MINISTRY OF EDUCATION AND TRAINING
QUY NHON UNIVERSITY

VUONG TRUNG DUNG

SOME DISTANCE FUNCTIONS IN QUANTUM INFORMATION THEORY AND RELATED PROBLEMS

SPECIALITY: CALCULUS
CODE NO.: 9 46 01 02

SUMMARY OF DOCTORAL THESIS IN MATHEMATICS

BINH DINH - 2024

The work has been completed at:
Quy Nhon University

The Board of Supervisors:
1. Assoc. Prof. Dr. Le Cong Trinh
2. Assoc. Prof. Dr Dinh Trung Hoa

Reviewer 1: Prof. Dang Duc Trong

Reviewer 2: Prof. Pham Tien Son

Reviewer 3: Assoc. Prof. Pham Quy Muoi

The thesis shall be defended at the University level Thesis Assessment Council at Quy Nhon University at : ....

The thesis can be found at:
-National Library of Vietnam
- Learning Resources Center Quy Nhon University

UNDERTAKING FORM

This thesis has been completed within Quy Nhon University, under the supervisor of Assoc. Prof. Dr. Le Cong Trinh and Assoc. Prof. Dinh Trung Hoa. I hereby assure that this research project is mine. All results are honest, have been approved by co-authors and have not been released by anyone else before.

Author

Vuong Trung Dung

ACKNOWLEDGEMENTS

This thesis was undertaken during my years as a PhD student at the Department of Mathematics and Statistics, Quy Nhon University. Upon the completion of this thesis, I am deeply indebted to numerous individuals. On this occasion, I would like to extend my sincere appreciation to all of them.

First and foremost, I would like to express my sincerest gratitude to Assoc. Prof. Dr. Dinh Trung Hoa, who guided me into the realm of matrix analysis and taught me right from the early days. Not only that, but he also devoted a significant amount of valuable time to engage in discussions, and provided problems for me to solve. He motivated me to participate in workshops and establish connections with senior researchers in the field. He guided me to find enjoyment in solving mathematical problems and consistently nurtured my enthusiasm for my work. I can't envision having a more exceptional advisor and mentor than him.

The second person I would like to express my gratitude to is Assoc. Prof. Dr. Le Cong Trinh, who has been teaching me since my undergraduate days and also introduced me to Prof. Hoa. From the early days of sitting in lecture halls at university, Prof. Trinh has been instilling inspiration and a love for mathematics in me. It's fortunate that now I have the opportunity to be mentored by him once again. He has always provided enthusiastic support not only in my work but also in life. Without that dedicated support, it would have been difficult for me to complete this thesis.

I would like to extend a special thank you to the educators at both the Department of Mathematics-Statistic and the Department of Graduate Training at Quy Nhon University for providing the optimal environment for a postgraduate student who comes from a distant location like myself. Binh Dinh is also my hometown and the place where I have spent all my time from high school to university. The privilege and personal happiness of coming back to Quy Nhon University for advanced studies cannot be overstated.

I am grateful to the Board and Colleagues of VNU-HCM High School for the Gifted for providing me much supports to complete my PhD study. Especially, I would like to extend my heartfelt gratitude to Dr. Nguyen Thanh Hung, who has provided assistance to me in both material and spiritual aspects since the very first days I set foot in Saigon. He is not only a mentor and colleague but also a second father to me, who not only supported me financially and emotionally during challenging times but also constantly encouraged me to pursue a doctoral degree. Without this immense support and encouragement, I wouldn't be where I am today.

I also want to express my gratitude to Su for the wonderful time we've spent together, which has been a driving force for me to complete the PhD program and strive for even greater achievements that I have yet to attain.

Lastly, and most significantly, I would like to express my gratitude to my family. They have always been by my side throughout work, studies, and life. I want to thank my parents for giving birth to me and nurturing me to adulthood. This thesis is a gift I dedicate to them.

CONTENTS

Introduction

Quantum information stands at the confluence of quantum mechanics and information theory, wielding the mathematical elegance of both realms to delve into the profound nature of information processing at the quantum level. In classical information theory, bits are the fundamental units representing 0 and 1. Quantum information theory, however, introduces the concept of qubits, the quantum counterparts to classical bits. Unlike classical bits, qubits can exist in a superposition of states, allowing them to be both 0 and 1 simultaneously. This unique property empowers quantum computers to perform certain calculations exponentially faster than classical computers.

Entanglement is a crucial phenomenon in quantum theory where two or more particles become closely connected. When particles are entangled, changing the state of one immediately affects the state of the other, no matter the distance between them. This has important implications for quantum information and computing, offering new possibilities for unique ways of handling

Quantum algorithms, such as Shor's algorithm for factoring large numbers and Grover's algorithm for quantum search, exemplify the power of quantum information in tackling complex computational tasks with unparalleled efficiency.

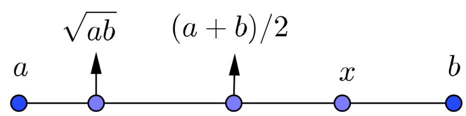
In order to treat information processing in quantum systems, it is necessary to mathematically formulate fundamental concepts such as quantum systems, states, and measurements, etc. Useful tools for researching quantum information are functional analysis and matrix theory. First, we consider the quantum system. It is described by a Hilbert space $\mathcal{H}$, which is called a *representation space*. This will be advantageous because it is not only the underlying basis of quantum mechanics but is also as helpful in introducing the special notation used for quantum mechanics. *The (pure) physical states* of the system correspond to unit vectors of the Hilbert space. This correspondence is not 1-1. When f_1 and f_2 are unit vectors, then the corresponding states are identical if $f_1 = zf_2$ for a complex number z of modulus 1 . Such z is often called *phase*. The pure physical state of the system determines a corresponding state vector up to a phase. Traditional quantum mechanics distinguishes between pure states and *mixed states*. Mixed states are described by *density matrices*. A density matrix or statistical operator is a positive matrix of trace 1 on the Hilbert space. This means that the space has a basis consisting of eigenvectors of the statistical operator and the sum of eigenvalues is 1. In quantum information theory, distance functions are used to measure the distance between two mixed states. Additionally, these distance functions can be employed to characterize the properties of a given quantum state. For instance, they can quantify the quantum entanglement between two parts of a state, representing the shortest distance between the state and the set of all separable states. These distance functions naturally extend to the set of positive semi-definite matrices, which is also the main focus of this thesis.

Nowadays, the significance of matrix theory has been widely recognized across various fields, including engineering, probability and statistics, quantum information, numerical analysis, biological and social sciences. In image processing (subdivision schemes), medical imaging (MRI), radar signal processing, statistical biology (DNA/genome), and machine learning, data from numerous experiments are stored as positive definite matrices. To work with each set of data, we need to select its representative element. In other words, we need to compute the average of the corresponding positive definite matrices. Therefore, considering global solutions of the least-squares problems for matrices is of paramount importance (refer to [2, 8, 18, 28, 67, 73] for examples).

Let $0 < a \leq x \leq b$. Consider the following least squares problem:

$$d^2(x, a) + d^2(x, b) \to \min, \quad x \in [a, b],$$

where $d := d_E(x, y) = |y - x|$, or, $d := d_R(x, y) := |\log(y) - \log(x)|$.



The arithmetic mean $(a+b)/2$ and the geometric mean $\sqrt{ab}$ are unique solutions to the above problem with respect to d_E and d_R distance, respectively. Moreover, based on the AM-GM inequality for two non-negative numbers a and b, we have a new distance as follows

$$d(a, b) = \frac{a + b}{2} - \sqrt{ab}.$$

For $A, B \in \mathbb{P}_n$, some matrix analogs of scalar distances are:

- Euclidean distance induced from Euclidean/Frobenius inner product $\langle A, B\rangle = \operatorname{Tr}(A^*B)$. The associated norm is $\|A\|_F = \langle A, A\rangle^{1/2} = (\operatorname{Tr}(A^*A))^{1/2}$.
- The Riemann distance [12] is $\delta_R(A, B) = || \log(A^{-1}B)||_2 = \left(\sum_{i=1}^{n} \log^2 \lambda_i(A^{-1}B)\right)^{1/2}$.
- Bures-Wasserstein distance [13] in the theory of optimal transport :

$$d_b(A, B) = \left(\operatorname{Tr}(A + B) - 2\operatorname{Tr}\left(\left(A^{1/2}BA^{1/2}\right)^{1/2}\right)\right)^{1/2}.$$

- The Log-Determinant metric [75] in machine learning and quantum information:

$$d_l(A, B) = \log\det\frac{A + B}{2} - 2\log\det(AB).$$

- The Hellinger metric or Bhattacharya metric [73] in quantum information :

$$d_h(A, B) = \left(\operatorname{Tr}(A + B) - 2\operatorname{Tr}\left(A^{1/2}B^{1/2}\right)\right)^{1/2}.$$

In applications, one are sometimes interested in distance-like functions that provide distance between two data points. Such functions are not necessarily symmetric; and the triangle inequality does not need to be true. Divergences [11] are such distance-like functions .

Definition. *A smooth function* $\Phi : \mathbb{P}_n \times \mathbb{P}_n \to \mathbb{R}^+$ *is called a quantum divergence if*

(i) $\Phi(A, B) = 0$ *if and only if* $A = B$.

(ii) The derivative $D\Phi$ *with respect to the second variable vanishes on the diagonal, i.e.,*

$$D\Phi(A, X)|_{X=A} = 0.$$

(iii) The second derivative $D^2\Phi$ *is positive on the diagonal, i.e.,*

$$D^2\Phi(A, X)|_{X=A}(Y, Y) \geq 0 \quad \text{for all Hermitian matrix } Y.$$

Some divergences that have recently received a lot of attention are in [11, 14, 35, 56].

Now let us revisit the scalar mean theory which serves as a starting point for our next problem in this thesis.

A *scalar mean* of non-negative numbers is a function from $\mathbb{R}^+ \times \mathbb{R}^+$to $\mathbb{R}^+$such that:

1) $M(x, x) = x$ for every $x \in \mathbb{R}^+$.

2) $M(x, y) = M(y, x)$ for every $x, y \in \mathbb{R}^+$.

3) If $x < y$, then $x < M(x, y) < y$.

4) If $x < x_0$ and $y < y_0$, then $M(x, y) < M(x_0, y_0)$.

5) $M(x, y)$ is continuous.

6) $M(tx, ty) = tM(x, y)$ for $t, x, y \in \mathbb{R}^+$.

A two-variable function $M(x, y)$ satisfying condition 6) can be reduced to a one-variable function $f(x) := M(1, x)$. Namely, $M(x, y)$ is recovered from f as $M(x, y) = xf(x^{-1}y)$. Notice that the function f, corresponding to M is monotone increasing on $\mathbb{R}^+$. And this relation forms a one-to-one correspondence between means and monotone increasing functions on $\mathbb{R}^+$.

The following are some desired properties of any object that is called a *mean* M on $\mathbb{H}_n^+$.

(A1). Positivity: $A, B \geqslant 0 \Rightarrow M(A, B) \geqslant 0$.

(A2). Monotonicity: $A \geqslant A', B \geqslant B' \Rightarrow M(A, B) \geqslant M(A', B')$.

(A3). Positive homogeneity: $M(kA, kB) = kM(A, B)$ for $k \in \mathbb{R}^+$.

(A4). Transformer inequality: $X^*M(A, B)X \leqslant M(X^*AX, X^*BX)$ for $X \in B(\mathcal{H})$.

(A5). Congruence invariance: $X^*M(A, B)X = M(X^*AX, X^*BX)$ for invertible $X \in B(\mathcal{H})$.

(A6). Concavity: $M\left(tA+(1-t)B, tA'+(1-t)B'\right) \geqslant tM\left(A, A'\right)+(1-t)M\left(B, B'\right)$ for $t \in [0,1]$.

(A7). Continuity from above: if $A_n \downarrow A$ and $B_n \downarrow B$, then $M\left(A_n, B_n\right) \downarrow M(A, B)$.

(A8). Betweenness: if $A \leqslant B$, then $A \leqslant M(A, B) \leqslant B$.

(A9). Fixed point property: $M(A, A)=A$.

To study matrix or operator means in general, we must first consider three classical means in mathematics: arithmetic, geometric, and harmonic means. These means are defined in the following manner, respectively,

$$A \nabla B=\frac{1}{2}(A+B),$$

$$A \sharp B=A^{1/2}\left(A^{-1/2} B A^{-1/2}\right)^{1/2} A^{1/2},$$

and

$$A!B=2(A^{-1}+B^{-1})^{-1}.$$

In the above definitions, if matrix A is not invertible, we replace A with $A_\epsilon=A+\epsilon I$ and then let ϵ tend to 0 (similarly for matrix B). It can be seen that the arithmetic, harmonic and geometric means share the properties (A1)-(A9) in common. In 1980, Kubo and Ando [54] developed an axiomatic theory of operator mean on $\mathbb{H}_n^+$. At first, they defined a *connection* of two matrices as follows (the term connection comes from the study of electrical network connections).

Definition. *A connection on $\mathbb{H}_n^+$ is a binary operation σ on $\mathbb{H}_n^+$ satisfying the following axioms for all $A, A', B, B', C \in \mathbb{H}_n^+$:*

(M1). Monotonicity: $A \leqslant A', B \leqslant B' \Longrightarrow A \sigma B \leqslant A' \sigma B'$.

(M2). Transformer inequality: $C(A \sigma B) C \leqslant (CAC) \sigma (CBC)$.

(M3). Joint continuity from above: if $A_n, B_n \in B(\mathcal{H})^+$satisfy $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \sigma B_n \downarrow A \sigma B$.

A *mean* is a connection with normalization condition

(M4) $I \sigma I=I$.

To each connection σ corresponds its *transpose* σ' defined by $A \sigma' B=B \sigma A$. A connection σ is *symmetric* by definition if $\sigma=\sigma'$. The *adjoint* of σ, denoted by σ^*, is defined by $A \sigma^* B=\left(A^{-1} \sigma B^{-1}\right)^{-1}$, for invertible A, B. When σ is a non-zero connection, its *dual*, in symbol $\sigma^{\perp}$, is defined by $\sigma^{\perp}=\left(\sigma'\right)^*=\left(\sigma^*\right)'$.

However, Kubo-Ando theory of means still has many limitations. In applied and engineering fields, people need more classes of means that are non Kubo-Ando. For some non Kubo-Ando means we refer the interested readers to [17, 23, 25, 35, 37].

One of the famous non-Kubo-Ando means is the spectral geometric mean [37], denoted as $A\natural B$, introduced in 1997 by Fiedler and Ptk . It is called the spectral geometric mean because $(A\natural B)^2$ is similar to AB and that the eigenvalues of their spectral mean are the positive square roots of the corresponding eigenvalues of AB. In 2015, Kim and Lee [52] defined the weighted spectral mean:

$$A\natural_t B := \left(A^{-1}\sharp B\right)^t A \left(A^{-1}\sharp B\right)^t, \quad t \in [0,1].$$

In this thesis we focus on two problems:

1. **Distance function generated by operator means.** We introduce some new distance on the set of positive definite matrices in the relation to operator means, and their applications. In addition, we also study some geometric properties for means such as the in-betweenness property, and data processing inequality in quantum information.

2. **A new weighted spectral geometric mean.** We introduce a new weighted spectral geometric mean, denoted by $\mathcal{F}_t(A,B)$ and study basic properties for this quantity. We also establish a weak log-majorization relation involving $\mathcal{F}_t(A,B)$ and the Lie-Trotter formula for $\mathcal{F}_t(A,B)$.

The main tools in our research are the spectral theorem for Hermitian matrices and the theory of Kubo-Ando means. Some fundamental techniques in the theory of operator monotone functions and operator convex functions are also utilized in the dissertation. We also employ basic knowledge in matrix theory involving unitarily invariant norms, trace, etc.

The main results in this thesis are presented in the following articles:

1. Vuong T.D., Vo B.K (2020), "An inequality for quantum fidelity", *Quy Nhon Univ. J. Sci.*, 4 (3).

2. Dinh T.H., Le C.T., Vo B.K, Vuong T.D. (2021), "Weighted Hellinger distance and in betweenness property", *Math. Ine. Appls.,* 24, 157-165.

3. Dinh T.H., Le C.T., Vo B.K., Vuong T.D. (2021), "The α-z-Bures Wasserstein divergence", *Linear Algebra Appl.,* 624, 267-280.

4. Dinh T.H., Le C.T., Vuong T.D., α-z-fidelity and α-z-weighted right mean, *Submitted.*

5. Dinh T.H., Tam T.Y., Vuong T.D, On new weighted spectral geometric mean, *Submitted.*

They were presented on the seminars at the Department of Mathematics and Statistics at Quy Nhon University and at the following international workshops and conferences as follows:

1. First SIBAU-NU Workshop on Matrix Analysis and Linear Algebra, 15-17 October, 2021.

2. 20th Workshop on Optimization and Scientific Computing, April 21-23, 2022 - Ba Vi, Vietnam.
3. International Workshop on Matrix Analysis and Its Applications, June 4, 2022, Quy Nhon, Viet Nam.
4. The second international workshop on Matrix Theory and Applications, AKFA University, November, 2022.
5. International Workshop on Matrix Analysis and Its Applications, July 7-8, 2023, Quy Nhon, Viet Nam.
6. 10th Viet Nam Mathematical Congress, August 8-12, 2023, Da Nang, Viet Nam.

This thesis has introduction, three chapters, conclusion, further investigation, a list of the author's papers related to the thesis and preprints related to the topics of the thesis, and a list of references.

The introduction provides a background on the topics covered in this work and explains why they are meaningful and relevant. It also briefly summarizes the content of the thesis and highlights the main results from the main three chapters.

Chapter 1

Preliminaries

1.1 Matrix theory fundamentals

Let $\mathbb{N}$ be the set of all natural numbers. For each $n \in \mathbb{N}$, we denote by $\mathbb{M}_n$ the algebra of all $n \times n$ complex matrices, $\mathbb{H}_n$ is the set of all $n \times n$ Hermitian matrices, $\mathbb{H}_n^+$ is the set of $n \times n$ positive semi-definite matrices, $\mathbb{P}_n$ is the cone of positive matrices in $\mathbb{M}_n$, and $\mathcal{D}_n$ is the set of density matrices which are the positive matrices with trace equal to one. Denote by I and O the identity and zero elements of $\mathbb{M}_n$, respectively. This thesis deals with problems for matrices, which are operators in finite-dimensional Hilbert spaces $\mathcal{H}$. We will indicate if the case is infinite-dimensional.

Recall that for two vectors $x = (x_j), y = (y_j) \in \mathbb{C}^n$ the inner product $\langle x, y \rangle$ of x and y is defined as $\langle x, y \rangle \equiv \sum_j x_j \bar{y}_j$. Now let A be a matrix in $\mathbb{M}_n$, the conjugate transpose or the adjoint A^* of A is the complex conjugate of the transpose A^T. We have, $\langle Ax, y \rangle = \langle x, A^* y \rangle$.

Definition 1.1.1. A matrix $A = (a_{ij})_{i,j=1}^n \in \mathbb{M}_n$ is said to be:

(i) diagonal if $a_{ij} = 0$ when $i \neq j$.

(ii) invertible if there exists an matrix B of order $n \times n$ such that $AB = I_n$. In this situation A has a unique inverse matrix $A^{-1} \in \mathbb{M}_n$ such that $A^{-1}A = AA^{-1} = I_n$.

(iii) normal if $AA^* = A^*A$.

(iv) unitary if $AA^* = A^*A = I_n$.

(v) Hermitian if $A = A^*$.

(vi) positive semi-definite if $\langle Ax, x \rangle \geq 0$ for all $x \in \mathbb{C}^n$.

(vii) positive definite if $\langle Ax, x \rangle > 0$ for all $x \in \mathbb{C}^n \backslash \{0\}$.

Definition 1.1.2 (Lowner's Order [86])**.** Let A and B be two Hermitian matrices of same order n. We say that $A \geq B$ if and only if $A - B$ is a positive semi-definite matrix.

Let $A \in \mathbb{M}_n$, we denote the eigenvalues of A by $\lambda_j(A)$, for $j = 1, 2, ..., n$. For a matrix $A \in \mathbb{M}_n$, the notation $\lambda(A) \equiv (\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A))$ means that $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$.

The *absolute value* of matrix $A \in \mathbb{M}_n$ is the square root of matrix A^*A and denoted by

$$|A| = (A^*A)^{\frac{1}{2}}.$$

We call the eigenvalues of $|A|$ by the *singular value* of A and denote as $s_j(A)$, for $j = 1, 2, ..., n$. For a matrix $A \in \mathbb{M}_n$, the notation $s(A) \equiv (s_1(A), s_2(A), \ldots, s_n(A))$ means that $s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A)$.

The *trace* of a matrix $A = (a_{ij}) \in \mathbb{M}_n$, denoted by $\mathrm{Tr}(A)$, is the sum of all diagonal entries, or, we often use the sum of all eigenvalues $\lambda_i(A)$ of A, i.e.,

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i(A)$$

Related to the trace of the matrix, we recall the Araki-Lieb-Thirring trace inequality [18] used consistently throughout the thesis.

Theorem 1.1.1. *Let A and B be two positive semi-definite matrices, and let $q > 0$, we have*

$$\mathrm{Tr}\left[\left(B^{\frac{r}{2}}A^rB^{\frac{r}{2}}\right)^{\frac{q}{r}}\right] \leq \mathrm{Tr}\left[\left(B^{\frac{1}{2}}AB^{\frac{1}{2}}\right)^{q}\right], \ \textit{if } r \in (0, 1],$$

and

$$\mathrm{Tr}\left[\left(B^{\frac{r}{2}}A^rB^{\frac{r}{2}}\right)^{\frac{q}{r}}\right] \geq \mathrm{Tr}\left[\left(B^{\frac{1}{2}}AB^{\frac{1}{2}}\right)^{q}\right], \ \textit{if } r \geq 1.$$

The *determinant* of A is denoted and defined by

$$\det(A) = \sum_{\rho \in \mathbb{S}_n} \left(\mathrm{sgn}(\rho) \prod_{i=1}^{n} a_{i\rho_i}\right) = \prod_{j=1}^{n} \lambda_j.$$

where $\mathbb{S}_n$ is the set of all permutations ρ of the set $\mathbb{S} = \{1, 2, \ldots, n\}$.

A function $\|\cdot\| : \mathbb{M}_n \to \mathbb{R}$ is said to be a matrix norm if for all $A, B \in \mathbb{M}_n$ and $\forall \alpha \in \mathbb{C}$ we have:

(i) $\|A\| \geq 0$.

(ii) $\|A\| = 0$ if and only if $A = 0$.

(iii) $\|\alpha A\| = |\alpha| \cdot \|A\|$.

(iv) $\|A + B\| \leq \|A\| + \|B\|$.

In addition, a matrix norm is said to be sub-multiplicative matrix norm if

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

A matrix norm is said to be a *unitarily invariant norm* if for every $A \in \mathbb{M}_n$, we have $\|UAV\| = \|A\|$ for all $U, V \in \mathbb{U}_n$ unitary matrices. It is denoted as $|||\cdot|||$.

These are some important norms in $\mathbb{M}_n$.

The operator norm of A, defined by

$$|||A|||_{op} = \sqrt{\lambda_1\left(A^*A\right)} = s_1(A).$$

The Ky Fan k-norm is the sum of all singular values, i.e.,

$$\|A\|_k = \sum_{i=1}^{k} s_i(A).$$

The Schatten p-norm is defined as

$$\|A\|_p = \left(\sum_{i=1}^{n} s_i^p(A)\right)^{1/p}.$$

When $p = 2$, we have the Frobenius norm or sometimes called the Hilbert-Schmidt norm :

$$\|A\|_2 = \left(\operatorname{Tr}|A|^2\right)^{1/2} = \left(\sum_{j=1}^{n} s_j^2(A)\right)^{1/2}.$$

Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ be in $\mathbb{R}^n$. Let $x^{\downarrow} = \left(x_{[1]},\, x_{[2]}, \ldots, x_{[n]}\right)$ denote a rearrangement of the components of x such that $x_{[1]} \geqslant x_{[2]} \geqslant \ldots \geqslant x_{[n]}$. We say that x is majorized by y, denoted by $x \prec y$, if

$$\sum_{i=1}^{k} x_{[i]} \leqslant \sum_{i=1}^{k} y_{[i]}, \quad k = 1, 2, \ldots, n-1, \quad \text{and} \quad \sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}.$$

We say that x is weakly majorized by y if $\sum_{i=1}^{k} x_{[i]} \leqslant \sum_{i=1}^{k} y_{[i]}, k = 1, 2, \ldots, n$, denoted by $x \prec_w y$. If $x > 0$ (i.e., $x_i > 0$ for $i = 1, \ldots, n$) and $y > 0$, we say that x is log-majorized by y, denoted by $x \prec_{\log} y$, if

$$\prod_{i=1}^{k} x_{[i]} \leqslant \prod_{i=1}^{k} y_{[i]}, \quad k = 1, 2, \ldots, n-1, \quad \text{and} \quad \prod_{i=1}^{n} x_{[i]} = \prod_{i=1}^{n} y_{[i]}.$$

In other words, $x \prec_{\log} y$ if and only if $\log x \prec \log y$.

1.2 Matrix function and matrix mean

Now let us recall the spectral theorem which is one of the most important tools in functional analysis and matrix theory.

Theorem 1.2.1 (Spectral decomposition, [9])**.** *Let $\lambda_1 > \lambda_2 \ldots > \lambda_k$ be eigenvalues of a Hermitian matrix A. Then*

$$A = \sum_{j=1}^{k} \lambda_j P_j,$$

where P_j is the orthogonal projection onto the subspace spanned by the eigenvectors associated to the eigenvalue λ_j.

For a real-valued function f defined on some interval $K \subset \mathbb{R}$ and for a self-adjoint matrix $A \in \mathbb{M}_n$ with spectrum in K, the matrix $f(A)$ is defined by means of the functional calculus, i.e.,

$$A = \sum_{j=1}^{k} \lambda_j P_j \quad \Longrightarrow \quad f(A) := \sum_{j=1}^{k} f(\lambda_j) P_j.$$

In another words, if $A = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^*$ is a spectral decomposition of A (where U is some unitary), then

$$f(A) := U \operatorname{diag}(f(\lambda_1), \cdots, f(\lambda_n)) U^*.$$

We are now at the stage where we will discuss matrix/operator functions. Loewner was the first to study operator monotone functions in his seminal papers [63] in 1930. In the same time, Kraus investigated the notion operator convex function [55].

Definition 1.2.1 ([63])**.** A continuous function f defined on an interval $K (K \subset \mathbb{R})$ is said to be *operator monotone of order* n on K if for two Hermitian matrices A and B in $\mathbb{M}_n$ with spectras in K, one has

$$A \leq B \quad \text{implies} \quad f(A) \leq f(B).$$

If f is operator monotone of any orders then f is called *operator monotone.*

Theorem 1.2.2 (Lowner-Heinz's Inequality, [86])**.** *The function $f(t) = t^r$ is operator monotone on $[0, \infty)$ for $0 \leq r \leq 1$. More specifically, for two positive semi-definite matrices such that $A \leq B$. Then*

$$A^r \leq B^r, \quad 0 \leq r \leq 1.$$

Definition 1.2.2 ([55])**.** A continuous function f defined on an interval $K (K \subset \mathbb{R})$ is said to be *operator convex of order* n on K if for any Hermitian matrices A and B in $\mathbb{M}_n$ with spectra in K and for all real numbers $0 \leq \lambda \leq 1$,

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

If f is operator convex of any order n then f is called *operator convex.* If $-f$ is operator convex then we call f is operator concave.

Theorem 1.2.3 ([10])**.** *Function $f(t) = t^r$ in $[0, \infty)$ is operator convex when $r \in [-1, 0] \cup [1, 2]$. More specifically, for any positive semi-definite matrices A, B and for any $\lambda \in [0, 1]$,*

$$(\lambda A + (1-\lambda)B)^r \leq \lambda A^r + (1-\lambda) B^r.$$

Another important example is the function $f(t) = \log t$, which is operator monotone on $(0, \infty)$ and the function $g(t) = t \log t$ is operator convex. The relations between operator monotone and operator convex via the theorem below.

Theorem 1.2.4 ([9])**.** *Let f be a (continuous) real function on the interval $[0, \alpha)$. Then the following two conditions are equivalent:*

(i) f is operator convex and $f(0) \leq 0$.

(ii) The function $g(t) = \dfrac{f(t)}{t}$ is operator monotone on $(0, \alpha)$.

Definition 1.2.3 ([10])**.** Let $f(A, B)$ be a real valued function of two matrix variables. Then, f is called *jointly concave*, if for all $0 \leq \alpha \leq 1$,

$$f(\alpha A_1 + (1-\alpha)A_2, \alpha B_1 + (1-\alpha)B_2) \geq \alpha f(A_1, B_1) + (1-\alpha) f(A_2, B_2)$$

for all A_1, A_2, B_1, B_2. If $-f$ is jointly concave, we say f is *jointly convex.*

We will review very quickly some basic concepts of the Fretchet differential calculus, with special emphasis on matrix analysis. Let X, Y be real Banach spaces, and let $\mathcal{L}(X, Y)$ be the space of bounded linear operators from X to Y. Let U be an open subset of X. A continuous map f from U to Y is said to be differentiable at a point u of U if there exists $T \in \mathcal{L}(X, Y)$ such that

$$\lim_{v \to 0} \frac{\|f(u+v) - f(u) - Tv\|}{\|v\|} = 0.$$

It is clear that if such a T exists, it is unique. If f is differentiable at u, the operator T above is called the derivative of f at u. We will use for it the notation $Df(u)$, of $\partial f(u)$. This is sometimes called the Frchet derivative. If f is differentiable at every point of U, we say that it is differentiable on U. One can see that, if f is differentiable at u, then for every $v \in X$,

$$Df(u)(v) = \left.\frac{d}{dt}\right|_{t=0} f(u+tv).$$

This is also called the directional derivative of f at u in the direction v. If f_1, f_2 are two differentiable maps, then $f_1 + f_2$ is differentiable and

$$D\left(f_1 + f_2\right)(u) = Df_1(u) + Df_2(u).$$

The composite of two differentiable maps f and g is differentiable and we have the chain rule

$$D(g \circ f)(u) = Dg(f(u)) \cdot Df(u).$$

One important rule of differentiation for real functions is the product rule: $(fg)' = f'g + gf'$. If f and g are two maps with values in a Banach space, their product is not defined - unless the range is an algebra as well. Still, a general product rule can be established. Let f, g be two differentiable maps from X into Y_1, Y_2, respectively. Let B be a continuous bilinear map from $Y_1 \times Y_2$ into Z. Let φ be the map from X to Z defined as $\varphi(x) = B(f(x), g(x))$. Then for all u, v in X

$$D\varphi(u)(v) = B(Df(u)(v), g(u)) + B(f(u), Dg(u)(v)).$$

This is the product rule for differentiation. A special case of this arises when $Y_1 = Y_2 = \mathcal{L}(Y)$, the algebra of bounded operators in a Banach space Y. Now $\varphi(x) = f(x)g(x)$ is the usual product of two operators. The product rule then is

$$D\varphi(u)(v) = [Df(u)(v)] \cdot g(u) + f(u) \cdot [Dg(u)(v)]$$

Higher order Frchet derivatives can be identified with multilinear maps. Let f be a differentiable map from X to Y. At each point u, the derivative $Df(u)$ is an element of the Banach space $\mathcal{L}(X, Y)$. Thus we have a map Df from X into $\mathcal{L}(X, Y)$, defined as $Df : u \to Df(u)$. If this map is differentiable at a point u, we say that f is twice differentiable at u. The derivative of the map Df at the point u is called the second derivative of f at u. It is denoted as $D^2 f(u)$. This is an element of the space $\mathcal{L}(X, \mathcal{L}(X, Y))$. Let $\mathcal{L}_2(X, Y)$ be the space of bounded bilinear maps from $X \times X$ into Y. The elements of this space are maps f from $X \times X$ into Y that are linear in both variables, and for whom there exists a constant c such that

$$\|f(x_1, x_2)\| \leq c \|x_1\| \|x_2\|$$

for all $x_1, x_2 \in X$. The infimum of all such c is called $\|f\|$. This is a norm on the space $\mathcal{L}_2(X, Y)$, and the space is a Banach space with this norm. If φ is an element of $\mathcal{L}(X, \mathcal{L}(X, Y))$, let

$$\tilde{\varphi}(x_1, x_2) = [\varphi(x_1)](x_2) \text{ for } x_1, x_2 \in X.$$

Then $\tilde{\varphi} \in \mathcal{L}_2(X, Y)$. It is easy to see that the map $\varphi \to \tilde{\varphi}$ is an isometric isomorphism. Thus the second derivative of a twice differentiable map f from X to Y can be thought of as a bilinear map from $X \times X$ to Y. It is easy to see that this map is symmetric in the two variables; i.e.,

$$D^2 f(u)(v_1, v_2) = D^2 f(u)(v_2, v_1)$$

for all u, v_1, v_2. Derivatives of higher order can be defined by repeating the above procedure. The p th derivative of a map f from X to Y can be identified with a p-linear map from the space $X \times X \times \cdots \times X$ (p copies) into Y. A convenient method of calculating the p th derivative of f is provided by the formula

$$D^p f(u)(v_1, \ldots, v_p) = \left.\frac{\partial^p}{\partial t_1 \cdots \partial t_p}\right|_{t_1 = \cdots = t_p = 0} f(u + t_1 v_1 + \cdots + t_p v_p).$$

In connections with electrical engineering, Anderson and Duffin [3] defined the *parallel sum* of two positive definite matrices A and B by

$$A : B = \left(A^{-1} + B^{-1}\right)^{-1}.$$

The harmonic mean is $2(A:B)$ which is the dual of the *arithmetic mean* $A\nabla B = \frac{A+B}{2}$. In this period time, Pusz and Woronowicz [69] introduced the *geometric mean* as

$$A\sharp B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}.$$

They also proved that the geometric mean is the unique positive solution of the Riccati equation

$$XA^{-1}X = B.$$

In 2005, Moakher [65] conducted a study, and then in 2006, Bhatia and Holbrook [14] investigated the structure of the Riemannian manifold $\mathbb{H}_n^+$. They showed that the curve

$$\gamma(t) = A\sharp_t B = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2} \quad (t \in [0,1])$$

is the unique geodesic joining A and B, and called t-geometric mean or weighted geometric mean. The weighted harmonic and the weighted arithmetic means are defined by

$$A!_t B = \left(tA^{-1} + (1-t)B^{-1}\right)^{-1},$$

and

$$A\nabla_t B = tA + (1-t)B.$$

The well-known inequality related to these quantities is the harmonic, geometric, and arithmetic means inequality [47, 60] , that is,

$$A!_t B \leq A\sharp_t B \leq A\nabla_t B.$$

These three means are Kubo-Ando means. Let's collect the main content of the Kubo-Ando means theory in the general case [54]. For $x > 0$ and $t \geq 0$, the function $\phi(x,t) = \frac{x(1+t)}{x+t}$ is bounded and continuous on the extended half-line $[0,\infty]$. The Lowner theory ([9, 45]) on operator-monotone functions states that the map $m \mapsto f$, defined by

$$f(x) = \int_{[0,\infty]} \phi(x,t)dm(t) \text{ for } x > 0,$$

establishes an affine isomorphism from the class of positive Radon measures on $[0,\infty]$ onto the class of operator-monotone functions. In the representation abvove, $f(0) = \inf_x f(x) = m(\{0\})$ and $\inf_x f(x)/x = m(\{\infty\})$.

Theorem 1.2.5. *[Kubo-Ando] For each operator connection σ, there exists a unique operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}^+$, satisfying*

$$f(t)I_n = I_n\sigma(tI_n), t > 0,$$

and for $A, B > 0$ the formula

$$A\sigma B = A^{\frac{1}{2}}f(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})A^{\frac{1}{2}}$$

holds, with the right hand side defined via functional calculus, and extended to $A, B \geq 0$ as follows

$$A\sigma B = \lim_{\epsilon\to 0}(A+\epsilon I_n)\sigma(B+\epsilon I_n).$$

We call f the representing function of σ.

The next theorem follows from the integral representation of matrix monotone functions and from the previous theorem.

Theorem 1.2.6. *The map, $m \mapsto \sigma$, defined by*

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

where

$$a = m(\{0\}) \text{ and } b = m(\{\infty\}),$$

establishes an affine isomorphism from the class of positive Radon measures on $[0, \infty]$ onto the class of connections.

If P and Q are two projections, then the explicit formulation for $P\sigma Q$ is simpler.

Theorem 1.2.7. *If σ is a mean, then for every pair of projections P and Q*

$$P\sigma Q = a(P - P \wedge Q) + b(Q - P \wedge Q) + P \wedge Q,$$

where

$$a = 1\sigma 0 \quad \text{and} \quad b = \lim_{x\to\infty} (1\sigma x)/x.$$

An immediate consequence of the above theorem is the following relation for projections P and Q

$$P!Q = P \wedge Q \quad \text{and} \quad P\#Q = P \wedge Q.$$

Let f be the representing function of σ. Since $xf(x^{-1})$ is the representing function of the transpose σ', then σ is symmetric if and only if $f(x) = xf(x^{-1})$. The next theorem gives the representation for a symmetric connection.

Theorem 1.2.8. *The map, $n \mapsto \sigma$, defined by*

$$A\sigma B = \frac{c}{2}(A + B) + \int_{(0,1]} \frac{1+t}{2t}\{(tA) : B + A : (tB)\} dn(t)$$

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

Chapter 2

Weighted Hellinger distance

In recent years, many researchers have paid attention to different distance functions on the set $\mathbb{P}_n$ of positive definite matrices. Along with the traditional Riemannian metric $d_R(A,B) = \left(\sum_{i=1}^{n} \log^2 \lambda_i(A^{-1}B)\right)^{1/2}$ (where $\lambda_i(A^{-1}B)$ are eigenvalues of the matrix $A^{-1/2}BA^{-1/2}$), there are other important functions. Two of them are the Bures-Wasserstein distance, which are adapted from the theory of optimal transport [13]:

$$d_b(A,B) = \left(\operatorname{Tr}(A+B) - 2\operatorname{Tr}((A^{1/2}BA^{1/2})^{1/2})\right)^{1/2},$$

and the Hellinger metric or Bhattacharya metric in quantum information [75]:

$$d_h(A,B) = \left(\operatorname{Tr}(A+B) - 2\operatorname{Tr}(A^{1/2}B^{1/2})\right)^{1/2}.$$

Notice that the metric d_h is the same as the Euclidean distance between $A^{1/2}$ and $B^{1/2}$, i.e., $\|A^{1/2} - B^{1/2}\|_F$.

Recently, Ha [43] introduced the Alpha Procrustes distance as follows: For $\alpha > 0$ and for two positive semi-definite matrices A and B,

$$d_{b,\alpha} = \frac{1}{\alpha} d_b(A^{2\alpha}, B^{2\alpha}).$$

He showed that the Alpha Procrustes distances are the Riemannian distances corresponding to a family of Riemannian metrics on the manifold of positive definite matrices, which encompass both the Log-Euclidean and Wasserstein Riemannian metrics. Since the Alpha Procrustes distances are defined based on the Bures-Wasserstein distance, we also call them the *weighted Bures-Wasserstein distances*. In that flow, in this chapter we can define the *weighted Hellinger metric* for two positive semi-definite matrices as follows:

$$d_{h,\alpha}(A,B) = \frac{1}{\alpha} d_h(A^{2\alpha}, B^{2\alpha}),$$

then investigate its properties within this framework.

The results of this chapter are taken from [32].

2.1 Weighted Hellinger distance

Definition 2.1.1. For two positive semi-definite matrices A and B and for $\alpha > 0$, the weighted Hellinger distance between A and B is defined as

$$d_{h,\alpha}(A,B) = \frac{1}{\alpha} d_h(A^{2\alpha}, B^{2\alpha}) = \frac{1}{\alpha} (\operatorname{Tr}(A^{2\alpha} + B^{2\alpha}) - 2\operatorname{Tr}(A^{\alpha} B^{\alpha}))^{\frac{1}{2}}.$$

Proposition 2.1.1. For two positive semi-definite matrices A and B,

$$\lim_{\alpha \to 0} d_{h,\alpha}^2(A,B) = ||\log(A) - \log(B)||_F^2.$$

Proposition 2.1.2. For two positive semi-definite matrices A and B,

$$d_{b,\alpha}(A,B) \leq d_{h,\alpha}(A,B) \leq \sqrt{2} d_{b,\alpha}(A,B).$$

2.2 In-betweenness property

In 2016, Audenaert introduced the in-betweenness property of matrix means [5]. We say that a matrix mean σ satisfies the *in-betweenness property* with respect to the metric d if for any pair of positive definite operators A and B,

$$d(A, A\sigma B) \leq d(A,B).$$

In [34], the authors introduced and studied the in-sphere property of matrix means. Dinh, Franco and Dumitru also published several papers [26, 28] on geometric properties of the matrix power mean $\mu_p(t; A, B) := (tA^p + (1-t)B^p)^{1/p}$ with respect to different distance functions. They also considered the case of the matrix power mean in the sense of Kubo-Ando [54] which is defined as

$$P_p(t, A, B) = A^{1/2} \left(tI + (1-t)(A^{-1/2} B A^{-1/2})^p \right)^{1/p} A^{1/2}.$$

In this section, we focus our study on the in-betweenness properties of the matrix power means with respect to the weighted Bures-Wasserstein and weighted Hellinger distances. As a consequence of the equivalence, using the operator convexity and concavity of the power functions, we show that the matrix power mean satisfies the in-betweenness property with respect to $d_{h,\alpha}$ (Theorem 2.2.1) and $d_{b,\alpha}$ (Theorem 2.2.2). We also show that among symmetric means, the arithmetic mean is the only one that satisfies the in-betweenness property in the weighted Bures-Wasserstein and weighted Hellinger distances.

Theorem 2.2.1. *Let $0 < p/2 \leq \alpha \leq p$ and $0 \leq t \leq 1$. Then*

$$d_{h,\alpha}(A, \mu_p(t; A, B)) \leq d_{h,\alpha}(A,B),$$

for all $A, B \in \mathbb{H}_n^+$.

Theorem 2.2.2. *Let $0 < p/2 \leq \alpha \leq p$ and $1/2 \leq t \leq 1$. Then,*

$$d_{b,\alpha}(A, \mu_p(t; A, B)) \leq d_{b,\alpha}(A, B),$$

for all $A, B \in \mathbb{H}_n^+$.

In [28, Theorem 2] the authors proved that the matrix Kubo-Ando power mean $P_p(t, A, B)$ satisfies the in-betweenness property which follows from the fact that the function $g(t) = \mathrm{Tr}(A^{1/2} P_p(t; A, B)^{1/2})$ is concave. Note that $P_t(A, B) \neq P_t(B, A)$, i.e., P_t is not symmetric. However, for the symmetric means we may have the following result whose proof is adapted from [22].

Theorem 2.2.3. *Let σ be a symmetric mean and assume that one of the following inequalities holds for any pair of positive definite matrices A and B:*

$$d_{h,\alpha}(A, A\sigma B) \leq d_{h,\alpha}(A, B) \tag{2.1}$$

or

$$d_{b,\alpha}(A, A\sigma B) \leq d_{b,\alpha}(A, B). \tag{2.2}$$

Then σ is the arithmetic mean.

In this chapter, we introduce a new distance called the weighted Hellinger distance and investigate its properties. This distance is constructed based on Minh's approach when he constructed the weighted Bures distance. The weighted Bures distance is an extended version with one parameter of the Bures distance. In the next chapter, we introduce a new quantum divergence called the α-z-Bures Wasserstein divergence, which is considered as an extension with two parameters of the Bures distance.

Chapter 3

The α-z-Bures Wasserstein divergence

It is well-known that in the Riemannian manifold of positive definite matrices, the weighted geometric mean $A\sharp_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$ is the unique geodesic joining A and B, where $A, B \in \mathbb{P}_n$. For $t = 1/2$, $A\sharp_{1/2}B$ is called the geometric mean of A and B. It is obvious that $A\sharp_{1/2}B$ is a matrix generalization of the geometric mean $\sqrt{ab}$ of positive numbers a and b. Let $A_1, A_2, \cdots, A_m$ be positive definite matrices. In 2004, Moakher [65] and then Bhatia and Holbrook [14] studied the following least squares problem

$$\min_{X>0} \sum_{i=1}^{m} \delta_2^2(X, A_i), \tag{3.1}$$

where $\delta_2(A, B) = ||\log(A^{-1}B)||_2$ is the Riemannian distance between A and B. They showed that (3.1) has a unique solution which is called the Karcher mean of $A_1, A_2, \cdots, A_m$. In literature, this mean has different names such as: Fréchet mean, Cartan mean, Riemannian center of mass. It turns out that the solution of (3.1) is the unique positive definite solution of the Karcher equation

$$\sum_{i=1}^{m} \log(X^{1/2}A_iX^{1/2}) = 0. \tag{3.2}$$

In [60], Lim and Palfia showed that the solution of (3.2) is nothing but the limit of the solution of the following matrix equation as $t \to 0$,

$$X = \sum_{i=1}^{m} \frac{1}{m} X\sharp_t A_i. \tag{3.3}$$

Recently, Franco and Dumitru [38] introduced the so-called Renyi power means of matrices. More precisely, for $0 < \alpha_i \le z_i \le 1$ and for positive definite matrices A_i, B_i, using the approach in [60] developed by Lim and Pálfia, they showed that the following equation

$$X = \sum_{i=1}^{m} \omega_i P_{\alpha_i, z_i}(X, A_i) \tag{3.4}$$

had a unique positive definite solution, where (ω_i) is a probability vector and $P_{\alpha,z}(A,B) = (B^{\frac{1-\alpha}{2z}} A^{\frac{\alpha}{z}} B^{\frac{1-\alpha}{2z}})^z$-the matrix function in the α-z-Renyi relative entropy introduced by Audenaert and Datta [7] in 2015. Notice that if we replace $P_{\alpha_i,z_i}(X,A_i)$ in (3.4) with the weighted geometric mean $X\sharp_t A_i$, the solution of the corresponding matrix equation is the weighted power mean.

Now, notice that if we change the distance function in (3.1), the solution may be different, if exists. Interestingly, in applications people sometimes are interested in distance-like functions that provide distance between two data points. Such functions are not necessarily symmetric; and the triangle inequality does not need to be true. Divergences are such distance-like functions. An important example of divergences is the Bures-Wasserstein metric studied by Bhatia and coauthors [13] as follows:

$$d_b(A,B) = (\mathrm{Tr}((A+B)/2) - \mathrm{Tr}(A^{1/2}BA^{1/2})^{1/2})^{1/2},$$

where $\mathrm{Tr}((A^{1/2}BA^{1/2})^{1/2})$ is the quantum fidelity of two positive definite matrices A and B. They showed that d_b^2 is a quantum divergence and solved the least squares problem with respect to the Bures-Wasserstein divergence. In another paper [14], these authors introduced so called the weighted Bures-Wasserstein distance as

$$d_{b,t}(A,B) = (\mathrm{Tr}((1-t)A + tB) - \mathrm{Tr}(F_t(A,B))^{1/2},$$

where $F_t(A,B) = \mathrm{Tr}(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}})^t$ is the sandwiched quasi-relative entropy [59, 79]. They also solved the least squares problem with respect to this divergence. Mention that $(A^{1/2}BA^{1/2})^{1/2}$ and $(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}})^t$ are matrix generalizations of the geometric mean $\sqrt{ab}$ and the weighted geometric mean $a^{1-t}b^t$ of positive numbers a and b, respectively.

Motivated by works mentioned above, in this chapter, we introduce and study different properties of the α-z-Bures Wasserstein divergence defined as

$$\Phi(A,B) = \mathrm{Tr}((1-\alpha)A + \alpha B) - \mathrm{Tr}(Q_{\alpha,z}(A,B)), \tag{3.5}$$

whenever A and B are positive definite matrices, and $Q_{\alpha,z}(A,B) = P_{\alpha,z}(B,A)$. Note that $Q_{\alpha,z}(A,B)$ is also a parameterized matrix version of the weighted geometric mean $a^{1-\alpha}b^\alpha$.

The results of this chapter are taken from [30, 31, 32, 77].

3.1 The α-z-Bures Wasserstein divergence and the least squares problem

Theorem 3.1.1. *Let $0 \le \alpha \le z \le 1$. Then the quantity*

$$\Phi(X,Y) = \mathrm{Tr}((1-\alpha)X + \alpha Y) - \mathrm{Tr}(Q_{\alpha,z}(X,Y)) \quad (X,Y>0)$$

is a divergence.

We also solve the least square problem with respect to $\Phi(A, B)$ and showed that the solution of this problem is exactly the unique positive definite solution of the matrix equation $\sum_{i=1}^{m} w_i Q_{\alpha,z}(X, A_i) = X$.

Theorem 3.1.2. *For $0 \le \alpha \le z \le 1$, the function*

$$F(X) = \sum_{i=1}^{m} \omega_i \Phi(A_i, X)$$

attains minimum at X_0, where X_0 is the unique positive definite solution of the following matrix equation

$$\sum_{i=1}^{m} w_i Q_{\alpha,z}(X, A_i) = X.$$

In [49], M. Jeong and co-authors investigated this solution and denoted it by $\mathcal{R}_{\alpha,z}(\omega, \mathbb{A})$-called α-z-weighted right mean. Then, we continued studying this quantity and get some new results.

Theorem 3.1.3. *Let $0 \le \alpha \le z \le 1, \alpha \ne 1, z \ne 0$. Let $\mathbb{A} = (A_1, ..., A_m)$ be an m-tuple of positive definite matrices, and $\omega = (w_1, ..., w_m)$ a probability vector. We have*

$$\frac{1+z-\alpha}{1-\alpha} I - \frac{z}{1-\alpha} \sum_{j=1}^{m} w_j A_j^{-\frac{1-\alpha}{z}} \le \mathcal{R}_{\alpha,z}(\omega, \mathbb{A}) \le \Big(\frac{1+z-\alpha}{1-\alpha} I - \frac{z}{1-\alpha} \sum_{j=1}^{m} w_j A_j^{\frac{1-\alpha}{z}} \Big)^{-1}.$$

The second inequality holds when $(1 + z - \alpha)I - z \sum_{j=1}^{m} w_j A_j^{\frac{1-\alpha}{z}}$ is invertible.

Hwang and Kim [48] proved that for any weighted m-mean $\mathcal{G}_m$ between arithmetic mean and geometric mean, the function $\mathcal{G}_m^{\omega} := \mathcal{G}_m(\omega, \cdot) : \mathbb{P}^m \to \mathbb{P}$ is differentiable at $\mathbb{I} = (I, ..., I)$ with

$$D\mathcal{G}_n^{\omega}(\mathbb{I})(X_1, ..., X_m) = \sum_{j=1}^{m} w_j X_j.$$

Notice that the α-z-weighted right mean does not satisfy the above condition. However, we do have a similar result as follows.

Theorem 3.1.4. *Let $\omega = (w_1, ..., w_m)$ be a probability vector and let $\mathcal{R}_{\alpha,z}^{\omega} := \mathcal{R}_{\alpha,z}(\omega, \cdot) : \mathbb{P}_n^m \longrightarrow \mathbb{P}_n$. Then $\mathcal{R}_{\alpha,z}^{\omega}$ is differentiable at $\mathbb{I} = (I, ..., I)$, and*

$$D\mathcal{R}_{\alpha,z}^{\omega}(\mathbb{I})(X_1, ..., X_m) = \sum_{j=1}^{m} w_j X_j.$$

Lastly, we prove that $\mathcal{R}_{\alpha,z}(\omega, \mathbb{A})$ is a multivariate Lie-Trotter mean.

Theorem 3.1.5. *The $\mathcal{R}_{\alpha,z}(\omega, \mathbb{A})$ is the multivariate Lie-Trotter mean, that means, for any probability vector $\omega = (w_1, w_2, ., w_m)$, we have*

$$\lim_{s \to 0} \mathcal{R}_{\alpha,z}(\omega, \gamma_1(s), ..., \gamma_m(s))^{1/s} = \exp \Big(\sum_{j=1}^{m} w_j \gamma_j'(0) \Big),$$

where for $\varepsilon > 0, \gamma_j : (-\varepsilon, \varepsilon) \longrightarrow \mathbb{P}_n$ are differentiable curves with $\gamma_j(0) = I$, for all $j = 1, 2, ..., m$.

3.2 Data processing inequality and in-betweenness property

In this section, we show that this divergence satisfies the data processing inequality (DPI) in quantum information. The data processing inequality is an information-theoretic concept that states that the information content of a signal cannot be increased via a local physical operation. This can be expressed concisely as post-processing cannot increase information, that is, for any completely positive trace preserving map $\mathcal{E}$ and for any positive semi-definite matrices A and B,

$$\Phi(\mathcal{E}(A), \mathcal{E}(B)) \leq \Phi(A, B).$$

Furthermore, we show that the matrix power mean $\mu(t, A, B) = ((1-t)A^p + tB^p)^{1/p}$ satisfies the in-betweenness property with respect to the α-z-Bures Wasserstein divergence.

Theorem 3.2.2. *Let $A, B \in \mathbb{P}_n, 0 < \alpha \leq z \leq 1, 1/2 \leq p \leq 1$ and $\alpha \leq zp$. Then for any positive definite matrices A and B,*

$$\Phi(A, \mu_p) \leq \Phi(A, B).$$

3.3 Quantum fidelity and its parameterized versions

Quantum fidelity is an important quantity in quantum information theory and quantum chaos theory. It is a distance measure between density matrices, which are considered as quantum states. Although it is not a metric, it has many useful properties that can be used to define a metric on the space of density matrices. In the next section, we give some properties for quantum fidelity and its extended version. An important results is we establish some variational principles for the quantum α-z-fidelity

$$f_{\alpha,z}(\rho, \sigma) := \operatorname{Tr}\left(\rho^{\alpha/2z}\sigma^{(1-\alpha)/z}\rho^{\alpha/2z}\right)^z = \operatorname{Tr}\left(\sigma^{(1-\alpha)/2z}\rho^{\alpha/z}\sigma^{(1-\alpha)/2z}\right)^z,$$

where ρ and σ are two postitive definite matrices. That is, it is the extremal value of two matrix functions

$$P(X) = z\operatorname{Tr}\left(\sigma^{\frac{z-\alpha}{2z}}\rho^{\frac{\alpha}{z}}\sigma^{\frac{z-\alpha}{2z}}X\right) - (z-1)\operatorname{Tr}\left(\sigma^{\frac{z-1}{2z}}X\sigma^{\frac{z-1}{2z}}\right)^{\frac{z}{z-1}},$$

and

$$Q(X) = \left(\operatorname{Tr}(\sigma^{\frac{z-\alpha}{2z}}\rho^{\frac{\alpha}{z}}\sigma^{\frac{z-\alpha}{2z}}X)\right)^z . \left(\operatorname{Tr}(\sigma^{\frac{z-1}{2z}}X\sigma^{\frac{z-1}{2x}})^{\frac{z}{z-1}}\right)^{1-z}.$$

Theorem 3.3.4. *Let ρ, σ be positive definite matrices and $0 < \alpha < z < 1$. We have*

(i) $f_{\alpha,z}(\rho, \sigma) = \min_{X>0} P(X)$.

(ii) $f_{\alpha,z}(\rho, \sigma) = \min_{X>0} Q(X)$.

Furthermore, the minimum is achieved at $X_0 = \sigma^{\frac{1-z}{2z}}(\sigma^{\frac{1-\alpha}{2z}}\rho^{\frac{\alpha}{z}}\sigma^{\frac{1-\alpha}{2z}})^{z-1}\sigma^{\frac{1-z}{2z}}$.

3.4 The α-z-fidelity between unitary orbits

At the last section of this chapter, we use $f_{\alpha,z}$ to measure the distance between two quantum orbits and prove that the set of these distace is a close interval in $\mathbb{R}^+$.

Theorem 3.4.2. *Let ρ and $\sigma \in \mathcal{D}_n$, the α-z-fidelity $f_{\alpha,z}(\rho,\sigma) = \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2z}}\rho^{\frac{\alpha}{z}}\sigma^{\frac{1-\alpha}{2z}}\right)^z$ between the unitary orbits U_ρ and U_σ satifies*

$$\max_{U \in U(\mathbb{H})} f_{\alpha,z}(\rho, U\sigma U^*) = \sum_{i=1}^{n} \lambda_i^{\downarrow}(\rho)^{\alpha}\lambda_i^{\downarrow}(\sigma)^{1-\alpha},$$

and

$$\min_{U \in U(\mathbb{H})} f_{\alpha,z}(\rho, U\sigma U^*) = \sum_{i=1}^{n} \lambda_i^{\downarrow}(\rho)^{\alpha}\lambda_i^{\uparrow}(\sigma)^{1-\alpha},$$

where $\lambda(\rho) = (\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of ρ and $\lambda^{\downarrow}(\rho)$ (resp. $\lambda^{\uparrow}(\rho)$) is a rearrangement of $\lambda(\rho)$ in decreasing order (resp. increasing order).

Theorem 3.4.3. *For $0 \leq \alpha \leq z \leq 1$,*

$$\{f_{\alpha,z}(\rho, U\sigma U^*) : U \in U(\mathbb{H})\} = \Big[\sum_{i=1}^{n} \lambda_i^{\downarrow}(\rho)^{\alpha}\lambda_i^{\uparrow}(\sigma)^{1-\alpha}, \sum_{i=1}^{n} \lambda_i^{\downarrow}(\rho)^{\alpha}\lambda_i^{\downarrow}(\sigma)^{1-\alpha}\Big].$$

In this chapter, we introduce a new quantum divergence called the α-z-Bures Wasserstein distance, which is an extension with two parameters of the Bures distance. Then we investigate its properties. In particular, we solve the least square problem with respect to this divergence and study its solution. In the next chapter, we introduce a new weighted spectral geometric mean denoted by $\mathcal{F}_t(A,B)$ and study the properties of this quantity. Additionally, we provide some comparisons between $\mathcal{F}_t(A,B)$ and $A \diamond_t B$, which is the solution to the least square problem with respect to the Bures Wasserstein distance.

Chapter 4

A new weighted spectral geometric mean

It is well-known [10] that the geometric mean $A\sharp B$ is the midpoint of the geodesic

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

joining A and B under the Riemannian metric $\delta_R(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm [11].

The spectral geometric mean of $A, B \in \mathbb{P}_n$ was introduced by Fiedler and Pták in 1997 [37], and one of its formulations is

$$A\natural B := (A^{-1}\sharp B)^{1/2} A (A^{-1}\sharp B)^{1/2}. \tag{4.1}$$

It is called *the spectral geometric mean* because $(A\natural B)^2$ is similar to AB and that the eigenvalues of their spectral mean are the positive square roots of the corresponding eigenvalues of AB [37, Theorem 3.2].

Kim and Lee [52] defined the weighted spectral mean:

$$A\natural_t B := \left(A^{-1}\sharp B\right)^t A \left(A^{-1}\sharp B\right)^t, \quad t \in [0,1]. \tag{4.2}$$

It is obvious that $A\natural_t B$ is a curve joining A and B. They studied the relative operator entropy related to the spectral geometric mean and several properties similar to those of the relative entropy of Tsallis operator defined via the matrix geometric mean. Recently, Gan, Liu, and Tam [41] and Gan and Tam [40] studied $A\natural_t B$ and obtained some nice properties.

Note that in (4.2) the geometric mean $A^{-1}\sharp B$ is a main component of the weighted spectral mean $A\natural_t B$ while the middle term is A, independent of t.

Following that sequence of events, in this chapter we define a new weighted mean, called $\mathcal{F}$-mean.

The results of this chapter are taken from [33].

4.1 A new weighted spectral geometric mean and its basic properties

Definition 4.1.1. Let $A, B \in \mathbb{P}_n$. Define

$$\mathcal{F}_t(A,B) := (A^{-1}\sharp_t B)^{1/2} A^{2-2t} (A^{-1}\sharp_t B)^{1/2}, \quad t \in [0,1]. \tag{4.3}$$

It is obvious that $\mathcal{F}_0(A,B) = A$ and $\mathcal{F}_1(A,B) = B$, and hence $\mathcal{F}_t(A,B)$ is a curve joining A and B. For $t = \frac{1}{2}$, $\mathcal{F}_{\frac{1}{2}}(A,B)$ is the spectral geometric mean (4.1). We call $\mathcal{F}_t(A,B)$ weighted $\mathcal{F}$-mean and it is different from (4.2).

From the Riccati equation, it is obvious that $A\sharp X = B$ if and only if $X = BA^{-1}B$. Therefore, $F_t(A,B)$ is the unique positive definite solution X to

$$A^{2(t-1)}\sharp X = (A^{-1}\sharp_t B)^{1/2}.$$

Proposition 4.1.1. Let $A, B \in \mathbb{P}_n$. The following properties hold for all $t \in [0,1]$.

1. $\mathcal{F}_t(A,B) = A^{1-t}B^t$ if A and B commute.
2. $\mathcal{F}_t(aA, bB) = a^{1-t}b^t\mathcal{F}_t(A,B)$ for $a, b > 0$.
3. $U^*\mathcal{F}_t(A,B)U = \mathcal{F}_t(U^*AU, U^*BU)$ for $U \in U(n)$.
4. $\mathcal{F}_t^{-1}(A,B) = \mathcal{F}_t(A^{-1}, B^{-1})$.
5. $\det \mathcal{F}_t(A,B) = (\det A)^{1-t}(\det B)^t$.
6. $2((1-t)A + tB^{-1})^{-1/2} - A^{2(t-1)} \leq \mathcal{F}_t(A,B) \leq [2((1-t)A^{-1} + tB)^{-1/2} - A^{-2(t-1)}]^{-1}$.

4.2 The Lie-Trotter formula and weak log-majorization

Theorem 4.2.1. *Let $A, B \in \mathbb{H}_n$ and $t \in [0,1]$. Then*

$$\lim_{p\to 0} \mathcal{F}_t^{1/p}(e^{pA}, e^{pB}) = e^{(1-t)A+tB}.$$

Theorem 4.2.2. *Let $(\alpha_1, ..., \alpha_{m-1}) \in \mathbb{R}^{m-1}$, and $X_1, X_2, ..., X_m \in \mathbb{H}_n$. The curve*

$$\gamma(t) := \mathcal{F}_{\alpha_{m-1}}\Big(e^{tX_m}, \mathcal{F}_{\alpha_m-2}\Big(e^{tX_{m-1}}, \mathcal{F}_{\alpha_m-3}(...\mathcal{F}_{\alpha_1}(e^{tX_2}, e^{tX_1})...\Big)\Big)$$

is a differentiable curve with $\gamma(0) = I$ and

$$\gamma'(0) = \sum_{k=1}^{m}\prod_{i=k}^{m}\alpha_i\,(1-\alpha_{k-1})\,X_k,$$

where $\alpha_0 = 0$ *and* $\alpha_m = 1$. *In particular, if* $\alpha_k = \dfrac{k}{k+1}$, *for* $k = 1, 2, ..., m-1$ *then* $\gamma'(0) = \dfrac{1}{m}\sum_{k=1}^{m} X_k$.

At the end of this chapter, we compare the weak-log majorization between the $\mathcal{F}$-mean and the Wasserstein mean, which is the solution to the least square problem with respect to the Bures distance or Wasserstein distance.

Theorem 4.2.3. *Let* $A, B \in \mathbb{P}_n$ *and* $t \in [0, 1]$.

(i) If $0 \le t \le \frac{1}{2}$ *then*

$$\mathcal{F}_t(A, B) \prec_{w\log} A \diamond_t B;$$

(ii) If $\frac{1}{2} \le t \le 1$ *then*

$$\mathcal{F}_{1-t}(B, A) \prec_{w\log} A \diamond_t B.$$

Conclusion

This thesis obtained the following main results:

1. We introduce a new Weighted Hellinger distance, denoted as $d_{h,\alpha}(A, B)$, and prove that it acts as an interpolating metric between the Log-Euclidean and Hellinger metrics. Additionally, we establish the equivalence between the weighted Bures-Wasserstein and weighted Hellinger distances. Moreover, we demonstrate that both distances satisfy the in-betweenness property. Moreover, we also show that among symmetric means, the arithmetic mean is the only one that satisfies the in-betweenness property in the weighted Bures-Wasserstein and weighted Hellinger distances.

2. We construct a new quantum divergence called the α-z-Bures-Wasserstein divergence and demonstrate that this divergence satisfies the in-betweenness property and the data processing inequality in quantum information theory. Furthermore, we solve the least squares problem with respect to this divergence and establish that the solution to this problem corresponds exactly to the unique positive solution of the matrix equation

$$\sum_{i=1}^{m} w_i Q_{\alpha,z}(X, A_i) = X,$$

where $Q_{\alpha,z}(A, B) = \left(A^{\frac{1-\alpha}{2z}} B^{\frac{\alpha}{z}} A^{\frac{1-\alpha}{2z}}\right)^z$ and $0 < \alpha \leq z \leq 1$. Afterwards, we proceed to study the properties of the solution to this problem and achieve several significant results. In addition, we provide an inequality for quantum fidelity and its parameterized versions. Then, we utilize α-z-fidelity to measure the distance between two quantum orbits.

3. We introduce a new weighted geometric mean, called the $\mathcal{F}$-mean. We establish some properties for the $\mathcal{F}$-mean and prove that it satisfies the Lie-Trotter formula, Furthermore, we provide a comparison in weak-log majorization between the $\mathcal{F}$-mean and the Wasserstein mean.

Further investigation

In the future, we intend to continue the investigation in the following directions:

- Construct some new distance function based on non-Kubo-Ando means.
- Construct a new distance function between two matrices with different dimensions.
- For $X, Y > 0$ and $0 < t < 1$, verify whether the two quantities

  $$\Phi_1(X,Y) = \mathrm{Tr}((1-t)X + tY) - \mathrm{Tr}\left(X \natural_t Y\right)$$

  and

  $$\Phi_2(X,Y) = \mathrm{Tr}((1-t)X + tY) - \mathrm{Tr}\left(F_t(X,Y)\right)$$

  are divergences and simultaneously solve related problems.
- Quantity $\mathcal{F}_t(X,Y)$ is new; therefore, we need to establish new properties for this quantity while also comparing it with the previously known means.

List of Author's related to the thesis

1. Vuong T.D., Vo B.K (2020), "An inequality for quantum fidelity", *Quy Nhon Univ. J. Sci.*, 4 (3).

2. Dinh T.H., Le C.T., Vo B.K, Vuong T.D. (2021), "Weighted Hellinger distance and in betweenness property", *Math. Ine. Appls.,* 24, 157-165.

3. Dinh T.H., Le C.T., Vo B.K., Vuong T.D. (2021), "The α-z-Bures Wasserstein divergence", *Linear Algebra Appl.,* 624, 267-280.

4. Dinh T.H., Le C.T., Vuong T.D., α-z-fidelity and α-z-weighted right mean, *Submitted*.

5. Dinh T.H., Tam T.Y., Vuong T.D, On new weighted spectral geometric mean, *Submitted*.