Matrix Means and Geometric Constructions

Miklós Pálfia

A Thesis submitted for the degree of Doctor of Philosophy

Supervisor: prof. István Vajk

Department of Automation and Applied Informatics
Budapest University of Technology and Economics

August 2011

Talet är tänkandets början och slut.
Med tanken föddes talet.
Utöfver talet når tanken icke.

Numbers are the beginning and end of thinking.
With thoughts were numbers born.
Beyond numbers thought does not reach.

Magnus Gustaf Mittag-Leffler, 1903

Abstract

In this thesis we consider means of several positive definite matrices based on the 2-variable forms of mean functions. In other words the problem is to extend a 2-variable mean to multiple variables. Our basis of study will be the Kubo-Ando theory of 2-variable matrix means. This theory has been formulated in 1980 and since then it has been an open problem to provide an axiomatic theory of multiple variable matrix means. Here we give three different axiomatic extensions for every possible 2-variable matrix mean and thus providing a solution to the extension problem.

Two of these extension methods were considered quite recently by Ando-Li-Mathias and Bini-Meini-Poloni. Both of these methods are based on so called symmetrization procedures that extends a mean function to n+1 variables as a limit using the n-variable forms of the mean function. The applicability of these two procedures were proved only for the geometric mean by these researchers. Proving applicability means showing the convergence of the process to a limit point for all n, where n denotes the number of matrices. These two procedures mentioned above are defined recursively, so we get the n+1-variable mean as a limit point of a process that depends on the n-variable form of the mean function. This potentially leads to computational problems, since it seems to be almost impossible to get an explicit formula for the mean function even for 3 matrices. Therefore we study here a third procedure as well first considered by the author which directly extends from the 2-variable formulas of matrix means, since these mean are explicitly given by the Kubo-Ando theory. Although we end up with a mean that is much easier to compute, we have to pay the price, we loose permutation invariancy of the n-variables. We also prove that the procedure converges for all possible 2-variable matrix means.

This direct process may be considered on complete metric spaces as well so that it provides a mean function in this setting. The 2-variable formulas are understood here as the unique metric midpoints between any two points of the metric space. The aforementioned procedures were also considered before in complete metric spaces of nonpositive curvature. Here we advance further by proving applicability of the procedure for complete metric spaces with certain upper curvature bounds, i.e. we let the curvature to take positive values as well. Since these problems are more natural in this metric geometric setting we will also investigate the problem of finding all possible matrix means that are metric midpoints on certain affinely connected manifolds. During this process we will completely classify all such matrix means and their corresponding manifolds.

In the metric geometric setting one also faces the problem of finding the center of mass of points. The same problem can also be found in the case of the geometric mean of matrices. Certain real life problems end up with the calculation of the center of mass. This leads us to the practical applications of our results, since the new results can be used to approximate the center of mass in metric spaces.

Contents

1 Introduction

The use of certain mean functions dates back to the antiquities. For example the three Pythagorean means, the arithmetic mean

$$A(x_1, \ldots, x_n) = \frac{\sum_{i=1}^n x_i}{n}, \tag{1.1}$$

the geometric mean

$$G(x_1, \ldots, x_n) = \sqrt[n]{\prod_{i=1}^n x_i}, \tag{1.2}$$

and the harmonic

$$H(x_1, \ldots, x_n) = \left(\frac{\sum_{i=1}^n x_i^{-1}}{n} \right)^{-1} \tag{1.3}$$

of positive real numbers have been well known since the ancient Greeks. Several properties of these means have been long known as well, for instance the chain of inequalities $H(x_1, \ldots, x_n) \leq G(x_1, \ldots, x_n) \leq A(x_1, \ldots, x_n)$ between them, the permutation invariance in their variables and that they are monotone functions in their variables.

In the 1970s and 80s researchers in matrix theory started to consider means of positive definite matrices, due to their usage in electric circuits theory [2, 3, 4]. A so called n-pole is the generalization of the resistor, which is a 2-pole, but with n-poles. In this case if we consider the currents and potentials (with respect to a reference point) at each node, by assuming linearity of the system, we have a matrix correspondence between the vector formed by the currents at each node I and the vector of potentials U as $U = RI$, where R is an n-by-n matrix and it is called the resistance matrix of the network. Suppose we choose $n/2$ of the poles as input poles and another $n/2$ as output poles. Then it is possible to consider the series connection of two n-poles and one may ask the question what is the overall resistance matrix of the network. It will be two times the arithmetic mean $\frac{A+B}{2}$ of the two resistance matrices. If we consider parallel connection then the overall resistance matrix will be two times the harmonic mean $2(A^{-1} + B^{-1})^{-1}$ of the two resistance matrices.

The generalization of these two means of positive definite matrices to several variables is straightforward, we just have to use the several variable formulas mentioned above for numbers. However it turns out that even the 2-variable version of the geometric mean of positive matrices is not straightforward. At first glance we have the problem of non commutativity of the matrix multiplication therefore the scalar formula is not permutation invariant. There are also other more serious problems with the classical formula that we will discuss later.

So, all in all, it was the study of electrical networks that derived the interest in means of positive matrices. Several 2-variable functions were considered as candidates of mean functions of two positive matrix. Basic requirements were posed for such functions, for instance monotonicity in their variables and continuity. These basic requirements led to the theory of Kubo and Ando, which

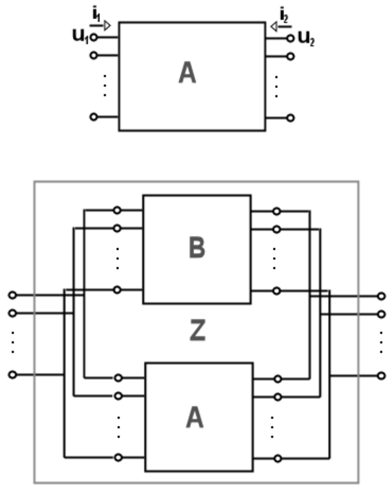

Figure 1: Parallel connection of two n-poles.

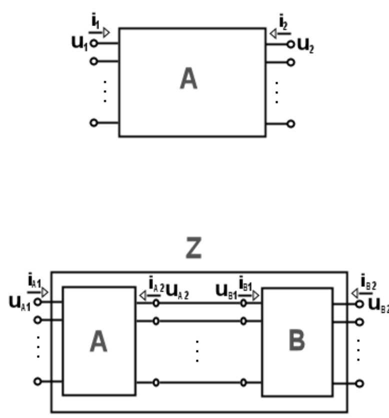

Figure 2: Series connection of two n-poles.

described 2-variable matrix means completely, by providing a characterization of each matrix mean by an operator monotone function. These functions have very strong analytical properties that we will discuss in the following section.

Kubo-Ando theory has been formulated in [36] in 1980, but since then, it was an open problem to provide an axiomatic characterization of all n-variable matrix means based on the 2-variable formulas. We will solve this problem here by providing three different extensions for all 2-variable matrix means in Section 7. One of these extensions will also be considered in a metric geometric setting in Section 5, yielding us an n-variable mean in complete metric spaces with some upper curvature bounds. Also in Section 6 we will find out when is a symmetric matrix mean a midpoint operation on a certain manifold and during this process we will carry out a previously unknown one, real parameter family of affinely connected manifolds that have a midpoint operation which is also a matrix mean. In the next section we build up the theory of operator monotone functions and use it to carry out the theory of 2-variable matrix means of Kubo and Ando in Section 3. After this we will move on to the study of the geometric mean in Section 4. In Section 4 we will meet with the recent diverse ideas of researchers that have been used to extend the 2-variable geometric mean to several variables. After Section 4 we will concentrate on the new results of the author. Section 9 contains the main results that were carried out by the author in the same order that they appear in this thesis. All new theorems and definitions of the author will be indicated by the name of the author and the corresponding publication.

2 Operator Monotone Functions

In this section we will follow the lines of [9]. First of all we define functions of hermitian matrices.

Definition 2.1. Let f be a real function on an interval I. If D is a diagonal matrix $D = diag(\lambda_1, \dots, \lambda_n)$ with entries λ_i belonging to I, then $f(D) = diag(f(\lambda_1), \dots, f(\lambda_n))$. If A is hermitian, then we use the spectral theorem and choose unitary U to have $A = U^*DU$, where D is diagonal, and then define $f(A) = U^*f(D)U$.

We will use the partial order $\geq$ on the set of hermitian matrices defined as $B \geq A$ if and only if $B - A$ is positive semi-definite, that is $\langle x, (B-A)x\rangle \geq 0$ for all vectors x, $\langle \cdot, \cdot \rangle$ denoting the usual hermitian inner product.

Definition 2.2 (Operator Monotone Function)**.** A function f is matrix monotone of order n (or matrix n-monotone) if for all $n \times n$ hermitian matrices $B \geq A$ we have $f(B) \geq f(A)$. If f is monotone for all order n, then it is said to be operator monotone (or matrix monotone).

Similarly to the real case, we have convexity and concavity of functions.

Definition 2.3 (Operator Convexity/Concavity)**.** A function f is matrix convex if and only if for all hermitian matrices A, B and real $0 \leq \lambda \leq 1$ we have

$$f((1-\lambda)A + \lambda B) \leq (1-\lambda)f(A) + \lambda f(B). \tag{2.1}$$

If f is continuous as well, then this condition is equivalent to requiring

$$f\left(\frac{A+B}{2}\right) \leq \frac{f(A)+f(B)}{2}. \tag{2.2}$$

Conversely we say that a function f is operator concave if $-f$ is operator convex, that is, we have reversed inequalities above for f.

It is obvious that the set of operator monotone and the set of operator convex functions are closed under taking convex combinations, and taking pointwise limits of functions. One might also suspect that being operator monotone of order n for a fixed order is less restrictive than being so for all orders. Actually this is true, but in our case we will be focusing on functions which are operator monotone for all orders, since these functions have very strong properties that we will exhibit later in this section.

We will use further notations, $\rho(A)$ will denote the spectral radius of an arbitrary operator A, i.e.

$$\rho(A) = \max\left\{|\lambda| : \lambda \text{ is an eigenvalue of } A\right\}, \tag{2.3}$$

while $\|A\|$ will denote its operator norm, $\|A\| = \sup_{\|x\|=1}\|Ax\|$. It is easy to see that if A is positive, then $A \leq I$ if and only if $\rho(A) \leq 1$. Also an operator will be called a contraction if and only if $\|A\| \leq 1$, equivalently $A^*A \leq I$.

Lemma 2.1. *If $B \geq A$, then for every operator X we have $X^*BX \geq X^*AX$.*

Proof. For arbitrary vector r we have

$$\langle r, X^*BXr\rangle = \langle Xr, BXr\rangle \geq \langle Xr, AXr\rangle = \langle r, X^*AXr\rangle. \tag{2.4}$$

□

The two functions below provide our first easy examples of operator monotone functions.

Proposition 2.2. *The function $f(t) = -1/t$ is operator monotone on $(0, \infty)$, while $g(t) = t^{1/2}$ is operator monotone on $[0, \infty)$.*

Proof. The operator monotonicity of f follows from the order-reversing property of multiplication by -1 and taking inverses.

For g let $B \geq A \geq 0$ and suppose that B is invertible. Then

$$1 \geq \left\|A^{1/2}B^{-1/2}\right\| \geq \rho(A^{1/2}B^{-1/2}) = \rho(B^{-1/4}A^{1/2}B^{-1/4}), \tag{2.5}$$

that is $I \geq B^{-1/4}A^{1/2}B^{-1/4}$, so $B^{1/2} \geq A^{1/2}$. If B is not invertible then $B + \epsilon I$ is for all $\epsilon > 0$. Repeating the above argument and letting $\epsilon \to 0$ we obtain the operator monotonicity of g on $[0, \infty)$ as well. □

2.1 Some Properties

Let K be a contraction. Let $L = (I - KK^*)^{1/2}$ and $L = (I - K^*K)^{1/2}$. Then the operators U, V given as

$$U = \begin{bmatrix} K & L \\ M & -K^* \end{bmatrix}, V = \begin{bmatrix} K & -L \\ M & K^* \end{bmatrix} \tag{2.6}$$

are unitary. Also for $0 \leq \lambda \leq 1$

$$W = \begin{bmatrix} \lambda^{1/2} I & -(1-\lambda)^{1/2} I \\ (1-\lambda)^{1/2} I & \lambda^{1/2} I \end{bmatrix} \tag{2.7}$$

is unitary as well.

Theorem 2.3. *Let I be an interval with $0 \in I$ and f be a real function on I. Then the following are equivalent:*

1. *f is operator convex on I and $f(0) \leq 0$.*
2. *$f(K^*AK) \leq K^*f(A)K$ for all contractions K and hermitian A with eigenvalues in I.*
3. *$f(K_1^*AK_1 + K_2^*BK_2) \leq K_1^*f(A)K_1 + K_2^*f(B)K_2$ for all operators K_1, K_2 such that $K_1^*K_1 + K_2^*K_2 \leq I$ and for all hermitian A, B with eigenvalues in I.*
4. *$f(PAP) \leq P(A)P$ for all projections P and hermitian A with eigenvalues in I.*

Proof. (1) $\Rightarrow$ (2): Let $T = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}$ and let U, V be unitary operators defined in (2.6). Then

$$U^*TU = \begin{bmatrix} K^*AK & K^*AL \\ LAK & LAL \end{bmatrix}, V^*TV = \begin{bmatrix} K^*AK & -K^*AL \\ -LAK & LAL \end{bmatrix}, \tag{2.8}$$

so

$$\begin{bmatrix} K^*AK & 0 \\ 0 & LAL \end{bmatrix} = \frac{U^*TU + V^*TV}{2} \tag{2.9}$$

and

$$\begin{aligned} \begin{bmatrix} f(K^*AK) & 0 \\ 0 & f(LAL) \end{bmatrix} &= f\left(\frac{U^*TU + V^*TV}{2}\right) \leq \\ \leq \frac{f(U^*TU) + f(V^*TV)}{2} &= \frac{U^*f(T)U + V^*f(T)V}{2} = \\ = \frac{1}{2}\left\{U^* \begin{bmatrix} f(A) & 0 \\ 0 & f(0) \end{bmatrix} U + V^* \begin{bmatrix} f(A) & 0 \\ 0 & f(0) \end{bmatrix} V\right\} &\leq \\ \leq \frac{1}{2}\left\{U^* \begin{bmatrix} f(A) & 0 \\ 0 & \end{bmatrix} U + V^* \begin{bmatrix} f(A) & 0 \\ 0 & 0 \end{bmatrix} V\right\} &= \\ = \begin{bmatrix} K^*f(A)K & 0 \\ 0 & Lf(A)L \end{bmatrix}. \end{aligned} \tag{2.10}$$

That is $f(K^*AK) \le K^* f(A)K$.

(2) $\Rightarrow$ (3): Let $T = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}, K = \begin{bmatrix} K_1 & 0 \\ K_2 & 0 \end{bmatrix}$. Then K is a contraction. We have

$$K^*TK = \begin{bmatrix} K_1^*AK_1 + K_2^*BK_2 & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.11}$$

so

$$\begin{aligned} \begin{bmatrix} f(K_1^*AK_1 + K_2^*BK_2) & 0 \\ 0 & f(0) \end{bmatrix} &= f(K^*TK) \le K^*f(T)K = \\ &= \begin{bmatrix} K_1^*f(A)K_1 + K_2^*f(B)K_2 & 0 \\ 0 & f(0) \end{bmatrix}. \end{aligned} \tag{2.12}$$

(3) $\Rightarrow$ (4) is trivial.

(4) $\Rightarrow$ (1): Let A, B be hermitian with eigenvalues in I and $0 \le \lambda \le 1$. Let $T = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}, P = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ and let W be the unitary operator defined by (2.7). Then

$$PW^*TWP = \begin{bmatrix} \lambda A + (1-\lambda)B & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.13}$$

so

$$\begin{aligned} \begin{bmatrix} f(\lambda A + (1-\lambda)B) & 0 \\ 0 & f(0) \end{bmatrix} &= f(PW^*TWP) \le Pf(W^*TW)P = \\ = PW^*f(T)WP &= \begin{bmatrix} \lambda f(A) + (1-\lambda)f(B) & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{2.14}$$

so f is operator convex, and $f(0) \le 0$. □

Theorem 2.4. *Let f be a function mapping $[0, \infty]$ into itself. Then f is operator monotone if and only if it is operator concave.*

Proof. Suppose f is operator monotone. If $f(K^*AK) \ge K^*f(A)K$ for all positive A and contraction K, then from Theorem 2.3 it would follow that f is operator concave. Let $T = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}$ and let U be the unitary operator defined by (2.6). Then $U^*TU = \begin{bmatrix} K^*AK & K^*AL \\ LAK & LAL \end{bmatrix}$. We can find $\lambda > 0$ for any $\epsilon > 0$ such that

$$U^*TU \le \begin{bmatrix} K^*AK + \epsilon I & 0 \\ 0 & \lambda I \end{bmatrix}. \tag{2.15}$$

Replace T by $f(T)$ to get

$$\begin{bmatrix} K^*f(A)K & K^*f(A)L \\ Lf(A)K & Lf(A)L \end{bmatrix} \le \begin{bmatrix} f(K^*AK + \epsilon I) & 0 \\ 0 & f(\lambda)I \end{bmatrix} \tag{2.16}$$

by the operator monotonicity of f. Since ϵ is arbitrary we have $K^*f(A)K \leq f(K^*AK)$.

Conversely, let f be operator concave. Let $0 \leq A \leq B$. Then for any $0 < \lambda < 1$ we have

$$\lambda B = \lambda A + (1-\lambda)\frac{\lambda}{1-\lambda}(B-A). \tag{2.17}$$

Operator concavity of f then yields

$$f(\lambda B) \geq \lambda f(A) + (1-\lambda) f\left(\frac{\lambda}{1-\lambda}(B-A)\right). \tag{2.18}$$

Now $f(X)$ is positive for every positive X, so $f(\lambda B) \geq \lambda f(A)$ that is, by letting $\lambda \to 1$, $f(B) \geq f(A)$. □

Corollary 2.5. *Let f be a continuous function from $(0,\infty)$ to itself. Then if f is operator monotone then $g(t) = 1/f(t)$ is operator convex.*

Corollary 2.6. *Let I be an interval such that $0 \in I$, and let f be a real function on I with $f(0) \leq 0$. Then for every hermitian A with spectrum in I and for all projections P*

$$f(PAP) \leq Pf(PAP) = Pf(PAP)P. \tag{2.19}$$

Corollary 2.7. *Let f be a continuous real fucntion on $[0,\infty)$. Then for all positive operators A and projections P*

$$f\left(A^{1/2}PA^{1/2}\right)A^{1/2}P \leq A^{1/2}Pf(PAP). \tag{2.20}$$

Theorem 2.8. *Let f be a real function on the interval $[0,\alpha)$. Then the following are equivalent:*

1. *f is operator convex and $f(0) \leq 0$.*
2. *$g(t) = f(t)/t$ is operator monotone on $(0,\alpha)$.*

Proof. (1) $\Rightarrow$ (2): Let $0 < A \leq B$ be matrices. Then $0 < A^{1/2} \leq B^{1/2}$, so $B^{-1/2}A^{1/2}$ is a contraction by using the operator monotonicity of the square root function, so using Theorem 2.3

$$f(A) = f(A^{1/2}B^{-1/2}BB^{-1/2}A^{1/2}) \leq A^{1/2}B^{-1/2}f(B)B^{-1/2}A^{1/2}, \tag{2.21}$$

which implies that

$$A^{-1/2}f(A)A^{-1/2} \leq B^{-1/2}f(B)B^{-1/2}. \tag{2.22}$$

This is equivalent to $A^{-1}f(A) \leq B^{-1}f(B)$, in other words, g is operator monotone.

(2) $\Rightarrow$ (1): Since g is operator monotone on $(0,\alpha)$, we have $f(0) \leq 0$. We will show that f satisfies condition (4) of Theorem 2.3. Let P be an arbitrary

projection and let A be positive with eigenvalues in $(0, \alpha)$. Then there exists an $\epsilon > 0$ such that $(1+\epsilon)A$ has all its eigenvalues in $(0, \alpha)$ as well. Now $(1+\epsilon)P \leq (1+\epsilon)I$, so $A^{1/2}(P+\epsilon I)A^{1/2} \leq (1+\epsilon)A$. So considering the operator monotonicity of g we get

$$\begin{aligned} A^{-1/2}(P+\epsilon I)^{-1}A^{-1/2} f\left(A^{1/2}(P+\epsilon I)A^{1/2}\right) &\leq (1+\epsilon)^{-1}A^{-1} f\left((1+\epsilon)A\right) \\ A^{-1/2} f\left(A^{1/2}(P+\epsilon I)A^{1/2}\right) A^{1/2}(P+\epsilon I) &\leq \\ &\leq (1+\epsilon)^{-1}(P+\epsilon I) f\left((1+\epsilon)A\right)(P+\epsilon I). \end{aligned} \tag{2.23}$$

Letting $\epsilon \to 0$, this gives

$$A^{-1/2} f\left(A^{1/2}PA^{1/2}\right) A^{1/2}P \leq Pf(A)P. \tag{2.24}$$

By the previous two corollaries, we get

$$f(PAP) \leq Pf(A)P. \tag{2.25}$$

□

To advance further, we have to introduce some further notations related to derivatives of certain functions.

Definition 2.4 (Divided Differences)**.** Let f be a continuously differentiable function. Then the function $f^{[1]}$ is defined as

$$\begin{aligned} f^{[1]}(\lambda, \mu) &= \frac{f(\lambda) - f(\mu)}{\lambda - \mu}, \text{ if } \lambda \neq \mu, \\ f^{[1]}(\lambda, \mu) &= f'(\lambda), \text{ if } \lambda = \mu. \end{aligned} \tag{2.26}$$

The function $f^{[1]}(\lambda, \mu)$ is called the first divided differences of f at (λ, μ). If Γ is a diagonal matrix with diagonal entries λ_i, then we denote by $f^{[1]}(\Gamma)$ the matrix whose (i, j) entry is $f^{[1]}(\lambda_i, \lambda_j)$ and if $A = U^*DU$ is hermitian with unitary U and diagonal D, then $f^{[1]}(A) = U^* f^{[1]}(D)U$.

Similarly we define second divided differences $f^{[2]}$ for a twice continuously differentiable function f as

$$f^{[2]}(\lambda_1, \lambda_2, \lambda_3) = \frac{f^{[1]}(\lambda_1, \lambda_2) - f^{[1]}(\lambda_1, \lambda_3)}{\lambda_2 - \lambda_3} \tag{2.27}$$

for distinct $\lambda_1, \lambda_2, \lambda_3$, otherwise we define

$$f^{[2]}(\lambda, \lambda, \lambda) = \frac{1}{2} f''(\lambda) \tag{2.28}$$

by using continuity.

We will consider the derivative of functions considered over the space of hermitian matrices. That is

Definition 2.5. We call a function f Fréchet-differentiable at A if there exists a linear operator $Df[A]$ on the space of hermitian matrices such that for all H

$$\|f(A+H)-f(A)-Df[A][H]\|\, o(\|H\|). \tag{2.29}$$

Then the linear operator $Df[A]$ is called the Fréchet-differential or derivative of f at A. It follows that if f has a derivative at A, then

$$Df[A][H] = \left.\frac{d}{dt}\right|_{t=0} f(A+tH). \tag{2.30}$$

Now we will exhibit the connection between the derivative $Df[A]$ and the matrix $f^{[1]}(A)$.

Lemma 2.9. *Let f be a polynomial. Then for all diagonal Γ and hermitian matrix H, we have*

$$Df[\Gamma][H] = f^{[1]}(\Gamma)\circ H, \tag{2.31}$$

where $\circ$ denotes the Schur-product.

Proof. Both sides of (2.31) is linear in f, so it is enough to prove it for powers. So let $f(t)=t^n$. Then

$$Df[\Gamma][H] = \sum_{k=1}^{n} \Gamma^{k-1}H\Gamma^{n-k}. \tag{2.32}$$

This is a matrix with (i,j) entries equal to $\sum_{k=1}^{n}\Gamma_{ii}^{k-1}\Gamma_{jj}^{n-k}H_{ij}$. We also have that the (i,j) entry of $f^{[1]}(\Gamma)$ is $\sum_{k=1}^{n}\Gamma_{ii}^{k-1}\Gamma_{jj}^{n-k}$. □

Corollary 2.10. *Let f be a polynomial. Then if $A=U\Gamma U^*$*

$$Df[A][H] = U\left[f^{[1]}(\Gamma)\circ U^*HU\right]U^*. \tag{2.33}$$

Proof. Since

$$\left.\frac{d}{dt}\right|_{t=0} f(U\Gamma U^*+tH) = U\left[\left.\frac{d}{dt}\right|_{t=0} f(\Gamma+tU^*HU)\right]U^*, \tag{2.34}$$

and the assertion follows from Lemma 2.9. □

Theorem 2.11. *Let $f\in C^1(I)$ and A a hermitian matrix with eigenvalues in I. Then*

$$Df[A][H] = f^{[1]}(A)\circ H, \tag{2.35}$$

where $\circ$ denotes the Schur-product in a basis where A is diagonal.

Proof. Let $A = U\Gamma U^*$, where Γ is diagonal. We claim that

$$Df[A][H] = U\left[f^{[1]}(\Gamma) \circ U^*HU\right]U^*. \tag{2.36}$$

We have already proved this for all polynomials. Now we prove it for all $f \in C^1$.

Let us denote the right hand side of (2.36) by $df[A][H]$. By definition $df[A]$ is a linear map on hermitian matrices. Also all entries of the matrix $f^{[1]}(\Gamma)$ are bounded by $\max_{|t|\leq\|A\|}$ by the mean value theorem. Hence

$$\|df[A][H]\| \leq \max_{|t|\leq\|A\|} \|H\|. \tag{2.37}$$

Let H be a hermitian matrix with such norm that the eigenvalues of $A+H$ are in I. Choose a closed interval $[a,b]$ in I such that the eigenvalues of A and $A+H$ are contained in it. Choose a sequence of polynomials such that $f_n \to f$ and $f_n' \to f'$ uniformly on $[a,b]$. Let L be the line segment connecting A and $A+H$ in the space of hermitian matrices. Now the mean value theorem for Fréchet derivatives yields

$$\begin{aligned}\big\|f_m(A+H) - f_n(A+H) - f_(A) + f_n(A)\big\| &\leq \\ \leq \|H\| \sup_{X\in L} \|Df_m(X) - Df_n(X)\| &= \\ = \|H\| \sup_{X\in L} \|df_m(X) - df_n(X)\|&,\end{aligned} \tag{2.38}$$

since $Df_n = df_n$ holds.

Let ϵ be any positive real number. Then by (2.37) there exists a positive integer N_0 such that for all $m, n \geq N_0$

$$\sup_{X\in L} \|df_m(X) - df_n(X)\| \leq \frac{\epsilon}{3} \tag{2.39}$$

and also

$$\sup_{X\in L} \|df_n(A) - df(A)\| \leq \frac{\epsilon}{3} \tag{2.40}$$

hold. Let $m \to \infty$ and use (2.38) and (2.39) to obtain

$$\|f(A+H) - f(A) - (f_n(A+H) - f_n(A))\| \leq \frac{\epsilon}{3}\|H\|. \tag{2.41}$$

If $\|H\|$ is sufficiently small, then by the definition of the Fréchet derivative

$$\|f_n(A+H) - f_n(A) - df_n[A][H]\| \leq \frac{\epsilon}{3}\|H\|, \tag{2.42}$$

so we have, using the triangle inequality

$$\begin{aligned}\|f_n(A+H)\ &-f_n(A) - df[A][H]\| \leq \\ &\leq \|f(A+H) - f(A) - (f_n(A+H) - f_n(A))\| + \\ &\quad + \|f_n(A+H) - f_n(A) - df_n[A][H]\| + \\ &\quad + \|(df[A] - df_n[A])[H]\|,\end{aligned} \tag{2.43}$$

and use the above estimations to conclude that

$$\|f(A+H) - f(A) - df[A][H]\| \leq \epsilon \|H\|, \tag{2.44}$$

which is $Df[A] = df[A]$ for sufficiently small $\|H\|$. □

Theorem 2.12. *Let $f \in C^1(I)$. Then f is operator monotone on I if and only if, for every hermitian matrix A with eigenvalues in I, $f^{[1]}(A)$ is positive.*

Proof. Let f be operator monotone, and let A be hermitian with eigenvalues in I. Let H be the matrix with 1 entries. H is positive and $A + tH \geq 0$ if $t \geq 0$, hence $f(A+tH) - f(A)$ is positive for small t, so $Df[A][H] \geq 0$. So $f^{[1]}(A) \circ H \geq 0$ by Theorem 2.11, in other words $f^{[1]}(A) \geq 0$.

For the converse implication, let $A \geq B$ be hermitian with eigenvalues in I. Let $X(t) = (1-t)A + tB$, for $0 \leq t \leq 1$, so $X(t)$ has eigenvalues in I as well. So by assumption $f^{[1]}(X(t)) \geq 0$ for all t. Since $X'(t) = B - A \geq 0$ and the Schur-product of two positive matrices is positive, $f^{[1]}(X(t)) \circ X'(t)$ is also positive. By the previous theorem $f^{[1]}(X(t)) \circ X'(t) = Df[X(t)][X'(t)]$, so

$$f(B) - f(A) = f(X(1)) - f(X(0)) = \int_0^1 f^{[1]}(X(t)) \circ X'(t) dt \geq 0. \tag{2.45}$$

□

Lemma 2.13. *If f is continuous and operator monotone of $(-1,1)$, then for each $-1 \leq \lambda \leq 1$, the function $g_\lambda(t) = (t+\lambda)f(t)$ is operator convex.*

Proof. We will use Theorem 2.8 to prove this. Assume that f is operator monotone and continuous on $[-1,1]$. Then the function $f(t-1)$ is operator monotone on $[0,2)$. Let $g(t) = tf(t-1)$, so $g(0) = 0$ and $g(t)/t$ is operator monotone on $(0,2)$. So by Theorem 2.8 $g(t)$ is operator convex on $[0,2)$, which in turn implies that the function $h_1(t) = g(t+1) = (t+1)f(t)$ is operator convex on $[-1,1)$. If we apply the same argument for $-f(-t)$ which happens to be operator monotne as well on $[-1,1]$, we see that the function $h_2(t) = -(t+1)f(-t)$ is operator convex as well on $[-1,1)$. So changing signes of t preserves convexity, therefore the function $h_3(t) = h_2(-t)$ is also operator convex. Hence for $|\lambda| \leq 1$, $g_\lambda(t) = \frac{1+\lambda}{2}h_1(t) + \frac{1-\lambda}{2}h_2(t)$ is also operator convex, since its a convex combination of operator convex functions.

For operator monotone and continuous f on $(-1,1)$, the function $f((1-\epsilon)t)$ is continuous and operator monotone on $[-1,1]$ for all $\epsilon > 0$. So by the argument above $(t+\lambda)f((1-\epsilon)t)$ is operator convex. So by letting $\epsilon \to 0$ we get that $(t+\lambda)f(t)$ is operator convex. □

The next theorem shows that every operator monotone function is necessarily continuously differentiable on its domain. This is the first step toward exhibiting the strong smoothness properties of such functions. In order to be able to prove this assertion we have to introduce a new tool. This is essentially a smoothing technique, the so called regularization of a function using mollifiers and convolution.

Definition 2.6 (Mollifier)**.** Let ϕ be a real function of C^∞ class with the following properties: $\phi \geq 0$, ϕ is even, the support of ϕ is $[-1,1]$ and $\int \phi = 1$. For each $\epsilon > 0$ let $\phi_\epsilon(x) = \frac{1}{\epsilon}\phi\left(\frac{x}{\epsilon}\right)$. Then the support of ϕ is $[-\epsilon, \epsilon]$ and ϕ_ϵ has all the other properties listed above. The functions ϕ_ϵ are called mollifiers.

Definition 2.7 (Regularization)**.** If f is locally integrable function, then

$$f_\epsilon(x) = (f * \phi_\epsilon)(x) = \int f(x-y)\phi_\epsilon(y)dy \tag{2.46}$$

is defined to be its regularization.

The following nice properties are fulfilled by the family f_ϵ:

1. Every f_ϵ is a C^∞ function.
2. If the support of f is contained in a compact set, then the support of f_ϵ is contained in an ϵ-neighborhood of the same compact set.
3. If f is continuous at x_0 then $f(x_0) = \lim_{\epsilon \downarrow x_0} f_\epsilon(x_0)$.
4. If f has a first order singularity at x_0, then $\lim_{\epsilon \downarrow x_0} f_\epsilon(x_0) = \frac{f(x_0+)+f(x_0-)}{2}$.
5. If f is continuous at x, then $f_\epsilon(x)$ converges to $f(x)$ uniformly on every compact set, as $\epsilon \to 0$.
6. If f is differentiable, then $(f_\epsilon)' = (f')_\epsilon$.
7. If f is monotone, then $f_\epsilon'(x) \to f'(x)$ as $\epsilon \to 0$, if $f'(x)$ exists.

Theorem 2.14. *Every operator monotone function f on I is in the class C^1.*

Proof. Let f_ϵ be a regularization of f of order ϵ for $0 < \epsilon < 1$. Then f_ϵ is in the class C^∞ on $(-1+\epsilon, 1-\epsilon)$. It is also clearly operator monotone. Let $\bar{f}(t) = \lim_{\epsilon\to 0} f_\epsilon(t)$. Then $\bar{f}(t) = \frac{f(t+)+f(t-)}{2}$.

Now let $g_\epsilon(t) = (t+1)f_\epsilon(t)$. Then by Lemma 2.13, g_ϵ is operator convex. Let $\bar{g}(t) = \lim_{\epsilon\to 0} g_\epsilon(t)$, then also $\bar{g}(t)$ is operator convex. Since every convex function is continuous, therefore $\bar{g}(t)$ is continuous as well. This in turn implies that $\bar{f}(t)$ is continuous, which tells us that $\bar{f}(t) = f(t)$, hence $f(t)$ is continuous.

Let $g(t) = (t+1)f(t)$. Then g is a convex function on I, so it is left and right differentiable and the one-sided derivatives satisfy the properties

$$g_-'(t) \leq g_+'(t), \lim_{s\downarrow t} g_\pm'(s) = g_+'(t), \lim_{s\uparrow t} g_\pm'(s) = g_-'(t). \tag{2.47}$$

But $g_\pm'(t) = f(t) + (t+1)f_\pm'(t)$, and since $t+1 > 0$ the derivatives $f_\pm'(t)$ also satisfy the above relations.

Let $A = \begin{bmatrix} s & 0 \\ 0 & t \end{bmatrix}$, $s, t \in (-1, 1)$. If ϵ is small enough, then $s, t \in (-1 + \epsilon, 1-\epsilon)$. Since f_ϵ is operator monotone on this interval, the matrix $f_\epsilon^{[1]}(A)$ is positive by Theorem 2.12, which implies that

$$\left(\frac{f_\epsilon(s) - f_\epsilon(t)}{s-t}\right)^2 \leq f_\epsilon'(s)f_\epsilon'(t). \tag{2.48}$$

Since $f_\epsilon \to f$ uniformly on compact sets, $f_\epsilon(s) - f_\epsilon(t) \to f(s) - f(t)$. Also $f'_\epsilon(s) \to \frac{f'_+(t)+f'_-(t)}{2}$, so the above inequality gives, taking the limit $\epsilon \to 0$, that

$$\left(\frac{f(s)-f(t)}{s-t}\right)^2 \leq \frac{1}{4}\left[f'_+(s)+f'_-(s)\right]\left[f'_+(t)+f'_-(t)\right]. \tag{2.49}$$

Now as we let $s \downarrow t$, and considering the fact that the derivatives of f satisfy similar relations as (2.47), we get

$$\left[f'_+(t)\right]^2 \leq \frac{1}{4}\left[f'_+(t)+f'_-(t)\right]\left[f'_+(t)+f'_-(t)\right], \tag{2.50}$$

which implies that $f'_+(t) = f'_-(t)$, so f is differentiable, and also f' satisfies relations like (2.47), so it is continuous as well. □

We move on to study properties of operator convex functions, which could be done via the study of their second divided differences mentioned earlier in the section. We state the following three propositions without proofs. Their proofs involve some straightforward calculation or similar techniques discussed earlier in the preceding assertions.

Proposition 2.15. *If $\lambda_1, \lambda_2, \lambda_3$ are distinct, then $f^{[2]}(\lambda_1,\lambda_2,\lambda_3)$ is the quotient of the two determinants*

$$\begin{vmatrix} f(\lambda_1) & f(\lambda_2) & f(\lambda_3) \\ \lambda_1 & \lambda_2 & \lambda_3 \\ 1 & 1 & 1 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ 1 & 1 & 1 \end{vmatrix}, \tag{2.51}$$

so the function $f^{[2]}$ is permutation invariant in its variables.

Proposition 2.16. *If $f(t) = t^n$ for $n = 2, 3, \ldots$ we have that*

$$f^{[2]}(\lambda_1,\lambda_2,\lambda_3) = \sum_{\substack{0 \leq p,q,r \\ p+q+r=n-2}} \lambda_1^p \lambda_2^q \lambda_3^r. \tag{2.52}$$

Proposition 2.17. *Let $f(t) = t^n$, for $n \geq 2$ integer. Suppose that A is a diagonal matrix with eigenvalues λ_i and P_i denote the projections onto the coordinate axes. Then for every hermitian H*

$$\begin{aligned} \left.\frac{d^2}{dt^2}\right|_{t=0} f(A+tH) &= 2 \sum_{p+q+r=n-2} A^p H A^q H A^r = \\ &= 2 \sum_{i,j,k} f^{[2]}(\lambda_i,\lambda_j,\lambda_k) P_i H P_j H P_k, \end{aligned} \tag{2.53}$$

which also holds for all C^2 function f.

Theorem 2.18. *If $f \in C^2(I)$ and f is operator convex, then for each $\mu \in I$ the function $g(t) = f^{[1]}(\mu, t)$ is operator monotone.*

Proof. Since $f \in C^2$, $g \in C^1$, therefore by Theorem 2.12, it is enough to show that the matrix with (i, j) entries $f^{[1]}(\lambda_i, \lambda_j)$ is positive for all $\lambda_i \in I$.

Choose any $\lambda_1, \ldots, \lambda_{n+1} \in I$. Let A be diagonal with entries $\lambda_1, \ldots, \lambda_{n+1}$. Since f is operator convex and it is in the C^2 class, for every hermitian H, $\frac{d^2}{dt^2}\Big|_{t=0} f(A + tH)$ must be positive. Let P_i denote the projections onto the coordinate axes, so we have an explicit expression for this in (2.53). Let H be of the form

$$\begin{bmatrix} 0 & 0 & \cdots & & \bar{z}_1 \\ 0 & 0 & \cdots & & \bar{z}_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ z_1 & z_2 & \cdots & z_n & 0 \end{bmatrix}, \tag{2.54}$$

where z_i are arbitrary complex numbers. Let x denote the $(n + 1)$-vector $(1, \ldots, 1, 0)$. Then we have

$$\langle x, P_i H P_j H P_k x\rangle = z_k \bar{z}_i \delta_{j,n+1} \tag{2.55}$$

for $1 \le i, j, k \le n + 1$ and $\delta_{i,j}$ is the Kronecker-symbol. So then we have by the positivity of the matrix $\frac{d^2}{dt^2}\Big|_{t=0} f(A + tH)$ and the above that

$$\begin{aligned} 0 \le & \sum_{1\le i,j,k\le n+1} f^{[2]}(\lambda_i, \lambda_j, \lambda_k) \langle x, P_i H P_j H P_k x\rangle = \\ = & \sum_{1\le i,k\le n+1} f^{[2]}(\lambda_i, \lambda_{n+1}, \lambda_k) z_k \bar{z}_i. \end{aligned} \tag{2.56}$$

We also have that

$$\begin{aligned} f^{[2]}(\lambda_i, \lambda_{n+1}, \lambda_k) &= \frac{f^{[1]}(\lambda_{n+1}, \lambda_i) - f^{[1]}(\lambda_{n+1}, \lambda_k)}{\lambda_i - \lambda_k} = \\ &= g^{[1]}(\lambda_i, \lambda_k). \end{aligned} \tag{2.57}$$

So we get that

$$0 \le \sum_{1\le i,k\le n+1} g^{[1]}(\lambda_i, \lambda_k) z_k \bar{z}_i. \tag{2.58}$$

Since z_i is arbitrary, this is equivalent to the positivity of the matrix with (i, j) entries $g^{[1]}(\lambda_i, \lambda_j)$. □

Corollary 2.19. *If $f \in C^2(I)$, $f(0) = 0$ and f is operator convex, then the function $g(t) = \frac{f(t)}{t}$ is operator monotone.*

Proof. By the above theorem $f^{[1]}(0, t)$ is operator monotone, which is just $f(t)/t$ in this case. □

Corollary 2.20. *If f is operator monotone on I and $f(0) = 0$, then the function $g(t) = \frac{t+\lambda}{t} f(t)$ is operator monotone for all $|\lambda| \le 1$.*

Proof. Let us assume that $f \in C^2$. By Lemma 2.13 the function $g_\lambda(t) = (t + \lambda)f(t)$ is operator convex. By the previous corollary $g(t)$ is operator monotone. For the case if f is not in the class of C^2, we consider its regularization f_ϵ, and apply the same argument to $f_\epsilon(t) - f_\epsilon(0)$, and then let $\epsilon \to 0$. □

Corollary 2.21. *If f is operator monotone on I and $f(0) = 0$, then f is twice differentiable at* 0*.*

Proof. By the previous corollary, the function $g(t) = \left(1 + \frac{1}{t}\right) f(t)$ is operator monotone, and by Theorem 2.14 it is continuously differentiable. Therefore the function $h(t) = \frac{1}{t} f(t)$, $h(0) := f'(0)$ is continuously differentiable, which yields that f is twice differentiable at 0. □

2.2 Loewner's Characterization

Consider all functions f on the interval $I = (-1, 1)$ that are operator monotone and satisfy the conditions

$$f(0) = 0, f'(0) = 1. \tag{2.59}$$

Let K be the collection of all such functions. Clearly, K is a convex set. We will show that this set is compact in the topology of pointwise convergence and will find its extreme points. This will enable us to write an integral representation for functions in K.

Lemma 2.22. *If $f \in K$, then*

$$\begin{aligned} f(f) &\leq \frac{t}{1-t} \text{ for } 0 \leq t < 1, \\ f(f) &\geq \frac{t}{1+t} \text{ for } -1 < t < 0, \\ |f''(f)| &\leq 2. \end{aligned} \tag{2.60}$$

Proof. Let $A = \begin{bmatrix} t & 0 \\ 0 & 0 \end{bmatrix}$. Then by Theorem 2.12, the matrix

$$f^{[1]}(A) = \begin{bmatrix} f'(t) & f(t)/t \\ f(t)/t & 1 \end{bmatrix} \tag{2.61}$$

is positive. Hence

$$\frac{f(t)^2}{t^2} \leq f'(t). \tag{2.62}$$

Let $g_\pm(t) = (t \pm 1)f(t)$. By Lemma 2.13, both functions $g_\pm(t)$ are convex, hence their derivatives are monotonically increasing functions. Since $g'_\pm(t) = f(t) + (t \pm 1)f'(t)$ and $g'_\pm(0) = \pm 1$, this implies that

$$\begin{aligned} f(t) + (t-1)f'(t) &\geq -1 \text{ for } t > 0 \\ f(t) + (t+1)f'(t) &\leq 1 \text{ for } t < 0. \end{aligned} \tag{2.63}$$

Thus we obtain

$$f(t)+1 \geq \frac{(1-t)f(t)^2}{t^2} \text{ for } t>0. \tag{2.64}$$

Now suppose that for some $0<t<1$ we have $f(t)>\frac{t}{1-t}$. Then $f(t)^2>\frac{t}{1-t}f(t)$, so from the above we get $f(t)+1>\frac{f(t)}{t}$. But this gives the inequality $f(t)<\frac{t}{1-t}$, which contradicts our assumption. This shows that $f(t)\leq \frac{t}{1-t}$ for $0\leq t<1$. The second inequality of the lemma is obtained by the same argument using the other inequality.

We have already seen in the proof of Corollary 2.21 that

$$f'(0)+\frac{1}{2}f''(0)=\lim_{t\to 0}\frac{(1+t^{-1})f(t)-f'(0)}{t}. \tag{2.65}$$

Let $t\downarrow 0$ and use the first inequality of the lemma to conclude that this limit is smaller than 2. Let $t\uparrow 0$ and use the second inequality to conclude that it is bigger than 0. Together these two imply that $|f''(0)|\leq 2$. □

Proposition 2.23. *The set K is compact in the topology of pointwise convergence.*

Proof. Let f_i be a net in K. By the above lemma the set $f_i(t)$ is bounded for each t. So, by Thychonoff's Theorem, there exists a subnet f_{i_k} that converges pointwise to a bounded function f. The limit function f is operator monotone and $f(0)=0$. We show that $f'(0)=1$ so that $f\in K$, and hence K is compact.

By Corollary 2.20 each of the functions $\left(1+\frac{1}{t}\right)f_i(t)$ is monotone on $(-1,1)$. Since for all i, $\lim_{t\to 0}\left(1+\frac{1}{t}\right)f_i(t)=f_i'(0)=1$, we see that $\left(1+\frac{1}{t}\right)f_i(t)\geq 1$ if $t\geq 0$ and is ≤ 1 if $t\leq 0$. Hence if $t>0$ we have $\left(1+\frac{1}{t}\right)f(t)\geq 1$, and if $t<0$ we have the opposite inequality. Since f is continuously differentiable, this shows that $f'(0)=1$. □

Proposition 2.24. *All extreme points of K have the form*

$$f(t)=\frac{t}{1-\alpha t}, \textit{ where } \alpha=\frac{1}{2}f''(0). \tag{2.66}$$

Proof. Let $f\in K$. For each $-1<\lambda<1$ let

$$g_\lambda(t)=\left(1+\frac{\lambda}{t}\right)f(t)-\lambda. \tag{2.67}$$

By Corollary 2.20, g_λ is operator monotone. Note that $g_\lambda(0)=0$, since $f(0)=0$ and $f'(0)=1$. Also, $g_\lambda'(0)=1+\frac{1}{2}\lambda f''(0)$, so the function h_λ defined as

$$h_\lambda(t)=\frac{1}{1+\frac{1}{2}\lambda f''(0)}\left[\left(1+\frac{\lambda}{t}\right)f(t)-\lambda\right] \tag{2.68}$$

is in K. Since $|f''(0)|\leq 2$, we see that $|\frac{1}{2}\lambda f''(0)|<1$. We can write

$$f=\frac{1}{2}\left[1+\frac{1}{2}\lambda f''(0)\right]h_\lambda+\frac{1}{2}\left[1-\frac{1}{2}\lambda f''(0)\right]h_{-\lambda}. \tag{2.69}$$

So, if f is an extreme point of K, we must have $f = h_\lambda$. This says that

$$\left[1 + \frac{1}{2}\lambda f''(0)\right] f(t) = \left(1 + \frac{\lambda}{t}\right) f(t) - \lambda, \tag{2.70}$$

from which we have that

$$f(t) = \frac{t}{1 - \frac{1}{2} f''(0) t}. \tag{2.71}$$

□

Theorem 2.25. *For each f in K there exists a unique probability measure μ on $[-1,1]$ such that*

$$f(t) = \int_{-1}^{1} \frac{t}{1 - \lambda t} d\mu(\lambda). \tag{2.72}$$

Proof. For $-1 \le \lambda \le 1$, consider the functions $h_\lambda(t) = \frac{t}{1-\lambda t}$. By Proposition 2.24, the extreme points of K are included in the family h_λ. Since K is compact and convex, it is the closed convex hull of its extreme points by the Krein-Milman Theorem. Finite convex combinations of elements of the family $\{h_\lambda : -1 \le \lambda \le 1\}$ can also be written as $\int h_\lambda d\nu(\lambda)$, where ν is a probability measure on $[-1,1]$ with finite support. Since f is in the closure of these combinations, there exists a net ν_i of finitely supported probability measure on $[-1,1]$ such that the net $f_i(t) = \int h_\lambda d\nu_i(\lambda)$ converges to $f(t)$. Since the space of the probability measure is weak* compact, the net ν_i has an accumulation point μ. In other words, a subnet of $\int h_\lambda d\nu_i(\lambda)$ converges to $\int h_\lambda d\mu(\lambda)$, so $f(t) = \int h_\lambda d\mu(\lambda) = \int \frac{t}{1-\lambda t} d\mu(\lambda)$.

Now suppose that there are two measure μ_1 and μ_2 for which the representation (2.72) is valid. Expand the integrand as a power series $\frac{t}{1-\lambda t} = \sum_{n=0}^{\infty} t^{n+1}\lambda^n$ convergent uniformly in $|\lambda| < 1$ for every fixed t with $|t| < 1$. This shows that

$$\sum_{n=0}^{\infty} t^{n+1} \int_{-1}^{1} \lambda^n d\mu_1(\lambda) = \sum_{n=0}^{\infty} t^{n+1} \int_{-1}^{1} \lambda^n d\mu_2(\lambda) \tag{2.73}$$

for all $|t| < 1$. The identity theorem for power series shows that

$$\int_{-1}^{1} \lambda^n d\mu_1(\lambda) = \int_{-1}^{1} \lambda^n d\mu_2(\lambda) \tag{2.74}$$

for all $n = 0, 1, 2, \ldots$, which is only possible when $\mu_1 = \mu_2$. □

We assumed that the normalizations (2.59) hold for K in order to make the set K compact. At this point we may remove these conditions to get the following

Corollary 2.26. *Let f be a nonconstant operator monotone function on $(-1,1)$. Then there exists a unique probability measure μ on $[-1,1]$ such that*

$$f(t) = f(0) + f'(0) \int_{-1}^{1} \frac{t}{1 - \lambda t} d\mu(\lambda). \tag{2.75}$$

Proof. We have that f is monotone and nonconstant, so $f'(0) \neq 0$. So the function $\frac{f(t)-f(0)}{f'(0)}$ is in K. □

The above corollary can be extended to any operator monotone function over an arbitrary interval (a, b), since f is operator monotone on (a, b) if and only if $f\left(\frac{b-a}{2}t + \frac{a+b}{2}\right)$ is operator monotone on $(-1, 1)$.

Using Corollary 2.26 we may also analytically extend an operator monotone f on $(-1, 1)$ by replacing t with complex z. In this way we may define f on the whole complex plane excluding $(-\infty, -1] \cup [1, \infty)$. Since

$$\Im \frac{z}{1-\lambda z} = \frac{\Im z}{|1-\lambda z|^2}, \tag{2.76}$$

so f maps the upper half-plane into itself and maps the lower half-plane into itself as well. Similarly $f(z) = \overline{f(\overline{z})}$, so it is invariant under reflections over the real line. The converse is also true, an analytic function that maps the upper half-plane into itself and is analytically continued to the lower half-plane via reflection across the real line, then it is operator monotone.

We will omit the further study of such functions in detail from the point of view of complex analysis, since the characterization obtained so far is sufficient for our purposes. Actually such analytically continued functions have a very rich theory, one may consult the class of Pick functions and their characterization due to a theorem of Nevanlinna [9].

Furthermore consider the following nice

Example 2.1. By contour integration using the Residuum Theorem we have that

$$\int_0^\infty \frac{\lambda^{r-1}}{1+\lambda} = \pi \csc r\pi,\ 0 < r < 1. \tag{2.77}$$

By change of variables we obtain from this that

$$t^r = \frac{\sin r\pi}{\pi} \int_0^\infty \frac{t}{t+\lambda} \lambda^{r-1} d\lambda \tag{2.78}$$

for all $t > 0$ and $0 < r < 1$. That is, t^r is operator monotone for all $r \in [0, 1]$.

Actually it turns out that for other values of r, this function is not operator monotone.

3 Matrix Means and Operator Monotone Functions

In this section we present the theory of Kubo and Ando, which characterizes matrix means by operator monotone functions. We denote by $P(n, \mathbb{C})$ the open convex cone of $n \times n$ positive definite matrices and by $H(n, \mathbb{C})$ the space of $n \times n$ hermitian matrices over the complex field $\mathbb{C}$.

Definition 3.1 (Matrix Mean)**.** A two-variable function $M\colon P(n,\mathbb{C})\times P(n,\mathbb{C}) \mapsto P(n,\mathbb{C})$ is called a matrix mean if

(i) $M(I,I) = I$ where I denotes the identity,

(ii) if $A \leq A'$ and $B \leq B'$, then $M(A,B) \leq M(A',B')$,

(iii) $CM(A,B)C \leq M(CAC,CBC)$,

(iv) if $A_n \downarrow A$ and $B_n \downarrow B$ then $M(A_n,B_n) \downarrow M(A,B)$.

The above definition were considered by Kubo and Ando in [36]. Actually they considered the above definition without the normalization property (i), and called such functions an operator connection with notation $A\sigma B$. For the case of matrix means they included property (i) as well. An immediate consequence of property (iii) is that for all invertible C we have

$$CM(A,B)C = M(CAC,CBC). \tag{3.1}$$

Yet another consequence of the properties is that if $A \leq B$ then

$$A = M(A,A) \leq M(A,B) \leq M(B,B) = B. \tag{3.2}$$

The importance of operator connections comes from electric circuit theory as it was mentioned in the first section. A remarkable property of operator connections is that they can be characterized by operator monotone functions.

Theorem 3.1 (Kubo-Ando [36])**.** *For each connection σ and $x > 0$ real number, the operator $1\sigma x$ is a scalar. Furthermore the map, $\sigma \mapsto f$, defined by*

$$f(x) = 1\sigma x \tag{3.3}$$

for $x > 0$, is an affine order-isomorphism from the class of operator connections onto the class of operator monotone functions.

Proof. Let σ be a connection. Suppose that P is a projection that commutes with positive operators A and B. Then commutativity implies

$$PAP = AP \leq A \text{ and } PBP = BP \leq B. \tag{3.4}$$

Using property (ii) and (iii), it follows that

$$P(A\sigma B)P \leq (PAP)\sigma(PBP) = (AP)\sigma(BP) \leq A\sigma B, \tag{3.5}$$

so the operator $A\sigma B - P(A\sigma B)P$ is positive and also has a vanishing diagonal block, hence

$$(I-P)\left[A\sigma B - P(A\sigma B)P\right]P, \tag{3.6}$$

in other words P and $A\sigma B$ commute as well. Similarly P commutes with $(AP)\sigma(BP)$, so what follows is that

$$\left[(AP)\sigma(BP)\right]P = (A\sigma B)P. \tag{3.7}$$

Since each scalar commutes with all projections, so does the the operator $1\sigma x$, hence it is a scalar. So $f(x) := 1\sigma x$ defines a real function. We will show that it is operator monotone. Let $0 \leq A \leq B$ be arbitrary with spectral decompositions $A = \sum_i a_i P_i$ and $B = \sum_i b_i Q_i$. Then it follows from (3.7) that

$$I\sigma A = \sum_i \left[P_i \sigma(a_i P_i)\right] P_i = \sum_i (1\sigma a_i) P_i = \sum_i f(a_i) P_i = f(A), \tag{3.8}$$

and similarly $I\sigma B = f(B)$, so by property (ii) we get $f(A) \leq f(B)$, i.e. f is operator monotone. The above implies also using (3.1) that

$$A\sigma B = A^{1/2} f\left(A^{-1/2} B A^{-1/2}\right) A^{1/2}. \tag{3.9}$$

What remains to prove is that every operator monotone function is obtained in the form (3.3). Let f be an operator monotone function. Then it has an integral representation which can be written in the form

$$f(x) = \int_{[0,\infty]} \frac{x(1+t)}{x+t} dm(t) \tag{3.10}$$

for $x > 0$ and m is a positive Radon measure. Then we define a binary operation σ by

$$A\sigma B = aA + bB + \int_{(0,\infty)} \frac{1+t}{t} \left((tA)^{-1} + B^{-1}\right)^{-1} dm(t), \tag{3.11}$$

where $a = m(\{0\})$ and $b = m(\{\infty\})$. Since $\left((tA)^{-1} + B^{-1}\right)^{-1}$ and $aA + bB$ satisfy conditions (ii), (iii) and (iv), the operation σ satisfies condition (ii) and (iii) by convexity of the class of operator connections, while property (iv) is proved by using the Monotone Convergence Theorem in measure theory, so σ is a connection. Finally for $x > 0$

$$1\sigma x = f(x) \tag{3.12}$$

as well, so we obtain the function f from the connection. □

By the above theorem we say that f is the representing function of a connection (or a mean if property (i) is fulfilled as well). In the case of matrix means we have the normalization condition $f(1) = 1$, which follows from property (i). Operator monotone functions which have that $f(1) = 1$ are called normalized operator monotone functions. It is also trivial that matrix means fulfill the property $M(A, A) = A$. Actually it turns out that a connection is a mean if and only if its representing function's Radon measure is a probability measure.

By the above integral representation we have the following

Corollary 3.2. *Every connection σ has the following properties:*

1. $(A\sigma B) + (C\sigma D) \leq (A + C)\sigma(B + D)$.

2. $S^*(A\sigma B)S \leq (S^*AS)\sigma(S^*BS)$ *for not necessarily hermitian* S.

Definition 3.2. We say that a connection σ is symmetric if and only if $A\sigma B = B\sigma A$ for arbitrary positive A, B. Symmetricity is similarly defined for matrix means as well.

Theorem 3.3. *The map,* $n \mapsto \sigma$*, defined by*

$$A\sigma B = \frac{c}{2}(A+B) + \int_{(0,1]} \frac{1+t}{2t}\left[\left((tA)^{-1}+B^{-1}\right)^{-1} + \left(A^{-1}+(tB)^{-1}\right)^{-1}\right] dn(t), \tag{3.13}$$

where $c = n(\{0\})$*, establishes an affine isomorphism from the class of positive Radon measures on the interval* $[0,1]$ *onto the class of symmetric connections.*

Proof. The fact that (3.13) is a symmetric connection is straightforward. Conversely, let σ be a symmetric connection with representing function f. It is not hard to see that $f(x) = xf(1/x)$ (actually a connection is symmetric if and only if this holds). Hence

$$\begin{aligned} f(x) &= \frac{f(x)+xf(1/x)}{2} = \\ &= \frac{a+b}{2}(1+x) + \int_{(0,\infty)} (1+t)\left(\frac{x}{x+t} + \frac{x}{xt+1}\right) dm(t) = \\ &= \frac{a+b}{2}(1+x) + \int_{(0,\infty)} \frac{1+t}{2}\left(\frac{x}{x+t} + \frac{x}{xt+1}\right) dn(t), \end{aligned} \tag{3.14}$$

where $dn(t) = dm(t) + dm(t^{-1})$, and $n(\{0\}) = a+b$.

It remains to prove that a measure n producing σ is unique. We may consider the measure $dm(t) = \frac{1}{2}dn(t)$ or $dm(t) = \frac{1}{2}dn(t^{-1})$ on $[0,\infty]$ according as $0 < t < 1$ or $1 < t < \infty$, and $m(\{1\}) = n(\{1\})$, $m(\{0\}) = m(\{\infty\}) = \frac{1}{2}n(\{0\})$. Now due to Theorem 3.1 and 2.25 the uniqueness of m, hence of n follows. □

In the above theorem to a symmetric mean corresponds a probability measure. Thus we obtain

Theorem 3.4. *Arithmetic mean is the maximum of all symmetric means, while the harmonic mean is the minimum.*

Proof. We have the inequality

$$\frac{2x}{1+x} \leq \frac{1+t}{2}\left(\frac{x}{x+t} + \frac{x}{xt+1}\right) \leq \frac{1+x}{2} \tag{3.15}$$

for $x, t > 0$, which yields

$$2(A^{-1}+B^{-1})^{-1} \leq \frac{1+t}{2t}\left[\left((tA)^{-1}+B^{-1}\right)^{-1} + \left(A^{-1}+(tB)^{-1}\right)^{-1}\right] \leq \frac{A+B}{2}. \tag{3.16}$$

The integration with respect to the probability measure n yields the assertion. □

So far we have met with the two basic matrix means, the arithmetic mean $\frac{A+B}{2}$ and the harmonic mean $2(A^{-1}+B^{-1})^{-1}$. But what would be the geometric mean of two positive matrices? Kubo-Ando theory tells us that we should choose the representing operator monotone function $t^{1/2}$, since the geometric mean of 1 and an arbitrary positive real number t is $t^{1/2}$. This provides us the geometric mean of two positive matrices

$$G(A,B) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}. \tag{3.17}$$

At first glance this does not seem to be symmetric, however it is easy to check that it is so. It has other remarkable properties that we should study later, for instance that it is the metric midpoint of the geodesic line connecting A and B with respect to a Riemannian metric given on the differentiable manifold $P(n,\mathbb{C})$.

4 Extension of the Geometric Mean to Multiple Variables

So far we have only met 2-variable matrix means. Kubo-Ando theory in the preceding section exhaustively characterizes all matrix means by relating every one of them to a normalized operator monotone function. The theory of operator monotone functions is very rich, as we saw in section two, however no such theory has been developed in several variables. A similar theory seems to be very far away at the moment.

The problem to extend a 2-variable matrix mean to several variables is straightforward if we consider the arithmetic or harmonic mean. In this case the several variable formulas coincide with the scalar formulas. The arithmetic mean is just $\frac{\sum_{i=1}^{n} X_i}{n}$, while the harmonic mean is $\left(\frac{\sum_{i=1}^{n} X_i^{-1}}{n}\right)^{-1}$. Most of the properties fulfilled by the 2-variable forms are inherited by these two several variable functions. For instance operator monotonicity is preserved, we also have invariance under permutations of the variables. Property (i), (iii) and (iv) in Definition 3.1 are also preserved. This gives us the motivation of the following

Definition 4.1 (Multivariable Matrix Mean)**.** Let $M : P(r,\mathbb{C})^n \mapsto P(r,\mathbb{C})$. Then M is called a matrix mean if the following conditions hold

1. $M(X,\dots,X) = X$ for every $X \in P(r,\mathbb{C})$,
2. $M(X_1,\dots,X_n)$ is invariant under the permutation of its variables,
3. $\min(X_1,\dots,X_n) \leq M(X_1,\dots,X_n) \leq \max(X_1,\dots,X_n)$ if min and max exist with respect to the positive definite order,
4. If $X_i \leq X_i'$, then $M(X_1,\dots,X_n) \leq M(X_1',\dots,X_n')$,
5. $M(X_1,\dots,X_n)$ is continuous,

6. $CM(X_1, \ldots, X_n)C^* \leq M(CX_1C^*, \ldots, CX_nC^*)$.

The above properties are fulfilled by the n-variable harmonic and arithmetic mean. But what about the geometric mean (3.17)? It is not even straightforward anymore how to define the n-variable geometric mean. This is a nontrivial problem, actually there are several competing definitions, which are indeed different and have nice properties. In order to understand these extensions, we have to exhibit some of the special properties which are possessed by the geometric mean. First of all the convex cone $P(r, \mathbb{C})$ carries a unique Riemannian structure which is related to the geometric mean.

4.1 The Riemannian Structure on $P(r, \mathbb{C})$

We will follow the lines of [11]. The set $P(r, \mathbb{C})$ is an open subset of the vector space of complex squared matrices, hence it is a differentiable manifold. This vector space can be equipped with a norm called the Frobenius norm, which is of the form

$$\|A\|_2 = \sqrt{Tr\{A^2\}}, \tag{4.1}$$

where Tr denotes the trace of a squared matrix, that is $TrA = \sum_i A_{i,i}$, where $A_{i,j}$ denotes the (i, j) entry of the matrix A. Note that the set $H(r, \mathbb{C})$ is a real vector space with the norm $\|\cdot\|_2$ as well. Now consider the following Riemannian metric

$$\langle X, Y \rangle_p = Tr\left\{p^{-1}Xp^{-1}Y\right\}, \tag{4.2}$$

where $p \in P(r, \mathbb{C})$ and $X, Y \in H(r, \mathbb{C})$. The above inner product is positive definite for every p and is a smooth function in p. As it turns out, the tangent space at every p is the space $H(r, \mathbb{C})$. Using this Riemannian metric, we may write it in the infinitesimal form

$$ds = \sqrt{\langle dp, dp \rangle_p} = \left\|p^{-1/2}dpp^{-1/2}\right\|_2 = \sqrt{Tr\left\{(p^{-1}dp)^2\right\}}. \tag{4.3}$$

If we have a piecewise differentiable path $\gamma : [a, b]$ in $P(r, \mathbb{C})$, we define its length by

$$L(\gamma) = \int_a^b \left\|\gamma^{-1/2}(t)\gamma'(t)\gamma^{-1/2}(t)\right\|_2 dt. \tag{4.4}$$

Now let us denote the group of invertible $r \times r$ matrices over the complex field by $GL(r, \mathbb{C})$. The first important property of the above defined metric is captured in the following

Proposition 4.1. *For each $X \in GL(r, \mathbb{C})$ and for each differentiable path γ, the transformation $p \mapsto X^*pX$ is an isometry of $P(r, \mathbb{C})$, that is*

$$L(\gamma) = L(X^*\gamma X), \tag{4.5}$$

and similarly the transformation $p \mapsto p^{-1}$ is also an isometry.

Proof. We have for each t that

$$\begin{aligned}
&\left\|(X^*\gamma(t)X)^{-1/2}(X^*\gamma'(t)X)(X^*\gamma(t)X)^{-1/2}\right\|_2^2 = \\
&= Tr\left\{(X^*\gamma(t)X)^{-1}(X^*\gamma'(t)X)(X^*\gamma(t)X)^{-1}(X^*\gamma'(t)X)\right\} = \\
&= Tr\left\{X^{-1}\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}\gamma'(t)X\right\} = \\
&= Tr\left\{\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}\gamma'(t)\right\} = \\
&= \left\|\gamma(t)^{-1/2}\gamma'(t)\gamma(t)^{-1/2}\right\|_2^2.
\end{aligned} \tag{4.6}$$

A similar calculation leads to the same argument for the map $p \mapsto p^{-1}$ using that the Fréchet differential of this is

$$\left(\gamma(t)^{-1}\right)' = -\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}. \tag{4.7}$$

□

For any two points $A, B \in P(r, \mathbb{C})$ we define the distance function

$$d(A, B) = \inf\left\{L(\gamma) : \gamma \text{ is a path from } A \text{ to } B\right\}. \tag{4.8}$$

Indeed it is a distance function, since the triangle inequality is fulfilled.

One of the crucial properties called the infinitesimal exponential metric increasing property (IEMI) of this metric is captured in the following

Proposition 4.2 (IEMI). *For all $X, Y \in H(r, \mathbb{C})$ we have*

$$\left\|\exp(X)^{-1/2} D\exp[X][Y]\exp(X)^{-1/2}\right\|_2 \geq \|Y\|_2, \tag{4.9}$$

where $D\exp[X]$ denotes the Fréchet derivative of exp.

Proof. Let X have eigenvalues denoted by λ_i. Then by Theorem 2.11

$$\begin{aligned}
&\exp(X)^{-1/2} D\exp[X][Y]\exp(X)^{-1/2} = \\
&= diag(\exp(-\lambda_i/2))\exp^{[1]}(X) \circ Y\, diag(\exp(-\lambda_i/2)) = \\
&= \left[\frac{\exp\left(\frac{\lambda_i-\lambda_j}{2}\right) - \exp\left(-\frac{\lambda_i-\lambda_j}{2}\right)}{\lambda_i - \lambda_j}\right]
\end{aligned} \tag{4.10}$$

and the assertion follows form the fact that $\frac{\exp(t/2)-\exp(-t/2)}{t} \geq 1$ for all t. □

Corollary 4.3. *Let $H(t)$ be an arbitrary path in $H(r, \mathbb{C})$ with $a \leq t \leq b$, and let $\gamma(t) = \exp H(t)$. Then*

$$L(\gamma) \geq \int_a^b \|H'(t)\|_2\, dt. \tag{4.11}$$

Proof. By the chain rule $\gamma'(t) = D\exp[H(t)][H'(t)]$, so the inequality follows from the definition of $L(\gamma)$ and IEMI. $\square$

Now if $\gamma(t)$ is a path connecting $A, B \in P(r,\mathbb{C})$, then $H(t) = \log\gamma(t)$ is a path connecting $\log A$ and $\log B$ in $H(r,\mathbb{C})$. The shortest path connecting these two points in the vector space $H(r,\mathbb{C})$ is a straight line, which has length $\|\log A - \log B\|_2$. Considering the above corollary we get that

$$L(\gamma) \geq \|\log A - \log B\|_2 , \tag{4.12}$$

which yields us the exponential metric increasing property (EMI):

Proposition 4.4 (EMI)**.** *For any two points $A, B \in P(r,\mathbb{C})$*

$$d(A,B) \geq \|\log A - \log B\|_2 . \tag{4.13}$$

Definition 4.2 (Geodesic)**.** Let $A, B \in P(r,\mathbb{C})$. A path γ connecting A and B is called a geodesic if $L(\gamma) = d(A,B)$.

Proposition 4.5. *Let $A, B \in P(r,\mathbb{C})$ be commuting matrices. Then* exp *maps the line segment $H(t) = (1-t)\log A + t\log B$ to the geodesic connecting A and B in $P(r,\mathbb{C})$.*

Proof. We have to verify that

$$\gamma(t) = \exp\left(H(t)\right) \tag{4.14}$$

is the unique shortest path joining A and B in the metric space $(P(r,\mathbb{C}), d)$. Since A, B commutes, we have $\gamma(t) = A^{1-t}B^t$ and $\gamma'(t) = (\log B - \log A)\gamma(t)$. Then we have

$$L(\gamma) = \int_0^1 \|\log A - \log B\|_2 \, dt = \|\log A - \log B\|_2 . \tag{4.15}$$

But EMI says that no path can be shorter than this.

For uniqueness suppose $\tilde{\gamma}$ is another path that joins A and B. Then $\log\tilde{\gamma(t)}$ is a path in $H(r,\mathbb{C})$ that joins $\log A$ and $\log B$. By Corollary 4.3 it has length $\|\log A - \log B\|_2$, but in the Euclidean space $H(r,\mathbb{C})$, the unique shortest path, which is a straight line connecting $\log A$ and $\log B$ has the same length, which is a reparametrization of $\log\gamma(t)$. $\square$

It is also straightforward, that the arc-length parametrization of $\gamma(t)$ when A, B commute is indeed

$$\gamma(t) = A^{1-t}B^t \tag{4.16}$$

for $0 \leq t \leq 1$.

Theorem 4.6. *Let $A, B \in P(r,\mathbb{C})$. Then there exists a unique geodesic $\gamma(t)$ connecting A and B with*

$$\gamma(t) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2},\ 0 \leq t \leq 1, \tag{4.17}$$

and $\gamma(t)$ has arc-length parametrization, i.e.

$$d(A, \gamma(t)) = td(A, B). \tag{4.18}$$

Moreover we have

$$d(A, B) = \left\| \log\left(A^{-1/2} B A^{-1/2}\right) \right\|_2 . \tag{4.19}$$

Proof. The matrices I and $A^{-1/2}BA^{-1/2}$ commute, so the geodesic connecting I and $A^{-1/2}BA^{-1/2}$ is arc-length parametrized as

$$\gamma_0(t) = \left(A^{-1/2} B A^{-1/2}\right)^t, \ 0 \leq t \leq 1. \tag{4.20}$$

We apply the isometry $p \mapsto A^{1/2}pA^{1/2}$ according to Proposition 4.1 to obtain the path

$$\gamma(t) = A^{1/2} \left(A^{-1/2} B A^{-1/2}\right)^t A^{1/2} \tag{4.21}$$

connecting the points A and B, so it must be the geodesic connecting the points A and B and also (4.18) follows. □

What follows here from the above assertion is that the Riemannian distance function is given in the form

$$d(A, B) = \left[Tr \left\{ \log(A^{-1/2} B A^{-1/2})^2 \right\} \right]^{1/2} \tag{4.22}$$

on $P(r, \mathbb{C})$. We may go the other way around and calculate the geodesic equations corresponding to the metric (4.2). The geodesic equations will have the form

$$\gamma'' = \gamma' \gamma^{-1} \gamma', \tag{4.23}$$

and with given initial data $\gamma(0) = p \in P(r, \mathbb{C})$ and $\gamma'(0) = X \in H(r, \mathbb{C})$, one gets the solution as

$$\gamma(t) = p^{1/2} \exp\left(p^{-1/2} X p^{-1/2} t\right) p^{1/2}. \tag{4.24}$$

If we consider the above geodesics for a fixed p and let X take arbitrary values from the tangent space at p we arrive at the exponential map of this manifold

$$\exp_p(X) = p^{1/2} \exp\left(p^{-1/2} X p^{-1/2}\right) p^{1/2}. \tag{4.25}$$

We will discuss exponential maps of affinely connected manifolds later. The inverse of the exponential map gives back the logarithm map, which is in this case

$$\log_p(q) = p^{1/2} \log\left(p^{-1/2} q p^{-1/2}\right) p^{1/2}. \tag{4.26}$$

Since for general Riemannian manifolds the distance function is given by

$$d(p, q)^2 = \left\langle \log_p(q), \log_p(q) \right\rangle_p , \tag{4.27}$$

we again end up with the same distance function (4.22) corresponding to the Riemannian metric (4.2).

At this point we must note that the geometric mean (3.17) is the midpoint of the geodesic line connecting A and B, according to Theorem 4.6. This is a very important observation, since in such a way the geometric mean has a corresponding Riemannian metric with respect to it is the midpoint operation. This is also the case with the arithmetic and harmonic mean as well. The corresponding Riemannian metric to the arithmetic mean given on $P(r, \mathbb{C})$ is just the Euclidean metric

$$\langle X, Y \rangle_p = Tr\{XY\} \tag{4.28}$$

for $X, Y \in H(r, \mathbb{C})$. This metric is the induced metric of the Frobenius norm $\|\cdot\|_2$ defined on the vector space of complex squared matrices. The geodesics of this metric (connecting arbitrary A, B) are the straight lines in the space of complex squared matrices

$$\gamma(t) = (1-t)A + tB \tag{4.29}$$

and the midpoint operation is the arithmetic mean.

The harmonic mean is the midpoint operation of the Riemannian metric given in the form

$$\langle X, Y \rangle_p = Tr\{p^{-2}Xp^{-2}Y\}. \tag{4.30}$$

This metric is isometric to the Euclidean vector space given above, which corresponds to the arithmetic mean. The isometry is given by the function $f(X) = X^{-1}$ over the set $P(r, \mathbb{C})$. Since the metric is isometric to a Euclidean space it is itself Euclidean.

Let us turn back to the Riemnnian metric (4.2) corresponding to the geometric mean. We have seen that the Riemannian metrics corresponding to the arithmetic and harmonic mean is Euclidean. What about the metric (4.2) corresponding to the geometric mean? We have to investigate further properties related to this metric to address this question.

Proposition 4.7. *If for some $A, B \in P(r, \mathbb{C})$, the identity matrix I lies on the geodesic connecting A and B, then A and B commute and*

$$\log B = -\frac{1-s}{s} \log A, \tag{4.31}$$

where $s = d(A, I)/d(A, B)$.

Proof. From Theorem 4.6 we know that

$$I = A^{1/2} \left(A^{-1/2}BA^{-1/2}\right)^s A^{1/2}, \tag{4.32}$$

where $s = d(A, I)/d(A, B)$, thus

$$B = A^{1/2}A^{-1/s}A^{1/2} = A^{-(1-s)/s}, \tag{4.33}$$

so A, B commute and (4.31) holds. □

By the above assertion and Proposition 4.5 it follows that exp is isometric on straight line segments in $H(r,\mathbb{C})$ passing through the 0 matrix. Additionally EMI tells us that exp is metric non-decreasing, which tells us that the Riemannian manifold $P(r,\mathbb{C})$ with the metric (4.2) is nonpositively curved, refer to [16].

An equivalent way (in the class of Riemannian manifolds [16]) to formulate this is showing that the semiparallelogram law holds.

Theorem 4.8. *[Semiparallelogram Law] Let $A, B \in P(r,\mathbb{C})$ be arbitrary, and let $M = G(A,B)$ be the midpoint of the geodesic connecting A, B. Then for all $C \in P(r,\mathbb{C})$ we have*

$$d(M,C)^2 \leq \frac{d(A,C)^2 + d(B,C)^2}{2} - \frac{1}{4}d(A,B)^2. \tag{4.34}$$

Proof. Applying the isometry $p \mapsto M^{-1/2}pM^{-1/2}$ to all matrices involved, we may assume $M = I$. Now I is the midpoint of the geodesic connecting A, B so we have by Proposition 4.7 that $\log B = -\log A$ and

$$d(A,B) = \|\log A - \log B\|_2 . \tag{4.35}$$

We have the same for $M = I$ and C,

$$d(M,C) = \|\log M - \log C\|_2 . \tag{4.36}$$

Since $H(r,\mathbb{C})$ is a vector space, it is Euclidean with the norm $\|\cdot\|_2$, hence it satisfies the parallelogram law

$$\|\log M - \log C\|_2^2 = \frac{\|\log A - \log C\|_2^2 + \|\log B - \log C\|_2^2}{2} - \frac{1}{4}\|\log A - \log B\|_2^2 . \tag{4.37}$$

Since $d(M,C) = \|\log M - \log C\|_2$ and $d(A,B) = \|\log A - \log B\|_2$, EMI leads us to the inequality of the assertion. □

Now we know enough about the metric (4.2) to turn back to the problem of extending the geometric mean to several variables. First of all we should be looking for extension methods which gives back the n-variable arithmetic and harmonic means, when we try to extend them from their 2-variable formulas. The first idea is to look for some external characterizations of the n-variable arithmetic and harmonic means.

4.2 Matrix Means defined as The Center of Mass

Suppose W is a complete Riemannian manifold with metric tensor $\langle\cdot,\cdot\rangle_p$ and Riemannian distance function $d(\cdot,\cdot)$. Then we define the center of mass of $p_i \in W$ for $1 \leq i \leq n$ as the minimizer of the function

$$C(x) = \sum_{i=1}^{n} d(x,p_i)^2. \tag{4.38}$$

If a minimizer exists and it is unique we denote it by $\arg\min_{x\in W} C(x)$. Firstly we will show the following

Proposition 4.9 (M. Pálfia [60]). *In the complete metric space* (W, d) *a minimizer of* $C(x)$ *exists and it is unique, if the metric space is nonpositively curved, i.e. the semiparallelogram law holds* (4.34).

Proof. Let $\gamma(t)$ be an arc-length parametrized geodesic connecting $x, y \in W$. Then it is not hard to show using the semiparallelogram law that we have for all $0 \leq t \leq 1$ and $z \in W$ that

$$d(\gamma(t), z)^2 \leq (1-t)d(x,z)^2 + td(y,z)^2 - t(1-t)d(x,y)^2. \tag{4.39}$$

In order to show this first consider the above for dyadic rationals t, i.e. $t = c2^{-j}$ which are dense in $[0, 1]$, then use a continuity argument to obtain it for general t.

So using the above inequality we get

$$C(\gamma(t)) \leq (1-t)C(\gamma(0)) + tC(\gamma(1)) - t(1-t)nd(\gamma(0), \gamma(1))^2. \tag{4.40}$$

Now let $\alpha := \inf_z C(z)$ and let z_l be a sequence of points with $\lim_{l\to\infty} C(z_l) = \alpha$. Let $z_{l,k}$ be the midpoint between z_l and z_k. Then for $l, k \to \infty$

$$\alpha \leq C(z_{l,k}) \leq \frac{C(z_l) + C(z_k)}{2} - \frac{1}{4}nd(z_l, z_k)^2. \tag{4.41}$$

Consequently, $d(z_l, z_k) \to 0$, i.e. z_l is a Cauchy sequence, by completeness it has a limit point $\hat{z}$. Moreover by continuity of $C(x)$ we have $C(\hat{z}) = \inf_z C(z)$.

For uniqueness assume $C(z_0) = C(z_1) = \inf_z C(z) = \alpha$ and $z_0 \neq z_1$. For the midpoint $z_{\frac{1}{2}}$ between z_0, z_1 we get a contradiction, since $\alpha \leq C(z_{\frac{1}{2}}) < \frac{C(z_0)+C(z_1)}{2} = \alpha$. □

We can further characterize the center of mass, since we already now that it exists and is a unique point, by calculating the gradient of $C(x)$. We need a

Definition 4.3. Let W be a Riemannian manifold with metric tensor $\langle \cdot, \cdot \rangle_p$. Then we define the exponential map $\exp_p$ of W at point $p \in W$ as a function mapping from the tangent space at p to the manifold W as follows. Let X_p be an element of the tangent space at p. Then $\exp_p(X_p)$ is the point $\gamma(1)$ on the geodesic emanating from p in the direction of X_p with arc-length parametrization, i.e. $\gamma(0) = p$ and $\gamma'(0) = X_p$. The inverse of the exponential map $\exp_p(X_p)$ at p is called the logarithm map and is denoted by $\log_p(q)$, if we have the above parametrization for $\gamma(t)$ and $\gamma(1) = q$ then $\log_p(q) = X_p$.

By the above definition, it is not hard to see that

$$d(p,q) = \sqrt{\left\langle \log_p(q), \log_p(q) \right\rangle_p} = \sqrt{\left\langle \log_q(p), \log_q(p) \right\rangle_q}. \tag{4.42}$$

We will see later that it is possible to define $\exp_p$ for non-Riemannian manifolds as well, if they are equipped with an affine connection.

Proposition 4.10. *Let W be a Riemannian manifold with metric $\langle \cdot, \cdot \rangle_p$ and Riemannian distance function $d(\cdot, \cdot)$. Then*

$$gradC(x) = -2 \sum_{i=1}^{n} \log_x(p_i). \tag{4.43}$$

Proof. Let f be a smooth function on W. Then the gradient $gradf(p)$ of f in the direction of the vector field X at point p is defined as

$$\langle gradf(p), X_p \rangle_p = \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0}, \tag{4.44}$$

where $\gamma(t)$ is a smooth curve with $\gamma(0) = p$ and $\gamma'(0) = X_p$.

Since *grad* is a linear map, it is enough to calculate the gradient of $f(x) = d(a, x)^2$. Let $\gamma(t)$ be a smooth curve and let

$$c_a(s, t) = \exp_a\left(s \log_a(\gamma(t))\right). \tag{4.45}$$

We will use $\dot{c_a}$ to denote differentiation with respect to t and c_a' to denote differentiation with respect to s of c_a. Then $\frac{D}{ds} c_a'(s,t) = 0$ and $c_a'(0,t) = \log_a(\gamma(t))$, where $\frac{D}{ds}$ denotes covariant differentiation with respect to s. Since $\frac{D}{ds} c_a'(s,t) = 0$ we have

$$2 \left\langle \frac{D}{ds} c_a'(s,t), c_a'(s,t) \right\rangle_{c_a(s,t)} = 0. \tag{4.46}$$

Since covariant differentiation is compatible with the metric by the Fundamental Theorem of Riemmanian geometry, this is equivalent to

$$2 \left\langle \frac{D}{ds} c_a'(s,t), c_a'(s,t) \right\rangle_{c_a(s,t)} = \frac{d}{ds} 2 \left\langle c_a'(s,t), c_a'(s,t) \right\rangle_{c_a(s,t)} = 0, \tag{4.47}$$

that is $\|c_a'(s,t)\|^2_{c_a(s,t)}$ is independent of s. We also have that

$$\langle c_a'(0,t), c_a'(0,t) \rangle_{c_a(0,t)} = \langle \log_a(\gamma(t)), \log_a(\gamma(t)) \rangle_a \tag{4.48}$$

and by the independence of $\|c_a'(s,t)\|^2_{c_a(s,t)}$ from s we get that

$$\begin{aligned} d(a, \gamma(t))^2 &= \langle \log_a(\gamma(t)), \log_a(\gamma(t)) \rangle_a = \langle c_a'(0,t), c_a'(0,t) \rangle_{c_a(0,t)} = \\ &= \langle c_a'(s,t), c_a'(s,t) \rangle_{c_a(s,t)}. \end{aligned} \tag{4.49}$$

Now we calculate

$$\begin{aligned} \left. \frac{d}{dt} d(a, \gamma(t))^2 \right|_{t=0} &= \left. \frac{d}{dt} \langle c_a'(s,t), c_a'(s,t) \rangle_{c_a(s,t)} \right|_{t=0} = \\ &= 2 \left. \left\langle \frac{D}{dt} c_a'(s,t), c_a'(s,t) \right\rangle_{c_a(s,t)} \right|_{t=0} = \end{aligned} \tag{4.50}$$

now we use the fact that covariant derivatives commute with ordinary partial derivatives, i.e. $\frac{D}{dt}\frac{d}{ds} = \frac{D}{ds}\frac{d}{dt}$

$$= 2\left\langle \frac{D}{ds}\dot{c}_a(s,t), c_a'(s,t)\right\rangle_{c_a(s,t)}\Bigg|_{t=0}. \tag{4.51}$$

Since $d(a,\gamma(t))^2$ is independent of s we have that

$$\int_0^1 d(a,\gamma(t))^2 ds = d(a,\gamma(t))^2. \tag{4.52}$$

Hence

$$\begin{aligned}
\frac{d}{dt}d(a,\gamma(t))^2\Big|_{t=0} &= \frac{d}{dt}\int_0^1 d(a,\gamma(t))^2 ds\Big|_{t=0} = \\
&= \int_0^1 2\left\langle \frac{D}{ds}\dot{c}_a(s,t), c_a'(s,t)\right\rangle_{c_a(s,t)} ds\Bigg|_{t=0} = \\
&= \int_0^1 2\left\langle \frac{D}{ds}\dot{c}_a(s,t), c_a'(s,t)\right\rangle_{c_a(s,t)} + \left\langle \dot{c}_a(s,t), \underbrace{\frac{D}{ds}c_a'(s,t)}_{=0}\right\rangle_{c_a(s,t)} ds\Bigg|_{t=0} = \\
&= \int_0^1 2\,\frac{d}{ds}\left\langle \dot{c}_a(s,t), c_a'(s,t)\right\rangle_{c_a(s,t)} ds\Big|_{t=0} = \\
&= 2\left\langle \dot{c}_a(1,t), c_a'(1,t)\right\rangle_{c_a(1,t)} - \left\langle \dot{c}_a(0,t), c_a'(0,t)\right\rangle_{c_a(0,t)}\Big|_{t=0} = \\
&= 2\left\langle \dot{c}_a(1,0), c_a'(1,0)\right\rangle_{c_a(1,0)} - \left\langle \dot{c}_a(0,0), c_a'(0,0)\right\rangle_{c_a(0,0)}.
\end{aligned} \tag{4.53}$$

Now since $\dot{c}_a(0,0) = 0$, $\dot{c}_a(1,0) = \gamma'(0)$, $c_a'(0,0) = \log_a(\gamma(0))$, $c_a'(1,0) = -\log_{\gamma(0)}(a)$ and $c_a(1,0) = \gamma(0)$, we have that

$$\frac{d}{dt}d(a,\gamma(t))^2\Big|_{t=0} = 2\left\langle \gamma'(0), -\log_{\gamma(0)}(a)\right\rangle_{\gamma(0)}. \tag{4.54}$$

This shows that $grad d(a,p)^2 = -2\log_p(a)$. □

Corollary 4.11. *An immediate consequence of the above proposition is that if* $\arg\min_{x\in W} C(x)$ *exists and is unique, it can be found by solving the equation*

$$0 = gradC(x) = -2\sum_{i=1}^{n}\log_x(p_i). \tag{4.55}$$

Let us do this in the case of the arithmetic mean. Consider the convex cone $P(r,\mathbb{C})$ as a subset of the vector space $H(r,\mathbb{C})$. The norm $\|\cdot\|_2$ on $H(r,\mathbb{C})$ yields us the Euclidean metric

$$d_E(A,B) = \sqrt{Tr\left\{(A-B)^2\right\}} \tag{4.56}$$

on the vector space $H(r, \mathbb{C})$. The restriction of this metric to $P(r, \mathbb{C})$ is also Euclidean and we have already mentioned that the 2-variable arithmetic mean is the geodesic midpoint operation on this space.

Corollary 4.12. *The n-variable arithmetic mean* $\frac{\sum_{i=1}^{n} A_i}{n}$ *is the center of mass of the points* $A_1, \ldots, A_n \in P(r, \mathbb{C})$ *with respect to the Euclidean metric* (4.56).

Proof. Proposition 4.9 tells us that the center of mass exists and is unique since the metric (4.56) is Euclidean, therefore the semiparallelogram law holds with equality (parallelogram law) mentioned earlier. By Corollary 4.11 we need to solve the equation

$$-2 \sum_{i=1}^{n} (X - A_i) = 0 \tag{4.57}$$

for $X \in P(r, \mathbb{C})$, since in this case $\log_p(q) = q - p$. The solution is the n-variable arithmetic mean. □

Proposition 4.13 (M. Pálfia [64])**.** *Let* $d(X, Y)$ *be defined as*

$$d(X, Y) = d_E(f(X), f(Y)), \tag{4.58}$$

where $f : P(r, \mathbb{C}) \mapsto P(r, \mathbb{C})$ *is a diffeomorphism. Then the unique minimizer* $\hat{X}$ *of the function*

$$C(X) = \sum_{i=1}^{n} d(X, X_i)^2 \tag{4.59}$$

is given as

$$\hat{X} = f^{-1}\left(\frac{\sum_{i=1}^{n} f(X_i)}{n}\right). \tag{4.60}$$

Proof. Since the corresponding metric $d(\cdot, \cdot)$ is a pullback of the Euclidean metric over the space of squared complex matrices it is also Euclidean. Using the isometric embedding f, the object function of the minimization problem is of the form

$$\sum_{i=1}^{n} d(X, X_i)^2 = \sum_{i=1}^{n} d_E\left(f(X), f(X_i)\right)^2. \tag{4.61}$$

But since by the previous corollary the Riemannian center of mass of the set $S = \{f(X_1), \ldots, f(X_n)\}$ in the Euclidean space of squared complex matrices is the arithmetic mean of the points $\{f(X_1), \ldots, f(X_n)\}$, therefore

$$A = \frac{\sum_{i=1}^{n} f(X_i)}{n} \tag{4.62}$$

minimizes the functional $\sum_{i=1}^{n} d_E\left(X, f(X_i)\right)^2$, so $\hat{X} = f^{-1}(A)$ minimizes $\sum_{i=1}^{n} d_E\left(f(X), f(X_i)\right)^2$. □

If we choose $f(X) = X^{-1}$ in the above proposition, we get that the n-variable harmonic mean is also characterized as the center of mass on a Riemannian manifold.

Now since the geometric mean $G(A, B)$ has also a corresponding Riemannian metric (4.2) where it is the center of mass of the two points A, B, we may define the n-variable geometric mean as the center of mass similarly to the arithmetic mean, since by Theorem 4.8 the metric space is nonpositively curved and Proposition 4.9 ensures the existence and uniqueness of the center of mass [47]. Since the logarithm map has the form (4.26) the center of mass of the points $X_1 \dots, X_n \in P(r, \mathbb{C})$ with repsect to the metric (4.2) is the unique solution $X \in P(r, \mathbb{C})$ of the equation

$$0 = \sum_{i=1}^{n} \log_{X_i}(X) = \sum_{i=1}^{n} \log\left(X_i^{-1/2} X X_i^{-1/2}\right). \tag{4.63}$$

This is a nonlinear matrix equation and it has not yet been solved analytically so far, however if we consider it for mutually commuting X_i, we can easily solve it analytically and the solution is

$$X = \prod_{i=1}^{n} X_i^{1/n}, \tag{4.64}$$

which is the usual geometric mean of positive numbers. The invariance under the permutations of the X_i of the center of mass is trivial, while operator monotonicity in its variables was an open question for several years, it has been solved very recently in [39] using the Riemannian structure (4.2) and its nonpositive curvature combined with a characteirization of the center of mass using probability theory.

4.3 Symmetrization Procedures and Weighted Means

Many researchers were focusing on the extension of the 2-variable geometric mean to several variables, since it has the corresponding Riemannian structure (4.2). This Riemannian metric space structure gives a very strong tool to extend the geometric mean. We have already seen the analogy to the arithmetic and harmonic means via the center of mass characterization. This idea essentially appeared first in [47]. We mention a few other constructions very soon. But before that we spend a few words on 2-variable weighted means. First of all it must be noted that Kubo-Ando theory characterizes matrix means and gives lower and upper bounds on possible symmetric means, however it tells nothing further about how to "weight" a symmetric mean. In the case of the arihtmetic, harmonic and geometric means, this is more or less straightforward, we can use the geodesic lines of the Riemannian structure to define 2-variable weighted means. In this case for $t \in [0, 1]$ the weighted arithmetic mean is given as

$$A_t(A, B) = (1 - t)A + tB, \tag{4.65}$$

while the weighted harmonic mean is given as

$$H_t(A,B) = \left[(1-t)A^{-1} + tB^{-1}\right]^{-1}. \tag{4.66}$$

Using the Riemannian structure the weighted geometric mean is

$$G_t(A,B) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2}. \tag{4.67}$$

We can see again the importance of the Riemannian structures corresponding to symmetric matrix means, since it provides us with weighted matrix means corresponding to symmetric ones as geodesic lines. Now we turn to other mean extension procedures. Consider the following procedure called the Ando-Li-Mathias procedure [5].

Definition 4.4. [ALM iteration] Let $X = (X_1^0, \dots, X_n^0)$ where $X_i^0 \in P(r,\mathbb{C})$ and define the mapping $M(X_1, \dots, X_n)$ inductively as follows. If $n = 2$ assume that $M(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \dots, X_{n-1})$ is already defined. Then using $M(X_1, \dots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M\left(Z_{\neq i}\left(X_1^l, \dots, X_n^l\right)\right), \tag{4.68}$$

where $Z_{\neq i}(X_1^l, \dots, X_n^l) = X_1^l, \dots, X_{i-1}^l, X_{i+1}^l, \dots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l\to\infty} X_i^l = M(X_1^0, \dots, X_n^0). \tag{4.69}$$

Theorem 4.14 (Theorem 3.2 [5])**.** *The limit in Definition 4.4 starting with $M(A,B) := G(A,B)$, $M(X_1, \dots, X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.*

The above proof relies heavily on the Riemannian structure (4.2). In [38] it was proved by Lawson and Lim that the above procedure converges in nonpositively curved metric spaces, using the midpoint operation of the space as the 2-variable mean to extend from. The convergence in Theorem 4.14 was shown to be linear. Temesi and Petz in [68] showed that the ALM procedure converges for orderable tuples of matrices, however the general case still remained open. Since the procedure recursively relies on itself, it is quite ineffective even for small n. Hence in [15] the following similar procedure was defined. Both of the above and the following procedures are referred to, in general, as symmetrization procedures.

Definition 4.5. [BMP iteration] Let $X = (X_1^0, \dots, X_n^0)$ where $X_i^0 \in P(r,\mathbb{C})$ and define the mapping $M(X_1, \dots, X_n)$ inductively as follows. If $n = 2$ assume that $M_t(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \dots, X_{n-1})$ is already defined. Then using $M(X_1, \dots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M_{\frac{n-1}{n}}\left(X_i^l, M\left(Z_{\neq i}\left(X_1^l, \dots, X_n^l\right)\right)\right), \tag{4.70}$$

where $Z_{\neq i}(X_1^l, \dots, X_n^l) = X_1^l, \dots, X_{i-1}^l, X_{i+1}^l, \dots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l\to\infty} X_i^l = M(X_1^0, \dots, X_n^0). \tag{4.71}$$

Theorem 4.15 (Theorem 3.1 [15])**.** *The limit in Definition 4.5 starting with $M(A,B) := G(A,B)$, $M(X_1,\dots,X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.*

The most important property of this procedure is essentially summarized in

Theorem 4.16 (Theorem 3.2 [15])**.** *The procedure in Definition 4.5 considered for the geometric mean converges cubically.*

The proofs of the above theorems were also relying on the metric structure (4.2). The important properties of the ALM- and BMP-procedures considered for the geometric mean are that their limit points fulfill the properties listed in Definition 4.1 and also some additional properties which are intuitively expected from a geometric mean [5, 15].

One of the major drawbacks, from the computational point of view of both symmetrization procedures, is their recursivity. Namely in order to be able to compute an n-mean we have to provide the (n-1)-variable version which is itself defined as a limit point of the same iteration. Therefore even if $n = 4$ we run into serious computational difficulties, since it is very hard to find the limit point of either symmetrization procedures in closed form of their initial variables (to be more precise this has not even been achieved yet). The author in [63] defined a procedure which is similar to the above two, but relies directly on the 2-variable form of means. In the next chapter we will discuss this procedure in a metric geometric setting and then later for every matrix mean.

5 Means in Complete k-convex Metric Spaces

We have already seen that the geometric mean is the midpoint operation on a Riemannian manifold of nonpositive curvature (4.2). This manifold is a complete metric space and we have already mentioned that in [38] Lawson and Lim considered the ALM-process in complete metric spaces of nonpositive curvature, which is a generalization of the geometric mean to metric geometric setting.

We will define our procedure in a more general metric setting, namely we will define our mean on complete metric spaces with a certain positive upper curvature bound. In other words, our procedure will not only work in nonpositively curved complete metric spaces, but also in positively curved metric spaces. These spaces will be called k-convex metric spaces.

5.1 k-convexity of Metric Spaces

Let (X,d) be a metric space. Let $I \subset (R)$, then the length $L(\gamma)$ of a curve $\gamma : I \to X$ is defined as the supremum of $\sum_{i=1}^{n} d(\gamma(t_{i-1}), \gamma(t_i))$ where $t_0 \leq$

$t_1 \leq \cdots \leq t_n$ and $t_0, \ldots, t_n \in I$. A curve $\gamma : I \to X$ is called a geodesic if and only if $d(\gamma(s), \gamma(t)) = d(\gamma(s), \gamma(r)) + d(\gamma(r), \gamma(t))$ for all $s < r < t \in I$. The metric space X is called a geodesic length space or geodesic metric space if any two points x, y can be connected with a geodesic γ such that $L(\gamma) = d(x, y)$. We denote the open ball by $B(x, r)$ with circumcenter x and radius r and with $\bar{B}(x, r)$ its closed counterpart.

The following definition of k-convexity is due to Ohta in [53]. We will establish our results for spaces with such properties below.

Definition 5.1. Let $k \in (0, 2]$.

- An open set U in a geodesic metric space (X, d) is called a C_k-domain if for any three points x, y, z, any geodesic $\gamma : [0, 1] \mapsto X$ between x, y and for all $t \in [0, 1]$ we have

$$d(z, \gamma(t))^2 \leq (1-t)d(z, x)^2 + td(z, y)^2 - \frac{k}{2}t(1-t)d(x, y)^2. \tag{5.1}$$

- A geodesic metric space (X, d) is k-convex if it is itself a $C_k - domain$.
- A geodesic metric space (X, d) is locally k-convex if every point in X is contained in a C_k-domain.

If the inequality (5.1) holds for $t = 1/2$ then it holds for all $t \in [0, 1]$. A k-convex metric space becomes a $CAT(0)$ space if the above inequality holds for $k = 2$. In this case the space is said to have nonpositive curvature in the sense of Alexandrov, in other words the semiparallelogram law holds.

We begin with recalling and investigating properties of k-convex spaces.

Lemma 5.1 (Lemma 2.2 in [53])**.** *If an open ball $B(x, r) \subset X$ is a C_k-domain then for any two points in $B(x, r)$ a geodesic is unique between them. In particular any two points in a k-convex metric space are connected by a unique geodesic.*

Proof. Fix two points $y, z \in B(x, r)$ and let $\alpha(t), \beta(t) : [0, 1] \mapsto X$ be two geodesics between y and z. For each $t \in (0, 1)$ take a minimal geodesic $\gamma : [0, 1] \mapsto X$ from $\alpha(t)$ to $\beta(t)$. It then follows from the k-convexity that

$$\begin{aligned} d(y, \gamma(1/2))^2 &\leq \frac{1}{2}d(y, \alpha(t))^2 + \frac{1}{2}d(y, \beta(t))^2 - \frac{k}{8}d(\alpha(t), \beta(t))^2 \\ &= t^2 d(y, z)^2 - \frac{k}{8}d(\alpha(t), \beta(t))^2 \end{aligned} \tag{5.2}$$

and similarly

$$d(z, \gamma(1/2))^2 \leq (1-t)^2 d(y, z)^2 - \frac{k}{8}d(\alpha(t), \beta(t))^2. \tag{5.3}$$

Therefore we have by using the triangle inequality that

$$\begin{aligned} d(y, z) \leq &\sqrt{t^2 d(y, z)^2 - \frac{k}{8}d(\alpha(t), \beta(t))^2} \\ &+ \sqrt{(1-t)^2 d(y, z)^2 - \frac{k}{8}d(\alpha(t), \beta(t))^2}, \end{aligned} \tag{5.4}$$

i.e., $\alpha(t) = \beta(t)$. □

We will also say that a geodesic metric space is uniquely geodesic if any two points can be connected by a unique distance minimizing geodesic. According to the above lemma we have unique geodesics between two points, therefore we also have unique metric midpoints as well, which will be the midpoints of these geodesics.

The existence of minimal balls containing a bounded subset in a C_k-domain is guaranteed by (5.1). The proof follows the case when $k = 2$.

Lemma 5.2 (M. Pálfia [60]). *Let S be a bounded subset of a complete k-convex metric space X. Then there exists a unique closed ball with minimal radius r containing S.*

Proof. We use a similar technique as the one in Proposition 5.10 in [6]. Let $r(x, S) = \sup_{y\in S} d(x, y)$ for $x \in X$ and $r(S) = \inf_{x\in X} r(x, S)$. For all $x, y \in X$ we have

$$r(m, S)^2 \leq \frac{r(x, S)^2 + r(y, S)^2}{2} - \frac{k}{8} d(x, y)^2, \tag{5.5}$$

where $m = \gamma(1/2)$ is the unique midpoint between x and y. From the above it is easy to conclude the following inequalities

$$\begin{aligned} d(x, y)^2 &\leq \frac{4}{k}\left[r(x, S)^2 + r(y, S)^2\right] - \frac{8}{k} r(m, S)^2 \\ &\leq \frac{4}{k}\left[r(x, S)^2 + r(y, S)^2\right] - \frac{8}{k} r(S)^2. \end{aligned} \tag{5.6}$$

From the above the uniqueness of the circumcenter of the metric ball is obvious, since if we had two circumcenters $c_1, c_2 \in X$ then from the above inequality we get $d(c_1, c_2) = 0$. We also have that a sequence x_n with $r(x_n, S) \to r(S)$ is Cauchy since by the inequalities above

$$d(x_m, x_n)^2 \leq \frac{4}{k}\left[r(x_m, S)^2 + r(x_n, S)^2\right] - \frac{8}{k} r(S)^2. \tag{5.7}$$

Hence by completeness x_n has a limit point x, which is the circumcenter of the ball with minimal radius. So we have $S \subset \bar{B}(x, r(S))$. □

It also follows that any metric ball in a C_k-domain is also a geodesically convex set.

Lemma 5.3 (M. Pálfia [60]). *Let $B(x, r) \subset D$ where D is a C_k-domain. Then $B(x, r)$ is a geodesically convex set which means that every geodesic which connects two points in $B(x, r)$ is also a subset of $B(x, r)$.*

Proof. Let $y, z \in B(x, r)$ so we have $d(x, y) \leq r$ and $d(x, z) \leq r$. Let $\gamma : [0, 1] \mapsto D$ be a geodesic (which is unique according to Lemma 5.1) such that $\gamma(0) = y$

and $\gamma(1) = z$. Now for arbitrary $t \in [0, 1]$

$$\begin{aligned} d(x, \gamma(t))^2 &\leq (1-t)d(x,y)^2 + td(x,z)^2 - \frac{k}{2}t(1-t)d(y,z)^2 \\ &\leq (1-t)d(x,y)^2 + td(x,z)^2 \\ &\leq (1-t)r^2 + tr^2 \\ &\leq r^2, \end{aligned} \tag{5.8}$$

so we have $d(x, \gamma(t)) \leq r$. This means that $\gamma(t) \subset B(x, r)$ so $B(x, r)$ is geodesically convex. $\square$

An important consequence of the above lemmas is the existence of a geodesic convex hull of a bounded set.

Definition 5.2 (Geodesic convex hull)**.** The geodesic convex hull $GCH(S)$ is the intersection of all convex sets containing S.

We may construct this set in the following way:

Proposition 5.4 (Proposition 2.5.5 in [66])**.** *$GCH(S)$ can be obtained as $GCH(S) = \bigcup_{n \geq 0} F_n$, where $F_0 = S$ and for $n \geq 1$ the set F_n consists of all points which lie on geodesics with starting and ending points in F_{n-1}.*

Proof. The union $\bigcup_{n \geq 0} F_n$ is an increasing union. Using the fact that the convex hull $GCH(S)$ is a geodesically convex set, it is easy to see by induction that for each $n \geq 0$, the set F_n is contained in $GCH(S)$. Thus, we have $\bigcup_{n \geq 0} F_n \subset GCH(S)$. Conversely if $x, y \in \bigcup_{n \geq 0} F_n$, then they belong to a set F_n for some $n \geq 0$. So the geodesic connecting x, y is contained in F_{n+1}, therefore in $\bigcup_{n \geq 0} F_n$. Thus, the set $\bigcup_{n \geq 0} F_n$ is geodesically convex and therefore it contains $GCH(S)$, i.e. $GCH(S) = \bigcup_{n \geq 0} F_n$. $\square$

What follows from Lemma 5.3 is that the geodesic convex hull of a bounded set is contained in a convex metric ball.

We will base the results in the next section on the above properties of k-convex metric spaces.

5.2 The Iterative Mean

In this section we will provide an extension of midpoint maps as means between two points on a k-convex metric space to several variables. This new mean is called the Iterative mean. Important properties of this mean will also be investigated here.

We will use the following notation to denote the unique midpoint between two points in a uniquely geodesic metric space as

$$a \sharp b = \gamma_{a,b}(1/2). \tag{5.9}$$

Definition 5.3 (Iterative process, M. Pálfia [60]). Let $Q_1^0, \ldots, Q_n^0$ be points in a uniquely geodesic metric space X and $\pi = \{\pi_0, \pi_1, \ldots\}$ be an infinite sequence of permutations, where each π_i is a permutation of the letters $\{1, \ldots, n\}$. With respect to the infinite sequence of permutations π let

$$Q_i^{l+1} = \begin{cases} Q_{\pi_l(i)}^l \sharp Q_{\pi_l(i+1)}^l & \text{if } 1 \le i < n, \\ Q_{\pi_l(n)}^l \sharp Q_{\pi_l(1)}^l & \text{else.} \end{cases} \tag{5.10}$$

The above procedure yields a sequence of n-tuple of points.

Theorem 5.5 (Iterative mean, M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1^0, \ldots, Q_n^0$ be points in the metric space X. Let us set up the iteration in Definition 5.3 on these points in X with respect to an infinite sequence of permutations $\pi = \{\pi_0, \pi_1, \ldots\}$. Then the sequences Q_i^l converge to a common limit point.*

Proof. We begin showing that the distances $d(Q_i^l, Q_j^l)$ are converging to zero, after that we will show that the Q_i^l sequences are themselves convergent. For the sake of simplicity of notations from now on we define $\pi_l(n+1) := \pi_l(1)$.

Let us consider one iteration step in Definition 5.3. From the k-convexity of X for every Q_i^1 and for arbitrary $x \in X$ we have

$$d\left(x, Q_i^1\right)^2 \le \frac{d\left(x, Q_{\pi_0(i)}^0\right)^2 + d\left(x, Q_{\pi_0(i+1)}^0\right)^2}{2} - \frac{k}{8} d\left(Q_{\pi_0(i)}^0, Q_{\pi_0(i+1)}^0\right)^2, \tag{5.11}$$

where $Q_i^1 = Q_{\pi_0(i)}^0 \sharp Q_{\pi_0(i+1)}^0$. If we consider the sum of these equations above for every i we arrive at

$$\sum_{i=1}^n d\left(x, Q_i^1\right)^2 \le \sum_{i=1}^n d\left(x, Q_i^0\right)^2 - \frac{k}{8} \sum_{i=1}^n d\left(Q_{\pi_0(i)}^0, Q_{\pi_0(i+1)}^0\right)^2, \tag{5.12}$$

as can be seen in Figure 3. Applying this to every iteration step we get

$$\underbrace{\sum_{i=1}^n d\left(x, Q_i^{l+1}\right)^2}_{a_{l+1}(x)} \le \underbrace{\sum_{i=1}^n d\left(x, Q_i^l\right)^2}_{a_l(x)} - \frac{k}{8} \underbrace{\sum_{i=1}^n d\left(Q_{\pi_l(i)}^l, Q_{\pi_l(i+1)}^l\right)^2}_{e_l}. \tag{5.13}$$

Note that the above is valid for every possible infinite sequence of permutations π.

Now the sequence $a_l(x) \ge 0$ measures the sum of the squared distances from an arbitrary point x and the points of the n-tuple in the lth iteration step. By (5.13) we have

$$a_{l+1}(x) \le a_l(x) - (k/8)e_l \quad \text{for all } x \in X. \tag{5.14}$$

In other words the sequence $a_l(x)$ indexed by l is monotone decreasing, since $e_l \ge 0$, and it is also bounded from below by 0 and above too by the initial

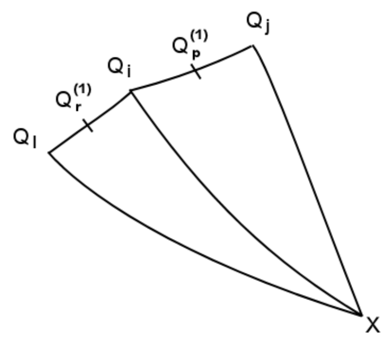

Figure 3: One iteration step in the Theorem 5.5 in a k-convex metric space.

finite value $a_0(x)$, therefore it is convergent. From the convergence of $a_l(x)$ and by (5.14) it is easy to see that $e_l \to 0$. This means that the points Q_i^l are approaching one another.

Let us consider the geodesic convex hull of the starting n points. This geodesic convex hull exists and is bounded because it is a subset of a convex metric ball according to Lemma 5.3. Set $A^l = GCH\left(\left\{Q_1^l, \ldots, Q_n^l\right\}\right)$. By the definition and construction of the convex hull in Proposition 5.4 we have $A^l \supseteq A^{l+1}$. We will show that the limit set $A = \lim_{l\to\infty} A^l = \bigcap A^l$ is of diameter zero, therefore a singleton according to Cantor's intersection theorem [72] (Theorem C of Ch. 2.12).

Firstly we will show that for arbitrary points x, y in the geodesic convex hull A^l we have

$$d\left(x, y\right)^2 \leq \max_{1 \leq p,q \leq n} d\left(Q_p^l, Q_q^l\right)^2 . \tag{5.15}$$

The geodesic convex hull A^l itself can be obtained by the method given in Proposition 5.4. Hence it is enough to show that for any two geodesics $\alpha(t)$ and $\beta(s)$ parametrized by arc-length with ending points in the set F_j we have

$$d\left(\alpha(t), \beta(s)\right)^2 \leq \max_{1 \leq p,q \leq n} d\left(Q_p^l, Q_q^l\right)^2 . \tag{5.16}$$

One way to obtain this with the notations $\alpha(0) = a, \alpha(1) = b, \beta(0) = c, \beta(1) = d$, is the following

$$
\begin{aligned}
d(\alpha(t),\beta(s))^2 &\le (1-t)d(a,\beta(s))^2 + td(b,\beta(s))^2 - t(1-t)\frac{k}{2}d(a,b)^2 \\
d(a,\beta(s))^2 &\le (1-s)d(a,c)^2 + sd(a,d)^2 - s(1-s)\frac{k}{2}d(c,d)^2 \\
d(b,\beta(s))^2 &\le (1-s)d(b,c)^2 + sd(b,d)^2 - s(1-s)\frac{k}{2}d(c,d)^2 .
\end{aligned}
\tag{5.17}
$$

Substituting into the first inequality above one arrives at the following

$$
\begin{aligned}
d(\alpha(t),\beta(s))^2 \le & (1-t)(1-s)d(a,c)^2 + (1-t)sd(a,d)^2 \\
& + t(1-s)d(b,c)^2 + tsd(b,d)^2 \\
& - t(1-t)\frac{k}{2}d(a,b)^2 - s(1-s)\frac{k}{2}d(c,d)^2 \\
\le & \max\left\{d(a,c)^2, d(a,d)^2, d(b,c)^2, d(b,d)^2\right\}.
\end{aligned}
\tag{5.18}
$$

Applying the above inequality recursively in every step of the construction of the convex hull in Proposition 5.4 one derives (5.16).

Now using the triangle inequality one automatically obtains the bound

$$
\sum_{i=1}^{n} d\left(Q^l_{\pi_l(i)}, Q^l_{\pi_l(i+1)}\right) \ge 2 \max_{1\le p,q\le n} d\left(Q^l_p, Q^l_q\right). \tag{5.19}
$$

Now as $e_l \to 0$ we have $d\left(Q^l_{\pi_l(i)}, Q^l_{\pi_l(i+1)}\right) \to 0$, so one obtains easily

$$
\lim_{l\to\infty} \max_{1\le p,q\le n} d\left(Q^l_p, Q^l_q\right) = 0. \tag{5.20}
$$

Thus also $diam(A^l) \to 0$ so by Cantor's intersection theorem the limit A is a singleton, which by completeness implies that any sequence of points $x^l \in A^l$ converges to this singleton A, so also every Q^l_i converges to this singleton as well. □

The above theorem ensures the convergence of the sequences Q^l_i, but it does not tell anything about the asymptotic rate of convergence to the common limit point. The next theorem ensures that the convergence rate is at least linear.

Theorem 5.6 (M. Pálfia [60])**.** *Let (X,d) be a complete k-convex geodesic metric space. Let $Q^0_1,\dots,Q^0_n$ be points in the metric space X. Let us set up the iteration in Definition 5.3 on these points in X. Let R denote the common limit point of these sequences. Then*

$$
\frac{a_{l+1}(R)}{a_l(R)} \le 1 - \frac{k}{2n^2}, \tag{5.21}
$$

so the points Q^l_i are converging to R linearly.

Proof. We will give a lower bound on $e_l/a_l(R)$ and use (5.14) to provide an upper bound on $a_{l+1}(R)/a_l(R)$.

Again by the triangle inequality one automatically obtains a lower bound on e_l as

$$\sum_{i=1}^{n} d\left(Q^l_{\pi_l(i)}, Q^l_{\pi_l(i+1)}\right) \geq 2 \max_{1\leq p,q\leq n} d\left(Q^l_p, Q^l_q\right). \tag{5.22}$$

Using the Cauchy-Schwarz inequality and the fact that all the above terms are positive one gets

$$\sum_{i=1}^{n} d\left(Q^l_{\pi_l(i)}, Q^l_{\pi_l(i+1)}\right) \leq \sqrt{n}\sqrt{\sum_{i=1}^{n} d\left(Q^l_{\pi_l(i)}, Q^l_{\pi_l(i+1)}\right)^2} = \sqrt{n}\sqrt{e_l}. \tag{5.23}$$

Hence we obtain the lower bound

$$\frac{4}{n} \max_{1\leq p,q\leq n} d\left(Q^l_p, Q^l_q\right)^2 \leq e_l. \tag{5.24}$$

According to the preceding proof, for arbitrary points x, y in the geodesic convex hull $A^l = GCH\left(\left\{Q^l_1, \ldots, Q^l_n\right\}\right)$ we have

$$d\left(x, y\right)^2 \leq \max_{1\leq p,q\leq n} d\left(Q^l_p, Q^l_q\right)^2. \tag{5.25}$$

Now from (5.24) and (5.16) we get

$$\frac{e_l}{a_l(R)} \geq \frac{4}{n^2}, \tag{5.26}$$

which together with (5.14) prove the theorem. □

If we consider the proof of the above theorem and (5.14), it is easy to see that for different permutations π_l between iteration steps in Theorem 5.5 we may get slower and faster rates of convergence. One way to speed up the rate of convergence to the common limit point R - which itself may depend on the chosen sequence of permutations π in Theorem 5.5 - is to maximize the error term e_l by which $a_l(R)$ at least decreases.

Taking into account the above mentioned, we can modify our iterational scheme with adding some heuristics, making it adaptive to the geometry of the sets given by the points in every iterational step. The function Idealmapping defined by Algorithm 1 by the author in [63] returns the array ma which contains the indices of points $Q^l_1, \ldots, Q^l_n$ in such an order, that if we set up one iterational step as letting $\pi_l(i) = ma(i)$, we can sufficiently reduce the distance between them.

For numerical results consult Figure 4, where we used the Riemannian manifold $P(r, \mathbb{C})$ with (4.2) as the complete k-convex metric space. Iteration 1 was performed using the same permutation π_0 for all iterations steps, i.e. $\pi_l = \pi_0$ for all $l \geq 0$. At the same time Iteration 3 used the heuristic function Idealmapping in every iteration step. We have measured the distance from the the "solution", which itself was calculated by letting Iteration 3 converge to under a certain treshold. We have also calulated the center of mass of the starting points by using two procedures to approximate it, a gradient method and a newton method, refer to [46].

Algorithm 1 Idealmapping

```
Require: x_1, ..., x_n
d ⇐ n(n − 1)/2
i ⇐ 1, j ⇐ 2
for k = 0 to d do
    r[k, 1] ⇐ d(x_i, x_j)
    r[k, 2] ⇐ i, r[k, 3] ⇐ j
    if j = n then
        j ⇐ n − i + 2, i ⇐ 1
    else
        i ⇐ i + 1, j ⇐ j + 1
    end if
end for
sort r by r[k, 1] descending
ma[1] ⇐ r[1, 2]
j ⇐ r[1, 3]
for k = 2 to n do
    find largest r[i, 1] for such i that (r[i, 2] = j or r[i, 2] = j) and r[i, 2] ∉ ma and r[i, 3] ∉ ma
    if r[i, 2] = j then
        j ⇐ r[i, 3], ma[k] ⇐ r[i, 2]
    else
        j ⇐ r[i, 2], ma[k] ⇐ r[i, 3]
    end if
end for
return ma
```

It is also crucial to point out that the common limit point R appears to depend on the infinite sequence of permutations π as numerical experiments suggest. So therefore one might prefer to use the notation R_π to express the dependence on the sequence of permutations π.

A major improvement of the Iterative mean over the ALM- and BMP-mean is that it relies only on the 2-variable form of means, so this procedure is feasible even for large number of variables. Also it is proved here that it has linear convergence rate, like the ALM-process. We will see later that it also satisfies almost all of the properties which are required from a matrix mean, except permutation invariance. We will discuss these features in later sections, when we consider this process for all possible matrix means.

In the next section we will use the extension method given in this section to approximate the center of mass of the points $Q_1^0, \dots, Q_n^0$.

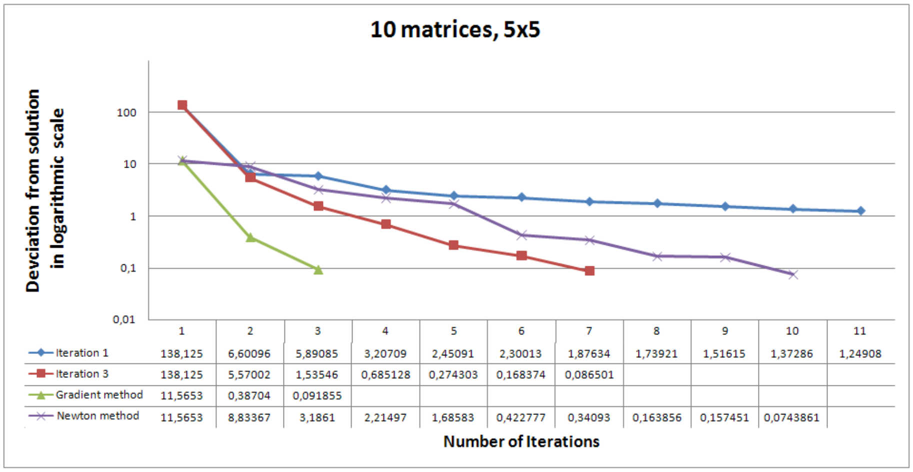

Figure 4: Convergence rate results for 10, 5×5 matrices.

5.3 Center of mass and k-convexity

Before one can study the center of mass of points in a metric space, it must be made sure that it is a unique point and it exists. If the metric space is complete and it has nonpositive curvature then this point exists and is unique as we have seen it before in Proposition 4.9.

The first steps will be to show the uniqueness of the center of mass in k-convex metric spaces.

Theorem 5.7 (M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1, \ldots, Q_n$ be points in the metric space X. Then the center of mass*

$$\arg\min_{x \in X} \sum_{i=1}^{n} d(x, Q_i)^2 \tag{5.27}$$

exists and is a unique point in the space X where $\arg\min_{x \in X} C(x)$ denotes the unique point that minimizes a function $C(x)$.

Proof. The proof of this is essentially the same as the proof of Proposition 4.9. □

Using (5.14) we can control the distance of an arbitrary point $x \in X$ and the limit point R_π of the procedure in the following

Theorem 5.8 (M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1, \ldots, Q_n$ be points in the metric space X. Then for arbitrary*

$x \in X$ and for the limit point R_π of the procedure in Theorem 5.5 set up on the points $Q_1, \ldots, Q_n$ we have the following inequality

$$d(R_\pi, x) \leq \sqrt{\frac{\sum_{i=1}^{n} d(x, Q_i)^2 - \frac{k}{8}\sum_{l=1}^{\infty} e_l}{n}}. \tag{5.28}$$

Proof. From (5.14) for all $x \in X$ we have

$$\frac{k}{8} e_l \leq a_l(x) - a_{l+1}(x). \tag{5.29}$$

Taking a finite sum of the above equations one arrives at the following

$$S_m = \frac{k}{8} \sum_{l=1}^{m} e_l \leq a_1(x) - a_{m+1}(x). \tag{5.30}$$

We can take the limit of the sums S_m as it is a monotone increasing sequence bounded from above, so we conclude that

$$\lim_{m\to\infty} a_m(x) \leq a_1(x) - \frac{k}{8}\sum_{l=1}^{\infty} e_l, \tag{5.31}$$

but $\lim_{m\to\infty} a_m(x)$ is nothing but $nd(x, R_\pi)^2$. This yields

$$d(x, R_\pi) \leq \sqrt{\frac{\sum_{i=1}^{n} d(x, Q_i)^2 - \frac{k}{8}\sum_{l=0}^{\infty} e_l}{n}} \text{ for all x.} \tag{5.32}$$

□

Now since the center of mass is unique in complete k-convex metric spaces, one can consider the following corollary of the above result.

Corollary 5.9 (M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1, \ldots, Q_n$ be points in the metric space X. Then the center of mass*

$$Y = \arg\min_{x\in X} \sum_{i=1}^{n} d(x, Q_i)^2 \tag{5.33}$$

and the limit point R_π of the procedure in Theorem 5.5 set up on the points $Q_1, \ldots, Q_n$ fulfill the following inequality

$$d(R_\pi, Y) \leq \sqrt{\frac{\sum_{i=1}^{n} d(Y, Q_i)^2 - \frac{k}{8}\sum_{l=1}^{\infty} e_l}{n}}. \tag{5.34}$$

It is interesting to consider the fact that if X is a Euclidean space then it has zero curvature which turns (5.1) with $k = 2$ into an equality which is the parallelogram law of the Euclidean space

$$d(z, \gamma_{x,y}(1/2))^2 = \frac{d(z,x)^2 + d(z,y)^2}{2} - \frac{1}{4} d(x,y)^2. \tag{5.35}$$

So in this case (5.14) turns into an equality as well

$$a_{l+1}(x) = a_l(x) - \frac{1}{4}e_l \quad \text{for all } x \in X. \tag{5.36}$$

One may minimize both sides with respect to x and conclude

$$\arg\min_{x\in X} a_{l+1}(x) = \arg\min_{x\in X} a_l(x). \tag{5.37}$$

Here we used the basic fact that the error term e_l is independent of x so therefore it acts as a constant term with respect to the above minimization.

Therefore we have just proved the following

Proposition 5.10 (M. Pálfia [60])**.** *If X is a Euclidean space then the limit point R_π of the procedure in Theorem 5.5 is the center of mass of the starting points for every possible infinite sequence of permutations π.*

Following the path of the above proposition one can conclude that in certain special situations even more is true.

Proposition 5.11 (M. Pálfia [60])**.** *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1, \ldots, Q_n$ be points in the metric space X that lie on a single geodesic segment. Then the limit point R_π of the procedure in Theorem 5.5 set up on the points $Q_1, \ldots, Q_n$ is the center of mass of the points $Q_1, \ldots, Q_n$.*

Proof. A single geodesic segment equipped with the distance function inherited from the space X is a Euclidean space, therefore it is just the case of Proposition 5.10. □

It is worth noting that one must be aware of the fact that the limit point R_π of the procedure depend on the chosen infinite sequence π. If (5.1) turns into an equality, as in the case of a single geodesic segment or Euclidean space, then the possibly different limit points depending on π of the procedure will collapse onto one unique point, the center of mass.

Bhatia and Holbrook studied the question in [13] that whether the ALM-mean is the same as the center of mass. It turned out by numerical simulations that they are generally slightly different. In this context we can say a bit more about this. Since the manifold $P(r, \mathbb{C})$ has nonpositive curvature (Theorem 4.8), it is automatically k-convex for $k = 2$, therefore we may use all the above machinery presented above. Now the ALM-mean is the same as the one given by Theorem 5.5 for $n = 3$ variables. By Theorem 5.11 we know that the above two points are the same as long as the starting points lie on a single geodesic segment. In the other cases we have an upper bound on their distance according to Corollary 5.9. It seems so that the difference of the two points may be due to nonzero curvature. Theorem 5.5 gives an extended geometric mean for several matrices which is the same as the one defined by Jung-Lee-Yamazaki in [32] for one particular infinite sequence of permutations π.

This is also the case for the special orthogonal group which is also an actively studied manifold in terms of averaging. The references are [46], [48]. $SO(n)$

is locally k-convex which is the consequence of the following two important propositions in [53].

Proposition 5.12 (Ohta)**.** *A $CAT(1)$-space (X, d) with $\operatorname{diam} X \leq \pi/2 - \epsilon, \epsilon \in (0, \pi/2)$ is k-convex for $k = (\pi - 2\epsilon)\sin\epsilon/\cos\epsilon$.*

Proposition 5.13 (Ohta)**.** *An Alexandrov space with a local upper curvature bound is locally k-convex for any $k \in (0, 2)$.*

For these kind of spaces we may also use the machinery presented here in the previous sections since they are themselves C_k-domains or they have subsets small enough to be C_k-domains. Therefore the above claims for the case of the geometric mean can be carried out also for these spaces. In these spaces we can define means of several points as an extension of the midpoint maps in the spaces as the limit point of the procedure in Theorem 5.5. The above two propositions due to Ohta tells us how to translate the requirement of k-convexity to the language of curvature. As we can see, an upper curvature bound suffices.

Interestingly enough in the case of the arithmetic and harmonic means, the ALM-mean, the BMP-mean and the Iterative mean are all the same. This follows from Corollary 4.12 which tells us that the n-variable arihtmetic mean is the center of mass on $P(r, \mathbb{C})$ with its inherited Euclidean metric, while similarly by Proposition 4.13 with diffeomorphism $f(X) = X^{-1}$, we have that the n-variable harmonic mean is the center of mass as well on the manifold $P(r, \mathbb{C})$ diffeomorphic to itself by $f(X)$. Furthermore Proposition 5.10 tells us that since these spaces are Euclidean, our Iterative procedure gives back the center of mass. Actually Proposition 5.10 does not prove this explicitly for the ALM- and BMP-process, but using the same ideas, similar proofs can be carried out for these two procedures as well. While this argument work for these means which are related to Euclidean spaces, it does not work in the case of the geometric mean. Even more it seems so that if the metric space is not Euclidean, then all these means are different.

6 Symmetric Matrix Means as Metric Midpoints

The kind reader probably noticed how important are these metric structures corresponding to matrix means. Especially in the case of the geometric mean, where the corresponding space is non-Euclidean, therefore the above discussed extension problems are far from being trivial. The original program of the author was to show that each symmetric matrix mean is a midpoint operation on a complete metric space, thus all the matrix mean extension problems are treatable through the above metric geometric framework. However it turned out that this program was too ambitious, there are actually very few symmetric matrix means which are midpoint operations on Riemannian manifolds. Even there are very few if we just want the corresponding manifolds to be affinely connected, but not necessarily metrizable.

In this section we will answer these questions. We will classify all possible affinely connected manifolds which have a midpoint operation that happens to be a symmetric matrix mean. This classification will also show us which symmetric matrix means have a corresponding weighted mean that is also a geodesic in some affinely connected manifold. We will begin with some general geometrical constructions, which will be applied later in the case of symmetric matrix means. With the help of these tools we exhibit some symmetries of the possible affine connections that can occur, which ultimately lead to their classification. In fact it turns out here at the end that all symmetric matrix means which are midpoint operations on $P(r,\mathbb{C})$ have its corresponding affine connection of the form

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2}\left(X_p p^{-1} Y_p + Y_p p^{-1} X_p\right), \tag{6.1}$$

where $0 \leq \kappa \leq 2$ and the tangent space is $H(r,\mathbb{C})$, the space of hermitian matrices, at every point $p \in P(r,\mathbb{C})$. This result is summarized in Theorem 6.14 in the end of this section. During the classification process we will exhaustively study the properties of these possible connections, namely their parallel transports, metrizability, symmetricity, etc.

6.1 Affinely Connected Manifolds and the Exponential Map

Let W be a smooth manifold. The tangent bundle TW is the disjoint union of all the tangent spaces T_pW at point p, i.e.

$$TW = \bigcup_{p\in W} \{p\} \times T_pW. \tag{6.2}$$

Definition 6.1. [Affine Connection] An affine connection (or Koszul connection) ∇ on a smooth manifold W is a mapping

$$\begin{aligned} C^\infty(W,TW) \times C^\infty(W,TW) &\mapsto C^\infty(W,TW) \\ (X,Y) &\mapsto \nabla_X Y \end{aligned} \tag{6.3}$$

of smooth vector fields $X, Y \in C^\infty(W,TW)$, which satisfies the following properties:

1. $\nabla_{fX} Y = f\nabla_X Y$, that is, ∇ is $C^\infty(W,\mathbb{R})$-linear in the first variable.
2. $\nabla_X(fY) = df[X] + f\nabla_X Y$, that is, it satisfies the Leibniz-rule in the second variable.
3. $\nabla_X(Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2$, that is, linearity in the second variable.

The geodesics of an affine connection can also be defined as smooth curves $\gamma(t)$ satisfying

$$\nabla_{\gamma'(t)}\gamma'(t) = 0. \tag{6.4}$$

In this case the exponential map $exp_p(X_p)$ is defined to be

$$\exp_p(X_p t) = \gamma(t), \tag{6.5}$$

where $\gamma'(0) = X_p$. Similarly to the Riemannian case, the logarithm map $\log_p(q)$ is defined as the inverse of $\exp_p(X_p)$.

We also define the parallel transport vector field $X(t)$ of a given vector $X_{\gamma(0)} \in T_{\gamma(0)}W$ along a smooth curve $\gamma(t)$ as the solution of the ODE

$$\nabla_{\gamma'(t)} X(t) = 0. \tag{6.6}$$

A Riemannian structure automatically leads to a distinguished affine connection, the Levi-Civita connection. The only connection which is compatible with the metric $\langle \cdot, \cdot \rangle_p$, according to the Fundamental Theorem of Riemannian geometry [27, 25]. The above definitions are given in the modern, index-less notation. We may state them fixing a coordinate frame using indices. In particular to an affine connection ∇, in the fixed coordinate frame $e_i = \frac{\partial}{\partial x^i}$, we have corresponding Christoffel-symbols Γ^i_{jk} given as

$$\nabla_{e_j} e_k = \Gamma^i_{jk} e_i. \tag{6.7}$$

This gives the equivalence between the index-less and the classical notation. If we have a Riemannian metric g_{ij}, that is, a given positive definite tensor at every tangent space, smoothly varying over the manifold W, the corresponding metric compatible Levi-Civita connection is determined by the assumption that

$$\nabla_{e_l} g_{ik} = 0. \tag{6.8}$$

From this we obtain the Christoffel-symbols in the form

$$\Gamma^i_{jk} = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right), \tag{6.9}$$

where g^{ik} denotes the inverse of g_{ik}. It follows that the Levi-Civita connection is a symmetric connection (torsion-free), i.e. $\Gamma^i_{jk} = \Gamma^i_{kj}$.

The covariant derivative of a vector field $X^m E_m$ is given as

$$\nabla_{e_l} X^m = \frac{\partial X^m}{\partial x^l} + \Gamma^m_{kl} X^k. \tag{6.10}$$

Similarly we define the covariant derivative of tensors as

$$\nabla_{e_l} A^{ik} = \frac{\partial A^{ik}}{\partial x^l} + \Gamma^i_{ml} A^{mk} + \Gamma^k_{ml} A^{im}. \tag{6.11}$$

For covariant tensors we have a negative sign before each Γ^i_{jk} and the indices are lowered accordingly.

In the remaining of this section we reconstruct the exponential map of an arbitrary affinely connected differentiable manifold based on its midpoint map. Without loss of generality we fix a base point p as the starting point of the geodesics. The basics of the exponential map of a manifold can be found for example in Chapter I. paragraph 6 [27].

Theorem 6.1 (M. Pálfia [56]). *Let M be an affinely connected smooth manifold diffeomorphically embedded into a vector space V. Suppose that the midpoint map $m(p,q) = \exp_p(1/2\log_p(q))$ is known in every normal neighborhood where the exponential map $\exp_p(X)$ is a diffeomorphism. Then in these normal neighborhoods the inverse of the exponential map $\log_p(q)$ can be fully reconstructed from the midpoint map in the form*

$$\log_p(q) = \lim_{n\to\infty} \frac{m(p,q)^{\circ n} - p}{\frac{1}{2^n}}, \tag{6.12}$$

where we use the notation $m(p,q)^{\circ n} \equiv m\left(p, m(p,q)^{\circ(n-1)}\right)$.

Proof. We will use some basic properties of the differential of the exponential map to construct the inverse of it, the logarithm map. Since in small enough normal neighborhoods the exponential map is a diffeomorphism, it can be given as the inverse of the logarithm map $\log_p(q)$.

By the basic properties of the exponential map we have

$$\left.\frac{\partial \exp_p(Xt)}{\partial t}\right|_{t=0} = \lim_{t\to 0} \frac{\exp_p(Xt) - p}{t} = X, \tag{6.13}$$

where $X \in T_pM$. Here we used the fact that we have an embedding into a vector space. Suppose $\exp_p(X) = q$ is in the normal neighborhood. We are going to provide the limit on the right hand side of the above equation. The limit clearly exists in the normal neighborhood so

$$\lim_{t\to 0} \frac{\exp_p(Xt) - p}{t} = \lim_{n\to\infty} \frac{\exp_p\left(X\frac{1}{2^n}\right) - p}{\frac{1}{2^n}} = \lim_{n\to\infty} \frac{m(p,q)^{\circ n} - p}{\frac{1}{2^n}}. \tag{6.14}$$

Here we use the notation $m(p,q)^{\circ n} \equiv m\left(p, m(p,q)^{\circ(n-1)}\right)$. We are in a normal neighborhood so the exponential map has an inverse, the logarithm map, so the limit may be written as

$$X = \lim_{t\to 0} \frac{\exp_p(Xt) - p}{t} = \lim_{n\to\infty} \frac{m(p,q)^{\circ n} - p}{\frac{1}{2^n}} = \log_p(q). \tag{6.15}$$

□

In the above assertion we used the midpoint map to reconstruct the exponential map, but we can use any map that yields a point, other then the ending points, on the geodesic connecting two points in the normal neighborhood. This is summarized in the following

Proposition 6.2 (M. Pálfia [56]). *Let M be an affinely connected smooth manifold diffeomorphically embedded into a vector space V. In every normal neighborhood N let $\gamma_{a,b}(t)$ denote the geodesic connecting $a, b \in N$ with parametrization $\gamma_{a,b}(0) = a$ and $\gamma_{a,b}(1) = b$. Suppose that the map $m(a,b)_{t_0} = \gamma_{a,b}(t_0) = \exp_p(t_0 \log_p(q))$ is known for a $t_0 \in (0,1)$ in every normal neighborhood N where* *the exponential map is a diffeomorphism and $a, b \in N$. Then in these normal neighborhoods the logarithm map can be fully reconstructed as*

$$\log_p(q) = \lim_{n\to\infty} \frac{m(p,q)_{t_0}^{\circ n} - p}{t_0^n}, \tag{6.16}$$

with the notation $m(p,q)_{t_0}^{\circ n} \equiv m\left(p, m(p,q)_{t_0}^{\circ(n-1)}\right)_{t_0}$. We also obtain the exponential map by inverting $\log_p(q)$.

We are going to use this construction in the next sections to characterize matrix means which occur as midpoint maps on affinely connected manifolds.

6.2 Geometric Constructions Applied to Matrix Means

As we already know since Section 3, that an important consequence of Kubo-Ando theory is that every matrix mean can be written in the form

$$M(A,B) = A^{1/2} f\left(A^{-1/2} B A^{-1/2}\right) A^{1/2}, \tag{6.17}$$

where $f(t)$ is a normalized operator monotone function. For symmetric means, we have $f(t) = tf(1/t)$ which implies that $f'(1) = 1/2$. Recall from Section 2 the integral characterization that an operator monotone function $f(t)$, which is defined over the interval $(0, \infty)$, possesses:

$$f(t) = \alpha + \beta t + \int_0^\infty \left(\frac{\lambda}{\lambda^2+1} - \frac{1}{\lambda+t}\right) d\mu(\lambda), \tag{6.18}$$

where α is a real number, $\beta \geq 0$ and μ is a positive measure on $(0, \infty)$ such that

$$\int_0^\infty \frac{1}{\lambda^2+1} d\mu(\lambda) < \infty. \tag{6.19}$$

We will use this integral characterization at several points in order to show that certain functions are analytic.

We are interested in finding all possible symmetric matrix means which are also geodesic midpoint operations on smooth manifolds. We will call such a matrix mean affine [64]:

Definition 6.2 (Affine matrix mean, M. Pálfia [56])**.** An affine matrix mean $M : W^2 \mapsto W$ is a symmetric matrix mean which is at the same time a geodesic midpoint operation $M(A,B) = \exp_A(1/2 \log_A(B))$ on a smooth manifold $W \supseteq P(n, \mathbb{C})$ equipped with an affine connection, where B is assumed to be in the injectivity radius of the exponential map $\exp_A(x)$ of the connection given at the point A. The mapping $\log_A(x)$ is just the inverse of the exponential map at the point $A \in W$.

The following assertion will show that if a matrix mean is affine then the exponential map of the corresponding smooth manifold has a special structure. We will use similarly the notation $M(A,B)^{\circ n} = M\left(A, M(A,B)^{\circ(n-1)}\right)$ as before.

Theorem 6.3 (M. Pálfia [56]). *Let $M(A,B)$ be a symmetric matrix mean. Then*

$$\lim_{n\to\infty} \frac{M(A,B)^{\circ n} - A}{\frac{1}{2^n}} = A^{1/2} \log_I \left(A^{-1/2} B A^{-1/2} \right) A^{1/2} \tag{6.20}$$

where the limit exists and is uniform for all $A, B \in P(n, \mathbb{C})$ and $\log_I(t)$ is an operator monotone function on the interval $(0, \infty)$.

Proof. We will prove the convergence to a continuous function $\log_I(t)$ in a more general setting. The operator monotonicity in the matrix mean case will be a particularization.

First of all note that by the repeated usage of (6.17) we can reduce the above problem to the right hand side of the following formula:

$$\frac{M(A,B)^{\circ n} - A}{\frac{1}{2^n}} = A^{1/2} \frac{f\left(A^{-1/2} B A^{-1/2}\right)^{\circ n} - I}{\frac{1}{2^n}} A^{1/2}. \tag{6.21}$$

From now on we will explicitly use the notation $g(t)^{\circ n} = g\left(g(t)^{\circ(n-1)}\right)$ for arbitrary function $g(t)$ where this notation is straightforward.

Due to the above formula it is enough to prove the assertion for a single operator monotone function $f(t)$. If the corresponding matrix mean is symmetric then we have $f(t) = tf(1/t)$ which implies that the derivative of the operator monotone function $f(t)$ is $1/2$ at the identity, so $f'(1) = 1/2$. Actually this is just the special case of this problem considered for arbitrary concave, analytic functions $f(t)$ given in the following form

$$\lim_{n\to\infty} \frac{f(X)^{\circ n} - I}{f'(1)^n}, \tag{6.22}$$

for $X \in P(n, \mathbb{C})$. As every operator monotone function which maps $(0, \infty)$ to $(0, \infty)$, is analytic on $(0, \infty)$ and has an analytic continuation on the complex half-plane, we can consider the functional calculus for hermitian matrices in the above equations. Therefore we can further reduce the problem to the set of the positive reals by diagonalizing X and considering the convergence for every distinct diagonal element separately.

Without loss of generality we may shift the function $f(t)$ by 1 so it is enough to show the assertion for

$$\lim_{n\to\infty} \frac{g(t)^{\circ n}}{g'(0)^n}, \tag{6.23}$$

where $g(t) = f(t+1) - 1$ and so $g(t)^{\circ n} = f(t+1)^{\circ n} - 1$. From now on we will be considering the shifted problem for the sake of simplicity. At this point we must emphasize the fact that the function $g(t)$ must have 0 as an attractive and only fixed point on the interval of interest $(-1, \infty)$. In the unshifted case this is equivalent to $f(t)$ having 1 as the only attractive fixed point on the interval $(0, \infty)$, which is the case by Banach's fixed point theorem for normalized operator monotone functions $f(t)$ with $f'(1) = 1/2$ (operator monotone functions are also concave, so $f''(t) \leq 0$). We can also assume that $0 < g'(0) < 1$. The rest of the argument will be based on the claim that the above limit of analytic functions of the form $g(t)^{\circ n}/g'(0)^n$ is uniform Cauchy therefore the limit function exists and is continuous.

First of all we have 0 as the attractive and only fixed point of $g(t)$, so for arbitrary $x \in (-1, \infty)$ the sequence $x_n = g(x)^{\circ n}$ converges to 0. We have $g(0) = 0$ and by the mean value theorem we have

$$x_n = g(x)^{\circ n} = g'(t_n) g(x)^{\circ(n-1)} = \prod_{i=1}^{n} g'(t_i) x, \tag{6.24}$$

where $t_i \in \left[0, g(x)^{\circ(i-1)}\right]$ if $x \geq 0$ or $t_i \in \left[g(x)^{\circ(i-1)}, 0\right]$ if $x < 0$, since $g(t)$ is a concave function on $(-1, \infty)$. As $x_n \to 0$ for arbitrary x we have $g'(t_i) \to g'(0)$. Now we have to obtain a suitable upper bound on

$$\left| \frac{g(x)^{\circ n}}{g'(0)^n} - \frac{g(x)^{\circ m}}{g'(0)^m} \right|. \tag{6.25}$$

We argue as follows

$$\begin{aligned}
&\left| \frac{g(x)^{\circ n}}{g'(0)^n} - \frac{g(x)^{\circ m}}{g'(0)^m} \right| = \frac{|g(x)^{\circ n} - g'(0)^{n-m} g(x)^{\circ m}|}{g'(0)^n} \leq \\
&\leq \frac{\left|\prod_{i=m+1}^{n} g'(t_i) - g'(0)^{n-m}\right| \left|\prod_{i=1}^{m} g'(t_i)\right|}{g'(0)^n} |x| = \\
&= \left| \prod_{i=m+1}^{n} \frac{g'(t_i)}{g'(0)} - 1 \right| \left| \prod_{i=1}^{m} \frac{g'(t_i)}{g'(0)} \right| |x|.
\end{aligned} \tag{6.26}$$

Now uniform convergence follows if $|\prod_{i=1}^{\infty} g'(t_i)/g'(0)| < \infty$ because then the tail $\prod_{i=m+1}^{\infty} g'(t_i)/g'(0) \to 1$ so (6.25) can be arbitrarily small on any compact interval in $(-1, \infty)$ by choosing a uniform m. By the continuity of $g'(t)$ and $x_n \to 0$ we have $g'(t_i) \to g'(0)$ and by assumption $0 < g'(0) < 1$, therefore there exists N and q such that for all $i > N$ we have $0 < g'(t_i) \leq q < 1$. What follows here is that $\exists K_1, K_2 < \infty$ such that $|t_N| \leq K_1$ and $|g''(t_i)| \leq K_2$ for all $i > N$. This yields the bound $|t_i| \leq K_1 q^{i-N}$ for all $i > N$. Considering the Taylor expansion of $g'(t)$ around 0 we get

$$\frac{g'(t_i)}{g'(0)} = \frac{g'(0) + g''(t_i')t_i}{g'(0)} \tag{6.27}$$

for $0 < t_i' < t_i$. What follows from this is that

$$\left| \prod_{i=N}^{\infty} \frac{g'(t_i)}{g'(0)} \right| \leq \prod_{i=N}^{\infty} \left(1 + \frac{K_1 K_2}{g'(0)} q^{i-N} \right). \tag{6.28}$$

The infinite product on the right hand side converges because $\sum_{j=0}^{\infty} \frac{K_1 K_2}{g'(0)} q^j$ converges hence $|\prod_{i=1}^{\infty} g'(t_i)/g'(0)| < \infty$ for all x in the compact interval.

At this point we can easily establish the convergence for normalized operator monotone functions because they are concave functions by Theorem 2.4, so $f''(t) \leq 0$ and they have only one fixed point which is 1. The fact that the limit is operator monotone in this case follows from the operator monotonicity of the generating $f(t)$. □

Actually the above proof works for a larger class of functions then the family of normalized operator monotone functions. The limit in (6.22) exists and it is a continuous function if the twicely differentiable function $f(t)$ has 1 as the only attractive fixed point and the derivative $-1 < f'(t) < 1$.

Proposition 6.4 (M. Pálfia [56]). *The limit function* $\log_I(t)$ *in Theorem 6.3 maps* $P(n, \mathbb{C})$ *to* $H(n, \mathbb{C})$ *injectively and*

$$I - X^{-1} \leq \log_I(X) \leq X - I \tag{6.29}$$

for all $X \in P(n, \mathbb{C})$ *with respect to the positive definite order of matrices.*

Proof. By Theorem 2.12 we know that an operator monotone function has nonnegative derivative, also by Theorem 2.4 we have that its second derivative is nonpositive. Now suppose that $\log_I(t)$ has zero derivative at some point t_0 in its domain. Then by the preceeding two observations, for all $t \geq t_0$, $\log_I(t)$ must be constant. Since this function is analytic on $(0, \infty)$ and it has an analytic continuation by virtue of Corollary 2.26 to the upper half plane. So if it is constant for all $t \geq t_0$, then its power series consist of a constant term. The function, since it is analytic on $(0, \infty)$, equals to its power series on the domain of its analyticity, so it should be constant on the whole $(0, \infty)$ interval.

Now we will show that this cannot happen. Suppose we have two normalized operator monotone functions $f(t)$ and $g(t)$ corresponding to two symmetric matrix means and $f(X) \leq g(X)$ for all $X \in P(n, \mathbb{C})$. Then it is easy to see that $\log_{I,f}(X) \leq \log_{I,g}(X)$ for the two $\log_I(t)$ corresponding to $f(t)$ and $g(t)$ respectively in Theorem 6.3. By Theorem 3.4 we know that the smallest symmetric matrix mean is the harmonic mean, while the largest is the arithmetic mean, in other words

$$\left(\frac{A^{-1} + B^{-1}}{2}\right)^{-1} \leq M(A, B) \leq \frac{A + B}{2} \tag{6.30}$$

for all symmetric matrix means $M(A, B)$ and arbitrary $A, B \in P(n, \mathbb{C})$. This inequality is equivalent to

$$\left(\frac{I + X^{-1}}{2}\right)^{-1} \leq f(X) \leq \frac{I + X}{2} \tag{6.31}$$

at the level of the representing normalized operator monotone functions. Now the harmonic and the arithmetic means are affine means, in particular they correspond to Euclidean manifolds. The logarithm map is $\log_I(X) = X - I$ in the case of the arithmetic mean, while $\log_I(X) = I - X^{-1}$ in the case of the harmonic mean, by using Theorem 6.1 and 6.3 and the corresponding

Euclidean metric structures. Now again we have $\log_{I,f}(X) \leq \log_{I,g}(X)$ for two corresponding normalized operator monotone functions $f(t)$ and $g(t)$. This combined with inequality (6.31) yield (6.29). Now it remains an easy exercise to see that $\log_I(X)$ cannot be constant on $(0,\infty)$, since then it would violate inequality (6.29).

These observations yield that $\log_I(t)$ is injective, since it is nonconstant operator monotone, and it follows from the functional calculus that it maps $P(n,\mathbb{C})$ to $H(n,\mathbb{C})$. □

Since $\log_I(t)$ is operator monotone on $(0,\infty)$, it is also analytic there, so it has an analytic inverse $\exp_I(t)$ by Lagrange's Inversion Theorem, since its derivative is nonzero due to Theorem 6.4. It is also easy to see that $\log'_I(1) = 1$, so $\exp'_I(0) = 1$ and since $\log_I(1) = 0$ we have $\exp_I(0) = 1$. This follows from the fact that

$$\log'_I(t) = \lim_{n\to\infty} \frac{\prod_{i=0}^{n-1} f'\left(f(t)^{\circ i}\right)}{\frac{1}{2^n}}, \tag{6.32}$$

since $\log_I(t)$ is a uniform limit of analytic functions, therefore its differential is the limit of the differential of the functions $\frac{f(t)^{\circ n}-1}{1/2^n}$, which are also uniformly converging due to a similar argument to the one given in the proof of Theorem 6.3 and $f'(1) = 1/2$ by the symmetricity of the matrix mean. By these considerations we have just arrived at the following

Proposition 6.5 (M. Pálfia [56])**.** *If a symmetric matrix mean $M(A,B)$ is an affine mean, then the exponential map and its inverse, the logarithm map are of the following forms*

$$\begin{aligned} \exp_p(X) &= p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} \\ \log_p(X) &= p^{1/2}\log_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} \end{aligned} \tag{6.33}$$

for $p \in P(n,\mathbb{C})$, where $\exp_I(X)$ and $\log_I(X)$ are analytic functions such that $\exp_I : H(n,\mathbb{C}) \mapsto P(n,\mathbb{C})$ and $\log_I(X)$ is its inverse and $\log'_I(I) = I, \exp'_I(0) = I, \log_I(I) = 0, \exp_I(0) = I$.

Note that by Weierstrass's approximation theorem we also have

$$\begin{aligned} p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} &= p\exp_I\left(p^{-1}X\right) \\ p^{1/2}\log_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} &= p\log_I\left(p^{-1}X\right). \end{aligned} \tag{6.34}$$

In some cases, to ensure easier reading, similarly as in the above formulas, we will denote matrices with uppercase letters which are elements of some tangent space, while at the same time we will use lowercase letters for denoting matrices which are points of a differentiable manifold.

6.3 Constructions of Invariant Affine Connections

Let us recall the Riemannian manifold with given metric (4.2). This is actually the symmetric space $GL(n,\mathbb{C})/U(n,\mathbb{C})$, where $U(n,\mathbb{C})$ denotes the group of unitary transformations. We did not cover the theory of symmetric spaces, but we shall not need it, so the symmetricity of this space is just mentioned as a fact, although it is a very important one from a certain point of view [16]. We will turn back to this later. What we do need is that the Levi-Civita connection corresponding to this Riemannian manifold is

$$\nabla_{X_p}Y_p = DY[p][X_p] - \frac{1}{2}\left(X_p p^{-1}Y_p + Y_p p^{-1}X_p\right), \tag{6.35}$$

here $DY[p][X_p]$ denotes the Fréchet-differential of Y at the point p in the direction X_p. One important property fulfilled by symmetric spaces is that their connection is invariant under their parallel transport. So the above connection is also an invariant one.

The question that can be asked at this point is that are there other symmetric matrix means which correspond to symmetric spaces as midpoint maps on $P(n,\mathbb{C})$? Two other examples are known, these are the arithmetic mean and the harmonic mean. The symmetric spaces corresponding to these two means are Euclidean while the symmetric space corresponding to the geometric mean has nonpositive curvature. It has flat and negatively curved de Rham factors.

At this point we begin with the characterization of means that correspond to affine symmetric spaces in general. What we know at this point is that the two functions, which are of each others inverse, $\log_I(t)$ and $\exp_I(t)$ exist for all symmetric matrix means, as it was proved in Theorem 6.3. The calculation of the limit (6.22) might be complicated. We give examples where the limit function may be calculated relatively easily.

Example 6.1. Consider the one-parameter family of means

$$X^{1/2}\left(\frac{I+\left(X^{-1/2}YX^{-1/2}\right)^q}{2}\right)^{1/q}X^{1/2}. \tag{6.36}$$

These functions are actually matrix means if and only if $q \in [-1,1]$ as we will see later, but nonetheless we can consider the case now when q is an arbitrary nonzero real number. The corresponding generating functions are $f_q(t) = \sqrt[q]{(1+x^q)/2}$. We have

$$f_q(x)^{\circ 3} = \left(\frac{1+\frac{1+\frac{1+x^q}{2}}{2}}{2}\right)^{1/q} \tag{6.37}$$

Examining the continued fraction that occurs here, it is easy to justify the following

$$\lim_{n\to\infty}\frac{f_q(x)^{\circ n}-1}{f'(1)^n}=\lim_{n\to\infty}\frac{\left(\frac{x^q}{2^n}+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^n}\right)^{1/q}-1}{\frac{1}{2^n}}=$$
$$=\lim_{n\to\infty}\frac{\left(\frac{x^q}{2^n}+1-\frac{1}{2^n}\right)^{1/q}-1}{\frac{1}{2^n}}=\lim_{t\to 0}\frac{(tx^q-t+1)^{1/q}-1}{t}= \tag{6.38}$$
$$=\left.\frac{\partial(tx^q-t+1)^{1/q}}{\partial t}\right|_{t=0}=\frac{x^q-1}{q}.$$

In [25] and [27] there is an extensive study of affine connections on manifolds. A well known fact is that the affine connection on a manifold can be reconstructed by differentiating the parallel transport in the following way

$$\nabla_{X_p}Y_p=\lim_{t\to 0}\frac{\Gamma_t^0(\gamma)Y_{\gamma(t)}-Y_{\gamma(0)}}{t}, \tag{6.39}$$

where $\gamma(t)$ denotes an arbitrary smooth curve emanating from p in the direction $X_p=\partial\gamma(t)/\partial t|_{t=0}$ and $\Gamma_t^s(\gamma)Y$ denotes the parallel transport of the vector field Y along the curve γ from $\gamma(t)$ to $\gamma(s)$, refer to [25, 27]. The above limit does not depend on the curve itself, only on its initial direction vector and it depends on the vector field Y in an open neighborhood of p. On affine symmetric spaces the parallel transport from one point to another is given by the differential of the geodesic symmetries with a negative sign. The geodesic symmetry is given as

$$S_p(q)=\exp_p(-\log_p(q)). \tag{6.40}$$

On affine symmetric spaces this map is an affine transformation so one can conclude that

$$\Gamma_1^0(\gamma)Y=-\left.\frac{\partial S_{\gamma(1/2)}(\exp_q(Yt))}{\partial t}\right|_{t=0}, \tag{6.41}$$

where $\gamma(t)$ is a geodesic connecting $p=\gamma(0)$ and $q=\gamma(1)$.

We have already proved the following formulas for the exponential and logarithm map at the end of the preceding section

$$\begin{aligned}\exp_p(X)&=p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2}=p\exp_I\left(p^{-1}X\right)\\ \log_p(X)&=p^{1/2}\log_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2}=p\log_I\left(p^{-1}X\right).\end{aligned} \tag{6.42}$$

The above identities already specify the geodesic symmetries with the notation $S_I(X)=\exp_I(-\log_I(X))$ as

$$S_p(q)=\exp_p(-\log_p(q))=p^{1/2}S_I\left(p^{-1/2}qp^{-1/2}\right)p^{1/2}=pS_I\left(p^{-1}q\right). \tag{6.43}$$

Now we are in position to prove the following

Theorem 6.6 (M. Pálfia [56]). *Let $P(n,\mathbb{C})$ be subset of an affine symmetric space with affine geodesic symmetries given as* (6.43)*. Then the invariant affine connection has the form*

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2}\left(X_p p^{-1} Y_p + Y_p p^{-1} X_p\right), \qquad (6.44)$$

where $\kappa = S_I''(1)/2$.

Proof. We are going to use (6.41) to obtain the connection (6.44). We make the assumption that the geodesic symmetries are of the form (6.43). The functions $\exp_p(X)$ and $\log_p(X)$ are of the form (6.42), where $\exp_I(t)$ and $\log_I(t)$ are analytic functions on a disk centered around 0 and 1 respectively. We also have that $\log_I(1) = 0$, $\exp_I(0) = 1$ and furthermore

$$\left.\frac{\partial \exp_I(t)}{\partial t}\right|_{t=0} = 1. \qquad (6.45)$$

First of all we have to differentiate the map $S_p(q)$ given in (6.43) to obtain $\Gamma_1^0(\gamma)Y = T_{q\to p}Y$, where $\gamma(t)$ is a geodesic connecting $p = \gamma(0)$ and $q = \gamma(1)$.

$$\begin{aligned}\left.\frac{\partial S_p(\exp_q(Yt))}{\partial t}\right|_{t=0} &= \left.\frac{\partial p S_I(p^{-1}\exp_q(Yt))}{\partial t}\right|_{t=0} = \\ &= pDS_I\left[p^{-1}q\right]\left[p^{-1}Y\right]\end{aligned} \qquad (6.46)$$

We used the fact that $\partial \exp_q(Yt)/\partial t|_{t=0} = Y$ which is a consequence of $\exp_I'(0) = 1$.

Now we are going to differentiate the parallel transport as given by (6.41) to get back the connection. We use the holomorphic functional calculus to express the Fréchet-differential in (6.46) as

$$DS_I[X][U] = \frac{1}{2\pi i}\int_g S_I(z)[zI - X]^{-1}U[zI - X]^{-1}dz. \qquad (6.47)$$

It also easy to see that $DS_I[I][I] = S_I'(1) = -1$, so we may express the limit (6.41) by the following differential

$$\nabla_{\gamma'(0)} Y_{\gamma(0)} = -\left.\frac{\partial \gamma(t/2) DS_I\left[\gamma(t/2)^{-1}\gamma(t)\right]\left[\gamma(t/2)^{-1}Y_{\gamma(t)}\right]}{\partial t}\right|_{t=0} = \qquad (6.48)$$

we massage this further by using the holomorphic functional calculus

$$= -\frac{\partial}{\partial t}\gamma(t/2)\frac{1}{2\pi i}\int_g S_I(z)[zI-\gamma(t/2)^{-1}\gamma(t)]^{-1}\gamma(t/2)^{-1}Y_{\gamma(t)}$$
$$[zI-\gamma(t/2)^{-1}\gamma(t)]^{-1}dz\big|_{t=0} = -\frac{1}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}DS_I[I][I]-$$
$$-\gamma(0)\frac{1}{2\pi i}\int_g S_I(z)\left\{[zI-I]^{-1}\frac{1}{2}\gamma(0)^{-1}\gamma'(0)[zI-I]^{-1}\gamma(0)^{-1}Y_{\gamma(0)}[zI-I]^{-1}+\right.$$
$$+[zI-I]^{-1}\gamma(0)^{-1}Y_{\gamma(0)}[zI-I]^{-1}\frac{1}{2}\gamma(0)^{-1}\gamma'(0)[zI-I]^{-1}+$$
$$+[zI-I]^{-1}\left[-\gamma(0)^{-1}\frac{1}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}+\gamma(0)^{-1}DY[\gamma(0)][\gamma'(0)]\right]$$
$$\left.[zI-I]^{-1}\right\}dz = \tag{6.49}$$

by using the fact that $DS_I[I][I]$ and $[zI-I]^{-1}$ commutes with every matrix we get

$$= -\frac{DS_I[I][I]}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}-$$
$$-\gamma(0)\frac{1}{2\pi i}\int_g \frac{S_I(z)dz}{(z-1)^3}\frac{1}{2}\gamma(0)^{-1}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}-$$
$$-\gamma(0)\frac{1}{2\pi i}\int_g \frac{S_I(z)dz}{(z-1)^3}\frac{1}{2}\gamma(0)^{-1}Y_{\gamma(0)}\gamma'(0)\gamma(0)^{-1}-$$
$$-\gamma(0)\frac{1}{2\pi i}\int_g \frac{S_I(z)dz}{(z-1)^2}\left[-\frac{1}{2}\gamma(0)^{-1}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}+\gamma(0)^{-1}DY[\gamma(0)][\gamma'(0)]\right] \tag{6.50}$$

at this point we are going to use the integral representation

$$S_I^{(j)}(1) = \frac{j!}{2\pi i}\int_g \frac{S_I(z)}{(z-1)^{j+1}}dz$$

to further simplify the above.

$$\nabla_{\gamma'(0)}Y_{\gamma(0)} = -\frac{S_I''(1)}{4}\left[\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}+Y_{\gamma(0)}\gamma(0)^{-1}\gamma'(0)\right]-$$
$$-\frac{S_I'(1)}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}-\frac{S_I'(1)}{2}\left[-\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}+2DY[\gamma(0)][\gamma'(0)]\right]=$$
$$= -S_I'(1)DY[\gamma(0)][\gamma'(0)]-\frac{S_I''(1)}{4}\left[\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}+Y_{\gamma(0)}\gamma(0)^{-1}\gamma'(0)\right]. \tag{6.51}$$

So we have that $\kappa = S_I''(1)/2$.

□

The above clearly tells us that all symmetric spaces occurring in such a way that their midpoint operation is a matrix mean, have invariant affine connections in the form (6.44). We are going to study these connections as κ being a parameter. We will find out in the next section for which values of κ are these spaces symmetric. Also for arbitrary real κ (6.44) defines an affine connection with corresponding exponential and logarithm map which are of the form (6.42). This fact follows from considering the geodesic equations for the curves $\gamma_1(t) = \exp_I\left(p^{-1/2}Xp^{-1/2}t\right)$ and $\gamma_2(t) = \exp_p(Xt) = p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}t\right)p^{1/2}$. We will also determine if these connections are metric or not.

6.4 Properties of These Affine Connections

Our next step is to integrate the geodesic equations corresponding to the one parameter family of connections (6.44).

Theorem 6.7 (M. Pálfia [56]). *The geodesic equation corresponding to the affine connection* (6.44) *is*

$$\ddot{\gamma} = \kappa\dot{\gamma}\gamma^{-1}\dot{\gamma}. \tag{6.52}$$

The solution of this equation is the following one parameter family of functions

$$\exp_I(X) = \begin{cases} [(1-\kappa)X+1]^{\frac{1}{1-\kappa}} & \textit{if } \kappa \neq 1, \\ \exp(X) & \textit{else.} \end{cases} \tag{6.53}$$

Proof. First of all note that by (6.42) it is enough to solve the equation (6.52) for real numbers. Therefore the equation takes the form

$$\exp_I''(t) = \kappa\exp_I'(t)^2\exp_I(t)^{-1}. \tag{6.54}$$

If we transform the equation to the inverse function of $\exp_I(t)$ which will be the logarithm map $\log_I(t)$, then we get a separable first order differential equation of the form

$$\log_I''(t) = -\kappa\log_I'(t)t^{-1}. \tag{6.55}$$

Solving the above we get the logarithm map as

$$\log_I(X) = \begin{cases} \frac{X^{1-\kappa}-1}{1-\kappa} & \text{if } \kappa \neq 1, \\ \log(X) & \text{else.} \end{cases} \tag{6.56}$$

From this by inverting the above function we get the assertion.

□

Since we have integrated the geodesic equations we can get back the means which induce these affinely connected manifolds using (6.42)

$$\begin{aligned} M(X,Y) &= \exp_X\left(\frac{1}{2}\log_X(Y)\right) = \\ &= \begin{cases} X^{1/2}\left[\frac{I+\left(X^{-1/2}YX^{-1/2}\right)^{1-\kappa}}{2}\right]^{\frac{1}{1-\kappa}}X^{1/2} & \text{if } \kappa \neq 1, \\ X^{1/2}\left(X^{-1/2}YX^{-1/2}\right)^{1/2}X^{1/2} & \text{else.} \end{cases} \end{aligned} \tag{6.57}$$

The above functions are matrix means if $\kappa \in [0,2]$, see exercise 4.5.11 [11]. For other values of κ the corresponding functions fail to be operator monotone, however they still may be considered as means from a geometrical point of view.

If $\kappa = 0$ we get back the arithmetic mean as the midpoint operation, and the weighted arithmetic mean

$$A_t(A,B) = (1-t)A + tB \tag{6.58}$$

is the geodesic line connecting A and B with respect to the metric $\langle X,Y\rangle_p = Tr\{X^*Y\}$. If $\kappa = 2$ we get back the harmonic mean as the midpoint operation, and the weighted harmonic mean

$$H_t(A,B) = \left((1-t)A^{-1} + tB^{-1}\right)^{-1} \tag{6.59}$$

is also a geodesic with respect to the metric $\langle X,Y\rangle_p = Tr\left\{p^{-2}Xp^{-2}Y\right\}$. We have already mentioned that the second metric is isometric to the first one, so it is also Euclidean.

In the case when $\kappa = 1$ the midpoint is the geometric mean and the geodesics are given by the weighted geometric mean

$$G_t(A,B) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2}. \tag{6.60}$$

The corresponding Riemannian metric is $\langle X,Y\rangle_p = Tr\left\{p^{-1}Xp^{-1}Y\right\}$. This manifold, which is the symmetric space $GL(n,\mathbb{C})/U(n,\mathbb{C})$, satisfies the semi-parallelogram law (see Section 4.1), so is nonpositively curved while the other two has zero curvature.

So we already know that in the case of arithmetic, geometric and harmonic mean ($\kappa = 0,1,2$ respectively) we have a corresponding Riemannian metric. These metrics are of fundamental importance in the theory of matrix means as we have seen so far. Since all of the manifolds of this one-parameter family are analytic, we can omit the study of holonomy groups and study the problem directly using power series expansions as in [21]. It is also easy to see that these connections are symmetric and torsion free so all of them can possibly be a Levi-Civita connection of a Riemannian manifold.

Let W be an analytic manifold with a symmetric affine connection. Let R^i_{jkl} denote its Riemann curvature tensor with respect to a coordinate frame. Then W admits a Riemannian metric g_{ij} if and only if every solution g_{ij} of the following system of equations

$$g_{sl}R^s_{ikl} + g_{is}R^s_{jkl} = 0 \tag{6.61}$$

also satisfies the system of equations

$$g_{sl}R^s_{ikl;m} + g_{is}R^s_{jkl;m} = 0, \tag{6.62}$$

here we use the Einstein summation convention for repeated indices and the semicolon ; to denote the covariant differentiation with repsect to the index which follows the semicolon. The above theorem can be found in [75] as Theorem 1.3. Similarly one may also consult the classical paper [21].

In our case it turns out that

$$\begin{aligned}\Gamma^i_{jk}E_i =& -\frac{\kappa}{2}(E_jp^{-1}E_k + E_kp^{-1}E_j)\\ R^i_{jkl}E_i =& \left(\frac{\kappa}{2}-\frac{\kappa^2}{4}\right)\left(E_jp^{-1}E_kp^{-1}E_l + E_lp^{-1}E_kp^{-1}E_j - \right.\\ & \left. -E_jp^{-1}E_lp^{-1}E_k - E_kp^{-1}E_lp^{-1}E_j\right),\end{aligned} \tag{6.63}$$

where the E_i form the standard basis of the vector space of hermitian matrices. In order to determine which of these manifolds are symmetric spaces it is sufficient to calculate the covariant differential $R^s_{jkl;m}$, since it vanishes everywhere if and only if the underlying manifold is a symmetric space. So we have by definition

$$\begin{aligned}R^i_{jkl;m}E_i =& \frac{\partial R^i_{jkl}}{\partial x^m}E_i + \Gamma^i_{nm}R^n_{jkl}E_i - \\ & - \Gamma^n_{jm}R^i_{nkl}E_i - \Gamma^n_{km}R^i_{jnl}E_i - \Gamma^n_{lm}R^i_{jkm}E_i.\end{aligned} \tag{6.64}$$

After some calculations one gets the following formulas

$$\begin{aligned}&\frac{\partial R^i_{jkl}}{\partial x^m}E_i = \\ &= \left(\frac{\kappa}{2}-\frac{\kappa^2}{4}\right)\left(-E_jp^{-1}E_mp^{-1}E_kp^{-1}E_l - E_jp^{-1}E_kp^{-1}E_mp^{-1}E_l + \right.\\ &+ E_jp^{-1}E_mp^{-1}E_lp^{-1}E_k + E_jp^{-1}E_lp^{-1}E_mp^{-1}E_k - \\ &- E_lp^{-1}E_mp^{-1}E_kp^{-1}E_j - E_lp^{-1}E_kp^{-1}E_mp^{-1}E_j + \\ &\left. + E_kp^{-1}E_mp^{-1}E_lp^{-1}E_j + E_kp^{-1}E_lp^{-1}E_mp^{-1}E_j\right)\\ &\Gamma^i_{nm}R^n_{jkl}E_i = \\ &= -\frac{\kappa}{2}\left(\frac{\kappa}{2}-\frac{\kappa^2}{4}\right)\left(E_jp^{-1}E_kp^{-1}E_lp^{-1}E_m - E_jp^{-1}E_lp^{-1}E_kp^{-1}E_m + \right.\\ &+ E_lp^{-1}E_kp^{-1}E_jp^{-1}E_m - E_kp^{-1}E_lp^{-1}E_jp^{-1}E_m + \\ &+ E_mp^{-1}E_jp^{-1}E_kp^{-1}E_l - E_mp^{-1}E_jp^{-1}E_lp^{-1}E_k + \\ &\left. + E_mp^{-1}E_lp^{-1}E_kp^{-1}E_j - E_mp^{-1}E_kp^{-1}E_lp^{-1}E_j\right).\end{aligned} \tag{6.65}$$

It is possible to check using the above that $R^s_{jkl;m} = 0$ everywhere if and only if $\kappa = 0, 1, 2$. This proves the following

Proposition 6.8 (M. Pálfia [56])**.** *The only symmetric matrix means which are affine means corresponding to symmetric spaces are the arithmetic, harmonic and geometric means.*

Now we turn to the metrization problem. First of all we compute the parallel transport map over a geodesic connecting an arbitrary point and the identity.

Proposition 6.9 (M. Pálfia [56]). *Let $c(t)$ be a geodesic with respect to the connection* (6.44) *and $c(0) = I, c(1) = p$. Then the unique solution of $\nabla_{\dot{c}(t)} Y = 0$ with respect to the connection* (6.44) *and the initial condition $Y_{c(0)} = Y_0$ is the vector field*

$$Y(t) = c(t)^{\frac{\kappa}{2}} Y_0 c(t)^{\frac{\kappa}{2}}. \tag{6.66}$$

Proof. We have to integrate the equation $\nabla_{c'(t)} Y_{c(t)} = 0$. This is equivalent to

$$DY[c(t)][c'(t)] - \frac{\kappa}{2}\left(c'(t)c(t)^{-1}Y_{c(t)} + Y_{c(t)}c(t)^{-1}c'(t)\right) = 0. \tag{6.67}$$

We will be looking for the solution $Y_{c(t)} = Y(t)$ in the form

$$Y(t) = f(c(t))Y_0 f(c(t)), \tag{6.68}$$

for some analytic function $f(x)$. We have for the Fréchet-differential

$$DY[c(t)][c'(t)] = \frac{dY(t)}{dt} = \frac{df(c(t))}{dt} Y_0 f(c(t)) + f(c(t))Y_0 \frac{df(c(t))}{dt}. \tag{6.69}$$

Now substituting into the equation of the parallel transport above, we get

$$\frac{df(c(t))}{dt} Y_0 f(c(t)) + f(c(t))Y_0 \frac{df(c(t))}{dt} = \frac{\kappa}{2}\left(c'(t)c(t)^{-1}Y_{c(t)} + Y_{c(t)}c(t)^{-1}c'(t)\right). \tag{6.70}$$

Since $c(t) = \exp_I(t \log_I(p))$, it has a power series expansion, as has $f(x)$, so we have by commutativity that

$$\frac{\kappa}{2} c'(t)c(t)^{-1} f(c(t)) = \frac{df(c(t))}{dt} = Df[c(t)][c'(t)] = f'(c)c'(t). \tag{6.71}$$

Since everything on the left and right hand side commutes with each other, we arrive at the following separable differential equation

$$\frac{\kappa}{2} c^{-1} = f'(c) f(c)^{-1}, \tag{6.72}$$

which has its solution in the form $f(c) = c^{\kappa/2}$.

□

By the above proposition we should have the Riemannian metric in the form

$$\left\langle p^{-\kappa/2} X p^{-\kappa/2}, p^{-\kappa/2} Y p^{-\kappa/2} \right\rangle \tag{6.73}$$

for some positive definite bilinear form $\langle \cdot, \cdot \rangle$ given on the tangent space at I. Due to the properties of R^i_{jkl} we conclude that a trivial solution of (6.61) is the mapping $Tr\{XY\}$ at I, since R^i_{jkl} is the same for all κ except for a constant multiple and for $\kappa = 1$ we have the connection of $GL(n, \mathbb{C})/U(n, \mathbb{C})$, for which we have the solution $Tr\{XY\}$ at I. But it is easy to see that $Tr\{XY\}$ is not a solution of (6.62) at I if $\kappa \neq 0, 1, 2$. Similarly $Tr\left\{p^{-1}Xp^{-1}Y\right\}$ is a solution of (6.61) at arbitrary p for $\kappa \neq 0, 2$, but it is not a solution of (6.62) if $\kappa \neq 1$. So we conclude

Proposition 6.10 (M. Pálfia [56]). *The smooth manifolds with affine connection* (6.44) *do not carry a Riemannian metric unless* $\kappa = 0, 1, 2$.

So remarkably we have not found any previously unknown matrix mean so far which is the midpoint map on a Riemannian manifold, although we have already found a previously unknown, generally non-metrizable, one parameter family of affinely connected manifolds where the midpoint operations are symmetric matrix means. Due to the above one would expect that these Riemannian manifolds are sparse. Actually in the next section we show that this one parameter family of connections is exhaustive, there exists no other affinely connected manifold where the midpoint map is a symmetric matrix mean.

6.5 Classification of Affine Matrix Means

Due to Proposition 6.5 we have the exponential and logarithm map in the form

$$\begin{aligned} \exp_p(X) &= p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} \\ \log_p(X) &= p^{1/2}\log_I\left(p^{-1/2}Xp^{-1/2}\right)p^{1/2} \end{aligned} \tag{6.74}$$

for $p \in P(n, \mathbb{C})$, where $\exp_I(X)$ and $\log_I(X)$ are analytic functions. The function $\exp_I : H(n, \mathbb{C}) \mapsto P(n, \mathbb{C})$ and $\log_I(X)$ is its inverse, $\log_I'(I) = I, \exp_I'(0) = I, \log_I(I) = 0, \exp_I(0) = I$. Suppose that (6.74) represent the exponential and logarithm map of an affinely connected manifold. Then the analytic function $\exp_I(t)$ is the solution of some geodesic equations

$$\exp_I''(t) + \Gamma\left(\exp_I'(t), \exp_I'(t), \exp_I(t)\right) = 0, \tag{6.75}$$

where $\Gamma(\cdot, \cdot, \cdot) : H(n, \mathbb{C}) \times H(n, \mathbb{C}) \times P(n, \mathbb{C}) \mapsto H(n, \mathbb{C})$ is a smooth function in all variables and linear in the first two. By Propostion 15 and Corollary 16 of Chapter 6 in [73] we know that connections which have the same torsion and geodesics are identical and for an arbitrary connection there is a unique connection with vanishing torsion and with the same geodesics. If we have an affine connection with non-symmetric Christoffel symbols Γ^i_{jk}, it has the same geodesics as its symmetric part $\frac{\Gamma^i_{jk}+\Gamma^i_{kj}}{2}$, so without loss of generality we may assume in our case that all connections are symmetric, so we will be considering mappings $\Gamma(\cdot, \cdot, \cdot)$ which are symmetric in their first two arguments.

Proposition 6.11 (M. Pálfia [56]). *Suppose that* $\Gamma(\cdot, \cdot, \cdot), \exp_I(\cdot), \exp_p(\cdot)$ *are functions given with the above properties. Then*

$$\Gamma(X, X, p) = p^{1/2}\Gamma\left(p^{-1/2}Xp^{-1/2}, p^{-1/2}Xp^{-1/2}, I\right)p^{1/2} \tag{6.76}$$

for $p \in P(n, \mathbb{C})$ *and* $X \in H(n, \mathbb{C})$.

Proof. Let $\gamma(t) = \exp_I\left(p^{-1/2}Xp^{-1/2}t\right)$. Since $\exp_I$ is an analytic function we have

$$\begin{aligned} \gamma'(t) &= p^{-1/2}Xp^{-1/2}\exp_I'\left(p^{-1/2}Xp^{-1/2}t\right) \\ \gamma''(t) &= p^{-1/2}Xp^{-1/2}\exp_I''\left(p^{-1/2}Xp^{-1/2}t\right)p^{-1/2}Xp^{-1/2}. \end{aligned} \tag{6.77}$$

By the geodesic equations we have

$$\begin{aligned}\gamma''(t) &= -\Gamma\left(\gamma'(t),\gamma'(t),\gamma(t)\right)\\ \exp_I''\left(p^{-1/2}Xp^{-1/2}t\right) &= -p^{1/2}X^{-1}p^{1/2}\Gamma\left(p^{-1/2}Xp^{-1/2}\exp_I'\left(p^{-1/2}Xp^{-1/2}t\right),\right.\\ &\left. p^{-1/2}Xp^{-1/2}\exp_I'\left(p^{-1/2}Xp^{-1/2}t\right),\exp_I\left(p^{-1/2}Xp^{-1/2}t\right)\right)p^{1/2}X^{-1}p^{1/2}.\end{aligned} \tag{6.78}$$

If we consider the geodesic equations for $\gamma(t) = \exp_p(Xt)$ we get

$$\begin{aligned}\exp_I''\left(p^{-1/2}Xp^{-1/2}t\right) &= -p^{1/2}X^{-1}\Gamma\left(Xp^{-1/2}\exp_I'\left(p^{-1/2}Xp^{-1/2}t\right)p^{1/2},\right.\\ &\left. p^{1/2}\exp_I'\left(p^{-1/2}Xp^{-1/2}t\right)p^{-1/2}X, p^{1/2}\exp_I\left(p^{-1/2}Xp^{-1/2}t\right)p^{1/2}\right)X^{-1}p^{1/2}.\end{aligned} \tag{6.79}$$

The left hand sides of the two equations above are the same so as the right hand sides. Taking $t = 0$ and that $\exp_I'(0) = I, \exp_I(0) = I$ we get for all $p \in P(n,\mathbb{C}), X \in H(n,\mathbb{C})$ that

$$\begin{aligned}p^{1/2}X^{-1}p^{1/2}\Gamma\left(p^{-1/2}Xp^{-1/2}, p^{-1/2}Xp^{-1/2}, I\right)p^{1/2}X^{-1}p^{1/2} &=\\ = p^{1/2}X^{-1}\Gamma\left(X,X,p\right)X^{-1}p^{1/2},&\end{aligned} \tag{6.80}$$

which proves the assertion.

□

By the above result we have just reduced the problem of characterizing $\Gamma(X,X,p)$ to the characterization of $\Gamma(X,X,I)$. Now we will show that $\Gamma(X,X,p)$ is invariant under similarity transformations.

Proposition 6.12 (M. Pálfia [56])**.** *For all $p \in P(n,\mathbb{C})$ and $X \in H(n,\mathbb{C})$ and invertible S we have*

$$\Gamma\left(SXS^{-1}, SXS^{-1}, SpS^{-1}\right) = S\Gamma\left(X,X,p\right)S^{-1}. \tag{6.81}$$

Proof. We have by the geodesic equations

$$\begin{aligned}X^2\exp_I''(Xt) &= -\Gamma\left(X\exp_I'(Xt), X\exp_I'(Xt), \exp_I(Xt)\right)\\ SX^2\exp_I''(Xt)S^{-1} &= -S\Gamma\left(X\exp_I'(Xt), X\exp_I'(Xt), \exp_I(Xt)\right)S^{-1}.\end{aligned} \tag{6.82}$$

Similarly if we consider the geodesic equations for the curve $\exp_I\left(SXS^{-1}t\right)$ we get

$$\begin{aligned}SX^2S^{-1}\exp_I''(SXS^{-1}t) &= -\Gamma\left(SXS^{-1}\exp_I'(SXS^{-1}t), SXS^{-1}\exp_I'(SXS^{-1}t),\right.\\ &\left.\exp_I(SXS^{-1}t)\right)\\ SX^2\exp_I''(Xt)S^{-1} &= -\Gamma\left(SX\exp_I'(Xt)S^{-1}, SX\exp_I'(Xt)S^{-1},\right.\\ &\left. S\exp_I(Xt)S^{-1}\right).\end{aligned} \tag{6.83}$$

Again since the above two equations are identical we get the assertion.

□

By the above proposition we have for hermitian X that

$$\Gamma\left(X,X,I\right)=U\Gamma\left(D,D,I\right)U^{*}, \tag{6.84}$$

for some diagonal D and unitary U, so it is enough to characterize $\Gamma\left(X,X,I\right)$ for diagonal X.

Theorem 6.13 (M. Pálfia [56])**.** *Let D be diagonal with real coefficients. Then*

$$\Gamma\left(D,D,I\right)=-cD^{2}, \tag{6.85}$$

for some real valued constant c.

Proof. First we will show that $\Gamma\left(I,I,I\right)=cI$ for some real constant c. Consider the case when $\gamma(t)=\exp_I(\lambda It)$ for some real λ. Then by the geodesic equations for $\gamma(t)$ we have

$$\lambda^2\exp_I''(\lambda It)=-\Gamma\left(\lambda\exp_I'(\lambda It),\lambda\exp_I'(\lambda It),\exp_I(\lambda It)\right). \tag{6.86}$$

By linearity of $\Gamma(\cdot,\cdot,\cdot)$ in the first two variables, this is equivalent to

$$\lambda^2\exp_I''(\lambda It)=-\lambda^2\Gamma\left(\exp_I'(\lambda It),\exp_I'(\lambda It),\exp_I(\lambda It)\right). \tag{6.87}$$

Letting $t=0$ we get

$$cI=-\Gamma\left(I,I,I\right), \tag{6.88}$$

where $c=\exp_I''(0)$ is a real number, since $\exp_I:H(n,\mathbb{C})\mapsto P(n,\mathbb{C})$ is an analytic function with real coefficients in its Taylor series.

The next step is to show that for a projection $P=P^2=P^*$ we have $\Gamma\left(P,P,I\right)=-cP$. Consider again $\gamma(t)=\exp_I(Pt)$. Then the geodesic equations read

$$P^2\exp_I''(Pt)=-\Gamma\left(P\exp_I'(Pt),P\exp_I'(Pt),\exp_I(Pt)\right). \tag{6.89}$$

Since $P^2=P$ and again letting $t=0$ we get

$$cP=-\Gamma\left(P,P,I\right), \tag{6.90}$$

where c is trivially the same constant as determined above for $\Gamma\left(I,I,I\right)$. Now suppose that we have two mutually orthogonal projections P_1,P_2 such that $P_1P_2=0$. Then we have for the projection P_1+P_2 using linearity of $\Gamma(\cdot,\cdot,\cdot)$ in the first two variables that

$$\begin{aligned}&\Gamma\left(P_1,P_1,I\right)+\Gamma\left(P_2,P_2,I\right)=-c(P_1+P_2)=\Gamma\left(P_1+P_2,P_1+P_2,I\right)=\\&=\Gamma\left(P_1,P_1,I\right)+\Gamma\left(P_1,P_2,I\right)+\Gamma\left(P_2,P_1,I\right)+\Gamma\left(P_2,P_2,I\right),\end{aligned} \tag{6.91}$$

which yields that for mutually orthogonal projections P_1,P_2 we get the orthogonality relation

$$\Gamma\left(P_1,P_2,I\right)=0. \tag{6.92}$$

Finally since a diagonal D can be written as $D = \sum_i \lambda_i P_i$ for mutually orthogonal projections P_i, we have

$$\begin{aligned}\Gamma(D,D,I) &= \Gamma\left(\sum_i \lambda_i P_i, \sum_i \lambda_i P_i, I\right) = \\ &= \sum_i \lambda_i^2 \Gamma(P_i,P_i,I) = -\sum_i \lambda_i^2 c P_i = \\ &= -cD^2,\end{aligned} \tag{6.93}$$

which is what needed to be shown.

□

The above three theorems with the other preceding results presented here, lead us to the concluding

Theorem 6.14 (M. Pálfia [56]). *All affine matrix means $M(X,Y)$ are of the form*

$$M(X,Y) = \begin{cases} X^{1/2}\left[\frac{I+\left(X^{-1/2}YX^{-1/2}\right)^{1-\kappa}}{2}\right]^{\frac{1}{1-\kappa}} X^{1/2} & \text{if } \kappa \neq 1, \\ X^{1/2}\left(X^{-1/2}YX^{-1/2}\right)^{1/2} X^{1/2} & \text{if } \kappa = 1, \end{cases} \tag{6.94}$$

where $0 \leq \kappa \leq 2$. The symmetric affine connections corresponding to these means are

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2}\left(X_p p^{-1} Y_p + Y_p p^{-1} X_p\right). \tag{6.95}$$

Proof. By Proposition 6.11, 6.12 and Theorem 6.13 we have that the functions $\Gamma(\cdot,\cdot,\cdot) : H(n,\mathbb{C}) \times H(n,\mathbb{C}) \times P(n,\mathbb{C}) \mapsto H(n,\mathbb{C})$ representing the Christoffel symbols are of the form

$$\Gamma(X,X,p) = -cXp^{-1}X. \tag{6.96}$$

This formula determines the functions that are the symmetric parts of the possible connections, and these connections have geodesics determined by Theorem 6.7 in the form (6.94). Again by Proposition 15 and Corollary 16 of Chapter 6 in [73] we know that connections which have the same torsion and geodesics are identical and for an arbitrary connection there is a unique connection with vanishing torsion and with the same geodesics. So in other words since the connections (6.95) are symmetric, affine and have the same geodesics, therefore they give *the* sought symmetric connections for each κ if we choose $c = \kappa$.

The corresponding midpoint operations have the form (6.57), and these are matrix means if and only if $\kappa \in [0,2]$, since the representing functions $f(t)$ in (6.17) turn out to be operator monotone only in these cases (see again exercise 4.5.11 [11]). This gives us the complete classification of affine matrix means. □

It turned out so far that only 3 symmetric matrix means are midpoint operations on a Riemannian manifold. Even more, only a one real parameter family of symmetric matrix means are midpoint operations on affinely connected manifolds. These cases although do not cover the possible Finslerian manifolds, or other metric spaces, however the sparsness of them suggests us that there should be only slight hope for finding other such families of symmetric matrix means. Even in this one real parameter family it is not hard to see that they cannot be midpoints of a Finslerian space, since in that case the space should be Riemannian, which is generally not the case, except the cases of the arithmetic, geometric and harmonic means.

The above might be a disappointing result, since we cannot use the metric space machinery of the previous section in general, however we will see in the following sections, that we do not have to. We will prove that the Iterative process the ALM-process and even the BMP-process converges for all symmetric matrix means. The BMP-process requires a weighted counterpart for each symmetric matrix mean. At the moment, these are only given for affine matrix means as geodesic lines, but we will show that it is possible to define a weighted counterpart for every symmetric matrix mean without affine geodesy. Essentially the next section will be about how to mimic these geometric structures and proofs, or in other words: how to do geometry without geometric structures.

7 Extensions of Matrix Means without Metric Structures

We have seen so far how well applicable are these metric structures. But this purely geometrical framework works only in the case of the geometric mean, the other two known cases are not of deep interest, the arithmetic and the harmonic means are trivially extendable to several variables. At this point our problem is that we cannot use the metric geometric machinery for other symmetric matrix means to extend them to multiple variables. Clearly we cannot talk about the center of mass, since without metric structures, it does not exist. However we will be able to use somehow the Euclidean structure of $P(n, \mathbb{C})$ which corresponds to the arithmetic mean. We will combine it with the partial order of $P(n, \mathbb{C})$, the positive definite order. First we will consider the Iterative mean process given in Definition 5.3 for all symmetric matrix means and prove its convergence in general.

7.1 Iterative Mean for all Matrix Means

We will be considering not necessarily symmetric means, so we have to generalize the procedure given in Definition 5.3. The following algorithm is by M. Pálfia [54].

The next section is devoted to the proof of the assertion that the sequences $(X_i^k)_{k\geq 0}$, $i = 1, \dots, n$, generated by Algorithm 2 are convergent and have the same limit point. Note that this limit point depends on the data X and M,

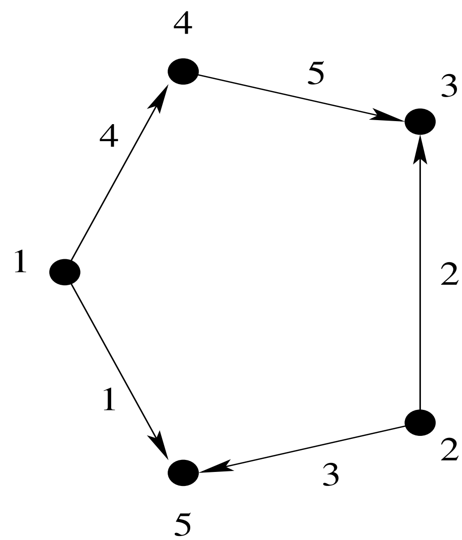

Figure 5: Example of graph G^k for step 4 of Algorithm 2. If this graph is chosen in step 4, then step 6 yields $X_1^{k+1} = M(X_1^k, X_5^k)$, $X_2^{k+1} = M(X_2^k, X_3^k)$, etc.

Algorithm 2 Iterative extension of a 2-variable matrix mean

1: Data: $X = (X_1, \dots, X_n) \in P(r, \mathbb{C})^n$; a 2-variable matrix mean $M : P(r, \mathbb{C}) \times P(r, \mathbb{C}) \to P(r, \mathbb{C})$.
2: Initialization: $X_i^0 := X_i$, $i = 1, \dots, n$.
3: **for** $k = 0, 1, \dots$ **do**
4:     Choose a directed graph G^k with n vertices labelled from 1 to n and n edges labelled from 1 to n, such that G^k is connected as an undirected graph and every vertex has exactly two edges connected to it; see figure 5 for an example.
5:     **for** $i = 1, \dots, n$ **do**
6:         $X_i^{k+1} := M(X_{j_i}^k, X_{l_i}^k)$, where j_i is the tail vertex and l_i the head vertex of edge i in G^k.
7:     **end for**
8: **end for**

---

and also, in general, on the graphs G^k chosen in step 4 of Algorithm 2. In order to emphasize this dependence, we will later on denote this limit point by $M_G(X_1, \ldots, X_n)$, where G stands for $(G^k)_{k\geq 0}$.

7.1.1 Proof of the convergence of the procedure

In this section we will prove the convergence of the sequences of matrices given in Algorithm 2 to a common limit point. We begin with showing the boundedness of sequences in Algorithm 2, then we prove an inequality which is similar to the semi-parallelogram law at a certain point in the set $P(r, \mathbb{C})$. After these preparations we move on to the main proof. At the end of the section, some convergence related questions of the limit point of the sequences in Algorithm 2 are covered as well.

Lemma 7.1 (M. Pálfia [54])**.** *The sequences given in Algorithm 2 are bounded for all n.*

Proof. Let D_i^0 be matrices such that $D_i^0 \geq X_i^0$ and $D_1^0 \leq D_2^0, \ldots \leq D_n^0$. Set up the iteration given in Algorithm 2 on the D_i^0 matrices with the same infinite sequence of graphs as given for the X_i^k matrices. Considering property (3.2) and the monotonicity property, it is easy to see that $X_i^k \leq D_i^k$ and $D_i^k \leq D_n^0$ for all i and k. Therefore every X_i^k is bounded above by D_n^0 for all i and k. In the same way we may construct the lower bound as well. □

We advance further by showing that a "semi-parallelogram law"

$$d(X, M(A,B))^2 \leq \frac{d(X,A)^2 + d(X,B)^2}{2} - \frac{1}{4}d(A,B)^2 \tag{7.1}$$

holds for $X = 0$ for the distance function defined by

$$d(A,B)^2 = Tr\left\{(A-B)^*(A-B)\right\}, \tag{7.2}$$

which is induced by the Hilbert-Schmidt inner product.

First of all, note that we are in the set $P(r, \mathbb{C})$. Each element of this set of finite norm is of finite distance from the 0 matrix measured with the above distance function.

Lemma 7.2 (M. Pálfia [54])**.** *For the distance function* (7.2) *and for any matrix mean function $M(A,B) \leq \frac{A+B}{2}$ the following holds*

$$d(0, M(A,B))^2 \leq \frac{d(0,A)^2 + d(0,B)^2}{2} - \frac{1}{4}d(A,B)^2. \tag{7.3}$$

Proof. We will show that the above inequality can be reduced to an easier one, which can be easily proved to be true. So using the distance function (7.2), the above equation is equivalent to

$$Tr\left\{M(A,B)^*M(A,B)\right\} \leq \frac{Tr\left\{A^*A\right\} + Tr\left\{B^*B\right\}}{2} - \\ -\frac{Tr\left\{A^*A\right\} + Tr\left\{B^*B\right\} - Tr\left\{A^*B\right\} - Tr\left\{B^*A\right\}}{4}.$$

Noticing the fact that all of the matrices are hermitian we get

$$Tr\left\{M(A,B)^2\right\} \le \frac{Tr\left\{A^2\right\}+Tr\left\{B^2\right\}+Tr\left\{AB+BA\right\}}{4} = Tr\left\{\left(\frac{A+B}{2}\right)^2\right\}$$

$$0 \le Tr\left\{\left(\frac{A+B}{2}\right)^2 - M(A,B)^2\right\}$$

$$0 \le Tr\left\{\left(\frac{A+B}{2} - M(A,B)\right)\left(\frac{A+B}{2} + M(A,B)\right)\right\} \tag{7.4}$$

Thus we see that (7.4) is equivalent with the inequality of the assertion.

From the condition $\frac{A+B}{2} \ge M(A,B)$ it is easy to see that $\frac{A+B}{2} - M(A,B)$ and $\frac{A+B}{2} + M(A,B)$ are both positive definite as well, so

$$0 \le \left(\frac{A+B}{2} + M(A,B)\right)^{1/2}\left(\frac{A+B}{2} - M(A,B)\right)\left(\frac{A+B}{2} + M(A,B)\right)^{1/2}.$$

This yields (7.4), which proves the assertion. □

We will also need the following preparatory lemma, which involves the arithmetic mean.

Lemma 7.3 (M. Pálfia [54])**.** *Let X_i^k be sequences given in Algorithm 2 with $M(A,B) \le \frac{A+B}{2}$, then*

$$\frac{\sum_{i=1}^n X_i^{k+1}}{n} \le \frac{\sum_{i=1}^n X_i^k}{n}. \tag{7.5}$$

Proof. In one iteration step we have the following for every i

$$X_i^{k+1} = M\left(X_{j_i}^k, X_{l_i}^k\right) \le \frac{X_{j_i}^k + X_{l_i}^k}{2}, \tag{7.6}$$

It must be noted that every X_l^k appears twice when the X_i^{k+1}-s are computed since every vertex is one of the ending points of exactly two distinct edges in the graph G^k. So summing up these equations for every i we arrive at (7.5). □

Now we are ready to prove the main theorem of this section.

Theorem 7.4 (M. Pálfia [54])**.** *Let the matrix mean M in Algorithm 2 satisfy $M(A,B) \le \frac{A+B}{2}$ for all $A,B \in P(r,\mathbb{C})$. Then the n sequences $(X_i^k)_{k\ge 0}$, $i = 1,\dots,n$, generated by Algorithm 2 converge and have the same limit point.*

Proof. We begin with showing that the distances $d(X_i^k, X_j^k)$ are converging to zero, where $d(\cdot,\cdot)$ is defined by (7.2). Later on we will show that the X_i^k sequences are themselves convergent.

Let us consider one iteration step in Algorithm 2, which actually maps pairs of X_i^k to a X_l^{k+1} through a graph by taking the mean $M(X_i^k, X_j^k)$ of the matrices

X_i^k and X_j^k corresponding to the two vertices of an edge. From Lemma 7.2 we get

$$d\left(0, X_i^1\right)^2 \leq \frac{d\left(0, X_{j_i}\right)^2 + d\left(0, X_{l_i}\right)^2}{2} - \frac{1}{4} d\left(X_{j_i}, X_{l_i}\right)^2, \tag{7.7}$$

where $X_i^1 = M(X_{j_i}, X_{l_i})$. Each vertex of the graph G^k (step 4 of Algorithm 2) has exactly two edges connected to it. So if we sum up the equations above for every edge we arrive at

$$\sum_{i=1}^{n} d\left(0, X_i^1\right)^2 \leq \sum_{i=1}^{n} d\left(0, X_i\right)^2 - \frac{1}{4} \sum_{i=1}^{n} d\left(X_{j_i}, X_{l_i}\right)^2. \tag{7.8}$$

Applying this to every iteration step we get

$$\underbrace{\sum_{i=1}^{n} d\left(0, X_i^{k+1}\right)^2}_{a_{k+1}} \leq \underbrace{\sum_{i=1}^{n} d\left(0, X_i^k\right)^2}_{a_k} - \frac{1}{4} \underbrace{\sum_{i=1}^{n} d\left(X_{j_i}^k, X_{l_i}^k\right)^2}_{e_k}. \tag{7.9}$$

Note that the above is valid for every possible infinite sequence of graphs.

Now the sequence $a_k \geq 0$ measures the sum of the squared distances from 0 and the matrices of the n-tuple in every iteration step. This sequence is monotonic decreasing and bounded from below by 0 and above by the initial finite value a_0, therefore it is convergent. From the convergence of a_k and (7.9) we have

$$a_{k+1} \leq a_k - (1/4) e_k, \tag{7.10}$$

which means that $e_k = \sum_{i=1}^n d\left(X_{j_i}^k, X_{l_i}^k\right)^2 \to 0$. Hence we know that the X_i^k matrices are approaching one another.

Now from Lemma 7.1 it is easy to see that the matrix sequences in Algorithm 2 are bounded, therefore they have a convergent subsequence of n-tuples $X_i^{k_j}$ which must have the same limit point A. Let $X_i^{l_j}$ be another convergent subsequence of tuples but with another limit point B. Without loss of generality $k_j > l_j$, we have from Lemma 7.3

$$\frac{\sum_{i=1}^n X_i^{k_j}}{n} \leq \frac{\sum_{i=1}^n X_i^{l_j}}{n}. \tag{7.11}$$

But we may choose a subsequence of subsequences as $k_r < l_r$, then

$$\frac{\sum_{i=1}^n X_i^{k_r}}{n} \geq \frac{\sum_{i=1}^n X_i^{l_r}}{n}. \tag{7.12}$$

Taking the limits we have $A \leq B$ and $A \geq B$, so $A = B$. Hence every convergent subsequence of tuples has the same limit point $(A, \ldots, A)$. Since the whole sequence of tuples is bounded, it converges to $(A, \ldots, A)$ as well. □

At this point due to the Kubo-Ando theory of matrix means [36], we already showed that the procedure converges for every symmetric matrix mean. This is so since the largest symmetric matrix mean is the arithmetic mean $\frac{A+B}{2}$, refer to Theorem (3.4).

Note that the above proof does not tell anything about the possibly different limit points of the iterative procedures corresponding to different sequences of graphs chosen in step 4 of Algorithm 2. These limit points generally seems to be different, they depend on the graphs G^k chosen in every iteration step k, similarly to the metric space case. Therefore we also introduce the following notation in order to express this dependence on the sequence of graphs. Let us denote the infinite sequence of graphs with

$$G = \left\{G^0, G^1, \ldots\right\}. \tag{7.13}$$

With this notation from now on we denote the common limit point of the sequences in Theorem 7.4 as $M_G(X_1, \ldots, X_n)$ to express the dependence of the limit point on the sequence of graphs G.

The other question that one can ask (motivated by Theorem 5.6 in the metric space case) is what is the rate of convergence of the sequences X_i^k to the common limit $M_G(X_1, \ldots, X_n)$. Or more specifically how does the infinite sequence of graphs $G = \left\{G^0, G^1, \ldots\right\}$ affect the rate of convergence. Generally numerical experiments show that the rate of convergence should be linear for all possible infinite sequences of graphs G. However it appears that the chosen graphs can greatly affect the quotient of this linear convergence, similarly to the metric space case, when we have the permutations π. The heuristic function Idealmapping speeds up the convergence of the iterative procedure here as well. Roughly speaking Idealmapping maximizes the length of the closed path in G^k. By length we mean the sum of the squared distances of the matrices X_i^k from one another measured over the edges of the closed path in G^k with the distance function (7.2). This is just the error term e_k introduced in (7.10). Now one can conclude that as X_i^k approaches the common limit point so does a_k its own limit which is a nonnegative number. Therefore if we maximize e_k in every step we can speed up the convergence. This argument tells us how Idealmapping works.

7.1.2 Properties of the extension $M_G(X_1, \ldots, X_n)$

Now that we have proved the convergence of this extension method, we advance further by showing some useful properties of the limit point $M_G(X_1, \ldots, X_n)$.

Proposition 7.5 (M. Pálfia [54])**.** *The limit point $M_G(X_1, \ldots, X_n)$ of the matrix sequences given in Algorithm 2 satisfies* 1., 3. *and* 4. *in Definition 4.1 with respect to an infinite sequence of graphs G.*

Proof. Poperty 1. is trivial. We prove property 4. Let $X_1, \ldots, X_n \in P(r, \mathbb{C})$ and $X_i \leq X_i' \in P(r, \mathbb{C})$. Let us consider one iteration step with respect to the mapping between the n-tuple of matrices and some graph g. Compute the means with respect to the graph g on the two n-tuple given as $X_1, \ldots, X_n$ and $X_1', \ldots, X_n'$. Considering the two iteration steps with respect to the same graph

g we get the following inequalities

$$M(X_i, X_j) \leq M(X'_i, X'_j), \tag{7.14}$$

for all i, j pairs by the monotonicity property of mean functions. Using again this property and (3.2), it is easy to see that throughout the iterative process we will have these kind of inequalities so we can see that the order of matrices is preserved by one iteration step, thus taking the limits we can see that the two limits will have the same order as well.

Finally poperty 3. is an easy consequence of property 1. and 4., if minimum and maximum exist. Setting up the same iteration on the new n-tuple formed by the minimal element we get the inequality on the left in property 3., similarly we can obtain the inequality on the right as well. □

Proposition 7.6 (M. Pálfia [54])**.** *If $M(A, B) \leq N(A, B) \leq (A + B)/2$ are matrix means, then the same ordering is true for the induced limit points $M_G(X_1, \ldots, X_n)$ and $N_G(X_1, \ldots, X_n)$ of the matrix sequences given in Algorithm 2 with respect to an infinite sequence of graphs G.*

Proof. The proof of this assertion is very similar to the above one, the only difference is that after one iteration we get $X_i^1 \leq (X'_i)^1$, where $X_i^1 = M(X_j, X_l)$ and $(X'_i)^1 = N(X_j, X_l)$ for all i. Now again considering the fact that the monotonicity is preserved by one iteration step, we get the same ordering for the limits. □

Proposition 7.7 (M. Pálfia [54])**.** *The limit point $M_G(X_1, \ldots, X_n)$ of the matrix sequences given in Algorithm 2 satisfies property* 6. *in Definition 4.1.*

Proof. Let $(X'_i)^0 = CX_i^0C^*$ and set up the same iteration on $(X'_1)^0, \ldots, (X'_n)^0$ as on $X_1^0, \ldots, X_n^0$. Equation (3.1) implies

$$CX_i^{k+1}C^* = CM\left(X_{j_i}^k, X_{l_i}^k\right)C^* = M\left(CX_{j_i}^kC^*, CX_{l_i}^kC^*\right). \tag{7.15}$$

Applying this recursively in every iteration step we get

$$CX_i^kC^* = (X'_i)^k. \tag{7.16}$$

Taking the limit $k \to \infty$ the assertion follows. □

Proposition 7.8 (M. Pálfia [54])**.** *The limit point $M_G(X_1, \ldots, X_n)$ of the matrix sequences given in Algorithm 2 is a continuous function in its $X_1, \ldots, X_n$ variables.*

Proof. We know that for a function $f : Y_1 \to Y_2$ between two metric spaces (Y_1, d_1) and (Y_2, d_2) sequential continuity and the usual topological continuity are equivalent. A proof can be found for example in [44]. We will show that sequential continuity holds therefore arriving at the desired result.

We will make use of the following multiplicative metric on $P(r, \mathbb{C})$

$$R(A, B) = \max\left\{\rho(A^{-1}B), \rho(B^{-1}A)\right\} \tag{7.17}$$

for all $A, B \in P(r, \mathbb{C})$ and $\rho(A)$ denotes the spectral radius of A. The above metric has the following properties [15]

(i) $R(A,B) \geq 1$,

(ii) $R(A,B) = 1$ iff $A = B$,

(iii) $R(A,C) \leq R(A,B)R(B,C)$,

(iv) $R(A,B)^{-1}A \leq B \leq R(A,B)A$,

(v) $\|A - B\| \leq (R(A,B) - 1)\|A\|$.

An extension of this metric to $P(r,\mathbb{C})^n$ can be given as follows. Let $X = (X_1, \cdots, X_n) \in P(r,\mathbb{C})^n$ and $Y = (Y_1, \cdots, Y_n) \in P(r,\mathbb{C})^n$, then we define

$$R_n(X,Y) = \max_{1\leq i\leq n} \left\{R(X_i, Y_i)\right\}. \tag{7.18}$$

Now suppose we have a convergent sequence of tuples $X^k = \left(X_1^k, \ldots, X_n^k\right) \in P(r,\mathbb{C})^n$ for which $\left(X_1^k, \ldots, X_n^k\right) \to (X_1, \ldots, X_n) = X \in P(r,\mathbb{C})^n$. Using property (iv) of $R(A,B)$ we have the following inequalities

$$R_n(X^k, X)^{-1} X_i^k \leq X_i \leq R_n(X^k, X) X_i^k. \tag{7.19}$$

Now applying the monotonicity property of M_G proved in Proposition 7.5 we have with the notation $c_k = R_n(X^k, X)$ the following

$$M_G\left(c_k^{-1}X_1^k, \ldots, c_k^{-1}X_n^k\right) \leq M_G\left(X_1, \ldots, X_n\right) \leq M_G\left(c_k X_1^k, \ldots, c_k X_n^k\right). \tag{7.20}$$

Using Proposition 7.7 we conclude that

$$c_k^{-1} M_G\left(X_1^k, \ldots, X_n^k\right) \leq M_G\left(X_1, \ldots, X_n\right) \leq c_k M_G\left(X_1^k, \ldots, X_n^k\right). \tag{7.21}$$

Taking the limit $k \to \infty$ we have $c_k \to 1$. This shows that

$$\lim_{k\to\infty} M_G\left(X_1^k, \ldots, X_n^k\right) = M_G\left(\lim_{k\to\infty} X_1^k, \ldots, \lim_{k\to\infty} X_n^k\right) \tag{7.22}$$

which is sequential continuity for M_G. □

Actually we have proved more above, we basically showed that for a function

Corollary 7.9 (M. Pálfia [64])**.** *$F : P(r,\mathbb{C})^n \mapsto P(r,\mathbb{C})$ which satisfies properties*

1. *if $X_i \leq X_i'$ for all i, then $F(X_1, \ldots, X_n) \leq F(X_1', \ldots, X_n')$,*

2. *$F(cX_1, \ldots, cX_n) = cF(X_1, \ldots, X_n)$ for real $c > 0$,*

it follows that F is continuous.

This also shows

Corollary 7.10 (M. Pálfia [64])**.** *Property (iv) is superfluous in Definition 3.1 of Kubo-Ando connection theory.*

We would like to point out that this is the second time so far that we have seen an inequality of the form

$$a_{k+1} \leq a_k - (k/8)e_k. \tag{7.23}$$

This inequality played a fundamental role in the proofs of Theorem 5.5 and 7.4 with $k = 2$ in the second case. We will meet with this inequality for a couple of times later as well.

7.2 Weighted Means Revisited

In the previous subsection we have seen that it is possible to show that the Iterative mean procedure works for all symmetric matrix means, without invoking a strong geometrical framework attached to each symmetric matrix mean. We will see in this section that even weighted counterparts corresponding to a symmetric matrix mean are constructable without affine geodesic structures. In this case as well, we will follow geometric intuition, although the proofs will be of matrix analytical nature.

First of all we define a procedure for every symmetric matrix mean $M(A, B)$ and for all $t \in [0, 1]$ which will be our weighted mean. Our procedure will be based on the fact that every $t \in [0, 1]$ can be approximated by dyadic rationals $\frac{m}{2^n}$ since dyadic rationals are dense in $[0, 1]$.

Definition 7.1 (Weighted mean process, M. Pálfia [64])**.** Let $M(\cdot, \cdot)$ be a symmetric matrix mean, $A, B \in P(r, \mathbb{C})$ and $t \in [0, 1]$. Let $a_0 = 0$ and $b_0 = 1$, $A_0 = A$ and $B_0 = B$. Define a_n, b_n and A_n, B_n recursively by the following procedure for all $n = 0, 1, 2, \ldots$:

if $a_n = t$ **then**
  $a_{n+1} = a_n$ and $b_{n+1} = a_n$, $A_{n+1} = A_n$ and $B_{n+1} = A_n$
else if $b_n = t$ **then**
  $a_{n+1} = b_n$ and $b_{n+1} = b_n$, $A_{n+1} = B_n$ and $B_{n+1} = B_n$
else if $\frac{a_n+b_n}{2} \leq t$ **then**
  $a_{n+1} = \frac{a_n+b_n}{2}$ and $b_{n+1} = b_n$, $A_{n+1} = M(A_n, B_n)$ and $B_{n+1} = B_n$
else
  $b_{n+1} = \frac{a_n+b_n}{2}$ and $a_{n+1} = a_n$, $B_{n+1} = M(A_n, B_n)$ and $A_{n+1} = A_n$
end if

According to the above a_{n+1}, b_{n+1} and A_{n+1}, B_{n+1} are clearly defined with respect to a_n, b_n and A_n, B_n recursively.

This algorithm may also be regarded as a kind of binary search with recurrence relation:

if $t = \frac{t_1+t_2}{2}$ **then**
  $M_t(A, B) = M\left(M_{t_1}(A, B), M_{t_2}(A, B)\right)$
end if

Theorem 7.11 (M. Pálfia [64])**.** *The sequences A_n and B_n given in Definition 7.1 are convergent and have the same limit point.*

Proof. In the case if $t = m2^{-k}$ for some integer m and k then there is nothing to prove, the procedure converges after finite steps. So suppose that t is not a dyadic rational. We will make use the already introduced (7.17) multiplicative metric on $P(r,\mathbb{C})$ [5]

$$R(A,B) = \max\left\{\rho(A^{-1}B), \rho(B^{-1}A)\right\} \tag{7.24}$$

for all $A, B \in P(r,\mathbb{C})$ and $\rho(A)$ denotes the spectral radius of A. Since $R(A,B) = R(I, A^{-1/2}BA^{-1/2})$ we have

$$R\left(A, M(A,B)\right) = R\left(I, f\left(A^{-1/2}BA^{-1/2}\right)\right), \tag{7.25}$$

where $f(t)$ is the corresponding normalized operator monotone function. Now since $M(A,B)$ is symmetric Theorem 3.4 holds. From this for the corresponding normalized operator monotone function $f(t)$ we have

$$2\left(I + X^{-1}\right)^{-1} \leq f(X) \leq \frac{I+X}{2}. \tag{7.26}$$

This also yields that $R(I, f(X)) \leq \max\left\{\rho(\frac{I+X}{2}), \rho(\frac{I+X^{-1}}{2})\right\} = \frac{1+R(I,X)}{2}$ for every $X \in P(r,\mathbb{C})$, so

$$R\left(A, M(A,B)\right) \leq \frac{1+R(A,B)}{2}. \tag{7.27}$$

By the above inequality we can easily conclude the following for the sequences A_n, B_n

$$\begin{aligned} R(A_{n+1}, B_{n+1}) &\leq \frac{1+R(A_n,B_n)}{2} = 1 + \frac{1}{2}\left[R(A_n,B_n) - 1\right] \\ R(A_n, B_n) &\leq 1 + \frac{1}{2^n}\left[R(A_0,B_0) - 1\right] \\ R(A_n, A_{n+1}) &\leq 1 + \frac{1}{2}\left[R(A_n,B_n) - 1\right] \\ R(A_n, A_{n+1}) &\leq 1 + \frac{1}{2^n}\left[R(A_0,B_0) - 1\right]. \end{aligned} \tag{7.28}$$

There exists $K \in P(r,\mathbb{C})$ such that $A \leq K, B \leq K$ and by property (ii) of matrix means $A_n \leq K, B_n \leq K$ so by property (v). of $R(\cdot,\cdot)$

$$\begin{aligned} \|A_{n+1} - A_n\| &\leq \left(R(A_{n+1}, A_n) - 1\right)\|K\| \\ \|A_{n+1} - A_n\| &\leq \frac{1}{2^n}\left[R(A_0,B_0) - 1\right]\|K\| \\ \sum_{n=0}^{\infty}\|A_{n+1} - A_n\| &\leq \sum_{n=0}^{\infty}\frac{1}{2^n}\left[R(A_0,B_0) - 1\right]\|K\| = \\ &= 2\left[R(A_0,B_0) - 1\right]\|K\|. \end{aligned} \tag{7.29}$$

This means that A_n is a Cauchy sequence therefore convergent and by the above we also have that $\|A_n - B_n\| \to 0$ so both A_n and B_n converge to the same limit point. □

We will base our weighted mean on the above theorem.

Definition 7.2 (Weighted mean, M. Pálfia [64]). The common limit point of A_n, B_n in Theorem 7.11 will be denoted by $M_t(A, B)$ and from now on in the article is considered as the corresponding weighted mean to a symmetric matrix mean $M(\cdot, \cdot)$.

What are the properties of this weighted mean? First of all it is not hard to prove the following

Proposition 7.12 (M. Pálfia [64]). *$M_t(A, B)$ yields the correct corresponding weighted means in the case of the arithmetic, geometric, harmonic means.*

The above is a consequence of the affine geodesy of the corresponding manifolds mentioned above. There are further important properties which are fulfilled by $M_t(A, B)$:

Proposition 7.13 (M. Pálfia [64]). *$M_t(A, B)$ for $A, B \in P(r, \mathbb{C})$ and $t \in [0, 1]$ fulfills the following properties*

(i') $M_t(I, I) = I$,

(ii') *if* $A \leq A'$ *and* $B \leq B'$, *then* $M_t(A, B) \leq M_t(A', B')$,

(iii') $CM_t(A, B)C \leq M_t(CAC, CBC)$,

(iv') *if* $A_n \downarrow A$ *and* $B_n \downarrow B$ *then* $M_t(A_n, B_n) \downarrow M_t(A, B)$,

(v') *if* $N(A, B) \leq M(A, B)$ *then* $N_t(A, B) \leq M_t(A, B)$,

(vi') $M_{1/2}(A, B) = M(A, B)$,

(vii') $M_t(A, B)$ *is continuous in* t,

Proof. Property (i') and (ii') are trivial consequences of the similar properties for symmetric matrix means.

For property (iii') consider $A' = CAC$ and $B' = CBC$ and start the procedure in the definition of $M_t(\cdot, \cdot)$ for the pair A, B and A', B'. Then we have $CA_1C = CM(A_0, B_0)C \leq M(CA_0C, CB_0C) = A'_1$ if $t > 1/2$ or $CB_1C = CM(A_0, B_0)C \leq M(CA_0C, CB_0C) = B'_1$. Now for every n we use property (ii) for symmetric matrix means so we have $CA_nC \leq A'_n$ and $CB_nC \leq B'_n$ for every $n \geq 1$. Taking the limits we conclude the assertion of property (iii').

What immediately follows from this property is that $M(CXC^*, CYC^*) = CM(X, Y)C^*$ for all invertible C. Now using Corollary 7.9 see that $M_t(A, B)$ is continuous in A, B so by property (ii') and the continuity we get property (iv') as a consequence.

At this point we already have by the Kubo-Ando theory of matrix means that $M_t(A, B)$ is a matrix mean as well, so it fulfills the additional properties (v)-(ix). Consequently it has a representation with a normalized operator monotone function.

Property (v') is an easy consequence of repeated usage of property (ii) for matrix means for every n. Property (vi') is also trivial.

To prove property (vii') we have to do a bit more work. We have to show that if $|t_1 - t_2| \to 0$ then also $\|M_{t_1}(A,B) - M_{t_2}(A,B)\| \to 0$. Suppose $t_1 < t_2$ and take the smallest j for which we have $t_1 \leq m2^{-j} \leq t_2$ for some m. Let us set up the iterative procedure in Definition 7.1 on A, B with t_1 and t_2 respectively. Let us denote the yielded matrix sequences in the case of t_1 with $A_i^{t_1}, B_i^{t_1}$ and in the case of t_2 with $A_i^{t_2}, B_i^{t_2}$ and similarly for the numbers with $a_i^{t_1}, b_i^{t_1}$ and $a_i^{t_2}, b_i^{t_2}$. Notice that the iterative procedure in the jth step for t_1 will yield $b_j^{t_1} = m2^{-j}$ and similarly $a_j^{t_2} = m2^{-j}$ in the jth step for t_2. Suppose $t_1 \neq m2^{-j}$. Then there exists $i \geq j$ such that $a_i^{t_1} \leq t_1 \leq b_i^{t_1}$ but $(a_i^{t_1} + b_i^{t_1})/2 \geq t_1$, this means that $b_{i+1}^{t_1} \neq b_i^{t_1}$. If $t_1 = m2^{-j}$ then we have $a_p^{t_1} = b_p^{t_1} = t_1$ for $p > j$ and in this case we define $i := +\infty$. Similarly either there exists a smallest $l \geq j$ such that $a_{l+1}^{t_2} \neq a_l^{t_2}$, or we have $t_2 = m2^{-j}$ so $a_p^{t_2} = b_p^{t_2} = t_2$ for $p > j$ and again then we define $l := +\infty$. Notice that i and l cannot be infinite at the same time, so we define $k := \min\{i, l\}$. It is easy to see that as $t_1 \to t_2$, $k \to \infty$. We also have that $B_k^{t_1} = A_k^{t_2}$, so we can bound the distance of the limit points $M_{t_1}(A,B)$ and $M_{t_2}(A,B)$ from $B_k^{t_1} = A_k^{t_2}$ as follows:

$$\begin{aligned} \left\| B_k^{t_1} - \lim_{j\to\infty} B_j^{t_1} \right\| &\leq \sum_{i=k}^{\infty} \left\| B_{i+1}^{t_1} - B_i^{t_1} \right\| \\ \sum_{i=k}^{\infty} \left\| B_{i+1}^{t_1} - B_i^{t_1} \right\| &\leq \frac{1}{2^k} 2 \left[R(A_0, B_0) - 1 \right] \|K\| = \\ &= \frac{1}{2^{k-1}} \left[R(A_0, B_0) - 1 \right] \|K\| . \end{aligned} \tag{7.30}$$

We also have the same bound for $\left\| A_k^{t_2} - \lim_{j\to\infty} A_j^{t_2} \right\|$. Since $B_k^{t_1} = A_k^{t_2}$, we have

$$\begin{aligned} \|M_{t_1}(A_0,B_0) - M_{t_2}(A_0,B_0)\| \leq & \left\| M_{t_1}(A_0,B_0) - B_k^{t_1} \right\| + \\ & + \left\| M_{t_2}(A_0,B_0) - A_k^{t_2} \right\| \leq \\ \leq & \frac{1}{2^{k-2}} \left[R(A_0, B_0) - 1 \right] \|K\| . \end{aligned} \tag{7.31}$$

Since $k \to \infty$ as $t_1 \to t_2$, by the above $\|M_{t_1}(A,B) - M_{t_2}(A,B)\| \to 0$. □

By the above proposition we have that $M_t(A,B)$ is a continuous function in t. So $M_t(A,B)$ is a one parameter family of matrix means corresponding to every symmetric matrix mean. Since every matrix mean by virtue of property (ix) is representable by a normalized operator monotone function $f(x)$, we may represent such one parameter family of matrix means by a one parameter family of normalized operator monotone functions $f_t(x), t \in [0,1]$. So in other words we have the following

Corollary 7.14 (M. Pálfia [64])**.** *For every symmetric matrix mean $M(A,B)$ there is a corresponding one parameter family of weighted means $M_t(A,B)$ for*

$t \in [0,1]$. Let $f(x)$ be the normalized operator monotone function corresponding to $M(A,B)$. Then similarly we have a one parameter family of normalized operator monotone functions $f_t(x)$ corresponding to $M_t(A,B)$. The family $f_t(x)$ is continuous in t, and $f_0(x) = 1$ and $f_1(x) = x$ are the two extremal points, so $f_t(x)$ interpolates between these two points.

Based on this phenomenon we can conclude the following

Proposition 7.15 (M. Pálfia [64])**.** *Let $M(A,B)$ be a symmetric matrix mean. Then*

$$\left((1-t)A^{-1} + tB^{-1}\right)^{-1} \leq M_t(A,B) \leq (1-t)A + tB, \tag{7.32}$$

where $M_t(A,B)$ is the weighted version of $M(A,B)$.

Proof. For every symmetric matrix mean $M(A,B)$ we have by Theorem 3.4 that

$$\left(\frac{A^{-1} + B^{-1}}{2}\right)^{-1} \leq M(A,B) \leq \frac{A+B}{2}. \tag{7.33}$$

Now what follows from Propostion 7.12 is that the harmonic mean on the left hand side above has the weighted harmonic mean $H_t(A,B)$ defined in (4.66) as its weighted counterpart, and similarly we have the weighted arithmetic mean $A_t(A,B)$ defined in (4.65) as the weighted counterpart for the arithmetic mean on the right hand side above. Thus by property (v') in Proposition 7.13 and the above inequality we have

$$H_t(A,B) \leq M_t(A,B) \leq A_t(A,B). \tag{7.34}$$

□

We are going to use the above definition $M_t(A,B)$ of a weighted matrix mean to set up the Bini-Meini-Poloni procedure for every symmetric matrix mean, but before we do that we turn to the Ando-Li-Mathias procedure in the following section and prove its convergence for every symmetric matrix mean.

7.3 Ando-Li-Mathias Procedure Revisited

In this section we will prove the convergence of the Ando-Li-Mathias procedure for every possible symmetric matrix mean. In order to do that we will generalize the argument given for the Iterative mean in the case of all symmetric matrix means by applying induction. First of all let us recall Definition 4.4, the Ando-Li-Mathias procedure [5]:

Definition 7.3 (ALM iteration)**.** Let $X = (X_1^0, \ldots, X_n^0)$ where $X_i^0 \in P(r, \mathbb{C})$ and define the mapping $M(X_1, \ldots, X_n)$ inductively as follows. If $n = 2$ assume that $M(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \ldots, X_{n-1})$ is already defined. Then using $M(X_1, \ldots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M\left(Z_{\neq i}\left(X_1^l, \ldots, X_n^l\right)\right), \tag{7.35}$$

where $Z_{\neq i}(X_1^l, \dots, X_n^l) = X_1^l, \dots, X_{i-1}^l, X_{i+1}^l, \dots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l\to\infty} X_i^l = M(X_1^0, \dots, X_n^0). \tag{7.36}$$

Theorem 7.16 (M. Pálfia [64])**.** *Let $F : P(r,\mathbb{C})^2 \mapsto P(r,\mathbb{C})$ and suppose that $F(A,B)$ fulfills one of the inequalities below:*

$$\left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1} \leq F(A,B) \leq \left[\frac{A^2+B^2}{2} - \frac{k}{8}(A-B)^2\right]^{1/2} \tag{7.37}$$

for a $k \in (0,2]$, or

$$F(A,B) \leq \frac{A+B}{2}. \tag{7.38}$$

Then in Definition 4.4 starting with $M(A,B) := F(A,B)$, $M(X_1,\dots,X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.

Before we prove the above theorem we mention a few remarks and several lemmas which we will make use of later. First of all condition (7.37) might seem a bit strange at first glance although it immediately becomes straightforward if we consider $k = 2$, since in this case the right hand side becomes the arithmetic mean. If $k = 0$ then the right hand side is the square mean $\left(\frac{A^2+B^2}{2}\right)^{1/2}$. This literally means that the above theorem automatically covers every symmetric matrix mean due to Theorem 3.4 as a special case.

Now we have to study some properties of the square mean in order to prepare the necessary steps for the proof of the above theorem. First of all one should notice that the square mean is an affine mean. The underlying manifold is a Riemannian manifold defined as a pullback metric of the Euclidean metric $\langle A,B\rangle_E = Tr\{A^*B\}$ over the space of squared complex matrices. This Euclidean metric has corresponding distance function

$$\begin{aligned} d_E(A,B)^2 = \langle A-B, A-B\rangle_E = \\ = Tr\left\{(A-B)^*(A-B)\right\}. \end{aligned} \tag{7.39}$$

The isometry is $f(x) = x^2$ and it embeds $P(r,\mathbb{C})$ into $P(r,\mathbb{C})$. The distance function of the pullback metric on $P(r,\mathbb{C})$ is

$$\begin{aligned} d_{1/2}(A,B)^2 = \langle f(A)-f(B), f(A)-f(B)\rangle_E = \\ = Tr\left\{\left(A^2-B^2\right)^*\left(A^2-B^2\right)\right\}. \end{aligned} \tag{7.40}$$

The geodesics of this metric are of the form

$$\gamma_{A,B}(t) = f^{-1}\left[(1-t)f(A)+tf(B)\right] = \left[(1-t)A^2+tB^2\right]^{1/2}. \tag{7.41}$$

This shows that square mean is an affine mean, so the weighted mean process $M_t(A,B)$ for the square mean gives back the corresponding point on the geodesic above. Furthermore since the above metric is a pullback of a Euclidean metric, it is also Euclidean.

Actually the isometry $f(x)$ can be chosen arbitrarily, particularly any diffeomorphism will suffice. We are going to derive some properties of the ALM-procedure with $F(A,B)=\left(\frac{A^2+B^2}{2}\right)^{1/2}$ and $F(A,B)=\left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1}$ on $P(r,\mathbb{C})$ endowed with the above corresponding pullback metrics. We are going to denote the general pullback of the distance function $d_E(\cdot,\cdot)$ for an arbitrary f by

$$\begin{aligned} d_f(A,B)^2 &= \langle f(A)-f(B), f(A)-f(B)\rangle_E = \\ &= Tr\left\{[f(A)-f(B)]^*\,[f(A)-f(B)]\right\}. \end{aligned} \tag{7.42}$$

The metric space $P(r,\mathbb{C})$ with the distance function (7.42) is Euclidean, since its metric is a pullback metric of the standard Euclidean metric on the space of complex $r\times r$ matrices. Let $x_i\in P(r,\mathbb{C})$ for $i\in\{1,\dots,n\}$ and define $S=\{x_1,\dots,x_n\}$. We already know that the function

$$b(x)=\sum_{i=1}^{n} d(x,x_i)^2 \tag{7.43}$$

has a minimum for $d(\cdot,\cdot)=d_f(\cdot,\cdot)$ and this minimal value is attained at a unique point $\hat{x}$ which is called the center of mass of S. Moreover the center of mass is explicitly given for these metric spaces on $P(r,\mathbb{C})$ by Proposition 4.13. If we perform one ALM-iteration step on n points in the space $P(r,\mathbb{C})$ with this map then the iteration leaves the center of mass of the points invariant.

Proposition 7.17 (M. Pálfia [64])**.** *Let $X_i^0\in P(r,\mathbb{C})$ for $i=1,\dots,n$. Then the ALM-procedure (Definition 4.4) set up on the matrices $X_1^0,\dots,X_n^0$ with the $n-1$ variable function $M(x_1,\dots,x_{n-1})=f^{-1}\left(\frac{\sum_{i=1}^{n-1}f(x_i)}{n-1}\right)$ leaves the Riemann centroid of the points $X_1^0,\dots,X_n^0$ invariant with respect to the distance function* (7.42)*.*

Proof.

$$\begin{aligned} f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^1\right)}{n}\right) &= f^{-1}\left[\sum_{i=1}^{n}\frac{f\left(M\left(Z_{\neq i}(X_1^0,\dots,X_n^0)\right)\right)}{n}\right] = \\ = f^{-1}\left[\sum_{i=1}^{n}\frac{\sum_{j=1,j\neq i}^{n-1}\frac{f\left(X_j^0\right)}{n-1}}{n}\right] &= f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^0\right)}{n}\right) \end{aligned} \tag{7.44}$$

Similarly we obtain the above equality for every iteration step, so

$$f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^{l+1}\right)}{n}\right)=f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^{l}\right)}{n}\right)=f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^{0}\right)}{n}\right). \tag{7.45}$$

□

We turn our attention to more general functions than these pullback means. The following inequalities will turn out to be useful tools later.

Lemma 7.18 (M. Pálfia [64]).

$$\begin{aligned}&\left[(1-t)A^2+tB^2-\frac{k_2}{2}t(1-t)(A-B)^2\right]^{1/2}\leq\\ &\leq\left[(1-t)A^2+tB^2-\frac{k_1}{2}t(1-t)(A-B)^2\right]^{1/2}\end{aligned} \tag{7.46}$$

for any $A, B \in P(r, \mathbb{C})$ *and* $t \in [0, 1]$ *if* $k_1 \leq k_2$.

Proof.

$$0 \leq \frac{k_2-k_1}{2}t(1-t)(A-B)^2$$

$$-\frac{k_2}{2}t(1-t)(A-B)^2 \leq -\frac{k_1}{2}t(1-t)(A-B)^2$$

$$\begin{aligned}(1-t)A^2+tB^2-\frac{k_2}{2}t(1-t)(A-B)^2 &\leq\\ &\leq(1-t)A^2+tB^2-\frac{k_1}{2}t(1-t)(A-B)^2\end{aligned}$$

Taking the square root of both sides and considering the fact that the the square root is operator monotone we get the inequality of the assertion.

□

Notice that for $k_1 = 0$ and $k_2 = 2$ we get the weighted arithmetic-square mean inequality

$$(1-t)A+tB \leq \left[(1-t)A^2+tB^2\right]^{1/2}. \tag{7.47}$$

We will prove an important inequality which will play a fundamental role in our further investigations. An important part of the proof of the convergence of the ALM- and BMP-process will rely on this inequality.

Lemma 7.19 (M. Pálfia [64]). *Let* $k \in [0, 2]$, $F : P(r, \mathbb{C})^2 \mapsto P(r, \mathbb{C})$ *and*

$$F(A,B) \leq \left[(1-t)A^2+tB^2-\frac{k}{2}t(1-t)(A-B)^2\right]^{1/2}. \tag{7.48}$$

Then with the distance function $d_E(A,B)^2 = Tr\left\{(A-B)^*(A-B)\right\}$,

$$d_E\left(0,F(A,B)\right)^2 \leq (1-t)d_E(0,A)^2+td_E(0,B)^2-\frac{k}{2}t(1-t)d_E(A,B)^2. \tag{7.49}$$

Proof. By substitution the assertion has the following form

$$Tr\left\{F(A,B)^2\right\} \le Tr\left\{(1-t)A^2 + tB^2 - \frac{k}{2}t(1-t)(A-B)^2\right\}. \tag{7.50}$$

This holds, since we have the following identity for hermitian positive definite $X \le Y$

$$0 \le Tr\left\{Y^2 - X^2\right\} = Tr\left\{(Y-X)(Y+X)\right\}. \tag{7.51}$$

By choosing $X = F(A,B) - \left[(1-t)A^2 + tB^2 - \frac{k}{2}t(1-t)(A-B)^2\right]^{1/2}$ and $Y = F(A,B) + \left[(1-t)A^2 + tB^2 - \frac{k}{2}t(1-t)(A-B)^2\right]^{1/2}$ we get the assertion.

□

Notice that the above lemma is already true for every matrix mean $M(A,B)$ and their weighted $M_t(A,B)$ counterparts by Lemma 7.15 and (7.47). By Lemma 7.18 we also have a relatively wide family of functions which fulfills the conditions of the above lemma.

Now we are in position to prove Theorem 7.16.

Proof. (Theorem 7.16) The proof will be based on induction on the number of matrices n. We are going to measure the sum of the squared distances of the matrices X_i^l from the zero matrix with respect to the distance function (7.39) with

$$a_n^l = \sum_{i=1}^{n} d_E(0, X_i^l)^2 = \sum_{i=1}^{n} Tr\left\{(X_i^l)^2\right\}. \tag{7.52}$$

We will also measure sum of the squared distances of the X_i^l from one another. We will form this sum over all possible pairs of X_i^l as

$$e_n^l = \sum_{1 \le i < j \le n} d_E(X_i^l, X_j^l)^2 = \sum_{1 \le i < j \le n} Tr\left\{(X_i^l - X_j^l)^2\right\}. \tag{7.53}$$

We are going to denote the common limit point of the sequences X_i^l by $F_n(X_1^0, \dots, X_n^0)$ for n. We will need the following lemmas which will be proved by induction as well on the number of matrices n, so we have to embed these lemmas into this proof of Theorem 7.16. All three lemmas will be proved by assuming that they hold for n matrices and also that the ALM-procedure converges to common limit for n matrices. Making this assumption we show that the lemmas hold for $n+1$ and that the ALM procedure converges to common limit for $n+1$ as well. For the first step of the induction ($n = 3$) we will prove the lemmas directly. First we are going to treat the case of the first inequality (7.37).

Lemma 7.20 (Monotone Iteration, M. Pálfia [64])**.** *In the first case of inequality* (7.37) *we have*

$$\left(\frac{\sum_{i=1}^{n}\left(X_i^{l+1}\right)^{-1}}{n}\right)^{-1} \ge \left(\frac{\sum_{i=1}^{n}\left(X_i^{l}\right)^{-1}}{n}\right)^{-1}. \tag{7.54}$$

In the second case of inequality (7.38) *we have*

$$\frac{\sum_{i=1}^{n} X_i^{l+1}}{n} \leq \frac{\sum_{i=1}^{n} X_i^{l}}{n}. \tag{7.55}$$

Proof. We argue by induction on the number of matrices n by making the assumption that the ALM-procedure converges for $n \geq 3$ to common limit point $F_n(X_1^0, \dots, X_n^0)$, in other words $X_i^l \to F_n(X_1^0, \dots, X_n^0)$ for n and that the lemma holds for n. Consider the first case of inequality (7.37). Then the inequality of the lemma for n implies that

$$\left(\frac{\sum_{i=1}^{n} \left(X_i^l\right)^{-1}}{n}\right)^{-1} \geq \left(\frac{\sum_{i=1}^{n} \left(X_i^0\right)^{-1}}{n}\right)^{-1} \tag{7.56}$$

and if we take the limit on the left hand side for n we get the inequality

$$\left(\frac{\sum_{i=1}^{n} \left(F_n(X_1^0, \dots, X_n^0)\right)^{-1}}{n}\right)^{-1} = F_n(X_1^0, \dots, X_n^0) \geq \left(\frac{\sum_{i=1}^{n} \left(X_i^0\right)^{-1}}{n}\right)^{-1}. \tag{7.57}$$

The above inequality also holds directly for $n = 2$ by the assumption of inequality (7.37), so this will also provide the first step in our induction.

Now we prove the lemma for $n + 1$ if it is true for n. By (7.57) we have

$$X_i^{l+1} = F_n\left(Z_{\neq i}(X_1^l, \dots, X_{n+1}^l)\right) \geq \left(\frac{\sum_{j=1, j\neq i}^{n+1} \left(X_i^l\right)^{-1}}{n}\right)^{-1}. \tag{7.58}$$

The n+1-variable harmonic mean is operator monotone in its variables, therefore if we take the $n+1$-variable harmonic mean of the above on the left and right hand side, we get

$$\left(\frac{\sum_{i=1}^{n+1} F_n\left(Z_{\neq i}(X_1^l, \dots, X_{n+1}^l)\right)^{-1}}{n+1}\right)^{-1} \geq \left(\frac{\sum_{i=1}^{n+1} \left(H_i^l\right)^{-1}}{n+1}\right)^{-1}, \tag{7.59}$$

where $H_i^l = \left(\frac{\sum_{j=1, j\neq i}^{n+1} \left(X_i^l\right)^{-1}}{n}\right)^{-1}$. Then by Proposition 7.17 with $f(t) = t^{-1}$, the harmonic mean of the $n+1$ matrices is left invariant on the right hand side, so this is equivalent to

$$\left(\frac{\sum_{i=1}^{n+1} \left(X_i^{l+1}\right)^{-1}}{n+1}\right)^{-1} \geq \left(\frac{\sum_{i=1}^{n+1} \left(X_i^l\right)^{-1}}{n+1}\right)^{-1}. \tag{7.60}$$

The second case given by inequality (7.38) is very similar to the proof of the first case. Instead of inequality (7.56) we have

$$\frac{\sum_{i=1}^{n} X_i^{l}}{n} \leq \frac{\sum_{i=1}^{n} X_i^{0}}{n} \tag{7.61}$$

and instead of (7.57) we have

$$\frac{\sum_{i=1}^n F_n(X_1^0,\dots,X_n^0)}{n} = F_n(X_1^0,\dots,X_n^0) \le \frac{\sum_{i=1}^n X_i^0}{n}. \tag{7.62}$$

The rest of the argument is just the same, although we have the n-variable arithmetic mean replacing the n-variable harmonic mean, and the inequalities are reversed. The lemma is proved. □

Lemma 7.21 (Decreasing Distances, M. Pálfia [64])**.** *We have*

$$a_n^{l+1} \le a_n^l - \frac{k}{8} z_n e_n^l \tag{7.63}$$

in the case of (7.37)*, or we have* (7.63) *with* $k = 2$ *in the case of* (7.38)*. In both cases* $z_n = \frac{2}{n-1}$*.*

Proof. We will prove this for the case (7.37). The second case of (7.38) is just the same with $k = 2$, we will only use that the right hand side of (7.37) holds, so we do not have to treat the second case (7.38) separately due to (7.47) with $t = 1/2$. The first step is to show the above for $n = 3$. By Lemma 7.19 and that the right hand side of (7.37) is equivalent to the assumption of the lemma for $t = 1/2$, we have

$$\begin{aligned} d_E\left(0, F(X_i^l, X_j^l)\right)^2 &\le \frac{d_E(0,X_i^l)^2 + d(0,X_j^l)^2}{2} - \frac{k}{8} d_E(X_i^l, X_j^l)^2 \\ d_E\left(0, X_s^{l+1}\right)^2 &\le \frac{d_E(0,X_i^l)^2 + d(0,X_j^l)^2}{2} - \frac{k}{8} d_E(X_i^l, X_j^l)^2, \end{aligned} \tag{7.64}$$

where $i, j, s \in \{1,2,3\}$ and $i \ne j \ne s, s \ne i$. There are 3 distinct inequalities of the above for $s = 1,2,3$. By summing these inequalities for s we get (7.63) for $n = 3$ and $z_3 = 1$.

Now suppose (7.63) holds for n and that X_i^l converge to a common limit point for n denoted again by $F_n(X_1^0,\dots,X_n^0)$. Then we have

$$a_n^l \le a_n^0 - \frac{k}{8} z_n e_n^0 \tag{7.65}$$

and by taking the limit on the left hand side we get

$$\begin{aligned} \lim_{l\to\infty} a_n^l = n d_E\left(0, F_n(X_1^0,\dots,X_n^0)\right)^2 &\le a_n^0 - \frac{k}{8} z_n e_n^0 \\ d_E\left(0, F_n(X_1^0,\dots,X_n^0)\right)^2 &\le \frac{a_n^0}{n} - \frac{k}{8}\frac{z_n}{n} e_n^0. \end{aligned} \tag{7.66}$$

Then set up the ALM-procedure on $X_i^0 \in P(r,\mathbb{C}); i = 1,2\dots,n+1$ with $M_n(X_1,\dots,X_n) := F_n(X_1,\dots,X_n)$. Inequality (7.66) can be applied in any of the iteration steps, so we get

$$\begin{aligned} & d_E\left(0, F_n\left(Z_{\ne i}\left(X_1^l,\dots,X_{n+1}^l\right)\right)\right)^2 \le \\ & \le \frac{\sum_{j=1,j\ne i}^{n+1} d_E(0,X_j^l)^2}{n} - \frac{k}{8}\frac{z_n}{n} \sum_{1\le j<s\le n+1, j\ne i, s\ne i} d_E(X_j^l, X_s^l)^2. \end{aligned} \tag{7.67}$$

If we sum these inequalities for i we arrive at the following

$$\sum_{i=1}^{n+1} d_E\left(0, F_n\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)^2 \leq$$
$$\leq \sum_{i=1}^{n+1} \frac{\sum_{j=1,j\neq i}^{n+1} d_E(0, X_j^l)^2}{n} - \frac{k}{8}\frac{z_n}{n}\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2. \tag{7.68}$$

The left hand side of the above is just a_{n+1}^{l+1}. The first term on the right hand side is easily written as

$$\sum_{i=1}^{n+1} \frac{\sum_{j=1,j\neq i}^{n+1} d_E(0, X_j^l)^2}{n} = \sum_{i=1}^{n+1} d_E(0, X_i^l)^2 = a_{n+1}^l. \tag{7.69}$$

We have to carefully analyze the second term

$$\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2. \tag{7.70}$$

Consider the complete graph K_{n+1} on $n+1$ vertices labelled from 1 to $n+1$. In this way we have a natural bijective mapping between the matrices X_i^l and the vertices of K_{n+1}. Then for every squared distance $d_E(X_j^l, X_s^l)^2$ we have a corresponding edge in K_{n+1} of the form (j, s). Then the sum $\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2$ is just the sum of the squared distances corresponding to the edges of the complete graph K_n given on the vertices $\{1, \ldots, i-1, i+1, \ldots, n+1\}$. This is almost $\sum_{1\leq j<s\leq n+1} d_E(X_j^l, X_s^l)^2$, but we leave out from the sum every squared distance corresponding to an edge that has the vertex i as an ending vertex. So actually (7.70) almost equals to

$$\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1} d_E(X_j^l, X_s^l)^2 = (n+1)e_{n+1}^l, \tag{7.71}$$

but in the sum (7.70) every vertex has been left out once, so every squared distance corresponding to an edge has been left out twice, hence

$$\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2 = (n-1)e_{n+1}^l. \tag{7.72}$$

This shows us that $z_{n+1} = \frac{n-1}{n}z_n$ and also $z_3 = 1$, so in other words by solving the recursion we get

$$z_n = \frac{2}{n-1}. \tag{7.73}$$

This concludes the lemma for every n. □

Lemma 7.22 (Boundedness, M. Pálfia [64]). *The matrix sequences X_i^l are bounded for all n.*

Proof. We have the trivial lower bound $X_i^l \geq 0$, since by assumption $F(A,B) \geq 0$, so we have $X_i^l \geq 0$ for $n=3$. Now assume again that the ALM-procedure converges to common limit point denoted by $F_n(X_1^0,\ldots,X_n^0)$ for n and $F_n(X_1^0,\ldots,X_n^0) \geq 0$. Then trivially for $n+1$ the sequences $X_i^l \geq 0$ since $F_n(X_1^0,\ldots,X_n^0) \geq 0$. This also shows that if the sequences converge for $n+1$ to a common limit $F_{n+1}(X_1^0,\ldots,X_n^0)$, then this limit is also bounded from below, so $F_{n+1}(X_1^0,\ldots,X_n^0) \geq 0$.

Now we provide a suitable upper bound as well. By the previous assertion Lemma Decreasing Distances we have (7.63) for $n \geq 3$. In particularly for $n=3$ it holds providing the first step, while for $n>3$ we need the inductional hypothesis that the ALM-procedure converges to a common limit point $F_{n-1}(X_1^0,\ldots,X_{n-1}^0)$ for $n-1$. The rest of the argument is just the same for all $n \geq 3$. So by (7.63) we have $a_n^{l+1} \leq a_n^l$ which means that the sequence is monotone decreasing. So we have the bound

$$a_n^l \leq a_n^0 = b. \tag{7.74}$$

By this above we get for arbitrary i that

$$d\left(0,X_i^l\right)^2 = a_n^l - \sum_{j=1,j\neq i}^{n} d\left(0,X_j^l\right)^2 \leq a_n^l \leq b. \tag{7.75}$$

This means that the norm $\left\|X_i^l\right\|$ is bounded from above by b since

$$\left\|X_i^l\right\|^2 = Tr\left\{\left(X_i^l\right)^2\right\} = d\left(0,X_i^l\right)^2. \tag{7.76}$$

This concludes the proof of the lemma. □

Now we move on to the final step of the induction. We prove that for $n=3$ the ALM-procedure converges and that if it converges for n then it converges for $n+1$ in both cases of inequalities (7.37) and (7.38). This last step will be based on the three lemmas: Lemma Monotone Iteration, Lemma Decreasing Distances and Lemma Boundedness. It is not necessary to prove separately the $n=3$ case since these three lemmas hold for $n=3$ and the argument will be the same as for general $n+1$ requiring the inductional hypothesis, the convergence of the procedure to a common limit point for n.

So by Lemma Decreasing Distances we have

$$a_n^{l+1} \leq a_n^l - \frac{k}{8} z_n e_n^l, \tag{7.77}$$

in other words a_n^l is a decreasing nonnegative sequence in l, therefore convergent. Since $z_n > 0$ and has fixed value for each n by Lemma Decreasing Distances, this means that $e_n^l \to 0$ as $l \to \infty$, so the matrices X_i^l are approaching one another. By Lemma Boundedness we have that these sequences are bounded, hence they have convergent subsequences. But since $e_n^l \to 0$ these subsequences are converging to a common limit point. Let $X_i^{s_l}$ denote a subsequence converging to say A and $X_i^{r_l}$ another subsequence converging to B. Without loss of generality we can take $s_l > r_l$. By Lemma Monotone Iteration for the case of inequality (7.37) we have

$$\left(\frac{\sum_{i=1}^{n}(X_i^{s_l})^{-1}}{n}\right)^{-1} \geq \left(\frac{\sum_{i=1}^{n}(X_i^{r_l})^{-1}}{n}\right)^{-1}. \tag{7.78}$$

Now choose a subsequence of subsequences $s_j < r_j$ so then again by the lemma

$$\left(\frac{\sum_{i=1}^{n}\left(X_i^{s_j}\right)^{-1}}{n}\right)^{-1} \leq \left(\frac{\sum_{i=1}^{n}\left(X_i^{r_j}\right)^{-1}}{n}\right)^{-1}. \tag{7.79}$$

Taking the limits we have $A \geq B$ and $A \leq B$ so $A = B$. In the second case (7.38) we have the same argument but using the n-variable arithmetic mean instead of the n-variable harmonic above.

Now this argument shows the convergence of the ALM-procedure to a common limit point directly for $n = 3$ and inductively for n assuming convergence to common limit for $n - 1$.

□

We are going to study some properties of this limit point later, jointly with the case of the BMP-mean, after showing that the BMP procedure converges. In the next section we will show a similar theorem to Theorem 7.16 for the BMP-procedure.

7.4 Bini-Meini-Poloni Procedure Revisited

In this section we will treat the case of the Bini-Meini-Poloni procedure. We may do that for matrix means since we have defined a weighted mean $M_t(A, B)$ corresponding to any symmetric matrix mean $M(A, B)$. The outline of the proof of the convergence of the BMP-procedure will roughly follow the one of the ALM-procedure, although some lemmas will be formulated differently.

Firstly let us recall again Definition 4.5, the Bini-Meini-Poloni procedure [15]:

Definition 7.4 (BMP iteration)**.** Let $X = (X_1^0, \ldots, X_n^0)$ where $X_i^0 \in P(r, \mathbb{C})$ and define the mapping $M(X_1, \ldots, X_n)$ inductively as follows. If $n = 2$ assume that $M_t(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \ldots, X_{n-1})$ is already defined. Then using $M(X_1, \ldots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M_{\frac{n-1}{n}}\left(X_i^l, M\left(Z_{\neq i}\left(X_1^l, \ldots, X_n^l\right)\right)\right), \tag{7.80}$$

where $Z_{\neq i}(X_1^l, \ldots, X_n^l) = X_1^l, \ldots, X_{i-1}^l, X_{i+1}^l, \ldots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l\to\infty} X_i^l = M(X_1^0, \ldots, X_n^0). \tag{7.81}$$

Theorem 7.23 (M. Pálfia [64]). *Let $F : [0,1] \times P(r,\mathbb{C})^2 \mapsto P(r,\mathbb{C})$ and suppose that $F_t(A,B)$ fulfills one of the inequalities below:*

$$\begin{aligned}&\left[(1-t)A^{-1}+tB^{-1}\right]^{-1} \leq F_t(A,B) \leq \\ &\leq \left[(1-t)A^2+tB^2-\frac{k}{2}t(1-t)(A-B)^2\right]^{1/2}\end{aligned} \tag{7.82}$$

for a $k \in (0,2]$ and every $t \in [0,1]$, or

$$F_t(A,B) \leq (1-t)A+tB, \tag{7.83}$$

for every $t \in [0,1]$. Then in Definition 4.5 starting with $M_t(A,B) := F_t(A,B)$, $M(X_1,\dots,X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.

Before we turn to the proof of the above theorem, we again consider some lemmas which will be similar to the ALM case. Let us recall again the metric space $P(r,\mathbb{C})$ with the distance function (7.42). We already know that the minimum of

$$b(x) = \sum_{i=1}^{n} d_f(x,x_i)^2 \tag{7.84}$$

is attained at a unique point in $P(r,\mathbb{C})$ denoted by $\hat{x}$ and we also know that

$$\hat{x} = f^{-1}\left(\frac{\sum_{i=1}^{n} f(x_i)}{n}\right). \tag{7.85}$$

We will need a similar theorem to Proposition 7.17.

Proposition 7.24 (M. Pálfia [64]). *Let $X_i^0 \in P(r,\mathbb{C})$ for $i=1,\dots,n$. Then the BMP-procedure (Definition 4.5) set up on the matrices $X_1^0,\dots,X_n^0$ with the weighted mean function $M_t(A,B) := f^{-1}((1-t)f(A)+tf(B))$ and the $n-1$ variable function $M(x_1,\dots,x_{n-1}) := f^{-1}\left(\frac{\sum_{i=1}^{n-1} f(x_i)}{n-1}\right)$ leaves the Riemann centroid of the points $X_1^0,\dots,X_n^0$ invariant with respect to the distance function (7.42).*

Proof.

$$\begin{aligned}&f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^1\right)}{n}\right) = f^{-1}\left[\sum_{i=1}^{n}\frac{f\left(M_{\frac{n-1}{n}}\left(X_i^0, M\left(Z_{\neq i}(X_1^0,\dots,X_n^0)\right)\right)\right)}{n}\right] = \\ &= f^{-1}\left[\sum_{i=1}^{n}\frac{\frac{f(X_i^0)}{n}+\frac{n-1}{n}\sum_{j=1,j\neq i}^{n}\frac{f(X_j^0)}{n-1}}{n}\right] = f^{-1}\left(\frac{\sum_{i=1}^{n} f\left(X_i^0\right)}{n}\right)\end{aligned} \tag{7.86}$$

Similarly we obtain the above equality for every iteration step, so

$$f^{-1}\left(\frac{\sum_{i=1}^n f\left(X_i^{l+1}\right)}{n}\right) = f^{-1}\left(\frac{\sum_{i=1}^n f\left(X_i^{l}\right)}{n}\right) = f^{-1}\left(\frac{\sum_{i=1}^n f\left(X_i^{0}\right)}{n}\right). \tag{7.87}$$

□

Proof. (Theorem 7.23) The proof again will be based on induction on the number of matrices n. We will use the same notations to denote the sum of the squared distances of the matrices X_i^l from the zero matrix with respect to the distance function (7.39), so

$$a_n^l = \sum_{i=1}^n d_E(0, X_i^l)^2 = \sum_{i=1}^n Tr\left\{(X_i^l)^2\right\} \tag{7.88}$$

$$e_n^l = \sum_{1\le i<j\le n} d_E(X_i^l, X_j^l)^2 = \sum_{1\le i<j\le n} Tr\left\{(X_i^l - X_j^l)^2\right\}. \tag{7.89}$$

We will denote by $F(X_1^0, \ldots, X_n^0)$ the common limit point of the sequences X_i^l for n. The proof will rely on similar three lemmas to the ones in the proof of the ALM-procedure. First we are going to treat the case of the first inequality (7.82).

Lemma 7.25 (Monotone Iteration, M. Pálfia [64])**.** *In the first case of inequality* (7.82) *we have*

$$\left(\frac{\sum_{i=1}^n \left(X_i^{l+1}\right)^{-1}}{n}\right)^{-1} \ge \left(\frac{\sum_{i=1}^n \left(X_i^{l}\right)^{-1}}{n}\right)^{-1}. \tag{7.90}$$

In the second case of inequality (7.83) *we have*

$$\frac{\sum_{i=1}^n X_i^{l+1}}{n} \le \frac{\sum_{i=1}^n X_i^{l}}{n}. \tag{7.91}$$

Proof. The proof uses similar ideas to the case of Lemma Monotone Iteration for the ALM-process. We again argue by induction on the number of matrices n. Consider the first case of inequality (7.82). Suppose that the BMP-procedure converges for $n \ge 3$ to common limit point $F(X_1^0, \ldots, X_n^0)$ in other words $X_i^l \to F(X_1^0, \ldots, X_n^0)$ for n. Also the inequality of the lemma for n implies that

$$\left(\frac{\sum_{i=1}^n \left(X_i^{l}\right)^{-1}}{n}\right)^{-1} \ge \left(\frac{\sum_{i=1}^n \left(X_i^{0}\right)^{-1}}{n}\right)^{-1} \tag{7.92}$$

and if we take the limit on the left hand side for n we get the inequality

$$\left(\frac{\sum_{i=1}^n \left(F(X_1^0, \ldots, X_n^0)\right)^{-1}}{n}\right)^{-1} = F(X_1^0, \ldots, X_n^0) \ge \left(\frac{\sum_{i=1}^n \left(X_i^{0}\right)^{-1}}{n}\right)^{-1}. \tag{7.93}$$

The above inequality also holds for $n = 2$ by the assumption of inequality (7.82), so this provides the first step for $n = 3$ in our induction.

Now we prove the lemma for $n+1$ if it is true for n. Similarly to the case of Lemma Monotone Iteration in the ALM process, we make use of the operator monotonicity of the $n+1$-variable harmonic mean, and use (7.93). Then we use Proposition 7.24 with $f(t) = t^{-1}$ and the same argument as in the ALM case, performed using instead one BMP iteration step, yields

$$\left(\frac{\sum_{i=1}^{n+1}\left(X_i^{l+1}\right)^{-1}}{n+1}\right)^{-1} \geq \left(\frac{\sum_{i=1}^{n+1}\left(X_i^{l}\right)^{-1}}{n+1}\right)^{-1}. \tag{7.94}$$

The second case given by inequality (7.83) again can be treated similarly to the ALM-case. □

Lemma 7.26 (Decreasing Distances, M. Pálfia [64]). *We have*

$$a_n^{l+1} \leq a_n^l - \frac{k}{8} z_n e_n^l, \tag{7.95}$$

in the case of (7.82), *or we have* (7.95) *with* $k = 2$ *in the case of* (7.83). *In both cases* $z_n = \frac{4}{(n-1)n}$.

Proof. The situation is similar again to the ALM case. We will prove this for the case (7.82), the second case of (7.83) is just the same with $k = 2$, since we will only use that the right hand side of (7.37) holds, so we do not have to treat the second case (7.83) separately due to (7.47). The first step is to show the above for $n = 3$. By Lemma 7.19 and that the right hand side of (7.82) is equivalent to the assumption of the lemma, we get

$$\begin{aligned} d_E\left(0, F(X_i^l, X_j^l)\right)^2 \leq& \frac{d_E(0, X_i^l)^2 + d(0, X_j^l)^2}{2} - \frac{k}{8} d_E(X_i^l, X_j^l)^2 \\ d_E\left(0, F_{2/3}\left(X_s^l, F(X_i^l, X_j^l)\right)\right)^2 \leq& \frac{1}{3} d_E(0, X_s^l)^2 + \frac{2}{3} d(0, F(X_i^l, X_j^l))^2 \\ &- \frac{k}{2}\frac{1}{3}\frac{2}{3} d_E(X_s^l, F(X_i^l, X_j^l))^2 \\ \leq& \frac{1}{3} d_E(0, X_s^l)^2 + \frac{2}{3} d(0, F(X_i^l, X_j^l))^2 \\ \leq& \frac{d_E(0, X_s^l)^2 + d_E(0, X_i^l)^2 + d(0, X_j^l)^2}{3} \\ &- \frac{k}{8}\frac{2}{3} d_E(X_i^l, X_j^l)^2 \end{aligned} \tag{7.96}$$

in other words the last inequality is equivalent to

$$d_E\left(0, X_s^{l+1}\right)^2 \leq \frac{d_E(0, X_s^l)^2 + d_E(0, X_i^l)^2 + d(0, X_j^l)^2}{3} - \frac{k}{8}\frac{2}{3} d_E(X_i^l, X_j^l)^2, \tag{7.97}$$

where $i, j, s \in \{1,2,3\}$ and $i \neq j \neq s, s \neq i$. There are 3 distinct inequalities of the above for $s = 1, 2, 3$. By summing these inequalities for s we get (7.95) for $n = 3$ and $z_3 = \frac{2}{3}$.

Now suppose (7.95) holds for n and that X_i^l converge to a common limit point for n. Then we have

$$a_n^l \leq a_n^0 - \frac{k}{8} z_n e_n^0 \tag{7.98}$$

and by taking the limit on the left hand side we get

$$\begin{aligned} \lim_{l\to\infty} a_n^l = n d_E\left(0, F(X_1^0, \ldots, X_n^0)\right)^2 &\leq a_n^0 - \frac{k}{8} z_n e_n^0 \\ d_E\left(0, F(X_1^0, \ldots, X_n^0)\right)^2 &\leq \frac{a_n^0}{n} - \frac{k}{8}\frac{z_n}{n} e_n^0. \end{aligned} \tag{7.99}$$

Then set up the BMP-procedure on $X_i^0 \in P(r, \mathbb{C}); i = 1, 2 \ldots, n+1$ with $M_n(X_1, \ldots, X_n) := F(X_1, \ldots, X_n)$. Inequality (7.99) can be applied in any of the iteration steps, so we get

$$\begin{aligned} & d_E\left(0, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)^2 \leq \\ & \leq \frac{\sum_{j=1, j\neq i}^{n+1} d_E(0, X_j^l)^2}{n} - \frac{k}{8}\frac{z_n}{n} \sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2. \end{aligned} \tag{7.100}$$

Then we have to compute $X_i^{l+1} = F_{\frac{n}{n+1}}\left(X_i^l, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)$ and bound its squared distance from the zero matrix

$$\begin{aligned} d_E\left(0, F_{\frac{n}{n+1}}\left(X_i^l, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)\right)^2 &\leq \frac{1}{n+1} d_E\left(0, X_i^l\right)^2 \\ &+ \frac{n}{n+1} d_E\left(0, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)^2 \\ -\frac{k}{8}\frac{1}{n+1}\frac{n}{n+1} d_E\left(X_i^l, F_{\frac{n}{n+1}}\left(X_i^l, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)\right)^2&. \end{aligned} \tag{7.101}$$

We drop the last term, as it seems that it is hard to estimate it from below, and substitute in using inequality (7.100), we get

$$\begin{aligned} & d_E\left(0, X_i^{l+1}\right)^2 \leq \frac{1}{n+1} d_E\left(0, X_i^l\right)^2 + \frac{n}{n+1} d_E\left(0, F\left(Z_{\neq i}\left(X_1^l, \ldots, X_{n+1}^l\right)\right)\right)^2 \\ & \leq \frac{1}{n+1} d_E\left(0, X_i^l\right)^2 \\ & + \frac{n}{n+1}\left[\frac{\sum_{j=1, j\neq i}^{n} d_E(0, X_j^l)^2}{n} - \frac{k}{8}\frac{z_n}{n} \sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2\right] \\ & \leq \frac{\sum_{j=1, j\neq i}^{n+1} d_E(0, X_j^l)^2}{n+1} - \frac{k}{8}\frac{z_n}{n+1} \sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2. \end{aligned} \tag{7.102}$$

If we sum these inequalities for i we arrive at the following

$$\begin{aligned} &\sum_{i=1}^{n+1} d_E\left(0, X_i^{l+1}\right)^2 \leq \\ &\leq \sum_{i=1}^{n+1} \frac{\sum_{j=1}^{n+1} d_E(0, X_j^l)^2}{n+1} - \frac{k}{8}\frac{z_n}{n+1}\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2, \end{aligned} \tag{7.103}$$

which is equivalent to

$$\begin{aligned} &\sum_{i=1}^{n+1} d_E\left(0, X_i^{l+1}\right)^2 \leq \\ &\leq \sum_{i=1}^{n+1} d_E(0, X_i^l)^2 - \frac{k}{8}\frac{z_n}{n+1}\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2. \end{aligned} \tag{7.104}$$

The left hand side of the above is just a_{n+1}^{l+1} and the first term on the right hand side is a_{n+1}^l. By the proof of the convergence of the ALM-process the second term

$$\sum_{i=1}^{n+1}\sum_{1\leq j<s\leq n+1, j\neq i, s\neq i} d_E(X_j^l, X_s^l)^2 = (n-1)e_{n+1}^l. \tag{7.105}$$

Thus $z_{n+1} = \frac{n-1}{n+1}z_n$ and also $z_3 = \frac{2}{3}$, by solving the recursion we get

$$z_n = \frac{4}{(n-1)n}. \tag{7.106}$$

This concludes the lemma for every n. □

Lemma 7.27 (Boundedness, M. Pálfia [64])**.** *The matrix sequences X_i^l are bounded for all n.*

Proof. We have the trivial lower bound $X_i^l \geq 0$, since by assumption $F_t(A, B) \geq 0$, so we have $X_i^l \geq 0$ for $n = 3$. Now similarly to the case of the ALM-process we assume again that the BMP-procedure converges to common limit point denoted by $F(X_1^0, \ldots, X_n^0)$ for n and $F(X_1^0, \ldots, X_n^0) \geq 0$. Then again if the sequences converge for $n+1$ to a common limit $F(X_1^0, \ldots, X_{n+1}^0)$, then this limit $F(X_1^0, \ldots, X_{n+1}^0) \geq 0$.

We provide the suitable upper bound similarly to the ALM case. We again have $a_{n+1}^l \leq a_{n+1}^0 = b$ so we similarly get $\left\|X_i^l\right\| \leq b$ for $n+1$ if the procedure converges to common limit for n. This finishes the proof of the lemma. □

Now the last step of the proof is exactly the same as in the case of the ALM-procedure, we just have to use the three lemmas: Lemma Monotone Iteration, Decreasing Distances and Boundedness adapted for the case of the BMP iteration.

□

Notice that for matrix means we have the weighted mean procedure $M_t(A,B)$ introduced here. By Proposition 7.13 we have that every such mean is smaller than the weighted arithmetic mean and larger than the weighted harmonic mean. So as a consequence of Theorem 7.23 we get that the BMP-procedure converges for every symmetric matrix mean if we identify their weighted counterparts with our weighted mean $M_t(A,B)$.

7.5 Properties of the ALM and BMP mean

We will show that the limit point of the ALM and BMP processes, denoted by $M_{ALM}(X_1,\ldots,X_n)$ and $M_{BMP}(X_1,\ldots,X_n)$ respectively, as extensions of symmetric matrix means, fulfill the following properties.

Theorem 7.28 (M. Pálfia [64]). *If $M(A,B)$ is a symmetric matrix mean, then the $M := M_{ALM}(X_1,\ldots,X_n)$ and $M := M_{BMP}(X_1,\ldots,X_n)$ extensions fulfill the following properties*

(I) $M(X,\ldots,X) = X$ *for every* $X \in P(r,\mathbb{C})$*,*

(II) $M(X_1,\ldots,X_n)$ *is invariant under the permutation of its variables,*

(III) $\min(X_1,\ldots,X_n) \le M(X_1,\ldots,X_n) \le \max(X_1,\ldots,X_n)$ *if* min *and* max *exist with respect to the positive definite order,*

(IV) If $X_i \le X'_i$*, then* $M(X_1,\ldots,X_n) \le M(X'_1,\ldots,X'_n)$*,*

(V) $M(X_1,\ldots,X_n)$ *is continuous,*

(VI) $M(CX_1C^*,\ldots,CX_nC^*) = CM(X_1,\ldots,X_n)C^*$ *for all invertible* C*.*

Proof. The proof of each property will be based on induction. Each of them trivially holds for $n=2$ by properties of matrix means. So it remains to prove them for $n+1$ assuming that they hold for n.

Property (I) and (II) trivially holds for $n+1$ if it holds for n. We prove property (IV). Let $X_1^0,\ldots,X_{n+1}^0 \in P(r,\mathbb{C})$ and $X_i^0 \le (X'_i)^0 \in P(r,\mathbb{C})$. If we iterate by the ALM process, it is easy to see that the order $X_i^0 \le (X'_i)^0$ is preserved due to the inductional hypothesis on property (IV), so $X_i^l \le (X'_i)^l$. Taking the limits $l \to \infty$ we get the assertion. In case of the BMP-process the argument is similar but we have to use also that $M_t(A,B) \le M_t(A',B')$ if $A \le A'$ and $B \le B'$.

Property (III) is an easy consequence of property (I) and (IV), if minimum and maximum exist. Setting up the same iteration on the new n-tuple formed by the minimal element we get the inequality on the left in property (III), similarly we can obtain the inequality on the right as well.

To prove property (VI) let $(X'_i)^0 = CX_i^0C^*$ and set up the ALM or BMP process on $(X'_1)^0,\ldots,(X'_n)^0$ as on $X_1^0,\ldots,X_n^0$. Property (VI) implies in the case of ALM

$$\begin{aligned} CX_i^{l+1}C^* &= CM\left(Z_{\neq i}\left(X_1^l,\ldots,X_{n+1}^l\right)\right)C^* = \\ &= M\left(Z_{\neq i}\left(CX_1^lC^*,\ldots,CX_{n+1}^lC^*\right)\right). \end{aligned} \tag{7.107}$$

In the case of the BMP process we have similarly

$$
\begin{aligned}
CX_i^{l+1}C^* &= CM_{\frac{n}{n+1}}\left(X_i^l, M\left(Z_{\neq i}\left(X_1^l,\dots,X_{n+1}^l\right)\right)\right)C^* = \\
&= M_{\frac{n}{n+1}}\left(CX_i^lC^*, M\left(Z_{\neq i}\left(CX_1^lC^*,\dots,CX_{n+1}^lC^*\right)\right)\right).
\end{aligned}
\tag{7.108}
$$

Applying the above recursively in every iteration step we get

$$CX_i^lC^* = (X_i')^l. \tag{7.109}$$

Taking the limit $l\to\infty$ the assertion follows.

Property (V) is a consequence of properties (IV) and (VI) by Lemma 7.9. □

Now we can see that the assumed properties in Definition 4.1 are fulfilled in general by these two means as well.

We also have that the ALM and BMP procedures preserve the ordering of functions. So we have for the ALM process the following

Proposition 7.29 (M. Pálfia [64])**.** *If $M(A,B)\leq N(A,B)$ are functions satisfying the properties of $F(A,B)$ in Theorem 7.16, then the same ordering is true for the ALM limit points $M(X_1,\dots,X_n)$ and $N(X_1,\dots,X_n)$.*

Proof. Again we argue by induction. The inequality

$$M(X_1,\dots,X_n)\leq N(X_1,\dots,X_n) \tag{7.110}$$

holds for $n=2$ by assumption. Let us denote the matrices in the ALM iteration steps performed with $M(X_1,\dots,X_{n-1})$ and $N(X_1,\dots,X_{n-1})$ on $X_1^0,\dots,X_n^0\in P(r,\mathbb{C})$ by X_i^l and $(X_i')^l$ respectively. Now again we have $M(X_1,\dots,X_{n-1})\leq N(X_1,\dots,X_{n-1})$ by the inductional hypothesis so we have $X_i^l\leq (X_i')^l$. Taking the limits we get the assertion. □

A similar, although a bit different assertion holds for the BMP process.

Proposition 7.30 (M. Pálfia [64])**.** *If $M_t(A,B)\leq N_t(A,B)$ are functions satisfying the properties of $F_t(A,B)$ in Theorem 7.23, then the same ordering is true for the BMP limit points $M(X_1,\dots,X_n)$ and $N(X_1,\dots,X_n)$.*

Proof. We have an inductional argument similarly to the preceding case of the ALM process. The inequality $M(A,B)\leq N(A,B)$ holds for by assumption since $M(A,B)=M_{1/2}(A,B)$ and $N(A,B)=N_{1/2}(A,B)$. Let us denote the matrices in the BMP iteration steps performed with $M(X_1,\dots,X_{n-1})$ and $N(X_1,\dots,X_{n-1})$ on $X_1^0,\dots,X_n^0\in P(r,\mathbb{C})$ by X_i^l and $(X_i')^l$ respectively. Now again we have $M(X_1,\dots,X_{n-1})\leq N(X_1,\dots,X_{n-1})$ by the inductional hypothesis and also $M_t(A,B)\leq N_t(A,B)$ so we have $X_i^l\leq (X_i')^l$. Taking the limits we get the assertion. □

In the next section we will consider some convergence rate properties fulfilled by the BMP process.

7.6 Convergence rate of the BMP process

A nice property of the BMP process, discussed in Theorem 4.16, is its cubic convergence rate in a small neighborhood of its limit point. This is an advantage over the ALM process which is known to converge linearly. In this section we will show that the BMP process generally converges cubically for every possible matrix mean in a small neighborhood of the limit point of the process. The proof will be similar to the one presented in [15]. In order to be able to use such an argument we have to obtain a series expansion for the weighted mean $M_t(A, B)$ in the neighborhood of the identity matrix I.

We are going to use the big O notation. This means that we have $X = Y + O(\epsilon^k)$ if and only if there exist constants $\epsilon_0 < 1$ and θ such that for each $0 < \epsilon < \epsilon_0$ we have $\|X - Y\| \leq \theta\epsilon^k$.

Proposition 7.31 (M. Pálfia [64])**.** *Let $M(A, B)$ be a symmetric matrix mean and $f(t)$ be its corresponding normalized operator monotone function. Let $f(t)$ have a series expansion around I as*

$$f(X) = I + \frac{X - I}{2} + \sum_{k=2}^{\infty} b_k (X - I)^k. \tag{7.111}$$

Then we have a series expansion for $M_t(I, X) = f_t(X)$ whenever $\|X - I\| \leq \epsilon < 1$ in the form

$$f_t(X) = I + t(X - I) + 4b_2 t(1 - t)(X - I)^2 + O(\epsilon^3). \tag{7.112}$$

Proof. Since $M_t(A, B)$ is a matrix mean if the generating $M(A, B)$ is a symmetric matrix mean therefore it has the following representation

$$M_t(A, B) = A^{1/2} f_t \left(A^{-1/2} B A^{-1/2}\right) A^{1/2} = A f_t \left(A^{-1} B\right), \tag{7.113}$$

where $f_t(X)$ is a normalized operator monotone function in X, therefore analytic on $(0, \infty)$, hence we have the second equality as well. Since it is generally analytic only on $(0, \infty)$, we expect (7.111) to be convergent only for $\|X - I\| < 1$. By the above representation for $M_t(A, B)$ and the fact that $M_{1/2}(A, B) = M(A, B)$ by definition, it is enough to show that the expansion in the assertion holds for $f_t(X)$. In other words we have to consider the mean $M_t(A_0, B_0)$ of $A_0 = I$ and an arbitrary $B_0 = X$. We also have a natural expansion in the neighborhood of I for the inverse function as

$$X^{-1} = \sum_{k=0}^{\infty} (-1)^k (X - I)^k, \tag{7.114}$$

which is convergent for $\|X - I\| < 1$. Now in every step of the Weighted mean process we have to compute a symmetric mean of two matrices and by the assumption of the assertion we have $\|X - I\| \leq \epsilon < 1$. Without loss of generality we may write A_j and B_j in the following forms

$$\begin{aligned} A_j &= I + y_1^j (X - I) + y_2^j (X - I)^2 + O(\epsilon^3) \\ B_j &= I + z_1^j (X - I) + z_2^j (X - I)^2 + O(\epsilon^3). \end{aligned} \tag{7.115}$$

Now we will make use of the above expansions to get an expansion for $M(A_j, B_j)$ up to the $O(\epsilon^3)$ term as follows

$$M(A_j, B_j) = A_j f(A_j^{-1}B_j) = \left(I + y_1^j(X-I) + y_2^j(X-I)^2 + O(\epsilon^3)\right)$$
$$f\left[A_j^{-1}\left(I + z_1^j(X-I) + z_2^j(X-I)^2 + O(\epsilon^3)\right)\right] =$$
$$= \left(I + y_1^j(X-I) + y_2^j(X-I)^2 + O(\epsilon^3)\right) f\left[\left(I - y_1^j(X-I) + \right.\right.$$
$$\left.\left. +((y_1^j)^2 - y_2^j)(X-I)^2 + O(\epsilon^3)\right)\left(I + z_1^j(X-I) + z_2^j(X-I)^2 + O(\epsilon^3)\right)\right] \tag{7.116}$$

where we have used (7.114) to express A_j^{-1} and (7.111) to express $f(X)$ up to $O(\epsilon^3)$ terms. After some calculation and taking into account that the terms $(X-I)^k$ with $k \geq 3$ are of $O(\epsilon^3)$, we get that

$$M(A_j, B_j) = I + \frac{y_1^j + z_1^j}{2}(X-I) + \left[\frac{y_2^j + z_2^j}{2} + b_2(y_1^j - z_1^j)^2\right](X-I)^2 + O(\epsilon^3). \tag{7.117}$$

Note that since $A_0 = I$ and $B_0 = X$ we have $y_2^0 = 0$ and $z_2^0 = 0$. Hence it can be easily proved by induction that the terms

$$\begin{aligned} y_2^j &= b_2 p_j(y_1^0, z_1^0, t) \\ z_2^j &= b_2 q_j(y_1^0, z_1^0, t), \end{aligned} \tag{7.118}$$

where p_j and q_j are functions which do not depend on b_2. Also since $f_t(X)$ is an analytic function due to Kubo-Ando theory, therefore the limits $p = \lim_{j\to\infty} p_j = \lim_{j\to\infty} q_j$ and this limit function is also independent of b_2.

Now since the weighted geometric mean

$$G_t(A, B) = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2} = A\left(A^{-1}B\right)^t \tag{7.119}$$

is an affine mean, therefore the Weighted mean process gives back $G_t(A, B)$ for every $t \in [0, 1]$ and $A, B \in P(r, \mathbb{C})$. In other words in this case if we expand the function X^t into a Taylor series around I we get

$$G_t(I, X) = X^t = I + t(X-I) - \frac{t(1-t)}{2}(X-I)^2 + O(\epsilon^3) \tag{7.120}$$

and this equation for $t = 1/2$ gives that $b_2 = -1/8$. Since $G_t(A, B)$ is an affine mean and p does not explicitly depend on b_2 we get $p = 4t(1-t)$. Similar consideration can be applied in the case of the linear term $t(X-I)$. □

The above proposition tells us that no matter how we choose the symmetric matrix mean $M(A, B)$, the series expansion of $M_t(A, B)$ around I will have similar structure up to the $(X-I)^2$ term:

$$M_t(A, B) = (1-t)A + tB + 4b_2 t(1-t)A^{1/2}\left(A^{-1/2}BA^{-1/2} - I\right)^2 A^{1/2} + \dots \tag{7.121}$$

Actually $b_2 \leq 0$ for every matrix mean since the corresponding operator monotone function is concave. It is not hard to prove using Proposition 7.15 that $-1/2 \leq b_2 \leq 0$ for every matrix mean.

A remarkable consequence of the above series expansion is that the BMP procedure converges at least cubically if the matrices are sufficiently close to each other for all symmetric matrix means.

To prove this assertion we will prove the following more precisely formulated

Theorem 7.32 (M. Pálfia [64])**.** *Let $0 < \epsilon < 1$, $K, X_i^0 \in P(r, \mathbb{C}), i = 1, \ldots, n$, and $E_i = K^{-1}X_i^0 - I$. Set up the the BMP process on the X_i^0 with a symmetric matrix mean $M(A, B)$ which has series expansion given as in Proposition 7.31. Now suppose that $\|E_i\| \leq \epsilon$ for all i. Then for the matrices X_i^1 the following hold.*

C1: We have

$$K^{-1}X_i^1 - I = T_n + O(\epsilon^3), \tag{7.122}$$

where

$$T_n = \frac{1}{n}\sum_{j=1}^{n} E_j + \frac{2b_2}{n^2}\sum_{i,j=1}^{n}(E_i - E_j)^2. \tag{7.123}$$

C2: There are positive constants θ, σ and $\bar{\epsilon} < 1$ (all of which may depend on n) such that for all $\epsilon \leq \bar{\epsilon}$,

$$\left\|K_1^{-1}X_i^1 - I\right\| \leq \theta\epsilon^3 \tag{7.124}$$

for a suitable matrix K_1 satisfying $\left\|K^{-1}K_1 - I\right\| \leq \sigma\epsilon$.

C3: The BMP iteration in Definition 4.5 converges at least cubically.

C4: We have

$$K_1^{-1}M_{BMP}(X_1^0, \ldots, X_n^0) - I = O(\epsilon^3), \tag{7.125}$$

where $M_{BMP}(X_1^0, \ldots, X_n^0)$ denotes the limit point of the BMP-process in Definition 4.5.

Proof. Let $A, B \in P(r, \mathbb{C})$ such that $K^{-1}A = I + F_1$ and $K^{-1}B = I + F_2$ and $\|F_1\| \leq \delta$, $\|F_2\| \leq \delta$, $0 < \delta < 1$. Then we have by the series expansion in Proposition 7.31 that

$$\begin{aligned}
K^{-1}M_t(A, B) &= K^{-1}Af_t(A^{-1}B) = (I + F_1)f_t\left((I + F_1)^{-1}(I + F_2)\right)\\
&= (I + F_1)f_t\left((I - F_1 + F_1^2 + O(\delta^3))(I + F_2)\right)\\
&= (I + F_1)f_t\left(I + F_2 - F_1 - F_1F_2 + F_1^2 + O(\delta^3)\right)\\
&= I + (1 - t)F_1 + tF_2 + 4b_2t(t - 1)(F_2 - F_1)^2 + O(\delta^3).
\end{aligned} \tag{7.126}$$

Now we will prove the theorem by induction on n. Let Ci_n denote the assertion Ci of the theorem ($i = 1, 2, 3, 4$) for a given value of n. Then we show

(1) $C1_2$ holds,

(2) $C1_n \Rightarrow C2_n$,

(3) $C2_n \Rightarrow C3_n, C4_n$,

(4) $C4_n \Rightarrow C1_{n+1}$.

(1): This is equation (7.126) for $t = 1/2$.
(2): It is obvious that $T_n = O(\epsilon)$, so choosing $K_1 = K(I + T_n)$ we have

$$X_i^1 = K(I + T_n + O(\epsilon^3)) = K_1(I + (I + T_n)^{-1}O(\epsilon^3)) = K_1(I + O(\epsilon^3)). \quad (7.127)$$

Using explicit constants in the big-O estimates, we get

$$\left\|K_1^{-1}X_i^1 - I\right\| \le \theta\epsilon^3, \left\|K^{-1}K_1 - I\right\| \le \sigma\epsilon \quad (7.128)$$

for some constants θ, σ.

(3): Suppose now that ϵ is small enough to have $\theta\epsilon^3 \le \epsilon$ and let $\epsilon_1 = \theta\epsilon^3$. Now we apply $C2$ with starting matrices X_i^1, with ϵ_1 replacing ϵ and K_1 replacing K, getting

$$\left\|K_2^{-1}X_i^1 - I\right\| \le \theta\epsilon_1^3, \left\|K_1^{-1}K_2 - I\right\| \le \sigma\epsilon_1. \quad (7.129)$$

Repeating this for all steps of the iterative process, we get for all $l = 0, 1, \dots$ that

$$\left\|K_l^{-1}X_i^l - I\right\| \le \theta\epsilon_{l-1}^3 = \epsilon_l, \left\|K_l^{-1}K_{l+1} - I\right\| \le \sigma\epsilon_l \quad (7.130)$$

with $\epsilon_{l+1} = \theta\epsilon_l^3$.

Let us introduce the notation

$$d(X, Y) = \left\|X^{-1}Y - I\right\| \quad (7.131)$$

for any $X, Y \in P(r, \mathbb{C})$. Notice that

$$\|X - Y\| \le \|X\| \left\|X^{-1}Y - I\right\| \le \|X\| \, d(X, Y). \quad (7.132)$$

Similarly we also have

$$\begin{aligned} d(X, Z) &= \left\|(X^{-1}Y - I)(Y^{-1}Z - I) + X^{-1}Y - I + Y^{-1}Z - I\right\| \\ &\le d(X, Y)d(Y, Z) + d(X, Y) + d(Y, Z). \end{aligned} \quad (7.133)$$

Using the introduced notation we have according to (7.130) that

$$d(K_l, X_i^l) \le \epsilon_l, d(K_l, K_{l+1}) \le \sigma\epsilon_l. \quad (7.134)$$

Now we will show by induction that, for ϵ smaller than a fixed constant, it follows that

$$d(K_l, K_{l+t}) \le \left(2 - \frac{1}{2^t}\right)\sigma\epsilon_l. \quad (7.135)$$

First of all, for all $t \ge 1$

$$\epsilon_{l+t} = \theta^{\frac{3^t - 1}{2}}\epsilon^{3^t}, \quad (7.136)$$

which, for ϵ smaller than $\min(1/8, \theta^{-1})$, implies

$$\frac{\epsilon_{l+t}}{\epsilon_l} \leq \epsilon_l^{\frac{3^t-1}{2}} \leq \frac{1}{2^{t+2}}. \tag{7.137}$$

Let us suppose that $\epsilon \leq \sigma^{-1}$. Then by (7.133) we have

$$\begin{aligned} d(K_l, K_{l+t+1}) \leq & d(K_l, K_{l+t}) d(K_{l+t}, K_{l+t+1}) \\ & + d(K_l, K_{l+t}) + d(K_{l+t}, K_{l+t+1}) \\ \leq & \left(2 - \frac{1}{2^t}\right) \sigma\epsilon_l + \sigma\epsilon_l \left(\sigma\epsilon_{l+t} - \frac{\epsilon_{l+t}}{\epsilon_l}\right) \\ \leq & \left(2 - \frac{1}{2^t}\right) \sigma\epsilon_l + \sigma\epsilon_l \frac{1}{2^{t+1}} = \left(2 - \frac{1}{2^{t+1}}\right) \sigma\epsilon_l. \end{aligned} \tag{7.138}$$

So we have for each $t \leq 0$ that

$$\|K_t - K\| \leq \|K\| \left\|K^{-1}K_t - I\right\| \leq 2\sigma \|K\| \epsilon, \tag{7.139}$$

which yields $\|K_t\| \leq 2 \|K\|$ for all t. By a similar argument we have

$$\|K_{l+t} - K_l\| \leq \|K_l\| d(K_{l+t}, K_l) \leq 2\sigma \|K\| \epsilon_l. \tag{7.140}$$

What follows here from the bounds already imposed on ϵ, the sequence ϵ_l tends monotonically to zero with cubic convergence rate, thus K_l is a Cauchy sequence as well and therefore converges. Let us denote its limit point with K^*. K_l also converges cubically since if we let $t \to \infty$ in the above inequality, we get

$$\|K^* - K_l\| \leq 2\sigma \|K\| \epsilon_l. \tag{7.141}$$

Now using the relations in (7.130), we get

$$\begin{aligned} \left\|X_i^l - K^*\right\| & \leq \left\|X_i^l - K_l\right\| + \|K_l - K^*\| \\ & \leq 2 \|K\| d(K_l, X_i^l) + 2\sigma \|K\| \epsilon_l \\ & \leq (2\sigma + 2) \|K\| \epsilon_l, \end{aligned} \tag{7.142}$$

which means that X_i^l converges to K^* with cubic convergence rate, so $C3$ is proved. By (7.130), (7.133) and (7.135) we get that

$$\begin{aligned} d(K_1, X_i^t) & \leq d(K_1, K_t) d(K_t, X_i^t) + d(K_1, K_t) + d(K_t, X_i^t) \\ & \leq 2\sigma\epsilon_1\epsilon_t + 2\sigma\epsilon_1 + \epsilon_t \leq (4\sigma + 1)\epsilon_1 = O(\epsilon^3), \end{aligned} \tag{7.143}$$

which is $C4$.

(4): Using $C4_n$ and (7.126) with $F_1 = E_{n+1}, F_2 = K^{-1}M_{BMP}(X_1^0, \ldots, X_n^0) = T_n + O(\epsilon^3), \delta = 2n\epsilon$, we have

$$\begin{aligned} K^{-1}X_{n+1}^1 = & K^{-1}\left(M_{\frac{n}{n+1}}(X_{n+1}^0, M_{BMP}(X_1^0, \ldots, X_n^0))\right) \\ = & I + \frac{1}{n+1}E_{n+1} + \frac{n}{n+1}T_n \\ & + \frac{4b_2 n}{(n+1)^2}\left(E_{n+1} - \frac{1}{n}\sum_{i=1}^{n} E_i\right)^2 + O(\epsilon^3). \end{aligned} \tag{7.144}$$

Notice that

$$T_n = \frac{1}{n} S_n + 4b_2 \frac{(n-1)Q_n - P_n}{n^2} \tag{7.145}$$

where $S_n = \sum_{i=1}^n E_i$, $Q_n = \sum_{i=1}^n E_i^2$, $P_n = \sum_{i,j=1,i\neq j}^n E_i E_j$. We also have $S_n^2 = P_n + Q_n$ and $S_{n+1} = S_n + E_{n+1}$, $Q_{n+1} = Q_n + E_{n+1}^2$, $P_{n+1} = P_n + E_{n+1} S_n + S_n E_{n+1}$, which yields according to the above that

$$\begin{aligned} K^{-1} X_{n+1}^1 &= I + \frac{1}{n+1} S_{n+1} + \frac{4b_2 n}{(n+1)^2} Q_{n+1} - \frac{4b_2}{2(n+1)^2} P_{n+1} + O(\epsilon^3) \\ &= I + T_{n+1} + O(\epsilon^3). \end{aligned} \tag{7.146}$$

Since this expression is symmetric with respect to E_i, it follows that it is the same for all X_i^1. □

The above proof is almost the same as the one presented for a similar theorem in [15] for the geometric mean. The major differences are in the series expansions for the weighted mean $M_t(A, B)$ and the limit point $M_{BMP}(X_1, \ldots, X_n)$.

8 Practical Applications

Calculation of means naturally appear in practical applications when smoothing of multisampled data is needed. The arithmetic mean plays a central role in statistics and the arithmetic mean of matrices appears in multivariate statistics. Probability theory and statistics also benefits from certain inequalities between classical means. For instance the inequality between the multi-variable forms of the arithmetic and the square mean can be used to prove that a certain random variable that has vanishing variance must have Dirac-delta distribution, i. e. the support of its distribution function is a point.

We will consider two areas of possible practical applications. One of them is the problem of calculating averages of points in a complete metric space, the other is calculating averages of positive definite matrices, which itself in a way is a subcase of the former problem.

8.1 Averaging in complete metric spaces

In many applications, such as the study of plate tectonics [70] or sequence-dependent continuum modeling of DNA [45], the experimental data is given as a sequence of three dimensional orentation data. This set of data is usually thought of as three dimensional orthogonal matrices, i. e. elements of the group $SO(3)$. This group is a connected, simply connected compact Lie group, therefore also a Riemannian manifold with a bi-invariant metric. The curvature of the manifold in the case of $SO(3)$ is constant, its value is $1/4$ while in the higher dimensional cases it varies over the manifold. As we have mentioned before, by Proposition 5.12 and 5.13 of Ohta we already know that $SO(r)$ is a complete, locally k-convex metric space.

To do averaging in $SO(r)$, one can use the fact that this manifold is automatically embended into the space of squared complex matrices. So if we have $Q_i \in SO(r)$ for $i = 1, \ldots, n$, we may take the arithmetic mean of Q_i, although then we will end up with a matrix that is not an element of $SO(r)$. One possibility is to orthogonally project back to the manifold $SO(r)$. This method is widely used although has several major drawbacks. For instance we may end up with a projected element of $SO(r)$ that is not in the convex hull of the initial data with respect to the Riemannian structure of $SO(r)$. Moreover if we consider this method on $P(r, \mathbb{C})$ with Riemannian metric (4.2), we may end up with points which are on the boundary of the cone $P(r, \mathbb{C})$, although any boundary point is of infinite Riemannian distance from any inner points of the cone $P(r, \mathbb{C})$. This is a serious problem which is discussed in [23, 47]. However if we consider the Iterative mean on $SO(r)$ we can eliminate this problem by using the Iterative mean, if the initial points are in a small enough metric ball according to Proposition 5.12.

The other widely used mean is the center of mass defined by (5.27). Theorem 5.7 ensures the existence and uniqueness of the center of mass in a small enough metric ball. Then according to Proposition 4.10 the center of mass can be calculated as the solution of the equation

$$\sum_{i=1}^{n} \log_X(Q_i) = 0. \tag{8.1}$$

This is a nontrivial and often nonlinear equation. A gradient or newton method may be applied for finding the solution of it, although one must ensure the convergence of the methods. A gradient method combined with a certain line search algorithm ensures the global convergence, although the line search rule significantly increases the computational time. One may use the Iterative mean to approximate the center of mass, since the distance of the two points are generally bounded due to Corollary 5.9. Numerical experiments suggests that the centroid is very close to the Iterative mean, so we can speed up the gradient method with line search rule by starting from the Iterative mean, i. e. we approximate the Iterative mean sufficiently by its defining iteration, then start the gradient method from the approximating point. This method can be applied in the case of the geometric mean as well, since $P(r, \mathbb{C})$ is a nonpositively curved Riemannian manifold with the trace metric (4.2).

8.2 Averaging elements of $P(r, \mathbb{C})$

Averaging points in the nonpositively curved Riemannian manifold $P(r, \mathbb{C})$ can be regarded as calculating matrix means. In 1980 Kubo and Ando formulated the axiomatic theory of 2-variable matrix means, see Section 3 again. Since then several researchers were trying to extend the theory to several variables. Here we have given three different axiomatic extensions, the Iterative mean, the ALM mean and the BMP mean. The properties considered by Kubo and Ando carries over to these n-variable extensions nicely. This provides us with matrix inequalities between n-variable means, which can possibly be used in several situations. Some of these means can be used in the future for certain averaging problems where the arithmetic mean does not fit well to the problem. Some of these situations have already been mentioned earlier.

9 Summary

Means of two positive matrices were characterized by Kubo and Ando in 1980. Their theory was based on the classical Loewner theory of operator monotone functions. They showed that every two-variable matrix mean is isomorphic to a normalized operator monotone function. Since then the theory has found many applications in quantum information theory and operator theory.

However since it has been an open problem to extend the axiomatic theory to three or more variables. The arithmetic and harmonic means were trivial in several variables even for positive definite matrices, but no other matrix mean can be easily extended to several variables. The first ideas were given for the logarithmic mean by Carson [19]. Then Horwitz [30] considered a so called symmetrization method, which were considered by Ando, Li and Mathias (the ALM process) [5] as well to extend the geometric mean to several variables. Later the convergence of this process were proved in nonpositively curved metric spaces by Lawson and Lim [38]. Later other geometric means were proposed, for instance the Riemannian mean of Moakher [47] and the Bini-Meini-Poloni mean which is again a sort of symmetrization process (BMP-process) [15].

Even after these successes it has been not known whether these procedures can be applied to all matrix means. We will give a general theory here which solves this problem. The whole theory has a geometrical picture which makes it possibe to consider the theory in complete metric spaces with a certain curvature bound. Therefore the first results will be presented in this metric geometric setting. We will denote explicitly which theorems and definitions were given and proved by the author.

9.1 Means in Complete k-convex Metric Spaces

The results in this section form the first thesis group. The following definition of k-convexity is due to Ohta in [53]. We establish our results for spaces with such properties below.

Definition 9.1. Let $k \in (0, 2]$.

- An open set U in a geodesic metric space (X, d) is called a C_k-domain if for any three points x, y, z, any geodesic $\gamma : [0, 1] \mapsto X$ between x, y and for all $t \in [0, 1]$ we have

$$d(z, \gamma(t))^2 \leq (1-t)d(z, x)^2 + td(z, y)^2 - \frac{k}{2}t(1-t)d(x, y)^2. \tag{9.1}$$

- A geodesic metric space (X, d) is k-convex if it is itself a $C_k - domain$.

- A geodesic metric space (X, d) is locally k-convex if every point in X is contained in a C_k-domain.

If the inequality (9.1) holds for $t = 1/2$ then it holds for all $t \in [0, 1]$. A k-convex metric space becomes a $CAT(0)$ space if the above inequality holds for $k = 2$. In this case the space is said to have nonpositive curvature in the sense of Alexandrov, in other words the semiparallelogram law holds.

The following definition gives a process, which can be applied to extend means to several variables.

Definition 9.2 (Iterative process, M. Pálfia [60]). Let $Q_1^0, \dots, Q_n^0$ be points in a uniquely geodesic metric space X and $\pi = \{\pi_0, \pi_1, \dots\}$ be an infinite sequence of permutations, where each π_i is a permutation of the letters $\{1, \dots, n\}$. With respect to the infinite sequence of permutations π let

$$Q_i^{l+1} = \begin{cases} Q_{\pi_l(i)}^l \sharp Q_{\pi_l(i+1)}^l & \text{if } 1 \leq i < n, \\ Q_{\pi_l(n)}^l \sharp Q_{\pi_l(1)}^l & \text{else.} \end{cases} \tag{9.2}$$

The above procedure yields a sequence of n-tuple of points.

The following result ensures the convergence of the process in general.

Theorem 9.1 (Iterative mean, M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1^0, \dots, Q_n^0$ be points in the metric space X. Let us set up the iteration in Definition 9.2 on these points in X with respect to an infinite sequence of permutations $\pi = \{\pi_0, \pi_1, \dots\}$. Then the sequences Q_i^l converge to a common limit point.*

The rate of convergence is linear due to the following

Theorem 9.2 (M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1^0, \dots, Q_n^0$ be points in the metric space X. Let us set up the iteration in Definition 9.2 on these points in X. Let R denote the common limit point of these sequences. Then*

$$\frac{a_{l+1}(R)}{a_l(R)} \leq 1 - \frac{k}{2n^2}, \tag{9.3}$$

so the points Q_i^l are converging to R linearly.

The heuristic function Idealmapping defined by Algorithm 1 gives us a tool to speed up the rate of convergence to the common limit point. It must be noted however that the limit point depends on the infinite sequence of permutations π in Definition 9.2, so therefore it is denoted by R_π.

The following theorems partially answer a question proposed by Bhatia and Holbrook in [13] that whether the center of mass is the same point as the limit of certain symmetrization procedures.

Corollary 9.3 (M. Pálfia [60]). *Let (X, d) be a complete k-convex geodesic metric space. Let $Q_1, \dots, Q_n$ be points in the metric space X. Then the center of mass*

$$Y = \arg\min_{x \in X} \sum_{i=1}^{n} d\left(x, Q_i\right)^2 \tag{9.4}$$

and the limit point R_π of the procedure in Theorem 9.1 set up on the points $Q_1, \dots, Q_n$ fulfill the following inequality

$$d(R_\pi, Y) \leq \sqrt{\frac{\sum_{i=1}^{n} d\left(Y, Q_i\right)^2 - \frac{k}{8}\sum_{l=1}^{\infty} e_l}{n}}. \tag{9.5}$$

Proposition 9.4 (M. Pálfia [60])**.** *If X is a Euclidean space then the limit point R_π of the procedure in Theorem 9.1 is the center of mass of the starting points for every possible infinite sequence of permutations π.*

The limit point R_π of the procedure depends on the chosen infinite sequence π. If (9.1) turns into an equality, as in the case of a single geodesic segment or Euclidean space, then the possibly different limit points depending on π of the procedure will collapse onto one unique point, the center of mass.

Theorem 9.1 gives a mean for the special orthogonal group as well which is also an actively studied manifold in terms of averaging [46], [48]. $SO(n)$ is locally k-convex due to propositions in [53]. These propositions due to Ohta tells us how to translate the requirement of k-convexity to the language of curvature. It turns out that an upper curvature bound suffices. So this mean exists not only in nonpositively curved spaces as the ALM mean does which was shown by Lawson and Lim [38], but also in positively curved metric spaces as well.

In the next section we solve the problem, that whether the arithmetic, harmonic and geometric means are the only matrix means which are midpoint operations on certain manifolds.

9.2 Symmetric Matrix Means as Metric Midpoints

This section contains the second thesis group. Here $P(n, \mathbb{C})$ denotes the convex cone of positive definite n-by-n matrices over the complex field $\mathbb{C}$ and similarly $H(n, \mathbb{C})$ denotes the vector space of hermitian n-by-n matrices over $\mathbb{C}$. We begin with general theorems for affinely connected manifolds.

Theorem 9.5 (M. Pálfia [56])**.** *Let M be an affinely connected smooth manifold diffeomorphically embedded into a vector space V. Suppose that the midpoint map $m(p, q) = \exp_p(1/2 \log_p(q))$ is known in every normal neighborhood where the exponential map $\exp_p(X)$ is a diffeomorphism. Then in these normal neighborhoods the inverse of the exponential map $\log_p(q)$ can be fully reconstructed from the midpoint map in the form*

$$\log_p(q) = \lim_{n\to\infty} \frac{m(p, q)^{\circ n} - p}{\frac{1}{2^n}}, \tag{9.6}$$

where we use the notation $m(p, q)^{\circ n} \equiv m\left(p, m(p, q)^{\circ(n-1)}\right)$.

By the Kubo-Ando theory of 2-variable matrix means it is known that every matrix mean can be written in the form

$$M(A, B) = A^{1/2} f\left(A^{-1/2} B A^{-1/2}\right) A^{1/2}, \tag{9.7}$$

where $f(t)$ is a normalized operator monotone function. For symmetric means, we have $f(t) = tf(1/t)$ which implies that $f'(1) = 1/2$. Recall from Loewner theory [9] the integral characterization that an operator monotone function $f(t)$, which is defined over the interval $(0, \infty)$, possesses:

$$f(t) = \alpha + \beta t + \int_0^\infty \left(\frac{\lambda}{\lambda^2 + 1} - \frac{1}{\lambda + t} \right) d\mu(\lambda), \tag{9.8}$$

where α is a real number, $\beta \geq 0$ and μ is a positive measure on $(0, \infty)$ such that

$$\int_0^\infty \frac{1}{\lambda^2 + 1} d\mu(\lambda) < \infty. \tag{9.9}$$

We are interested in finding all possible symmetric matrix means which are also geodesic midpoint operations on smooth manifolds. We call such a matrix mean affine [64]:

Definition 9.3 (Affine matrix mean, M. Pálfia [56])**.** An affine matrix mean $M : W^2 \mapsto W$ is a symmetric matrix mean which is at the same time a geodesic midpoint operation $M(A, B) = \exp_A(1/2 \log_A(B))$ on a smooth manifold $W \supseteq P(n, \mathbb{C})$ equipped with an affine connection, where B is assumed to be in the injectivity radius of the exponential map $\exp_A(x)$ of the connection given at the point A. The mapping $\log_A(x)$ is just the inverse of the exponential map at the point $A \in W$.

The following assertion shows that if a matrix mean is affine then the exponential map of the corresponding smooth manifold has a special structure. We will use similarly the notation $M(A, B)^{\circ n} = M\left(A, M(A, B)^{\circ(n-1)}\right)$ as before.

Theorem 9.6 (M. Pálfia [56])**.** *Let $M(A, B)$ be a symmetric matrix mean. Then*

$$\lim_{n\to\infty} \frac{M(A, B)^{\circ n} - A}{\frac{1}{2^n}} = A^{1/2} \log_I \left(A^{-1/2} B A^{-1/2}\right) A^{1/2} \tag{9.10}$$

where the limit exists and is uniform for all $A, B \in P(n, \mathbb{C})$ and $\log_I(t)$ is an operator monotone function on the interval $(0, \infty)$.

As a consequence of the above we conclude the following

Proposition 9.7 (M. Pálfia [56])**.** *If a symmetric matrix mean $M(A, B)$ is an affine mean, then the exponential map and its inverse, the logarithm map are of the following forms*

$$\begin{aligned} \exp_p(X) &= p^{1/2} \exp_I \left(p^{-1/2} X p^{-1/2}\right) p^{1/2} \\ \log_p(X) &= p^{1/2} \log_I \left(p^{-1/2} X p^{-1/2}\right) p^{1/2} \end{aligned} \tag{9.11}$$

for $p \in P(n, \mathbb{C})$, where $\exp_I(X)$ and $\log_I(X)$ are analytic functions such that $\exp_I : H(n, \mathbb{C}) \mapsto P(n, \mathbb{C})$ and $\log_I(X)$ is its inverse and $\log_I'(I) = I, \exp_I'(0) = I, \log_I(I) = 0, \exp_I(0) = I$.

After investigating the properties of the possible affine connections that can occur we derive the main result of the section.

Theorem 9.8 (M. Pálfia [56])**.** *All affine matrix means $M(X,Y)$ are of the form*

$$M(X,Y) = \begin{cases} X^{1/2}\left[\frac{I+\left(X^{-1/2}YX^{-1/2}\right)^{1-\kappa}}{2}\right]^{\frac{1}{1-\kappa}} X^{1/2} & \text{if } \kappa \neq 1, \\ X^{1/2}\left(X^{-1/2}YX^{-1/2}\right)^{1/2} X^{1/2} & \text{if } \kappa = 1, \end{cases} \tag{9.12}$$

where $0 \leq \kappa \leq 2$. The symmetric affine connections corresponding to these means are

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2}\left(X_p p^{-1} Y_p + Y_p p^{-1} X_p\right). \tag{9.13}$$

These connections are only metrizable (Riemannian) if and only if $\kappa = 0, 1, 2$, in which case we get back the arithmetic, geometric and harmonic means respectively.

We can see due to the above result, that generally we cannot treat the extension problem of matrix means to several variables as a purely metric geometric problem. Although some of the geometric ideas can be applied somehow. The following sections are all based around this idea. This is the longest and possibly most complicated part of the thesis. The results here can be found in the section "Extensions of Matrix Means without Metric Structures".

9.3 Iterative Mean for all Matrix Means

In general our goal is to construct several variable functions with the following properties. Most of these properties were considered by Ando, Li and Mathias in [5] and also by Petz and Temesi [68].

Definition 9.4 (Multivariable Matrix Mean)**.** Let $M : P(r,\mathbb{C})^n \mapsto P(r,\mathbb{C})$. Then M is called a matrix mean if the following conditions hold

1. $M(X,\ldots,X) = X$ for every $P(r,\mathbb{C})$,
2. $M(X_1,\ldots,X_n)$ is invariant under the permutation of its variables,
3. $\min(X_1,\ldots,X_n) \leq M(X_1,\ldots,X_n) \leq \max(X_1,\ldots,X_n)$ if min and max exist with respect to the positive definite order,
4. If $X_i \leq X_i'$, then $M(X_1,\ldots,X_n) \leq M(X_1',\ldots,X_n')$,
5. $M(X_1,\ldots,X_n)$ is continuous,
6. $CM(X_1,\ldots,X_n)C^* \leq M(CX_1C^*,\ldots,CX_nC^*)$.

The next algorithm is the extension of the Iterative mean for metric spaces to the matrix mean setting.

The next result is the first general result which gives a solution to the long standing problem of axiomatic extension of 2-variable matrix means.

Algorithm 3 Iterative extension of a 2-variable matrix mean

```
Data: X = (X_1, ..., X_n) ∈ P(r,ℂ)^n; a 2-variable matrix mean M : P(r,ℂ) × P(r,ℂ) → P(r,ℂ).
Initialization: X_i^0 := X_i, i = 1, ..., n.
for k = 0, 1, ... do
   Choose a directed graph G^k with n vertices labelled from 1 to n and n edges labelled from 1 to n, such that every vertex has exactly two edges connected to it.
   for i = 1, ..., n do
      X_i^{k+1} := M(X_{j_i}^k, X_{l_i}^k), where j_i is the tail vertex and l_i the head vertex of edge i in G^k.
   end for
end for
```

Theorem 9.9 (M. Pálfia [54])**.** *Let the matrix mean M in Algorithm 3 satisfy $M(A,B) \leq \frac{A+B}{2}$ for all $A, B \in P(r,\mathbb{C})$. Then the n sequences $(X_i^k)_{k\geq 0}$, $i = 1,\dots,n$, generated by Algorithm 3 converge and have the same limit point.*

These limit points generally seems to be different, they depend on the graphs G^k chosen in every iteration step k, similarly to the metric space case. Therefore we also introduce the following notation in order to express this dependence on the sequence of graphs. Let us denote the infinite sequence of graphs with

$$G = \left\{G^0, G^1, \dots\right\}. \tag{9.14}$$

With this notation from now on we denote the common limit point of the sequences in Theorem 9.9 as $M_G(X_1,\dots,X_n)$ to express the dependence of the limit point on the sequence of graphs G.

The limit points clearly provides us with n-variable extensions that possess the required properties.

Proposition 9.10 (M. Pálfia [54])**.** *The limit point $M_G(X_1,\dots,X_n)$ of the matrix sequences given in Algorithm 3 satisfies* 1.*,* 3. *and* 4. *in Definition 9.4 with respect to an infinite sequence of graphs G.*

Proposition 9.11 (M. Pálfia [54])**.** *If $M(A,B) \leq N(A,B) \leq (A+B)/2$ are matrix means, then the same ordering is true for the induced limit points $M_G(X_1,\dots,X_n)$ and $N_G(X_1,\dots,X_n)$ of the matrix sequences given in Algorithm 3 with respect to an infinite sequence of graphs G.*

Proposition 9.12 (M. Pálfia [54])**.** *The limit point $M_G(X_1,\dots,X_n)$ of the matrix sequences given in Algorithm 3 satisfies property* 6. *in Definition 9.4.*

Proposition 9.13 (M. Pálfia [54])**.** *The limit point $M_G(X_1,\dots,X_n)$ of the matrix sequences given in Algorithm 3 is a continuous function in its $X_1,\dots,X_n$ variables.*

During the proof of the last theorem the author also carried out a notable generally applicable result.

Theorem 9.14 (M. Pálfia [64])**.** *Let $F : P(r,\mathbb{C})^n \mapsto P(r,\mathbb{C})$ which satisfies properties*

1. *if $X_i \leq X_i'$ for all i, then $F(X_1,\dots,X_n) \leq F(X_1',\dots,X_n')$,*
2. *$F(cX_1,\dots,cX_n) = cF(X_1,\dots,X_n)$ for real $c > 0$.*

Then F is continuous.

9.4 Weighted 2-variable Matrix Means

Kubo-Ando theory does not give us any hint on how to find the weighted 2-variable versions of a symmetric matrix mean. Here we provide a suitable process that defines us a weighted 2-variable matrix mean corresponding to every symmetric one.

Definition 9.5 (Weighted mean process, M. Pálfia [64])**.** Let $M(\cdot,\cdot)$ be a symmetric matrix mean, $A, B \in P(r,\mathbb{C})$ and $t \in [0,1]$. Let $a_0 = 0$ and $b_0 = 1$, $A_0 = A$ and $B_0 = B$. Define a_n, b_n and A_n, B_n recursively by the following procedure for all $n = 0,1,2,\dots$:

if $a_n = t$ **then**
  $a_{n+1} = a_n$ and $b_{n+1} = a_n$, $A_{n+1} = A_n$ and $B_{n+1} = A_n$
else if $b_n = t$ **then**
  $a_{n+1} = b_n$ and $b_{n+1} = b_n$, $A_{n+1} = B_n$ and $B_{n+1} = B_n$
else if $\frac{a_n+b_n}{2} \leq t$ **then**
  $a_{n+1} = \frac{a_n+b_n}{2}$ and $b_{n+1} = b_n$, $A_{n+1} = M(A_n, B_n)$ and $B_{n+1} = B_n$
else
  $b_{n+1} = \frac{a_n+b_n}{2}$ and $a_{n+1} = a_n$, $B_{n+1} = M(A_n, B_n)$ and $A_{n+1} = A_n$
end if

According to the above a_{n+1}, b_{n+1} and A_{n+1}, B_{n+1} are clearly defined with respect to a_n, b_n and A_n, B_n recursively.

This algorithm may also be regarded as a kind of binary search with recurrence relation:

if $t = \frac{t_1+t_2}{2}$ **then**
  $M_t(A,B) = M\left(M_{t_1}(A,B), M_{t_2}(A,B)\right)$
end if

Theorem 9.15 (M. Pálfia [64])**.** *The sequences A_n and B_n given in Definition 9.5 are convergent and have the same limit point.*

Definition 9.6 (Weighted mean, M. Pálfia [64])**.** The common limit point of A_n, B_n in Theorem 9.15 will be denoted by $M_t(A,B)$ and from now on is considered as the corresponding weighted mean to a symmetric matrix mean $M(\cdot,\cdot)$.

The following results will give us the nice porperties which a weighted matrix mean should possess.

Proposition 9.16 (M. Pálfia [64])**.** *$M_t(A,B)$ yields the correct corresponding weighted means in the case of the arithmetic, geometric, harmonic means.*

The above is a consequence of the affine geodesy of the corresponding manifolds mentioned above. There are further important properties which are fulfilled by $M_t(A,B)$:

Proposition 9.17 (M. Pálfia [64])**.** *$M_t(A,B)$ for $A,B \in P(r,\mathbb{C})$ and $t \in [0,1]$ fulfills the following properties*

(i') $M_t(I,I) = I$,

(ii') *if* $A \leq A'$ *and* $B \leq B'$, *then* $M_t(A,B) \leq M_t(A',B')$,

(iii') $CM_t(A,B)C \leq M_t(CAC,CBC)$,

(iv') *if* $A_n \downarrow A$ *and* $B_n \downarrow B$ *then* $M_t(A_n,B_n) \downarrow M_t(A,B)$,

(v') *if* $N(A,B) \leq M(A,B)$ *then* $N_t(A,B) \leq M_t(A,B)$,

(vi') $M_{1/2}(A,B) = M(A,B)$,

(vii') $M_t(A,B)$ *is continuous in* t,

Corollary 9.18 (M. Pálfia [64])**.** *For every symmetric matrix mean $M(A,B)$ there is a corresponding one parameter family of weighted means $M_t(A,B)$ for $t \in [0,1]$. Let $f(x)$ be the normalized operator monotone function corresponding to $M(A,B)$. Then similarly we have a one parameter family of normalized operator monotone functions $f_t(x)$ corresponding to $M_t(A,B)$. The family $f_t(x)$ is continuous in t, and $f_0(x) = 1$ and $f_1(x) = x$ are the two extremal points, so $f_t(x)$ interpolates between these two points.*

Based on this phenomenon we can conclude the following

Proposition 9.19 (M. Pálfia [64])**.** *Let $M(A,B)$ be a symmetric matrix mean. Then*

$$\left((1-t)A^{-1} + tB^{-1}\right)^{-1} \leq M_t(A,B) \leq (1-t)A + tB, \tag{9.15}$$

where $M_t(A,B)$ is the weighted version of $M(A,B)$.

We are going to use the big O notation. This means that we have $X = Y + O(\epsilon^k)$ if and only if there exist constants $\epsilon_0 < 1$ and θ such that for each $0 < \epsilon < \epsilon_0$ we have $\|X - Y\| \leq \theta\epsilon^k$.

Proposition 9.20 (M. Pálfia [64])**.** *Let $M(A,B)$ be a symmetric matrix mean and $f(t)$ be its corresponding normalized operator monotone function. Let $f(t)$ have a series expansion around I as*

$$f(X) = I + \frac{X - I}{2} + \sum_{k=2}^{\infty} b_k (X - I)^k. \tag{9.16}$$

Then we have a series expansion for $M_t(I,X) = f_t(X)$ whenever $\|X - I\| \leq \epsilon < 1$ in the form

$$f_t(X) = I + t(X - I) + 4b_2 t(1-t)(X-I)^2 + O(\epsilon^3). \tag{9.17}$$

This important result will ultimately lead us to the cubic convergence of the BMP-process later.

9.5 Ando-Li-Mathias Procedure for all Matrix Means

Here in this section we will give an affirmative answer to the conjecture formulated by Petz and Temesi in [68, 69] that the ALM-process converges for all matrix means. The convergence of this process for the geometric mean were proved by Ando, Li and Mathias [5] and later by Petz and Temesi [68]. The last two were also able to prove the convergence of the process for orderable tuples. A general proof however was out of reach at that time.

Definition 9.7 (ALM iteration)**.** Let $X = (X_1^0, \ldots, X_n^0)$ where $X_i^0 \in P(r,\mathbb{C})$ and define the mapping $M(X_1, \ldots, X_n)$ inductively as follows. If $n = 2$ assume that $M(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \ldots, X_{n-1})$ is already defined. Then using $M(X_1, \ldots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M\left(Z_{\neq i}\left(X_1^l, \ldots, X_n^l\right)\right), \tag{9.18}$$

where $Z_{\neq i}(X_1^l, \ldots, X_n^l) = X_1^l, \ldots, X_{i-1}^l, X_{i+1}^l, \ldots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l\to\infty} X_i^l = M(X_1^0, \ldots, X_n^0). \tag{9.19}$$

Theorem 9.21 (M. Pálfia [64])**.** *Let $F : P(r,\mathbb{C})^2 \mapsto P(r,\mathbb{C})$ and suppose that $F(A,B)$ fulfills one of the inequalities below:*

$$\left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1} \leq F(A,B) \leq \left[\frac{A^2+B^2}{2} - \frac{k}{8}(A-B)^2\right]^{1/2} \tag{9.20}$$

for a $k \in (0,2]$, or

$$F(A,B) \leq \frac{A+B}{2}. \tag{9.21}$$

Then in Definition 9.7 starting with $M(A,B) := F(A,B)$, $M(X_1, \ldots, X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.

The next section concludes a similar result for the BMP process.

9.6 Bini-Meini-Poloni Procedure for all Matrix Means

This result is similar to the above one and also gives us an axiomatic extension of 2-variable matrix means in general. This process were considered by Bini, Meini and Poloni in [15]. They were only able to prove the convergence of this process for the geometric mean. Here we prove it in full generality.

Definition 9.8 (BMP iteration)**.** Let $X = (X_1^0, \ldots, X_n^0)$ where $X_i^0 \in P(r, \mathbb{C})$ and define the mapping $M(X_1, \ldots, X_n)$ inductively as follows. If $n = 2$ assume that $M_t(X_1, X_2)$ is already given. For general $n > 2$ assume that $M(X_1, \ldots, X_{n-1})$ is already defined. Then using $M(X_1, \ldots, X_{n-1})$, set up the iteration

$$X_i^{l+1} = M_{\frac{n-1}{n}}\left(X_i^l, M\left(Z_{\neq i}\left(X_1^l, \ldots, X_n^l\right)\right)\right), \tag{9.22}$$

where $Z_{\neq i}(X_1^l, \ldots, X_n^l) = X_1^l, \ldots, X_{i-1}^l, X_{i+1}^l, \ldots, X_n^l$. If the sequences X_i^l converge to a common limit point for every i, then define

$$\lim_{l \to \infty} X_i^l = M(X_1^0, \ldots, X_n^0). \tag{9.23}$$

Theorem 9.22 (M. Pálfia [64])**.** *Let $F : [0,1] \times P(r, \mathbb{C})^2 \mapsto P(r, \mathbb{C})$ and suppose that $F_t(A, B)$ fulfills one of the inequalities below:*

$$\begin{aligned} &\left[(1-t)A^{-1} + tB^{-1}\right]^{-1} \leq F_t(A, B) \leq \\ &\leq \left[(1-t)A^2 + tB^2 - \frac{k}{2}t(1-t)(A-B)^2\right]^{1/2} \end{aligned} \tag{9.24}$$

for a $k \in (0, 2]$ and every $t \in [0, 1]$, or

$$F_t(A, B) \leq (1-t)A + tB, \tag{9.25}$$

for every $t \in [0,1]$. Then in Definition 9.8 starting with $M_t(A, B) := F_t(A, B)$, $M(X_1, \ldots, X_n)$ exists for all n, in other words the sequences converge to a common limit point for all n.

Proposition 9.20 tells us that no matter how we choose the symmetric matrix mean $M(A, B)$, the series expansion of $M_t(A, B)$ around I will have similar structure up to the $(X - I)^2$ term:

$$M_t(A, B) = (1-t)A + tB + 4b_2 t(1-t)A^{1/2}\left(A^{-1/2}BA^{-1/2} - I\right)^2 A^{1/2} + \ldots \tag{9.26}$$

Actually $b_2 \leq 0$ for every matrix mean since the corresponding operator monotone function is concave.

A remarkable consequence of the above series expansion is that the BMP procedure converges at least cubically if the matrices are sufficiently close to each other for all symmetric matrix means. This was proved by Bini, Meini and Poloni [15] for the geometric mean and this result was a major improvement over the ALM mean which were known to converge only linearly.

To prove this assertion we will prove the following more precisely formulated

Theorem 9.23 (M. Pálfia [64])**.** *Let $0 < \epsilon < 1$, $K, X_i^0 \in P(r, \mathbb{C}), i = 1, \ldots, n$, and $E_i = K^{-1}X_i^0 - I$. Set up the the BMP process on the X_i^0 with a symmetric matrix mean $M(A, B)$ which has series expansion given as in Proposition 9.20. Now suppose that $\|E_i\| \leq \epsilon$ for all i. Then for the matrices X_i^1 the following hold.*

C1: We have

$$K^{-1}X_i^1 - I = T_n + O(\epsilon^3), \tag{9.27}$$

where

$$T_n = \frac{1}{n}\sum_{j=1}^{n} E_j + \frac{2b_2}{n^2}\sum_{i,j=1}^{n}(E_i - E_j)^2. \tag{9.28}$$

C2: There are positive constants θ, σ and $\bar{\epsilon} < 1$ (all of which may depend on n) such that for all $\epsilon \leq \bar{\epsilon}$,

$$\left\|K_1^{-1}X_i^1 - I\right\| \leq \theta\epsilon^3 \tag{9.29}$$

for a suitable matrix K_1 satisfying $\left\|K^{-1}K_1 - I\right\| \leq \sigma\epsilon$.

C3: The BMP iteration in Definition 9.8 converges at least cubically.

C4: We have

$$K_1^{-1}M_{BMP}(X_1^0, \ldots, X_n^0) - I = O(\epsilon^3), \tag{9.30}$$

where $M_{BMP}(X_1^0, \ldots, X_n^0)$ denotes the limit point of the BMP-process in Definition 9.8.

The last section ensures that the properties mentioned in Definition 9.4 are fulfilled.

9.7 Properties of the ALM and BMP mean

We will show that the limit point of the ALM and BMP processes, denoted by $M_{ALM}(X_1, \ldots, X_n)$ and $M_{BMP}(X_1, \ldots, X_n)$ respectively, as extensions of symmetric matrix means, fulfill the following required properties.

Theorem 9.24 (M. Pálfia [64])**.** *If $M(A, B)$ is a symmetric matrix mean, then the $M := M_{ALM}(X_1, \ldots, X_n)$ and $M := M_{BMP}(X_1, \ldots, X_n)$ extensions fulfill the following properties*

(I) $M(X, \ldots, X) = X$ for every $X \in P(r, \mathbb{C})$,

(II) $M(X_1, \ldots, X_n)$ is invariant under the permutation of its variables,

(III) $\min(X_1, \ldots, X_n) \leq M(X_1, \ldots, X_n) \leq \max(X_1, \ldots, X_n)$ if $\min$ and $\max$ exist with respect to the positive definite order,

(IV) If $X_i \leq X_i'$, then $M(X_1, \ldots, X_n) \leq M(X_1', \ldots, X_n')$,

(V) $M(X_1, \ldots, X_n)$ is continuous,

(VI) $M(CX_1C^*, \dots, CX_nC^*) = CM(X_1, \dots, X_n)C^*$ *for all invertible* C*.*

Proposition 9.25 (M. Pálfia [64])**.** *If* $M(A, B) \leq N(A, B)$ *are functions satisfying the properties of* $F(A, B)$ *in Theorem 9.21, then the same ordering is true for the ALM limit points* $M(X_1, \dots, X_n)$ *and* $N(X_1, \dots, X_n)$*.*

Proposition 9.26 (M. Pálfia [64])**.** *If* $M_t(A, B) \leq N_t(A, B)$ *are functions satisfying the properties of* $F_t(A, B)$ *in Theorem 9.22, then the same ordering is true for the BMP limit points* $M(X_1, \dots, X_n)$ *and* $N(X_1, \dots, X_n)$*.*

10 Acknowledgements

I would like to express my gratitude to Mihály Rácz who has given me the first examples of mathematical beauty. Without him I would not dare to consider mathematics as a topic of interest. During the first university years professor György Serény has made it possible for me to start any sort of research work in mathematics. I am very grateful for that and the vise advices that he has given. I am especially thankful to professor Dénes Petz for giving me this very nice research topic of matrix means. I must also emphasize the importance of certain situations when he persuaded me to work hard on this topic. I would also like to thank my supervisor professor István Vajk for guiding me throughout the struggles that I had. He was always willing to sacrifice his time to listen to and give advice on my problems and thoughts which made it possible to synthesize the ideas. During my time at CEU, professor Pál Hegedűs has given me crucial advices on how to publish articles. I am very grateful for that and for the courage he has given me all the time.

At last but not least I am very thankful to my parents who provided me ideal working conditions all the time. Without them I would be nowhere.

References

[1] E. Ahn, S. Kim and Y. Lim, *An extended LieTrotter formula and its applications*, Linear Algebra and its Appl., 427 (2007), pp. 190–196.

[2] W. N. Anderson Jr., *Shorted operators*, SIAM J. Appl. Math., 20 (1971), pp. 520–525.

[3] W. N. Jr. Anderson and R. J. Duffin, *Series and parallel addition of matrices*, J. Math. Anal. Appl., 26 (1969), pp. 576-594.

[4] W. N. Jr. Anderson and G. E. Trapp, *Shorted Operators II*, SIAM J. Appl. Math., 28 (1975), pp. 60-71.

[5] T. Ando, C-K. Li and R. Mathias, *Geometric means*, Linear Algebra Appl., 385 (2004), pp. 305–334.

[6] W. Ballmann, *Lectures On Spaces Of Nonpositive Curvature*, Birkhäuser, 1995, ISBN: 978-3-7643-5242-4.

[7] F. Barbaresco, *Innovative tools for radar signal processing based on Cartan's geometry of symmetric positive definite matrices and information geometry*, in Proceedings of the IEEE International Radar Conference, Rome, Italy, 2008.

[8] F. Barbaresco, *Interactions between symmetric cone and information geometry: Bruhat-Tits and siegel space models for higher resolution autoregressive dopper imagery*. In *Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, 2008. Revised Invited Papers*, pages 124-163. Springer-Verlag, 2009.

[9] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1996.

[10] R. Bhatia, *On the exponential metric increasing property*, Linear Algebra Appl., 375 (2003), pp. 211-220.

[11] R. Bhatia, *Positive Definite Matrices*, Princeton University Press, 2007.

[12] R. Bhatia and J. Holbrook, *Noncommutative geometric means*, Math. Intelligencer 28 (2006), pp. 32–39.

[13] R. Bhatia and J. Holbrook, *Riemannian geometry and matrix geometric means*, Linear Alg. Appl., 413 (2006), pp. 594-618.

[14] D. A. Bini and B. Iannazzo, *Computing the Karcher mean of symmetric positive definite matrices*, preprint.

[15] D. A. Bini, B. Meini and F. Poloni, *An effective matrix geometric mean satisfying the Ando-Li-Mathias properties*, Math. Comp., 79 (2010), pp. 437–452.

[16] M. R. Bridson and A. Häfliger, *Metric Spaces of Non-Positive Curvature*, 643 p. Springer-Verlag, Berlin Heidelberg (1999), ISBN: 978-3-540-64324-1.

[17] D. Burago, Yu D. Burago and S. Ivanov, *A Course in Metric Geometry*, American Mathematical Society, 2001, ISBN 0-8218-2129-6.

[18] É. Cartan, *La topologie des espaces representatifs des groupes de Lie*, Enseignment Math., 35 (1936), pp. 177-200, Exposés de Géometrie VIII.

[19] B́. Carson, *The logarithmic mean*, Amer. Math. Monthly 79 (1972), pp. 615-618.

[20] C. Conde, *Differential Geometry for Nuclear Positive Operators*, Integral Equations and Operator Theory, Vol. 57, Num. 4 (2006), pp. 451-471.

[21] L. P. Eisenhart and O. Veblen, *The Riemann Geometry and Its Generalization*, Proceedings of the National Academy of Sciences of the United States of America, Vol. 8, No. 2 (1922), pp. 19–23.

[22] M. Fiedler and V. Pták, *A new positive definite geometric mean of two positive definite matrices*, Linear Algebra Appl. 251 (1997), pp. 1-20.

[23] P. Fillard, V. Arsigny, X. Pennec and N. Ayache, *Joint estimation and smoothing of clinical DT-MRI with a Log-Euclidean metric*, Research Report RR-5607, INRIA, Sophia-Antipolis, France, June 2005.

[24] P. D. Fletcher and S. Joshi, *Riemannian geometry for the statistical analysis of difusion tensor data*, Signal Processing 87 (2007), pp. 250–262.

[25] H. W. Guggenheimer, *Differential Geometry*, 378 p. Mc Graw-Hill, New York, (1977), ISBN: 978-0-486-63433-3.

[26] T. Hayashi, *Non-commutative arithmetic-geometric mean inequality*, Proc. Amer. Math. Soc. 137 (2009), pp. 3399–3406.

[27] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, 641 p. American Mathematical Society (2001), ISBN: 978-0-821-82848-9.

[28] F. Hiai and D. Petz, *Riemannian metrics on positive definite matrices related to means*, Linear Algebra Appl., 430 (2009), pp. 3105–3130.

[29] F. Hiai and H. Kosaki, *Means of Hilbert space operators*, Lecture Notes In Maths. 1820 (2003), Springer, ISBN: 978-3-540-40680-8.

[30] A. Horwitz, *Invariant means*, J. Math. Anal Appl. 270 (2002), pp. 499-518.

[31] J. Jost, *Nonpositive Curvature: Geometric and Analytic Aspects*, Birkh auser-Verlag, Basel, 1997.

[32] C. Jung, H. Lee and T. Yamazaki, *On a new construction of geometric mean of n-operators*, Linear Alg. Appl., vol. 431 (2009), pp. 1477–1488.

[33] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math., vol. 30 (1977), pp. 509-541.

[34] S. Kobayashi and N. Nomizu, *Foundations of Differential Geometry Vol. 1.*, 344 p. Wiley-Interscience (1996), ISBN: 978-0-471-15733-5.

[35] S. Kobayashi and N. Nomizu, *Foundations of Differential Geometry Vol. 2.*, 488 p. Wiley-Interscience (1996), ISBN: 978-0-471-15732-8.

[36] F. Kubo and T. Ando, *Means of positive linear operators*, Math. Ann., 246 (1980), pp. 205–224.

[37] J. Lawson, H. Lee and Y. Lim, *Weighted geometric means*, to appear in Forum Mathematicum.

[38] J. Lawson and Y. Lim, *A general framework for extending means to higher orders*, Colloq. Math., 113 (2008), pp. 191–221.

[39] J. Lawson and Y. Lim, *Monotonic properties of the least squares mean*, Math. Ann. (2010), doi:10.1007/s00208-010-0603-6.

[40] J. Lawson and Y. Lim, *Metric convexity of symmetric cones*, Osaka J. Math. 44 (2007), pp. 795–816.

[41] J. Lawson and Y. Lim, *Symmetric spaces with convex metrics*, Forum Mathematicum 19 (2007), pp. 571–602.

[42] H. Lee, Y. Lim and T. Yamazaki, *Multi-variable weighted geometric means of positive definite matrices*, Linear Algebra Appl., to appear [Doi:10.1016/j.laa.2011.01.026].

[43] Y. Lim, *Best approximation in Riemannian geodesic submanifolds of positive definite matrices*, Can. J. Math. 56 (2004), pp. 776–793.

[44] I. J. Maddox, *Elements of functional analysis*, Cambridge University Press, Cambridge, 1988.

[45] R. S. Manning, J. H. Maddocks and J. D. Kahn, *A continuum rod model of sequence-dependent DNA structure*, J. Chem. Phys., 105 (1996), pp. 5626–5646.

[46] J. H. Manton, *A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups*, in Eighth International Conference on Control, Automation, Robotics and Vision, Kunming, China, December 2004.

[47] M. Moakher, *A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices*, SIMAX, 26 (2005), pp. 735–747.

[48] M. Moakher, *Means and averaging in the group of rotations*, SIAM J. Matrix Anal. Appl., vol. 24, no. 1 (2002), pp. 1-16.

[49] M. Moakher, *On the averaging of symmetric positive-definite tensors*, Journal of Elasticity 82 (2006), pp. 273–296.

[50] M. Moakher and P. G. Batchelor, *Symmetric positive-definite matrices: from geometry to applications and visualization*, 452, Math. Vis., Springer, Berlin (2006), pp. 285-298.

[51] K.-H. Neeb, *A CartanHadamard Theorem for BanachFinsler Manifolds*, Geom. Ded., vol. 95, No. 1 (2002), pp. 115–156.

[52] K. Nomizu, *Invariant Affine Connections on Homogeneous Spaces*, American Journal of Mathematics, Vol. 76, No. 1 (1954), pp. 33–65.

[53] S.-I. Ohta, *Convexities of metric spaces*, Geom. Dedicata, 125 (2007), 225-250.

[54] M. Pálfia, *A multivariable extension of two-variable matrix means*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 385–393. (IF'10:2.411)

[55] M. Pálfia, *An iterative approach to means of matrices*, Proceedings of the Automation and Applied Computer Science Workshop (AACS 2009), pp. 271–284, ISBN: 978–963–420–978–2.

[56] M. Pálfia, *Classification of affine matrix means*, submitted to Diff. Geom. Appl. (2011).

[57] M. Pálfia, *Direct Rendering of B-spline Surfaces*, Proceedings of the Automation and Applied Computer Science Workshop (AACS 2007), pp. 307–313, ISBN 978–963–420–909–6.

[58] M. Pálfia, *Geometric means*, Proceedings of the Automation and Applied Computer Science Workshop (AACS 2010), pp. 267–274.

[59] M. Pálfia, *Matrix Means and the Riemannian Center of Mass*, Proceedings of the Automation and Applied Computer Science Workshop (AACS 2008), pp. 35–49, ISBN 978–963–420–955–3.

[60] M. Pálfia, *Means in metric spaces and the center of mass*, J. Math. Anal. Appl., 381 (2011), pp. 383–391. (IF'10:1.225)

[61] M. Pálfia, *On means of several positive matrices and numbers*, Proceedings of the Automation and Applied Computer Science Workshop (AACS 2007), pp. 283–291.

[62] M. Pálfia, *The Riemann Barycenter Computation and Means of Several Matrices*, Proceedings of World Academy of Science, Engineering and Technology, Vol. 30, July 2008, ISSN: 1307-6884, pp. 60–66.

[63] M. PÁLFIA, *The Riemann barycenter computation and means of several matrices*, Int. J. Comput. Math. Sci., 3(3) (2009), pp. 128–133.

[64] M. PÁLFIA, *Weighted matrix means and symmetrization procedures*, Linear Alg. Appl. (2011), in press, DOI: 10.1016/j.laa.2011.06.036. (IF'10:1.073)

[65] M. PÁLFIA AND D. PETZ, *Unitary Dilatations and Means of Matrices (Unitér dilatációk és mátrixok közepei)*, Polygon, vol. XV, no. 2 (2007), pp. 1–16.

[66] A. PAPADOPOULOS, *Metric spaces, Convexity and Nonpositive Curvature*, IRMA Lectures in Mathematics and Theoretical Physics, 6. European Mathematical Society (EMS), Zürich, 2005.

[67] D. PETZ, Matrix Analysis with some Applications, http://bolyai.cs.elte.hu/p̃etz/matrixbme.pdf.

[68] D. PETZ AND R. TEMESI, *Means of positive numbers and matrices*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 712–720.

[69] D. PETZ, *Means of positive matrices: Geometry and a conjecture*, Annales Mathematicae et Informaticae 32 (2005), pp. 129-139.

[70] M. J. PRENTICE, *Fitting smooth paths to rotation data*, Appl. Statist., 36 (1987), pp. 325–331.

[71] M. SAGAE AND K. TANABE, *Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices*, Linear and Multilinear Algebras 37 (1994), pp. 279–282.

[72] G. F. SIMMONS, *Introduction to Topology and Mondern Analysis*, McGraw-Hill Book Company Inc., 1966.

[73] M. SPIVAK, *A comprehensive introduction to differential geometry Vol. 2*, Publish or Perish, Houston, Texas (1979), ISBN: 0-914098-88-8.

[74] K. T. STURM, *Probability measures on metric spaces of nonpositive curvature*, in: Pascal Auscher, et al. (Eds.), Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs, April 16July 13, 2002, Paris, France, in: Contemp. Math., vol. 338, Amer. Math. Soc., Providence, RI, (2003), 357-390.

[75] A. VANŽUROVÁ AND P. ŽÁČKOVÁ, *Metrization of linear connections*, Journal of Applied Mathematics, Vol. 2, No. 1 (2009), pp. 151–164.

[76] T. YAMAZAKI, *The Riemannian mean and matrix inequalities related to the Ando-Hiai inequality and Chaotic order*, to appear in Operators and Matrices.

[77] M. Zerai and M. Moakher, *The Riemannian geometry of the space of positive-defite matrices and its application to the regularization of diffusion tensor MRI data*, J. Math. Imaging Vision, submitted.