

# On Affine Matrix Means

## MSc Thesis

Miklós Pálffia

May 31, 2011

Supervisor: Pál Hegedűs  
Department of Mathematics and its Applications  
Central European University

## Contents

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b>Introduction</b>   | <b>3</b>  |
| <b>2</b>  | <b>Operator Monotone Functions</b>                          | <b>5</b>  |
| 2.1       | Some Properties . . . . .                                   | 6         |
| 2.2       | Loewner's Characterization . . . . .                        | 17        |
| <b>3</b>  | <b>Matrix Means and Operator Monotone Functions</b>         | <b>20</b> |
| <b>4</b>  | <b>Extension of Matrix Means to Multiple Variables</b>      | <b>24</b> |
| 4.1       | The Riemannian Structure on $P(r, \mathbb{C})$ . . . . .    | 25        |
| 4.2       | Matrix Means defined as The Center of Mass . . . . .        | 30        |
| 4.3       | Symmetrization Procedures and Weighted Means . . . . .      | 35        |
| <b>5</b>  | <b>Affinely Connected Manifolds and the Exponential Map</b> | <b>38</b> |
| <b>6</b>  | <b>Geometric Constructions Applied to Matrix Means</b>      | <b>41</b> |
| <b>7</b>  | <b>Constructions of Invariant Affine Connections</b>        | <b>45</b> |
| <b>8</b>  | <b>Properties of These Affine Connections</b>               | <b>49</b> |
| <b>9</b>  | <b>Classification of Affine Matrix Means</b>                | <b>54</b> |
| <b>10</b> | <b>Conclusions</b>  | <b>57</b> |

# 1 Introduction

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

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

the geometric mean

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

and the harmonic

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

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  $H(x_1, \dots, x_n) \leq G(x_1, \dots, x_n) \leq A(x_1, \dots, x_n)$  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. In order to understand monotonicity of matrix functions, we give an introduction into their theory in the next section.

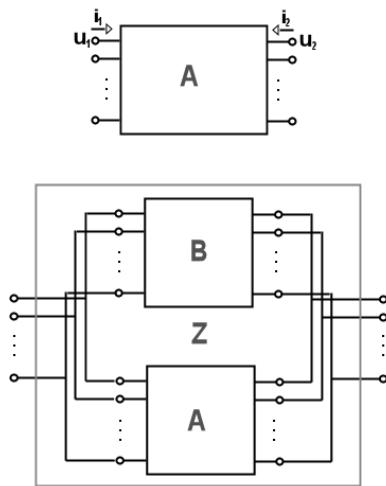


Figure 1: Parallel connection of two n-poles.

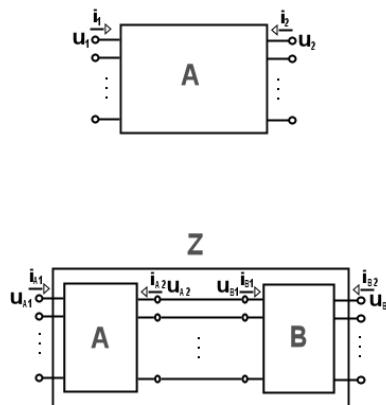


Figure 2: Series connection of two n-poles.

## 2 Operator Monotone Functions

In this section we will follow the lines of [7]. 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  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  with entries  $\lambda_i$  belonging to  $I$ , then  $f(D) = \text{diag}(f(\lambda_1), \dots, f(\lambda_n))$ . 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

$$f((1 - \lambda)A + \lambda B) \leq (1 - \lambda)f(A) + \lambda f(B). \quad (2.1)$$

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

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

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.

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

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

$$\langle r, X^*BXr \rangle = \langle Xr, BXr \rangle \geq \langle Xr, AXr \rangle = \langle r, X^*AXr \rangle. \quad (2.4)$$

□

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

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

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  $\epsilon \rightarrow 0$  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  $M = (I - K^*K)^{1/2}$ . Then the operators  $U, V$  given as

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

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

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

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

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

so

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

and

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

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

(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

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

so

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

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

(4)  $\Rightarrow$  (1): Let  $A, B$  be hermitian with eigenvalues in  $I$  and  $0 \leq \lambda \leq 1$ . 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

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

so

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

so  $f$  is operator convex, and  $f(0) \leq 0$ .  $\square$

**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  $f(K^*AK) \geq K^*f(A)K$  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

$$U^*TU \leq \begin{bmatrix} K^*AK + \epsilon I & 0 \\ 0 & \lambda I \end{bmatrix}. \quad (2.15)$$

Replace  $T$  by  $f(T)$  to get

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

by the operator monotonicity of  $f$ . Since  $\epsilon$  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

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

Operator concavity of  $f$  then yields

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

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

**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  $P$*

$$f(PAP) \leq Pf(PAP) = Pf(PAP)P. \quad (2.19)$$

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

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

**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

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

which implies that

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

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

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

Letting  $\epsilon \rightarrow 0$ , this gives

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

By the previous two corollaries, we get

$$f(PAP) \leq Pf(A)P. \quad (2.25)$$

□

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

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

The function  $f^{[1]}(\lambda, \mu)$  is called the first divided differences of  $f$  at  $(\lambda, \mu)$ . If  $\Gamma$  is a diagonal matrix with diagonal entries  $\lambda_i$ , 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

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

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

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

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  $H$

$$\|f(A + H) - f(A) - Df[A][H]\| = o(\|H\|). \quad (2.29)$$

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

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

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  $\Gamma$  and hermitian matrix  $H$ , we have

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

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

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

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  $\sum_{k=1}^n \Gamma_{ii}^{k-1} \Gamma_{jj}^{n-k}$ .  $\square$

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

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

*Proof.* Since

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

and the assertion follows from Lemma 2.9.  $\square$

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

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

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

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

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

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

$$\|df[A][H]\| \leq \max_{|t| \leq \|A\|} \|H\|. \quad (2.37)$$

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  $f_n \rightarrow f$  and  $f'_n \rightarrow f'$  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

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

since  $Df_n = df_n$  holds.

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

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

and also

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

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

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

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

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

so we have, using the triangle inequality

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

and use the above estimations to conclude that

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

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

**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

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

$\square$

**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 monotone 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 signs 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  $\epsilon \rightarrow 0$  we get that  $(t+\lambda)f(t)$  is operator convex.  $\square$

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

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

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

is defined to be its regularization.

The following nice properties are fulfilled by the family  $f_\epsilon$ :

1. Every  $f_\epsilon$  is a  $C^\infty$  function.
2. If the support of  $f$  is contained in a compact set, then the support of  $f_\epsilon$  is contained in an  $\epsilon$ -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  $\epsilon \rightarrow 0$ .
6. If  $f$  is differentiable, then  $(f_\epsilon)' = (f')_\epsilon$ .

7. If  $f$  is monotone, then  $f'_\epsilon(x) \rightarrow f'(x)$  as  $\epsilon \rightarrow 0$ , if  $f'(x)$  exists.

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

*Proof.* Let  $f_\epsilon$  be a regularization of  $f$  of order  $\epsilon$  for  $0 < \epsilon < 1$ . Then  $f_\epsilon$  is in the class  $C^\infty$  on  $(-1 + \epsilon, 1 - \epsilon)$ . It is also clearly operator monotone. Let  $\bar{f}(t) = \lim_{\epsilon \rightarrow 0} f_\epsilon(t)$ . 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,  $g_\epsilon$  is operator convex. Let  $\bar{g}(t) = \lim_{\epsilon \rightarrow 0} g_\epsilon(t)$ , 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

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

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  $\epsilon$  is small enough, then  $s, t \in (-1 + \epsilon, 1 - \epsilon)$ . Since  $f_\epsilon$  is operator monotone on this interval, the matrix  $f_\epsilon^{[1]}(A)$  is positive by Theorem 2.12, which implies that

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

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

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

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

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

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.  $\square$

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*

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

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

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

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

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

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

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  $\lambda_1, \dots, \lambda_{n+1} \in I$ . Let  $A$  be diagonal with entries  $\lambda_1, \dots, \lambda_{n+1}$ . 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

$$\begin{bmatrix} 0 & 0 & \dots & \bar{z}_1 \\ 0 & 0 & \dots & \bar{z}_2 \\ \cdot & \cdot & \dots & \cdot \\ z_1 & z_2 & \dots & z_n & 0 \end{bmatrix}, \quad (2.54)$$

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

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

for  $1 \leq i, j, k \leq n + 1$  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

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

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

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

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

**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.  $\square$

**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  $|\lambda| \leq 1$ .*

*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  $f_\epsilon$ , and apply the same argument to  $f_\epsilon(t) - f_\epsilon(0)$ , and then let  $\epsilon \rightarrow 0$ .  $\square$

**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  $g(t) = (1 + \frac{1}{t}) f(t)$  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.  $\square$

## 2.2 Loewner's Characterization

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

$$f(0) = 0, f'(0) = 1. \quad (2.59)$$

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*

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

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

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

is positive. Hence

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

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

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

Thus we obtain

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

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

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

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  $|f''(0)| \leq 2$ .  $\square$

**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  $(1 + \frac{1}{t}) f_i(t)$  is monotone on  $(-1, 1)$ . Since for all  $i$ ,  $\lim_{t \rightarrow 0} (1 + \frac{1}{t}) f_i(t) = f'_i(0) = 1$ , we see that  $(1 + \frac{1}{t}) f_i(t) \geq 1$  if  $t \geq 0$  and  $\leq 1$  if  $t \leq 0$ . Hence if  $t > 0$  we have  $(1 + \frac{1}{t}) f(t) \geq 1$ , and if  $t < 0$  we have the opposite inequality. Since  $f$  is continuously differentiable, this shows that  $f'(0) = 1$ .  $\square$

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

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

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

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

By Corollary 2.20,  $g_\lambda$  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  $h_\lambda$  defined as

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

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

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

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

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

from which we have that

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

$\square$

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

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

*Proof.* For  $-1 \leq \lambda \leq 1$ , 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  $h_\lambda$ . 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  $\{h_\lambda : -1 \leq \lambda \leq 1\}$  can also be written as  $\int h_\lambda d\nu(\lambda)$ , where  $\nu$  is a probability measure on  $[-1, 1]$  with finite support. Since  $f$  is in the closure of these combinations, there exists a net  $\nu_i$  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  $\nu_i$  has an accumulation point  $\mu$ . 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  $\mu_1$  and  $\mu_2$  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

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

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

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

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

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  $\mu$  on  $[-1, 1]$  such that*

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

*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  $K$ .  $\square$

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  $f(\frac{b-a}{2}t + \frac{a+b}{2})$  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

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

so  $f$  maps the upper half-plane into itself and maps the lower half-plane into itself as well. Similarly  $f(z) = \overline{f(\bar{z})}$ , 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 [7].

Furthermore consider the following nice

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

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

By change of variables we obtain from this that

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

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.

**Definition 3.1** (Matrix Mean). A two-variable function  $M: P(n, \mathbb{C}) \times P(n, \mathbb{C}) \mapsto P(n, \mathbb{C})$  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 [26]. 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

$$CM(A, B)C = M(CAC, CBC). \quad (3.1)$$

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 [26]). *For each connection  $\sigma$  and  $x > 0$  real number, the operator  $1\sigma x$  is a scalar. Furthermore the map,  $\sigma \mapsto f$ , defined by*

$$f(x) = 1\sigma x \quad (3.2)$$

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

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

$$PAP = AP \leq A \text{ and } PBP = BP \leq B. \quad (3.3)$$

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

$$P(A\sigma B)P \leq (PAP)\sigma(PBP) = (AP)\sigma(BP) \leq A\sigma B, \quad (3.4)$$

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

$$(I - P)[A\sigma B - P(A\sigma B)P]P, \quad (3.5)$$

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

$$[(AP)\sigma(BP)]P = (A\sigma B)P. \quad (3.6)$$

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 (3.6) that

$$I\sigma A = \sum_i [P_i\sigma(a_i P_i)]P_i = \sum_i (1\sigma a_i)P_i = \sum_i f(a_i)P_i = f(A), \quad (3.7)$$

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

$$A\sigma B = A^{1/2} f \left( A^{-1/2} B A^{-1/2} \right) A^{1/2}. \quad (3.8)$$

What remains to prove is that every operator monotone function is obtained in the form (3.2). Let  $f$  be an operator monotone function. Then it has an integral representation which can be written in the form

$$f(x) = \int_{[0, \infty]} \frac{x(1+t)}{x+t} dm(t) \quad (3.9)$$

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

$$A\sigma B = aA + bB + \int_{(0, \infty)} \frac{1+t}{t} ((tA)^{-1} + B^{-1})^{-1} dm(t), \quad (3.10)$$

where  $a = m(\{0\})$  and  $b = m(\{\infty\})$ . Since  $((tA)^{-1} + B^{-1})^{-1}$  and  $aA + bB$  satisfy conditions (ii), (iii) and (iv), the operation  $\sigma$  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  $\sigma$  is a connection. Finally for  $x > 0$

$$1\sigma x = f(x) \quad (3.11)$$

as well, so we obtain the function  $f$  from the connection.  $\square$

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  $\sigma$  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  $\sigma$  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*

$$A\sigma B = \frac{c}{2}(A+B) + \int_{(0,1]} \frac{1+t}{2t} \left[ ((tA)^{-1} + B^{-1})^{-1} + (A^{-1} + (tB)^{-1})^{-1} \right] dn(t), \quad (3.12)$$

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 (3.12) is a symmetric connection is straightforward. Conversely, let  $\sigma$  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

$$\begin{aligned} f(x) &= \frac{f(x) + xf(1/x)}{2} = \\ &= \frac{a+b}{2}(1+x) + \int_{(0,\infty)} (1+t) \left( \frac{x}{x+t} + \frac{x}{xt+1} \right) dm(t) = \\ &= \frac{a+b}{2}(1+x) + \int_{(0,\infty)} \frac{1+t}{2} \left( \frac{x}{x+t} + \frac{x}{xt+1} \right) dn(t), \end{aligned} \quad (3.13)$$

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

It remains to prove that a measure  $n$  producing  $\sigma$  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.  $\square$

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

$$\frac{2x}{1+x} \leq \frac{1+t}{2} \left( \frac{x}{x+t} + \frac{x}{xt+1} \right) \leq \frac{1+x}{2} \quad (3.14)$$

for  $x, t > 0$ , which yields

$$2(A^{-1} + B^{-1})^{-1} \leq \frac{1+t}{2t} \left[ ((tA)^{-1} + B^{-1})^{-1} + (A^{-1} + (tB)^{-1})^{-1} \right] \leq \frac{A+B}{2}. \quad (3.15)$$

The integration with respect to the probability measure  $n$  yields the assertion.  $\square$

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

$$G(A, B) = A^{1/2} \left( A^{-1/2} B A^{-1/2} \right) A^{1/2}. \quad (3.16)$$

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 Matrix Means 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** (Multi-variable 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  $P(r, \mathbb{C})$ ,
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.  $CM(X_1, \dots, X_n)C^* \leq M(CX_1C^*, \dots, CX_nC^*)$ .

The above properties are fulfilled by the n-variable harmonic and arithmetic mean. But what about the geometric mean (3.16)? 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 [9]. 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

$$\|A\|_2 = \sqrt{\text{Tr}\{A^2\}}, \quad (4.1)$$

where  $\text{Tr}$  denotes the trace of a squared matrix, that is  $\text{Tr}A = \sum_i A_{i,i}$ , 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

$$\langle X, Y \rangle_p = \text{Tr}\{p^{-1}Xp^{-1}Y\}, \quad (4.2)$$

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

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

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

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

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  $\gamma$ , the transformation  $p \mapsto X^*pX$  is an isometry of  $P(r, \mathbb{C})$ , that is*

$$L(\gamma) = L(X^* \gamma X), \quad (4.5)$$

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

*Proof.* We have for each  $t$  that

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

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

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

□

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

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

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  $\lambda_i$ . Then by Theorem 2.11

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

and the assertion follows from 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

$$L(\gamma) \geq \|\log A - \log B\|_2, \quad (4.12)$$

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

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

$$d(A, B) \geq \|\log A - \log B\|_2. \quad (4.13)$$

**Definition 4.2** (Geodesic). Let  $A, B \in P(r, \mathbb{C})$ . A path  $\gamma$  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

$$\gamma(t) = \exp(H(t)) \quad (4.14)$$

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

$$L(\gamma) = \int_0^1 \|\log A - \log B\|_2 dt = \|\log A - \log B\|_2. \quad (4.15)$$

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

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

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*

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

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

$$d(A, \gamma(t)) = td(A, B). \quad (4.18)$$

Moreover we have

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

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

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

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

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

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

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

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

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

$$\gamma'' = \gamma' \gamma^{-1} \gamma', \quad (4.23)$$

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

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

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

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

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

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

Since for general Riemannian manifolds the distance function is given by

$$d(p, q)^2 = \langle \log_p(q), \log_p(q) \rangle_p, \quad (4.27)$$

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 (3.16) 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

$$\langle X, Y \rangle_p = \text{Tr}\{XY\} \quad (4.28)$$

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

$$\gamma(t) = (1-t)A + tB \quad (4.29)$$

and the midpoint operation is the arithmetic mean.

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

$$\langle X, Y \rangle_p = \text{Tr}\{p^{-2}Xp^{-2}Y\}. \quad (4.30)$$

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 Riemannian 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*

$$\log B = -\frac{1-s}{s} \log A, \quad (4.31)$$

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

*Proof.* From Theorem 4.6 we know that

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

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

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

so  $A, B$  commute and (4.31) holds.  $\square$

By the above assertion and Proposition 4.5 it follows that  $\exp$  is isometric on straight line segments in  $H(r, \mathbb{C})$  going 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 [12].

An equivalent way (in the class of Riemannian manifolds) 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*

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

*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

$$d(A, B) = \|\log A - \log B\|_2. \quad (4.35)$$

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

$$d(M, C) = \|\log M - \log C\|_2. \quad (4.36)$$

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

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

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.  $\square$

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

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

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.** *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

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

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

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

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

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

Consequently,  $d(z_l, z_k) \rightarrow 0$ , 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  $\alpha \leq C(z_{\frac{1}{2}}) < \frac{C(z_0) + C(z_1)}{2} = \alpha$ .  $\square$

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

**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

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

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*

$$\text{grad}C(x) = -2 \sum_{i=1}^n \log_x(p_i). \quad (4.43)$$

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

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

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

Since  $\text{grad}$  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

$$c_a(s, t) = \exp_a(s \log_a(\gamma(t))). \quad (4.45)$$

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

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

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

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

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

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

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

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

Now we calculate

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

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

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

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

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

Hence

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

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

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

This shows that  $gradd(a, p)^2 = -2 \log_p(a)$ .  $\square$

**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 find by solving the equation*

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

Let us do this for 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

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

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  $A_1, \dots, A_n \in P(r, \mathbb{C})$  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

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

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

**Proposition 4.13.** *Let  $d(X, Y)$  be defined as*

$$d(X, Y) = d_E(f(X), f(Y)), \quad (4.58)$$

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

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

is given as

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

*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

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

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

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

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

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. 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 respect to the metric (4.2) is the unique solution  $X \in P(r, \mathbb{C})$  of the equation

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

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

$$X = \prod_{i=1}^n X_i^{1/n}, \quad (4.64)$$

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 [28] using the Riemannian structure (4.2) and its non-positive curvature combined with a characterization 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 [30]. 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 arithmetic, 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

$$A_t(A, B) = (1 - t)A + tB, \quad (4.65)$$

while the weighted harmonic mean is given as

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

Using the Riemannian structure the weighted geometric mean is

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

Although in [36] it has been shown that it is possible to define a weighted mean corresponding to a symmetric one without the use of a corresponding Riemannian structure. This weighted mean procedure gives back the above weighted means corresponding to their symmetric counterparts. 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

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

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

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

**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 [27] it was proved 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. In [36] it was also shown that the above procedure converges for every symmetric matrix mean. The convergence in Theorem 4.14 was shown to be linear, and it seems that this is the case for general symmetric matrix means. Since the procedure recursively relies on itself, it is quite ineffective even for small  $n$ . Hence in [11] the following procedure was defined. The following procedure is similar to the ALM-process. Both of these procedures are referred to in general as symmetrization procedures.

**Definition 4.5.** [BMP iteration] Let  $X = (X_1^0, \dots, X_n^0)$  where  $X_i^0 \in P(r)$  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

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

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

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

**Theorem 4.15** (Theorem 3.1 [11]). *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 important property of this procedure is essentially summarized in

**Theorem 4.16** (Theorem 3.2 [11]). *The procedure in Definition 4.5 converges cubically.*

The proofs of the above theorems were also relying on the metric structure (4.2), however in [36] the above two theorems were proved for all matrix means without using explicitly corresponding metric structures. The important properties of the ALM- and BMP-procedures are that their limit points fulfill the properties listed in Definition 4.1.

In [34] a new procedure was defined which directly extended 2-variable symmetric matrix means to several variables with a similar symmetrization procedure. It was shown that the properties in Definition 4.1 are also fulfilled except the permutation invariance. This procedure in [35] were also considered and proved to converge in complete geodesic metric spaces with a certain upper curvature bound.

Interestingly enough all the above extension procedures are the same for the arithmetic and harmonic means, they all give back the expected  $n$ -variable formulas. However in the case of the geometric mean they are all different as was pointed out for the first time in [10]. There are certain special situations, for instance when the matrices  $X_i$  commute, when they are also the same, and give back the usual geometric mean of scalars. It seems that curvature controls this phenomenon. If we have a corresponding metric structure which is Euclidean, then it can be shown that all the above extension are the same, the symmetrization procedures converge to the center of mass. Even spaces with constant curvature does not have this property [35].

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 these extension problems are far from being trivial. This leads us at last to the main question of this thesis, what are those matrix means which have corresponding Riemannian metric structures? Or at least which one of them has an affine geometric structure? In the following sections 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, providing the solution of this problem raised in [36]. 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 that all symmetric matrix means which are midpoint operations on  $P(r, \mathbb{C})$  have its corresponding affine connection of the form

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2} (X_p p^{-1} Y_p + Y_p p^{-1} X_p), \quad (4.72)$$

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 9.4 in the last section. During the classification process we will exhaustively study the properties of these possible connections, namely their parallel transports, metrizability, symmetricity, etc.

## 5 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  $T_p W$  at point  $p$ , i.e.

$$TW = \bigcup_{p \in W} \{p\} \times T_p W. \quad (5.1)$$

**Definition 5.1.** [Affine Connection] An affine connection (or Koszul connection)  $\nabla$  on a smooth manifold  $W$  is a mapping

$$\begin{aligned} C^\infty(W, TW) \times C^\infty(W, TW) &\mapsto C^\infty(W, TW) \\ (X, Y) &\mapsto \nabla_X Y \end{aligned} \quad (5.2)$$

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,  $\nabla$  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

$$\nabla_{\gamma'(t)} \gamma'(t) = 0. \quad (5.3)$$

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

$$\nabla_{\gamma'(t)} X(t) = 0. \quad (5.4)$$

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 [19, 18]. 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  $\nabla$ , in the fixed coordinate frame  $e_i = \frac{\partial}{\partial x^i}$ , we have corresponding Christoffel-symbols  $\Gamma_{jk}^i$  given as

$$\nabla_{e_j} e_k = \Gamma_{jk}^i e_i. \quad (5.5)$$

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

$$\nabla_{e_l} g_{ik} = 0. \quad (5.6)$$

From this we obtain the Christoffel-symbols in the form

$$\Gamma_{jk}^i = \frac{1}{2} g^{im} \left( \frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right), \quad (5.7)$$

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

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

$$\nabla_{e_l} X^m = \frac{\partial X^m}{\partial x^l} + \Gamma_{kl}^m X^k. \quad (5.8)$$

Similarly we define the covariant derivative of tensors as

$$\nabla_{e_l} A^{ik} = \frac{\partial A^{ik}}{\partial x^l} + \Gamma_{ml}^i A^{mk} + \Gamma_{ml}^k A^{im}. \quad (5.9)$$

For covariant tensors we have a negative sign before each  $\Gamma_{jk}^i$  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 [19].

**Theorem 5.1.** *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*

$$\log_p(q) = \lim_{n \rightarrow \infty} \frac{m(p, q)^{\circ n} - p}{\frac{1}{2^n}}, \quad (5.10)$$

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

*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

$$\frac{\partial \exp_p(Xt)}{\partial t} \Big|_{t=0} = \lim_{t \rightarrow 0} \frac{\exp_p(Xt) - p}{t} = X, \quad (5.11)$$

where  $X \in T_p M$ . 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

$$\lim_{t \rightarrow 0} \frac{\exp_p(Xt) - p}{t} = \lim_{n \rightarrow \infty} \frac{\exp_p\left(X \frac{1}{2^n}\right) - p}{\frac{1}{2^n}} = \lim_{n \rightarrow \infty} \frac{m(p, q)^{\circ n} - p}{\frac{1}{2^n}}. \quad (5.12)$$

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

$$X = \lim_{t \rightarrow 0} \frac{\exp_p(Xt) - p}{t} = \lim_{n \rightarrow \infty} \frac{m(p, q)^{\circ n} - p}{\frac{1}{2^n}} = \log_p(q). \quad (5.13)$$

□

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 than the ending points, on the geodesic connecting two points in the normal neighborhood. This is summarized in the following

**Proposition 5.2.** *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*

$$\log_p(q) = \lim_{n \rightarrow \infty} \frac{m(p, q)_{t_0}^{\circ n} - p}{t_0^n}, \quad (5.14)$$

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 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

$$M(A, B) = A^{1/2} f \left( A^{-1/2} B A^{-1/2} \right) A^{1/2}, \quad (6.1)$$

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:

$$f(t) = \alpha + \beta t + \int_0^\infty \left( \frac{\lambda}{\lambda^2 + 1} - \frac{1}{\lambda + t} \right) d\mu(\lambda), \quad (6.2)$$

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

$$\int_0^\infty \frac{1}{\lambda^2 + 1} d\mu(\lambda) < \infty. \quad (6.3)$$

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 [36]:

**Definition 6.1** (Affine matrix mean). 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  $M(A, B)^{\circ n} = M(A, M(A, B)^{\circ(n-1)})$  as before.

**Theorem 6.1.** *Let  $M(A, B)$  be a symmetric matrix mean. Then*

$$\lim_{n \rightarrow \infty} \frac{M(A, B)^{\circ n} - A}{\frac{1}{2^n}} = A^{1/2} \log_I \left( A^{-1/2} B A^{-1/2} \right) A^{1/2} \quad (6.4)$$

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 (6.1) we can reduce the above problem to the right hand side of the following formula:

$$\frac{M(A, B)^{\circ n} - A}{\frac{1}{2^n}} = A^{1/2} \frac{f(A^{-1/2} B A^{-1/2})^{\circ n} - I}{\frac{1}{2^n}} A^{1/2}. \quad (6.5)$$

From now on we will explicitly use the notation  $g(t)^{\circ n} = g(g(t)^{\circ(n-1)})$  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  $f(t) = t f(1/t)$  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

$$\lim_{n \rightarrow \infty} \frac{f(X)^{\circ n} - I}{f'(1)^n}, \quad (6.6)$$

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

$$\lim_{n \rightarrow \infty} \frac{g(t)^{\circ n}}{g'(0)^n}, \quad (6.7)$$

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

$$x_n = g(x)^{\circ n} = g'(t_n)g(x)^{\circ(n-1)} = \prod_{i=1}^n g'(t_i)x, \quad (6.8)$$

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

$$\left| \frac{g(x)^{\circ n}}{g'(0)^n} - \frac{g(x)^{\circ m}}{g'(0)^m} \right|. \quad (6.9)$$

We argue as follows

$$\begin{aligned} \left| \frac{g(x)^{\circ n}}{g'(0)^n} - \frac{g(x)^{\circ m}}{g'(0)^m} \right| &= \frac{|g(x)^{\circ n} - g'(0)^{n-m}g(x)^{\circ m}|}{g'(0)^n} \leq \\ &\leq \frac{|\prod_{i=m+1}^n g'(t_i) - g'(0)^{n-m}| |\prod_{i=1}^m g'(t_i)|}{g'(0)^n} |x| = \\ &= \left| \prod_{i=m+1}^n \frac{g'(t_i)}{g'(0)} - 1 \right| \left| \prod_{i=1}^m \frac{g'(t_i)}{g'(0)} \right| |x|. \end{aligned} \quad (6.10)$$

Now uniform convergence follows if  $|\prod_{i=1}^{\infty} g'(t_i)/g'(0)| < \infty$  because then the tail  $\prod_{i=m+1}^{\infty} g'(t_i)/g'(0) \rightarrow 1$  so (6.9) can be arbitrarily small on any compact interval in  $(-1, \infty)$  by choosing a uniform  $m$ . By the continuity of  $g'(t)$  and  $x_n \rightarrow 0$  we have  $g'(t_i) \rightarrow g'(0)$  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

$$\frac{g'(t_i)}{g'(0)} = \frac{g'(0) + g''(t'_i)t_i}{g'(0)} \quad (6.11)$$

for  $0 < t'_i < t_i$ . What follows from this is that

$$\left| \prod_{i=N}^{\infty} \frac{g'(t_i)}{g'(0)} \right| \leq \prod_{i=N}^{\infty} \left( 1 + \frac{K_1 K_2}{g'(0)} q^{i-N} \right). \quad (6.12)$$

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  $f(t)$ .  $\square$

Actually the above proof works for a larger class of functions than the family of normalized operator monotone functions. The limit in (6.6) exists and it is a continuous function if the twice differentiable function  $f(t)$  has 1 as the only attractive fixed point and the derivative  $-1 < f'(t) < 1$ .

**Proposition 6.2.** *The limit function  $\log_I(t)$  in Theorem 6.1 maps  $P(n, \mathbb{C})$  to  $H(n, \mathbb{C})$  injectively and*

$$I - X^{-1} \leq \log_I(X) \leq X - I \quad (6.13)$$

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 non-negative 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 preceding 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 6.1. 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

$$\left( \frac{A^{-1} + B^{-1}}{2} \right)^{-1} \leq M(A, B) \leq \frac{A + B}{2} \quad (6.14)$$

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

$$\left( \frac{I + X^{-1}}{2} \right)^{-1} \leq f(X) \leq \frac{I + X}{2} \quad (6.15)$$

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 5.1 and 6.1 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 (6.15) yield (6.13). Now it remains an easy exercise to see that  $\log_I(X)$  cannot be constant on  $(0, \infty)$ , since then it would violate inequality (6.13).

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  $H(n, \mathbb{C})$ .  $\square$

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 6.2. 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

$$\log'_I(t) = \lim_{n \rightarrow \infty} \frac{\prod_{i=0}^{n-1} f'(f(t)^{\circ i})}{\frac{1}{2^n}}, \quad (6.16)$$

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 6.1 and  $f'(1) = 1/2$  by the symmetricity of the matrix mean. By these considerations we have just arrived at the following

**Proposition 6.3.** *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*

$$\begin{aligned} \exp_p(X) &= p^{1/2} \exp_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2} \\ \log_p(X) &= p^{1/2} \log_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2} \end{aligned} \quad (6.17)$$

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

$$\begin{aligned} p^{1/2} \exp_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2} &= p \exp_I(p^{-1} X) \\ p^{1/2} \log_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2} &= p \log_I(p^{-1} X). \end{aligned} \quad (6.18)$$

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.

## 7 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 very important, from certain point of views [12]. We will turn back to this later. What we do need is that the Levi-Civita connection corresponding to this Riemannian space is

$$\nabla_{X_p} Y_p = D Y[p][X_p] - \frac{1}{2} (X_p p^{-1} Y_p + Y_p p^{-1} X_p), \quad (7.1)$$

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 6.1. The calculation of the limit (6.6) might be complicated. We give examples where the limit function may be calculated relatively easily.

*Example 7.1.* Consider the one-parameter family of means

$$X^{1/2} \left( \frac{I + (X^{-1/2} Y X^{-1/2})^q}{2} \right)^{1/q} X^{1/2}. \quad (7.2)$$

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

$$f_q(x)^{\circ 3} = \left( \frac{1 + \frac{1 + \frac{1 + x^q}{2}}{2}}{2} \right)^{1/q} \quad (7.3)$$

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

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{f_q(x)^{\circ n} - 1}{f'(1)^n} &= \lim_{n \rightarrow \infty} \frac{\left( \frac{x^q}{2^n} + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^n} \right)^{1/q} - 1}{\frac{1}{2^n}} = \\ &= \lim_{n \rightarrow \infty} \frac{\left( \frac{x^q}{2^n} + 1 - \frac{1}{2^n} \right)^{1/q} - 1}{\frac{1}{2^n}} = \lim_{t \rightarrow 0} \frac{(tx^q - t + 1)^{1/q} - 1}{t} = \\ &= \frac{\partial(tx^q - t + 1)^{1/q}}{\partial t} \Big|_{t=0} = \frac{x^q - 1}{q}. \end{aligned} \quad (7.4)$$

In [18] and [19] 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

$$\nabla_{X_p} Y_p = \lim_{t \rightarrow 0} \frac{\Gamma_t^0(\gamma) Y_{\gamma(t)} - Y_{\gamma(0)}}{t}, \quad (7.5)$$

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  $\gamma$  from  $\gamma(t)$  to  $\gamma(s)$ , refer to [18, 19]. 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

$$S_p(q) = \exp_p(-\log_p(q)). \quad (7.6)$$

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

$$\Gamma_1^0(\gamma)Y = -\left.\frac{\partial S_{\gamma(1/2)}(\exp_q(Yt))}{\partial t}\right|_{t=0}, \quad (7.7)$$

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

$$\begin{aligned} \exp_p(X) &= p^{1/2} \exp_I(p^{-1/2}Xp^{-1/2})p^{1/2} = p \exp_I(p^{-1}X) \\ \log_p(X) &= p^{1/2} \log_I(p^{-1/2}Xp^{-1/2})p^{1/2} = p \log_I(p^{-1}X). \end{aligned} \quad (7.8)$$

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

$$S_p(q) = \exp_p(-\log_p(q)) = p^{1/2}S_I(p^{-1/2}qp^{-1/2})p^{1/2} = pS_I(p^{-1}q). \quad (7.9)$$

Now we are in position to prove the following

**Theorem 7.1.** *Let  $P(n, \mathbb{C})$  be subset of an affine symmetric space with affine geodesic symmetries given as (7.9). Then the invariant affine connection has the form*

$$\nabla_{X_p}Y_p = DY[p][X_p] - \frac{\kappa}{2}(X_pp^{-1}Y_p + Y_pp^{-1}X_p), \quad (7.10)$$

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

*Proof.* We are going to use (7.7) to obtain the connection (7.10). We make the assumption that the geodesic symmetries are of the form (7.9). The functions  $\exp_p(X)$  and  $\log_p(X)$  are of the form (7.8), 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

$$\left.\frac{\partial \exp_I(t)}{\partial t}\right|_{t=0} = 1. \quad (7.11)$$

First of all we have to differentiate the map  $S_p(q)$  given in (7.9) to obtain  $\Gamma_1^0(\gamma)Y = T_{q \rightarrow p}Y$ , where  $\gamma(t)$  is a geodesic connecting  $p = \gamma(0)$  and  $q = \gamma(1)$ .

$$\begin{aligned} \frac{\partial S_p(\exp_q(Yt))}{\partial t} \Big|_{t=0} &= \frac{\partial pS_I(p^{-1}\exp_q(Yt))}{\partial t} \Big|_{t=0} = \\ &= pDS_I[p^{-1}q][p^{-1}Y] \end{aligned} \quad (7.12)$$

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 (7.7) to get back the connection. We use the holomorphic functional calculus to express the Fréchet-differential in (7.12) as

$$DS_I[X][U] = \frac{1}{2\pi i} \int_g S_I(z)[zI - X]^{-1}U[zI - X]^{-1}dz. \quad (7.13)$$

It is also easy to see that  $DS_I[I][I] = S'_I(1) = -1$ , so we may express the limit (7.7) by the following differential

$$\nabla_{\gamma'(0)}Y_{\gamma(0)} = - \frac{\partial \gamma(t/2)DS_I[\gamma(t/2)^{-1}\gamma(t)][\gamma(t/2)^{-1}Y_{\gamma(t)}]}{\partial t} \Big|_{t=0} = \quad (7.14)$$

we massage this further by using the holomorphic functional calculus

$$\begin{aligned} &= -\frac{\partial}{\partial t} \gamma(t/2) \frac{1}{2\pi i} \int_g S_I(z)[zI - \gamma(t/2)^{-1}\gamma(t)]^{-1}\gamma(t/2)^{-1}Y_{\gamma(t)} \\ &[zI - \gamma(t/2)^{-1}\gamma(t)]^{-1}dz \Big|_{t=0} = -\frac{1}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)}DS_I[I][I] - \\ &-\gamma(0)\frac{1}{2\pi i} \int_g S_I(z) \left\{ [zI - I]^{-1} \frac{1}{2}\gamma(0)^{-1}\gamma'(0)[zI - I]^{-1}\gamma(0)^{-1}Y_{\gamma(0)}[zI - I]^{-1} + \right. \\ &+ [zI - I]^{-1}\gamma(0)^{-1}Y_{\gamma(0)}[zI - I]^{-1}\frac{1}{2}\gamma(0)^{-1}\gamma'(0)[zI - I]^{-1} + \\ &+ [zI - I]^{-1} \left[ -\gamma(0)^{-1}\frac{1}{2}\gamma'(0)\gamma(0)^{-1}Y_{\gamma(0)} + \gamma(0)^{-1}DY[\gamma(0)][\gamma'(0)] \right] \\ &\left. [zI - I]^{-1} \right\} dz = \end{aligned} \quad (7.15)$$

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

get

$$\begin{aligned}
&= -\frac{DS_I[I][I]}{2} \gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} - \\
&\quad - \gamma(0) \frac{1}{2\pi i} \int_g \frac{S_I(z) dz}{(z-1)^3} \frac{1}{2} \gamma(0)^{-1} \gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} - \\
&\quad - \gamma(0) \frac{1}{2\pi i} \int_g \frac{S_I(z) dz}{(z-1)^3} \frac{1}{2} \gamma(0)^{-1} Y_{\gamma(0)} \gamma'(0) \gamma(0)^{-1} - \\
&\quad - \gamma(0) \frac{1}{2\pi i} \int_g \frac{S_I(z) dz}{(z-1)^2} \left[ -\frac{1}{2} \gamma(0)^{-1} \gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} + \gamma(0)^{-1} D Y[\gamma(0)][\gamma'(0)] \right]
\end{aligned} \tag{7.16}$$

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.

$$\begin{aligned}
\nabla_{\gamma'(0)} Y_{\gamma(0)} &= -\frac{S_I''(1)}{4} [\gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} + Y_{\gamma(0)} \gamma(0)^{-1} \gamma'(0)] - \\
&\quad - \frac{S_I'(1)}{2} \gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} - \frac{S_I'(1)}{2} [-\gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} + 2 D Y[\gamma(0)][\gamma'(0)]] = \\
&= -S_I'(1) D Y[\gamma(0)][\gamma'(0)] - \frac{S_I''(1)}{4} [\gamma'(0) \gamma(0)^{-1} Y_{\gamma(0)} + Y_{\gamma(0)} \gamma(0)^{-1} \gamma'(0)].
\end{aligned} \tag{7.17}$$

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 (7.10). We are going to study these connections as  $\kappa$  being a parameter. We will find out in the next section for which values of  $\kappa$  are these spaces symmetric. Also for arbitrary real  $\kappa$  (7.10) defines an affine connection with corresponding exponential and logarithm map which are of the form (7.8). This fact follows from considering the geodesic equations for the curves  $\gamma_1(t) = \exp_I(p^{-1/2} X p^{-1/2} t)$  and  $\gamma_2(t) = \exp_p(Xt) = p^{1/2} \exp_I(p^{-1/2} X p^{-1/2} t) p^{1/2}$ . We will also determine if these connections are metric or not.

## 8 Properties of These Affine Connections

Our next step is to integrate the geodesic equations corresponding to the one parameter family of connections (7.10).

**Theorem 8.1.** *The geodesic equation corresponding to the affine connection (7.10) is*

$$\ddot{\gamma} = \kappa \dot{\gamma} \gamma^{-1} \dot{\gamma}. \quad (8.1)$$

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

$$\exp_I(X) = \begin{cases} [(1 - \kappa)X + 1]^{\frac{1}{1-\kappa}} & \text{if } \kappa \neq 1, \\ \exp(X) & \text{else.} \end{cases} \quad (8.2)$$

*Proof.* First of all note that by (7.8) it is enough to solve the equation (8.1) for real numbers. Therefore the equation takes the form

$$\exp_I''(t) = \kappa \exp_I'(t)^2 \exp_I(t)^{-1}. \quad (8.3)$$

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

$$\log_I''(t) = -\kappa \log_I'(t)t^{-1}. \quad (8.4)$$

Solving the above we get the logarithm map as

$$\log_I(X) = \begin{cases} \frac{X^{1-\kappa}-1}{1-\kappa} & \text{if } \kappa \neq 1, \\ \log(X) & \text{else.} \end{cases} \quad (8.5)$$

From this by inverting the above function we get the assertion.  $\square$

Since we have integrated the geodesic equations we can get back the means which induce these affinely connected manifolds using (7.8)

$$\begin{aligned} M(X, Y) &= \exp_X \left( \frac{1}{2} \log_X(Y) \right) = \\ &= \begin{cases} X^{1/2} \left[ \frac{I + (X^{-1/2} Y X^{-1/2})^{1-\kappa}}{2} \right]^{\frac{1}{1-\kappa}} X^{1/2} & \text{if } \kappa \neq 1, \\ X^{1/2} (X^{-1/2} Y X^{-1/2})^{1/2} X^{1/2} & \text{else.} \end{cases} \end{aligned} \quad (8.6)$$

The above functions are matrix means if  $\kappa \in [0, 2]$ , see exercise 4.5.11 [9]. For other values of  $\kappa$  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

$$A_t(A, B) = (1 - t)A + tB \quad (8.7)$$

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

$$H_t(A, B) = ((1 - t)A^{-1} + tB^{-1})^{-1} \quad (8.8)$$

is also a geodesic with respect to the metric  $\langle X, Y \rangle_p = \text{Tr} \{ p^{-2} X p^{-2} Y \}$ . 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

$$G_t(A, B) = A^{1/2} \left( A^{-1/2} B A^{-1/2} \right)^t A^{1/2}. \quad (8.9)$$

The corresponding Riemannian metric is  $\langle X, Y \rangle_p = \text{Tr} \{ p^{-1} X p^{-1} Y \}$ . 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 [16]. 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  $R_{jkl}^i$  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

$$g_{sl} R_{ikl}^s + g_{is} R_{jkl}^s = 0 \quad (8.10)$$

also satisfies the system of equations

$$g_{sl} R_{ikl;m}^s + g_{is} R_{jkl;m}^s = 0, \quad (8.11)$$

here we use the Einstein summation convention for repeated indices and the semicolon ; to denote the covariant differentiation with respect to the index which follows the semicolon. The above theorem can be found in [40] as Theorem 1.3. Similarly one may also consult the classical paper [16].

In our case it turns out that

$$\begin{aligned} \Gamma_{jk}^i E_i &= -\frac{\kappa}{2} (E_j p^{-1} E_k + E_k p^{-1} E_j) \\ R_{jkl}^i E_i &= \left( \frac{\kappa}{2} - \frac{\kappa^2}{4} \right) (E_j p^{-1} E_k p^{-1} E_l + E_l p^{-1} E_k p^{-1} E_j - \\ &\quad - E_j p^{-1} E_l p^{-1} E_k - E_k p^{-1} E_l p^{-1} E_j), \end{aligned} \quad (8.12)$$

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  $R_{jkl;m}^s$ , since it vanishes everywhere

if and only if the underlying manifold is a symmetric space. So we have by definition

$$\begin{aligned} R_{jkl;m}^i E_i &= \frac{\partial R_{jkl}^i}{\partial x^m} E_i + \Gamma_{nm}^i R_{jkl}^n E_i - \\ &\quad - \Gamma_{jm}^n R_{nkl}^i E_i - \Gamma_{km}^n R_{jnl}^i E_i - \Gamma_{lm}^n R_{jkm}^i E_i. \end{aligned} \quad (8.13)$$

After some calculations one gets the following formulas

$$\begin{aligned} \frac{\partial R_{jkl}^i}{\partial x^m} E_i &= \\ &= \left( \frac{\kappa}{2} - \frac{\kappa^2}{4} \right) (-E_j p^{-1} E_m p^{-1} E_k p^{-1} E_l - E_j p^{-1} E_k p^{-1} E_m p^{-1} E_l + \\ &\quad + E_j p^{-1} E_m p^{-1} E_l p^{-1} E_k + E_j p^{-1} E_l p^{-1} E_m p^{-1} E_k - \\ &\quad - E_l p^{-1} E_m p^{-1} E_k p^{-1} E_j - E_l p^{-1} E_k p^{-1} E_m p^{-1} E_j + \\ &\quad + E_k p^{-1} E_m p^{-1} E_l p^{-1} E_j + E_k p^{-1} E_l p^{-1} E_m p^{-1} E_j) \\ \Gamma_{nm}^i R_{jkl}^n E_i &= \\ &= -\frac{\kappa}{2} \left( \frac{\kappa}{2} - \frac{\kappa^2}{4} \right) (E_j p^{-1} E_k p^{-1} E_l p^{-1} E_m - E_j p^{-1} E_l p^{-1} E_k p^{-1} E_m + \\ &\quad + E_l p^{-1} E_k p^{-1} E_j p^{-1} E_m - E_k p^{-1} E_l p^{-1} E_j p^{-1} E_m + \\ &\quad + E_m p^{-1} E_j p^{-1} E_k p^{-1} E_l - E_m p^{-1} E_j p^{-1} E_l p^{-1} E_k + \\ &\quad + E_m p^{-1} E_l p^{-1} E_k p^{-1} E_j - E_m p^{-1} E_k p^{-1} E_l p^{-1} E_j). \end{aligned} \quad (8.14)$$

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

**Proposition 8.2.** *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 8.3.** *Let  $c(t)$  be a geodesic with respect to the connection (7.10) and  $c(0) = p, c(1) = I$ . Then the unique solution of  $\nabla_{\dot{c}(t)} Y = 0$  with respect to the connection (7.10) and the initial condition  $Y_{c(0)} = Y_0$  is the vector field*

$$Y(t) = c(t)^{-\frac{\kappa}{2}} Y_0 c(t)^{-\frac{\kappa}{2}}. \quad (8.15)$$

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

$$DY[c(t)][c'(t)] - \frac{\kappa}{2} (c'(t)c(t)^{-1} Y_{c(t)} + Y_{c(t)} c(t)^{-1} c'(t)) = 0. \quad (8.16)$$

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

$$Y(t) = f(c(t)) Y_0 f(c(t)), \quad (8.17)$$

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

$$DY[c(t)][c'(t)] = \frac{dY(t)}{dt} = \frac{df(c(t))}{dt} Y_0 f(c(t)) + f(c(t)) Y_0 \frac{df(c(t))}{dt}. \quad (8.18)$$

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

$$\frac{df(c(t))}{dt} Y_0 f(c(t)) + f(c(t)) Y_0 \frac{df(c(t))}{dt} = \frac{\kappa}{2} (c'(t) c(t)^{-1} Y_{c(t)} + Y_{c(t)} c(t)^{-1} c'(t)). \quad (8.19)$$

Since  $c(t) = \exp_I((1-t) \log_I(p))$ , it has a power series expansion, as has  $f(x)$ , so we have by commutativity that

$$\frac{\kappa}{2} c'(t) c(t)^{-1} f(c(t)) = \frac{df(c(t))}{dt} = Df[c(t)][c'(t)] = f'(c) c'(t). \quad (8.20)$$

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

$$\frac{\kappa}{2} c^{-1} = f'(c) f(c)^{-1}, \quad (8.21)$$

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

$$\left\langle p^{-\kappa/2} X p^{-\kappa/2}, p^{-\kappa/2} Y p^{-\kappa/2} \right\rangle \quad (8.22)$$

for some positive definite bilinear form  $\langle \cdot, \cdot \rangle$  given on the tangent space at  $I$ . Due to the properties of  $R_{jkl}^i$  we conclude that a trivial solution of (8.10) is the mapping  $\text{Tr}\{XY\}$  at  $I$ , since  $R_{jkl}^i$  is the same for all  $\kappa$  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  $\text{Tr}\{XY\}$  at  $I$ . But it is easy to see that  $\text{Tr}\{XY\}$  is not a solution of (8.11) at  $I$  if  $\kappa \neq 0, 1, 2$ . Similarly  $\text{Tr}\{p^{-1} X p^{-1} Y\}$  is a solution of (8.10) at arbitrary  $p$  for  $\kappa \neq 0, 2$ , but it is not a solution of (8.11) if  $\kappa \neq 1$ . So we conclude

**Proposition 8.4.** *The smooth manifolds with affine connection (7.10) does not carry a Riemannian metric unless  $\kappa = 0, 1, 2$ .*

So remarkably we have not found any previously unknown matrix means so far which are midpoint maps 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.

## 9 Classification of Affine Matrix Means

Due to Proposition 6.3 we have the exponential and logarithm map in the form

$$\begin{aligned}\exp_p(X) &= p^{1/2} \exp_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2} \\ \log_p(X) &= p^{1/2} \log_I \left( p^{-1/2} X p^{-1/2} \right) p^{1/2}\end{aligned}\tag{9.1}$$

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 (9.1) 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

$$\exp''_I(t) + \Gamma(\exp'_I(t), \exp'_I(t), \exp_I(t)) = 0, \tag{9.2}$$

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 Proposition 15 and Corollary 16 of Chapter 6 in [39] 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  $\Gamma^i_{jk}$ , 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 9.1.** *Suppose that  $\Gamma(\cdot, \cdot, \cdot)$ ,  $\exp_I(\cdot)$ ,  $\exp_p(\cdot)$  are functions given with the above properties. Then*

$$\Gamma(X, X, p) = p^{1/2} \Gamma \left( p^{-1/2} X p^{-1/2}, p^{-1/2} X p^{-1/2}, I \right) p^{1/2} \tag{9.3}$$

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

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

$$\begin{aligned}\gamma'(t) &= p^{-1/2} X p^{-1/2} \exp'_I \left( p^{-1/2} X p^{-1/2} t \right) \\ \gamma''(t) &= p^{-1/2} X p^{-1/2} \exp''_I \left( p^{-1/2} X p^{-1/2} t \right) p^{-1/2} X p^{-1/2}.\end{aligned}\tag{9.4}$$

By the geodesic equations we have

$$\begin{aligned}\gamma''(t) &= -\Gamma(\gamma'(t), \gamma'(t), \gamma(t)) \\ \exp''_I \left( p^{-1/2} X p^{-1/2} t \right) &= -p^{1/2} X^{-1} p^{1/2} \Gamma \left( p^{-1/2} X p^{-1/2} \exp'_I \left( p^{-1/2} X p^{-1/2} t \right), \right. \\ &\quad \left. p^{-1/2} X p^{-1/2} \exp'_I \left( p^{-1/2} X p^{-1/2} t \right), \exp_I \left( p^{-1/2} X p^{-1/2} t \right) \right) p^{1/2} X^{-1} p^{1/2}.\end{aligned}\tag{9.5}$$

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

$$\begin{aligned} \exp_I''(p^{-1/2}Xp^{-1/2}t) &= -p^{1/2}X^{-1}\Gamma\left(Xp^{-1/2}\exp_I'(p^{-1/2}Xp^{-1/2}t)p^{1/2},\right. \\ &\quad \left.p^{1/2}\exp_I'(p^{-1/2}Xp^{-1/2}t)p^{-1/2}X, p^{1/2}\exp_I(p^{-1/2}Xp^{-1/2}t)p^{1/2}\right)X^{-1}p^{1/2}. \end{aligned} \quad (9.6)$$

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

$$\begin{aligned} p^{1/2}X^{-1}p^{1/2}\Gamma\left(p^{-1/2}Xp^{-1/2}, p^{-1/2}Xp^{-1/2}, I\right)p^{1/2}X^{-1}p^{1/2} &= \\ &= p^{1/2}X^{-1}\Gamma(X, X, p)X^{-1}p^{1/2}, \end{aligned} \quad (9.7)$$

which proves the assertion.  $\square$

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 9.2.** *For all  $p \in P(n, \mathbb{C})$  and  $X \in H(n, \mathbb{C})$  and invertible  $S$  we have*

$$\Gamma(SXS^{-1}, SXS^{-1}, SpS^{-1}) = S\Gamma(X, X, p)S^{-1}. \quad (9.8)$$

*Proof.* We have by the geodesic equations

$$\begin{aligned} X^2\exp_I''(Xt) &= -\Gamma(X\exp_I'(Xt), X\exp_I'(Xt), \exp_I(Xt)) \\ SX^2\exp_I''(Xt)S^{-1} &= -S\Gamma(X\exp_I'(Xt), X\exp_I'(Xt), \exp_I(Xt))S^{-1}. \end{aligned} \quad (9.9)$$

Similarly if we consider the geodesic equations for the curve  $\exp_I(SXS^{-1}t)$  we get

$$\begin{aligned} SX^2S^{-1}\exp_I''(SXS^{-1}t) &= -\Gamma(SXS^{-1}\exp_I'(SXS^{-1}t), SXS^{-1}\exp_I'(SXS^{-1}t), \\ &\quad \exp_I(SXS^{-1}t)) \\ SX^2\exp_I''(Xt)S^{-1} &= -\Gamma(SX\exp_I'(Xt)S^{-1}, SX\exp_I'(Xt)S^{-1}, \\ &\quad S\exp_I(Xt)S^{-1}). \end{aligned} \quad (9.10)$$

Again since the above two equations are identical we get the assertion.  $\square$

By the above proposition we have for hermitian  $X$  that

$$\Gamma(X, X, I) = U\Gamma(D, D, I)U^*, \quad (9.11)$$

for some diagonal  $D$  and unitary  $U$ , so it is enough to characterize  $\Gamma(X, X, I)$  for diagonal  $X$ .

**Theorem 9.3.** *Let  $D$  be diagonal with real coefficients. Then*

$$\Gamma(D, D, I) = -cD^2, \quad (9.12)$$

for some real valued constant  $c$ .

*Proof.* First we will show that  $\Gamma(I, I, I) = cI$  for some real constant  $c$ . Consider the case when  $\gamma(t) = \exp_I(\lambda It)$  for some real  $\lambda$ . Then by the geodesic equations for  $\gamma(t)$  we have

$$\lambda^2 \exp_I''(\lambda It) = -\Gamma(\lambda \exp_I'(\lambda It), \lambda \exp_I'(\lambda It), \exp_I(\lambda It)). \quad (9.13)$$

By linearity of  $\Gamma(\cdot, \cdot, \cdot)$  in the first two variables, this is equivalent to

$$\lambda^2 \exp_I''(\lambda It) = -\lambda^2 \Gamma(\exp_I'(\lambda It), \exp_I'(\lambda It), \exp_I(\lambda It)). \quad (9.14)$$

Letting  $t = 0$  we get

$$cI = -\Gamma(I, I, I), \quad (9.15)$$

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  $\Gamma(P, P, I) = -cP$ . Consider again  $\gamma(t) = \exp_I(Pt)$ . Then the geodesic equations read

$$P^2 \exp_I''(Pt) = -\Gamma(P \exp_I'(Pt), P \exp_I'(Pt), \exp_I(Pt)). \quad (9.16)$$

Since  $P^2 = P$  and again letting  $t = 0$  we get

$$cP = -\Gamma(P, P, I), \quad (9.17)$$

where  $c$  is trivially the same constant as determined above for  $\Gamma(I, I, I)$ . Now suppose that we have two mutually orthogonal projections  $P_1, P_2$  such that  $P_1 P_2 = 0$ . Then we have for the projection  $P_1 + P_2$  using linearity of  $\Gamma(\cdot, \cdot, \cdot)$  in the first two variables that

$$\begin{aligned} \Gamma(P_1, P_1, I) + \Gamma(P_2, P_2, I) &= -c(P_1 + P_2) = \Gamma(P_1 + P_2, P_1 + P_2, I) = \\ &= \Gamma(P_1, P_1, I) + \Gamma(P_1, P_2, I) + \Gamma(P_2, P_1, I) + \Gamma(P_2, P_2, I), \end{aligned} \quad (9.18)$$

which yields that for mutually orthogonal projections  $P_1, P_2$  we get the orthogonality relation

$$\Gamma(P_1, P_2, I) = 0. \quad (9.19)$$

Finally since a diagonal  $D$  can be written as  $D = \sum_i \lambda_i P_i$  for mutually orthogonal projections  $P_i$ , we have

$$\begin{aligned} \Gamma(D, D, I) &= \Gamma\left(\sum_i \lambda_i P_i, \sum_i \lambda_i P_i, I\right) = \\ &= \sum_i \lambda_i^2 \Gamma(P_i, P_i, I) = -\sum_i \lambda_i^2 c P_i = \\ &= -cD^2, \end{aligned} \quad (9.20)$$

which is what needed to be shown.  $\square$

The above three theorems with the other preceding results presented here, lead us to the concluding

**Theorem 9.4.** *All affine matrix means  $M(X, Y)$  are of the form*

$$M(X, Y) = \begin{cases} X^{1/2} \left[ \frac{I + (X^{-1/2} Y X^{-1/2})^{1-\kappa}}{2} \right]^{\frac{1}{1-\kappa}} X^{1/2} & \text{if } \kappa \neq 1, \\ X^{1/2} (X^{-1/2} Y X^{-1/2})^{1/2} X^{1/2} & \text{if } \kappa = 1, \end{cases} \quad (9.21)$$

where  $0 \leq \kappa \leq 2$ . The symmetric affine connections corresponding to these means are

$$\nabla_{X_p} Y_p = DY[p][X_p] - \frac{\kappa}{2} (X_p p^{-1} Y_p + Y_p p^{-1} X_p). \quad (9.22)$$

*Proof.* By Proposition 9.1, 9.2 and Theorem 9.3 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

$$\Gamma(X, X, p) = -c X p^{-1} X. \quad (9.23)$$

This formula determines the functions that are the symmetric parts of the possible connections, and these connections have geodesics determined by Theorem 8.1 in the form (9.21). Again by Proposition 15 and Corollary 16 of Chapter 6 in [39] 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 (9.22) are symmetric, affine and have the same geodesics, therefore they give the sought symmetric connections for each  $\kappa$  if we choose  $c = \kappa$ .

The corresponding midpoint operations have the form (8.6), and these are matrix means if and only if  $\kappa \in [0, 2]$ , since the representing functions  $f(t)$  in (6.1) turn out to be operator monotone only in these cases (see again exercise 4.5.11 [9]). This gives us the complete classification of affine matrix means.  $\square$

## 10 Conclusions

The results presented in Section 5-9 are completely new and have been taken from the article [37] of the author. From the point of view of the applicability of Riemannian structures in the theory of matrix means, it is a bit disappointing result, since it turns out that there exist no other Riemannian structures than the ones corresponding to the arithmetic, the harmonic and the geometric mean. However it may turn out that there are Finsler structures, this can be taken as a future research project. If the kind reader consults [1, 5, 10, 11, 15, 17, 30, 34, 35, 36], it is evident that the extension of 2-variable matrix means to several variables are governed by geometrical analogies and most of the constructions are inherited from a geometrical framework provided by the geometric mean. However the proofs tend to become more difficult if we abandon the realm of metric geometry.

## 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] R. BHATIA, *Matrix Analysis*, Springer-Verlag, New York, 1996.
- [8] R. BHATIA, *On the exponential metric increasing property*, Linear Algebra Appl., 375 (2003), pp. 211-220.
- [9] R. BHATIA, *Positive Definite Matrices*, Princeton University Press, 2007.
- [10] R. BHATIA AND J. HOLBROOK, *Riemannian geometry and matrix geometric means*, Linear Alg. Appl., 413 (2006), pp. 594-618.
- [11] 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.
- [12] 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.
- [13] D. BURAGO, YU D. BURAGO AND S. IVANOV, *A Course in Metric Geometry*, American Mathematical Society, 2001, ISBN 0-8218-2129-6.
- [14] É. CARTAN, *La topologie des espaces representatifs des groupes de Lie*, Enseignement Math., 35 (1936), pp. 177-200, Exposés de Géometrie VIII.
- [15] C. CONDE, *Differential Geometry for Nuclear Positive Operators*, Integral Equations and Operator Theory, Vol. 57, Num. 4 (2006), pp. 451-471.
- [16] 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.

- [17] 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.
- [18] H. W. GUGGENHEIMER, *Differential Geometry*, 378 p. Mc Graw-Hill, New York, (1977), ISBN: 978-0-486-63433-3.
- [19] S. HELGASON, *Differential Geometry, Lie Groups, and Symmetric Spaces*, 641 p. American Mathematical Society (2001), ISBN: 978-0-821-82848-9.
- [20] F. HIAI AND D. PETZ, *Riemannian metrics on positive definite matrices related to means*, Linear Algebra Appl., 430 (2009), pp. 3105–3130.
- [21] F. HIAI AND H. KOSAKI, *Means of Hilbert space operators*, Lecture Notes In Maths. 1820 (2003), Springer, ISBN: 978-3-540-40680-8.
- [22] J. JOST, *Nonpositive Curvature: Geometric and Analytic Aspects*, Birkhäuser-Verlag, Basel, 1997.
- [23] H. KARCHER, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math., vol. 30 (1977), pp. 509-541.
- [24] S. KOBAYASHI AND N. NOMIZU, *Foundations of Differential Geometry Vol. 1.*, 344 p. Wiley-Interscience (1996), ISBN: 978-0-471-15733-5.
- [25] S. KOBAYASHI AND N. NOMIZU, *Foundations of Differential Geometry Vol. 2.*, 488 p. Wiley-Interscience (1996), ISBN: 978-0-471-15732-8.
- [26] F. KUBO AND T. ANDO, *Means of positive linear operators*, Math. Ann., 246 (1980), pp. 205–224.
- [27] J. LAWSON AND Y. LIM, *A general framework for extending means to higher orders*, Colloq. Math., 113 (2008), pp. 191–221.
- [28] J. LAWSON AND Y. LIM, *Monotonic properties of the least squares mean*, Math. Ann. (2010), doi:10.1007/s00208-010-0603-6.
- [29] I. J. MADDOX, *Elements of functional analysis*, Cambridge University Press, Cambridge, 1988.
- [30] M. MOAKHER, *A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices*, SIMAX, 26 (2005), pp. 735–747.
- [31] M. MOAKHER, *Means and averaging in the group of rotations*, SIAM J. Matrix Anal. Appl., vol. 24, no. 1 (2002), pp. 1-16.
- [32] K.-H. NEEB, *A CartanHadamard Theorem for BanachFinsler Manifolds*, Geom. Ded., vol. 95, No. 1 (2002), pp. 115–156.
- [33] K. NOMIZU, *Invariant Affine Connections on Homogeneous Spaces*, American Journal of Mathematics, Vol. 76, No. 1 (1954), pp. 33–65.

- [34] M. PÁLFIA, *A multi-variable extension of two-variable matrix means*, SIAM Journal on Matrix Anal. Appl. (2011), in press.
- [35] M. PÁLFIA, *Means in metric spaces and the center of mass*, J. Math. Anal. Appl., 381 (2011), pp. 383–391.
- [36] M. PÁLFIA, *Weighted matrix means and symmetrization procedures*, submitted to Linear Alg. Appl. (2011).
- [37] M. PÁLFIA, *Classification of affine matrix means*, submitted to Mathematische Annalen (2011).
- [38] D. PETZ AND R. TEMESI, *Means of positive numbers and matrices*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 712–720.
- [39] M. SPIVAK, *A comprehensive introduction to differential geometry Vol. 2*, Publish or Perish, Houston, Texas (1979), ISBN: 0-914098-88-8.
- [40] A. VANŽUROVÁ AND P. ŽÁČKOVÁ, *Metrization of linear connections*, Journal of Applied Mathematics, Vol. 2, No. 1 (2009), pp. 151–164.