$$\boxed{\textbf{ELA}}$$

CONSTRUCTING MATRIX GEOMETRIC MEANS*

FEDERICO POLONI†

Abstract. In this paper, we analyze the process of "assembling" new matrix geometric means from existing ones, through function composition or limit processes. We show that for $n = 4$ a new matrix mean exists which is simpler to compute than the existing ones. Moreover, we show that for $n > 4$ the existing proving strategies cannot provide a mean computationally simpler than the existing ones.

Key words. Matrix geometric mean, Positive definite matrix, Invariance properties, Groups of permutations.

AMS subject classifications. 65F30, 15A48, 47A64, 20B35.

1. Introduction.

Literature review. In the last few years, several papers have been devoted to defining a proper way to generalize the concept of geometric mean to $n \geq 3$ Hermitian, positive definite $m \times m$ matrices. A seminal paper by Ando, Li and Mathias [1] defined the mathematical problem by stating ten properties that a "good" matrix geometric mean should satisfy. However, these properties do not uniquely define a multivariate matrix geometric mean; thus several different definitions appeared in literature.

Ando, Li and Mathias [1] first proposed a mean whose definition for n matrices is based on a limit process involving several geometric means of $n-1$ matrices. Later Bini, Meini and Poloni [4] noted that the slow convergence speed of this method prevents its use in applications; its main shortcoming is the fact that its complexity grows as $O(n!)$ with the number of involved matrices. In the same paper, they proposed a similar limit process with increased convergence speed, but still with complexity $O(n!)$. Pálfia [9] proposed a mean based on a similar process involving only means of 2 matrices, and thus much simpler and cheaper to compute, but lacking property P3 (permutation invariance) from the ALM list. Lim [6] proposed a family of matrix geometric means that are based on an iteration requiring at each step the computation of a mean of $m \geq n$ matrices. Since the computational complexity for all known means greatly increases with n, the resulting family is useful as an example but highly

ELA

impractical for numerical computations.

At the same time, Moakher [7, 8] and Bhatia and Holbrook [3, 2] proposed a completely different definition, which we shall call the *Riemannian centroid* of $A_1, A_2, \ldots, A_n$. The Riemannian centroid $G^R(A_1, A_2, \ldots, A_n)$ is defined as the minimizer of a sum of squared distances,

$$G^R(A_1, A_2, \ldots, A_n) = \arg\min_X \sum_{i=1}^n \delta^2(A_i, X), \qquad (1.1)$$

where δ is the geodesic distance induced by a natural Riemannian metric on the space of symmetric positive definite matrices. The same X is the unique solution of the equation

$$\sum_{i=1}^n \log(A_i^{-1}X) = 0, \qquad (1.2)$$

involving the matrix logarithm function. While most of the ALM properties are easy to prove, it is still an open problem whether it satisfies P4 (monotonicity). The computational experiments performed up to now gave no counterexamples, but the monotonicity of the Riemannian centroid is still a conjecture [3], up to our knowledge.

Moreover, while the other means had constructive definitions, it is not apparent how to compute the solution to either (1.1) or (1.2). Two methods have been proposed, one based on a fixed-point iteration [8] and one on the Newton methods for manifolds [9, 8]. Although both seem to work well on "tame" examples, their computational results show a fast degradation of the convergence behavior as the number of matrices and their dimension increase. It is unclear whether on more complicated examples there is convergence in the first place; unlike the other means, the convergence of these iteration processes has not been proved, as far as we know.

Notations. Let us denote by $\mathbb{P}_m$ the space of Hermitian positive-definite $m \times m$ matrices. For all $A, B \in \mathbb{P}_m$, we shall say that $A < B$ ($A \leq B$) if $B - A$ is positive definite (semidefinite). With A^* we denote the conjugate transpose of A. We shall say that $\underline{A} = (A_i)_{i=1}^n \in (\mathbb{P}_m)^n$ is a *scalar n-tuple* of matrices if $A_1 = A_2 = \cdots = A_n$. We shall use the convention that both $Q(\underline{A})$ and $Q(A_1, \ldots, A_n)$ denote the application of the map $Q : (\mathbb{P}_m)^n \to \mathbb{P}_m$ to the n-tuple $\underline{A}$.

ALM properties. Ando, Li and Mathias [1] introduced ten properties defining when a map $G : (\mathbb{P}_m)^n \to \mathbb{P}_m$ can be called a *geometric mean*. Following their paper, we report here the properties for $n = 3$ only, for the sake of simplicity; the generalization to different values of n is straightforward.

P1 (consistency with scalars) If A, B, C commute then $G(A, B, C) = (ABC)^{1/3}$.

P1' This implies $G(A, A, A) = A$.

P2 (joint homogeneity) $G(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{1/3}G(A, B, C)$, for each $\alpha, \beta, \gamma > 0$.

P2' This implies $G(\alpha A, \alpha B, \alpha C) = \alpha G(A, B, C)$.

P3 (permutation invariance) $G(A, B, C) = G(\pi(A, B, C))$ for all the permutations $\pi(A, B, C)$ of A, B, C.

P4 (monotonicity) $G(A, B, C) \geq G(A', B', C')$ whenever $A \geq A'$, $B \geq B'$, $C \geq C'$.

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) $G(S^*AS, S^*BS, S^*CS) = S^*G(A, B, C)S$ for any non-singular S.

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 $G(A, B, C) \geq \lambda G(A_1, B_1, C_1) + (1 - \lambda)G(A_2, B_2, C_2)$.

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)

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

The matrix geometric mean for $n = 2$. For $n = 2$, the ALM properties uniquely define a matrix geometric mean which can be expressed explicitly as

$$A \# B := A(A^{-1}B)^{1/2}. \tag{1.3}$$

This is a particular case of the more general map

$$A \#_t B := A(A^{-1}B)^t, \quad t \in \mathbb{R}, \tag{1.4}$$

which has a geometrical interpretation as the parametrization of the geodesic joining A and B for a certain Riemannian geometry on $\mathbb{P}_m$ [2].

The ALM and BMP means. Ando, Li and Mathias [1] recursively define a matrix geometric mean G_n^{ALM} of n matrices in this way. The mean G_2^{ALM} of two matrices coincides with (1.3); for $n \geq 3$, suppose the mean of $n - 1$ matrices G_{n-1}^{ALM} is already defined. Given $A_1, \ldots, A_n$, compute for each $j = 1, 2, \ldots$

$$A_i^{(j+1)} := G_{n-1}^{ALM}(A_1^{(j)}, A_2^{(j)}, \ldots A_{i-1}^{(j)}, A_{i+1}^{(j)}, \ldots A_n^{(j)}) \quad i = 1, \ldots, n, \tag{1.5}$$

where $A_i^{(0)} := A_i$, $i = 1, \ldots n$. The sequences $(A_i^{(j)})_{j=1}^{\infty}$ converge to a common (not depending on i) matrix, and this matrix is a geometric mean of $A_1^{(0)}, \ldots, A_n^{(0)}$.

The mean proposed by Bini, Meini and Poloni [4] is defined in the same way, but with (1.5) replaced by

$$A_i^{(j+1)} := G_{n-1}^{BMP}(A_1^{(j)}, A_2^{(j)}, \ldots A_{i-1}^{(j)}, A_{i+1}^{(j)}, \ldots A_n^{(j)}) \#_{1/n} A_i \quad i = 1, \ldots, n. \tag{1.6}$$

ELA

Though both maps satisfy the ALM properties, matrices A, B, C exist for which $G^{ALM}(A, B, C) \neq G^{BMP}(A, B, C)$.

While the former iteration converges linearly, the latter converges cubically, and thus allows one to compute a matrix geometric mean with a lower number of iterations. In fact, if we call p_k the average number of iterations that the process giving a mean of k matrices takes to converge (which may vary significantly depending on the starting matrices), the total computational cost of the ALM and BMP means can be expressed as $O(n!p_3p_4 \ldots p_n m^3)$. The only difference between the two complexity bounds lies in the expected magnitude of the values p_k. The presence of a factorial and of a linear number of factors p_k is undesirable, since it means that the problem scales very badly with n. In fact, already with $n = 7, 8$ and moderate values of m, a large CPU time is generally needed to compute a matrix geometric mean [4].

The Pálfia mean. Pálfia [9] proposed to consider the following iteration. Let again $A_i^{(0)} := A_i$, $i = 1, \ldots, n$. Let us define

$$A_i^{(k+1)} := A_i^{(k)} \# A_{i+1}^{(k)}, \quad i = 1, \ldots, n, \qquad (1.7)$$

where the indices are taken modulo n, i.e., $A_{n+1}^{(k)} = A_1^{(k)}$ for all k. We point out that the definition in the original paper [9] is slightly different, as it considers several possible orderings of the input matrices, but the means defined there can be put in the form (1.7) up to a permutation of the starting matrices $A_1, \ldots, A_n$.

As for the previous means, it can be proved that the iteration (1.7) converges to a scalar n-tuple; we call the common limit of all components $G^P(A_1, \ldots, A_n)$. As we noted above, this function does not satisfy P3 (permutation invariance), and thus it is not a geometric mean in the ALM sense.

Other composite means. Apart from the Riemannian centroid, all the other definitions follow the same pattern:

- build new functions of n matrices by taking nested compositions of the existing means—preferably using only means of less than n matrices;
- take the common limit of a set of n functions defined as in the above step.

The possibilities for defining new iterations following this pattern are endless. Ando, Li, Mathias, and Bini, Meini, Poloni chose to use in the first step composite functions using computationally expensive means of $n-1$ matrices; this led to poor convergence results. Pálfia chose instead to use more economical means of two variables as starting points; this led to better convergence (no $O(n!)$), but to a function which is not symmetric with respect to permutations of its entries (P3, permutation invariance).

ELA

As we shall see in the following, the property P3 is crucial: all the other ones are easily proved for a mean defined as composition/limit of existing means.

A natural question to ask is whether we can build a matrix geometric mean of n matrices as the composition of matrix means of less matrices, without the need of a limit process. Two such unsuccessful attempts are reported in the paper by Ando, Li and Mathias [1], as examples of the fact that it is not easy to define a matrix satisfying P1–P10. The first is

$$G^{4rec}(A, B, C, D) := (A \# B) \# (C \# D). \qquad (1.8)$$

Unfortunately, there are matrices such that $(A \# B) \# (C \# D) \neq (A \# C) \# (B \# D)$, so P3 fails. A second attempt is

$$G^{rec}(A, B, C) := (A^{4/3} \# B^{4/3}) \# C^{2/3}, \qquad (1.9)$$

where the exponents are chosen so that P1 (consistency with scalars) is satisfied. Again, this function is not symmetric in its arguments, and thus fails to satisfy P3.

A second natural question is whether an iterative scheme such as the ones for G^{ALM}, G^{BMP} and G^P can yield P3 without having a $O(n!)$ computational cost. For example, if we could build a scheme similar to the ALM and BMP ones, but using only means of $\frac{n}{2}$ matrices in the recursion, then the $O(n!)$ growth would disappear.

In this paper, we aim to analyze in more detail the process of "assembling" new matrix means from the existing ones, and show which new means can be found, and what cannot be done because of group-theoretical obstructions related to the symmetry properties of the composed functions. By means of a group-theoretical analysis, we will show that for $n = 4$ a new matrix mean exists which is simpler to compute than the existing ones; numerical experiments show that the new definition leads to a significant computational advantage. Moreover, we will show that for $n > 4$ the existing strategies of composing matrix means and taking limits cannot provide a mean which is computationally simpler than the existing ones.

2. Quasi-means and notation.

Quasi-means. Let us introduce the following variants to some of the Ando–Li–Mathias properties.

P1" Weak consistency with scalars. There are $\alpha, \beta, \gamma \in \mathbb{R}$ such that if A, B, C commute, then $G(A, B, C) = A^\alpha B^\beta C^\gamma$.

P2" Weak homogeneity. There are $\alpha, \beta, \gamma \in \mathbb{R}$ such that for each $r, s, t > 0$, $G(rA, sB, tC) = r^\alpha s^\beta t^\gamma G(A, B, C)$. Notice that if P1" holds as well, these must be the same α, β, γ (proof: substitute scalar values in P1").

ELA

P9' Weak determinant identity. For all $d > 0$, if $\det A = \det B = \det C = d$, then $\det G(A, B, C) = d$.

We shall call a *quasi-mean* a function $Q : (\mathbb{P}_m)^n \to (\mathbb{P}_m)$ that satisfies P1",P2", P4, P6, P7, P8, P9'. This models expressions which are built starting from basic matrix means but are not symmetric, e.g., $A \# G(B, C, D \# E)$, (1.8), and (1.9).

THEOREM 2.1. *If a quasi-mean Q satisfies P3 (permutation invariance), then it is a geometric mean.*

Proof. From P2" and P3, it follows that $\alpha = \beta = \gamma$. From P9', it follows that if $\det A = \det B = \det C = 1$,

$$2^m = \det Q(2A, 2B, 2C) = \det \left(2^{\alpha+\beta+\gamma} Q(A, B, C) \right) = 2^{m(\alpha+\beta+\gamma)},$$

thus $\alpha + \beta + \gamma = 1$. The two relations combined together yield $\alpha = \beta = \gamma = 1/3$. Finally, it is proved in Ando, Li and Mathias [1] that P5 and P10 are implied by the other eight properties P1–P4 and P6–P9. $\square$

For two quasi-means Q and R of n matrices, we shall write $Q = R$ if $Q(\underline{A}) = R(\underline{A})$ for each n-tuple $\underline{A} \in \mathbb{P}_m$

Group theory notation. The notation $H \leq G$ $(H < G)$ means that H is a subgroup (proper subgroup) of G. Let us denote by $\mathfrak{S}_n$ the symmetric group on n elements, i.e., the group of all permutations of the set $\{1, 2, \ldots, n\}$. As usual, the symbol $(a_1 a_2 a_3 \ldots a_k)$ stands for the permutation ("cycle") that maps $a_1 \mapsto a_2$, $a_2 \mapsto a_3$, $\ldots a_{k-1} \mapsto a_k$, $a_k \mapsto a_1$ and leaves the other elements of $\{1, 2, \ldots n\}$ unchanged. Different symbols in the above form can be chained to denote the group operation of function composition; for instance, $\sigma = (13)(24)$ is the permutation $(1, 2, 3, 4) \mapsto (3, 4, 1, 2)$. We shall denote by $\mathfrak{A}_n$ the alternating group on n elements, i.e., the only subgroup of index 2 of $\mathfrak{S}_n$, and by $\mathfrak{D}_n$ the dihedral group over n elements, with cardinality $2n$. The latter is identified with the subgroup of $\mathfrak{S}_n$ generated by the rotation $(1, 2, \ldots, n)$ and the mirror symmetry $(2, n)(3, n-1) \cdots (n/2, n/2+2)$ (for even values of n) or $(2, n)(3, n-1) \cdots ((n+1)/2, (n+3)/2)$ (for odd values of n).

Coset transversals. Let now $H \leq \mathfrak{S}_n$, and let $\{\sigma_1, \ldots, \sigma_r\} \subset \mathfrak{S}_n$ be a transversal for the right cosets $H\sigma$, i.e., a set of maximal cardinality $r = n!/|H|$ such that $\sigma_j \sigma_i^{-1} \notin H$ for all $i \neq j$. The group $\mathfrak{S}_n$ acts by permutation over the cosets $(H\sigma_1, \ldots, H\sigma_r)$, i.e., for each σ there is a permutation $\tau = \rho_H(\sigma)$ such that

$$(H\sigma_1\sigma, \ldots, H\sigma_r\sigma) = (H\sigma_{\tau(1)}, \ldots, H\sigma_{\tau(r)}).$$

It is easy to check that in this case $\rho_H : \mathfrak{S}_n \to \mathfrak{S}_r$ must be a group homomorphism. Notice that if H is a normal subgroup of $\mathfrak{S}_n$, then the action of $\mathfrak{S}_n$ over the coset space is represented by the quotient group $\mathfrak{S}_n/H$, and the kernel of ρ_H is H.

ELA

EXAMPLE 2.2. The coset space of $H = \mathfrak{D}_4$ has size $4!/8 = 3$, and a possible transversal is $\sigma_1 = e$, $\sigma_2 = (12)$, $\sigma_3 = (14)$. We have $\rho_H(\mathfrak{S}_4) \cong \mathfrak{S}_3$: indeed, the permutation $\sigma = (12) \in \mathfrak{S}_4$ is such that $(H\sigma_1\sigma, H\sigma_2\sigma, H\sigma_3\sigma) = (H\sigma_2, H\sigma_1, H\sigma_3)$, and therefore $\rho_H(\sigma) = (12)$, while the permutation $\tilde{\sigma} = (14) \in \mathfrak{S}_4$ is such that $(H\sigma_1\tilde{\sigma}, H\sigma_2\tilde{\sigma}, H\sigma_3\tilde{\sigma}) = (H\sigma_3, H\sigma_2, H\sigma_1)$, therefore $\rho_H(\tilde{\sigma}) = (13)$. Thus $\rho_H(\mathfrak{S}_4)$ must be a subgroup of $\mathfrak{S}_3$ containing (12) and (13), that is, $\mathfrak{S}_3$ itself.

With the same technique, noting that $\sigma_i^{-1}\sigma_j$ maps the coset $H\sigma_i$ to $H\sigma_j$, we can prove that the action ρ_H of $\mathfrak{S}_n$ over the coset space is transitive.

Group action and composition of quasi-means. We may define a right action of $\mathfrak{S}_n$ on the set of quasi-means of n matrices as

$$(Q\sigma)(A_1, \ldots, A_n) := Q(A_{\sigma(1)}, \ldots, A_{\sigma(n)}).$$

The choice of putting σ to the right, albeit slightly unusual, was chosen to simplify some of the notations used in Section 4.

When Q is a quasi-mean of r matrices and $R_1, R_2, \ldots R_r$ are quasi-means of n matrices, let us define $Q \circ (R_1, R_2, \ldots R_r)$ as the map

$$(Q \circ (R_1, R_2, \ldots R_r))\,(\underline{A}) := Q(R_1(\underline{A}), R_2(\underline{A}), \ldots, R_r(\underline{A})). \tag{2.1}$$

THEOREM 2.3. *Let $Q(A_1, \ldots, A_r)$ and $R_j(A_1, \ldots A_n)$ (for $j = 1, \ldots, r$) be quasi-means. Then,*

1. *For all $\sigma \in \mathfrak{S}_r$, $Q\sigma$ is a quasi-mean.*
2. *$(A_1, \ldots, A_r, A_{r+1}) \mapsto Q(A_1, \ldots, A_r)$ is a quasi-mean.*
3. *$Q \circ (R_1, R_2, \ldots R_r)$ is a quasi-mean.*

Proof. All properties follow directly from the monotonicity (P4) and from the corresponding properties for the means Q and R_j. □

We may then define the *isotropy group*, or *stabilizer group* of a quasi-mean Q

$$\mathrm{Stab}(Q) := \{\sigma \in \mathfrak{S}^n : Q = Q\sigma\}. \tag{2.2}$$

3. Means obtained as map compositions.

Reductive symmetries. Let us define the concept of *reductive symmetries* of a quasi-mean as follows.

- in the special case in which $G_2(A, B) = A \# B$, the symmetry property that $A \# B = B \# A$ is a reductive symmetry.

ELA

426   F. Poloni

- let $Q \circ (R_1, \ldots, R_r)$ be a quasi-mean obtained by composition. The symmetry with respect to the permutation σ (i.e., the fact that $Q = Q\sigma$) is a *reductive symmetry* for $Q \circ (R_1, \ldots, R_r)$ if this property can be formally proved relying only on the reductive symmetries of Q and $R_1, \ldots, R_r$.

For instance, if we take $Q(A, B, C) := A \# (B \# C)$, then we can deduce that $Q(A, B, C) = Q(A, C, B)$ for all A, B, C, but not that $Q(A, B, C) = Q(B, C, A)$ for all A, B, C. This does not imply that such a symmetry property does not hold: if we were considering the operator $+$ instead of $\#$, then it would hold that $A+B+C = B+C+A$, but there are no means of proving it relying only on the commutativity of addition — in fact, associativity is crucial.

As we stated in the introduction, Ando, Li and Mathias [1] showed explicit counterexamples proving that all the symmetry properties of G^{4rec} and G^{rec} are reductive symmetries. We conjecture the following.

CONJECTURE 1. *All the symmetries of a quasi-mean obtained by recursive composition from G_2 are reductive symmetries.*

In other words, we postulate that no "unexpected symmetries" appear while examining quasi-means compositions. This is a rather strong statement; however, the numerical experiments and the theoretical analysis performed up to now never showed any invariance property that could not be inferred by those of the underlying means.

We shall prove several result limiting the reductive symmetries that a mean can have; to this aim, we introduce the *reductive isotropy group*

$$\mathrm{RStab}(Q) = \{\sigma \in \mathrm{Stab}(Q) : Q = Q\sigma \text{ is a reductive symmetry}\}. \qquad (3.1)$$

We will prove that there is no quasi-mean Q such that $\mathrm{RStab}(Q) = \mathfrak{S}_n$. This shows that the existing "tools" in the mathematician's "toolbox" do not allow one to construct a matrix geometric mean (with full proof) based only on map compositions; thus we need either to devise a completely new construction or to find a novel way to prove additional invariance properties involving map compositions.

Reduction to a special form. The following results show that when looking for a reductive matrix geometric mean, i.e., a quasi-mean Q with $\mathrm{RStab}\,Q = \mathfrak{S}_n$, we may restrict our search to quasi-means of a special form.

THEOREM 3.1. *Let Q be a quasi-mean of $r + s$ matrices, and $R_1, R_2, \ldots, R_r$, $S_1, S_2, \ldots, S_s$ be quasi-means of n matrices such that $R_i \neq S_j\sigma$ for all i, j and every*

ELA

$\sigma \in \mathfrak{S}_n$. *Then,*

$$\mathrm{RStab}(Q \circ (R_1, R_2, \ldots, R_r, S_1, S_2, \ldots, S_s))$$
$$\subseteq \mathrm{RStab}(Q \circ (R_1, \ldots, R_r, R_1, R_1, \ldots, R_1)). \quad (3.2)$$

Proof. Let $\sigma \in \mathrm{RStab}(Q \circ (R_1, R_2, \ldots, R_r, S_1, S_2, \ldots S_s))$; since the only invariance properties that we may assume on Q are those predicted by its invariance group, it must be the case that

$$(R_1\sigma, R_2\sigma, \ldots, R_r\sigma, S_1\sigma, S_2\sigma, \ldots S_s\sigma)$$

is a permutation of $(R_1, R_2, \ldots, R_r, S_1, S_2, \ldots S_s)$ belonging to $\mathrm{RStab}(Q)$. Since $R_i \neq S_j\sigma$, this permutation must map the sets $\{R_1, R_2, \ldots, R_r\}$ and $\{S_1, S_2, \ldots S_s\}$ to themselves. Therefore, the same permutation maps

$$(R_1, R_2, \ldots, R_r, R_1, R_1, \ldots R_1)$$

to

$$(R_1\sigma, R_2\sigma, \ldots, R_r\sigma, R_1\sigma, R_1\sigma, \ldots R_1\sigma).$$

This implies that

$$Q(R_1, R_2, \ldots, R_r, R_1, R_1, \ldots R_1) \quad = \quad Q(R_1\sigma, R_2\sigma, \ldots, R_r\sigma, R_1\sigma, R_1\sigma, \ldots R_1\sigma)$$

as requested. ◻

THEOREM 3.2. *Let* $M_1 := Q \circ (R_1, R_2, \ldots, R_r)$ *be a quasi-mean. Then there is a quasi-mean* M_2 *in the form*

$$\tilde{Q} \circ (\tilde{R}\sigma_1, \tilde{R}\sigma_2, \ldots, \tilde{R}\sigma_{\tilde{r}}), \quad (3.3)$$

where $(\sigma_1, \sigma_2, \ldots, \sigma_{\tilde{r}})$ *is a right coset transversal for* $\mathrm{RStab}(\tilde{R})$ *in* $\mathfrak{S}_n$, *such that* $\mathrm{RStab}(M_1) \subseteq \mathrm{RStab}(M_2)$.

Proof. Set $\tilde{R} = R_1$. For each $i = 2, 3, \ldots, r$ if $R_i \neq \tilde{R}\sigma$, we may replace it with $\tilde{R}$, and by Theorem 3.1 the restricted isotropy group increases or stays the same. Thus by repeated application of this theorem, we may reduce to the case in which each R_i is in the form $\tilde{R}\tau_i$ for some permutation τ_i.

Since $\{\sigma_i\}$ is a right transversal, we may write $\tau_i = h_i\sigma_{k(i)}$ for some $h_i \in H$ and $k(i) \in \{1, 2, \ldots, \tilde{r}\}$. We have $\tilde{R}h = \tilde{R}$ since $h \in \mathrm{Stab}\,\tilde{R}$, thus $R_i = \tilde{R}\sigma_{k(i)}$. The resulting quasi-mean is $Q \circ (\tilde{R}\sigma_{k(1)}, \ldots, \tilde{R}\sigma_{k(r)})$. Notice that we may have $k(i) =$

ELA

F. Poloni

$k(j)$, or some cosets may be missing. Let now $\tilde{Q}$ be defined as $\tilde{Q}(A_1, A_2, \ldots, A_{\tilde{r}}) := Q(A_{k(1)}, \ldots, A_{k(r)})$; then we have

$$\tilde{Q}(\tilde{R}\sigma_1, \ldots, \tilde{R}\sigma_{\tilde{r}}) = Q(\tilde{R}\sigma_{k(1)}, \ldots, \tilde{R}\sigma_{k(r)}) \tag{3.4}$$

and thus the isotropy groups of the left-hand side and right-hand side coincide. $\square$

For the sake of brevity, we shall define

$$Q \circ R := Q \circ (R\sigma_1, \ldots, R\sigma_r),$$

assuming a standard choice of the transversal for $H = \operatorname{Stab} R$. Notice that $Q \circ R$ depends on the ordering of the cosets $H\sigma_1, \ldots, H\sigma_r$, but not on the choice of the coset representative σ_i, since $Qh\sigma_i = Q\sigma_i$ for each $h \in H$.

EXAMPLE 3.3. The quasi-mean $(A, B, C) \mapsto (A \,\#\, B)\,\#(B \,\#\, C)$ is $Q \circ Q$, where $Q(X, Y, Z) = X \,\#\, Y$, $H = \{e, (12)\}$, and the transversal is $\{e, (13), (23)\}$.

EXAMPLE 3.4. The quasi-mean $(A, B, C) \mapsto (A \,\#\, B) \,\#\, C$ is not in the form (3.3), but in view of Theorem 3.1, its restricted isotropy group is a subgroup of that of $(A, B, C) \mapsto (A \,\#\, B) \,\#(A \,\#\, B)$.

The following theorem shows which permutations we can actually prove to belong to the reductive isotropy group of a mean in the form (3.3).

THEOREM 3.5. Let $H \leq \mathfrak{S}_n$, R be a quasi-mean of n matrices such that $\operatorname{RStab} R = H$ and Q be a quasi-mean of $r = n!/|H|$ matrices. Let $G \in \mathfrak{S}_n$ be the largest permutation subgroup such that $\rho_H(G) \leq \operatorname{RStab}(Q)$. Then, $G = \operatorname{RStab}(Q \circ R)$.

Proof. Let $\sigma \in G$ and $\tau = \rho_H(\sigma)$; we have

$$\begin{aligned} (Q \circ R)\sigma(\underline{A}) &= Q\big(R\sigma_1\sigma(\underline{A}), R\sigma_2\sigma(\underline{A}), \ldots, R\sigma_r\sigma(\underline{A})\big) \\ &= Q\big(R\sigma_{\tau(1)}(\underline{A}), R\sigma_{\tau(2)}(\underline{A}), \ldots, R\sigma_{\tau(r)}(\underline{A})\big) \\ &= Q\big(R\sigma_1(\underline{A}), R\sigma_2(\underline{A}), \ldots, R\sigma_r(\underline{A})\big), \end{aligned}$$

where the last equality holds because $\tau \in \operatorname{Stab}(Q)$.

Notice that the above construction is the only way to obtain invariance with respect to a given permutation σ: indeed, to prove invariance relying only on the invariance properties of Q, $(R\sigma_1\sigma, \ldots, R\sigma_r\sigma) = (R\sigma_{\tau(1)}, \ldots, R\sigma_{\tau(r)})$ must be a permutation of $(R\sigma_1, \ldots, R\sigma_r)$ belonging to $\operatorname{RStab} Q$, and thus $\rho_H(\sigma) = \tau \in \operatorname{Stab} Q$. Thus the reductive invariance group of the composite mean is precisely the largest subgroup G such that $\rho_H(G) \leq \operatorname{Stab} Q$. $\square$

EXAMPLE 3.6. Let $n = 4$, Q be any (reductive) geometric mean of three matrices (i.e., $\operatorname{RStab} Q = \mathfrak{S}_3$), and $R(A, B, C, D) := (A \,\#\, C) \,\#(B \,\#\, D)$. We have

ELA

$H = \text{RStab}\, R = \mathfrak{D}_4$, the dihedral group over four elements, with cardinality 8. There are $r = 4!/|H| = 3$ cosets. Since $\rho_H(\mathfrak{S}_4)$ is a subset of $\text{Stab}\, Q = \mathfrak{S}_3$, the isotropy group of $Q \circ R$ contains $G = \mathfrak{S}_4$ by Theorem 3.5. Therefore $Q \circ R$ is a geometric mean of four matrices.

Indeed, the only assertion we have to check explicitly is that $\text{RStab}\, R = \mathfrak{D}_4$. The isotropy group of R contains (24) and (1234), since by using repeatedly the fact that $\#$ is symmetric in its arguments we can prove that $R(A, B, C, D) = R(A, D, C, B)$ and $R(A, B, C, D) = R(D, A, B, C)$. Thus it must contain the subgroup generated by these two elements, that is, $\mathfrak{D}_4 \leq \text{RStab}\, R$. The only subgroups of $\mathfrak{S}_4$ containing $\mathfrak{D}_4$ as a subgroup are the two trivial ones $\mathfrak{S}_4$ and $\mathfrak{D}_4$. We cannot have $\text{RStab}\, R = \mathfrak{S}_4$, since R has the same definition as G^{4rec} of equation (1.8), apart from a reordering, and it was proved [1] that this is not a geometric mean.

It is important to notice that by choosing $G_3 = G_3^{ALM}$ or $G_3 = G_3^{BMP}$ in the previous example we may obtain a geometric mean of four matrices using a single limit process, the one needed for G_3. This is more efficient than G_4^{ALM} and G_4^{BMP}, which compute a mean of four matrices via several means of three matrices, each of which requires a limit process in its computation. We will return to this topic in Section 5.

Above four elements. Is it possible to obtain a reductive geometric mean of n matrices, for $n > 4$, starting from simpler means and using the construction of Theorem 3.5? The following result shows that the answer is no.

THEOREM 3.7. *Suppose $G := \text{RStab}(Q \circ R) \geq \mathfrak{A}_n$ and $n > 4$. Then $\mathfrak{A}_n \leq \text{RStab}(Q)$ or $\mathfrak{A}_n \leq \text{RStab}(R)$.*

Proof. Let us consider $K = \ker \rho_H$. It is a normal subgroup of $\mathfrak{S}_n$, but for $n > 4$ the only normal subgroups of $\mathfrak{S}_n$ are the trivial group $\{e\}$, $\mathfrak{A}_n$ and $\mathfrak{S}_n$ [5]. Let us consider the three cases separately.

1. $K = \{e\}$. In this case, $\rho_H(G) \cong G$, and thus $G \leq \text{RStab}\, Q$.
2. $K = \mathfrak{S}_n$. In this case, $\rho_H(\mathfrak{S}_n)$ is the trivial group. But the action of $\mathfrak{S}_n$ over the coset space is transitive, since $\sigma_i^{-1}\sigma_j$ sends the coset $H\sigma_i$ to the coset $H\sigma_j$. So the only possibility is that there is a single coset in the coset space, i.e., $H = \mathfrak{S}_n$.
3. $K = \mathfrak{A}_n$. As in the above case, since the action is transitive, it must be the case that there are at most two cosets in the coset space, and thus $H = \mathfrak{S}_n$ or $H = \mathfrak{A}_n$. ∎

Thus it is impossible to apply Theorem 3.5 to obtain a quasi-mean with reductive isotropy group containing $\mathfrak{A}_n$, unless one of the two starting quasi-means has a re-

ELA

ductive isotropy group already containing $\mathfrak{A}_n$.

4. Means obtained as limits.

An algebraic setting for limit means. We shall now describe a unifying algebraic setting in terms of isotropy groups, generalizing the procedures leading to the means defined by limit processes G^{ALM}, G^{BMP} and G^P.

Let $S : (\mathbb{P}_m)^n \to (\mathbb{P}_m)^n$ be a map; we shall say that S *preserves* a subgroup $H < \mathfrak{S}_n$ if there is a map $\tau : H \to H$ such that $Sh(\underline{A}) = \tau(h)S(\underline{A})$ for all $\underline{A} \in \mathbb{P}_m$.

THEOREM 4.1. *Let $S : (\mathbb{P}_m)^n \to (\mathbb{P}_m)^n$ be a map and $H < \mathfrak{S}_n$ be a permutation group such that*

1. $(\underline{A}) \to \big(S(\underline{A})\big)_i$ *is a quasi-mean for all $i = 1, \ldots, n$,*
2. S *preserves H,*
3. *for all $\underline{A} \in (\mathbb{P}_m)^n$, $\lim_{k \to \infty} S^k(\underline{A})$ is a scalar n-tuple[1],*

and let us denote by $S^\infty(\underline{A})$ the common value of all entries of the scalar n-tuple $\lim_{k \to \infty} S^k(\underline{A})$. Then, $S^\infty(\underline{A})$ is a quasi-mean with isotropy group containing H.

Proof. From Theorem 2.3, it follows that $\underline{A} \mapsto \big(S^k(\underline{A})\big)_i$ is a quasi-mean for each k. Since all the properties defining a quasi-mean pass to the limit, S^∞ is a quasi-mean itself.

Let us take $h \in H$ and $\underline{A} \in \mathbb{P}_n$. It is easy to prove by induction on k that $S^k h(\underline{A}) = \tau^k(h)\big(S^k(\underline{A})\big)$. Now, choose a matrix norm inducing the Euclidean topology on $\mathbb{P}_m$; let $\varepsilon > 0$ be fixed, and let us take K such that for all $k > K$ and for all $i = 1, \ldots, n$ the following inequalities hold:

- $\left\| \big(S^k(\underline{A})\big)_i - S^\infty(\underline{A}) \right\| < \varepsilon,$
- $\left\| \big(S^k h(\underline{A})\big)_i - S^\infty h(\underline{A}) \right\| < \varepsilon.$

We know that $\big(S^k h(\underline{A})\big)_i = \big(\tau^k(h)S^k(\underline{A})\big)_i = \big(S^k(\underline{A})\big)_{\tau^k(h)(i)}$, therefore

$$\|S^\infty(\underline{A}) - S^\infty h(\underline{A})\| \leq \left\| \big(S^k(\underline{A})\big)_{\tau^k(h)(i)} - S^\infty(\underline{A}) \right\|$$
$$+ \left\| \big(S^k h(\underline{A})\big)_i - S^\infty h(\underline{A}) \right\| < 2\varepsilon.$$

Since ε is arbitrary, the two limits must coincide. This holds for each $h \in H$, therefore $H \leq \operatorname{Stab} S^\infty$. ∎

---

[1]Here S^k denotes function iteration: $S^1 = S$ and $S^{k+1}(\underline{A}) = S(S^k(\underline{A}))$ for all k.

ELA

EXAMPLE 4.2. The map S defining G_4^{ALM} is

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \mapsto \begin{bmatrix} G_3^{ALM}(B,C,D) \\ G_3^{ALM}(A,C,D) \\ G_3^{ALM}(A,B,D) \\ G_3^{ALM}(A,B,C) \end{bmatrix}.$$

One can see that $S\sigma = \sigma^{-1}S$ for each $\sigma \in \mathfrak{S}_4$, and thus with the choice $\tau(\sigma) := \sigma^{-1}$ we get that S preserves $\mathfrak{S}_4$. Thus, by Theorem 4.1, $S^\infty = G_4^{ALM}$ is a geometric mean of four matrices. The same reasoning applies to G^{BMP}.

EXAMPLE 4.3. The map S defining G_4^P is

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \mapsto \begin{bmatrix} A \# B \\ B \# C \\ C \# D \\ D \# A \end{bmatrix}.$$

S preserves the dihedral group $\mathfrak{D}_4$. Therefore, provided the iteration process converges to a scalar n-tuple, S^∞ is a quasi-mean with isotropy group containing $\mathfrak{D}_4$.

Efficiency of the limit process. As in the previous section, we are interested in seeing whether this approach, which is the one that has been used to prove invariance properties of the known limit means [1, 4], can yield better results for a different map S.

THEOREM 4.4. *Let $S : (\mathbb{P}_m)^n \to (\mathbb{P}_m)^n$ preserve a group H. Then, the invariance group of each of its components S_i, $i = 1, \ldots, n$, is a subgroup of H of index at most n.*

Proof. Let i be fixed, and set $I_k := \{h \in H : \tau(h)(i) = k\}$. The sets I_k are mutually disjoint and their union is H, so the largest one has cardinality at least $|H|/n$, let us call it $I_{\bar{k}}$.

From the hypothesis that S preserves H, we get $S_i h(\underline{A}) = S_{\bar{k}}(\underline{A})$ for each $\underline{A}$ and each $h \in I_k$. Let $\bar{h}$ be an element of I_k; then $S_i h(\bar{h}^{-1}\underline{A}) = S_{\bar{k}}(\bar{h}^{-1}\underline{A}) = S_i(\underline{A})$. Thus the isotropy group of S_i contains all the elements of the form $h\bar{h}^{-1}$, $h \in I_k$, and those are at least $|H|/n$. □

The following result holds [5, page 147].

THEOREM 4.5. *For $n > 4$, the only subgroups of $\mathfrak{S}_n$ with index at most n are:*

- *the alternating group $\mathfrak{A}_n$,*
- *the n groups $T_k = \{\sigma \in \mathfrak{S}_n : \sigma(k) = k\}$, $k = 1, \ldots, n$, all of which are isomorphic to $\mathfrak{S}_{n-1}$,*

ELA

F. Poloni

- *for $n = 6$ only, another conjugacy class of 6 subgroups of index 6 isomorphic to $\mathfrak{S}_5$.*

Analogously, the only subgroups of $\mathfrak{A}_n$ with index at most n are:

- *the n groups $U_k = \{\sigma \in \mathfrak{A}_n : \sigma(k) = k\}$, $k = 1, \ldots, n$, all of which are isomorphic to $\mathfrak{A}_{n-1}$,*
- *for $n = 6$ only, another conjugacy class of 6 subgroups of index 6 isomorphic to $\mathfrak{A}_5$.*

This shows that whenever we try to construct a geometric mean of n matrices by taking a limit processes, such as in the Ando–Li–Mathias approach, the isotropy groups of the starting means must contain $\mathfrak{A}_{n-1}$. On the other hand, by Theorem 3.5, we cannot generate means whose isotropy group contains $\mathfrak{A}_{n-1}$ by composition of simpler means; therefore, there is no simpler approach than that of building a mean of n matrices as a limit process of means of $n-1$ matrices (or at least quasi-means with $\operatorname{Stab} Q = \mathfrak{A}_{n-1}$, which makes little difference). This shows that the recursive approach of G^{ALM} and G^{BMP} cannot be simplified while still maintaining P3 (permutation invariance).

5. Computational issues and numerical experiments.

A faster mean of four matrices. The results we have exposed up to now are negative results, and they hold for $n > 4$. On the other hand, it turns out that for $n = 4$, since $\mathfrak{A}_n$ is not a simple group, there is the possibility of obtaining a mean that is computationally simpler than the ones in use. Such a mean is the one we described in Example 3.6. Let us take any mean of three elements (we shall use G_3^{BMP} here since it is the one with the best computational results); the new mean is therefore defined as

$$G_4^{NEW}(A, B, C, D) := G_3^{BMP}\left((A \# B) \# (C \# D), (A \# C) \# (B \# D),\right.$$
$$\left.(A \# D) \# (B \# C)\right). \quad (5.1)$$

Notice that only one limit process is needed to compute the mean; conversely, when computing G_4^{ALM} or G_4^{BMP} we are performing an iteration whose elements are computed by doing four additional limit processes; thus we may expect a large saving in the overall computational cost.

We may extend the definition recursively to $n > 4$ elements using the construction described in (1.6), but with G^{NEW} instead of G^{BMP}. The total computational cost, computed in the same fashion as for the ALM and BMP means, is $O(n! p_3 p_5 p_6 \ldots p_n m^3)$. Thus the undesirable dependence from $n!$ does not disappear; the new mean should only yield a saving measured by a multiplicative constant in the

ELA

Data set (number of matrices)	BMP mean	New mean
$NaClO_3$ (5)	1.3E+00	3.1E-01
Ammonium dihydrogen phosphate (4)	3.5E-01	3.9E-02
Potassium dihydrogen phosphate (4)	3.5E-01	3.9E-02
Quartz (6)	2.9E+01	6.7E+00
Rochelle salt (4)	6.0E-01	5.5E-02

TABLE 5.1
CPU times for the elasticity data sets

	BMP mean	New mean
Outer iterations ($n = 4$)	3	none
Inner iterations ($n = 3$)	4×2.0 (avg.) per outer iteration	2
Matrix square roots (`sqrtm`)	72	15
Matrix p-th roots (`rootm`)	84	6

TABLE 5.2
Number of inner and outer iterations needed, and number of matrix roots needed (ammonium dihydrogen phosphate)

complexity bound.

Benchmarks. We have implemented the original BMP algorithm and the new one described in the above section with MATLAB® and run some tests on the same set of examples used by Moakher [8] and Bini *et al.* [4]. It is an example deriving from physical experiments on elasticity. It consists of five sets of matrices to average, with n varying from 4 to 6, and 6×6 matrices split into smaller diagonal blocks.

For each of the five data sets, we have computed both the BMP and the new matrix mean. The CPU times are reported in Table 5.1. As a stopping criterion for the iterations, we used

$$\max_{i,j,k} \left| (A_i^{(h)})_{jk} - (A_i^{(h+1)})_{jk} \right| < 10^{-13}.$$

As we expected, our mean provides a substantial reduction of the CPU time which is roughly by an order of magnitude.

Following Bini *et al.* [4], we then focused on the second data set (ammonium dihydrogen phosphate) for a deeper analysis; we report in Table 5 the number of iterations and matrix roots needed in both computations.

The examples in these data sets are mainly composed of matrices very close to each other; we shall consider here instead an example of mean of four matrices whose

ELA

F. Poloni

	BMP mean	New mean
Outer iterations ($n = 4$)	4	none
Inner iterations ($n = 3$)	4×2.5 (avg.) per outer iteration	3
Matrix square roots (`sqrtm`)	120	18
Matrix p-th roots (`rootm`)	136	9

TABLE 5.3

Number of inner and outer iterations needed, and number of matrix roots needed

Operation	Result
$\left\|G^{BMP}(M^{-2}, M, M^2, M^3) - M\right\|_2$	4.0E-14
$\left\|G^{NEW}(M^{-2}, M, M^2, M^3) - M\right\|_2$	2.5E-14
$\left\|\det(G^{BMP}(A, B, C, D)) - (\det(A)\det(B)\det(C)\det(D))^{1/4}\right\|$	5.5E-13
$\left\|\det(G^{BMP}(A, B, C, D)) - (\det(A)\det(B)\det(C)\det(D))^{1/4}\right\|$	2.1E-13

TABLE 5.4

Accuracy tests

mutual distances are larger:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 100 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 20 & 0 & -10 \\ 0 & 20 & 0 \\ -10 & 0 & 20 \end{bmatrix}.$$

(5.2)

The results regarding these matrices are reported in Table 5.3.

Accuracy. It is not clear how to check the accuracy of a limit process yielding a matrix geometric mean, since the exact value of the mean is not known *a priori*, apart from the cases in which all the A_i commute. In those cases, P1 yields a compact expression for the result. So we cannot test accuracy in the general case; instead, we have focused on two special examples.

As a first accuracy experiment, we computed $G(M^{-2}, M, M^2, M^3) - M$, where M is taken as the first matrix of the second data set on elasticity; the result of this computation should be zero according to P1. As a second experiment, we tested the validity of P9 (determinant identity) on the means of the four matrices in (5.2). The results of both computations are reported in Table 5.4; the results are well within the errors permitted by the stopping criterion, and show that both algorithms can reach a satisfying precision.

$$\boxed{\textbf{ELA}}$$

6. Conclusions.

Research lines. The results of this paper show that, by combining existing matrix means, it is possible to create a new mean which is faster to compute than the existing ones. Moreover, we show that using only function compositions and limit processes with the existing proof strategies, it is not possible to achieve any further significant improvement with respect to the existing algorithms. In particular, the dependency from $n!$ cannot be removed. New attempts should focus on other aspects, such as:

- proving new "unexpected" algebraic relations involving the existing matrix means, which would allow to break out of the framework of Theorem 3.5– Theorem 4.1.
- introducing new kinds of matrix geometric means or quasi-means, different from the ones built using function composition and limits.
- proving that the Riemannian centroid (1.1) is a matrix mean in the sense of Ando–Li–Mathias (currently P4 is an open problem), or providing faster and reliable algorithms to compute it.

It is an interesting question whether it is possible to construct a quasi-mean whose isotropy group is exactly $\mathfrak{A}_n$.

REFERENCES

[1] T. Ando, Chi-Kwong Li, and Roy Mathias. Geometric means. *Linear Algebra Appl.*, 385:305–334, 2004.

[2] Rajendra Bhatia. *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.

[3] Rajendra Bhatia and John Holbrook. Riemannian geometry and matrix geometric means. *Linear Algebra Appl.*, 413(2-3):594–618, 2006.

[4] Dario A. Bini, Beatrice Meini, and Federico Poloni. An effective matrix geometric mean satisfying the Ando-Li-Mathias properties. *Math Comp.*, 2009. In press.

[5] John D. Dixon and Brian Mortimer. *Permutation groups*, volume 163 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[6] Yongdo Lim. On Ando-Li-Mathias geometric mean equations. *Linear Algebra Appl.*, 428(8-9):1767–1777, 2008.

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

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

[9] Miklós Pálfia. The Riemann barycenter computation and means of several matrices. *Int. J. Comput. Math. Sci.*, 3(3):128–133, 2009.