cdots & A_k \end{bmatrix} Q_{h,j} \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix} \\
&= \begin{bmatrix} E_{1,2} & \cdots & E_{k-1,k} \end{bmatrix} \left(\frac{1}{2k} T^{-1} \otimes (A_h^{-1} E_{h,j} A_j^{-1}) \right) \begin{bmatrix} E_{1,2} \\ \vdots \\ E_{k-1,k} \end{bmatrix} \\
&= \frac{1}{2k} \sum_{r=1}^{k-1} \sum_{s=1}^{k-1} \gamma_{r,s} E_{r,r+1} A_h^{-1} E_{h,j} A_j^{-1} E_{s,s+1},
\end{aligned}
$$

with $T^{-1} = (\gamma_{r,s})$. Denoting $\gamma = \gamma(k) = \frac{1}{2k} \sum_{r,s} \gamma_{r,s}$ one has

$$\|E'_{h,j}\| \leqslant \|E_{h,j}\| \gamma(k) \max_r \|E_{r,r+1}\|^2 \|A_h^{-1}\| \cdot \|A_j^{-1}\|.$$

Therefore,

$$\max_{i,j} \|E'_{i,j}\| \leqslant \gamma(k) \max_{i,j} \|E_{i,j}\|^3 \cdot \max_j \|A_j^{-1}\|^2. \tag{2.6}$$

We synthesize the above discussion in the following result, where we give also a condition such that iteration (1.3) converges.

THEOREM 2.1. *If the sequences $A_i^{(\nu)}$ generated by* (1.3) *have a common limit G, then there exists a constant γ such that $\|A_i^{(\nu+1)} - A_j^{(\nu+1)}\| \leqslant \gamma \|A_i^{(\nu)} - A_j^{(\nu)}\|^3$, for any $i,j = 1,\dots,k$, i.e., convergence has order at least 3. If $\max_j \|A_j^{-1}\| \cdot \max_{i,j} \|A_i - A_j\| < \varepsilon$ for $i,j = 1,\dots,k$, where $0 < \varepsilon < 1/3$ then $\max_j \|(A_j^{(\nu)})^{-1}\| \cdot \max_{i,j} \|A_i^{(\nu)} - A_j^{(\nu)}\| < \varepsilon$ for $i,j = 1,\dots,k$, for any ν, moreover $\max_{i,j} \|A_i^{(\nu)} - A_j^{(\nu)}\| \leqslant (2\varepsilon/(1-\varepsilon))^\nu$, and the sequences $A_i^{(\nu)}$ converge to the same limit G.*

Proof. The first part of the theorem follows from (2.6). Concerning the second part, denote $\delta = \max_{i,j} \|A_i - A_j\|$, $\delta' = \max_{i,j} \|A'_i - A'_j\|$, $f = \max_i \|A_i^{-1}\|$, $f' = \max_i \|A_i'^{-1}\|$, and observe that $\|A_j^{-1}(A_i - A_j)\| \leqslant \delta f$. Let us prove that if $\delta f < \varepsilon$ with ε sufficiently small, then also $\delta' f' \leqslant \varepsilon$ so that $\|A_j'^{-1}(A'_i - A'_j)\| \leqslant \varepsilon$ as well. From (2.4) one finds that

$$\delta' \leqslant 2\delta^2 \max_i \|A_i^{-1}\| = 2\delta^2 f. \tag{2.7}$$

Now we provide an upper bound to f' by proving that

$$f' \leqslant f/(1 - \delta f). \tag{2.8}$$

We rely on the following inequalities which derive directly from the definition of the matrix functions exp and log by taking the norms of both sides of (2.3):

$$
\begin{aligned}
&\|\exp(X)\| \leqslant \exp(\|X\|) \\
&\|\log(I + X)\| \leqslant -\log(1 - \|X\|), \quad \text{if } \|X\| < 1.
\end{aligned} \tag{2.9}
$$

We note

$$\left\|A_i'^{-1}\right\| \leqslant \left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k} \log(A_i^{-1}A_j)\right)\right\| \cdot \|A_i^{-1}\| \leqslant \left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k} \log(A_i^{-1}A_j)\right)\right\| f. \tag{2.10}$$

By using (2.9) one finds that

$$\begin{aligned}\left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log(A_i^{-1}A_j)\right)\right\| &\leqslant \exp\left(\left\|\frac{1}{k}\sum_{j=1}^{k}\log(A_i^{-1}A_j)\right\|\right)\\ &\leqslant \exp\left(\frac{1}{k}\sum_{j=1}^{k}\left\|\log(A_i^{-1}A_j)\right\|\right)\\ &= \exp\left(\frac{1}{k}\sum_{j=1}^{k}\left\|\log(I+A_i^{-1}E_{j,i})\right\|\right)\end{aligned}$$

Now, since $\|A_i^{-1}E_{j,i}\| \leqslant \delta f \leqslant \varepsilon < 1$ we may apply (2.9) and get

$$\begin{aligned}\left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log(A_i^{-1}A_j)\right)\right\| &\leqslant \exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log(1-\|A_i^{-1}E_{j,i}\|\right)\\ &= \left(\prod_{j=1}^{k}(1-\|A_i^{-1}E_{j,i}\|)\right)^{-1/k}\\ &\leqslant (1-\delta f)^{-1}\end{aligned}$$

which in the view of (2.10) yields (2.8). Now we are ready to prove that if the condition $\delta f \leqslant \varepsilon$ is satisfied then $\delta' f' \leqslant \varepsilon$ as well. Combining (2.7) and (2.8) yields

$$\delta' f' \leqslant (\delta f)^2 \frac{2}{1-\delta f}.$$

Clearly, if $\varepsilon < 1/3$ then $\delta' f' < \varepsilon$ and from (2.7) one deduces that

$$\delta' \leqslant \frac{2}{3}\delta.$$

An inductive process completes the convergence proof. □

Proving global convergence is still an open problem. From the many numerical experiments that we have performed we have always observed convergence. It is interesting to point out that if the matrices A_i pairwise commute then convergence occurs in just one step for any k-tuple of positive definite matrices $A_1, \ldots, A_k$.

3. The ALM properties. A large number of the ALM properties are satisfied by the Cheap mean. In this section we give a formal proof for the properties P1, P2, P3, P6, P8, and P9, while for P4 we provide a counterexample which shows that monotonicity is not fulfilled by our mean. The proof of validity of P5, P7 and P10 is usually performed relying on P4. We have no counterexample for P5, P7 and P10.

For the sake of notational simplicity we provide the proofs in the case $k = 3$. The generalization to any k is straightforward. We show that starting with $A_0 = A$, $B_0 = B$ and $C_0 = C$, properties P1, P2, P6, P8 and P9, are held by A_1 itself (and B_1 and C_1). This fact can be used in an induction argument, proving that the same properties hold for A_k, B_k and C_k, for each $k > 0$ and thus for the limit.

P1 Consistency with scalars. If A, B, C commute, then

$$A_1 = A\exp\left(\frac{1}{3}(\log(A^{-1}B)+\log(A^{-1}C))\right) = A\exp\left(\frac{1}{3}(\log(A^{-1}BA^{-1}C)\right) = A(A^{-2}BC)^{1/3} = (ABC)^{1/3}.$$

where we have used Property 2 of Theorem 1.1. The same holds for B_1 and C_1.

P2 Joint homogeneity. If $\widehat{A}=\alpha A$, $\widehat{B}=\beta B$ and $\widehat{C}=\gamma C$, with $\alpha,\beta,\gamma>0$, then

$$\begin{aligned}\widehat{A}_1 =&\alpha A\exp\left(\frac{1}{3}\left(\log(A^{-1}B\frac{\beta}{\alpha})+\log(A^{-1}C\frac{\gamma}{\alpha})\right)\right)\\ =&\alpha A\exp\left(\frac{1}{3}\left(\log(A^{-1}B)+\log(A^{-1}C)+\log(\frac{\beta\gamma}{\alpha^2}I)\right)\right)\\ =&\alpha A\exp\left(\frac{1}{3}(\log(A^{-1}B)+\log(A^{-1}C))\right)\exp\log((\frac{\beta\gamma}{\alpha^2})^{1/3}) = (\alpha\beta\gamma)^{1/3}A_1,\end{aligned}$$

where we have used Properties 1 and 2 of Theorem 1.1. The same holds for B_1 and C_1.

P3 Permutation invariance. It follows immediately from the definition.

P4 Monotonicity. This property is not satisfied in general as it is shown by the following numerical counterexample.
Let

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

For $\varepsilon=0.0001$ and $0<h\leqslant 3$ it holds that $\widetilde{A}(h)\geqslant A$ and the matrix $G(\widetilde{A},B,C)-G(A,B,C)$ has a negative eigenvalue. For instance, for $h=1$ the eigenvalues are `-2.4131e-3, 2.2853e-2, 1.0826e-1`.

P6 Congruence invariance. Observe that starting from $\widehat{A}=S^*AS$, $\widehat{B}=S^*BS$, $\widehat{C}=S^*CS$ one has

$$\begin{aligned}\widehat{A}_1 =&\widehat{A}\exp\left(\frac{1}{3}(\log(\widehat{A}^{-1}\widehat{B})+\log(\widehat{A}^{-1}\widehat{C}))\right)\\ =&S^*AS\exp\left(\frac{1}{3}(\log(S^{-1}A^{-1}S^{-*}S^*BS)+\log(S^{-1}A^{-1}S^{-*}S^*BS)\right)\\ =&S^*ASS^{-1}\exp\left(\frac{1}{3}(\log(A^{-1}B)+\log(A^{-1}C))\right)S = S^*A_1S,\end{aligned}$$

where Property 3 of Theorem 1.1 has been used. The same holds for $\widehat{B}_1$ and $\widehat{C}_1$.

P8 Self duality. Observe that

$$\begin{aligned}A_1^{-1} =& \exp(-\frac{1}{3}(\log(A^{-1}B)+\log(A^{-1}C)))A^{-1}\\ =& \exp(\log(B^{-1}A)^{1/3}+\log(C^{-1}A)^{1/3})A^{-1}\\ =& A^{-1}\exp(\log(AB^{-1})^{1/3}+\log(AC^{-1})^{1/3}) = \widehat{A}_1,\end{aligned}$$

where $\widehat{A}_1$ is obtained from $\widehat{A}=A^{-1}$, $\widehat{B}=B^{-1}$ and $\widehat{C}=C^{-1}$, thus the self-duality holds for A_1. The same holds for B_1 and C_1.

P9 Determinant identity. The identity follows from $\det(e^{A+B}) = \det(e^A e^B)$, in fact for A, B and C

$$\begin{aligned}\det A_1 &= \det A \det(\exp(\log(A^{-1}B)^{1/3})) \det(\exp(\log(A^{-1}C)^{1/3})) \\ &= \det A \det(A^{-1}B)^{1/3} \det(A^{-1}C)^{1/3} = (\det A \det B \det C)^{1/3},\end{aligned}$$

where Property 4 of Theorem 1.1 has been used. The same holds for $\det B_1$ and $\det C_1$.

Observe that in the counterexample concerning monotonicity the matrices A, B and C are quite far from each other and do not satisfy the convergence conditions of Theorem 2.1.

We do not have any counterexample to the monotonicity where the matrices A_i satisfy the convergence conditions of Theorem 2.1, and we believe that monotonicity is satisfied "locally", i.e., if the set of matrices A_i, $i = 1, \ldots, k$ lie in a neighborhood of their mean.

Observe, moreover, that A_1 verifies properties P1, P2, P6, P8 and P9, thus, it can be viewed as a rough mean.

4. Numerical experiments. We have implemented the computation of the Cheap mean together with other algorithms for matrix means in the Matrix Means Toolbox [5] available for Matlab and Octave. Here we report part of the many numerical experiments that we have performed.

In the first set of tests we compare the execution times of computing the Cheap mean and the mean of [6], in the following BMP mean, which among the ALM means is the fastest available.

The test matrices are generated randomly with different values of their condition numbers according to the following Matlab commands:

```
n = 10; W = rand(n) - rand(n); X = W' * W; X = X - eye(n) * min(eig(X));

X = X/norm(X); X = X + eye(n)/(cnd - 1); X = X/norm(X);
```

so that the parameter `cnd` coincides with the condition number of `X`.

For various values of the condition number `cnd`, for $n = 4$ and $k = 3 : 10$, in Table 4 we report the CPU time required to compute the Cheap mean and the BMP mean together with the Euclidean distance of the two means. A "$*$" denotes a CPU time larger than 10^4 seconds. The number of iterations required to compute the Cheap mean as well as the number of outer iterations in the recursive process to compute the BMP mean has been between 4 and 5.

The exponential growth with k of the complexity of the BMP mean is evident, while the polynomial complexity of the Cheap mean makes the computation possible even for much larger values of k. It is interesting to observe that the Cheap mean and the BMP mean are not so far from each other.

The second bunch of tests compares the Cheap mean with the mean

$$G = \exp\left(\frac{1}{k}\sum_{i=1}^{k}\log(A_i)\right)$$

which, for simplicity we call ExpLog mean, in order to find out the cases where the monotonicity property is not satisfied. To this end, we consider a 3×3 diagonal matrix A_1 with diagonal entries $1, \delta, \delta^2$, for $0 < \delta < 1$ so that $\|A_1\| = 1$ and its condition

	cnd= 1.e2			cnd= 1.e4			cnd= 1.e8		
k	Cheap	BMP	Dist.	Cheap	BMP	Dist.	Cheap	BMP	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 4.1

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

ε	1e-6	1e-5	1e-4	1e-3	1e-2	1e-1	1	1e1
cnd=1e4	0	0	0	0	0	5	34	64
cnd=1e6	0	0	0	7	28	24	58	78
cnd=1e8	0	0	21	45	16	20	71	77
cnd=1e10	0	36	41	24	6	22	77	82
cnd=1e12	39	56	14	5	3	40	85	85

TABLE 4.2

Percentage of cases where the ExpLog mean of A_1, A_2, A_3 fails to be monotone where A_2 and A_3 are chosen in a neighborhood of A_1 of radius ε and A_1 has condition number `cnd`

number is $1/\delta^2$, and define $A_2 = A_1 + \varepsilon U_1$, $A_3 = A_1 + \varepsilon U_2$, where U_1, U_2 are positive definite random matrices with norm 1, generated as follows:

$$\mathtt{W} = \mathtt{rand}(3) - \mathtt{rand}(3);\ \mathtt{W} = \mathtt{W} * \mathtt{W}';\ \mathtt{U} = \mathtt{W}/\mathtt{norm}(\mathtt{W});$$

In this way the matrices A_2, A_3 stay in the sphere of center A_1 and radius ε. We have generated 100 random values and computed the number of cases where the matrix $G(A_1 + 0.01 * A_2, A_2, A_3) - G(A_1, A_2, A_3)$ is not positive definite. Tables 4.2 and 4.3 report these values according to the conditioning of A_1 and to the radius of the neighborhood of A_1. It is evident that the ExpLog mean fails to be monotone even for moderate values of the condition number and for relatively small neighborhoods of A_1, whereas the Cheap mean seems to be more robust.

It is not difficult to construct numerical examples showing that the ExpLog mean is not congruence invariant, for instance if $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ and $S = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$,

$$\begin{aligned} S^* \exp\left(\frac{1}{2}(\log(A) + \log(I))\right) S &= \begin{bmatrix} 2 & 2 \\ 2 & 8 \end{bmatrix}, \\ \exp\left(\frac{1}{2}(\log(S^*AS) + \log(S^*S))\right) &\approx \begin{bmatrix} 3.0 & 5.4 \\ 5.4 & 13.5 \end{bmatrix}. \end{aligned}$$

The last bunch of tests, taken from [4], reports the number of iterations needed

ε	1e-6	1e-5	1e-4	1e-3	1e-2	1e-1	1	1e1
cnd=1e4	0	0	0	0	0	0	0	0
cnd=1e6	0	0	0	0	0	0	1	0
cnd=1e8	0	0	0	0	0	4	33	23
cnd=1e10	0	0	0	0	0	4	33	23
cnd=1e12	0	0	0	1	1	18	70	67

TABLE 4.3

Percentage of cases where the Cheap mean of A_1, A_2, A_3 fails to be monotone where A_2 and A_3 are chosen in a neighborhood of A_1 of radius ε and A_1 has condition number `cnd`

	cond=1.e2			cond=1.e4		
$k\backslash X_0$	I	AM	Cheap	I	AM	Cheap
3	74	26	17	114	89	41
4	66	21	17	82	59	37
5	65	19	16	87	58	35
6	62	20	16	81	54	31
7	61	21	15	83	63	29
8	61	20	15	93	55	29
9	58	19	14	89	50	29
10	56	19	14	94	47	28

TABLE 4.4

Number of iterations needed to approximate the Karcher mean up to the error `1.e-11` *by means of the iteration* (4.1) *(a), starting with the identity matrix, the arithmetic mean*

and the Cheap mean.

for approximating the Karcher mean relying on the iteration

$$\begin{aligned} X_{\nu+1} &= g(X_\nu), \qquad \nu = 0, 1, \dots, \\ g(X) &= X - \vartheta X^{1/2} \sum_{i=1}^{k} \log(X^{1/2} A_i^{-1} X^{1/2}) X^{1/2}, \end{aligned} \tag{4.1}$$

starting with X_0 equal to the identity matrix, the arithmetic mean and the Cheap mean. Here we have choosen the value of ϑ which minimizes the number of iterations. It is interesting to point out that the number of iterations required is much larger than the number of iterations needed to approximate the Cheap mean which in all the treated cases is less than or equal to 5. Moreover, choosing as starting approximation the Cheap mean yields a faster convergence.

We conclude with an example showing the mutual distance of most of the means of interest. We consider the following matrices

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \tag{4.2}$$

and we compute the ALM, BMP, Cheap, ExpLog and the Karcher means of them. Moreover, we compute the arithmetic-harmonic-geometric (AHG) mean, that is the geometric mean of the arithmetic mean and the harmonic mean. The latter does not satisfy most of the ALM properties, but it is easy to compute. In Figure 4 we have plotted the corresponding points in the three dimensional space of 2×2 symmetric matrices. One can observe that the ALM, BMP, Cheap and Karcher means are very

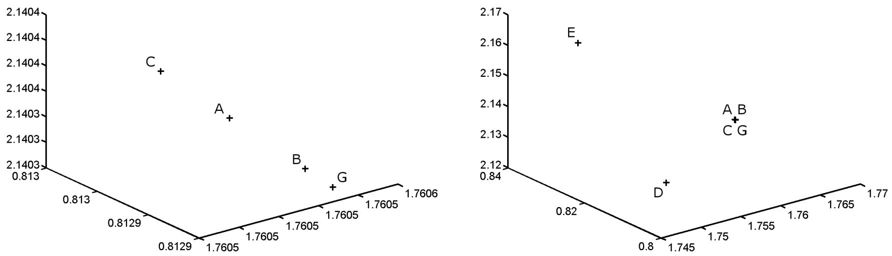

FIG. 4.1. *Localization of the ALM (A), BMP (B), Cheap (C), Arithmetic-Harmonic-Geometric (D), ExpLog (E) and Karcher (G) mean for A, B and C as in* (4.2)

near to each other, while ExpLog and AHG means are relatively far from the others. This is a typical situation that makes the Cheap mean preferable with respect to the ExpLog and AHG means.

5. Conclusion and open questions. We have introduced a new definition of geometric mean which, unlike the ALM means, can be computed with low computational effort even for a large number of input matrices (Cheap mean). We have proved its local convergence and that it fulfills most of the ALM properties.

Several problems remain open. A proof of global convergence of the iteration for the Cheap Mean is missing; concerning the lack of monotonicity, it would be interesting to find out under which conditions on the matrices A_i the Cheap mean keeps monotonicity. For instance, it seems reasonable that if the matrices A_i are close enough to each other then monotonicity should be satisfied.

Acknowledgments The authors wish to thank two anonymous referees for the constructive remarks which enabled to improve the presentation of the paper.

REFERENCES

[1] T. Ando, C.-K. Li, and R. Mathias. Geometric means. *Linear Algebra Appl.*, 385:305–334, 2004.
[2] R. Bhatia. *Positive definite matrices.* Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.
[3] R. Bhatia and J. Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra Appl.*, 413(2-3):594–618, 2006.
[4] D. Bini and B. Iannazzo. On the numerical solution of the matrix equation $\sum_{i=1}^{k} \log(XA_i^{-1}) = 0$. 16-th ILAS Conference, Pisa, June 21–25, 2010, www.dm.unipi.it/~ilas2010/abstracts-im4.pdf.
[5] D. A. Bini and B. Iannazzo. The Matrix Means Toolbox. http://bezout.dm.unipi.it/software/mmtoolbox/. Retrieved on May 7, 2010.
[6] D. A. Bini, B. Meini, and F. Poloni. An effective matrix geometric mean satisfying the Ando–Li–Mathias properties. *Math. Comp.*, 79(269):437–452, 2010.
[7] N. J. Higham. *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
[8] J. D. Lawson and Y. Lim. The geometric mean, matrices, metrics, and more. *Amer. Math. Monthly*, 108(9):797–812, 2001.
[9] J. H. Manton. A globally convergent numerical algorithm for computing the centre of mass on compact lie groups. In *Eighth International Conference on Control, Automation, Robotics and Vision, 2004. ICARCV 2004 8th, Kunming, China, December 2004*, 2004.

[10] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.*, 26(3):735–747, 2005.

[11] M. Moakher. On the averaging of symmetric positive-definite tensors. *J. Elasticity*, 82(3):273–296, 2006.

[12] F. Poloni. Constructing matrix geometric means, 2009. arXiv:0906.3132v1.

[13] G. Pusz and S. L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Rep. Math. Phys.*, 8:159–170, 1975.