Annali della
Scuola Normale Superiore di Pisa
Classe di Scienze

Roger D. Nussbaum
Joel E. Cohen

The arithmetic-geometric mean and its generalizations for noncommuting linear operators

Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 4e série, tome 15, nº 2 (1988), p. 239-308

<http://www.numdam.org/item?id=ASNSP_1988_4_15_2_239_0>

The Arithmetic-Geometric Mean and Its Generalizations for Noncommuting Linear Operators

ROGER D. NUSSBAUM* - JOEL E. COHEN+

Introduction

The arithmetic-geometric mean (to be defined in a moment) is the limit of an iterative process that operates recursively on pairs of positive real numbers. For over two centuries, an enormous amount of effort by some great mathematicians has been devoted to understanding and to generalizing the arithmetic-geometric mean. There have been two simple reasons why all this attention has been devoted to what is in essence a very humble idea. First, the limit has an important meaning or use that a priori could hardly be suspected from the definition of the iterative process. Specifically, the limit can be used to compute elliptic integrals, which are of substantial mathematical and scientific interest. Second, the iterative process converges to its limit with exceptional rapidity (quadratically-also to be defined later), so that very few iterative steps are required to approximate the limit very closely.

A large classical literature concerns generalizations of the arithmetic-geometric mean, or what could be called means and their iterations (see [6]). This paper concerns extensions of the arithmetic-geometric mean and of the classical generalizations from the case that the variables are real numbers to the case that the variables are linear operators. As in the case of positive numbers we are interested in three questions (for each generalization): the existence of a limit, the speed of convergence to the limit, and possible explicit formulas for the limit (for example, in terms of elliptic integrals). Though the machinery we have developed and the results we have obtained are substantial, as witnessed by the length of this paper, our success in achieving all three aims is not complete. First, we prove the existence of a limit for all the iterations we consider formally here. But for other interesting iterations, which appear to be plausible operator-theoretic generalizations of the arithmetic-geometric mean for

* Partially supported by National Science Foundation grant DMS 85-03316.
+ Partially supported by National Science Foundation grant BSR 84-07461.
Pervenuto alla Redazione il 20 Giugno 1987.

positive numbers, we observe numerically an apparent convergence to a limit but are unable to explain the observation mathematically. Hence we do not believe we have the last word on the existence of limits for generalizations of the arithmetic-geometric mean to linear operators. Second, for simplicity we establish a quadratic rate of convergence only for the "monster algorithm" considered in Section 3 below, although we believe a similar analysis can be used to determine, in all the other examples we treat, whether convergence is linear or quadratic. Third, we interpret the limiting linear operator in terms of elliptic integrals only for a small subset of the iterations whose limits we prove to exist. Even classically, explicit integral formulas for limits of iterated means are known only for a few examples which are very close to the arithmetic-geometric mean. However, in our case (see Section 4), we give a family, indexed by real numbers $\lambda \geq 1$, of reasonable definitions of the arithmetic-geometric mean of two linear operators A and B; but we only obtain an explicit integral formula when $\lambda = 1$.

On balance, the results of this paper are largely foundational: we prove the existence of limits for a wide variety of operator-theoretic generalizations, many apparently new, of the arithmetic-geometric mean. Though our success in finding explicit integral formulas for the limits is limited, it is possible that these results, and future extensions, will prove practically important for numerical algorithms to compute functions of matrices that can be derived from matrix elliptic integrals.

We now sketch the arithmetic-geometric mean and our results more precisely.

If A and B are positive real numbers, define a map f by

$$f(A,B) = ((A+B)/2,\ (AB)^{1/2}). \tag{0.1}$$

If f^k denotes the kth iterate of f, it is not hard to prove that there is a positive number $M = M(A,B)$ such that

$$\lim_{k\uparrow\infty} f^k(A,B) = (M,M).$$

The number M is called the "arithmetic-geometric mean of A and B" or the "AGM of A and B." First Landen, then Lagrange and finally Gauss observed independently that

$$\int_0^{\pi/2} (A^2\cos^2\Theta + B^2\sin^2\Theta)^{-1/2}d\Theta = (\pi/2)(M(A,B))^{-1}.$$

Lagrange and Legendre used this observation to compute elliptic integrals. Historical references to this work and to some of Gauss' deeper work on the AGM can be found in [6] and [15].

An enormous literature concerning "means and their iterations" touches on a wide range of mathematics [6]. For examples, if A and B are positive reals, $0 < \alpha,\ \beta < 1$ and [11]

$$f(A,B) = ((1-\alpha)A + \alpha B,\ A^{1-\beta}B^{\beta}) \tag{0.2}$$

or $p > 0,\ q > 0$ and

$$f(A,B) = (([A^p + B^p]/2)^{1/p},\ ([A^q + B^q]/2)^{1/q}), \tag{0.3}$$

then

$$\lim_{k\uparrow\infty} f^k(A,B) = (M,M) \tag{0.4}$$

where M is a positive number depending on A and B, and $\alpha,\ \beta$ or $p,\ q$. One can also study functions f which are functions of m variables, $m > 2$, and try to prove analogues of (0.4). Borchardt [9] considered the map

$$f(A,B,C,D) = \tag{0.5}$$

$$((A+B+C+D)/4,\ ((AB)^{1/2} + (CD)^{1/2})/2,$$

$$((AC)^{1/2} + (BD)^{1/2})/2,\ ((AD)^{1/2} + (BC)^{1/2})/2)$$

for positive reals $A,\ B,\ C$ and D and proved (this is the easy part of his work) that

$$\lim_{k\uparrow\infty} f^k(A,B,C,D) = (M,M,M,M), \tag{0.6}$$

where M is a positive number depending on $A,\ B,\ C$ and D. Many other examples are mentioned in Section 2 below.

A first goal of this paper is to describe reasonable analogues of the AGM and its generalizations when all the variables are positive definite, bounded, self-adjoint linear operators on a Hilbert space. Abbreviating the phrase "positive definite, bounded, self-adjoint linear operator" to "positive definite operator", the first question is: what should be the analogue of $A^{1/2}B^{1/2}$ (for A and B positive reals) when A and B are positive definite operators? More generally, if $\sigma \in \mathbb{R}^m$ satisfies $\sigma_i > 0,\ 1 \le i \le m$, and $\sum_{i=1}^{m} \sigma_i = 1$ and $A_i,\ 1 \le i \le m$, are positive reals, what is a reasonable analogue of $A_1^{\sigma_1} A_2^{\sigma_2} \cdots A_m^{\sigma_m}$ when the variables $A_i,\ 1 \le i \le m$, are positive definite operators? We suggest that a reasonable analogue of $\prod_{i=1}^{m} A_i^{\sigma_i}$ is

$$\exp(\sum_{j=1}^{m} \sigma_j \log A_j). \tag{0.7}$$

If, for positive definite operators A_j, $1 \leq j \leq m$, and a real number $r \neq 0$ we define

$$M_{r\sigma}(A_1, A_2, \cdots, A_m) = (\sum_{j=1}^{m} \sigma_j A_j^r)^{1/r},$$

one can prove that

$$\lim_{r \to 0} M_{r\sigma}(A_1, A_2, \cdots, A_m) = \exp(\sum_{j=1}^{m} \sigma_j \log A_j)$$

in the operator norm topology, so our suggestion dovetails nicely with certain reasonable means.

Using (0.7), we give operator-valued analogues of maps f like those mentioned before, and we prove convergence of $f^k(A, B)$ in the strong operator topology. For example, a very special case of results in Section 2 is that if $0 < \alpha$, $\beta < 1$, and A and B are positive definite operators, and

$$f(A, B) = ((1 - \alpha)A + \alpha B, \ \exp((1 - \beta)\log A + \beta \log B)), \tag{0.8}$$

then there exists a positive definite operator E such that

$$\lim_{k \uparrow \infty} f^k(A, B) = (E, E).$$

The first two sections of this paper deal with the convergence of very general operator-valued versions of extensions of the AGM. In Section 1, we give results (Theorems 1.1 and 1.2) which enable one to prove convergence in the strong operator topology of certain sequences of n-tuples of positive definite linear operators. An example would be $(A_k, B_k) = f^k(A, B)$ with f as in (0.8). The key idea in Sections 1 and 2 is to exploit the concavity of certain maps $A \to g(A)$, for positive definite A, and to use the beautiful classical theory of Loewner. In the applications in Section 2, we use only the concavity of the maps $A \to \log A$ and $A \to A^p (0 < p < 1)$ and the convexity of $A \to A^{-1}$; the full Loewner machinery is not needed.

The arguments simplify in the case of finite-dimensional matrices. Theorem 1.2, in particular, is not needed in the finite-dimensional case.

In Section 2, we use the convergence results of Section 1 to prove operator-valued versions of convergence theorems for iterates of many classical means. Our convergence theorems suggest that our conventions were reasonable and provide an answer to the question raised in [6, p. 196] of how to extend the usual means to noncommuting variables. The maps f we consider are not usually order-preserving, so the general convergence results in Section 4 of [26] (see [27] for a summary) are not applicable.

In Section 3, we extend the domain of $M(A, B)$, the operator-valued AGM of A and B, to pairs of bounded linear operators which are not necessarily positive definite and self-adjoint, and prove that $(A, B) \to M(A, B)$ is analytic.

The analogues of these questions are considered for a more general "monster algorithm" introduced in [6]. We also consider the commutative case ($AB = BA$) and prove an integral formula for $M(A,B)$ analogous to that when A and B are real. The commutative case was also treated in [32], but the discussion there seems incomplete.

There are already numerous papers concerning operator-valued versions of the AGM and other means. Section 4 of our paper displays the connection between our operator-valued definition of the AGM and one introduced by Fujii [19] and Ando and Kubo [5]. We prove that the two definitions are in general different. However, there is a continuum of "reasonable" definitions of an AGM, parametrized by $\lambda \geq 1$, such that $\lambda = 1$ corresponds to that of Fujii-Ando-Kubo and $\lambda = \infty$ corresponds to ours. For each $\lambda \geq 1$ and each pair of positive definite operators A and B there exists in the limit a positive definite operator E_λ which is the AGM of A and B for the algorithm corresponding to λ, and generally $E_\lambda \neq E_\mu$ if $\lambda \neq \mu$.

1. - Convergence criteria for sequences of linear operators

We recall some standard notation and results. If X and Y are complex Banach spaces, we denote by $\mathcal{L}(X,Y)$ the set of bounded complex linear operators from X to Y; $X^* = \mathcal{L}(X,\mathbb{C})$ will denote the continuous complex linear maps from X to $\mathbb{C}$, the complex numbers. If $X = Y$, we shall write $\mathcal{L}(X)$ instead of $\mathcal{L}(X,X)$. $\mathcal{L}(X,Y)$ is a Banach space in the standard norm, $\|A\| = \sup\{\|Ax\| : x \in X \text{ and } \|x\| \leq 1\}$. If $A \in \mathcal{L}(X,X)$, $\sigma(A)$ will denote the spectrum of A, so

$$\sigma(A) = \{\lambda \in \mathbb{C} : \lambda - A \text{ is not one-one and onto}\}.$$

If D is an open neighborhood of $\sigma(A)$ and $f : D \to \mathbb{C}$ is analytic, we define $f(A)$ in terms of Cauchy's integral formula:

$$f(A) = (2\pi i)^{-1} \int_\Gamma f(z)(z-A)^{-1} dz, \tag{1.1}$$

where Γ is a finite union of simple, closed rectifiable curves in D which contains $\sigma(A)$ in its interior. An exposition of the basic results about this functional calculus can be found in [17], [33] or [36].

Recall that if $\mathcal{A}$ denotes the algebra of functions which are analytic on an open neighborhood of $\sigma(A)$, then the map $f \to f(A)$ defined by (1.1) is an algebra homomorphism and $\sigma(f(A)) = f(\sigma(A))$. If g is analytic on an open neighborhood of $\sigma(B)$, where $B = f(A)$, then $g(B) = (g \circ f)(A)$.

If X and Y are real Banach spaces, $\mathcal{L}(X,Y)$ denotes the bounded, real linear maps from X to Y. If $A \in \mathcal{L}(X)$ and $\tilde{X}$ denotes the complexification of X, then A can be extended uniquely to a complex linear map $\tilde{A} : \tilde{X} \to \tilde{X}$, and we define $\sigma(A)$, the spectrum of A, to be $\sigma(\tilde{A})$. If f is analytic on a neighborhood of $\sigma(\tilde{A})$ and $f(\overline{z}) = \overline{f(z)}$ (where $\overline{w}$ denotes the complex conjugate of w), then $f(\tilde{A})(X) \subseteq X$, so we define $f(A) = f(\tilde{A})|X$ in this case.

Aside from the norm topology on $\mathcal{L}(X,X)$, there are also locally convex topologies called the "strong operator topology" and the "weak operator topology" (see [17], Chapter 6). If (A_k) is a sequence of bounded linear operators in $\mathcal{L}(X,X)$ and $A \in \mathcal{L}(X,X)$, then (A_k) approaches A in the strong operator topology as $k \to \infty$ if, for all $x \in X$,

$$\lim_{k\uparrow\infty} \|A_k(x) - A(x)\| = 0,$$

and A_k approaches A in the weak operator topology if, for all $x \in X$ and $\Psi \in X^*$,

$$\lim_{k\uparrow\infty} \Psi(A_k(x) - A(x)) = 0.$$

If (A_k) is a sequence of bounded linear operators which approaches a bounded linear operator A in the weak operator topology, we shall write

$$w\text{-}\lim_{k\uparrow\infty} A_k = A \text{ or } A_k \rightharpoonup A.$$

Similarly, if (A_k) converges to A in the strong operator topology we shall write

$$s\text{-}\lim_{k\uparrow\infty} A_k = A \text{ or } A_k \to A,$$

and if $\lim_{k\uparrow\infty} \|A_k - A\| = 0$ we shall write

$$n\text{-}\lim_{k\uparrow\infty} A_k = A \text{ or } A_k \Rightarrow A.$$

In this paper we shall also deal with sequences $(A^{(k)})$ of ordered m-tuples of bounded linear operators, so

$$A^{(k)} = (A_1^{(k)}, A_2^{(k)}, \cdots, A_m^{(k)}),$$

where $A_j^{(k)} \in \mathcal{L}(X,Y)$ for $1 \le j \le m$. We shall say the $A^{(k)}$ converges in the strong operator topology to the m-tuple $A = (A_1, A_2, \cdots, A_m)$ and write $A^{(k)} \to A$ if

$$s\text{-}\lim_{k\uparrow\infty} A_j^{(k)} = A_j \quad \text{for } 1 \le j \le m.$$

Similarly, we shall write $A^{(k)} \rightharpoonup A$ if $A_j^{(k)} \rightharpoonup A_j$ for $1 \le j \le m$ and $A^{(k)} \Rightarrow A$ if $A_j^{(k)} \Rightarrow A$ for $1 \le j \le m$.

If H is a complex Hilbert space with inner product $< x, \ y >$ and $A \in \mathcal{L}(H)$, then A is self-adjoint if $< Ax, \ y >=< x, \ Ay >$ for all $x, \ y \in H$. If A is self-adjoint and $f : \sigma(A) \subseteq \mathbb{R} \to \mathbb{C}$ is a continuous map, one can define $f(A)$. This definition agrees with that in (1.1) when f is analytic on an open neighborhood of A. If $A \in \mathcal{L}(H)$ is self-adjoint, we shall say that A is "positive semidefinite" (sometimes called nonnegative definite) if

$$< Ax, x > \geq 0 \qquad \text{for all } x \in H \tag{1.2}$$

and A is "positive definite" if there exists $\varepsilon > 0$ such that

$$< Ax, x > \geq \varepsilon \qquad \text{for all } x \in H \text{ with } \|x\| = 1. \tag{1.3}$$

We abbreviate "positive semidefinite" as "p.s.d.", "nonnegative definite" as "n.n.d." and "positive definite" as "p.d.". Positive semidefiniteness induces a partial ordering on the set of self-adjoint operators $A \in \mathcal{L}(H)$: if A and B are bounded, self-adjoint operators, we write $A \leq B$ if $B - A$ is positive semidefinite.

Henceforth, whenever we say that $A \in \mathcal{L}(H)$ is positive definite or positive semidefinite, it will be assumed that A is self-adjoint.

If K denotes the set of bounded, self-adjoint p.s.d. linear maps in $\mathcal{L}(H)$, then K is an example of a cone (with vertex at 0) in a Banach space $Y = \mathcal{L}(H)$; and K^0, the interior of K, is the set of self-adjoint, p.d. operators in $\mathcal{L}(H)$. In general, if C is a subset of a Banach space Z, we say that C is a cone (with vertex at 0) if C is a closed, convex subset of Z and (a) if $x \in C$, then $tx \in C$ for all real numbers $t \geq 0$ and (b) if $x \in C - \{0\}$, then $-x \notin C$. A cone C induces a partial ordering on Z by $x \leq y$ if and only if $y - x \in C$. If D is a subset of a Banach space Z_1, C_1 is a cone in Z_1 and C_2 is a cone in a Banach space Z_2 and $f : D \to Z_2$ is a map, we say that f is order-preserving (with respect to the partial orderings induced in Z_j by C_j) if for all x and y in D such that $x \leq y$ (in the partial ordering induced by C_1) one has $f(x) \leq f(y)$ (in the partial ordering induced by C_2). Usually we shall have $Z_1 = Z_2$ and $C_1 = C_2$. If D is convex, we say that $f : D \to Z_2$ is "convex" (with respect to the partial ordering induced by C_2) if for all x and y in D and all real numbers t with $0 \leq t \leq 1$, one has

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y). \tag{1.4}$$

We shall say that f is strictly convex if f is convex and for all $x \neq y$ in D

$$f((1-t)x + ty) \neq (1-t)f(x) + tf(y) \qquad \text{for } 0 < t < 1.$$

We shall say that f is concave (strictly concave) if $-f$ is convex (strictly convex).

Our first lemma is well-known for real-valued functions. The proof in our generality is essentially the same and we omit it.

LEMMA 1.1. *Let D be a convex subset of a Banach space Z_1, C_2 a cone in a Banach space Z_2 and $f : D \to Z_2$ a map which is continuous on line segments in D (so the map $t \to f((1-t)x + ty)$, $0 \leq t \leq 1$, is continuous for* *all x and y in D). If, with respect to the partial ordering induced by C_2, one has*

$$f((1/2)x + (1/2)y) \leq (1/2)f(x) + (1/2)f(y) \tag{1.5}$$

for all x and y in D, then f is convex. If $f : D \to Z_2$ is convex and for all x and y in D with $x \neq y$ one has

$$f((1/2)x + (1/2)y) \neq (1/2)f(x) + (1/2)f(y)$$

then f is strictly convex. If s_j, $1 \leq j \leq n$, are nonnegative real numbers such that $\sum_{j=1}^{n} s_j = 1$, and x_j, $1 \leq j \leq n$, are any points in D and f is convex, then

$$f\Big(\sum_{j=1}^{n} s_j x_j\Big) \leq \sum_{j=1}^{n} s_j \; f(x_j). \tag{1.6}$$

If f is strictly convex and $s_j > 0$ for $1 \leq j \leq n$, then equality holds in (1.6) *if and only if all the points x_j are equal for $1 \leq j \leq n$.*

We shall eventually need some continuity results for the strong operator topology.

LEMMA 1.2. *Suppose that (L_k) is a sequence of bounded linear maps of a complex Banach space X to itself and that (L_k) converges to a bounded linear operator L in the strong operator topology. Assume that $\sigma(L_k) \subseteq B$ and $\sigma(L) \subseteq B$, where B is a compact subset of the complex numbers, that D is a bounded open neighborhood of B such that $\Gamma = \partial D$ consists of a finite number of simple, closed rectifiable curves and that f is a complex-valued function which is defined and analytic on an open neighborhood of $\overline{D}$. If there exists a constant M such that*

$$\|(\lambda - L_k)^{-1}\| \leq M \text{ for all } \lambda \in \Gamma \text{ and all } k \geq 1, \tag{1.7}$$

then $f(L_k) \to f(L)$. If $f(z)$ is an entire function, then $f(L_k) \to f(L)$. If X is a Hilbert space and all the operators L_k are normal (or self-adjoint), $f(L_k) \to f(L)$.

PROOF. For $\lambda \in \Gamma$ and a fixed $x \in X$ one has

$$(\lambda - L_k)^{-1}x - (\lambda - L)^{-1}x = (\lambda - L_k)^{-1}(L_k - L)(\lambda - L)^{-1}x.$$

Applying the estimate in (1.7) yields

$$\|(\lambda - L_k)^{-1}x - (\lambda - L)^{-1}x\| \leq M\|(L_k - L)(\lambda - L)^{-1}x\|. \tag{1.8}$$

The uniform boundedness principle implies that there is a constant M_1 such that $\|L_k - L\| \leq M_1$ for all $k \geq 1$. Since $\lambda \to (\lambda - L)^{-1}$ is continuous in the operator norm, (1.8) implies that there is a constant M_2, independent of $\lambda \in \Gamma$, such that

$$\|(\lambda - L_k)^{-1}x - (\lambda - L)^{-1}x\| \le M_2 \qquad \text{for all } \lambda \in \Gamma. \tag{1.9}$$

Inequality (1.8) and the fact that $(L_k) \to L$ imply that

$$\lim_{k\uparrow\infty} \|(\lambda - L_k)^{-1}x - (\lambda - L)^{-1}x\| = 0.$$

A version of the Lebesgue dominated convergence theorem now implies that

$$\lim_{k\uparrow\infty} (2\pi i)^{-1} \int_\Gamma f(\lambda)(\lambda - L_k)^{-1}x d\lambda = \lim_{k\uparrow\infty} f(L_k)x$$
$$= (2\pi i)^{-1} \int_\Gamma f(\lambda)(\lambda - L)^{-1}x d\lambda = f(L)x,$$

so $f(L_k) \to f(L)$.

If X is a Hilbert space, $\lambda \in \Gamma$ and L_k is normal, it follows that $\lambda - L_k$ and $(\lambda - L_k)^{-1}$ are normal, so (see [36])

$$\|(\lambda - L_k)^{-1}\| = \sup\{|(\lambda - z)^{-1}| : z \in \sigma(L_k)\}. \tag{1.10}$$

Because $\sigma(L_k) \subseteq B$ and Γ are disjoint compact sets, (1.10) implies that (1.7) is satisfied, so $f(L_k) \to f(L)$ by the first part of the lemma.

Finally, suppose X is a complex Banach space and f is entire. If $R = \sup\{\|L_k\| : k \ge 1\}$, we can take $D = \{z : |z| < 2R\}$ and for $\lambda \in \partial D$, we have

$$(\lambda - L_k)^{-1} = \lambda^{-1} \sum_{j=0}^{\infty} (\lambda^{-1} L_k)^j,$$

so

$$\|(\lambda - L_k)^{-1}\| \le |\lambda|^{-1} \sum_{j=0}^{\infty} |\lambda^{-j}| \, \|L_k\|^j \le 2|\lambda|^{-1} = 2R^{-1}.$$

Thus the first part of the lemma implies that $f(L_k) \to f(L)$ in this case also. □

REMARK 1.1. The obvious analogue of Lemma 1.2 for the weak operator topology is false. Let H be l^2 and let $\{e_j : j \ge 1\}$ be the standard orthonormal basis for l^2. For $n \ge 1$, define a self-adjoint operator $A_n : H \to H$ by

$$A_n(e_j) = e_{n-j+1} \quad \text{for } 1 \le j \le n \text{ and}$$
$$A_n(e_j) = 0 \quad \text{for } j > n.$$

One can easily prove that $(A_n) \rightharpoonup 0$, but $(A_n^2) \rightharpoonup I$, the identity.

Before stating our first theorem we recall some basic facts about matrices with nonnegative entries. If M is an $n \times n$ matrix all of whose entries are nonnegative, M is called "irreducible" if, for each ordered pair (i, j) with $1 \leq i,\ j \leq n$, there exists an integer $p \geq 1$ (possibly dependent on (i, j)) such that the entry in row i and column j of M^p is strictly positive. The matrix M is called "primitive" if there exists an integer $p \geq 1$ such that all entries of M^p are strictly positive. If M is an irreducible matrix with nonnegative entries and r denotes the spectral radius of M, then $r > 0$ and there exists a unique (within scalar multiples) column vector u such that all entries of u are positive and

$$Mu = ru. \tag{1.11}$$

If M is primitive and if one defines $M_1 = r^{-1}M$ (where r is the spectral radius of M), then for any nonzero vector x, all of whose components are nonnegative, one has

$$\lim_{k\uparrow\infty} M_1^k x = \alpha u, \tag{1.12}$$

where u is the eigenvector in (1.11) and α is a positive number depending on x.

If $M = (m_{ij})$ is a matrix with nonnegative entries, M is called "column-stochastic" if

$$\sum_{i=1}^{n} m_{ij} = 1 \quad \text{for } 1 \leq j \leq n,$$

and M is "row-stochastic" if

$$\sum_{j=1}^{n} m_{ij} = 1 \quad \text{for } 1 \leq i \leq n.$$

It is an elementary exercise in the theory of nonnegative matrices that the spectral radius of any column-stochastic (or row-stochastic) matrix equals one. Furthermore, a trivial argument shows that the product of column-stochastic (or row-stochastic) matrices is column-stochastic (or row-stochastic).

Now suppose that M is a column-stochastic, primitive matrix with nonnegative entries and let u be a column vector, all of whose entries are positive, such that

$$Mu = u.$$

If we normalize u by demanding

$$\sum_{j=1}^{n} u_j = 1,$$

we know that u is unique. Define M_∞ to be the $n \times n$ matrix all of whose columns equal u. If e_j, $1 \leq j \leq n$, denotes the standard orthonormal basis of $\mathbb{R}^n$, we know that $M^k e_j$ is the jth column of M^k, and because M^k is column-stochastic, (1.12) implies that

$$\lim_{k\uparrow\infty} M^k e_j = u, \ 1 \leq j \leq n.$$

We conclude from the previous equation that, if M is column-stochastic and primitive, then

$$\lim_{k\uparrow\infty} M^k = M_\infty. \tag{1.13}$$

If K is a cone in a Banach space X, let C denote the cone which is the n-fold Cartesian product of K. Let Y denote the n-fold Cartesian product of X with any of the standard norms. If M is an $n \times n$ matrix with nonnegative entries, M induces a bounded linear map W of Y to Y by

$$W(x) = xM = y, \text{ where}$$

$$x = (x_1, x_2, \cdots, x_n), \ y = (y_1, y_2, \cdots, y_n), \ y_j = \sum_{i=1}^{n} m_{ij} x_i.$$

It is easy to check that $W(C) \subseteq C$, that $W(C - \{0\}) \subseteq C - \{0\}$ if no row of M is identically zero, and that (if K^0 is nonempty) $W(C^0) \subseteq C^0$ if no column of M is identically zero.

We are in a position to state our first theorem. For simplicity, we restrict ourselves to the cone of p.s.d. bounded linear operators on a Hilbert space, but versions of the following theorem can be given for more general cones.

THEOREM 1.1. *Let K denote the cone of positive semidefinite, self-adjoint bounded linear operators on a Hilbert space H. Let C denote the n-fold Cartesian product of K, $C = K \times K \times \cdots \times K$. Let Y denote the n-fold Cartesian product of $X = \mathcal{L}(H, H)$ with itself, $Y = X \times X \times \cdots \times X$. Suppose that $f : C^0 \to C^0$ is a continuous map and $\phi : K^0 \to X$ is a continuous map and define $\Phi : C^0 \to Y$ by*

$$\Phi(A_1, A_2, \cdots, A_n) = (\phi(A_1), \phi(A_2), \cdots, \phi(A_n)).$$

Assume that for every $A \in C^0$ there exist $B \in C^0$ and positive reals α and β such that

$$\alpha B \leq f^j(A) \leq \beta B, \tag{1.14}$$

and

$$\Phi(f^j(A)) \leq \Phi(\beta B) \tag{1.15}$$

for all $j \geq 0$, where the partial ordering in (1.14) and (1.15) is induced by C and f^j is the jth iterate of f. Assume that there exists an $n \times n$, primitive, column-stochastic matrix M such that, for every $A \in C^0$,

$$\Phi(f(A)) \geq \Phi(A)M. \tag{1.16}$$

Let u denote the unique column vector in $\mathbb{R}^n$ such that all components of u_i of u are positive and

$$Mu = u$$

and

$$\sum_{i=1}^{n} u_i = 1,$$

and let π_i denote the projection of Y onto its ith coordinate. If, for $A \in C^0$, we define

$$f^k(A) = (A_1^{(k)}, A_2^{(k)}, \cdots, A_n^{(k)}),$$

so $A_i^{(k)} = \pi_i(f^k(A))$, there exists $E \in K^0$ such that

$$s\text{-}\lim_{k\uparrow\infty} \sum_{i=1}^{n} u_i \ \phi(A_i^{(k)}) = E. \tag{1.17}$$

Furthermore, for $1 \leq i \leq n$, one has

$$w\text{-}\lim_{k\uparrow\infty} \ \phi(A_i^{(k)}) = E. \tag{1.18}$$

PROOF. If u is the eigenvector in the statement of the theorem, define, for $A = (A_1, A_2, \cdots, A_n) \in C^0$, a function $\Psi : C^0 \to K^0$ by the formula

$$\Psi(A) = \sum_{j=1}^{n} u_j \phi(A_j). \tag{1.19}$$

Inequality (1.16) implies, if $A = (A_1, A_2, \cdots, A_n)$, that

$$\begin{aligned} \Psi(f(A)) &= \sum_{j=1}^{n} u_j \phi(\pi_j f(A)) \\ \geq \sum_{j=1}^{n} u_j (\sum_{i=1}^{n} m_{ij} \phi(A_i)) &= \sum_{i=1}^{n} \phi(A_i)(\sum_{j=1}^{n} m_{ij} u_j) \\ &= \sum_{i=1}^{n} u_i \phi(A_i) = \Psi(A). \end{aligned} \tag{1.20}$$

If we define $E_k = \Psi(f^k(A))$, $E_k - E_1$ is an increasing sequence of bounded, self-adjoint operators, and (1.14) and (1.15) imply that $E_k - E_1$ is bounded above (in the partial ordering on K). Thus as $k \to \infty$, $E_k \to E$ with $E - E_1$ self-adjoint. By iterating inequality (1.16) we see that, for any $k \geq 1$,

$$\Phi(f^k(A)) \geq \Phi(A)M^k. \tag{1.21}$$

The remarks preceding the theorem imply that given any $\varepsilon > 0$, there exists $N_1 \geq 1$ such that for all $k \geq N_1$,

$$(1+\varepsilon)M_\infty \geq M^k \geq (1-\varepsilon)M_\infty, \tag{1.22}$$

where M_∞ is the matrix with all columns equal to u, and (1.22) means that for $1 \leq i, j \leq n$, the i, j entry of M^k is greater than or equal to the i, j entry of $(1-\varepsilon)M_\infty$. If $k \geq N_1$, and $1 \leq i \leq n$, it follows that

$$\begin{aligned} \pi_i(\Phi(f^k(A))) &\geq \pi_i(\Phi(f^{k-N_1}(A))M^{N_1}) = \pi_i(\Phi(f^{k-N_1}(A))M_\infty) \\ &\quad + \pi_i(\Phi(f^{k-N_1}(A))(M^{N_1} - M_\infty)) \\ &= E_{k-N_1} + \pi_i(\Phi(f^{k-N_1}(A))(M^{N_1} - M_\infty)). \end{aligned} \tag{1.23}$$

For a given $x \in H$ and $\varepsilon > 0$ there exists N_2 such that

$$< E_j x, x > \;\geq\; < Ex, x > -\varepsilon \qquad \text{for } j \geq N_2 \tag{1.24}$$

because $E_j \to E$. Combining inequalities (1.23) and (1.24) yields for $k \geq N_1 + N_2$,

$$\begin{aligned} < \phi(A_i^{(k)}x),\; x > \;&\geq\; < Ex, x > -\varepsilon \\ &\quad + < \pi_i(\Phi(f^{k-N_1}(A))(M^{N_1} - M_\infty))x, x > \end{aligned} \tag{1.25}$$

which implies that (using inequality (1.22) and recalling that $\Phi(f^{k-N_1}(A))$ is bounded in norm)

$$\lim_{k\to\infty} \inf \; < \phi(A_i^{(k)})x,\; x > \;\geq\; < Ex,\; x > . \tag{1.26}$$

If, for some $x \in H$ and some i, one has

$$\lim_{k\to\infty} \sup \; < \phi(A_i^{(k)}x),\; x > \;>\; < Ex,\; x >, \tag{1.27}$$

then inequalities (1.26) and (1.27) imply

$$\lim_{k\to\infty} \sup \; < E_k x,\; x > = \lim_{k\to\infty} \sup \sum_{j=1}^{n} u_j < \phi(A_i^{(k)})x,\; x > \;>\; < Ex,\; x > .$$

The above inequality contradicts the fact that

$$\lim_{k\uparrow\infty} \; < E_k x,\; x > \;=\; < Ex,\; x >,$$

so inequality (1.27) must be false. We conclude from (1.26) that

$$\lim_{k\uparrow\infty} < \phi(A_i^{(k)})x,\ x > \ = \ < Ex,\ x > . \tag{1.28}$$

Standard arguments using the polarization now imply that

$$w\text{-}\lim_{k\uparrow\infty} \phi(A_i^{(k)}) = E. \ \square$$

There are several obstacles to using Theorem 1.1. The first problem, of course, is to prove the existence of ϕ and M as in Theorem 1.1 for examples of interest to us. We shall use the classical results of C. Loewner concerning the concavity of order-preserving maps from the cone K of p.s.d. bounded linear operators of a Hilbert space H to $\mathcal{L}(H)$.

However, even assuming that we can establish the hypotheses of Theorem 1.1 in examples of interest, Theorem 1.1 provides inadequate information when H is infinite dimensional. If H is finite dimensional, the weak, strong and operator norm topologies on $\mathcal{L}(H)$ are identical, so Theorem 1.1 implies

$$\lim_{k\uparrow\infty} \|\phi(A_i^{(k)}) - E\| = 0 \qquad \text{for } 1 \le i \le n.$$

If ϕ is one-one with norm-continuous inverse, one concludes that

$$\lim_{k\uparrow\infty} \|A_i^{(k)} - \phi^{-1}(E)\| = 0.$$

If H is infinite dimensional, one would hope that there exists a p.d. bounded linear operator G such that

$$s\text{-}\lim_{k\uparrow\infty} A_i^{(k)} = G \qquad \text{for } 1 \le i \le n. \tag{1.29}$$

However, as Remark 1.1 shows, the weak operator convergence in (1.18) may be very far from the strong operator convergence hoped for in (1.29).

We now begin to address there deficiencies.

If K is the cone of p.s.d. bounded linear operators in $\mathcal{L}(H)$ and H is a Hilbert space, we need to know when certain maps defined on $D = K^0$ are order-preserving or concave or strictly concave. The first lemma is Loewner's theorem [23, 24] concerning order-preserving maps on K^0; an exposition of Loewner's theory is given in [16].

LEMMA 1.3. (C. Loewner) *Let K denote the cone of positive semidefinite, bounded linear maps of a Hilbert space H to itself. Suppose that $f : (0,\infty) \to \mathbb{R}$ is a continuous, real-valued map and that f has an analytic extension to $U = \{z \in \mathbb{C} : \mathrm{Im}(z) \neq 0$ or $(\mathrm{Im}(z) = 0$ and $\mathrm{Re}(z) > 0)\}$ such that $\mathrm{Im}(f(z)) > 0$ for all z such that $\mathrm{Im}(z) > 0$. Then the map $A \to f(A)$ for $A \in D = K^0$ is order-preserving with respect to the partial order induced by K.*

Functions that satisfy the hypotheses of Lemma 1.2 are $f(z) = \log(z)$ and $f(z) = z^p$, $0 < p \leq 1$, and $\log(z)$ will be the example of most interest here. One can give a simple and self-contained proof that if $A, B \in K^0$ and $A \leq B$, then $B^{-1} \leq A^{-1}$ (although we shall not do so). Using this fact, we now prove directly that the maps $A \to \log(A)$ and $A \to A^p$, $0 < p \leq 1$, are order-preserving on K^0. If $E \in K^0$ and $E_t \equiv (1-t)I + tE$, where I denotes the identity map, then one can easily prove that

$$\log(E) = \int_0^1 (d/dt)\ \log(E_t)dt = \int_0^1 E_t^{-1}(E - I)dt. \tag{1.30}$$

An algebraic manipulation gives

$$\log(E) = \int_0^1 [t^{-1}I - t^{-1}E_t^{-1}]dt. \tag{1.31}$$

If A and B are in K^0 and $A \leq B$ and $A_t = (1-t)I + tA$ and $B_t = (1-t)I + tB$ for $0 \leq t \leq 1$, then $A_t \leq B_t$ for $0 \leq t \leq 1$, so $t^{-1}A_t^{-1} \geq t^{-1}B_t^{-1}$ for $0 < t \leq 1$ and

$$t^{-1}I - t^{-1}A_t^{-1} \leq t^{-1}I - t^{-1}B_t^{-1}, \text{ for } 0 < t \leq 1. \tag{1.32}$$

Using (1.31) and (1.32) one finds that $\log(A) \leq \log(B)$. That $A^p \leq B^p$ if $A \leq B$ and $0 < p < 1$ follows by a similar argument from the formula (see [21], p. 286)

$$E^p = [\sin(\pi p)/\pi] \int_0^\infty \lambda^{p-1}E(\lambda + E)^{-1}d\lambda, \quad E \in K^0, \quad 0 < p < 1. \tag{1.33}$$

We also need concavity results for maps of K^0 to $\mathcal{L}(H)$.

LEMMA 1.4. (Ando [3]) *Suppose that K, H and f are as in Lemma* 1.3. *Then, for $A \in K^0$, the map $A \to f(A)$ is concave and the map $A \to Af(A)$ is convex* (*with respect to the partial ordering from K*).

Ando states Lemma 1.4 for finite-dimensional Hilbert spaces (Theorem 4 in [3]), but the same argument, based on Loewner's theory, works for general Hilbert spaces.

Lemma 1.4 is not quite adeguate for our purposes. We need *strict* concavity and convexity results, and in fact we shall need a property (see Theorem 1.2 below) analogous to the property of uniform convexity for norms. Such results will follow from the strict convexity of the map $A \to A^{-1}$ for $A \in K^0$, and this strict convexity was proved independently by P. Whittle [34, Lemma 1] and I. Olkin and J. Pratt [29]. Whittle's lemma is stated for finite-dimensional Hilbert spaces, but the proof applies in general and yields the following lemma.

LEMMMA 1.5. ([34] *and* [29]) *Let* H *be a Banach space and suppose that* $A, B \in \mathcal{L}(H)$. *If* λ *is a real number such that* $0 < \lambda < 1$ *and* A, B *and* $(1-\lambda)A + \lambda B$ *are one-one and onto, then*

$$(1.34)\qquad \begin{aligned} &(1-\lambda)A^{-1} + \lambda B^{-1} - ((1-\lambda)A + \lambda B)^{-1} = \\ &\lambda(1-\lambda)(A^{-1} - B^{-1})(\lambda A^{-1} + (1-\lambda)B^{-1})^{-1}(A^{-1} - B^{-1}). \end{aligned}$$

If H is a Hilbert space and K denotes the cone of positive semidefinite operators in $\mathcal{L}(H)$, then the map $A \to A^{-1}$ of K^0 to K^0 is strictly convex.

By exploiting Lemma 1.5 we obtain the following sharpening of Ando's theorem (Lemma 1.4). See also Bendat and Sherman [8].

LEMMA 1.6. *Let* f, K *and* H *be as in Lemma* 1.3. *If there do not exist real constants* α *and* β *such that* $f(x) = \alpha + \beta x$ *for all* $x > 0$, *then for* $A \in K^0$, *the map* $A \to f(A)$ *is strictly concave. If there do not exist real constants* α *and* γ *such that* $f(x) = \alpha + \gamma x^{-1}$ *for all* $x > 0$, *then* $A \to Af(A)$ *is strictly convex on* K^0.

PROOF. Loewner's theory implies that for λ not a negative real, $\lambda \neq 0$,

$$(1.35)\qquad f(\lambda) = \alpha + \beta\lambda + \int_{-\infty}^{0} (1 + \lambda t)(t - \lambda)^{-1} d\mu(t)$$

where α is real, $\beta \geq 0$ and μ is a finite nonnegative Borel measure with support in $(-\infty, 0]$. If we assume that $f(x)$ is not an affine map, then μ is not the zero measure. From (1.35) and the identity

$$(I + tA)(tI - A)^{-1} = -tI - (1 + t^2)(A - tI)^{-1},$$

we obtain

$$(1.36)\qquad f(A) = \alpha I + \beta A + \int_{-\infty}^{0} [-tI - (1 + t^2)(A - tI)^{-1}] d\mu(t).$$

The map $A \to \alpha I + \beta A$ is obviously concave and Lemma 1.5 implies that for each $t \leq 0$ and for $A \in K^0$ the map

$$A \to -tI - (1 + t^2)(A - tI)^{-1}$$

is strictly concave so (1.36) implies $A \to f(A)$ is concave (Lemma 1.4). To prove that $A \to f(A)$ is strictly concave, take A, $B \in K^0$ with $A \neq B$ and define

$$E = (A + B)/2.$$

Equation (1.36) gives

$$(1/2)f(A) + (1/2)f(B) - f(E)$$

(1.37)
$$= \int_{-\infty}^{0} [-(1/2)(A - tI)^{-1} - (1/2)(B - tI)^{-1}$$
$$+ (E - tI)^{-1}](1 + t^2)d\mu(t).$$

The strict concavity of $A \to -A^{-1}$ implies that there exist $t_0 \leq 0$, with t_0 in the support of μ, $x \in H$ and $\delta > 0$ such that

(1.38)
$$< (-1/2)(A - tI)^{-1}x - (1/2)(B - tI)^{-1}x + (E - tI)^{-1}x,\ x > \leq 0$$

for all $t \leq 0$ and strict inequality holds in inequality (1.38) for $|t - t_0| \leq \delta$. Using this information in (1.37) we find

$$< (1/2)f(A)x + (1/2)f(B)x - f(E)x,\ x > < 0$$

which proves strict concavity.

Starting from (1.36) we see that

(1.39)
$$Af(A) = \alpha A + \beta A^2 + \int_{-\infty}^{0} [-tA - (1 + t^2)A(A - tI)^{-1}]d\mu(t).$$

It is easy to prove directly that $A \to A^2$ is strictly convex on K. Therefore (because $\beta \geq 0$) the map $A \to \alpha A + \beta A^2$ is convex on K and strictly convex if $\beta > 0$. Some algebraic manipulation shows that

$$-tA - (1 + t^2)A(A - tI)^{-1} = -tA - (1 + t^2)I + |t|(1 + t^2)(A - tI)^{-1}$$

for $t \leq 0$. Lemma 1.5 implies that for each $t \leq 0$, the map

$$A \to -tA - (1 + t^2)I + |t|(1 + t^2)(A - tI)^{-1}$$

is convex and, in fact, strictly convex if $t < 0$. If the support of μ has nonempty intersection with $(-\infty, 0)$, the same kind of proof as used before shows that $A \to Af(A)$ is strictly convex $(A \in K^0)$. If the support of μ is $\{0\}$, (1.35) implies that

$$f(\lambda) = \alpha + \beta\lambda - \gamma\lambda^{-1}$$

for some $\gamma \geq 0$. If $\beta > 0$, the map $A \to Af(A)$ is given by $A \to \alpha A + \beta A^2 - \gamma I$ in this case and hence is strictly convex. If $\beta = 0$, $f(\lambda) = \alpha - \gamma\lambda^{-1}$, contrary to our assumption. □

When $f(z) = \log(z)$ or z^p, $0 < p < 1$, Lemma 1.6 can be proved directly by using (1.31) and (1.33) and Lemma 1.5.

An immediate consequence of Lemmas 1.1 and 1.6 is:

COROLLARY 1.1. *Let H be a Hilbert space. Suppose that A_j, $1 \leq j \leq m$, are bounded, self-adjoint, p.d. linear maps of H to H. If s_j, $1 \leq j \leq m$, are positive numbers such that $\sum_{j=1}^{m} s_j = 1$, it follows that*

$$\log\Big(\sum_{j=1}^{m} s_j A_j\Big) \geq \sum_{j=1}^{m} s_j \log(A_j). \tag{1.40}$$

Equality holds in (1.40) *if and only if all the operators A_j are equal. If $0 < \alpha < 1$,*

$$\Big(\sum_{j=1}^{m} s_j A_j\Big)^{\alpha} \geq \sum_{j=1}^{m} s_j A_j^{\alpha}. \tag{1.41}$$

Equality holds in (1.41) *if and only if all the operators A_j are equal.*

REMARK 1.2. Suppose that A_j and s_j are as above and that p and q are real numbers such that $1 \leq p \leq q$. Defining $\alpha = pq^{-1}$ and $B_j = A_j^q$, Corollary 1.1 implies

$$\Big(\sum_{j=1}^{m} s_j B_j\Big)^{\alpha} = \Big(\sum_{j=1}^{m} s_j A_j^q\Big)^{\alpha} \geq \sum_{j=1}^{m} s_j B_j^{\alpha} = \sum_{j=1}^{m} s_j A_j^p. \tag{1.42}$$

One only needs $0 < p < q$ to derive inequality (1.42). Since $p \geq 1$, Lemma 1.3 implies that $B \to B^{1/p}$ is order-preserving on K, so one obtains from (1.42)

$$\Big(\sum_{j=1}^{m} s_j A_j^q\Big)^{1/q} \geq \Big(\sum_{j=1}^{m} s_j A_j^p\Big)^{1/p}. \tag{1.43}$$

For positive real A_j, inequality (1.43) is a classical result [20].

Unfortunately, when H is infinite dimensional Lemma 1.6 is inadequate for our purposes. We need to exploit strict concavity and strict convexity in a more quantitative way, analogous to the idea of uniform convexity for a norm. The next lemma illustrates the sort of uniform convexity we need for the case of the strictly convex map $A \to A^{-1}$ when A is positive definite.

LEMMA 1.7. *Suppose that A_i, $1 \leq i \leq m$, are p.d., bounded linear maps of a Hilbert space H into itself and $\alpha I \leq A_i \leq \beta I$ for $1 \leq i \leq m$, where α and β are positive reals. Assume that σ_k, $1 \leq k \leq m$, are positive reals such that $\sum_{k=1}^{m} \sigma_k = 1$. Then, for $1 \leq i,\ j \leq m$,*

$$\sum_{k=1}^{m} \sigma_k A_k^{-1} - \Big(\sum_{k=1}^{m} \sigma_k A_k\Big)^{-1} \geq \sigma_i \sigma_j (\sigma_i + \sigma_j)^{-1} \alpha (A_i^{-1} - A_j^{-1})^2, \tag{1.44}$$

where the inequality refers to the partial ordering induced by the cone of positive semidefinite bounded linear maps.

PROOF. Because $A \to A^{-1}$ is convex, the left side of (1.44) is always greater than or equal to zero. To prove (1.44) it suffices (by relabelling) to prove it when $i = 1$, $j = 2$.

If $A \geq \alpha I$ and $B \geq \alpha I$, then the spectral mapping theorem implies $A^{-1} \leq \alpha^{-1} I$, $B^{-1} \leq \alpha^{-1} I$ and

$$\lambda A^{-1} + (1-\lambda)B^{-1} \leq \alpha^{-1} I \text{ and } (\lambda A^{-1} + (1-\lambda)B^{-1})^{-1} \geq \alpha I. \tag{1.45}$$

Using inequality (1.45) in (1.34) gives

$$(1-\lambda)A^{-1} + \lambda B^{-1} - ((1-\lambda)A + \lambda B)^{-1} \geq \alpha\lambda(1-\lambda)(A^{-1} - B^{-1})^2. \tag{1.46}$$

If we define $\lambda = \sigma_2/(\sigma_1 + \sigma_2)$, $A_1 = A$ and $A_2 = B$, inequality (1.46) gives

$$(1-\lambda)A_1^{-1} + \lambda A_2^{-1} - ((1-\lambda)A_1 + \lambda A_2)^{-1} \geq \alpha\lambda(1-\lambda)(A_1^{-1} - A_2^{-1})^2 \equiv R. \tag{1.47}$$

If we define $B_1 = (1-\lambda)A_1 + \lambda A_2$ with $\lambda = \sigma_2(\sigma_1 + \sigma_2)^{-1}$, inequality (1.47) and the convexity of $A \to A^{-1}$ give

$$\begin{aligned} \sum_{j=1}^{m} \sigma_j A_j^{-1} - \Big(\sum_{j=1}^{m} \sigma_j A_j\Big)^{-1} &\geq (\sigma_1 + \sigma_2)B_1^{-1} + \sum_{j=3}^{m} \sigma_j A_j^{-1} \\ &\quad - \Big[(\sigma_1 + \sigma_2)B_1 + \sum_{j=3}^{m} \sigma_j A_j\Big]^{-1} + (\sigma_1 + \sigma_2)R \geq (\sigma_1 + \sigma_2)R. \end{aligned} \tag{1.48}$$

Inequality (1.48) is precisely the statement of the lemma for $i = 1$ and $j = 2$. □

The next theorem, when combined with Theorem 1.1, will enable us to prove convergence in the strong operator topology.

THEOREM 1.2. *Let K be the cone of positive semidefinite bounded linear maps of a Hilbert space H into itself. Let C denote the m-fold Cartesian product of K with itself. Thus $C \subseteq Y$, where Y is the m-fold Cartesian product of $X = \mathcal{L}(H)$ with itself. Assume that $(B^{(k)})$, $k \geq 0$, is a sequence in C^0, and write $B^{(k)} = (B_1^{(k)}, B_2^{(k)}, \cdots, B_m^{(k)})$. Suppose that $\phi : (0, \infty) \to \mathbb{R}$ is a C^1, real-valued function such that $\lim_{x \to +\infty} d\phi(x)/dx = 0$ and such that ϕ has an analytic extension to $U = \{z \in \mathbb{C} : \mathrm{Im}(z) \neq 0$ or $(\mathrm{Im}(z) = 0$ and $\mathrm{Re}(z) > 0)\}$ and $\mathrm{Im}(\phi(z)) > 0$ for all z such that $\mathrm{Im}(z) > 0$. Assume that there exist positive numbers λ_p, $1 \leq p \leq m$, such that*

$$w\text{-}\lim_{k \uparrow \infty} \Big[\phi\Big(\sum_{p=1}^{m} \lambda_p B_p^{(k)}\Big) - \sum_{p=1}^{m} \lambda_p \phi(B_p^{(k)})\Big] = 0, \tag{1.49}$$

and that there are positive numbers α *and* β *such that*

$$\alpha I \leq B_p^{(k)} \leq \beta I \qquad \text{for } k \geq 1,\ 1 \leq p \leq m. \tag{1.50}$$

Then for any i *and* j *with* $1 \leq i, j \leq m$,

$$s\text{-}\lim_{k\uparrow\infty} [\phi(B_i^{(k)}) - \phi(B_j^{(k)})] = 0. \tag{1.51}$$

If there exist positive numbers u_i, $1 \leq i \leq m$, *and* $E \in K^0$ *such that* $\sum_{i=1}^m u_i = 1$ *and*

$$s\text{-}\lim_{k\uparrow\infty} \sum_{i=1}^m u_i \phi(B_i^{(k)}) = E,$$

then

$$s\text{-}\lim_{k\uparrow\infty} \phi(B_i^{(k)}) = E.$$

If the restriction of ϕ *to an open neighborhood of* $(0,\infty)$ *in* $\mathbb{C}$ *is one-one, then*

$$s\text{-}\lim_{k\uparrow\infty} B_i^{(k)} = \phi^{-1}(E).$$

PROOF. Loewner's theory (see [16]) implies that if $\lambda \neq 0$ and λ is not a negative real, then

$$\phi(\lambda) = \alpha_1 + \beta_1\lambda + \int_{-\infty}^{0} [-t - (1+t^2)(\lambda - t)^{-1}] d\mu(t), \tag{1.52}$$

where α_1 is real, $\beta_1 \geq 0$ and μ is a nonnegative, finite Borel measure. Using this formula one easily proves that the condition $\lim_{x\to\infty} d\phi(x)/dx = 0$ implies that $\beta_1 = 0$.

We claim first that for all i and j, $1 \leq i, j \leq m$, one has

$$s\text{-}\lim_{k\uparrow\infty} \phi(B_i^{(k)}) - \phi(B_j^{(k)}) = 0.$$

We shall prove this for $i = 1$ and $j = 2$, since the general argument is the same. We obtain from (1.52) that

$$\phi(\sum_{p=1}^m \lambda_p B_p^{(k)}) - \sum_{p=1}^m \lambda_p \phi(B_p^{(k)}) = \tag{1.53}$$

$$\int_{-\infty}^{0} (1+t^2)\{\sum_{p=1}^m \lambda_p (B_p^{(k)} - tI)^{-1} - (\sum_{p=1}^m \lambda_p (B_p^{(k)} - tI))^{-1}\}\mu(dt).$$

Inequality (1.50) implies that

$$B_p^{(k)} - tI \geq (\alpha - t)I,$$

so Lemma 1.7 and (1.53) give

$$\phi(\sum_{p=1}^{m} \lambda_p B_p^{(k)}) - \sum_{p=1}^{m} \lambda_p \phi(B_p^{(k)}) \geq \tag{1.54}$$

$$\lambda_1\lambda_2(\lambda_1+\lambda_2)^{-1} \int_{-\infty}^{0} (1+t^2)(\alpha - t)[(B_1^{(k)} - tI)^{-1} - (B_2^{(k)} - tI)^{-1}]^2 \mu(dt).$$

The left side of (1.54) is assumed to approach 0 in the weak operator topology as $k \to \infty$. Hence, for any $x \in H$,

$$\lim_{k\uparrow\infty} \int_{-\infty}^{0} (1+t^2)(\alpha - t)\|(B_1^{(k)} - tI)^{-1}x - (B_2^{(k)} - tI)^{-1}x\|^2 \mu(dt) = 0. \tag{1.55}$$

Now we return to (1.52). Using for the first time the fact that $\beta_1 = 0$, we obtain

$$\phi(B_1^{(k)})(x) - \phi(B_2^{(k)})(x) \tag{1.56}$$

$$= \int_{-\infty}^{0} (1+t^2)[(B_2^{(k)} - tI)^{-1}x - (B_1^{(k)} - tI)^{-1}x]\mu(dt).$$

Because $B_i^{(k)} \geq \alpha I$ for $k \geq 1$ and $1 \leq i \leq m$, it is easy to see that there exists M_1 such that

$$\|(B_i^{(k)} - tI)^{-1}\| \leq M_1|t|^{-1} \qquad \text{for all } t \leq -1,\ k \geq 1,\ 1 \leq i \leq m.$$

It follows that for $t \leq -1$

$$\|(B_1^{(k)} - tI)^{-1} - (B_2^{(k)} - tI)^{-1}\| = \tag{1.57}$$

$$\|(B_1^{(k)} - tI)^{-1}(B_2^{(k)})(B_2^{(k)} - tI)^{-1}\| \leq (2\beta)M_1^2 t^{-2}.$$

If we use this estimate in (1.56) and recall that μ is a finite measure, we find that for any $\varepsilon > 0$ and any $x \in H$, there exists a constant M depending only on ε, β, $\|x\|$ and μ such that for all $k \geq 1$

$$\int_{-\infty}^{-M} (1+t^2)\|(B_1^{(k)} - tI)^{-1}x - (B_2^{(k)} - tI)^{-1}x\|\mu(dt) < \varepsilon/2. \tag{1.58}$$

On the other hand, the Cauchy-Schwartz inequality implies that

$$\int_{-M}^{0} (1+t^2)\|(B_1^{(k)} - tI)^{-1}x - (B_2^{(k)} - tI)^{-1}x\|\mu(dt) \leq \tag{1.59}$$

$$\{\int_{-M}^{0} (1+t^2)\|(B_1^{(k)} - tI)^{-1}x - (B_2^{(k)} - tI)^{-1}x\|^2\mu(dt)\}^{1/2}(\int_{-M}^{0} (1+t^2)\mu(dt))^{1/2}.$$

(1.55) implies that, for fixed $x \in H$, the right side of (1.59) approaches zero as $k \to \infty$ and hence is less than $\varepsilon/2$ for k sufficiently large. Combining inequalities (1.58) and (1.59) gives, for k sufficiently large,

$$\int_{-\infty}^{0} (1+t^2)\|(B_1^{(k)} - tI)^{-1}x - (B_2^{(k)} - tI)^{-1}x\|\mu(dt) < \varepsilon.$$

Using (1.56), we see that

$$\lim_{k\uparrow\infty} \|\phi(B_1^{(k)})x - \phi(B_2^{(k)})x\| = 0.$$

As already remarked, the same argument shows that $\phi(B_i^{(k)}) - \phi(B_j^{(k)}) \to 0$ for any i and j.

Suppose now that there exist positive reals $u_1, u_2, \cdots, u_m$ such that $\sum_{p=1}^{m} u_p = 1$ and

$$s\text{-}\lim_{k\uparrow\infty} \sum_{p=1}^{m} u_p\phi(B_p^{(k)}) = E. \tag{1.60}$$

For any fixed i, $1 \leq i \leq m$, (1.60) can be rewritten as

$$s\text{-}\lim_{k\uparrow\infty}[\sum_{p=1}^{m} u_p\phi(B_i^{(k)}) + \sum_{p=1}^{m} u_p(\phi(B_p^{(k)}) - \phi(B_i^{(k)}))] = E,$$

which implies that

$$s\text{-}\lim_{k\uparrow\infty} \phi(B_i^{(k)}) = s\text{-}\lim_{k\uparrow\infty} \sum_{p=1}^{m} u_p\phi(B_i^{(k)}) = E.$$

Finally, if ϕ is one-one an open neighborhood of $(0,\infty)$ in $\mathbb{C}$, then ϕ^{-1} is defined and analytic on an open neighborhood of $\phi([\alpha, \beta])$ in $\mathbb{C}$. Lemma 1.2 implies that

$$s\text{-}\lim_{k\uparrow\infty} \phi^{-1}(\phi(B_i^{(k)})) = s\text{-}\lim_{k\uparrow\infty} B_i^{(k)} = \phi^{-1}(E). \ \square$$

REMARK 1.3. The functions $\phi(z) = \log(z)$ and $\phi(z) = z^p$, $0 < p < 1$, satisfy the hypotheses of Theorem 1.2.

REMARK 1.4. Suppose that K and H are as in Theorem 1.2, that f is as in Lemma 1.3 and that there do not exist real constants α and β such that $f(x) = \alpha + \beta x$ for all $x > 0$. Then Lemma 1.6 implies that $A \to f(A)$ is a strictly concave map from K^0 to $\mathcal{L}(H)$. Define $f = \phi$ and assume that (1.49) and (1.50) are satisfied. if H is finite dimensional, it follows from the strict concavity of f and a simple compactness argument that for all i and j, $1 \leq i, j \leq m$,

$$\lim_{k \uparrow \infty} \|B_i^{(k)} - B_j^{(k)}\| = 0.$$

Thus, if H is finite dimensional, Theorem 1.2 follows trivially from Lemma 1.6. Theorem 1.2 only provides new information when H is infinite dimensional.

Theorems 1.1 and 1.2 are examples of convergence results in particular cones. However, one can give versions of Theorem 1.1 which are valid for general classes of cones. Since we shall not use such results, we shall not prove them here, but it may be of interest to state the theorems.

THEOREM 1.3. *Let K be a cone with nonempty interior in a finite-dimensional Banach space X. Let C denote the n-fold Cartesian product of K. Let Y denote the n-fold Cartesian product of X. Suppose that $f : C^0 \to C^0$ and $\phi : K^0 \to X$ are continuous maps. For any $y \in C^0$ assume that there exists $z \in C^0$ (dependent on y) and positive constants α and β such that*

$$\alpha z \leq f^j(y) \leq \beta z, \qquad \textit{and}$$
$$\Phi(f^j(y)) \leq \Phi(\beta z)$$

for all $j \geq 0$, where $\Phi : C^0 \to Y$ is defined by $\Phi(x_1, x_2, \cdots, x_n) = (\phi(x_1), \phi(x_2), \cdots, \phi(x_n))$.

Assume that M is an $n \times n$, primitive column-stochastic matrix with nonnegative entries such that

$$\Phi(f(x)) \geq \Phi(x)M \qquad \textit{for all } x \in C^0.$$

Let u be the unique positive column vector such that $\sum_{i=1}^{n} u_i = 1$ and $Mu = u$. If, for a given $x \in C^0$, we write

$$f^k(x) = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)}),$$

then there exists $w \in X$, w dependent on x, such that

$$\lim_{k \uparrow \infty} \phi(x_i^{(k)}) = w \qquad \textit{for } 1 \leq i \leq n.$$

If ϕ is one-one, then $w \in \phi(K^0)$ and

$$\lim_{k\uparrow\infty} x_i^{(k)} = \phi^{-1}(w) \quad \text{for } 1 \le i \le n.$$

The proof of Theorem 1.3 is basically the same as that of Theorem 1.1. One uses finite dimensionality to insure that the cone is "normal", i.e. that for each $w \in K$, the set $\{v \in K : v \le w\}$ is bounded in norm. One also uses finite dimensionality to guarantee that for every $y \in C^0$, $\{f^j(y) : j \ge 1\}$ is precompact.

To generalize Theorem 1.3 to certain infinite dimensional cones, we need some terminology. Suppose that $\mathcal{T}$ is a topology on a Banach space X and that X becomes a Hausdorff, locally convex topological vector space in the $\mathcal{T}$ topology and the $\mathcal{T}$ topology is coarser than the norm topology (so every $\mathcal{T}$-open set is open in the norm topology). If K is a cone in X we shall say that a sequence (x_j) of elements of K is monotonic increasing (with respect to the ordering induced by K) if $x_j \le x_{j+1}$ for all $j \ge 1$, and we shall say that (x_j) is bounded above in the partial ordering induced by K if there exists $w \in K$ such that $x_j \le w$ for all $j \ge 1$. If K, X and $\mathcal{T}$ are as above, we shall say that K has the "monotone convergence property in the $\mathcal{T}$ topology" if every monotonic increasing sequence (x_j) in K such that (x_j) is bounded above in the partial ordering induced by K has a limit $z \in K$ in the $\mathcal{T}$ topology. In the situation of Theorem 1.1, $X = \mathcal{L}(H)$, K is the cone of p.s.d. operators in X, $\mathcal{T}$ is the strong operator topology and K has the monotonic convergence property in the $\mathcal{T}$ topology.

THEOREM 1.4. *Let notation and assumptions be as in Theorem* 1.3, *except do* not *assume that X is finite dimensional. Suppose that $\mathcal{T}$ is a topology on X, coarser than the norm topology, such that X is a Hausdorff, locally convex topological vector space in the $\mathcal{T}$ topology and K has the monotone convergence property in the $\mathcal{T}$ topology. If u is the unique normalized positive eigenvector of M and $f^k(x) = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})$, there exists $w \in X$ so that*

$$\lim_{k\uparrow\infty} \sum_{i=1}^{n} u_i \ \phi(x_i^{(k)}) = w,$$

where convergence is in the $\mathcal{T}$ topology and u_i is the ith component of u. If $L : X \to \mathbb{C}$ is any linear map such that $L(x) \ge 0$ for all $x \in K$ and L is continuous in the $\mathcal{T}$ topology, then

$$\lim_{k\uparrow\infty} L(\phi(x_i^{(k)})) = L(w) \qquad \text{for } 1 \le i \le n.$$

The proof of Theorem 1.4 is very similar to that of Theorem 1.1. If $y \in H$, the role of L in Theorem 1.4 is served by $L(A) = \langle Ay, y \rangle$ for $A \in \mathcal{L}(H)$.

2. - Convergence results for generalizations of the AGM

The original motivation for this paper was the problem of proving the convergence of $f^k(A,B)$ for the maps

$$f(A,B) = ((A+B)/2,\ \exp((\log A)/2 + (\log B)/2) \text{ or}$$
$$f(A,B) = (\alpha A + (1-\alpha)B,\ \exp(\beta \log A + (1-\beta)\log B)) \tag{2.1}$$

where A and B are positive definite, bounded linear operators on a Hilbert space. However, there are many generalizations of the classical Gauss-Lagrange-Legendre AGM, and to treat the extensions of these generalizations to the operator-valued case in a reasonably unified way it is necessary to consider much more general f than those in (2.1).

Thus, if H is a Hilbert space, let K denote the cone of p.s.d. operators in $\mathcal{L}(H)$. Let C denote the n-fold Cartesian product of K with itself. If $\sigma \in \mathbb{R}^n$, call σ a "probability vector" if all components σ_i of σ are nonnegative and $\sum_{i=1}^{n} \sigma_i = 1$. If r is a real number, σ is a probability vector and $A = (A_1, A_2, \cdots, A_n) \in C^0$, define a map $M_{r\sigma} : C^0 \to K^0$ by

$$M_{r\sigma}(A) = (\sum_{j=1}^{n} \sigma_j A_j^r)^{1/r}. \tag{2.2}$$

If $r = 0$, (2.2) does not make sense and we define

$$M_{0\sigma}(A) = \exp(\sum_{j=1}^{n} \sigma_j \log(A_j)). \tag{2.3}$$

If $A \in C^0$, then

$$\lim_{r \to 0} \|M_{r\sigma}(A) - M_{0\sigma}(A)\| = 0. \tag{2.4}$$

THEOREM 2.1. *Let K, H and C be as above. For each i, $1 \le i \le n$, let Γ_i be a finite collection of ordered pairs (r, σ) such that r is a nonnegative real and σ is a probability vector. For each $(r, \sigma) \in \Gamma_i$, let $c_{ir\sigma}$ be a positive real number. Define $f : C^0 \to C^0$ by*

$$f_j(A) = \sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} M_{r\sigma}(A), \tag{2.5}$$

where $f_j(A)$ denotes the jth component of $f(A)$. Assume that

$$\sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} = 1 \qquad \text{for } 1 \le j \le n. \tag{2.6}$$

If $\pi_i : \mathbb{R}^n \to \mathbb{R}$ *denotes the projection onto the ith component of a vector, define*

$$m_{ij} = \sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma}\pi_i(\sigma) \tag{2.7}$$

and assume that the $n \times n$ *matrix* $M = (m_{ij})$ *is primitive. Let* u *(*u *a column vector) denote the unique probability vector such that* $Mu = u$ *and let* $u_i = \pi_i(u)$*. If, for* $A = (A_1, A_2, \cdots, A_n) \in C^0$*, we write*

$$f^k(A) = (A_1^{(k)}, A_2^{(k)}, \cdots, A_n^{(k)}) = A^{(k)} \tag{2.8}$$

there exists $E \in K^0$ *such that*

$$s\text{-}\lim_{k\uparrow\infty} \sum_{i=1}^{n} u_i \log(A_i^{(k)}) = \log E \tag{2.9}$$

and

$$w\text{-}\lim_{k\uparrow\infty} \log(A_i^{(k)}) = \log E \quad \text{for } 1 \le i \le n. \tag{2.10}$$

If (r, σ) *and* (ρ, τ) *are both elements of* Γ_i *for some* i, $1 \le i \le n$*, then*

$$s\text{-}\lim_{k\uparrow\infty} [\log M_{r\sigma}(A^{(k)}) - \log M_{\rho\tau}(A^{(k)})] = 0. \tag{2.11}$$

If $(r, \sigma) \in \Gamma_i$ *for some* i *and* $r > 0$*, then for all* p *and* q *such that* $\pi_p(\sigma) > 0$ *and* $\pi_q(\sigma) > 0$ *one has*

$$s\text{-}\lim_{k\uparrow\infty} [\log(A_p^{(k)}) - \log(A_q^{(k)})] = 0. \tag{2.12}$$

If H *is finite dimensional,*

$$\lim_{k\uparrow\infty} \|A_i^{(k)} - E\| = 0 \text{ for } 1 \le i \le n. \tag{2.13}$$

If there exists $(\bar{r}, \bar{\sigma}) \in \Gamma_i$ *for some* i *such that* $\bar{r} > 0$ *and all components of* $\bar{\sigma}$ *are positive, then*

$$s\text{-}\lim_{k\uparrow\infty} A_i^{(k)} = E \qquad \text{for } 1 \le i \le n. \tag{2.14}$$

PROOF. If $A = (A_1, A_2, \cdots, A_n) \in C^0$ and $B = (B_1, B_2, \cdots, B_n) = f(A)$ and if we define $S_{1j} = S_{1j}(A)$ and $S_{2j} = S_{2j}(A)$ by

$$S_{1j} = \log\Big(\sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} M_{r\sigma}(A)\Big) - \sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} \log M_{r\sigma}(A) \tag{2.15}$$

and

$$(2.16) \qquad S_{2j} = \sum_{r\neq 0} \sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} r^{-1}[\log(\sum_{p=1}^{n} \pi_p(\sigma) A_p^r) - \sum_{p=1}^{n} \pi_p(\sigma)\log(A_p^r)],$$

we find that

$$(2.17) \qquad \log B_j - \sum_i m_{ij} \log A_i = S_{1j} + S_{2j}.$$

The concavity of $D \to \log D = \phi(D)$, for $D \in K^0$, implies that the right side of (2.17) is positive semidefinite. It follows that if $\Phi(A)$ is defined as in Theorem 1.1 and $\phi = \log$ we have

$$(2.18) \qquad \Phi(f(A)) \geq \Phi(A)M \qquad \text{for all } A \in C^0.$$

If $A = (A_1, A_2, \cdots, A_n) \in C^0$ and α and β are positive numbers such that

$$(2.19) \qquad \alpha I \leq A_j \leq \beta I,$$

the spectral mapping theorem implies that for $r > 0$

$$\alpha^r I \leq A_j^r \leq \beta^r I$$

and

$$(\log \alpha) I \leq \log A_j \leq (\log \beta) I.$$

If σ is any probability vector, it follows that

$$\alpha^r I \leq \sum_{j=1}^{n} \sigma_j A_j^r \leq \beta^r I$$

and

$$(\log \alpha) I \leq \sum_{j=1}^{n} \sigma_j \log A_j \leq (\log \beta) I,$$

and by applying the spectral mapping theorem again we conclude that

$$(2.20) \qquad \alpha I \leq M_{r\sigma}(A) \leq \beta I.$$

A variant of this argument shows that (2.20) also holds if $r < 0$. We obtain directly from (2.20) that

$$\alpha I \leq \sum_{(r,\sigma)\in\Gamma_j} c_{jr\sigma} M_{r\sigma}(A) \leq \beta I$$

or

$$\alpha I \le A_j^{(1)} \le \beta I \qquad \text{for } 1 \le j \le n.$$

A simple induction now shows that if A satisfies (2.19) and the notation is as in (2.8) then

$$\alpha I \le A_i^{(k)} \le \beta I \qquad \text{for } 1 \le i \le n \text{ and } k \ge 1, \tag{2.21}$$

and (2.21) implies that

$$(\log \alpha) I \le \log A_i^{(k)} \le (\log \beta) I, \ \ 1 \le i \le n \text{ and } k \ge 1. \tag{2.22}$$

Inequalities (2.20) - (2.22) verify the hypotheses of Theorem 1.1, so Theorem 1.1 gives (2.9) and (2.10). If H is finite dimensional, weak convergence implies norm convergence and by applying the exponential map to (2.10) we obtain (2.13).

If H is infinite dimensional, more care is necessary. If we replace A by $A^{(k)} = f^k(A)$ and B by $A^{(k+1)}$ in (2.17) then

$$\log A_j^{(k+1)} - \sum_i m_{ij} \log A_i^{(k)} = S_{1j}(A^{(k)}) + S_{2j}(A^{(k)}). \tag{2.23}$$

By using (2.10) and the fact that $\sum_i m_{ij} = 1$, we conclude that the left side of (2.23) converges to zero in the weak operator topology. Since $S_{1j}(A^{(k)})$ is p.s.d. and all summands of $S_{2j}(A^{(k)})$ are p.s.d., (2.23) implies that

$$w\text{-}\lim_{k\uparrow\infty} S_{1j}(A^{(k)}) = 0, \ \ 1 \le k \le n, \tag{2.24}$$

and

$$w\text{-}\lim_{k\uparrow\infty} [\log(\sum_{p=1}^{n} \pi_p(\sigma)(A_p^{(k)})^r) - \sum_{p=1}^{n} \pi_p(\sigma)\log((A_p^{(k)})^r)] = 0, \tag{2.25}$$

where (2.25) is satisfied if $(r,\sigma) \in \Gamma_j$ for some j and $r > 0$. We now apply Theorem 1.2 (recalling that $\phi(z) = \log z$ satisfies the hypotheses of Theorem 1.2). Using Theorem 1.2 and (2.24) we find that (2.11) holds, and Theorem 1.2 and (2.25) imply (2.12). If there exists $(\bar{r}, \bar{\sigma})$ as in the statement of the theorem, we obtain from (2.12) that

$$s\text{-}\lim_{k\uparrow\infty} [\log A_i^{(k)} - \log A_j^{(k)}] = 0 \qquad \text{for } 1 \le i, \ j \le n. \tag{2.26}$$

Combining (2.9) and (2.26) we obtain as in Theorem 1.2,

$$s\text{-}\lim_{k\uparrow\infty} \log A_i^{(k)} = \log E \qquad \text{for } 1 \le i \le n,$$

and (2.14) now follows easily. □

Theorem 2.1 provides insufficient information if H is infinite dimensional and $r = 0$ for all $(r, \sigma) \in \Gamma_j$, $1 \leq j \leq n$. We consider this case separately in the next theorem.

THEOREM 2.2. *Let the notation and assumptions be as in Theorem 2.1. For $1 \leq i \leq n$, assume that if $(r, \sigma) \in \Gamma_i$, then $r = 0$, so Γ_i can be considered a finite set of probability vectors and we can write*

$$f_i(A) = \sum_{\sigma \in \Gamma_i} c_{i\sigma} M_{0\sigma}(A), \qquad c_{i\sigma} = c_{i0\sigma}.$$

Let u be the unique probability vector such that $Mu = u$. Assume that there are $n-1$ pairs of probability vectors $\sigma^{(j)}$ and $\tau^{(j)}$, $2 \leq j \leq n$, such that $\sigma^{(j)}$ and $\tau^{(j)} \in \Gamma_i$ for some i depending on j and such that the $n-1$ vectors, $\alpha^{(j)} = \sigma^{(j)} - \tau^{(j)}$, $2 \leq j \leq n$, are linearly independent. Then for any $A \in C^0$, there exists $E \in K^0$ such that

$$s\text{-}\lim_{k \uparrow \infty} f^k(A) = (E, E, \cdots, E). \tag{2.27}$$

If $n = 2$ or if H is finite dimensional, (2.27) *remains valid without the assumption that there exist vectors $\alpha^{(j)}$, $2 \leq j \leq n$, as above.*

PROOF. Theorem 2.1 implies that if σ and τ are any two probability vectors in Γ_j, $1 \leq j \leq n$, then

$$s\text{-}\lim_{k \uparrow \infty} [\log M_{0\sigma}(A^{(k)}) - \log M_{0\tau}(A^{(k)})] = 0,$$

or equivalently

$$s\text{-}\lim_{k \uparrow \infty} \sum_{i=1}^{n} (\sigma_i - \tau_i) \log(A_i^{(k)}) = 0. \tag{2.28}$$

If u is the eigenvector of M in the statement of Theorem 2.1, let N be the $n \times n$ matrix whose first column is u and whose jth column is $\alpha^{(j)}$ for $2 \leq j \leq n$. Equations (2.9) and (2.28) imply that, in the notation of Theorem 2.1,

$$s\text{-}\lim_{k \uparrow \infty} \Phi(A^{(k)}) N = (\log E, 0, 0, \cdots, 0), \tag{2.29}$$

where

$$\Phi(A^{(k)}) = (\log A_1^{(k)}, \ \log A_2^{(k)}, \cdots, \log A_n^{(k)}).$$

Because the components of $\alpha^{(j)}$, $2 \leq j \leq n$, sum to zero and the components of u sum to one, it is easy to see that if we define $\alpha^{(1)} = u$, the n vectors

$\alpha^{(j)}$, $1 \leq j \leq n$, are linearly independent. This implies that N is invertible, and since

$$\Phi(E)N = (\log E, 0, 0, \cdots, 0),$$

we conclude that

$$s\text{-}\lim_{k\uparrow\infty} \Phi(A^{(k)}) = (\log E, 0, 0, \cdots, 0)N^{-1} = (\log E,\ \log E, \cdots, \log E),$$

which (with the aid of Lemma 1.2) gives (2.27).

If H is finite dimensional, (2.27) follows directly from Theorem 2.1 without any knowledge of the vectors $\alpha^{(j)}$. If $n = 2$ and Γ_1 or Γ_2 contains more than one element, there is a nonzero vector $\alpha^{(2)} = \sigma - \tau$ (σ and τ in Γ_i for $i = 1$ or 2) and the theorem follows from our previous remarks. Thus assume that $n = 2$ and that Γ_1 and Γ_2 each contains only one element, say $\Gamma_1 = \{\sigma\}$ and $\Gamma_2 = \{\tau\}$. If $\sigma = (\sigma_1, \sigma_2)$ and $\tau = (\tau_1, \tau_2)$, we find that

$$(\log A_1^{(k+1)},\ \log A_2^{(k+1)}) = (\log A_1^{(k)},\ \log A_2^{(k)})M,$$

where

$$M = \begin{pmatrix} \sigma_1 & \tau_1 \\ \sigma_2 & \tau_2 \end{pmatrix}$$

and M is assumed primitive. It follows that

$$(\log A_1^{(k)},\ \log A_2^{(k)}) = (\log A_1,\ \log A_2)M^k$$

and

$$n\text{-}\lim_{k\uparrow\infty} \log A_1^{(k)} = n\text{-}\lim_{k\uparrow\infty} \log A_2^{(k)} = u_1 \log A_1 + u_2 \log A_2$$

where the column vector $(u_1, u_2)^T$ is the unique probability vector which is an eigenvector for M. One obtains the norm convergence of $A_1^{(k)}$ and $A_2^{(k)}$ from the above equation. □

REMARK 2.1. Theorems 2.1 and 2.2 are not sharp results if H is infinite dimensional. It is possible that strong convergence of $(A_i^{(k)})$, $1 \leq i \leq n$, is valid with only the assumption that M is primitive, but we have not been able to prove this.

It is worth noting that the conclusions of Theorem 2.2 remain valid if there exist $n-1$ linearly independent vectors $\alpha^{(j)}$, $2 \leq j \leq n$, such that

$$s\text{-}\lim_{k\uparrow\infty} \sum_{i=1}^{n} \alpha_i^{(j)} \ \log A_i^{(k)} = 0$$

and

$$\sum_{i=1}^{n} \alpha_i^{(j)} = 0.$$

The vectors $\alpha^{(j)}$ do not have to arise as in Theorem 2.2.

REMARK 2.2. Let notation and assumptions be as in Theorem 2.2 but do not assume the existence of probability vectors $\sigma^{(j)}$ and $\tau^{(j)}$ as in Theorem 2.2. Instead suppose that

$$m_{ij} = \sum_{\sigma\in\Gamma_j} c_{j\sigma}\pi_i(\sigma) = u_i \quad \text{for } 1 \le i, j \le n. \tag{2.30}$$

Then the conclusion of Theorem 2.2 (in particular, (2.27)) still holds.

To see this, note that $M = (m_{ij})$ is the matrix in Theorem 2.1 and that the column vector u whose ith entry equals u_i (u_i as in (2.30)) satisfies $Mu = u$. Theorem 2.1 implies that for all σ, $\tau \in \Gamma_j$, $1 \le j \le n$,

$$s\text{-}\lim_{k\uparrow\infty} [\log M_{0\sigma}(A^{(k)}) - \log M_{0\tau}(A^{(k)})] = 0 \tag{2.31}$$

and there exists $E \in K^0$ so that

$$s\text{-}\lim_{k\uparrow\infty} [\sum_{j=1}^{n} u_j \ \log(A_j^{(k)})] = \log E.$$

Because $m_{ij} = u_i$ we have

$$\sum_{\sigma\in\Gamma_j} c_{j\sigma} \ \log M_{0\sigma}(A^{(k)}) = \sum_{i=1}^{n} u_i \ \log(A_i^{(k)}),$$

so

$$s\text{-}\lim_{k\uparrow\infty} [\sum_{\sigma\in\Gamma_j} c_{j\sigma} \ \log M_{0\sigma}(A^{(k)})] = \log E. \tag{2.32}$$

Combining (2.31) and (2.32) we obtain

$$s\text{-}\lim_{k\uparrow\infty} \log M_{0\sigma}(A^{(k)}) = \log E,$$

so Lemma 1.2 implies

$$s\text{-}\lim_{k\uparrow\infty} M_{0\sigma}(A^{(k)}) = E. \tag{2.33}$$

Using (2.33) we see that

$$s\text{-}\lim_{k\uparrow\infty} A_j^{(k+1)} = s\text{-}\lim_{k\uparrow\infty} \sum_{\sigma\in\Gamma_j} c_{j\sigma} M_{0\sigma}(A^{(k)}) = E, \tag{2.34}$$

which is the desired result.

Although the assumptions on M in Remark 2.2 are restrictive, they are satisfied in some important applications. Remark 2.2 also provides further evidence that Theorem 2.2 is far from best possible.

It is useful in some applications to allow functions which are the composition of functions like those in Theorems 2.1 or 2.2. One can give analogues of Theorems 2.1 and 2.2 for such functions even if H is infinite dimensional, but for simplicity we restrict ourselves to the finite dimensional case.

THEOREM 2.3. *Let K, H, C, ϕ and Φ be as in Theorem* 2.1 *and suppose that H is finite dimensional. Assume that $f : C^0 \to C^0$ and $g : C^0 \to C^0$ are continuous maps and define $h = g \circ f$. Assume that for every $A \in C^0$ there exist $B \in C^0$ and positive reals α and β such that*

$$\alpha B \leq h^j(A) \leq \beta B$$

and

$$\Phi(h^j(A)) \leq \Phi(\beta B)$$

for all $j \geq 0$. Assume that there exist $n \times n$ column-stochastic matrices M and N with nonnegative entries such that for every $A \in C^0$

$$\Phi(f(A)) \geq \Phi(A)M$$

and

$$\Phi(g(A)) \geq \Phi(A)N$$

and MN is primitive. Then there exists $E \in K^0$ such that, if $h^k(A) = (A_1^{(k)}, A_2^{(k)}, \cdots, A_n^{(k)})$,

$$\lim_{k\uparrow\infty} \|\phi(A_i^{(k)}) - \phi(E)\| = 0 \qquad \text{for } 1 \leq i \leq n.$$

If ϕ is one-one, one also obtains

$$\lim_{k\uparrow\infty} \|A_i^{(k)} - E\| = 0.$$

PROOF. By using the hypotheses on f and g one finds

$$\Phi(h(A)) \geq \Phi(f(A))N \geq \Phi(A)MN.$$

Thus h satisfies the hypotheses of Theorem 1.1, with MN replacing M in Theorem 1.1, and Theorem 2.3 follows immediately from Theorem 1.1. □

COROLLARY 2.1. *Let K, C and H be as in Theorem* 2.1 *and assume that H is finite dimensional. Let f be as in Theorem* 2.1, *but do not assume that the matrix M defined by* (2.7) *is primitive. Suppose that $g : C^0 \to C^0$ is like*

f. More precisely, for $1 \le i \le n$, let T_i be a finite collection of ordered pairs (s, τ) such that s is a nonnegative real and τ is a probability vector. For each $(s, \tau) \in T_i$, let $d_{is\tau}$ be a positive real number. Define $g : C^0 \to C^0$ by

$$g_j(A) = \sum_{(s,\tau)\in T_j} d_{js\tau} M_{s\tau}(A).$$

Assume that

$$\sum_{(s,\tau)\in T_j} d_{js\tau} = 1 \qquad \text{for } 1 \le j \le n$$

and define an $n \times n$ column-stochastic matrix P by

$$p_{ij} = \sum_{(s,\tau)\in T_j} d_{js\tau} \pi_i(\tau).$$

If MP is primitive and $h = g \circ f$, then for any $A \in C^0$ there exists $E \in K^0$ such that

$$\lim_{k\uparrow\infty} \|A_i^{(k)} - E\| = 0 \qquad \text{for } 1 \le i \le n,$$

where

$$h^k(A) = (A_1^{(k)}, A_2^{(k)}, \cdots, A_n^{(k)}).$$

PROOF. If $\phi = \log$, the proof of Theorem 2.1 shows that

$$\Phi(f(A)) \ge \Phi(A)M \text{ and } \Phi(g(A)) \ge \Phi(A)P$$

for all $A \in C^0$. Thus Corollary 2.1 follows easily from Theorem 2.3. Details are left to the reader. □

Theorems 2.1 and 2.2 provide no information if f is given as in (2.5) and $r < 0$ for some $(r, \sigma) \in \Gamma_j$. However, Theorems 1.1 and 1.2 provide information about certain functions of this type also.

THEOREM 2.4. *Let K, C and H be as in Theorem 2.1. For $1 \le j \le n$, let Γ_j be a finite collection of probability vectors and for each $\sigma \in \Gamma_j$ let $c_{j\sigma}$ be a positive real number such that*

$$\sum_{\sigma\in\Gamma_j} c_{j\sigma} = 1.$$

Define a map $f : C^0 \to C^0$ by

$$f_j(A) = \sum_{\sigma\in\Gamma_j} c_{j\sigma} M_{-1\sigma}(A),$$

where $f_j(A)$ denotes the jth component of $f(A)$ and

$$M_{-1\sigma}(A) = \Big(\sum_{i=1}^{n} \sigma_i A_i^{-1}\Big)^{-1}$$

for $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ *and* $A = (A_1, A_2, \cdots, A_n)$. *Define* m_{ij} *by*

$$m_{ij} = \sum_{\sigma \in \Gamma_j} \pi_i(\sigma) c_{j\sigma}$$

and assume the $n \times n$ *column-stochastic matrix* $M = (m_{ij})$ *is primitive. Finally assume that there are* $n-1$ *pairs of probability vectors* $\sigma^{(j)}$ *and* $\tau^{(j)}$, $2 \le j \le n$, *such that* $\sigma^{(j)}$ *and* $\tau^{(j)}$ *are in* Γ_i *for some* $i = i(j)$ *and the* $n-1$ *vectors* $\alpha^{(j)} = \sigma^{(j)} - \tau^{(j)}$, $2 \le j \le n$, *are linearly independent. Then for any* $A \in C^0$, *there exists* $E \in K^0$ *such that*

$$s\text{-}\lim_{k \uparrow \infty} f^k(A) = (E, E, \cdots, E).$$

PROOF. Define $\phi(z) = -z^{-1}$ and notice that $\phi(z)$ satisfies the conditions of Theorem 1.2 and that $B \to -B^{-1}$ ($B \in K^0$) is concave (see Section 1). Using the concavity of ϕ one easily sees that if $A \in C^0$ and $B = f(A)$,

$$\begin{aligned} &\phi(B_j) - \sum_{i=1}^{n} m_{ij}\phi(A_i) \\ &= \phi\Big(\sum_{\sigma \in \Gamma_j} c_{j\sigma} M_{-1\sigma}(A)\Big) - \sum_{\sigma \in \Gamma_j} c_{j\sigma}\phi(M_{-1\sigma}(A)) \end{aligned} \tag{2.35}$$

and the right side of the above equation is positive semidefinite. The rest of the proof follows from (2.35) by using Theorems 1.1 and 1.2 as in Theorem 2.2 and is left to the reader. □

REMARK 2.3. It is important to note that if σ is a probability vector, one can write (for $A \in C^0$)

$$M_{1\sigma}(A) = \sum_{i=1}^{n} \sigma_i A_i = \sum_{i=1}^{n} \sigma_i (A_i^{-1})^{-1} = \sum_{i=1}^{n} \sigma_i M_{-1\tau(i)}(A)$$

where $\tau(i)$ is the probability vector with 1 at the ith position. Thus if f is given as in (2.5) and $r = 1$ or $r = -1$ for each $(r, \sigma) \in \Gamma_j$, then by relabelling and redefining Γ_i one can assume that f is as in Theorem 2.4.

By using the above remark and Theorem 2.4 we obtain the following corollary.

COROLLARY 2.2. *Let the notation and assumptions be as in Theorem* 2.4 *except do not assume the existence of* $n-1$ *pairs* $\sigma^{(j)}$ *and* $\tau^{(j)}$ *as in Theorem* 2.4. *Assume that there exist* n *positive numbers* d_j, $1 \le j \le n$, *such that* $f_1(A)$, *the first component of* $f(A)$, *satisfies*

$$f_1(A) = f_1(A_1, A_2, \cdots, A_n) = \sum_{j=1}^{n} d_j A_j.$$

Then for any $A \in C^0$, *there exists* $E \in K^0$ *such that*

$$s\text{-}\lim_{k\uparrow\infty} f^k(A) = (E, E, \cdots, E).$$

PROOF. As noted in Remark 2.3, write

$$f_1(A) = \sum_{\sigma\in\Gamma_1} c_\sigma M_{-1\sigma}\ (A),$$

where Γ_1 comprises the n probability vectors $\tau(j)$, $1 \le j \le n$, and $\tau(j)$ has 1 in the jth position. Then

$$\alpha^{(j)} = \tau(1) - \tau(j),\ \ 2 \le j \le n,$$

are $n-1$ linearly independent vectors and the corollary follows from Theorem 2.4. □

A very special case of Corollary 2.2 is an arithmetic-harmonic mean (the case $\alpha = \beta = 1/2$ below) which has been considered by Fujii [19].

COROLLARY 2.3. *Let* H *be a Hilbert space,* K *the cone of positive semidefinite operators in* $\mathcal{L}(H)$, *and* $C = K \times K$. *If* α *and* β *are real numbers such that* $0 < \alpha,\ \beta < 1$, *define* $f : C^0 \to C^0$ *by*

$$f(A, B) = (\alpha A + (1-\alpha)B,\ \ (\beta A^{-1} + (1-\beta)B^{-1})^{-1}).$$

If $(A, B) \in C^0$, *there exists* $E \in K^0$ *such that*

$$s\text{-}\lim_{k\uparrow\infty} f^k(A, B) = (E, E).$$

Although we shall not prove this here, one can prove convergence in the operator norm under the assumptions of Corollary 2.3.

Similarly, as a direct corollary of Theorem 2.2 we obtain an operator valued extension of the AGM of the type suggested by (2.1).

COROLLARY 2.4. *Let* K, C *and* H *be as in Corollary* 2.3. *If* α *and* β *are real numbers such that* $0 < \alpha,\ \beta < 1$, *define* $f : C^0 \to C^0$ *by*

$$f(A, B) = (\alpha A + (1-\alpha)B, \exp(\beta \log A + (1-\beta)\log B)).$$

Then for any $(A, B) \in C^0$, *there exists* $E \in K^0$ *such that*

$$s\text{-}\lim_{k\uparrow\infty} f^k(A, B) = (E, E).$$

PROOF. Since A can be written as $\exp(\log A)$ and similarly for B, the mapping f is of the form considered in Theorem 2.2 and (in the notation of

Theorem 2.2) $n = 2$. One easily checks that

$$M = \begin{bmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{bmatrix},$$

so M is primitive and the corollary follows from Theorem 2.2. □

If A and B are positive real numbers in Corollary 2.4 and $\alpha = \beta = 1/2$, it was known classically (see [13]) that one can express the limit of $f^k(A, B)$ in terms of an elliptic integral. However, as D. Borwein and P.B. Borwein have pointed out [11], for general α and β (even if $\alpha = \beta$) there is no known integral formula for the limit of $f^k(A, B)$.

There are many other classical variants of the AGM. Carlson [13] gives a unified treatment of some of these results. In one example, Carlson defines (for a and b positive reals)

$$f_1(a,b) = (a+b)/2, \quad f_2(a,b) = (ab)^{1/2}, \quad f_3(a,b) = [a(a+b)/2]^{1/2}$$
$$\text{and } f_4(a,b) = [b(a+b)/2]^{1/2}.$$

He then defines a map $f_{ij}(a, b)$ by

$$f_{ij}(a,b) = (f_i(a,b), \; f_j(a,b)).$$

The case $i = 1$ and $j = 3$ is usually attributed to Borchardt (see [6], [9]). One can prove that

$$\lim_{k\uparrow\infty} f_{ij}^k(a,b) = (c,c), \quad c = L_{ij}(a,b),$$

and Carlson gives explicit integral formulas for $L_{ij}(a, b)$.

We wish to generalize the above convergence results to pairs of p.d. bounded linear operators A and B on a Hilbert space. If A and B commute or if, as in Section 4 below, one uses a different analogue of the square root of the product of two positive numbers, one can also generalize the integral formulas $L_{ij}(a, b)$. However, we shall only carry this out for the case of the AGM in Sections 3 and 4 below. Thus let H be a Hilbert space and K the cone of p.s.d. operators in $\mathcal{L}(H)$. If $A, B \in K^0$ and r is a real number such that $0 \leq r \leq 1$, define

$$h(A,B,r) = \exp(r \; \log(A) + (1-r) \; \log(B)). \tag{2.36}$$

Define a map f of $K^0 \times K^0$ into itself by

$$f(A,B) = (A_1, B_1), \tag{2.37}$$

where

$$A_1 = c_0 A + c_1 B + c_2 h(A,B,\gamma_2) + c_3 h(A,C,\gamma_3) + c_4 h(B,C,\gamma_4), \tag{2.38}$$

(2.39) $$B_1 = d_0A + d_1B + d_2h(A, B, \delta_2) + d_3h(A, C, \delta_3) + d_4h(B, C, \delta_4),$$

(2.40) $$C = (A + B)/2.$$

It will always be assumed that

(2.41) $$\sum_{j=0}^{4} c_j = \sum_{j=0}^{4} d_j = 1.$$

COROLLARY 2.5. *Let H be a Hilbert space and K the cone of p.s.d. operators in $\mathcal{L}(H)$. Let c_j and d_j, $0 \le j \le 4$, be nonnegative real numbers which satisfy* (2.41) *and let γ_j and δ_j, $2 \le j \le 4$, be real numbers such that $0 < \gamma_j,\ \delta_j < 1$ for $2 \le j \le 4$. In addition assume that $c_j < 1$ and $d_j < 1$ for $j = 0$ and $j = 1$. Define a map f of $K^0 \times K^0$ into itself by* (2.36) - (2.40). *Then for any $(A, B) \in K^0 \times K^0$, there exists $E \in K^0$ such that*

$$s\text{-}\lim_{k \uparrow \infty} f^k(A, B) = (E, E).$$

PROOF. Define $D = K \times K \times K$ and define a map $g : D^0 \to D^0$ by

(2.42) $$g(A, B, C) = (A_1, B_1, C_1),$$

where C is now an arbitrary element K^0, A_1 and B_1 are given by (2.38) and (2.39) respectively and

(2.43) $$C_1 = (1/2)(A_1 + B_1).$$

If π projects D^0 onto $K^0 \times K^0$ so that

$$\pi(A, B, C) = (A, B),$$

an easy induction argument shows that if $(A, B, C) \in D^0$ and $C = (A + B)/2$ then

(2.44) $$\pi(g^k(A, B, C)) = f^k(A, B) \qquad \text{for all } k \ge 1.$$

Thus to prove Corollary 2.5 it suffices to prove that for any $(A, B, C) \in D^0$ there exists $E \in K^0$ such that

(2.45) $$s\text{-}\lim_{k \uparrow \infty} g^k(A, B, C) = (E, E, E).$$

The map g is of the form in Theorem 2.2 (whereas f is not), and one could try to apply Theorem 2.2 and Remark 2.1. Such an approach requires slightly stronger assumptions than we have made, so we shall use a somewhat different argument.

We first eliminate some trivial cases. If we have

$$c_0 + c_1 + c_2 = 1 = d_0 + d_1 + d_2, \tag{2.46}$$

so $d_3 = d_4 = c_3 = c_4 = 0$, (A_1, B_1) is a function of (A, B), the function being of the type considered in Theorem 2.2. Thus we are in the case $n = 2$ of Theorem 2.2. The assumption that $c_j < 1$ and $d_j < 1$ for $j = 0$ and 1 insures that the corresponding 2×2 matrix M has all positive entries. Thus we are done if (2.46) is satisfied.

Similarly, if

$$d_0 + d_3 = 1 = c_0 + c_3, \tag{2.47}$$

(A_1, C_1) is a function of (A, C) and the corollary follows easily from the case $n = 2$ in Theorem 2.2. Also, if

$$d_1 + d_4 = 1 = c_1 + c_4, \tag{2.48}$$

(B_1, C_1) is a function of (B, C) and we return to the case $n = 2$ in Theorem 2.2.

Thus we can assume that (2.46), (2.47) and (2.48) are all not satisfied. Let M be the 3×3 column-stochastic matrix defined as in Theorem 2.2 for our map g. One can easily check from the defining equations for g that the third column of M is the arithmetic average of the first two columns. Because $c_j < 1$ and $d_j < 1$ for $j = 0$ and $j = 1$, each of the first two columns of M has at most one zero entry. If the first entries of columns one and two of M are both zero, then (2.48) is satisfied, contrary to assumption. Similarly, if the second entries of both columns one and two of M are both zero, (2.47) is satisfied, contrary to assumption. Finally, if the third entry of column one of M equals zero and the third entry of column two of M equals zero, (2.46) is satisfied, contrary to assumption. Thus the zero entry of column one of M (if it exists) is never in the same position as the zero entry of column two, so the third column of M, being the arithmetic average of columns one and two has all positive entries. Using this information, one easily checks that M^2 has all positive entries.

If u is the probability column vector such that $Mu = u$ and if we write

$$g^k(A, B, C) = (A_k, B_k, C_k),$$

Theorem 2.1 implies that there exists $E \in K^0$ such that

$$s\text{-}\lim_{k \uparrow \infty} (u_1 \log A_k + u_2 \log B_k + u_3 \log C_k) = \log E, \tag{2.49}$$

and $\log A_k$, $\log B_k$ and $\log C_k$ converge weakly to $\log E$. Because

$$w\text{-}\lim_{k\uparrow\infty} \log C_k = w\text{-}\lim_{k\uparrow\infty} \log((A_k + B_k)/2) = \log E,$$

we conclude that

$$w\text{-}\lim_{k\uparrow\infty}[\log((A_k + B_k)/2) - (\log A_k)/2 - (\log B_k)/2] = 0. \tag{2.50}$$

Using (2.50) and Theorem 1.2 we conclude that

$$s\text{-}\lim_{k\uparrow\infty}[\log A_k - \log B_k] = 0. \tag{2.51}$$

Because $C_k = (A_k + B_k)/2$ for $k \geq 1$, we have for $k \geq 1$

$$\begin{aligned}(\log A_{k+1},\ \log B_{k+1},\ \log C_{k+1}) &\geq (\log A_k,\ \log B_k,\ \log C_k)M \\ &\geq (\log A_k,\ \log B_k,\ (\log A_k)/2 + (\log B_k)/2)M.\end{aligned}$$

Using the above inequality we find that for $k \geq 1$

$$(\log A_{k+1},\ \log B_{k+1}) \geq (\log A_k,\ \log B_k)P,$$

where $P = (p_{ij})$ has elements

$$p_{ij} = m_{ij} + m_{3j}/2,\ \ 1 \leq i, j \leq 2.$$

Because each of the first two columns of M has at most one zero entry, all entries of P are positive, and P is obviously column-stochastic. If v is the unique column probability vector such that $Pv = v$, Theorem 2.1 implies

$$s\text{-}\lim_{k\uparrow\infty}(v_1 \log A_k + v_2 \log B_k) = \log E \tag{2.52}$$

for some $E \in K^0$. One obtains directly from (2.51) and (2.52) that

$$s\text{-}\lim_{k\uparrow\infty} \log A_k = s\text{-}\lim_{k\uparrow\infty} \log B_k = \log E,$$

so Lemma 1.2 implies

$$s\text{-}\lim_{k\uparrow\infty} A_k = s\text{-}\lim_{k\uparrow\infty} B_k = E = s\text{-}\lim_{k\uparrow\infty} C_k. \ \square$$

There are many other classical generalizations of the AGM. For example, G. Borchadt (see [6] and [9] for references) studied the map

$$f(a, b, c, d) = (a_1, b_1, c_1, d_1),$$

where

$$a_1 = (a+b+c+d)/4, \quad b_1 = ([ab]^{1/2} + [cd]^{1/2})/2, \tag{2.53}$$
$$c_1 = ([ac]^{1/2} + [bd]^{1/2})/2, \quad d_1 = ([ad]^{1/2} + [bc]^{1/2})/2$$

and a, b, c and d are (initially) positive reals. If f^n denotes the nth iterate of f and

$$f^n(a,b,c,d) = (a_n, b_n, c_n, d_n)$$

Borchardt proved (without great difficulty) that

$$\lim_{n\uparrow\infty} f^n(a,b,c,d) = (\gamma,\gamma,\gamma,\gamma), \quad \gamma > 0. \tag{2.54}$$

The number $\gamma = \mu(a,b,c,d)$ can be considered a generalized AGM. Borchardt established many properties of this AGM. If $a = c$ and $b = d$, then $a_n = c_n$ and $b_n = d_n$ for all $n \geq 1$ and (2.53) reduces to the original AGM.

As is pointed out in [6], Borchardt's algorithm is a special case of a more general construction, called a "monster algorithm" in [6]. Let G be a finite group of order n. If Θ_1 and Θ_2 are real-valued functions from G to $\mathbb{R}$, the convolution of Θ_1 and Θ_2, $\Theta_1 * \Theta_2$, is defined by

$$(\Theta_1 * \Theta_2)(t) = n^{-1} \sum_{s\in G} \Theta_1(s)\Theta_2(ts^{-1}). \tag{2.55}$$

(More generally, convolution can be defined with respect to a measure on a locally compact topological group). Define C to be the set of functions $\Theta : G \to [0,\infty)$, so C can be identified with the standard cone in $\mathbb{R}^n$, $n = |G|$. Define $F : C^0 \to C^0$ by

$$F(\Theta) = \Theta^{1/2} * \Theta^{1/2}, \text{ where } (\Theta^{1/2})(s) = [\Theta(s)]^{1/2}. \tag{2.56}$$

It is proved in [6] (at least for G abelian) that, if F^k denotes the kth iterate of F, then for any $\Theta \in C^0$

$$\lim_{k\uparrow\infty} F^k(\Theta) = \Theta_\infty,$$

where Θ_∞ is a positive, constant function. Borchardt's algorithm corresponds to $G = C_2 \times C_2$, where C_2 is a group of order 2. Another interesting example corresponds to a cyclic group of order 3 (see [6]).

We generalize this construction to operator-valued functions. Let K be the cone of p.s.d., bounded linear operators on a Hilbert space H. Let G be a finite group of order n. Let C denote the cone of maps $\Theta : G \to K$, so $C = \prod_{g\in G} K$. Define $F : C^0 \to C^0$ by

$$(F(\Theta))(t) = n^{-1} \sum_{s\in G} \exp((1/2)\log(\Theta(s)) + (1/2)\log(\Theta(ts^{-1}))). \tag{2.57}$$

This reduces to (2.56) when Θ is a real-valued.

COROLLARY 2.6. *Let the notation be as in the immediately preceding paragraph. Then for any* $\Theta \in C^0$,

$$s\text{-}\lim_{k\uparrow\infty} F^k(\Theta) = \Theta_\infty \in C^0,$$

where Θ_∞ *depends on* Θ *and* $\Theta_\infty : G \to K^0$ *is a constant function.*

PROOF. The cone C can be considered as the n-fold Cartesian product of K, and with this identification the map F is a special case of the maps considered in Theorem 2.2. We shall derive Corollary 2.6 from Remark 2.2.

Define $\phi : K^0 \to X = \mathcal{L}(H)$ by $\phi(A) = \log A$ and if Y denotes the Banach space of maps from G to X, define $\Phi : C^0 \to Y$ by

$$(\Phi(\Theta))(t) = \log(\Theta(t)) \qquad \text{for all } t \in G.$$

Applying ϕ to (2.57) and using the facts that

$$\sum_{s\in G} \log(\Theta(s)) = \sum_{s\in G} \log(\Theta(ts^{-1})) \tag{2.58}$$

and ϕ is concave gives

$$\log(F(\Theta)(t)) \geq n^{-1} \sum_{s\in G} \log(\Theta(s)) \tag{2.59}$$

or

$$\Phi(F(\Theta)) \geq \Phi(\Theta)M,$$

where M is the doubly stochastic matrix with all entries equal to n^{-1}. Thus we are in the situation of Remark 2.2 and the corollary follows. □

As already noted, we immediately obtain the operator analogue of Borchardt's algorithm from Corollary 2.6.

COROLLARY 2.7. *Let* K *denote the cone of* p.s.d., *bounded self-adjoint linear operators on a Hilbert space* H. *Define* $U = K \times K \times K \times K$. *Define* $f : U^0 \to U^0$ *by*

$$f(A, B, C, D) = (A_1, B_1, C_1, D_1)$$

where

$$A_1 = (1/4)(A + B + C + D),$$

$$B_1 = (1/2)[\exp((1/2)\log A + (1/2)\log B) + \exp((1/2)\log C + (1/2)\log D)],$$

$$C_1 = (1/2)[\exp((1/2)\log A + (1/2)\log C) + \exp((1/2)\log B + (1/2)\log D)], \text{ and}$$

$$D_1 = (1/2)[\exp((1/2)\log A + (1/2)\log D) + \exp((1/2)\log B + (1/2)\log C)].$$

Then if $f^m(A,B,C,D) = (A_m, B_m, C_m, D_m)$, *there exists* $E \in K^0$ *such that*

$$s\text{-}\lim_{m\uparrow\infty} (A_m, B_m, C_m, D_m) = (E, E, E, E).$$

Many other generalized means have operator-valued versions that can be analyzed by our methods. We mention only two more examples. Borchardt and Schwab (see [6]) considered the map

$$f(a,b) = (a_1, b_1), \ a_1 = (a+b)/2, \ b_1 = (a_1 b)^{1/2},$$

and the corresponding mean given by

$$\lim_{n\uparrow\infty} f^n(a,b) = (\delta, \delta).$$

Carlson [12] observed that the Borchardt-Schwab algorithm is naturally embedded in an algorithm involving three variables. Given positive numbers a, b and c, define α, β and γ by

$$\alpha = (b+c)/2, \ \beta = (a+c)/2, \ \gamma = (a+b)/2$$

and define $f(a,b,c) = (a_1, b_1, c_1)$, where

$$a_1 = (\beta\gamma)^{1/2}, \ b_1 = (\alpha\gamma)^{1/2} \text{ and } c_1 = (\alpha\beta)^{1/2}.$$

If $f^n(a,b,c) = (a_n, b_n, c_n)$, Carlson proved that

$$\lim_{n\uparrow\infty} (a_n, b_n, c_n) = (\delta, \delta, \delta).$$

He related the limit to certain integrals. If $b = c$, then $b_n = c_n$ for all $n \geq 1$ and one recovers the Borchardt-Schwab algorithm.

To generalize Carlson's algorithm to operator-valued maps, let K denote the cone of p.s.d., bounded self-adjoint linear maps of a Hilbert space H to itself. Define $E = K \times K \times K$. Define $f : E^0 \to E^0$ by

(2.60) $$f(A,B,C) = (A_1, B_1, C_1),$$

where

(2.61) $$A_1 = \exp((1/2)\log(\beta) + (1/2)\log(\gamma)), \ B_1 = \exp((1/2)\log(\alpha) + (1/2)\log(\gamma))$$
$$C_1 = \exp((1/2)\log(\alpha) + (1/2)\log(\beta))$$

and

(2.62) $$\alpha = (B+C)/2, \ \beta = (A+C)/2 \text{ and } \gamma = (A+B)/2.$$

COROLLARY 2.8. *If* $f : E^0 \to E^0$ *is defined by* (2.60) - (2.62), *then for any* $(A, B, C) \in E^0$,

$$s\text{-}\lim_{k \uparrow \infty} f^k(A, B, C) = (D, D, D), \tag{2.63}$$

where $D \in K^0$.

PROOF. Define $\Phi(A, B, C) = (\phi(A),\ \phi(B),\ \phi(C))$ for $(A, B, C) \in E^0$, where $\phi(A) = \log A$ for $A \in K^0$. It follows easily from the concavity of log that

$$\Phi(f(A, B, C)) \geq \Phi(A, B, C)M,$$

where

$$M = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

The reader can easily verify that the other hypotheses of Theorem 1.1 hold, so if $f^k(A, B, C) = (A_k, B_k, C_k)$, there exists $D \in K^0$ such that

$$\begin{aligned} &w\text{-}\lim_{k \uparrow \infty} \log A_k = w - \lim_{k \uparrow \infty} \log B_k \\ &= w\text{-}\lim_{k \uparrow \infty} \log C_k = \log D \end{aligned} \tag{2.64}$$

and

$$s\text{-}\lim_{k \uparrow \infty} (1/3)(\log A_k + \log B_k + \log C_k) = \log D. \tag{2.65}$$

The defining equation for f gives

$$\begin{aligned} &\log A_{k+1} - (1/2)\log A_k - (1/4)\log B_k - (1/4)\log(C_k) = \\ &(1/2)[\log((1/2)(A_k + C_k)) - (1/2)\log A_k - (1/2)\log(C_k)] \\ &+(1/2)[\log((1/2)(A_k + B_k)) - (1/2)\log A_k - (1/2)\log B_k]. \end{aligned} \tag{2.66}$$

The left side of (2.66) converges to zero in the weak operator topology, so Theorem 1.2 and (2.66) imply

$$s\text{-}\lim_{k \uparrow \infty} [\log A_k - \log C_k] = s\text{-}\lim_{k \uparrow \infty} [\log B_k - \log C_k] = 0. \tag{2.67}$$

(2.65) and (2.67) give

$$s\text{-}\lim_{k \uparrow \infty} \log A_k = s\text{-}\lim_{k \uparrow \infty} \log B_k = s\text{-}\lim_{k \uparrow \infty} \log C_k = \log D \tag{2.68}$$

and the corollary follows from (2.68) with the aid of Lemma 1.2. □

A final example is an algorithm of Meissel (see [6]). For positive real numbers a, b, c define

$$f(a,b,c) = \left([a+b+c]/3,\ \left([ab+ac+bc]/3\right)^{1/2},\ (abc)^{1/3}\right). \tag{2.69}$$

If K is the cone of p.s.d., self-adjoint linear operators on a Hilbert space H and $E = K \times K \times K$, one can define a map $f : E^0 \to E^0$ which is an analogue of the map in (2.69), namely,

$$f(A,B,C) = (A_1, B_1, C_1), \tag{2.70}$$

where

$$\begin{aligned} A_1 &= (A+B+C)/3,\ B_1 = [(1/3)\exp(\alpha+\beta) \\ &+ (1/3)\exp(\alpha+\gamma) + (1/3)\exp(\beta+\gamma)]^{1/2}, \\ C_1 &= \exp((\alpha+\beta+\gamma)/3),\ \alpha = \log A,\ \beta = \log B \text{ and } \gamma = \log C. \end{aligned} \tag{2.71}$$

COROLLARY 2.9. *If H and K are as above and $f : E^0 \to E^0$ is defined by* (2.70)-(2.71), *then for any $(A,B,C) \in E^0$, there exists $D \in K^0$ such that*

$$s\text{-}\lim_{n\to\infty} f^n(A,B,C) = (D,D,D). \tag{2.72}$$

PROOF. Corollary 2.9 follows by essentially the same argument used to prove Corollary 2.8 and is left to the reader. □

3. - Some elementary properties of the arithmetic-geometric mean

In this section we shall establish some basic properties of

$$M(A,B) = \lim_{k\uparrow\infty} f^k(A,B) \tag{3.1}$$

where

$$f(A,B) = ((A+B)/2,\ \exp((1/2)\log A + (1/2)\log B)). \tag{3.2}$$

Much of what we say extends to the more general examples considered in Section 2, but for simplicity we shall restrict ourselves to the above case or, occasionally, the "monster algorithm" of Corollary 2.6.

We begin with some generalities. If X is a complex Banach space, G is an open subset of X and $f : G \to X$ is continuous, f is called analytic if, whenever $B_r(u) = \{y : \|y-u\| < r\} \subseteq G$, $\Psi \in X^*$ is a complex linear functional and $v \in X$ is such that $\|v\| = 1$, then $\lambda \to \Psi(f(u + \lambda v))$ is complex analytic for all $\lambda \in \mathbb{C}$ such that $|\lambda| < r$.

If X is a complex Banach space, let Y denote the n-fold Cartesian product of X and for $y = (y_1, y_2, \cdots, y_n) \in Y$, define a seminorm $p(y)$ by

$$p(y) = \max\{\|y_i - y_j\| : 1 \le i, j \le n\}, \quad y = (y_1, y_2, \cdots, y_n). \tag{3.3}$$

Then $p(y) = 0$ if and only if $y \in S$, where

$$S = \{y = (y_1, y_1, \cdots, y_1) : y_1 \in X\}. \tag{3.4}$$

LEMMA 3.1. *Let X, Y, p and S be as above. For $y^0 \in S$ and $\delta > 0$ let $B_\delta(y^0) = \{y \in Y : \|y - y^0\| < \delta\} = U$ and suppose that $f : U \to Y$ is an analytic map such that $f(y^0) = y^0$. Assume that there exists $c < 1$ and a constant k such that*

$$p(f(y)) \le cp(y) \tag{3.5}$$

and

$$\|f(y) - y\| \le kp(y) \tag{3.6}$$

for all $y \in U$. Then there exists $r > 0$ such that $f^m(y) \in U$ for all $m \ge 1$ whenever $y \in \overline{B_r(y^0)}$ and $f^m(y)$ converges uniformly in $y \in \overline{B_r(y^0)}$ as $m \to \infty$ to a limit $g(y) \in S$ such that $f(g(y)) = g(y)$. The map $y \to g(y)$ is complex analytic on $B_r(y^0)$.

PROOF. For definiteness, define $\|y\| = \max_{1 \le i \le n} \|y_i\|$. If a sequence of analytic functions converges uniformly on an open set G to a limit g, g is analytic on G. Thus, in our case, to prove g is analytic on $B_r(y^0)$ for some $r > 0$, it suffices to prove that f^m (which is analytic) is defined and converges uniformly on $\overline{B_r(y^0)}$.

Take $r_0 > 0$ such that

$$r_0 + 2kr_0/(1 - c) < \delta$$

and note that $p(y) \le 2r_0$ for all $y \in B_{r_0}(y^0)$. Define

$$r_j = r_0 + 2kr_0\left(\sum_{i=0}^{j=1} c^i\right) < \delta, \quad j \ge 1,$$

and assume that if $y \in B_{r_0}(y^0)$, then $f^j(y) \in B_{r_j}(y^0)$ for $0 \le j \le m$ and $p(f^j(y)) \le c^j(2r_0)$ for $0 \le j \le m$. Then (3.5) and (3.6) imply that

$$\|f^{m+1}(y) - f^m(y)\| \le kp(f^m(y)) \le kc^m(2r_0),$$

so
$$\|f^{m+1}(y)-y^0\| \leq \|f^{m+1}(y)-f^m(y)\| + \|f^m(y)-y^0\|$$
$$\leq r_0 + 2kr_0\sum_{i=0}^{m-1} c^i + 2kr_0c^m = r_{m+1}.$$

By mathematical induction we find that $f^j(y) \in B_{r_j}(y^0)$ for all $j \geq 1$.

If $y \in B_{r_0}(y^0)$ and m and v are positive integers, $m < v$, then

$$\begin{aligned}\|f^m(y)-f^v(y)\| &\leq \sum_{i=1}^{v-m}\|f^{m+i}(y)-f^{m+i-1}(y)\| \\ &\leq k\sum_{i=1}^{v-m} p(f^{m+i-1}(y)) \\ &\leq k(2r_0)\sum_{i=1}^{v-m} c^{m+i-1} \\ &\leq k(2r_0)c^m/(1-c).\end{aligned} \tag{3.7}$$

(3.7) shows that $(f^m(y))$ is a Cauchy sequence with limit $g(y)$. If $v \to \infty$ in (3.7), then
$$\|f^m(y)-g(y)\| \leq 2r_0kc^m/(1-c),$$
so the convergence is uniform in $y \in B_{r_0}(y^0)$. Obviously $g(y) \in U$ and $f(g(y)) = g(y)$, and because $p(f(g(y))) = p(g(y)) \leq cp(g(y))$, $p(g(y)) = 0$ and $g(y) \in S$. □

Under the hypotheses of Lemma 3.1, convergence is "linear", whereas for the examples of interest to us, convergence is actually "quadratic" (see [35], Chapter 12 for definitions) and hence extremely rapid. The next lemma describes the situation we shall actually encounter.

LEMMA 3.2. *Let the notation and assumptions be as in Lemma* 3.1 *except instead of assuming that f satisfies inequality* (3.5), *suppose that there exists a constant M such that*

$$p(f(y)) \leq M(p(y))^2 \tag{3.8}$$

for all $y \in U$. Then the conclusions of Lemma 3.1 *are still valid.*

Furthermore, if δ in Lemma 3.1 *is so small that*

$$\sup_{y\in U} Mp(y) \leq c < 1, \tag{3.9}$$

then, setting $u = u_m = 2^m$,

$$\|f^m(y)-g(y)\| \leq (kM^{-1})c^u(1-c^u)^{-1} \tag{3.10}$$

for all $y \in \overline{B_{r_0}(y^0)}$ *and* $m \geq 0$.

PROOF. By decreasing δ we can assume that (3.9) is satisfied on $B_\delta(y^0)$. (3.8) then implies that (3.5) is satisfied, so the conclusions of Lemma 3.1 hold. An easy induction shows that if we define $u_j = 2^j$, then

$$p(f^m(y)) \leq M^{-1}(Mp(y))^{u_m} \leq M^{-1}c^{u_m}. \tag{3.11}$$

If we use (3.6) and (3.11) we find

$$\begin{aligned} \|f^m(y) - g(y)\| &= \lim_{v \uparrow \infty} \|f^m(y) - f^v(y)\| \\ &\leq \sum_{j=m}^{\infty} \|f^{j+1}(y) - f^j(y)\| \leq kM^{-1} \sum_{j=m}^{\infty} c^{u_j}. \end{aligned} \tag{3.12}$$

If we define $\rho = c^{2^m}$, it is easy to see that

$$\sum_{j=m}^{\infty} c^{u^j} \leq \sum_{i=1}^{\infty} \rho^i = \rho(1-\rho)^{-1}, \tag{3.13}$$

and (3.12) and (3.13) give (3.10). □

Next we establish a theorem which, as we shall see later, is applicable to the "monster algorithm" of Corollary 2.6.

THEOREM 3.1. *Let X be a complex Banach space and Y the n-fold Cartesian product of X with itself. Let p and S be as defined in* (3.3) *and* (3.4).

Suppose that V is an open subset of Y and $f : V \to Y$ is a complex analytic map. Define $W \subseteq V$ by

$$\begin{aligned} W = \{y \in V : f^m(y) \in V \text{ for all } m \geq 1 \text{ and there exists} \\ u \in V \cap S \text{ such that } \lim_{m \uparrow \infty} \|f^m(y) - u\| = 0\}. \end{aligned} \tag{3.14}$$

(The element u in (3.14) *depends on y). For each $u \in V \cap S$, assume there exist positive constants c, δ and k (dependent on u) such that $c < 1$ and*

$$p(f(y)) \leq cp(y) \quad \text{and} \quad \|f(y) - y\| \leq kp(y) \tag{3.15}$$

for all $y \in B_\delta(u) = \{y : \|y - u\| < \delta\}$. Then W is an open set and

$$g(y) = \lim_{m \uparrow \infty} f^m(y), \quad y \in W,$$

is an analytic function on W.

PROOF. If $y \in W$, select $u \in V \cap S$ such that $f^m(y)$ converges to u. If c, δ and k are as above, select r_0 as in Lemma 3.1, so that if $w \in \overline{B_{r_0}(u)}$, then $f^j(w) \in B_\delta(u)$ for all $j \geq 1$ and

$$\lim_{j\uparrow\infty} f^j(w) = g(w),$$

where g is an analytic function on $B_{r_0}(u)$. There exists an integer N so that $f^N(y) \in B_{r_0}(u)$, and by continuity of f^N, there exists $\delta_1 > 0$ so that $f^N(z) \in B_{r_0}(u)$ for all z such that $\|z - y\| < \delta_1$. It follows that if $\|z - y\| < \delta_1$,

$$g(z) = \lim_{m\uparrow\infty} f^m(z) = g(f^N(z)) \in S. \tag{3.16}$$

Thus we see that W is open. Also, because the restriction of g to $B_{r_0}(u)$ is analytic (by Lemma 3.1) and f^N is analytic, (3.16) implies that g is analytic on an open neighborhood of y. □

The argument of lemma 3.2 shows that to verify (3.15) in Theorem 3.1, it suffices to verify (3.6) and (3.8). To accomplish this for the examples of interest to us we need the next fact.

LEMMA 3.3. *Let H be a Banach space and $X = \mathcal{L}(H)$. Suppose that $A_0 \in X$ and $\sigma(A_0) \cap (-\infty, 0]$ is empty. Then there exists $\delta > 0$ and $M > 0$ such that* (1) *for all $A \in B_\delta(A_0)$, the open δ ball about A_0, $\sigma(A) \cap (-\infty, 0]$ is empty and* (2) *if $A_1, A_2, \cdots, A_m$ are any elements of $B_\delta(A_0)$ and $\alpha_1, \alpha_2, \cdots, \alpha_m$ are positive reals such that $\sum_{j=1}^{m} \alpha_j = 1$, then*

$$\|\log(\sum_{j=1}^{m} \alpha_j A_j) - \sum_{j=1}^{m} \alpha_j \log A_j\| \leq M \sum_{1\leq i\leq j\leq m} \|A_i - A_j\|^2. \tag{3.17}$$

PROOF. By assumption, $(1-t)A_0 + tI$ is invertible for $0 \leq t \leq 1$, and by continuity of the map $A \to A^{-1}$ on the set of invertible linear operators, $\|[(1-t)A_0 + tI]^{-1}\|$ is uniformly bounded for $0 \leq t \leq 1$, say by a constant M_1. If $\|A - A_0\| < \delta$ and $\delta < (2M_1)^{-1}$, then writing $A_{0t} = tA_0 + (1-t)I$ and $A_t = tA + (1-t)I$ for $0 \leq t \leq 1$, one has

$$A_t = A_{0t}[I + A_{0t}^{-1}t(A - A_0)],$$

so A_t is invertible (the product of invertible operators) and

$$\|A_t^{-1}\| \leq \|A_{0t}^{-1}\| \sum_{s=0}^{\infty} (\delta M_1)^j \leq 2\|A_{0t}^{-1}\| \leq 2M_1, \quad 0 \leq t \leq 1.$$

To prove (3.17), first assume that $m = 2$ and take A and B in $B_\delta(A_0)$ and α so that $0 < \alpha < 1$. If $C \in B_\delta(A_0)$ and we write $C_t = I + t(C - I)$ for $0 \leq t \leq 1$, we have

$$\log C = \int_0^1 (d/dt) \log(C_t) dt = \int_0^1 C_t^{-1}(C - I) dt \tag{3.18}$$
$$= \int_0^1 t^{-1}[I - C_t^{-1}] dt.$$

If we take $C = \alpha A + (1 - \alpha)B$, $C = A$ and $C = B$ in (3.18), we find after simplification that

$$\log(\alpha A + (1 - \alpha)B) - \alpha \log A - (1 - \alpha) \log B \tag{3.19}$$
$$= \int_0^1 t^{-1}\{\alpha A_t^{-1} + (1 - \alpha)B_t^{-1} - [\alpha A_t + (1 - \alpha)B_t]^{-1}\} dt,$$

where

$$A_t = tA + (1 - t)I \text{ and } B_t = tA + (1 - t)I. \tag{3.20}$$

By using Lemma 1.5 and (1.34) we obtain

$$\alpha A_t^{-1} + (1 - \alpha)B_t^{-1} - [\alpha A_t + (1 - \alpha)B_t]^{-1} \tag{3.21}$$
$$= \alpha(1 - \alpha)B_t^{-1}(A_t - B_t)A_t^{-1} \ [(1 - \alpha)A_t^{-1} + \alpha B_t^{-1}]^{-1} \ B_t^{-1}(A_t - B_t)A_t^{-1}.$$

Using the identity

$$[(1 - \alpha)A_t^{-1} + \alpha B_t^{-1}]^{-1} = A_t[(1 - \alpha)B_t + \alpha A_t]^{-1} B_t$$

in (3.21) and simplifying we find

$$\alpha A_t^{-1} + (1 - \alpha)B_t^{-1} - [\alpha A_t + (1 - \alpha)B_t]^{-1} \equiv D_t = \tag{3.22}$$
$$\alpha(1 - \alpha)t^2 B_t^{-1}(B - A)[(1 - \alpha)B_t + \alpha A_t]^{-1}(B - A)A_t^{-1}.$$

Since $\|B_t^{-1}\| \leq 2M_1$, $\|A_t^{-1}\| \leq 2M_1$ and

$$\|[(1 - \alpha)B_t + \alpha A_t]^{-1}\| \leq 2M_1,$$

we obtain from (3.22) (using also that $\alpha(1 - \alpha) \leq (1/4)$) that

$$\|D_t\| \leq 2M_1^3 t^2 \|B - A\|^2.$$

Using this estimate in (3.19) yields

$$(3.23)\qquad \begin{aligned} &\|\log(\alpha A + (1-\alpha)B) - \alpha \log A - (1-\alpha)\log B\| \\ &\leq 2M_1^3\|B-A\|^2 \int_0^1 t dt = M_1^3\|B-A\|^2. \end{aligned}$$

Our argument actually shows that if U is an open neighborhood of A_0 such that $((1-t)A + tI)^{-1}$ exists for all $A \in U$ and $0 \leq t \leq 1$ and

$$\|[(1-t)A + tI]^{-1}\| \leq 2M_1,$$

then (3.23) is satisfied.

We now proceed by induction. We have proved the lemma for $m = 2$, the constant M in (3.17) being M_1^3. Assume for some $m > 2$ that we have proved the lemma for $m-1$ and that the constant M in (3.17) can be taken to be M_1^3. Let A_j and α_j, $1 \leq j \leq m$, be as in the statement of Lemma 3.3 and define $A = A_1$,

$$B = (1-\alpha_1)^{-1}\sum_{j=2}^{m} \alpha_j A_j$$

and $\alpha = \alpha_1$. Then $A, B \in B_\delta(A_0)$ and (3.23) gives

$$(3.24)\qquad \begin{aligned} &\|\log(\sum_{j=1}^{m} \alpha_j A_j) - \alpha_1 \log A_1 - (1-\alpha_1)\log B\| \\ &\leq M_1^3\|B-A\|^2 = M_1^3\|\sum_{j=2}^{m}(\alpha_j/(1-\alpha_1))(A_j - A_1)\|^2. \end{aligned}$$

Using the Cauchy-Schwarz inequality,

$$(3.25)\qquad \begin{aligned} \|\sum_{j=2}^{m}(\alpha_j/(1-\alpha_1))(A_j - A_1)\|^2 &\leq [\sum_{j=2}^{m}(\alpha_j/(1-\alpha_1))\|A_j - A_1\|]^2 \\ &\leq \sum_{j=2}^{m}\|A_j - A_1\|^2. \end{aligned}$$

On the other hand, the inductive assumption implies

$$(3.26)\qquad \begin{aligned} &\|(1-\alpha_1)\log B - (1-\alpha_1)\sum_{j=2}^{m}(\alpha_j/(1-\alpha_1))\log A_j\| \\ &\leq (1-\alpha_1)M_1^3(\sum_{2\leq i\leq j\leq m}\|A_i - A_j\|^2). \end{aligned}$$

Combining (3.24) - (3.26) we find that

$$\|\log(\sum_{j=1}^{m}\alpha_j A_j) - \sum_{j=1}^{m}\alpha_j \log(A_j)\| \leq M_1^3 \sum_{1\leq i\leq j\leq m} \|A_i - A_j\|^2, \tag{3.27}$$

so the lemma has been proved by induction. □

With these preliminaries we return to the "monster algorithm" of Corollary 1.5.

THEOREM 3.2. *Let H be a complex Banach space, let $X = \mathcal{L}(H)$ and let $U = \{A \in X : \sigma(A) \cap (-\infty, 0]$ is empty$\}$, where $\sigma(A)$ denotes the spectrum of A. If G is a finite group of order $n \geq 2$, let Y denote the Banach space of maps from G to X, so Y can be identified with the n-fold Cartesian products of X. Define $V = \{\Theta \in Y : \Theta(s) \in U$ for all $s \in G\}$ and define $F : V \to Y$ by*

$$(F\Theta)(t) = n^{-1} \sum_{s\in G} \exp((1/2)\log \Theta(s) + (1/2)\log \Theta(ts^{-1})). \tag{3.28}$$

Define S to be the set of constant functions in Y and define W by

$$W = \{\Theta \in V : F^m(\Theta) \in V \text{ for all } m \geq 1 \text{ and there exists}$$
$$\Psi \in V \cap S \text{ such that } \lim_{m\uparrow\infty} \|F^m(\Theta) - \Psi\| = 0\},$$

where F^m denotes the mth iterate of F and Ψ depends on Θ. Then W is an open subset of Y. If $g(\Theta)$ is defined by

$$g(\Theta) = \lim_{m\uparrow\infty} F^m(\Theta) \tag{3.29}$$

for $\Theta \in W$, the map $\Theta \to g(\Theta)$ is analytic. The convergence in (3.29) is quadratic. If H is a finite dimensional Hilbert space, W contains W_1, where

$$W_1 = \{\Theta \in Y : \Theta(s) \text{ is positive definite and self-adjoint for all } s \in G\}.$$

PROOF. Select $\Psi \in V \cap S$. By Theorem 3.1 it suffices to prove that there exists $\delta > 0$ such that (3.6) and (3.8) are satisfied for all $\Theta \in B_\delta(\Psi) = \{\Theta \in Y : \sup_{s\in G} \|\Theta(s) - \Psi(s)\| < \delta\}$. The map $A \to \exp(A)$ is C^1 with bounded Fréchet derivative on bounded sets in X, so the map $A \to \exp(A)$ is Lipschitzian on bounded sets. Similarly, for δ small enough, $A \to \log A$ is Lipschitzian on $\overline{B_\delta(\Psi)}$. Thus to prove that F satisfies inequality (3.8) on $B_\delta(\Psi)$ for some $\delta > 0$, it suffices to prove that there exists M such that

$$p(\log F(\Theta)) \leq M(p(\log \Theta))^2 \tag{3.30}$$

for all $\Theta \in B_\delta(\Psi)$, where, for $\alpha \in Y$,

$$p(\alpha) = \sup_{s_1, s_2 \in G} \|\alpha(s_1) - \alpha(s_2)\|. \tag{3.31}$$

We know that for $\Theta \in V$

$$n^{-1} \sum_{s \in G} (1/2) \log \Theta(s) + (1/2) \log \Theta(ts^{-1}) = n^{-1} \sum_{\sigma \in G} \log \Theta(\sigma). \tag{3.32}$$

Using Lemma 3.3 we find that there exists $\delta > 0$ and a constant M_1 such that for all $\Theta \in B_\delta(\Psi)$ one has

$$\| \log F(\Theta(t)) - n^{-1} \sum_{s \in G} [(1/2) \log \Theta(s) + (1/2) \log \Theta(ts^{-1})] \| \leq M_1 \sum_{s_1 . s_2 \in G} \| \beta(s_1, t) - \beta(s_2, t) \|^2, \tag{3.33}$$

where

$$\beta(s, t) \equiv \exp((1/2) \log \Theta(s) + (1/2) \log \Theta(ts^{-1})). \tag{3.34}$$

Because the map $A \to \exp A$ is Lipschitzian on bounded sets, we find from (3.32) - (3.34) that there exists a constant M_2 such that

$$\| \log F(\Theta(t)) - n^{-1} \sum_{s \in G} \log \Theta(s) \| \leq M_2 (p(\log \Theta))^2 \tag{3.35}$$

for all $\Theta \in B_\delta(\Psi)$. It follows from (3.35) and the triangle inequality that for all $t_1, t_2 \in G$,

$$\| \log F(\Theta(t_1)) - \log F(\Theta(t_2)) \| \leq 2M_2 (p(\log \Theta))^2 \tag{3.36}$$

so

$$p(\log F(\Theta)) \leq 2M_2 (p(\log \Theta))^2.$$

It remains to prove (3.6). By using the Lipschitz nature of $A \to \exp A$ and $A \to \log A$ on appropriate sets, it suffices to prove that there exist $\delta > 0$ and a constant k such that

$$\| \log F(\Theta) - \log \Theta \| \leq kp(\log \Theta)$$

for all $\Theta \in B_\delta(\Psi)$. By using (3.35) and the triangle inequality we see that

$$\| \log F(\Theta(t)) - \log \Theta(t) \| \leq \| \log F(\Theta(t)) - n^{-1} \sum_{s \in G} \log \Theta(s) \| + \| n^{-1} \sum_{s \in G} (\log \Theta(s) - \log \Theta(t)) \| \leq M_2 (p(\log \Theta))^2 + p(\log \Theta). \tag{3.37}$$

If M_3 is chosen so that $p(\log \Theta) \leq M_3$ for all $\Theta \in B_\delta(\Psi)$, (3.37) implies that

$$\| \log F(\Theta) - \log \Theta \| \leq (M_2 M_3 + 1) p(\log \Theta).$$

The final assertion of Theorem 3.2 follows immediately from Corollary 2.6 and the fact that strong convergence implies norm convergence in finite dimensions. □

The *AGM* of Section 1 is a special case of the monster algorithm when the group is of order 2. Thus:

COROLLARY 3.1. *Let H be a complex Banach space, $X = \mathcal{L}(H)$ and $Y = X \times X$. Let $U = \{A \in X : \sigma(A) \cap (-\infty, 0) = \phi\}$ and let $V = \{(A,B) \in Y : A \in U$ and $B \in U\}$. For $(A,B) \in V$, define $f(A,B)$ by*

$$f(A,B) = ((A+B)/2,\ \exp((1/2)\log A + (1/2)\log B)).$$

Define W by

$$W = \{(A,B) \in V : f^m(A,B) \in V \text{ for all } m \geq 1 \text{ and there exist}$$
$$A_\infty \in U \text{ and } B_\infty \in U \text{ such that } \lim_{m \uparrow \infty} \| f^m(A,B) - (A_\infty, B_\infty) \| = 0\}.$$

Then W is open and if $g(A,B)$ is defined by

$$g(A,B) = \lim_{m \uparrow \infty} f^m(A,B)$$

for $(A,B) \in W$, then $(A,B) \to g(A,B)$ is analytic. If H is a finite dimensional Hilbert space, W contains W_1, where

$$W_1 = \{(A,B) : A \text{ and } B \text{ are positive definite, bounded and self adjoint}\}.$$

PROOF. Corollary 3.1 follows from Theorem 3.2 if one observes that $A_\infty = B_\infty$ whenever

$$\lim_{m \uparrow \infty} \| f^m(A,B) - (A_\infty, B_\infty) \| = 0$$

for some $A_\infty \in U$, $B_\infty \in U$. This is because the form of f implies

$$A_\infty = (A_\infty + B_\infty)/2. \ \square$$

REMARK 3.1. It would be interesting to obtain more information about the set W in Corollary 3.1, even for H finite dimensional. For instance, is it true that almost every pair $(A,B) \in Y$ (with respect to Lebesgue measure) belongs to W? Numerical studies for dim $H = 2,3,4,5$ suggest this may be true.

If H is a Hilbert space and $A \in \mathcal{L}(H) = X$, A is called accretive if $\mathrm{Re} \langle Ax, x \rangle \geq 0$ for all $x \in H$, and A is strictly accretive if there exists

$\alpha > 0$ so that Re$< Ax, x > \geq \alpha\|x\|^2$ for all $x \in H$. It is natural to conjecture that for almost every pair (A, B) such that A and B are strictly accretive one has $(A, B) \in W$. However, one can give an example of 2×2 upper triangular accretive matrices A and B such that if $(A_1, B_1) = f(A, B)$, B_1 is not accretive. This, of course, does not disprove the conjecture.

Because f is homogeneous of degree 1 and

$$f(S^{-1}AS,\ S^{-1}BS) = S^{-1}f(A, B)S,$$

it is obvious that if $(A, B) \in W$, then $(\lambda A, \lambda B) \in W$ for all $\lambda > 0$ and $(S^{-1}AS,\ S^{-1}BS) \in W$ for all invertible S. If H is finite dimensional and A and B are both upper triangular matrices with spectrum strictly in the right half plane, one can also prove that $(A, B) \in W$, though we omit the proof.

There remains one easy case in which one can prove $(A, B) \in W$, that is, when $AB = BA$. Stickel [32] has discussed the commutative case when A and B are matrices, but his argument seeems incomplete. We shall sketch an approach which works when H is a complex Banach space.

Suppose that H is a complex Banach space and $A, B \in \mathcal{L}(H) = X$ are commuting linear operators. Consider the algebra $\mathcal{A}$ of complex-valued functions g which are defined and analytic on $U_g \times V_g$, where U_g is an open neighborhood of $\sigma(A)$ and V_g is an open neighborhood of $\sigma(B)$. Two such functions g_1 and g_2 are identified if they agree on $U \times V$, where U and V are some open sets containing $\sigma(A)$ and $\sigma(B)$ respectively. If $g \in \mathcal{A}$ is defined on $U \times V$, let $\Gamma_1 \subseteq U$ be a finite union of simple, closed rectifiable curves which contain $\sigma(A)$ in the union of their interiors, and similarly for $\Gamma_2 \subseteq V$. Define $g(A, B) \in \mathcal{L}(H)$ by

$$g(A, B) = (2\pi i)^{-2} \int_{\Gamma_2}\int_{\Gamma_1} g(z_1, z_2)(z_1 - A)^{-1}(z_2 - B)^{-1}\ dz_1 dz_2. \tag{3.38}$$

The operator $g(A, B)$ defined by (3.38) does not depend on the particular choice of Γ_1 and Γ_2 as above. Furthermore, the map $g \longrightarrow g(A, B) \in \mathcal{L}(H)$ is an algebra homomorphism from $\mathcal{A}$ to $\mathcal{L}(H)$. The proof of this fact is a minor variant of the argument (see [33], [36]) for defining the functional calculus for a single operator and will not be given here. In fact, the functional calculus summarized by (3.38) is a special case of a much more subtle functional calculus developed by Shilov, Waelbrock, Arens-Calderon and Arens: see [7] for references. If $g_n : U \times V \longrightarrow \mathbb{C}$, $n \geq 1$, is a sequence of analytic functions and g_n converges uniformly on $U \times V$ to g, then by using (3.38) one can easily see that $g_n(A, B)$ converges in norm to $g(A, B)$.

If U and V are as above and $g : U \times V \longrightarrow \mathbb{C}$ is analytic, then by using the fact that $g \longrightarrow g(A, B)$ is an algebra homomorphism and that the function identically equal to 1 goes to the identity in $\mathcal{L}(H)$ one can see that

$$\sigma(g(A, B)) \subseteq g(\sigma(A) \times \sigma(B)). \tag{3.39}$$

Applying (3.39) to $g(z,w) = (z+w)/2$ and $g(z,w) = zw$ gives

$$\sigma((A+B)/2) \subseteq \{(u+v)/2 : u \in \sigma(A),\ v \in \sigma(B)\} \text{ and}$$
(3.40)
$$\sigma(AB) \subseteq \{uv : u \in \sigma(A),\ v \in \sigma(B)\}.$$

We can also use the composition of analytic functions. If $g \in \mathcal{A}$ and h is analytic on an open neighborhood of $g(\sigma(A) \times \sigma(B)) \subseteq \mathbb{C}$, then $j = \mathrm{h} \circ \mathrm{g} \in \mathcal{A}$ and (3.39) implies that $h(g(A,B))$ is defined and one can prove, as for the functional calculus in one variable, that

(3.41) $$j(A,B) = h(g(A,B)).$$

If $g(z,w) = (g_1(z,w),\ g_2(z,w))$, where $g_j \in \mathcal{A}$ for $j = 1,2$, one can define

$$g(A,B) = (g_1(A,B),\ g_2(A,B)) = (A_1, B_1),$$

and it is easy to see that A_1 and B_1 commute. If W_j is an open neighborhood of $g_j(\sigma(A) \times \sigma(B))$ and $h : W_1 \times W_2 \to \mathbb{C}^2$ is analytic, then $j = \mathrm{h} \circ \mathrm{g}$ is defined on $U_0 \times V_0$, where U_0 is an open neighborhood of $\sigma(A)$ and V_0 an open neighborhood of $\sigma(B)$ and

(3.42) $$j(A,B) = (\mathrm{h} \circ \mathrm{g})(A,B) = h(g(A,B)).$$

Now define open subsets $G \subseteq \mathbb{C}$ and $\mathcal{O} \subseteq \mathcal{L}(H)$ by

$$G = \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\} \text{ and}$$
(3.43)
$$\mathcal{O} = \{A \in \mathcal{L}(H) : \sigma(A) \subseteq G\}.$$

Let $z^{1/2}$ denote the standard single-valued branch of the square root function:

(3.44) $$z^{1/2} = \rho^{1/2} e^{i\Theta/2}, \text{ where } z = \rho e^{i\Theta},\ \rho > 0 \text{ and } -\pi < \Theta < \pi.$$

It is easy to see that for z and $w \in G$

$$\exp((1/2)\log z + (1/2)\log w) = z^{1/2} w^{1/2} = (zw)^{1/2}.$$

Thus, using the properties of the functional calculus, we see that if $(A,B) \in \mathcal{O} \times \mathcal{O}$ and $AB = BA$, then

(3.45) $$\exp((1/2)\log A + (1/2)\log B) = (AB)^{1/2}.$$

If $G_1 = \{z \in \mathbb{C} : z \neq 0 \text{ and } z \text{ is not a negative real}\}$, then if $(A,B) \in \mathcal{O} \times \mathcal{O}$, (3.40) implies that $\sigma(AB) \subseteq G_1$, and the spectral mapping theorem implies that

(3.46) $$\sigma((AB)^{1/2}) \subseteq \{z^{1/2} : z \in G_1\} = G.$$

(3.40) also implies that $\sigma((A+B)/2) \subseteq G$ if $(A, B) \in \mathcal{O} \times \mathcal{O}$. It follows that if we define $W = \{(A, B) \in \mathcal{O} \times \mathcal{O} : AB = BA\}$ and if $F : W \to \mathcal{L}(H) \times \mathcal{L}(H)$ is defined by

$$F(A, B) = ((A+B)/2, \ (AB)^{1/2}), \tag{3.47}$$

then $F(W) \subseteq W$. Thus if $(A, B) \in W$ we can consider $F^m(A, B)$, where F^m is the mth iterate of $F : W \to W$.

On the other hand, if we define $f(z, w) = ((z+w)/2, \ (zw)^{1/2})$, then $f(G \times G) \subseteq G \times G$. For $(A, B) \in W$, we have $f(A, B) = F(A, B)$, where $f(A, B)$ is defined by (3.38). If f^m denotes the mth iterate of $f : G \times G \to G \times G$, the previously mentioned properties of the functional calculus (particularly the rules of composition) imply that

$$f^m(A, B) = F^m(A, B), \tag{3.48}$$

where the left side of (3.48) is defined by (3.38) with $g = f^m$.

The classical theory of the AGM implies that for all $(z, w) \in G \times G$ one has

$$\lim_{m \uparrow \infty} f^m(z, w) = (\pi/2)(g(z, w)^{-1}, \ g(z, w)^{-1}), \tag{3.49}$$

where

$$g(z, w) = \int_0^\infty (s^2 + z^2)^{-1/2}(s^2 + w^2)^{-1/2} ds, \tag{3.50}$$

and that the convergence in (3.49) is uniform on compact subsets of $G \times G$. It follows from (3.49) and (3.50) that

$$\lim_{m \uparrow \infty} F^m(A, B) = (\pi/2)(g(A, B))^{-1}, \tag{3.51}$$

where $g(A, B)$ in (3.51) is defined by (3.38) and g is as in (3.50). Finally, a relatively simple argument (which we omit) shows that

$$g(A, B) = \int_0^\infty (s^2 + A^2)^{-1/2} \ (s^2 + B^2)^{-1/2} ds, \tag{3.52}$$

where the right side of (3.52) is interpreted as an improper Riemann integral with values in $\mathcal{L}(H)$. Thus:

PROPOSITION 3.1. *Let H be a complex Banach space,*

$$G = \{z \in \mathbb{C} : \mathrm{Re}(z) > 0\},$$

$$G_1 = \{z \in \mathbb{C} : z \neq 0 \text{ and } z \text{ is not a negative real}\}$$

and

$$W = \{(A,B) \in \mathcal{L}(H) \times \mathcal{L}(H) : AB = BA,\ \sigma(A) \subseteq G \text{ and } \sigma(B) \subseteq G\}.$$

If $(A,B) \in W$, *then* $\sigma((A+B)/2) \subseteq G$, $\sigma(AB) \subseteq G$, *and if, for* $(A,B) \in W$, $F(A,B)$ *is defined by*

$$F(A,B) = ((A+B)/2,\ \exp((1/2)\log A + (1/2)\log B)) = ((A+B)/2,\ (AB)^{1/2}),$$

then $F(W) \subseteq W$ *and*

$$n\text{-}\lim_{m\uparrow\infty} F^m(A,B) = (\pi/2)(g(A,B)^{-1},\ g(A,B)^{-1}), \tag{3.53}$$

where $g(A,B)$ *is defined by* (3.52).

4. - Alternate definitions of the AGM

In the previous section we have considered one type of generalization of the classical AGM. However, there are many possible "reasonable" generalizations of the AGM to pairs of bounded linear operators (A,B). In fact, as we shall see later, there is a continuum of arithmetic-geometric means, all of which are defined when A and B are p.d. and self-adjoint, all of which give the same value when $AB = BA$, but all of which in general give different values when $AB \neq BA$.

We begin with an observation which was made to the authors by the referee of our earlier paper [14]. If A and B are $n \times n$ Hermitian, positive definite matrices, define $f(A,B)$ by

$$f(A,B) = ((A+B)/2,\ B(B^{-1}A)^{1/2}), \tag{4.1}$$

where $(B^{-1}A)^{1/2}$ is defined by (1.1). The referee remarked that, despite appearances, the expression $B(B^{-1}A)^{1/2}$ is symmetric in A and B and is positive definite and self-adjoint. Furthermore, he observed that for A and B p.d. and self-adjoint, it is relatively easy to prove that

$$\lim_{k\uparrow\infty} f^k(A,B) = (C,C), \tag{4.2}$$

where C is p.d. and self-adjoint.

If A and B are p.d., self-adjoint bounded linear maps of a Hilbert space H to itself, J. Fujii [19] has defined a map g by

$$g(A,B) = ((A+B)/2,\ A\#B), \tag{4.3}$$

where $A\#B$ is a "geometric mean" introduced by Pusz and Woronowicz [31]. Pusz and Woronowicz proved (see also Theorem 2 in [3]) that

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}. \tag{4.4}$$

Fujii proves that if A and B are p.d., self-adjoint operators in $\mathcal{L}(H)$, then

$$s\text{-}\lim_{k\uparrow\infty} g^k(A,B) = (C,C), \tag{4.5}$$

where C is p.d. and self-adjoint.

We shall show first that

$$f(A,B) = g(A,B),$$

where $f(A,B)$ is defined by (4.1) and $g(A,B)$ by (4.3) and (4.4). This will imply of course that the limits defined in (4.2) and (4.5) are equal. We begin with a trivial lemma.

LEMMA 4.1. *Let H be a Hilbert space and suppose the $A, B \in \mathcal{L}(H)$ and A and B are p.d. and self-adjoint. Then $\sigma(AB) =$ spectrum of $AB \subseteq (0,\infty)$.*

PROOF. Because $\sigma(AB) = \sigma(A^{-1/2}(AB)A^{1/2}) = \sigma(A^{1/2}BA^{1/2})$, and $A^{1/2}BA^{1/2}$ is p.d. and self-adjoint, the lemma is proved. □

We now recall some basic fact about the functional calculus for linear operators. If H is a Hilbert space, $A \in \mathcal{L}(H)$ and f is analytic on an open neighborhood U of $\sigma(A) \cup \sigma(A^*)$ and $f(\bar{z}) = \overline{f(z)}$ for all z in U, then $f(A^*) = (f(A))^*$. If H is a Banach space, $A \in \mathcal{L}(H)$, f is analytic on an open neighborhood of $\sigma(A)$ and S is invertible, then

$$f(S^{-1}AS) = S^{-1}f(A)S. \tag{4.6}$$

We shall always use the standard single-valued branches of z^λ and $\log z$. Thus if

$$G_1 = \{z \in \mathbb{C} : z \neq 0 \text{ and } z \text{ is not a negative real}\}, \tag{4.7}$$

and $z = re^{i\Theta}$, $|\Theta| < \pi$, $r > 0$, then

$$z^\lambda = r^\lambda e^{i\lambda\Theta} \quad \text{and} \quad \log z = \log r + i\Theta.$$

It follows that if H is a complex Banach space, $A \in \mathcal{L}(H)$ and $\sigma(A) \subseteq G_1$, then for any real numbers λ and μ,

$$A^\lambda A^\mu = A^\lambda + \mu \tag{4.8}$$

and if λ and μ are real numbers such that $|\lambda| \leq 1$ (so $z^\lambda \in G_1$ for all $z \in G_1$)

$$(A^\lambda)^\mu = A^{\lambda\mu}, \quad |\lambda| \leq 1. \tag{4.9}$$

LEMMA 4.2. *Suppose H is a complex Banach space and A and B are in $\mathcal{L}(H)$, B is invertible and $\sigma(B^{-1}A) \subseteq G_1$, where G_1 is as in (4.7). Then A and BA^{-1} are invertible and for any real number λ,*

$$\begin{aligned} &\sigma(B^{-\lambda}AB^{\lambda-1}) = \sigma(B^{-1}A) \subseteq G_1 \text{ and} \\ &\sigma(A^{-\lambda}BA^{\lambda-1}) = \sigma(A^{-1}B) \subseteq G_1. \end{aligned} \tag{4.10}$$

Furthermore, for all real λ,

$$B^\lambda(B^{-\lambda}AB^{\lambda-1})^{1/2}B^{1-\lambda} = B(B^{-1}A)^{1/2}, \tag{4.11}$$

$$A^\lambda(A^{-\lambda}BA^{\lambda-1})^{1/2}A^{1-\lambda} = A(A^{-1}B)^{1/2}, \tag{4.12}$$

$$A(A^{-1}B)^{1/2} = B(B^{-1}A)^{1/2}. \tag{4.13}$$

If H is a Hilbert space and A and B are p.d. and self-adjoint, $B(B^{-1}A)^{1/2}$ is p.d. and self-adjoint.

PROOF. Because B is invertible and $B^{-1}A$ is invertible, $A = B(B^{-1}A)$ is invertible and $A^{-1}B$ and BA^{-1} are invertible. By using (4.8), we can write

$$B^{-\lambda}AB^{\lambda-1} = S(B^{-1}A)S^{-1}, \quad S = B^{1-\lambda},$$

so

$$\sigma(B^{-1}A) = \sigma(S(B^{-1}A)S^{-1}) \subseteq G_1.$$

By interchanging the roles of A and B, we obtain the other part of (4.10).

If $S = B^{1-\lambda}$, (4.6) and (4.8) yield

$$\begin{aligned} B^\lambda(B^{-\lambda}AB^{\lambda-1})^{1/2}B^{1-\lambda} &= BS^{-1}(B^{-\lambda}AB^{\lambda-1})^{1/2}S \\ &= B(S^{-1}B^{-\lambda}AB^{\lambda-1}S)^{1/2} = B(B^{-1}A)^{1/2}, \end{aligned}$$

which is (4.11). (4.12) is obtained by a similar argument.

(4.13) is equivalent to

$$(B^{-1}A)(A^{-1}B)^{1/2} = (B^{-1}A)^{1/2}. \tag{4.14}$$

However, (4.9) implies that

$$(A^{-1}B)^{1/2} = [(B^{-1}A)^{-1}]^{1/2} = (B^{-1}A)^{-1/2},$$

which yields (4.14).

If H is a Hilbert space and A and B are p.d. and self-adjoint, Lemma 4.1 implies that $\sigma(B^{-1}A) \subseteq (0,\infty)$. Thus the first part of the lemma is applicable, and taking $\lambda = 1/2$ in (4.11) and (4.12) gives

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

It is easy to see that $B\#A$ is self-adjoint and p.d., so $B(B^{-1}A)^{1/2}$ is also. □

Lemma 4.2 implies that the functions given by (4.1) and (4.3) respectively are equal when A and B are p.d., so the fact that (4.2) is valid (in the strong operator topology) follows from Fujii's theorem.

We shall now show that a much stronger result than (4.2) is valid. Let H be a complex Banach space and $X = \mathcal{L}(H)$. For k a positive integer, define

$$G_k = \{z = re^{i\Theta} : r > 0 \text{ and } |\Theta| < \pi/2^{k-1}\} \tag{4.15}$$

and

$$U_k = \{(A,B) \in X \times X : B \text{ is invertible and } \sigma(B^{-1}A) \subseteq G_k\}. \tag{4.16}$$

LEMMA 4.3. *Let H be a complex Banach space and let U_k be as in* (4.16) *for positive integers k. For $(A,B) \in U_1$, define $f(A,B)$ by*

$$f(A,B) = ((A+B)/2,\ B(B^{-1}A)^{1/2}). \tag{4.17}$$

If $(A,B) \in U_k$, then $f(A,B) = (A_1,B_1) \in U_{k+1}$.

PROOF. By assumption, $\sigma(B^{-1}A) \subseteq G_k$, so the spectral mapping theorem implies that

$$\sigma((B^{-1}A)^{1/2}) \subseteq \{z^{1/2} : z \in G_k\} = G_{k+1},$$

and $(B^{-1}A)^{1/2}$ is invertible. It is assumed that B is invertible, so

$$B_1 = B(B^{-1}A)^{1/2}$$

is invertible. Then

$$B_1^{-1}A_1 = (B^{-1}A)^{-1/2}B^{-1}((A+B)/2) = [(B^{-1}A)^{1/2} + (B^{-1}A)^{-1/2}]/2. \tag{4.18}$$

It follows from (4.18) and the spectral mapping theorem that

$$\begin{aligned} \sigma(B_1^{-1}A_1) &\subseteq \{(z^{1/2}+z^{-1/2})/2 : z \in \sigma(B^{-1}A)\} \\ &\subseteq \{(z^{1/2}+z^{-1/2})/2 : z \in G_k\}. \end{aligned} \tag{4.19}$$

If $z \in G_k$, $z^{1/2} \in G_{k+1}$, and if $w \in G_k$, $w^{-1} \in G_k$; also, G_{k+1} is convex for $k \geq 1$. Thus we conclude from (4.19) that $\sigma(B_1^{-1}A_1) \subseteq G_{k+1}$. □

LEMMA 4.4. *Let* U_1, f *and* H *be as in Lemma* 4.3. *If* C *and* D *are elements of* $\mathcal{L}(H)$ *and* $(A, B) \in U_1$, *define*

$$C(A, B) = (CA, CB) \text{ and } (A, B)D = (AD, BD).$$

If C *and* D *are invertible and* $(A, B) \in U_1$, *then* $C(A, B)D \in U_1$ *and*

$$f^m(C(A, B)D) = Cf^m(A, B)D \tag{4.20}$$

for all $m \geq 1$.

PROOF. If $(A, B) \in U_1$, B is invertible, so CBD is also invertible. Then

$$C(A, B)D = (CAD, CBD)$$

so $(CBD)^{-1}(CAD) = (D^{-1}B^{-1}C^{-1})(CAD) = D^{-1}(B^{-1}A)D$ and

$$\sigma((CBD)^{-1}(CAD)) = \sigma(D^{-1}(B^{-1}A)D) = \sigma(B^{-1}A) \subseteq G_1.$$

This shows that $C(A, B)D \subseteq U_1$.

The above calculation also shows that $((CBD)^{-1}(CAD))^{1/2} = D^{-1}(B^{-1}A)^{1/2}D$, so $(CBD)((CBD)^{-1}(CAD))^{1/2} = CB(B^{-1}A)^{1/2}D$. It follows that

$$f(C(A, B)D) = Cf(A, B)D,$$

and a simple mathematical induction (left to the reader) gives (4.20). □

With these preliminaries we can prove an extension of Fujii's theorem.

THEOREM 4.1. *Let* H *be a complex Banach space,* $X = \mathcal{L}(H)$, $G_1 = \{z = re^{i\Theta} \in \mathbb{C} : r > 0$ *and* $|\Theta| < \pi\}$, *and* $U_1 = \{(A, B) \in X \times X : B$ *is invertible and* $\sigma(B^{-1}A) \subseteq G_1\}$. *For* $(A, B) \in U_1$, *define*

$$f(A, B) = ((A + B)/2,\ B(B^{-1}A)^{1/2}).$$

Then $f(U_1) \subseteq U_1$ *and for every* $(A, B) \in U_1$ *there exists* $E \in X$, E *invertible, such that*

$$n\text{-}\lim_{k\uparrow\infty} f^k(A, B) = (E, E). \tag{4.21}$$

If H *is a Hilbert space and* A *and* B *are p.d. and self-adjoint, then* $(A, B) \in U_1$, *and if* $(A_1, B_1) = f(A, B)$, A_1 *and* B_1 *are p.d. and self-adjoint.*

PROOF. Lemma 4.3 implies that $f^k(A, B) \in U_{k+1} \subseteq U_1$, where U_k is given by (4.16). If we write $f^k(A, B) = (A_k, B_k)$, we obtain from Lemma 4.4 that

$$f^k(A, B) = A_1 f^{k-1}(I, A_1^{-1}B_1),\ \ k \geq 2. \tag{4.22}$$

If G_2 is as in (4.15) and C and D are any commuting bounded linear operators such that $\sigma(C) \subseteq G_2$ and $\sigma(D) \subseteq G_2$, the properties of the functional calculus for commuting operators (see the end of Section 3) imply that

$$f(C,D) = ((C+D)/2,\ D(D^{-1}C)^{1/2}) = ((C+D)/2,\ (CD)^{1/2}) = (C_1, D_1).$$

Thus, for such C and D, Proposition 3.1 implies that there exists an invertible $E \in X$ such that

$$n\text{-}\lim_{k\uparrow\infty} f^{k-1}(C,D) = (E,E). \tag{4.23}$$

Now take $C = I$ and $D = A_1^{-1}B_1$. Lemma 4.3 implies that $\sigma(D) \subseteq G_2$ and of course these C and D commute. Thus we conclude from (4.22) and (4.23) that

$$\lim_{k\uparrow\infty} \|f^k(A,B) - (E,E)\| = 0.$$

The statements in Theorem 4.1 concerning the self-adjoint case follow directly from lemmas 4.1 and 4.2. □

REMARK 4.1. The above argument shows that

$$\lim_{k\uparrow\infty} f^k(A,B) = (\pi/2)A_1(g(I, A_1^{-1}B_1)^{-1},\ g(I, A_1^{-1}B_1)^{-1}), \tag{4.24}$$

where $(A_1, B_1) = f(A,B)$ and $g(A,B)$ is defined in (3.52). If $\sigma(A^{-1}B)$ is contained in the right half plane, one can replace A_1 by A and B_1 by B in (4.24).

REMARK 4.2. Even if A and B are p.d. and self-adjoint, Theorem 4.1 gives new information: the convergence in (4.21) is in the operator norm, whereas in Fujii's theorem convergence is in the strong operator topology.

Theorem 4.1 gives a reasonable definition of the AGM which does not in general agree with the definition in Section 1. We now show that there is a family of reasonable definitions of the AGM, parametrized by $\lambda \geq 1$, which reduce to the definition in Theorem 4.1 for $\lambda = 1$ and give the definition in Section 1 for $\lambda = \infty$.

It is convenient to prove another lemma first.

LEMMA 4.5. *Let H be a complex Hilbert space and suppose that $A, B \in \mathcal{L}(H)$ and A and B are p.d. and self-adjoint. Then for every $\lambda \geq 1$*

$$B(B^{-1}A)^{1/2} \leq (A+B)/2 \leq ((A^\lambda + B^\lambda)/2)^{1/\lambda}. \tag{4.25}$$

If $B \leq A$, then also

$$B \leq B(B^{-1}A)^{1/2}. \tag{4.26}$$

PROOF. The right inequality in (4.25) is inequality (1.43) and has already been proved. By multiplying on the left and the right by $B^{-1/2}$, one sees that $B(B^{-1}A)^{1/2} \leq (A+B)/2$ if and only if

$$(4.27)\quad B^{1/2}(B^{-1}A)^{1/2}B^{-1/2} = (B^{-1/2}AB^{-1/2})^{1/2} \leq (B^{-1/2}AB^{-1/2}+I)/2.$$

If we define

$$L = (B^{-1/2}AB^{-1/2})^{1/2},$$

L is a p.d., self-adjoint operator, and (4.27) is equivalent to $2L \leq L^2 + I$, which is certainly true, because $0 \leq (L-I)^2 = L^2 - 2L + I$.

If $B \leq A$, then $I = B^{-1/2}BB^{-1/2} \leq B^{-1/2}AB^{-1/2}$ and

$$(4.28)\qquad I \leq (B^{-1/2}AB^{-1/2})^{1/2}.$$

Multiplying (4.28) on the left and right by $B^{1/2}$ and using Lemma 4.2 gives

$$B \leq B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2} = B(B^{-1}A)^{1/2}. \ \Box$$

If H is a complex Hilbert space and A, B are p.d., self-adjoint bounded linear operators, define, for fixed $\lambda \geq 1$,

$$(4.29)\qquad f_\lambda(A,B) = ((A+B)/2,\ [B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2}]^\lambda) = (A_1, B_1).$$

Since $A^{1/\lambda}$ and $B^{1/\lambda}$ are p.d. and self-adjoint, Lemma 4.2 implies that B_1 is p.d. and self- adjoint.

THEOREM 4.2. *Let H be a complex Hilbert space. For $A, B \in \mathcal{L}(H)$, A and B p.d. and self-adjoint, define $f_\lambda(A,B)$ by* (4.29). *Then for any pair (A,B) of p.d., bounded, self-adjoint linear operators,*

$$(4.30)\qquad s\text{-}\lim_{k\uparrow\infty} f_\lambda^k(A,B) = (E,E),$$

where E is p.d. and self-adjoint and f_λ^k denotes the kth iterate of f_λ.

PROOF. Define

$$g_\lambda(A,B) = ([(A^\lambda + B^\lambda)/2]^{1/\lambda},\ B(B^{-1}A)^{1/2}),$$

$$\Psi_\lambda(A,B) = (A^{1/\lambda},\ B^{1/\lambda}),$$

so

$$\Psi_\lambda^{-1}(A,B) = (A^\lambda, B^\lambda).$$

A simple calculation shows that

$$f_\lambda = \Psi_\lambda^{-1} g_\lambda \Psi_\lambda,$$

so

$$f_\lambda^k = \Psi_\lambda^{-1} g_\lambda^k \Psi_\lambda.$$

Thus it suffices to prove that if A and B are p.d. and self-adjoint, there exists E such that E is p.d. and self-adjoint and $g_\lambda^k(A, B) \to (E, E)$. If we define

$$g_\lambda^k(A, B) = (A_k, B_k),$$

Lemma 4.5 implies that

$$A_k \geq B_k \quad \text{for } k \geq 1. \tag{4.31}$$

Using (4.31) and Lemma 4.5 again,

$$B_{k+1} = B_k(B_k^{-1}A_k)^{1/2} \geq B_k, \tag{4.32}$$

so (B_k) is monotonic increasing for $k \geq 1$.

Select positive constants α and β so that

$$\alpha I \leq A_1, \ B_1 \leq \beta I.$$

Assume by mathematical induction that

$$\alpha I \leq A_k, \ B_k \leq \beta I \tag{4.33}$$

for some $k \geq 1$. The spectral mapping theorem implies that

$$\alpha^\lambda I \leq A_k^\lambda, \ B_k^\lambda \leq \beta^\lambda I$$

and

$$\alpha I \leq A_{k+1} = ((A_k^\lambda + B_k^\lambda)/2)^{1/\lambda} \leq \beta I.$$

Also

$$\alpha I \leq B_k \leq B_{k+1} \leq A_{k+1} \leq \beta I,$$

so (4.33) is satisfied for all $k \geq 1$ by mathematical induction.

From (4.32) and (4.33), there exists $E \in \mathcal{L}(H)$, E p.d. and self-adjoint so that for all $x \in H$,

$$\lim_{k \uparrow \infty} \|B_k x - Ex\| = 0.$$

In particular, for all $x \in H$

$$\lim_{k \uparrow \infty} (B_{k+1} - B_k)x = 0. \tag{4.34}$$

Using (4.11) to write B_{k+1} in a different form, we see that (4.34) implies

$$\lim_{k \uparrow \infty} B_k^{1/2}[(B_k^{-1/2} A_k B_k^{-1/2})^{1/2} - I] B_k^{1/2} x = 0 \tag{4.35}$$

for all $x \in H$. On the other hand, one can prove by using (4.33) that for $k \geq 1$

$$\|B_k^{1/2}[(B_k^{-1/2}A_kB_k^{-1/2})^{1/2} + I]B_k^{-1/2}\| \leq M, \tag{4.36}$$

where M is a constant independent of $k \geq 1$. Combining (4.35) and (4.36) we see that for all $x \in H$

$$\lim_{k\uparrow\infty}(A_k - B_k)x = \lim_{k\uparrow\infty}(H_kG_k)(x) = 0, \tag{4.37}$$

where

$$G_k = B_k^{1/2}[(B_k^{-1/2}A_kB_k^{-1/2})^{1/2} - I]B_k^{1/2}$$

and

$$H_k = B_k^{1/2}[(B_k^{-1/2}A_kB_k^{-1/2})^{1/2} + I]B_k^{-1/2}.$$

(4.37) implies that A_k also converges to E in the strong operator topology. □

The reason for considering the maps f_λ is:

THEOREM 4.3. *Let H be a complex Hilbert space and let $A, B \in \mathcal{L}(H)$ be p.d., self-adjoint operators such that*

$$\alpha I \leq A, \quad B \leq \beta I, \tag{4.38}$$

where α and β are positive reals. Then, for f_λ as in (4.29),

$$n\text{-}\lim_{\lambda\to\infty} f_\lambda(A, B) = \left(\frac{A+B}{2}, \ \exp\left(\left(\frac{1}{2}\right)\log A + \left(\frac{1}{2}\right)\log B\right)\right). \tag{4.39}$$

The convergence in (4.39) *is uniform for pairs A, B which satisfy* (4.38) *for fixed positive numbers α and β.*

PROOF. We shall use the standard "big oh" notation. Thus if $R_\lambda \in \mathcal{L}(H)$ is defined for all large λ, we shall write

$$R_\lambda = O(\lambda^{-p})$$

if there exists a constant M such that

$$\|R_\lambda\| \leq M\lambda^{-p} \tag{4.40}$$

for all $\lambda \geq \lambda_0$. In our case R_λ will always depend on operators A and B which satisfy (4.38), and the constant M in (4.40) and the number λ_0 can always be chosen to depend only on α and β.

The properties of the functional calculus imply that

$$A^{1/\lambda} = \exp((1/\lambda)\log A),$$
$$B^{-1/\lambda} = \exp((-1/\lambda)\log B).$$

By using the Taylor series for the exponential one obtains

$$B^{-1/\lambda} = I - (1/\lambda)\log B + O(\lambda^{-2}),$$
(4.41)
$$A^{1/\lambda} = I + (1/\lambda)\log A + O(\lambda^{-2}).$$

(4.41) gives

(4.42) $$B^{-1/\lambda}A^{1/\lambda} = I + (1/\lambda)(\log A - \log B) + O(\lambda^{-2}).$$

The binomial theorem is applicable to $(I + C)^{1/2}$ when $C \in \mathcal{L}(H)$ satisfies $\|C\| < 1$, so for λ large enouch (4.42) implies that

(4.43) $$(B^{-1/\lambda}A^{1/\lambda})^{1/2} = I - (1/[2\lambda])(\log B - \log A) + O(\lambda^{-2}).$$

By using (4.43) and the formula

$$B^{1/\lambda} = I + (1/\lambda)\log B + O(\lambda^{-2}),$$

we obtain

(4.44) $$B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2} = I + (1/[2\lambda])(\log A + \log B) + O(\lambda^{-2}).$$

If $C \in \mathcal{L}(H)$ and $\|C\| < 1$, one has the Taylor series

$$\log(I + C) = \sum_{k=1}^{\infty}(-1)^{k-1}C^k/k.$$

By applying this formula to (4.44), one finds that for large λ

(4.45) $$\log(B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2}) = (1/2\lambda)(\log A + \log B) + O(\lambda^{-2}).$$

The functional calculus implies that

(4.46) $$[B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2}]^{\lambda} = \exp[\lambda \log(B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2})],$$

and combining (4.45) and (4.46) yields

$$[B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2}]^{\lambda} = \exp((1/2)\log A + (1/2)\log B + O(1/\lambda))$$
(4.47)
$$= \exp((1/2)\log A + (1/2)\log B) + O(1/\lambda),$$

which completes the proof. □

REMARK 4.3. It is not hard to prove a direct analogue of Theorem 3.1 for the maps f_λ, $\lambda \geq 1$. In particular, one can prove that there exists a positive number ε, independent of $\lambda \geq 1$, such that if $C, D \in \mathcal{L}(H)$, $\|C\| < \varepsilon$ and $\|D\| < \varepsilon$ and $A = I + C$ and $B = I + D$, then $f_\lambda(A, B)$ is defined and

(4.48) $$n\text{-}\lim_{k \uparrow \infty} f_\lambda^k(A, B) = (M_\lambda(A, B),\ M_\lambda(A, B)),$$

and $M_\lambda(A, B)$ is an analytic function of (A, B). If we use the Lie bracket notation,

$$[A, B] = AB - BA$$

for operators A and B in $\mathcal{L}(H)$ and if we define $\mu = \lambda^{-1}$, then an unpleasant calculation (which we omit) proves that for $A = I + C$ and $B = I + D$ and ε sufficiently small (ε independent of $\lambda \geq 1$)

(4.49) $$\begin{aligned}[B^{1/\lambda}(B^{-1/\lambda}A^{1/\lambda})^{1/2}]^\lambda &= I + (1/2)(C + D) - (1/8)(C - D)^2 \\ &+ (1/16)(C - D)(C + D)(C - D) + ((\mu - 1)/24)([[C, D],\ C - D]) \\ &+ ((\mu - 1)^2/48)([[C, D],\ C - D]) + R_4(\mu, C, D).\end{aligned}$$

There exists a constant M independent of $\lambda \geq 1$ such that

(4.50) $$\|R_4(\mu, C, D)\| \leq M(\|C\| + \|D\|)^4,$$

where $R_4(\mu, C, D)$ is as in (4.49). By using (4.49), one can prove fairly easily that

(4.51) $$\begin{aligned}M_\lambda(A, B) &= I + (C + D)/2 - (1/16)(C - D)^2 \\ &+ (1/32)(C - D)(C + D)(C - D) \\ &+ ((\mu - 1)/48)([[C, D],\ C - D]) \\ &+ ((\mu - 1)^2/96)([[C, D],\ C - D]) + R(\mu, C, D),\end{aligned}$$

where

$$\|R(\mu, C, D)\| \leq M_1(\|C\| + \|D\|)^4$$

and M_1 is independent of $\lambda \geq 1$. By using (4.51) one can see that in general the operators $M_\lambda(A, B)$ are different for all $\lambda \geq 1$.

REMARK 4.4. If $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ is a probability vector and $A_1, A_2, \cdots, A_n$ are positive reals, one can consider $\prod_{i=1}^{n} A_i^{\sigma_i} \equiv A^\sigma$. This remark concerns what is a reasonable analogue of A^σ when $A_1, A_2, \cdots, A_n$ are positive definite linear operators. One possibility is (0.7). However, if $n = 2$, another possibility is

(4.51) $$(A_1, \sigma_1)\#(A_2, \sigma_2) \equiv A_1(A_1^{-1}A_2)^{\sigma_2}.$$

Using the methods of this section, one can easily show that the right side of (4.51) is positive definite and

$$(A_1, \sigma_1)\#(A_2, \sigma_2) = (A_2, \sigma_2)\#(A_1, \sigma_1).$$

However, if $n = 3$, there are at least three possible reasonable analogues of $A_1^{\sigma_1} A_2^{\sigma_2} A_3^{\sigma_3}$, where $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ is a probability vector. One is

$$(B^{\sigma_1+\sigma_2},\ \sigma_1 + \sigma_2)\#(A_3, \sigma_3), \tag{4.52}$$

where $B = (A_1, \sigma_1/(\sigma_1 + \sigma_2))\#(A_2, \sigma_2/(\sigma_1 + \sigma_2))$. Another is

$$(A_1, \sigma_1)\#(C^{\sigma_2+\sigma_3}, \sigma_2 + \sigma_3), \tag{4.53}$$

where $C = (A_2, \sigma_2/(\sigma_2+\sigma_3))\#(A_3, \sigma_3/(\sigma_2+\sigma_3))$. The formulas (4.52) and (4.53) define positive definite operators if A_1, A_2, and A_3 are positive definite, but numerical examples show that they are different in the absence of commutativity assumptions.

There does not appear to be a single "right" generalization of the AGM to three or more positive definite operators.

Acknowledgments

J.E.C. was supported in part by U.S. National Science Foundation grant BSR 84-07461, a Fellowship from the John D. and Catherine T. MacArthur Foundation, and the hospitality of Mr. and Mrs. William T. Golden.

REFERENCES

[1] W.N. Anderson Jr. - R.J. Duffin, *Series and parallel addition of matrices*, J. Math. Anal. Appl. **26** (1969), pp. 576-594.

[2] W.N. Anderson Jr. - G.E. Trapp, *Shorted operators II*, SIAM J. Appl. Math. **28** (1975), pp. 60-71.

[3] Tsuyoshi Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products*, Linear Algebra and its Applications **27** (1979), pp. 203-241.

[4] —, *On the arithmetic-geometric-harmonic mean inequality for positive definite matrices*, Linear Algebra and its Applications **52-53** (1983), pp. 31-37.

[5] TSUYOSHI ANDO - FUMIO KUBO, *Means of positive linear operators*, Math. Ann. **246** (1980), pp. 205-244.

[6] J. ARAZY - T. CLAESSON - S. JANSON - J. PEETRE, *Means and their iterations*, in Proceedings of the Nineteenth Nordic Congress of Mathematicians, Reykjavik, 1984, pp. 191-212 (published by the Icelandic Math. Soc., 1985).

[7] R. ARENS, *The analytic-functional calculus in commutative topological algebras*, Pac. J. Math. **11** (1961), pp. 405-429.

[8] J. BENDAT - S. SHERMAN, *Monotone and convex operator functions*, Trans. Amer. Math. Soc. **79** (1955), pp. 58-71.

[9] G. BORCHARDT, *Gesammelte Werke*, Herausgegeben von G. Hettner, Reimer. Berlin, 1888.

[10] J.M. BORWEIN - P.B. BORWEIN, *The arithmetic-geometric means and fast computation of elementary functions*, SIAM Review **26** (1984), pp. 351-366.

[11] D. BORWEIN - P.B. BORWEIN, *A generalized arithmetic-geometric mean, Problem 83-12*, SIAM Review **25** (1983), pp. 201.

[12] B.C. CARLSON, *Hidden symmetries of special functions*, SIAM Review **12** (1970), pp. 332-345.

[13] —, *Algorithms involving arithmetic and geometric means*, Amer. Math. Monthly **78** (1971), pp. 496-504.

[14] J.E. COHEN - R.D. NUSSBAUM, *Arithmetic-geometric means of positive matrices*, Math. Proc. Cambridge Phil. Soc. **101** (1987), pp. 209-219.

[15] D.A. COX, *The arithmetic-geometric mean of Gauss*, Enseignement Math. **30** (1984), pp. 275-330.

[16] W. DONOGHUE, *Monotone Matrix Functions and Analytic Continuation*, Springer-Verlag, New York, 1974.

[17] N. DUNFORD - J. SCHWARTZ, *Linear Operators*, vol. 1, Interscience, New York, 1958.

[18] C.J. EVERETT - N. METROPOLIS, *A generalization of the Gauss limit for iterated means*, Advances in Math. **7** (1971), pp. 297-300.

[19] JUN-ICHI FUJII, *Arithmetic-geometric means of operators*, Math. Japonica **23** (1979), pp. 667-669.

[20] G.H. HARDY - J.E. LITTLEWOOD - G. POLYA, *Inequalities*, 2nd ed., Cambridge University Press, 1959.

[21] T. KATO, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, New York, 1984.

[22] D.H. LEHMER, *On the compounding of certain means*, J. Math. Anal. Appl. **36** (1971), pp. 183-200.

[23] K. LOWNER - C. LOEWNER, *Uber monotone Matrixfunktionen*, Math. Zeit. **38** (1934), pp. 177-216.

[24] C. LOEWNER *Some classes of functions defined by difference or differential inequalities*, Bull. Amer. Math. Soc. **56** (1950), pp. 308-319.

[25] A.W. MARSHALL - I. OLKIN, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[26] R.D. NUSSBAUM, *Iterated nonlinear maps and Hilbert's projective metric*, Memoirs Amer. Math. Soc., no. **391** (1988).

[27] —, *Iterated nonlinear maps and Hilbert's projective metric: a summary*, in *Dynamics of Infinite Dimensional Systems*, edited by S.N. Chow and J.K. Hale, Springer Verlag, Heidelberg, 1987, pp. 231-249.

[28] —, *Convexity and* log *convexity for the spectral radius*, Linear Alg. and Appl. **73** (1986), pp. 59-122.

[29] I. OLKIN - J.W. PRATT, *A multivariable Tchebycheff inequality*, Annals of Math. Stat. **29** (1958), pp. 226-234.

[30] J. PEETRE, *Iteration of power means and other related means*, preprint, University of Lund, Sweden, 1984.

[31] W. PUSZ - S.L. WORONOWICZ, *Functional calculus for sesquilinear forms, and the purification map*, Rep. Math. Phys. **8** (1975), pp. 159-170.

[32] E.U. STICKEL, *Fast computation of matrix exponential and logarithm*, Analysis (München) **5** (1985), pp. 163-173.

[33] A.E. TAYLOR, *Introduction to Functional Analysis*, Wiley, New York, 1958.

[34] P. WHITTLE, *A multivariate generalization of Tchebychev's inequality*, Quarterly J. of Math. Oxford, Series 2, **9** (1958), pp. 232-240.

[35] J. WIMP, *Computation with Recurrence Relations*, Pittman Advanced Publishing Program, Boston, 1984.

[36] K. YOSHIDA, *Functional Analysis*, 6th ed., Springer-Verlag, New York, 1980.

Mathematics Department
Rutgers University
New Brunswick, New Jersey 08903

Rockefeller University
1230 York Avenue
New York, New York 10021-6399