An Introductory Course on

Non-commutative Information Geometry

Jan Naudts

Universiteit Antwerpen

ORCID: 0000-0002-4646-1190

2020 Mathematics Subject Classification: 62B12, 53B12

Preface

Information Geometry applies differential geometry to study statistical models. It considers manifolds of probability measures. If the model is parameterized then the natural metric on the tangent spaces is given by the Fisher information matrix. The work of Amari [27, 45] introduced the notion of a dually-flat geometry. It provides the structure for the present course. Recent books on Information geometry are (Amari, 2016) [64] and (Ay et al, 2017)[65].

Quantum Statistics finds its origin in Physics where Thermodynamics, a discipline developed in the nineteenth century, Statistical Physics as formulated in the book of (Gibbs, 1901) [1] and Quantum Mechanics (Heisenberg, 1925) [2] merge into Quantum Statistical Physics.

Quantum Statistics is a non-commutative extension of Statistics. It is therefore obvious to generalize Information Geometry to this non-commutative context. This is the subject of the present course. Early efforts in this direction include the works of Hasegawa [33, 40, 41], Petz [34, 41].

The course is introductory. It is not a review of the subject. Many interesting topics have been omitted. Many interesting papers are not cited. The topics that are discussed can often be treated in a more general way. A coherent picture of the subject is more important than a complete picture. I want to apologize when you feel I am giving too many details. My experience is that students may get lost on places where experts want to speed up.

Throughout the course the dimension of the involved Hilbert spaces is assumed to be finite. The commutative equivalent is the assumption that the event space is a finite set. This assumption reduces the technicality of the subject in a considerable manner. To be honest, another reason is that not all of the obstacles being raised by the infinite-dimensional case have been solved. At the end of the last chapter some of these difficulties are shortly mentioned.

An effort is done to make the text self-contained. Never the less it is assumed that the audience masters a substantial amount of mathematics. A good knowledge of

Functional Analysis, in particular the functional analysis of matrix algebras, is required. More evolved items can be found for instance in the textbook of Rudin [29]. At some places the books of Kato [9] on linear operators and of Bratelli-Robinson [22] on operator algebras are referred to. A basic knowledge of Differential Geometry and an introductory knowledge on Information Geometry are required as well.

Chapter 1 contains a short introduction on quantum statistics with emphasis on the C^*-algebraic approach. The need for a non-commutative extension of statistics is explained. As an example a historical experiment in the realm of quantum physics is reported. The proof is given that the experiment cannot be modeled using standard statistics. The example comes back in the final chapter where the quantum modeling is presented.

Chapter 2 introduces the Bures metric. There exists an extensive literature on the topic because of its use in Quantum Information theory. However, in the latter domain the original introduction by Bures [13] is largely neglected. Instead, the reformulation by Uhlmann [19] is used. The Bures version is followed here because it uses the C^*-algebraic context and because it is close in spirit to the Wasserstein metric of commutative statistics.

The m-connection is introduced. Its geodesics are convex combinations. Derivatives of geodesics define tangent vectors. The metric on the tangent planes is obtained by taking derivatives of the square of the Bures distance. The calculations are complicated by non-commuativity and require the notion of a symmetric logarithmic derivative.

Chapters 3 to 5 deal with the Bogoliubov metric. It is obtained by taking derivatives of Umegaki's relative entropy. The notion of an exponential arc is introduced. They are the geodesics of the e-connection. In the last section of Chapter 3 an alternative definition of an exponential arc is defined. The calculations make use of Tomita-Takesaki theory. An example shows that the two definitions are distinct.

Chapter 4 discusses the dually-flat geometry and its Legendre structure.

The topic of Chapter 5 is the notion of a quantum exponential family of states. They define a submanifold of the manifold of faithful states. It is shown that the hoistorical experiment discussed in Chapter 1 can be modeled by a quantum exponential family. The last section gives a short discussion of the difficulties encountered when one tries to generalize the present theory to the context of infinite-dimensional Hilbert spaces.

Contents

Chapter 1

Quantum Probability

1.1 Introduction

Quantum Probability and Quantum Statistics are being used in many areas beyond Quantum Mechanics. Some of these areas carry names such as Quantum Cognition, Quantum Social Science, Quantum Biology. Early papers are [36, 42, 53, 56, 60, 61, 63]. A recent mathematical paper applies the quantum formalism to Colorimetry [68].

Quantum statistics differs in a number of aspects from conventional statistics.

The most fundamental assumption of probability theory is that there exists a sample space X. The measurable subsets of X are called events. In some applications this assumption is violated by empirical data to such an extent that the concept of a sample space must be abandoned. This is the case in Quantum Mechanics.

It is quite common not to emphasize that random variables are functions of events. Rather one considers them as the basic quantities of probability theory. The generalization then consists of allowing that the product of two random variables is non-commutative. This can be done in two ways. Either one replaces the usual product of two functions by some non-commutative alternative. Or, one abandons the notion of a function and accepts that random variables are abstract quantities that receive their meaning from the fact that one can assign values to them by observation or experiment.

Section 1.3 focuses on the concept of conditional probability. The observation or measurement of two statistically independent quantities can introduce dependency. This is known as *breaking of statistical independence*. The obvious explanation is that the measurement of one quantity influences the outcome of subsequent measurements. In statistical terms this means that what is measured is a conditional probability given the outcome of the first measurement.

Quantum conditional expectations have been introduced in the context of Quantum Information Theory. See for instance [57]. However, the notion of conditional probability used when accumulating experimental data is the conventional one. Quantum conditional expectations are used in theoretical modeling only.

Theoretical arguments prove that only non-classical (i.e. quantum or non-commutative) models can explain quantum experiments. To illustrate this point one historical experiment featuring quantum entanglement is discussed in Section 1.4.

The next Section introduces some elements of quantum probability. The C^*-algebraic approach is highlighted. It is argued that classical probability theory is a special case of the more general C^*-algebraic formulation.

1.2 Quantum expectations

1.2.1 The algebra of n-by-n matrices

The product of two n-by-n matrices T and V is in general non-commutative. The space $\mathcal{B}(\mathbb{C}^n)$ of all n-by-n matrices with complex entries forms a Banach algebra for the norm defined by

$$||T|| = \sup\{|Tx| : x \in \mathbb{C}^n, |x| = 1\}. \tag{1.1}$$

Here, the length of a vector x in $\mathbb{C}^n$ is defined by

$$|x| = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

An important property of the supremum norm is that

$$||TV|| \leq ||T||\,||V|| \quad \text{for all } T, V.$$

The hermitian conjugate of a matrix T is denoted T^* and is called the *adjoint* of T. An operator T is self-adjoint if $T^* = T$. An operator T is *positive*, notation $T \geq 0$, if it is self-adjoint and all its eigenvalues are non-negative.

A special property of the norm (1.1), in relation to the adjoint operation, is that

$$||T||^2 = ||T^*T|| \quad \text{for any } T \text{ in } \mathcal{B}(\mathbb{C}^n).$$

This is called *the C^*-property of the norm.* It makes the Banach algebra $\mathcal{B}(\mathbb{C}^n)$ into a *C^*-algebra*.

By definition, a C^*-algebra $\mathcal{A}$ is a Banach algebra with an adjoint operation for which the C^*-property holds, in addition to some other obvious requirements. For the sake of simplicity only C^*-algebras are considered that have a unit element. The latter is denoted $\mathbb{I}$ in what follows.

An immediate consequence of the C^*-property is that $||T^*|| = ||T||$ for any T in $\mathcal{A}$. Further properties of positive elements in a C^*-algebra $\mathcal{A}$ are

- $T \geq 0$ if and only if there exists $V \geq 0$ in $\mathcal{A}$ such that $T = V^2$;
- $T^*T \geq 0$ for any T in $\mathcal{A}$;
- If $T \geq 0$ then $T \leq ||T||$ (i.e. $||T||\mathbb{I} - T \geq 0$).

In the context of C^*-algebras the terms 'hermitian matrix' and 'positive-definite matrix' are not used. They are replaced by 'self-adjoint operator', respectively 'strictly positive operator'. The 'hermitian conjugate' $A^\dagger$ becomes the 'adjoint operator' and is denoted A^*. Note that by definition a positive operator is also self-adjoint. An element A of a C^*-algebra $\mathcal{A}$ is positive if and only if there exists B in $\mathcal{A}$ such that $A = B^*B$.

A subclass of the C^*-algebras is formed by the von Neumann algebras. Instead of the abstract definition of the latter a practical characterization is given. The commutant $\mathcal{A}'$ of a set $\mathcal{A}$ of bounded linear operators on a Hilbert space $\mathscr{H}$ consists of all bounded linear operators B each of which commutes with all elements of $\mathcal{A}$. This is

$$\mathcal{A}' = \{B \in \mathcal{B}(\mathscr{H}) : BA = AB \text{ for all } A \in \mathcal{A}\}.$$

A C^*-algebra $A \subset \mathcal{B}(\mathscr{H})$ is a von Neumann algebra if it is equal to its bicommutant $\mathcal{A}''$. In general is $\mathcal{A} \subset \mathcal{A}''$. If $\mathcal{A}$ is a commutative C^*-algebra then one has $\mathcal{A} \subset \mathcal{A}' \cap \mathcal{A}''$.

1.2.2 Expectation values

In statistics the expectation value $\mathbb{E}\, f$ of a random variable f is determined by a probability measure μ.

$$\mathbb{E}\, f = \int_\Omega f(x)\mathrm{d}\mu(x).$$

It has the following properties.

- The map $f \mapsto \mathbb{E}\, f$ is a linear functional;

- The expectation $\mathbb{E}\, f$ of a positive random variable f cannot be negative;
- The expectation $\mathbb{E}\, 1$ of the constant function 1 equals 1.

These three properties are carried over to quantum statistics. The expectation value $\mathbb{E}\, T$ of a matrix T is determined by a state ω. The latter replaces the probability measure μ and inherits the above mentioned properties.

Definition 1 *A* state ω *of a* C^**-algebra* $\mathcal{A}$ *is a complex-valued linear function of* $\mathcal{A}$ *with the properties that*

- $\omega(T^*T) \geq 0$ *for all* T *in* $\mathcal{A}$ *(positivity);*
- $\omega(\mathbb{I}) = 1$ *(normalization).*

From the positivity condition it follows that

$$|\omega(T)| \leq ||T|| \quad \text{for all } T \text{ in } \mathcal{A}. \tag{1.2}$$

In particular, any state ω is continuous in norm.

As an example, consider a column vector x in $\mathbb{C}^n$ and assume it is normalized to 1, i.e. $x_{\mathrm{T}}x = 1$. Then a state ω_x on $\mathcal{B}(\mathbb{C}^n)$ is defined by

$$\omega_x(V) \;=\; x_{\mathrm{T}}Vx, \qquad \text{for any } n\text{-by-}n \text{ matrix } V.$$

Such a state is called a *vector state*.

1.2.3 Density matrices

A density matrix ρ is a non-negative-definite matrix with trace equal to 1. Its eigen values are non-negative and sum up to 1. Hence, they can be interpreted as probabilities.

More generally, a density matrix ρ is a positive traceclass operator with trace equal to 1.

Any state ω on the algebra $\mathcal{B}(\mathbb{C}^n)$ of n-by-n matrices is of the form

$$\omega(V) \;=\; \mathrm{Tr}\, \rho V,$$

where ρ is a density matrix. Conversely, given a density matrix ρ then the above expression defines a state ω on $\mathcal{B}(\mathbb{C}^n)$.

1.2.4 Classical probability

An important advantage of the C^*-algebraic approach is that it unifies quantum statistics with classical probability theory.

Let X be a locally-compact Hausdorff space. Let $C_0(X)$ denote the Banach algebra of complex-valued continuous functions on X that vanish at infinity. The norm is the supremum norm. The adjoint of a function is the complex-conjugate function. It is a C^*-algebra. If X is not compact then it is a C^*-algebra without unit. Then the constant functions must be added to make it into a C^*-algebra with unit.

Any probability measure μ on X defines a state ω on $C_0(X)$ by

$$\omega(A) = \int_X A(x)\mathrm{d}\mu(x).$$

Conversely, every state ω on $C_0(X)$ defines a Radon probability measure on X.

Alternatively, consider the algebra $\mathcal{A} = \mathscr{L}_\infty(\mathbb{R}^n, \mathbb{C})$ of all essentially bounded complex functions on Euclidean space $\mathbb{R}^n$ with its Lebesgue measure. It is a von Neumann algebra. Each function A in $\mathcal{A}$ acts by pointwise multiplication $(Af)(x) = A(x)f(x)$ as a bounded operator on the Hilbert space of square-integrable complex functions $\mathscr{L}_2(\mathbb{R}^n)$. The commutant $\mathcal{A}'$ of $\mathcal{A}$ coincides with $\mathcal{A}$. See [24], Part I, Ch. 7, Thm. 2.

Any normalized element f of $\mathscr{L}_2(\mathbb{R}^n)$ determines a vector state ω_f on $\mathcal{A}$ by

$$\omega_f(A) = \int_{\mathbb{R}^n} A(x)|f(x)|^2\mathrm{d}x.$$

1.2.5 Notes

In 1925 Heisenberg [2] proposed to represent observable quantities by matrices. This unconventional idea was the birth of Quantum Mechanics. By assuming that position and velocity of a quantum particle are non-commuting variables he could explain experimental data that classical mechanics cannot explain.

In the second half of the twentieth century mathematicians and mathematical physicists developed a formulation of quantum mechanics and quantum statistical mechanics in terms of C^*-algebras See the introduction of [22] for a short history.

C^*-algebras have been studied extensively by J. Dixmier [8, 20]. A subclass of the C^*-algebras is formed by the von Neumann algebras [10, 24]. These are also called W^*-algebras.

1.3 Conditional probabilities

1.3.1 Empirical data

Consider a probability space X, μ. The probability of an event $A \subset X$ is denoted $p(A)$ and is given by

$$p(A) = \int_X \mathbb{I}_A(x)\mathrm{d}\mu(x),$$

where $\mathbb{I}_A(x)$ equals 1 when $x \in A$ and 0 otherwise. The expectation of random variable f is denoted $\mathbb{E}_\mu f$ and is given by

$$\mathbb{E}_\mu f = \int_X f(x)\mathrm{d}\mu(x).$$

The conditional probability of an event B given an event A with non-vanishing probability is denoted $p(B|A)$ and is defined by

$$p(B|A) = \frac{p(A \cap B)}{P(A)}.$$

This is known as Kolmogorov's definition of conditional probability.

The conditional expectation of a random variable f given an event A with non-vanishing probability $p(A)$ is given by

$$\mathbb{E}_\mu f|A = \frac{1}{p(A)}\mathbb{E}_\mu f\,\mathbb{I}_A.$$

Statistical Inference is concerned with the analysis of empirical data with the intent to describe the data by a probability distribution on the sample space X. See for instance the book of Cox, [54]

In the most common approach a statistical model is available. It contains a few parameters the value of which one tries to estimate from the available data. This is known as parameter fitting. The obvious example is the linear regression model. In the present context one assumes that a non-commutative model is needed to describe the data.

Part of the course deals with model-independent statistical inference. One reason for abandoning models may be that in certain cases no model exists that explains the data. Another motivation is that the acceptation of a model introduces a bias. It is important that the acquisition of data and the preliminary analysis is not contaminated by model assumptions. The drawback of a model-free approach is that one looses the comfort of working with a small number of meaningful parameters. Currently available computer power can partly compensate for that.

1.3.2 Updating

Starting point is the availability of a prior probability measure μ either obtained from an 'educated guess' or as the result of previous observations. When new data become available an update procedure is used to select the posterior probability space. In what follows it is denoted X, ν. The corresponding probability of an event A is denoted $q(A)$.

Consider now the situation that two independent events A and B are measured and that the results of the measurements are used to verify their independence.

The outcome of repeated experiments is the empirical probability of the events A and B, denoted $p^{\text{emp}}(A)$ and $p^{\text{emp}}(B)$, and the empirical conditional probabilities $p^{\text{emp}}(A|B)$ and $p^{\text{emp}}(B|A)$. The question at hand is then to establish a criterion for finding an update ν of the probability distribution μ that is as close as possible to μ while reproducing the empirical results as well as possible.

The event A defines a partition A, A^{c} of the probability space X, μ. Here, A^{c} denotes the complement of A in X. In what follows a slightly more general situation is considered in which the event A is replaced by a partition $(A_i)_{i=1}^n$ of the measure space X, μ into subsets with non-vanishing probability. The notations p_i and μ_i are used, with

$$p_i = p(A_i) \quad \text{and} \quad \mathrm{d}\mu_i(x) = \frac{1}{p_i}\mathbb{I}_{A_i}(x)\mathrm{d}\mu(x). \tag{1.3}$$

Introduce the random variable g defined by $g(x) = i$ when $x \in A_i$. The conditional random variable $\mathbb{E}_\mu f|g$ is then defined by

$$\mathbb{E}_\mu f|g \quad = \quad \sum_i p_i \mathbb{E}_\mu f|A_i.$$

Repeated measurement of the random variable g yields the empirical probabilities

$$p_i^{\text{emp}} \quad = \quad \text{Emp Prob } \{g(x) = i\}.$$

They may deviate from the prior probabilities p_i. In addition one also measures the conditional probabilities

$$p^{\text{emp}}(B|A_i) \quad = \quad \text{Emp Prob of } B \text{ given that } g(x) = i.$$

1.3.3 Breaking of statistical independence

Breaking of statistical independence is a known phenomenon, illustrated by the following example.

Consider three binary variables. They take values 0 and 1 with equal probability. Let

$$
\begin{aligned}
A &= \{101, 111, 100, 110\}, & (1.4)\\
B &= \{011, 111, 010, 110\}, & (1.5)\\
C &= \{101, 010, 110\}.
\end{aligned}
$$

Then one has $p(A) = p(B) = 1/2$ and $p(A \cap B) = p(\{111, 110\}) = 1/4 = p(A)p(B)$. Hence, A and B are statistically independent. However, after conditioning on C one obtains

$$p(A \cap B|C) = \frac{1}{3} \neq p(A|C)\, p(B|C) = \frac{4}{9}.$$

One concludes that the conditioning on C breaks the independence of the events A and B. The third binary variable is irrelevant for the events A and B and is therefore called a *hidden variable*. It is relevant for the event C and makes A and B dependent after conditioning on C.

The EPR paradox of quantum mechanics [3, 7], discussed further on, concerns a breaking of statistical independence that cannot be explained by assuming one or more hidden variables [11, 21]. Experimental verification followed in [25]. The analysis of the experimental data is done in a model-independent way precisely because the intention is to show that no model of classical statistics can explain the data. Conditional probabilities are measured in an empirical manner without making use of the Kolmogorovian definition of conditional probabilities.

Two statistically independent events A and B are monitored in a sequence of experiments. If A holds then one verifies whether B or its complement B^{c} holds. This yields the empirical probabilities $p^{\text{emp}}(B|A)$ and $p^{\text{emp}}(B^{c}|A)$, the sum of which equals 1. Breaking of statistical independence is then verified by comparing $p^{\text{emp}}(B|A)$ to $p^{\text{emp}}(B)$.

1.4 A historical experiment

1.4.1 The EPR paradox

Let us analyze a historical experiment [25]. In the lab two photons leave at the same moment of time in opposite directions. The photons have opposite polarization. However, the actual value of the polarization is not known and cannot be known in advance. The point where the photons are generated is called the source. Two detectors are placed at equal distance from the source. Two random variables A

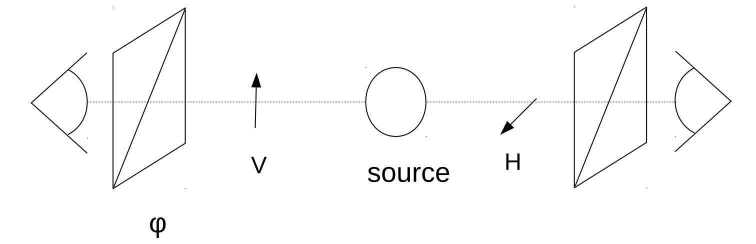

Figure 1.1: Experimental setup

and B are measured. Their value is either 1 if a photon is detected or 0 otherwise. Whether a photon is detected or not depends on whether the polarization of the photon matches the position of the polarizer. The experimental setup contains one free parameter ϕ, which is the angle of a polarizer placed in the beam reaching one of the two detectors. The other detector has a polarizer with a fixed orientation.

A puzzling situation arises when one tries to correlate the measurements of the two detectors. The experimental result [25] is compatible with the theoretical prediction [11] that the conditional probability of a coincidence of photons (i.e. $A = B = 1$) satisfies the relation

$$p^{\text{emp}}(B|A) = p^{\text{emp}}(B) + \kappa \cos(2\phi), \tag{1.6}$$

with $\kappa \neq 0$. However, (1.6) is not what one expects intuitively.

The photon leaves the source with arbitrary polarization. Half of the time it has the correct polarization to be detected. If the polarization is wrong then no photon is detected. Hence, the prior probabilities are $p(A) = p(B) = 1/2$. The two events A and B are statistically independent. This is,

$$p(A \cap B) = p(A)p(B).$$

From the Kolmogorovian definition

$$p(B|A) = \frac{p(A \cap B)}{p(A)} \tag{1.7}$$

one expects that the constant κ should vanish, which is experimentally not the case. The common explanation is that the detection of one of the events influences the other event. For Einstein, Podolsky and Rosen [3] such a result was not acceptable. The two measurements are statistically dependent, even if they are made at distant locations and do not communicate by any other means. Nowadays the phenomenon is well-established altough it remains difficult to understand what is going on.

$A = \{(1,0),(1,1)\}$	$p(A) = \frac{1}{2}$	$q(A\|A) = 1$	$q(A) = \frac{1}{2}$
$L = \{(1,0)\}$	$p(L) = \frac{1}{4}$	$q(L\|A) = \sin^2\phi$	$q(L) = \frac{1}{2}\sin^2\phi$
$H = \{(1,1)\}$	$p(H) = \frac{1}{4}$	$q(H\|A) = \cos^2\phi$	$q(H) = \frac{1}{2}\cos^2\phi$
$B = \{(0,1),(1,1)\}$	$p(B) = \frac{1}{2}$	$q(B\|A) = \cos^2\phi$	$q(B) = \frac{1}{2}$

Table 1.1: Prior and posterior probabilities in the case of the quantum example. The sets A and B are entangled by the measurement when $\cos^2\phi \neq 1/2$. Indeed, A and B are independent while after the measurement A and B are dependent because $q(A \cap B) = q(H) = (1/2)\cos^2\phi \neq q(A)q(B) = 1/4$.

The reader interested in exploring more recent developments in Quantum Optics is referred to the introductory text of [52].

Let us analyse the experimental setup in more detail.

The probability space is X, μ with $X = \{0,1\} \times \{0,1\}$ and probability $1/4$ for any of the atomic sets of X. The two events under consideration are

$$A = \{(1,0),(1,1)\} \quad \text{and} \quad B = \{(0,1),(1,1)\}.$$

They have equal probability and are independent of each other.

The measurement function g returns 1 on the sets containing $(1,0)$ and/or $(1,1)$ and 2 otherwise. The partition $(A_i)_i$ consists of 2 sets of equal probability $A_1 = A$ and $A_2 = A^c$. The measurement yields $p_1^{\text{emp}} = p_2^{\text{emp}} = 1/2 = p_1 = p_2$. The updated probability measure ν is expected to coincide with the prior probability measure μ because μ is confirmed by the locally available data. However, measurement of conditional probabilities shows that the posterior ν differs from the prior μ.

The empirical data are internally consistent. By this is meant that there exists a probability distribution ν which reproduces them. The complete specification of the updated measure ν is found in Table 1. In this table the conditional probabilities $q(\cdot|\cdot)$ are calculated starting from (1.6).

1.4.2 The Bell inequalities

For a review on Bell's inequalities, see [21].

Before the measurement takes place the events A and B are independent of each other. After the measurement they turn out to be dependent except when the angle ϕ, controlled by the experimenter, satisfies $\cos^2\phi = \sin^2\phi = 1/2$. The standing explanation is that both measurements influence each other. Quantum Mechanics

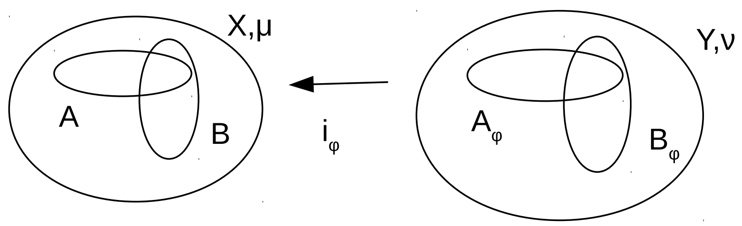

Figure 1.2: Illustration of the proof of the Bell inequality

explains this phenomenon adequately. Still, it remains difficult to understand why it happens.

One suggestion is that the measure space X, μ is a subspace of a larger measure space Y, ν and that the dependence of A and B originates from the projection onto the subspace. This is known in quantum mechanics as the *hidden variable* assumption. If the assumption holds then the projected probabilities satisfy the Bell inequalities [7]. The data of the experiment violate these inequalities. One concludes therefore that the hidden variable assumption, in combination with independence in the larger space, is not appropriate to explain the experimental outcome.

Let us reproduce here the mathematical argument. Introduce the notation

$$q(C) \quad = \quad \int_Y \mathbb{I}_C(y)\mathrm{d}\nu(y).$$

For each angle ϕ let i_ϕ be a measurable map from Y to X such that $p_\phi^{\mathrm{emp}} \circ i_\phi = q$. It follows that

$$p^{\mathrm{emp}}(B \cap A) = q(i_\phi^{-1}(B \cap A)) \leq q(B_\phi \cap A_\phi).$$

Independence of the inverse images $A_\phi = i_\phi^{-1}(A)$ and $B_\phi = i_\phi^{-1}(B)$ implies that $q(A_\phi \cap B_\phi) = q(A_\phi)q(B_\phi)$. Hence, one obtains

$$p^{\mathrm{emp}}(B \cap A) \leq q(B_\phi)q(A_\phi) = p^{\mathrm{emp}}(B)p^{\mathrm{emp}}(A).$$

The experimental data give $p^{\mathrm{emp}}(A) = p^{\mathrm{emp}}(B) = 1/2$ and

$$p^{\mathrm{emp}}(B \cap A) = \frac{1}{4} + \frac{\kappa}{2}\cos 2\phi.$$

The requirement that follows is that $\kappa \cos 2\phi \leq 0$ for all angles ϕ. The latter implies that $\kappa = 0$, which is experimentally violated.

Chapter 2

The Bures metric

2.1 Introduction

This course deals with the study of the manifold $\mathbb{M}_n$ of faithful states on the von Neumann algebra $\mathcal{B}(\mathbb{C}^n)$ of n-by-n matrices with complex entries. Starting point is a divergence function $D(\omega_1||\omega_2)$. It estimates how 'close' two states ω_1 and ω_2 are.

A state ω on a von Neumann algebra $\mathcal{A}$ is said to be faithful if $\omega(A^*A) = 0$ implies $A = 0$ for any A in $\mathcal{A}$. In the case of the algebra of n-by-n matrices it is faithful when the corresponding density matrix ρ is invertible.

A divergence function D on the manifold $\mathbb{M}_n$ is a smooth function $\mathbb{M}_n \times \mathbb{M}_n \mapsto [0, +\infty)$ satisfying

$$D(\omega_1||\omega_2) = 0 \quad \text{if and only if } \omega_1 = \omega_2.$$

Note that a divergence function is called a relative entropy in the Physics Literature.

From the divergence function one can derive an inner product on the tangent spaces of the manifold. Two choices of divergence function are discussed during the course. The topic of the present Chapter is the divergence function obtained by taking the square of the Bures distance.

The mathematical work of Bures [13] was picked up by Uhlmann [19, 46], by Jozsa [35] and by Dittmann [38] between others. Independently, Wootters [26] introduced a notion of quantum statistical distance. The Bures distance found wide acceptance in the domain of Quantum Information Theory [57]. As a consequence there exists an extensive literature on the topic. For a survey see Sommers and Życzkowski [48].

The classical analogue of the Bures distance is the Hellinger distance. The latter is sometimes called the statistical distance. The Bures distance is related to the Wasserstein distance. The Fubini-Study metric is related [39] to the Bures metric. It is not discussed in the present course.

The relative entropy of Umegaki [5] and the corresponding inner product of Bogoliubov are discussed in Chapter 3.

A rather detailed treatment of Bures' original work is given because it is based on the C^*-algebraic approach and fits well with the approach taken in later Chapters.

2.2 The Bures distance

2.2.1 Bures' definition

Consider the tensor product $\mathbb{C}^n \otimes \mathbb{C}^n$ of the Hilbert space $\mathbb{C}^n$ with itself. It is again a complex Hilbert space with inner product defined by linear extension of

$$(a \otimes b, c \otimes d) \quad = \quad (a, c)(d, b).$$

Note that it is linear in a and d and anti-linear in b and c.

Lemma 1 *The state ω determined by the density matrix ρ can be written as*

$$\omega(A) = \textit{Tr}\, \rho A = (A \otimes \mathbb{I}\, \Omega, \Omega), \qquad A \in \mathcal{A}. \tag{2.1}$$

with Ω a vector in $\mathbb{C}^n \otimes \mathbb{C}^n$.

Such a vector exists. Indeed, let $(f_i)_i$ be an orthonormal basis in which ρ is diagonal. Let $\rho f_i = p_i\, f_i$. Then one can take $\Omega = \sum_i \sqrt{p}_i\, f_i \otimes f_i$.

Let $S(\omega)$ denote the set of all vectors Ω in $\mathbb{C}^n \otimes \mathbb{C}^n$ for which (2.1) holds. The definition of Bures [13] applied to this special situation reads

$$d(\omega_1, \omega_2) \quad = \quad \inf\{||\Omega_1 - \Omega_2|| : \Omega_1 \in S(\omega_1), \Omega_2 \in S(\omega_2)\}.$$

Proposition 1 $d(\omega_1, \omega_2)$ *is a distance function.*

Proof

The proof is straightforward except for the non-degeneracy. Let $\epsilon > 0$. Then $d(\omega_1, \omega_2) = 0$ implies that $\Omega_1 \in S(\omega_1)$ and $\Omega_2 \in S(\omega_2)$ exist such that $||\Omega_1 - \Omega_2|| \leq \epsilon$. This gives

$$|\omega_1(A) - \omega_2(A)| \quad = \quad |(A \otimes \mathbb{I}\Omega_1, \Omega_1) - (A \otimes \mathbb{I}\Omega_2, \Omega_2)|$$

$$
\begin{aligned}
&= |(A \otimes \mathbb{I}\Omega_1, \Omega_1 - \Omega_2) - (A \otimes \mathbb{I}(\Omega_2 - \Omega_1), \Omega_2)| \\
&\leq |(A \otimes \mathbb{I}\Omega_1, \Omega_1 - \Omega_2)| + |(A \otimes \mathbb{I}(\Omega_2 - \Omega_1), \Omega_2)| \\
&\leq ||A||\,||\Omega_1 - \Omega_2|| + ||A||\,||\Omega_2 - \Omega_1|| \\
&\leq 2\epsilon||A||.
\end{aligned}
$$

This shows that $\omega_1 = \omega_2$.

□

2.2.2 Technicalities

In the next Section an explicit expression for Bures' distance is derived. Some technical results needed for that follow here.

Lemma 2 *If $U \in \mathcal{A}$ is unitary then*

$$| \mathit{Tr}\, AU| \quad \leq \quad \mathit{Tr}\, (A^*A)^{1/2}, \qquad A \in \mathcal{A}.$$

Proof

Let $A = J|A|$ denote the polar decomposition of the matrix A. It satisfies $J^*j = \mathbb{I}$ and $|A| = (A^*A)^{1/2}$.

Let $(f_i)_i$ be an orthonormal basis which diagonalizes $|A|$: $|A|f_i = a_i f_i$ with $a_i \geq 0$. One has

$$
\begin{aligned}
|\,\mathrm{Tr}\, AU| &= |\,\mathrm{Tr}\, J|A|U| \\
&= |\sum_i a_i(UJf_i, f_i)\,| \\
&\leq \sum_i a_i|(UJf_i, f_i)\,|.
\end{aligned}
$$

Because f_i and UJf_i both have unit length one has $|(UJf_i, f_i)| \leq 1$ so that

$$|\,\mathrm{Tr}\, AU| \leq \sum_i a_i = \mathrm{Tr}\, |A| = \mathrm{Tr}\, (A^*A)^{1/2}.$$

□

Lemma 3 *Choose an orthonormal basis $(f_i)_i$ in $\mathbb{C}^n$. For any vector Φ in $\mathbb{C}^n \otimes \mathbb{C}^n$ there exist $q_i \geq 0$ and unitary operators U and V such that*

$$\Phi \quad = \quad \sum_i \sqrt{q_i} U f_i \otimes V f_i.$$

Proof

By the Schmidt decomposition theorem there exist orthonormal sets $(s_i)_i$ and $(r_i)_i$ and complex numbers $(\lambda_i)_i$ such that

$$\Phi = \sum_i \lambda_i \, s_i \otimes r_i.$$

Take $q_i = |\lambda_i|^2$ and let $\lambda_i = e^{i\phi_i}\sqrt{q_i}$. There exists a unitary operator U which maps the basis vectors f_i onto the basis vectors $e^{i\phi_i} s_i$. Similarly, there exists a unitary operator V which maps the basis vectors f_i onto the basis vectors r_i. Hence one has

$$\begin{aligned}\Phi &= \sum_i \lambda_i \, s_i \otimes r_i \\ &= \sum_i \sqrt{q_i}(e^{i\phi_i} s_i) \otimes r_i \\ &= \sum_i \sqrt{q_i} U f_i \otimes V f_i.\end{aligned}$$

This proves the lemma.

□

Proposition 2 *Consider a pair of states ω_1, ω_2 on $\mathcal{B}(\mathbb{C}^n)$. For any Ω_1 in $S(\omega_1)$ is*

$$d(\omega_1, \omega_2) = \inf\{||\Omega_1 - \Omega_2|| : \Omega_2 \in S(\omega_2)\}.$$

Proof

Let ρ be the density matrix corresponding with ω_1. Let $(f_i)_i$ be a basis in which ρ is diagonal. Let $\rho f_i = p_i f_i$. Let $\Omega_1' = \sum_i \sqrt{p_i} f_i \otimes f_i$ and

$$\mathscr{H}_{\Omega_1'} = (\mathcal{B}(\mathbb{C}^n) \otimes \mathbb{I})\Omega_1' \subset \mathbb{C}^n \otimes \mathbb{C}^n.$$

A linear operator W on $\mathscr{H}_{\Omega_1'}$ is defined by

$$W(A \otimes \mathbb{I})\Omega_1' = (A \otimes \mathbb{I})\Omega_1, \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

It satisfies

$$||W(A \otimes \mathbb{I})\Omega_1'||^2 = ||(A \otimes \mathbb{I})\Omega_1||^2 = \omega_1(A^* A) = ||(A \otimes \mathbb{I})\Omega_1'||, \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

Hence, W is isometric and it is well-defined because $(A \otimes \mathbb{I})\Omega_1' = 0$ implies that $(\mathbb{I} \otimes A)\Omega_1 = 0$. Extend W from $\mathscr{H}_{\Omega_1'}$ to all of $\mathbb{C}^n \otimes \mathbb{C}^n$ in such a way that it becomes an unitary operator.

One verifies easily that W commutes with all of $\mathcal{B}(\mathbb{C}^n) \otimes \mathbb{I}$.

For any $\Omega_2 \in S(\omega_2)$ is

$$||\Omega_1 - \Omega_2|| = ||W\Omega_1' - \Omega_2|| = ||\Omega_1' - W^*\Omega_2||.$$

Note that $W^*\Omega_2$ belongs to $S(\omega_2)$. Indeed, one has

$$((A \otimes \mathbb{I})W^*\Omega_2, W^*\Omega_2) = (W^*(A \otimes \mathbb{I})\Omega_2, W^*\Omega_2) = ((A \otimes \mathbb{I})\Omega_2, \Omega_2), \qquad A \in \mathcal{B}(\mathbb{C}^n)$$

because W^* commutes with all of $\mathcal{A}$ and is an unitary operator. Hence one has

$$||\Omega_1 - \Omega_2|| \geq \inf\{||\Omega_1' - \Omega_2'|| : \Omega_2' \in S(\omega_2)\}.$$

Similarly is for any Ω_2' in $S(\omega_2)$

$$\begin{aligned} ||\Omega_1' - \Omega_2'|| &= ||W^*\Omega_1 - W^*\Omega_2'|| = ||\Omega_1 - W^*\Omega_2'|| \\ &\geq \inf\{||\Omega_1 - \Omega_2|| : \Omega_2 \in S(\omega_2)\}. \end{aligned}$$

Combination of both inequalities gives

$$\inf\{||\Omega_1 - \Omega_2|| : \Omega_2 \in S(\omega_2)\} = \inf\{||\Omega_1' - \Omega_2'|| : \Omega_2' \in S(\omega_2)\}.$$

Hence this quantity does not depend on the choice of Ω_1 in $S(\omega_1)$.

□

2.2.3 The theorem

Theorem 1 (Uhlmann) *Let ω_1 and ω_2 be states defined by the density matrices ρ_1, respectively ρ_2. Then the Bures distance between the two states is given by*

$$d(\omega_1, \omega_2) = \left[2 - 2\, Tr\left[\rho_1^{1/2}\rho_2\rho_1^{1/2}\right]^{1/2}\right]^{1/2}. \tag{2.2}$$

Proof

Let $\rho_1 f_i = p_i f_i$ and

$$\Omega_1 = \sum_i \sqrt{p_i}\, f_i \otimes f_i.$$

By Proposition 2 one has

$$d(\omega_1, \omega_2) = \inf\{||\Omega_1 - \Omega_2|| : \Omega_2 \in S(\omega_2)\},$$

By Lemma 3 there exist unitary operators U and V and numbers $q_i \geq 0$ such that

$$\Omega_2 = \sum_j \sqrt{q_j}\, U f_j \otimes V f_j.$$

Note that this implies that for any A in $\mathcal{A}$ one has

$$\begin{aligned}
\omega_2(A) &= (A \otimes \mathbb{I}\, \Omega_2, \Omega_2) \\
&= \sum_{j,k} \sqrt{q_j q_k}\, (A \otimes \mathbb{I}\, U f_j \otimes V f_j, U f_k \otimes V f_k) \\
&= \sum_j q_j\, (A U f_j, U f_j).
\end{aligned}$$

Hence, if F_j denotes the orthogonal projection onto $\mathbb{C} U f_j$ then $\sum_j q_j F_j$ equals ρ_2, the unique density matrix representing the state ω_2.

From the definition of the Bures distance it follows that $d(\omega_1, \omega_2)$ is the infimum of $||\Omega_1 - \Omega_2||$. Let us now show that $||\Omega_1 - \Omega_2||$ is always larger than or equal to the r.h.s. of (2.2).

One has

$$\begin{aligned}
||\Omega_1 - \Omega_2||^2 &= 2 - 2\Re(\Omega_1, \Omega_2) \\
&= 2 - 2\Re \sum_{i,j} \sqrt{p_i q_j}\, (f_i \otimes f_i, U f_j \otimes V f_j) \\
&= 2 - 2\Re \sum_{i,j} \sqrt{p_i q_j}\, (f_i, U f_j)\, (V f_j, f_i) \\
&= 2 - 2\Re \sum_{i,j} \sqrt{q_j}\, (\sqrt{\rho_1} f_i, U f_j)\, (V f_j, f_i) \\
&= 2 - 2\Re \sum_{i,j} \sqrt{q_j}\, (V f_j, f_i)\, (f_i, \sqrt{\rho_1} U f_j) \\
&= 2 - 2\Re \sum_j \sqrt{q_j}\, (V f_j, \sqrt{\rho_1} U f_j) \\
&= 2 - 2\Re \sum_j (V f_j, \sqrt{\rho_1}\sqrt{\rho_2} U f_j)
\end{aligned}$$

$$
\begin{aligned}
&= 2 - 2\Re \text{ Tr } U^* \sqrt{\rho_2}\sqrt{\rho_1} V \\
&\geq 2 - 2 \text{ Tr } \sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}} \qquad (2.3)
\end{aligned}
$$

This ends one half of the proof.

Next let us prove that the equality can be reached by an appropriate choice of Ω_2 in $S(\omega_2)$. Use the polar decomposition

$$\sqrt{\rho_2}\sqrt{\rho_1} = J|\sqrt{\rho_2}\sqrt{\rho_1}| = J\left[\sqrt{\rho_1}\rho_2\sqrt{\rho_1}\right]^{1/2}$$

to write

$$2 - 2 \text{ Tr } \left[\sqrt{\rho_1}\rho_2\sqrt{\rho_1}\right]^{1/2} = 2 - 2 \text{ Tr } J^* \sqrt{\rho_2}\sqrt{\rho_1}.$$

Now choose an orthonormal basis $(s_j)_j$ diagonalizing ρ_2, with $\rho_2 s_j = q_j s_j$ and let U be the unitary matrix for which $s_j = U f_j$. Further choose $V = J^* U$. Then Ω_2 given by

$$\Omega_2 = \sum_j \sqrt{q_j} U f_j \otimes V f_j$$

belongs to $S(\omega_2)$. Indeed, one has for A in $\mathcal{A}$

$$
\begin{aligned}
(A \otimes \mathbb{I}\, \Omega_2, \Omega_2) &= \sum_{j,k} \sqrt{q_j q_k}\, (A \otimes \mathbb{I}\, U f_j \otimes V f_j, U f_k \otimes V f_k) \\
&= \sum_{j,k} \sqrt{q_j q_k}\, (A U f_j, U f_k)\, (V f_j, V f_k) \\
&= \sum_j q_j\, (A s_j, s_j) \\
&= \text{Tr } \rho_2 A \\
&= \omega_2(A).
\end{aligned}
$$

Now calculate

$$
\begin{aligned}
||\Omega_1 - \Omega_2||^2 &= 2 - 2\Re(\Omega_1, \Omega_2) \\
&= 2 - 2\sum_{i,j} \sqrt{p_i q_j}\, (f_i \otimes f_i, s_j \otimes V f_j) \\
&= 2 - 2\Re(V f_j, f_i)(f_i, \sqrt{\rho_1}\sqrt{\rho_2} s_j) \\
&= 2 - 2\Re(V f_j, \sqrt{\rho_1}\sqrt{\rho_2} s_j) \\
&= 2 - 2\Re \text{ Tr } J^* \sqrt{\rho_2}\sqrt{\rho_1} \\
&= 2 - 2 \text{ Tr } \left[\rho_1^{1/2} \rho_2 \rho_1^{1/2}\right]^{1/2}
\end{aligned}
$$

$\square$

2.2.4 Notes

The quantity

$$F(\rho_1, \rho_2) = \operatorname{Tr}\left[\rho_1^{1/2}\rho_2\rho_1^{1/2}\right]^{1/2}$$

is called the fidelity [35]. See for instance Section 6.1 of [57]. The square of the fidelity is denoted $P(\rho_1, \rho_2)$ and is called the transition probability [39, 58]. These quantities are heavily used in Quantum Information Theory.

The relation between the fidelity of density matrices ρ_1 and ρ_2 and the square of the Bures distance between the corresponding states ω_1 and ω_2 follows from Uhlmann's Theorem

$$d^2(\omega_1, \omega_2) = 2(1 - F(\rho_1, \rho_2)).$$

In the commutative case the Bures distance reduces to a multiple of the Hellinger distance, also called statistical distance, between two discrete probability distributions. Indeed, when ρ_1 and ρ_2 commute then the expression for the square distance simplifies to

$$\begin{aligned} d(\rho_1, \rho_2)^2 &= 2\left(1 - \operatorname{Tr}\rho_1^{1/2}\rho_2^{1/2}\right) \\ &= \operatorname{Tr}\left[\rho_1^{1/2} - \rho_2^{1/2}\right]^2. \end{aligned}$$

Because the matrices ρ_1 and ρ_2 commute they can be diagonalized simultaneously. Let $(p_i)_i$, respectively $(q_i)_i$ their eigenvalues. The above expression becomes

$$d^2(\rho_1, \rho_2) = \sum_i [\sqrt{p_i} - \sqrt{q_i}]^2.$$

This is twice the square of the Hellinger distance between the two probability distributions p and q.

2.3 Geometry of the manifold of states

2.3.1 Tangent vectors

In the present section the state ω is fixed and is assumed to be faithful.

The space $T_\omega\mathbb{M}_n$ of tangent vectors at the point ω of $\mathbb{M}_n$ can be introduced in more than one way. On the tangent space an inner product is introduced. It can be combined with a connection at choice. For that reason a tangent vector is often defined

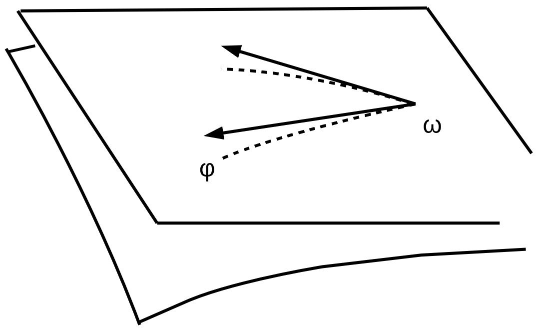

Figure 2.1: Tangent vector of a path from state ω to state ϕ

as an equivalence class of smooth paths on the manifold passing through the given point ω and having the same linearized behavior in a vicinity of ω. Alternatively, if a chart is available the tangent vectors can be introduced by taking derivatives w.r.t. the coordinates of the chart.

An obvious chart of the manifold consists of the matrix coefficients of the density matrix in a chosen basis. The dimension of the manifold is $n^2 - 1$. The -1 is because of the normalization condition.

Instead of starting from a chart the treatment given below introduces tangent vectors by choosing a particular connection and taking derivatives along a geodesic. This procedure does not take away the possibility of considering other connections later on. The choice made here is called the m-connection, with the 'm' from 'mixture'. The maps $\lambda \mapsto (1-\lambda)\omega + \lambda\phi$ are the geodesics of this geometry.

Because ω is faithful there exists for any state ϕ on $\mathcal{A}$ an $\epsilon > 0$ such that $\phi_\lambda = (1-\lambda)\omega + \lambda\phi$ belongs to $\mathbb{M}_n$ for all λ in the open interval $(-\epsilon, \epsilon)$. The derivative

$$\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0} \phi_\lambda \;=\; \phi - \omega$$

is a vector tangent to $\mathbb{M}_n$ at the point ω. It belongs to the tangent space $T_\omega\mathbb{M}_n$. The latter is in one-to-one correspondence with the space of all hermitian linear functionals f on $\mathcal{A}$ satisfying $f(\mathbb{I}) = 0$.

A special state on $\mathcal{A}$ is the tracial state, denoted ω^c. It corresponds with the choice of density matrix $\rho = \mathbb{I}/n$. The exponential map in the point ω^c of $\mathbb{M}_n$ maps a tangent vector $\omega - \omega^c$ onto the state ω. The image of $\mathbb{M}_n$ under the inverse map is an open convex environment of the origin of the Banach space of Hermitian functionals on $\mathcal{A}$ vanishing on $\mathbb{I}$. Hence it is a globally defined chart for the manifold $\mathbb{M}_n$. This observation is discussed further on in Section **??**.

2.3.2 The symmetric logarithmic derivative

Consider three density matrices ρ, σ and τ. In the subsequent section derivatives of

$$\mu \mapsto \left[(1-\mu)\sigma^{1/2}\rho\sigma^{1/2} + \mu\sigma^{1/2}\tau\sigma^{1/2})\right]^{1/2}$$

at $\mu = 0$ are needed. The square roots in this expression prohibit a straightforward calculation. The trick used below involves the concept of the Symmetric Logarithmic Derivative (SLD).

Let us start with a general result.

Proposition 3 *Let be given density matrices ρ and σ. Assume ρ is positive-definite. Let the matrix L be defined by*

$$L = 2\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}\sigma e^{-t\rho}. \tag{2.4}$$

It satisfies

(a) $2\sigma = \rho L + L\rho$;

(b) $\rho^2 + 2\sigma = (\rho + L)^2 - L^2$.

Proof

(a) Note that the integral in (2.4) converges in norm because ρ is positive-definite. Verify that

$$\begin{aligned} \rho L + L\rho &= 2\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}(\rho\sigma + \sigma\rho)e^{-t\rho} \\ &= -2\int_0^{+\infty} \mathrm{d}t\, \frac{\mathrm{d}}{\mathrm{d}t}\, e^{-t\rho}\sigma e^{-t\rho} \\ &= -2e^{-t\rho}\sigma e^{-t\rho}\Big|_0^{+\infty} \\ &= 2\sigma. \end{aligned}$$

(b) One has

$$(\rho + L)^2 - L^2 = \rho^2 + L\rho + \rho L = \rho^2 + 2\sigma.$$

$\square$

The equation for solving $2\sigma = (\rho L + L\rho)$ for L is known as the continuous Lyapunov equation.

By definition the SLD of $s \mapsto X_s$ is the solution L_s of

$$\frac{\mathrm{d}}{\mathrm{d}s} X_s = \frac{1}{2}\left(L_s X_s + X_s L_s\right) \tag{2.5}$$

By the previous proposition the solution L_s exists and is explicitly given by an expression of the form (2.4). If the matrices X_s at different values of s mutually commute with each other then the solution of this equation is $L_s = \mathrm{d}(\log X_s)/\mathrm{d}s$. This explains the name of SLD.

Corollary 1

$$\sigma_\lambda^{1/2} \rho \sigma_\lambda^{1/2} = (\rho + L_\lambda + O(\lambda^2))^2$$

with L_λ given by

$$L_\lambda = \int_0^{+\infty} \mathrm{d}t\, e^{-t\rho} \left[\sigma_\lambda^{1/2} \rho \sigma_\lambda^{1/2} - \rho^2\right] e^{-t\rho}. \tag{2.6}$$

Proof

Write

$$\sigma_\lambda^{1/2} \rho \sigma_\lambda^{1/2} = \rho^2 + \left[\sigma_\lambda^{1/2} \rho \sigma_\lambda^{1/2} - \rho^2\right].$$

Apply Proposition 3 to obtain

$$\sigma_\lambda^{1/2} \rho \sigma_\lambda^{1/2} = (\rho + L_\lambda)^2 - L_\lambda^2$$

with L_λ given by (2.6). Note that L_λ is of order λ as λ tends to zero. Hence, one obtains the result as stated.

□

Corollary 2

$$\sigma_\lambda^{1/2} \tau_\mu \sigma_\lambda^{1/2} = (\rho + L_\lambda + M_{\lambda,\mu} + O(\lambda^2) + O(\mu^2))^2.$$

with

$$M_{\lambda,\mu} = \mu \int_0^{+\infty} \mathrm{d}t\, e^{-t(\rho + L_\lambda)} \sigma_\lambda^{1/2} (\tau - \rho) \sigma_\lambda^{1/2} e^{-t(\rho + L_\lambda)}.$$

Proof

By the previous Corollary one has

$$\begin{aligned}\sigma_\lambda^{1/2}\tau_\mu\sigma_\lambda^{1/2} &= \sigma_\lambda^{1/2}\rho\sigma_\lambda^{1/2}+\mu\sigma_\lambda^{1/2}(\tau-\rho)\sigma_\lambda^{1/2}\\ &= (\rho+L_\lambda+\mathrm{O}(\lambda^2))^2+\mu\sigma_\lambda^{1/2}(\tau-\rho)\sigma_\lambda^{1/2}.\end{aligned}$$

Application of Proposition 3 now gives

$$\sigma_\lambda^{1/2}\tau_\mu\sigma_\lambda^{1/2} = (\rho+L_\lambda+\mathrm{O}(\lambda^2)+M_{\lambda,\mu})^2-M_{\lambda,\mu}^2.$$

This implies the stated result.

□

Corollary 3

$$\sigma_\lambda^{1/2} = \rho^{1/2}+\lambda\int_0^{+\infty}\mathrm{d}s\,e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}+O(\lambda^2).$$

Proof

Use Proposition 3 to obtain

$$\begin{aligned}\sigma_\lambda &= \left(\rho^{1/2}\right)^2+\lambda(\sigma-\rho)\\ &= (\rho^{1/2}+L)^2-L^2\end{aligned}$$

with L given by

$$L = \lambda\int_0^{+\infty}\mathrm{d}s\,e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}.$$

This implies the desired result.

□

2.3.3 Riemannian geometry

The Bures distance defines a metric on the manifold $\mathbb{M}_n$ of faithful states. What one needs in the context of Riemannian geometry is an inner product on the tangent planes. This can be obtained by linearizing the Bures distance. See for instance Dittman [38]. Here different techniques are used. The goal is to show that the inner product can be derived starting from Eguchi's work.

Following Eguchi [28, 30] an inner product on the tangent space $T_\omega\mathbb{M}_n$ can be derived from a divergence function by taking two derivatives. In the present context the divergence is chosen equal to the square of the Bures distance. Eguchi's expression becomes

$$(\phi-\omega,\psi-\omega)_\omega \;=\; -\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\mathrm{d}}{\mathrm{d}\mu}d^2(\phi_\lambda,\psi_\mu)\bigg|_{\lambda=\mu=0} \tag{2.7}$$

with

$$\phi_\lambda=(1-\lambda)\omega+\lambda\phi \quad \text{and} \quad \psi_\mu=(1-\mu)\omega+\mu\psi.$$

Because the distance $d(\omega,\phi)$ is symmetric in its arguments one has the following result.

Proposition 4 *One has*

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}d^2(\phi_\lambda,\phi_\mu)\bigg|_{\mu=\lambda} \;=\; -\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\mathrm{d}}{\mathrm{d}\mu}d^2(\phi_\lambda,\phi_\mu)\bigg|_{\lambda=\mu}. \tag{2.8}$$

This implies that

$$(\phi-\omega,\phi-\omega)_\omega \;=\; \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}d^2(\phi_\lambda,\omega)\bigg|_{\lambda=0}.$$

Proof

From $d(\phi_\lambda,\phi_\lambda)=0$ it follows that

$$0 \;=\; \frac{\mathrm{d}}{\mathrm{d}\lambda}d^2(\phi_\lambda,\phi_\mu)\bigg|_{\mu=\lambda}.$$

Take another derivative to obtain

$$\begin{aligned}0 &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{\mathrm{d}}{\mathrm{d}\lambda}d^2(\phi_\lambda,\phi_\mu)\bigg|_{\mu=\lambda}\right]\\ &= \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}d^2(\phi_\lambda,\phi_\mu)\bigg|_{\mu=\lambda}+\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\mathrm{d}}{\mathrm{d}\mu}d^2(\phi_\mu,\phi_\lambda)\bigg|_{\mu=\lambda}.\end{aligned}$$

This gives (2.8). The remainder of the proof is straightforward.

□

Theorem 2 *Let ω, ϕ and ψ be faithful states on the von Neumann algebra of bounded linear operators on $\mathbb{C}^n$. Let ρ, σ and τ be the corresponding density matrices. The inner product defined by (2.7) is given by*

$$(\phi - \omega, \psi - \omega)_\omega \;=\; \int_0^{+\infty} \mathrm{d}t \;\mathit{Tr}\, e^{-t\rho}(\sigma - \rho)e^{-t\rho}(\tau - \rho) \qquad (2.9)$$

The proof is found in the Appendix.

Expression (2.9) is known as Bures' metric. In the Literature it is often written as

$$(\phi - \omega, \psi - \omega)_\omega \;= \tfrac{1}{2} \quad \mathrm{Tr}\,(\sigma - \rho)G$$

with G the solution of

$$\frac{1}{2}\,(\rho G + G\rho) \;=\; \tau - \rho.$$

2.3.4 Affine coordinates

Let us now introduce a coordinate representation in which all connection coefficients Γ^k_{ij} of the mixture connection vanish.

The tangent space $T_\omega\mathbb{M}_n$ consists of all Hermitian functionals χ on $\mathcal{B}(\mathbb{C}^n)$ satisfying $\chi(\mathbb{I}) = 0$. Indeed, the tangent space $T_\omega\mathbb{M}_n$ is spanned by differences $\phi - \omega$ with ϕ any state in $\mathbb{M}_n$. Let ρ and σ be the density matrices of ω, respectively ϕ. Then one has

$$\phi(A) - \omega(A) \;=\; \mathrm{Tr}\,(\sigma - \rho)A, \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

The difference $\sigma - \rho$ is a Hermitian matrix with vanishing trace. It belongs to the space $\mathcal{A}^0_{\mathrm{sa}}$ of traceless Hermitian matrices. It is easy to show that there is a one-to-one correspondence between tangent vectors χ in $T_\omega\mathbb{M}_n$ and matrices X in $\mathcal{A}^0_{\mathrm{sa}}$. It is given by

$$\chi(A) = \;\mathrm{Tr}\,XA, \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

The space $\mathcal{A}^0_{\mathrm{sa}}$ is a real Hilbert space for the Hilbert-Schmidt inner product

$$(X, Y)_{\mathrm{HS}} = \;\mathrm{Tr}\,XY.$$

Hence one can construct an orthonormal set $(B_i)_i$ of vectors in $\mathcal{A}^0_{\mathrm{sa}}$. They satisfy $(B_i, B_j)_{\mathrm{HS}} = \delta_{i,j}$. A matrix X in $\mathcal{A}^0_{\mathrm{sa}}$ can then be expanded as

$$X \;=\; (X, B_i)_{\mathrm{HS}} B^i.$$

Any density matrix ρ of a state ω can be expanded as

$$\rho = \frac{1}{n}\mathbb{I} + (\,\mathrm{Tr}\,\rho B^i)B_i = \frac{1}{n}\mathbb{I} + \omega(B^i)B_i.$$

Introduce basis vectors $e_i^{(\mathrm{m})}$ in the tangent bundle by

$$[e_i^{(\mathrm{m})}]_\omega(A) \quad = \quad \mathrm{Tr}\, B_i A, \qquad A \in \mathcal{B}(\mathbb{C}^n),$$

independent of ω. The map $\omega \mapsto [e_i^{(\mathrm{m})}]_\omega$ is a constant vector field. The state ω now satisfies

$$\omega(A) = \,\mathrm{Tr}\,\rho A = \frac{1}{n}\,\mathrm{Tr}\,A + \omega(B^i)\,\mathrm{Tr}\,B_i A = \frac{1}{n}\,\mathrm{Tr}\,A + \omega(B^i)[e_i^{(\mathrm{m})}]_\omega(A), \quad A \in \mathcal{B}(\mathbb{C}^n). \tag{2.10}$$

Introduce coordinates $\xi(\omega)$ defined by $\xi^i(\omega) = \omega(B_i)$. The expression for a tangent vector $\phi - \omega$ in $T_\omega\mathbb{M}_n$ then becomes

$$\phi - \omega \quad = \quad \left(\xi^i(\phi) - \xi^i(\omega)\right)[e_i^{(\mathrm{m})}]_\omega.$$

The metric tensor of the Bures metric is given by

$$\begin{aligned} g_{ij}(\omega) &= (e_i^{(\mathrm{m})}, e_j^{(\mathrm{m})})_\omega \\ &= \int_0^{+\infty} \mathrm{d}t\ \,\mathrm{Tr}\,e^{-t\rho}B_i e^{-t\rho}B_j \end{aligned} \tag{2.11}$$

with ρ the density matrix of ω. The inner product takes on the form

$$(\phi - \omega, \psi - \omega)_\omega \quad = \quad (\xi^i(\phi) - \xi^i(\omega))\, g_{ij}(\omega)\, (\xi^j(\phi) - \xi^j(\omega)) \tag{2.12}$$

Consider now the path $\lambda \mapsto \phi_\lambda = \omega + \lambda(\phi - \omega)$, as before. Then one finds

$$\phi - \omega = \frac{\mathrm{d}}{\mathrm{d}\lambda}\phi_\lambda = \dot{\xi}^i\partial_i\phi_\lambda = \left[\dot{\xi}^i e_i^{(\mathrm{m})}\right]_{\phi_\lambda},$$

with $\dot{\xi}^i = \mathrm{d}\xi^i/\mathrm{d}\lambda$. Because the l.h.s. does not depend on λ and neither do the basis functions $e_i^{(\mathrm{m})}$ one obtains $\ddot{\xi}^i = 0$. This is a special case of the Euler-Lagrange equation

$$\ddot{\xi}^k + \Gamma^k_{ij}\dot{\xi}^i\dot{\xi}^j \quad = \quad 0,$$

with connection coefficients Γ^k_{ij} identically equal to 0. Hence the path $\lambda \mapsto \phi_\lambda$ is a geodesic.

2.3.5 Special role of the tracial state

The tracial state $\omega^{\rm c}$ occupies a special status. Its density matrix $\rho^{\rm c}$ is the identity operator $\mathbb{I}$ divided by n. The decomposition (2.10) of the state ω can therefore be written as

$$\omega = \omega^{\rm c} + \xi^i(\omega)[e_i^{\rm (m)}]_\omega.$$

The chart ξ is a global chart for the manifold $\mathbb{M}_n$. Note that $\xi(\omega^{\rm c}) = 0$. This shows that the chart ξ is centered at the tracial state $\omega^{\rm c}$.

2.4 The case n=2

2.4.1 Introduction

The manifolds $\mathbb{M}_n$ with $n = 2, 3$ have been studied in detail. In the case of 2-by-2 density matrices the geometry is clear. Early references are (Hübner 1992) [31] and (Jozsa, 1994) [35]. (Dittmann, 1999) [43] showed that for $n = 3$ the Riemannian curvature is not constant. A recent work on $n = 3$ is (Ercolessi, Schiavina 2013) [62]. The overall impression is that $n = 2$ is an exception and that it is not easy to analyse the manifold $\mathbb{M}_n$ for $n \geq 3$.

2.4.2 The Bloch sphere

The case $n = 2$ is especially important in Quantum Mechanics. It is known since long that the vector states on $\mathcal{B}(\mathbb{C}^2)$ can be represented by the points on the unit sphere of $\mathbb{R}^3$, called the Bloch sphere in this context. The interior points with faithful states. Any state on $\mathcal{B}(\mathbb{C}^2)$ is either a vector state or a faithful state. This is not anymore the case for $n \geq 3$.

A chart which maps any faithful density matrix ρ onto a point in $\mathbb{R}^3$ is given by

$$\rho = \frac{1}{2}\begin{pmatrix} 1+x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix} \qquad \text{with } |x| < 1.$$

It can be expressed in terms of the Pauli matrices σ_i as

$$\rho = \frac{1}{2}\mathbb{I} + x^i\sigma_i.$$

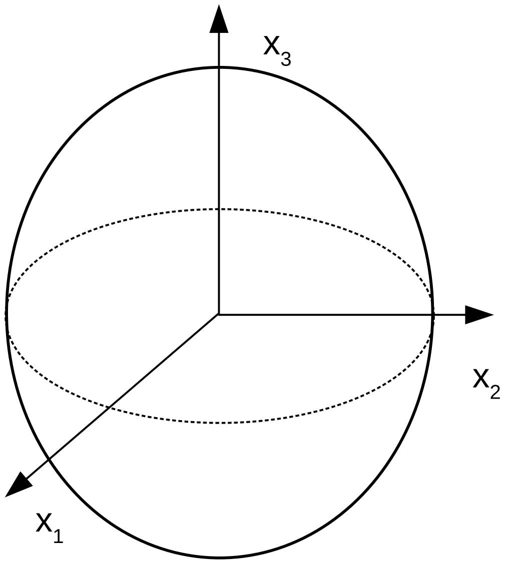

Figure 2.2: The Bloch sphere

The Pauli matrices are defined by

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

Choose now basis vectors B_i in Hilbert-Schmidt space proportional to the Pauli matrices

$$B_i \quad = \quad \frac{1}{\sqrt{2}}\sigma_i, \qquad i = 1, 2, 3.$$

It is straightforward to verify that $(B_i, B_j)_{\text{HS}} = \delta_{i,j}$ holds. The coordinates $\xi^i(\omega)$ introduced before are then found to equal $\xi^i(\omega) = x^i/\sqrt{2}$. This implies that the coordinates x^i are affine coordinates as well and that the geodesics of the m-connection correspond with straight lines through the Bloch sphere.

The calculation of an explicit expression for the metric tensor $g_{i,j}(\omega)$ is rather lengthy. A short discussion follows below.

The eigenvalues of ρ are $(1 \pm |x|)/2$. Hence, there exists a unitary matrix $U = U(\xi)$ such that

$$U^*\rho U \quad = \quad \frac{1}{2}\begin{pmatrix} 1 + |x| & 0 \\ 0 & 1 - |x| \end{pmatrix}.$$

This implies

$$U^* e^{-t\rho} U \quad = \quad \exp(-t(U^*\rho U))$$

$$
\begin{aligned}
&= \begin{pmatrix} \exp(-t(1+|x|)/2) & 0 \\ 0 & \exp(-t(1-|x|)/2) \end{pmatrix} \\
&= e^{-t/2} \left[\cosh \frac{t|x|}{2} \mathbb{I} - \sinh \frac{t|x|}{2} \sigma_3 \right].
\end{aligned}
$$

The expression (2.11) for the metric tensor becomes

$$
\begin{aligned}
g_{ij}(\omega) &= \int_0^{+\infty} \mathrm{d}t\, e^{-t}\, \mathrm{Tr}\, U \left[\cosh \frac{t|x|}{2} \mathbb{I} - \sinh \frac{t|x|}{2} \sigma_3 \right] U^* B_i \\
&\quad \times U \left[\cosh \frac{t|x|}{2} \mathbb{I} - \sinh \frac{t|x|}{2} \sigma_3 \right] U^* B_j \\
&= \frac{1}{2} \left[1 + \frac{1}{1-|x|^2} \right] \delta_{ij} \\
&- \frac{1}{2} \frac{|x|}{1-|x|^2} \left[\mathrm{Tr}\, U\sigma_3 U^* B_i B_j + \mathrm{Tr}\, B_i U \sigma_3 U^* B_j \right] \\
&+ \frac{1}{2} \frac{|x|^2}{1-|x|^2}\, \mathrm{Tr}\, U\sigma_3 U^* B_i U \sigma_3 U^* B_j.
\end{aligned}
$$

2.5 Appendix

Proof of Theorem 2.9 From Corollary 2 one obtains

$$
d^2(\phi_\lambda, \psi_\mu) = 2 - 2\, \mathrm{Tr}\, [\rho + L_\lambda + M_{\lambda,\mu}] + \mathrm{O}(\lambda^2) + \mathrm{O}(\mu^2).
$$

This implies

$$
\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu=0} d^2(\phi_\lambda, \psi_\mu) &= -2 \left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu=0} \mathrm{Tr}\, M_{\lambda,\mu} \\
&= -2 \int_0^{+\infty} \mathrm{d}t\ \mathrm{Tr}\, e^{-t(\rho+L_\lambda)} \sigma_\lambda^{1/2} (\tau - \rho) \sigma_\lambda^{1/2} e^{-t(\rho+L_\lambda)} + \mathrm{O}(\lambda^2) \\
&= -2 \int_0^{+\infty} \mathrm{d}t\ \mathrm{Tr}\, \sigma_\lambda^{1/2} e^{-2t(\rho+L_\lambda)} \sigma_\lambda^{1/2} (\tau - \rho) + \mathrm{O}(\lambda^2) \\
&= -\, \mathrm{Tr}\, \sigma_\lambda^{1/2} (\rho + L_\lambda)^{-1} \sigma_\lambda^{1/2} (\tau - \rho).
\end{aligned}
$$

Lemma 4

$$
\sigma_\lambda^{1/2} (\rho + L_\lambda)^{-1} \sigma_\lambda^{1/2} = \mathbb{I} - \rho^{-1/2} \left[L_\lambda - \lambda(\sigma - \rho) \right] \rho^{-1/2} + o(\lambda). \quad (2.13)
$$

Proof

Let

$$
Y = \sigma_\lambda^{1/2} (\rho + L_\lambda)^{-1} \sigma_\lambda^{1/2} - \mathbb{I}.
$$

Inversion gives

$$(\mathbb{I}+Y)^{-1} \;=\; \sigma_\lambda^{-1/2}(\rho+L_\lambda)\sigma_\lambda^{-1/2}. \qquad (2.14)$$

This can be written as

$$\begin{aligned}\rho+L_\lambda &= \sigma_\lambda^{1/2}(\mathbb{I}+Y)^{-1}\sigma_\lambda^{1/2}\\ &= \sigma_\lambda^{1/2}(\mathbb{I}-Y)\sigma_\lambda^{1/2}+\mathrm{o}(\lambda)\\ &= \sigma_\lambda+\sigma_\lambda^{1/2}Y\sigma_\lambda^{1/2}+\mathrm{o}(\lambda)\\ &= \rho+\lambda(\sigma-\rho)+\rho^{1/2}Y\rho^{1/2}+\mathrm{o}(\lambda).\end{aligned}$$

Here, it is used several times that Y is of order λ. The above expression can be written as

$$Y=\rho^{-1/2}(L_\lambda-\lambda(\sigma-\rho))\rho^{-1/2}.$$

The latter implies (2.13).

□

Now continue with the proof of the Theorem. With the help of the Lemma one obtains

$$\left.\frac{\mathrm{d}}{\mathrm{d}\mu}\right|_{\mu=0} d^2(\phi_\lambda,\psi_\mu) \;=\; \mathrm{Tr}\,\rho^{-1/2}\left[L_\lambda-\lambda(\sigma-\rho)\right]\rho^{-1/2}(\tau-\rho)+\mathrm{o}(\lambda).$$

Minus the derivative of the above expression w.r.t. λ gives the inner product

$$\begin{aligned}(\phi-\omega,\psi-\omega)_\omega &= -\,\mathrm{Tr}\,\rho^{-1/2}\left[\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0} L_\lambda\right]\rho^{-1/2}(\tau-\rho)\\ &+\;\mathrm{Tr}\,\rho^{-1/2}(\sigma-\rho)\rho^{-1/2}(\tau-\rho)\\ &= -\int_0^{+\infty}\mathrm{d}t\;\mathrm{Tr}\,\rho^{-1/2}e^{-t\rho}\left[\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0}\sigma_\lambda^{1/2}\rho\sigma_\lambda^{1/2}\right]e^{-t\rho}\rho^{-1/2}(\tau-\rho)\\ &+\;\mathrm{Tr}\,\rho^{-1/2}(\sigma-\rho)\rho^{-1/2}(\tau-\rho).\end{aligned}$$

From Corollary 3 one gets

$$\begin{aligned}&\int_0^{+\infty}\mathrm{d}t\,e^{-t\rho}\left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\right|_{\lambda=0}\sigma_\lambda^{1/2}\rho\sigma_\lambda^{1/2}e^{-t\rho}\\ =&\int_0^{+\infty}\mathrm{d}t\,e^{-t\rho}\int_0^{+\infty}\mathrm{d}s\,e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}\,\rho^{3/2}e^{-t\rho}\\ +&\int_0^{+\infty}\mathrm{d}t\,e^{-t\rho}\rho^{3/2}\int_0^{+\infty}\mathrm{d}s\,e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}e^{-t\rho}. \qquad (2.15)\end{aligned}$$

Three successive partial integrations yield

$$\begin{aligned}
\int_0^{+\infty} \mathrm{d}s\, e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}\,\rho^{3/2} &= -\int_0^{+\infty} \mathrm{d}s\, e^{-s\sqrt{\rho}}(\sigma-\rho)\frac{\mathrm{d}}{\mathrm{d}s}e^{-s\sqrt{\rho}}\,\rho \\
&= -\int_0^{+\infty} \mathrm{d}s\, \frac{\mathrm{d}}{\mathrm{d}s}\left[e^{-s\sqrt{\rho}}(\sigma-\rho)e^{-s\sqrt{\rho}}\right]\,\rho \\
&\quad +\int_0^{+\infty} \mathrm{d}s\, \left[\frac{\mathrm{d}}{\mathrm{d}s}e^{-s\sqrt{\rho}}\right](\sigma-\rho)e^{-s\sqrt{\rho}}\,\rho \\
&= (\sigma-\rho)\,\rho \\
&\quad -\int_0^{+\infty} \mathrm{d}s\, \rho^{1/2}(\sigma-\rho)e^{-s\sqrt{\rho}}\,\rho \\
&= (\sigma-\rho)\,\rho-\rho^{1/2}(\sigma-\rho)\rho^{1/2} \\
&\quad +\int_0^{+\infty} \mathrm{d}s\, \rho(\sigma-\rho)e^{-s\sqrt{\rho}}\,\rho^{1/2} \\
&= (\sigma-\rho)\,\rho-\rho^{1/2}(\sigma-\rho)\rho^{1/2}+\rho(\sigma-\rho) \\
&\quad -\int_0^{+\infty} \mathrm{d}s\, \rho^{3/2}(\sigma-\rho)e^{-s\sqrt{\rho}}.
\end{aligned}$$

Hence, (2.15) becomes

$$\begin{aligned}
&\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}\frac{\mathrm{d}}{\mathrm{d}\lambda}\bigg|_{\lambda=0}\sigma_\lambda^{1/2}\rho\sigma_\lambda^{1/2}e^{-t\rho} \\
=\ &\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}\left[(\sigma-\rho)\,\rho-\rho^{1/2}(\sigma-\rho)\rho^{1/2}+\rho(\sigma-\rho)\right]e^{-t\rho} \\
=\ &(\sigma-\rho)-\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}\rho^{1/2}(\sigma-\rho)\rho^{1/2}e^{-t\rho}.
\end{aligned}$$

The final result is

$$\begin{aligned}
(\phi-\omega,\psi-\omega)_\omega &= -\operatorname{Tr}\rho^{-1/2}\left[(\sigma-\rho)-\int_0^{+\infty} \mathrm{d}t\, e^{-t\rho}\rho^{1/2}(\sigma-\rho)\rho^{1/2}e^{-t\rho}\right]\rho^{-1/2}(\tau-\rho) \\
&\quad + \operatorname{Tr}\rho^{-1/2}(\sigma-\rho)\rho^{-1/2}(\tau-\rho) \\
&= \int_0^{+\infty} \mathrm{d}t\ \operatorname{Tr} e^{-t\rho}(\sigma-\rho)e^{-t\rho}(\tau-\rho).
\end{aligned}$$

Chapter 3

Exponential arcs

3.1 Introduction

Part of the present and subsequent Chapters is based on work of the author [66, 70, 74].

3.1.1 Motivation

In the previous chapter the chart $\omega \in \mathbb{M}_n \mapsto \xi(\omega)$ is introduced. It is an affine coordinate system for the m-connection. It maps $\mathbb{M}_n$ onto an open convex subset C of $\mathbb{R}^{n^2-1}$. The exponential map is only defined on the set C, not on all of $\mathbb{R}^{n^2-1}$. It takes some work to show that the Euclidean norm $||\xi(\omega)||$ is bounded above by the constant n. As a consequence any geodesic $\lambda \mapsto (1-\lambda)\omega + \lambda\phi$ is mapped onto a straight line $\lambda \mapsto \xi(1-\lambda)\omega + \lambda\phi$ of finite length. See the example $n = 2$ at the end of the previous chapter.

The present chapter introduces a different chart, denoted $\omega \mapsto x(\omega)$. It is shown that the range of this map is all of $\mathbb{R}^{n^2-1}$. A Riemannian manifold is said to be *geodesically complete* if for any tangent plane the exponential map is defined on all of the tangent plane. In the case of an affine coordinate system the tangent plane can be identified with the range of the chart. The goal of the present chapter is to introduce a chart which turns $\mathbb{M}_n$ into a geodesically complete manifold. The corresponding geometry is that of the e-connection.

A metric space is said to be *complete* if every Cauchy sequence has a limit point belonging to the space. The Hopf-Rinow theorem states that a Riemannian manifold is geodesically complete if and only if as a metric space it is complete.

3.1.2 Useful identities

A proof of the following result is found in the book of Amari-Nagaoka [45], p. 156. It is repeated here.

Proposition 5 *For any pair P, Q of positive-definite n-by-n matrices is*

$$P - Q \quad = \quad \int_0^1 \mathrm{d}u\, Q^u(\log P - \log Q)P^{1-u}. \tag{3.1}$$

Proof

Let us first prove that

$$1 - Q^t P^{-t} \quad = \quad \int_0^t \mathrm{d}u\, Q^u(\log P - \log Q)P^{-u}. \tag{3.2}$$

Expression (3.2) is clearly valid at $t = 0$. The derivatives w.r.t. t of both l.h.s. and r.h.s. equal

$$-Q^t(\log Q)P^{-t} + Q^t P^{-t}\log P.$$

Hence, (3.2) is valid at all t,

Take $t = 1$ in (3.2) and multiply from the right with P to obtain (3.1).

□

Proposition 6 *The identity*

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} e^{H+tA} \quad = \quad \int_0^1 \mathrm{d}u\, e^{uH}Ae^{(1-u)H} \tag{3.3}$$

holds for any pair of Hermitian matrices H and A in $\mathcal{B}(\mathbb{C}^n)$,

Proof

Take $P = \exp(H + tA)$ and $Q = \exp H$ in (3.1) to find

$$e^{H+tA} - e^H \quad = \quad t\int_0^1 \mathrm{d}u\, e^{uH}Ae^{(1-u)H}.$$

Divide by t and take the limit $t \to 0$ to obtain (3.3).

□

Note that by substitution of u by $1 - u$ one obtains

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} e^{H+tA} \quad = \quad \int_0^1 \mathrm{d}u\, e^{(1-u)H}Ae^{uH}. \tag{3.4}$$

3.1.3 The Kubo transform

The expression in the r.h.s. of 3.3 is known as a Kubo transform.

Definition 2 *Given a density matrix ρ in $\mathbb{M}_n$ the Kubo transform of a matrix A is defined by*

$$[A]^{\kappa}_{\rho} \;=\; \int_0^1 \mathrm{d}u\, \rho^u\, A\, \rho^{1-u}$$

Proposition 7 *The inverse of the Kubo transform exists and is given by*

$$A \;=\; \int_0^{+\infty} \mathrm{d}t\, \frac{1}{\rho+t}\, [A]^{\kappa}_{\rho}\, \frac{1}{\rho+t} \tag{3.5}$$

Proof

Choose an orthonormal basis $(f_i)_i$ which diagonalizes ρ. It satisfies $\rho f_i = \lambda_i f_i$ with $\lambda_i > 0$. One calculates

$$\begin{aligned} ([A]^{\mathrm{K}}_{\rho} f_i, f_j) &= (A f_i, f_j) \int_0^1 \mathrm{d}u\, \lambda_i^u \lambda_j^{1-u} \\ &= (A f_i, f_j)\, \frac{\lambda_i - \lambda_j}{\log \lambda_i - \log \lambda_j} \end{aligned}$$

and

$$\begin{aligned} &\int_0^{+\infty} \mathrm{d}t\, \left(\frac{1}{\rho+t}\, [A]^{\mathrm{K}}_{\rho}\, \frac{1}{\rho+t} f_i, f_j \right) \\ &= ([A]^{\mathrm{K}}_{\rho} f_i, f_j) \int_0^{+\infty} \mathrm{d}t\, \frac{1}{\lambda_i + t} \frac{1}{\lambda_j + t} \\ &= ([A]^{\mathrm{K}}_{\rho} f_i, f_j)\, \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j} \end{aligned}$$

Hence, one has for all i, j

$$([A]^{\mathrm{K}}_{\rho} f_i, f_j) \;=\; \int_0^{+\infty} \mathrm{d}t\, \left(\frac{1}{\rho+t}\, [A]^{\mathrm{K}}_{\rho}\, \frac{1}{\rho+t} f_i, f_j \right).$$

This implies (3.5).

□

The following properties of the Kubo transform can be derived easily.

- If $A = A^*$ then $([A]^{\mathrm{K}}_\rho)^* = [A]^{\mathrm{K}}_\rho$;
- Tr $A[B]^{\mathrm{K}}_\rho =$ Tr $[A]^{\mathrm{K}}_\rho B$;
- Tr $[A]^{\mathrm{K}}_\rho =$ Tr ρA.

The latter implies that a matrix A with vanishing expectation is transformed into a matrix $[A]^{\mathrm{K}}_\rho$ with vanishing trace.

Proof

One has

$$\begin{aligned} ([A]^{\mathrm{K}}_\rho)^* &= \int_0^1 \mathrm{d}u\, \rho^{1-u}\, A^*\, \rho^u \\ &= \int_0^1 \mathrm{d}u\, \rho^u\, A^*\, \rho^{1-u}. \end{aligned}$$

This shows the first claim.

One has

$$\text{Tr } [A]^{\mathrm{K}}_\rho B = \int_0^1 \mathrm{d}u \ \text{Tr } \rho^u\, A\, \rho^{1-u}\, B.$$

Replace u by $1-u$ and use cyclic permutation under the trace to obtain

$$\begin{aligned} \text{Tr } [A]^{\mathrm{K}}_\rho B &= \int_0^1 \mathrm{d}u \ \text{Tr } \rho^{1-u}\, A\, \rho^u\, B \\ &= \int_0^1 \mathrm{d}u \ \text{Tr } A\, \rho^u\, B\, \rho^{1-u} \end{aligned}$$

$$\text{Tr } \rho\, A\, [B]^{\mathrm{K}}_\rho.$$

This proves the second claim.

Finally, take $B = \mathbb{I}$ and use that $[\mathbb{I}]^{\mathrm{K}}_\rho = \rho$.

□

Note that the matrix $c_\rho(\sigma)$ satisfies Tr $\rho\, c_\rho(\sigma) = 0$. Its Kubo transform $[c_\rho(\sigma)]^{\mathrm{K}}_\rho$ is traceless. Hence, the corresponding linear functional belongs to the tangent plane $T_\omega \mathbb{M}_n$.

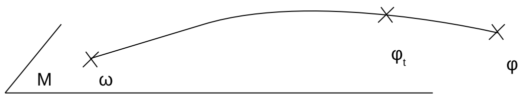

Figure 3.1: Exponential arc connecting the state ϕ to the state ω

3.2 Bogoliubov's metric

3.2.1 Umegaki's relative entropy

Umegaki's relative entropy [5, 17, 18] of the density matrix ρ relative to the density matrix σ is defined by

$$D(\rho||\sigma) \quad = \quad \text{Tr}\, \rho(\log\rho - \log\sigma). \tag{3.6}$$

The expression in the r.h.s. is well-defined for ρ, σ in $\mathbb{M}_n$.

In the mathematics literature a relative entropy is called a divergence. One sees immediately that $D(\rho||\rho) = 0$. The proof that always $D(\rho||\sigma) \geq 0$ is based on Klein's inequality. The proof that $D(\rho||\sigma) = 0$ implies that $\rho = \sigma$ follows from the inequality

$$D(\rho||\sigma) \quad \geq \quad \frac{1}{2}\,\text{Tr}\,(\rho-\sigma)^2.$$

Proofs are given in the Appendix.

In what follows the divergence of a pair of states ϕ, ω is used. It is given by the divergence of the corresponding density matrices σ, ρ, i.e. $D(\omega||\phi) = D(\rho||\sigma)$.

3.2.2 Exponential arcs

Definition 3 *An exponential arc connecting the state ϕ to the state $\omega \in \mathbb{M}_n$ is a map $t \in [0,1] \mapsto \phi_t$ with the property that the density matrices σ_t of the states ϕ_t are given by*

$$\sigma_t \quad = \quad \exp(\log\rho + t(\log\sigma - \log\rho) - \alpha(\phi_t))$$

with

$$\alpha(\phi_t) \quad = \quad \log\ \mathit{Tr}\ \exp(\log\rho + t(\log\sigma - \log\rho)).$$

Note that $\sigma_0 = \rho$, $\sigma_1 = \sigma$, $\alpha(0) = \alpha(1) = 0$.

The vector tangent to the path $t \mapsto \phi_t$ at $\phi_0 = \omega$ is the functional $[\chi_\phi]_\omega$ given by

$$\begin{aligned}[\chi_\phi]_\omega(A) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\phi_t(A)\right|_{t=0} \\ &= \mathrm{Tr}\,\left[\frac{\mathrm{d}}{\mathrm{d}t}\sigma_t\right]_{t=0} A.\end{aligned}$$

Use the identity (3.3) to calculate

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\sigma_t = \int_0^1 \mathrm{d}u\,\rho^u\left[\log\sigma - \log\rho - \left.\frac{\mathrm{d}}{\mathrm{d}t}\alpha(\phi_t)\right|_{t=0}\right]\rho^{1-u}.$$

As a result the tangent vector is given by

$$[\chi_\phi]_\omega(A) = \left[\log\sigma - \log\rho - \left.\frac{\mathrm{d}}{\mathrm{d}t}\alpha(\phi_t)\right|_{t=0}\right]_\rho^{\mathrm{K}} A, \qquad A \in \mathcal{A}. \tag{3.7}$$

Note that

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}t}\alpha(\phi_t)\right|_{t=0} &= \mathrm{Tr}\int_0^1 \mathrm{d}u\,\rho^u[\log\sigma - \log\rho]\rho^{1-u} \\ &= -\,\mathrm{Tr}\,\rho(\log\rho - \log\sigma) \\ &= -D(\rho||\sigma),\end{aligned}$$

with $D(\rho||\sigma)$ Umegaki's relative entropy. Hence, one can write

$$[\chi_\phi]_\omega(A) = \mathrm{Tr}\,c_\rho(\sigma)\left(\int_0^1 \mathrm{d}u\,\rho^{1-u}A\rho^u\right) = \mathrm{Tr}\,[c_\rho(\sigma)]_\rho^{\mathrm{K}} A, \qquad A \in \mathcal{B}(\mathbb{C}^n), \tag{3.8}$$

with the operator $c_\rho(\sigma)$ defined by

$$c_\rho(\sigma) = \log\sigma - \log\rho + D(\rho||\sigma).$$

3.2.3 Geodesic completeness

Proposition 8 *For any ω in $\mathbb{M}_n$ the map $\phi \in \mathbb{M}_n \mapsto [\chi_\phi]_\omega$ is one-to-one.*

Proof

Assume ϕ_t and ψ_t connect ϕ respectively ψ to ω and assume that $[\chi_\phi]_\omega = [\chi_\psi]_\omega$. Let ρ, σ, τ be the density matrices of ω, ϕ, ψ. One has for all A in $\mathcal{B}(\mathbb{C}^n)$

$$\mathrm{Tr}\,[c_\rho(\sigma)]_\rho^K A = [\chi_\phi]_\omega(A)$$

$$\begin{aligned} &= [\chi_\psi]_\omega(A) \\ &= \text{Tr}\, [c_\rho(\tau)]^K_\rho\, A. \end{aligned}$$

This implies that $[c_\rho(\sigma)]^K_\rho = [c_\rho(\tau)]^K_\rho$. Because the Kubo transform is invertible it follows that

$$\begin{aligned} \log\sigma - \log\rho + D(\rho||\sigma) &= c_\rho(\sigma) \\ &= c_\rho(\tau) \\ &= \log\tau - \log\rho + D(\rho||\tau). \end{aligned}$$

Multiply with ρ and take the trace to find that $D(\rho||\sigma) = D(\rho||\tau)$ and hence that $\log\sigma = \log\tau$. The latter implies that $\sigma = \tau$ and that $\phi = \psi$.

□

Theorem 3 *Fix ω in $\mathbb{M}_n$. One has*

(a) *For any Hermitian linear functional χ on $\mathcal{A}$ satisfying $\chi(\mathbb{I}) = 0$ there exists a state ϕ on $\mathcal{A}$ such that $\chi = [\chi_\phi]_\omega$.*

(b) *The exponential map $[\chi_\phi]_\omega \mapsto \omega_\phi$ is well-defined;*

(c) *(geodesic completeness) The exponential map is a one-to-one map from the tangent plane $T_\omega\mathbb{M}_n$ onto the manifold $\mathbb{M}_n$.*

Proof

(a) By Riesz' theorem there exists a self-adjoint element X in $\mathcal{A}$ such that

$$\chi(A) = (A, X)_{\text{HS}} = \text{Tr}\, XA, \qquad A \in \mathcal{A}.$$

Let Y be the inverse Kubo transform of X. Let σ be defined by

$$\sigma = \frac{\exp(\log\rho + Y)}{\text{Tr}\, \exp(\log\rho + Y)}.$$

Then σ is a density matrix. Let ϕ be the state defined by σ. It satisfies

$$\begin{aligned} c_\rho(\sigma) &= \log\sigma - \log\rho + D(\rho||\sigma) \\ &= Y - \log\, \text{Tr}\, \exp(\log\rho + Y) + D(\rho||\sigma). \end{aligned}$$

The vector at ω tangent to the exponential arc connecting ϕ to ω is given by

$$[\chi_\phi]_\omega(A) = \text{Tr}\, c_\rho(\sigma) \left(\int_0^1 \mathrm{d}u\, \rho^{1-u} A \rho^u \right)$$

$$\begin{aligned}
&= \text{Tr } [Y - \log \text{ Tr } \exp(\log\rho + Y) + D(\rho||\sigma)] \left(\int_0^1 \mathrm{d}u\, \rho^{1-u} A \rho^u\right) \\
&= \text{Tr}\, XA + [\log \text{ Tr } \exp(\log\rho + Y) - D(\rho||\sigma)]\, \omega(A) \\
&= \chi(A).
\end{aligned}$$

The term proportional to $\omega(A)$ drops out because by assumption $\chi(\mathbb{I}) = 0$ while also $[\chi_\phi]_\omega(\mathbb{I}) = 0$.

(b,c) From part (a) of the theorem it follows that the range of the map $\phi \mapsto [\chi_\phi]_\omega$ is all of $T_\rho\mathbb{M}_n$. The previous proposition shows that the map is one-to-one. Hence its inverse, which is the exponential map, is well-defined on all of $T_\rho\mathbb{M}_n$ and has $\mathbb{M}_n$ as its range.

□

3.2.4 Bogoliubov's inner product

Introduce an inner product for the tangent space $T_\omega\mathbb{M}_n$ defined by

$$(\chi_\phi, \chi_\psi)_\omega = -\frac{\mathrm{d}}{\mathrm{d}s}\frac{\mathrm{d}}{\mathrm{d}t} D(\phi_s||\psi_t)\Big|_{s=t=0}.$$

As before ϕ_s and ψ_t are the exponential arcs connecting ϕ, ψ to ω.

One derivative gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} D(\phi_s||\psi_t)\Big|_{t=0} &= -\frac{\mathrm{d}}{\mathrm{d}t} \text{ Tr}\, \sigma_s \log\tau_t \Big|_{t=0} \\
&= -\frac{\mathrm{d}}{\mathrm{d}t} \text{ Tr}\, \sigma_s\, [\log\rho + t(\log\tau - \log\rho) - \alpha(\psi_t)] \Big|_{t=0} \\
&= -\text{ Tr}\, \sigma_s\, [\log\tau - \log\rho] + \frac{\mathrm{d}}{\mathrm{d}t}\alpha(\psi_t)\Big|_{t=0} \qquad (3.9)
\end{aligned}$$

The latter term does not depend on s. Hence it does not contribute when taking the derivative w.r.t. s.

The second derivative gives

$$\begin{aligned}
(\chi_\phi, \chi_\psi)_\omega &= \frac{\mathrm{d}}{\mathrm{d}s} \text{ Tr}\, \sigma_s\, [\log\tau - \log\rho] \Big|_{s=0} \\
&= \int_0^1 \mathrm{d}u \text{ Tr}\, \sigma_s^u \left[\log\sigma - \log\rho - \frac{\mathrm{d}}{\mathrm{d}s}\alpha(\phi_s)\right] \sigma_s^{(1-u)}\ [\log\tau - \log\rho] \Big|_{s=0} \\
&= \int_0^1 \mathrm{d}u \text{ Tr}\, \rho^u\, [\log\sigma - \log\rho]\, \rho^{1-u}\, [\log\tau - \log\rho]
\end{aligned}$$

$$-\frac{\mathrm{d}}{\mathrm{d}s}\alpha(\phi_s)\bigg|_{s=0} \text{Tr}\,\rho\,[\log\tau - \log\rho]\,.$$

Take $s = 0$ in (3.9) to find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\alpha(\psi_t)\bigg|_{t=0} &= \text{Tr}\,\rho\,[\log\tau - \log\rho] \\ &= -D(\rho||\tau).\end{aligned} \tag{3.10}$$

One obtains

$$\begin{aligned}(\chi_\phi, \chi_\psi)_\omega &= \int_0^1 \mathrm{d}u\ \text{Tr}\,\rho^u\,[A - \text{Tr}\,\rho A]\,\rho^{1-u}\,[B - \text{Tr}\,\rho B] \\ &= \int_0^1 \mathrm{d}u\ \text{Tr}\,\rho^{1-u}\,[A - \text{Tr}\,\rho A]\,\rho^u\,[B - \text{Tr}\,\rho B]\end{aligned} \tag{3.11}$$

with

$$A = \log\sigma - \log\rho \quad \text{and} \quad B = \log\tau - \log\rho.$$

This is the Bogoliubov inner product adapted to the present notations. See [16, 32, 66].

Proposition 9 *The inner product defined by (3.11) is symmetric, positive and non-degenerate.*

Proof

The symmetry $(\chi_\phi, \chi_\psi)_\omega = (\chi_\psi, \chi_\phi)_\omega$ follows by cyclic permutation under the trace and substitution of u by $1 - u$.

The positivity of $(\chi_\phi, \chi_\phi)_\omega$ follows from

$$(\chi_\phi, \chi_\phi)_\omega = \int_0^1 \mathrm{d}u\ \text{Tr}\ \left(\rho^{(1-u)/2}c_\rho(\sigma)\rho^{u/2}\right)^\dagger\ \left(\rho^{(1-u)/2}c_\rho(\sigma)\rho^{u/2}\right).$$

Finally, non-degeneracy is shown as follows. The assumption that $(\chi_\phi, \chi_\phi)_\omega = 0$ implies that $\rho^{(1-u)/2}c_\rho(\sigma)\rho^{u/2} = 0$ for all u in $[0, 1]$. This is only possible if $c_\rho(\sigma) = 0$. The latter implies that $\sigma = \rho$ and hence that $\phi = \omega$.

□

3.3 Coordinate representation

3.3.1 Affine coordinates

One can write for any state ϕ with density matrix σ

$$\log\sigma = x^i(\phi)B_i + \text{Tr}\ \log\sigma,$$

with $(B_i)_i$ be the basis introduced in Section 2.3.4. The basis vectors span the subspace of $\mathscr{H}_{\text{HS}}$ consisting of the hermitian matrices with vanishing trace. Hence $\log\sigma -$ Tr $\log\sigma$ can be expanded in this basis and one has

$$x^i(\phi) \quad = \quad (\log\sigma, B^i)_{\text{HS}}.$$

The map $\phi \mapsto x(\phi) \in \mathbb{R}^{n^2-1}$ is a global chart for the manifold $\mathbb{M}_n$.

From the definition (3.7) of the tangent functional $[\chi_\phi]_\omega \in T_\omega\mathbb{M}_n$ it follows that

$$[\chi_\phi]_\omega(A) \quad = \quad [x^i(\phi) - x^i(\omega)]\,[e_i^{(e)}]_\omega(A), \qquad A \in \mathcal{A}. \tag{3.12}$$

with

$$[e_i^{(e)}]_\omega(A) \quad = \quad \text{Tr}\,[B_i - \text{ Tr}\,\rho B_i]_\rho^{\text{K}}\,A, \qquad A \in \mathcal{A}. \tag{3.13}$$

The functionals $[e_i^{(e)}]_\omega$ are basis vectors for the tangent plane $T_\omega\mathbb{M}_n$. Note that the basis vectors depend on the state ω.

For further use, note also that the density matrix σ of the state ϕ can be written as

$$\sigma \quad = \quad \frac{e^{x^i(\phi)B_i}}{\text{Tr}\,e^{x^i(\phi)B_i}}$$

and that the basis vectors satisfy the symmetry relation

$$\begin{aligned}
[e_i^{(e)}]_\omega(B_j) \quad &= \quad \int_0^1 \mathrm{d}u\ \text{Tr}\,\rho^u\,[B_i - \omega(B_i)]\,\rho^{1-u}\,[B_j - \omega(B_j)] \\
&= \quad \int_0^1 \mathrm{d}u\ \text{Tr}\,\rho^{1-u}\,[B_j - \omega(B_j)]\,\rho^u\,[B_j - \omega(B_j)] \\
&= \quad [e_j^{(e)}]_\omega(B_i).
\end{aligned}$$

The following result shows that the $x^i(\phi) - x^i(\omega)$ are coordinates centered at the state ω.

Proposition 10 *For any tangent vector χ_ϕ is*

$$[\chi_\phi]_\omega(A) \quad = \quad [x^i(\phi) - x^i(\omega)]\,[e_i^{(e)}]_\omega(A) \tag{3.14}$$

Proof

From

$$c_\rho(\sigma) \quad = \quad \log\sigma - \log\rho + D(\rho||\sigma)$$

$$
\begin{aligned}
&= [x^i(\phi) - x^i(\omega)]B_i + \text{ Tr } \log\sigma - \text{ Tr } \log\rho + D(\rho||\sigma) \\
&= [x^i(\phi) - x^i(\omega)]\,[B_i - \text{ Tr } \rho B_i]
\end{aligned}
\tag{3.15}
$$

one obtains for all A in $\mathcal{B}(\mathbb{C}^n)$

$$
\begin{aligned}
\text{Tr } [c_\rho(\sigma)]^{\kappa}_{\rho} A &= [x^i(\phi) - x^i(\omega)] \text{ Tr } [B_i - \text{ Tr } \rho B_i]^{\kappa}_{\rho} A \\
&= [x^i(\phi) - x^i(\omega)]\, [e^{(e)}_i]_\omega(A)
\end{aligned}
\tag{3.16}
$$

□

3.3.2 The metric tensor

The metric tensor of the Bogoliubov metric is given by

$$
g_{ij}(\omega) = (e^{(e)}_i, e^{(e)}_j)_\omega
$$

From the previous proposition it follows that the inner product (3.11) can be written as

$$
(\chi_\phi, \chi_\psi)_\omega \quad = \quad [x^i(\phi) - x^i(\omega)]\, g_{ij}(\omega)[x^j(\phi) - x^j(\omega)].
$$

The basis vectors $e^{(e)}_i$ are tangent vectors $[e^{(e)}_i]_\omega = [\chi_{\phi_i}]_\omega$ labeled with states ϕ_i whose density matrices σ_i depend on ω and are given by

$$
\sigma_i = \frac{e^{\log\rho + B_i}}{\text{Tr } e^{\log\rho + B_i}},
\tag{3.17}
$$

with ρ the density matrix of ω. The coordinates of these states ϕ_i are given by

$$
\begin{aligned}
x^i(\phi_j) &= (\log\sigma_j, B_i)_{\text{HS}} \\
&= x^i(\omega) + g^i_j.
\end{aligned}
$$

This implies that $\chi_{\phi_j} = [e^{(e)}_j]_\omega$. Use that $c_\rho(\sigma_j) = B_j - \text{ Tr } \rho B_j$ and (3.11) to calculate the metric tensor

$$
\begin{aligned}
g_{ij}(\omega) &= (e^{(e)}_i, e^{(e)}_j)_\omega \\
&= (\chi_{\phi_i}, \chi_{\phi_j}) \\
&= \int_0^1 \mathrm{d}u \text{ Tr } \rho^{1-u}\,[B_i - \omega(B_i)]\,\rho^u\,[B_j - \omega(B_j)] \\
&= [e^{(e)}_i]_\omega(B_j - \omega(B_j)) \\
&= [e^{(e)}_j]_\omega(B_i - \omega(B_i)).
\end{aligned}
\tag{3.18}
$$

3.4 An alternative approach

3.4.1 The GNS construction

Every state ω on a C^*-algebra $\mathcal{A}$ can be represented as a C^*-algebra of bounded operators on a Hilbert space $\mathscr{H}$ in such a way that the state ω becomes a vector state. This means that there exists a normalized vector Ω in $\mathscr{H}$ such that $\omega(A) = (A\Omega, \Omega)$ for all A in $\mathcal{A}$. The vector Ω may be assumed to be cyclic for $\mathcal{A}$, which means that $\mathcal{A}\Omega$ is dense in $\mathscr{H}$. This is the essence of the Gelfand-Naimark-Segal (GNS) construction.. The GNS-representation is unique up to unitary equivalence. See one of the references [8, 14, 22] for mathematical details.

The GNS-construction is general. The construction is rather abstract. In the present case the algebra $\mathcal{A}$ is the finite-dimensional algebra $\mathcal{B}(\mathbb{C}^n)$ of n-by-n matrices with complex entries. In this case the construction can be made more concrete using the Hilbert space $\mathbb{C}^n \otimes \mathbb{C}^n$ in the same way as in Chapter 2.

The construction starts like in Lemma 1 of Chapter 2 by selecting an orthonormal basis $(f_i)_i$ diagonalizing the density matrix ρ of the given state ω in $\mathbb{M}_n$. Let $\rho f_i = p_i f_i$ and choose $\Omega = \sum_i \sqrt{p_i}\, f_i \otimes f_i$. Then one has

$$\begin{aligned}(A \otimes \mathbb{I}\,\Omega, \Omega) &= \sum_{i,j} \sqrt{p_i p_j}(A \otimes \mathbb{I}\, f_i \otimes f_i, f_j \otimes f_j)\\ &= \sum_i p_i (A f_i, f_i) = \text{ Tr } \rho A = \omega(A)\end{aligned} \tag{3.19}$$

This shows that the state ω becomes a vector state.

Let us now show that the vector Ω is cyclic for $\mathcal{B}(\mathbb{C}^n)$. This then leads to the conclusion that the representation

Lemma 5 *The set* $\{A \otimes \mathbb{I}\,\Omega :\ A \in \mathcal{A}\}$ *is equal to the Hilbert space* $\mathscr{H} = \mathbb{C}^n \otimes \mathbb{C}^n$.

Note thatthe definition of the tensor product $\mathbb{C}^n \otimes \mathbb{C}^n$ is as in Chapter 2, linear in the former argument and anti-linear in the latter.

Proof

Let be given an arbitary Ψ in $\mathscr{H}$. It can be expanded as

$$\Psi = \sum_{k,i} \lambda_{k,i} f_k \otimes f_i.$$

Let us try to construct a matrix A such that $A \otimes \mathbb{I}\Omega = \Psi$.

One calculates

$$A \otimes \mathbb{I} \sum_i \sqrt{p_i} f_i \otimes f_i \quad = \quad \sum_{k,i} \lambda_{k,i} f_k \otimes f_i.$$

This simplifies to

$$\sqrt{p_i} A f_i \quad = \quad \sum_k \lambda_{k,i} f_k.$$

The choice $A_{k,i} = \lambda_{k,j}/\sqrt{p_i}$ solves the latter condition.

□

3.4.2 Tomita-Takesaki theory

The Tomita-Takesaki theory is a general theory about the duality between a von Neumann algebra $\mathcal{A}$ and its commutant $\mathcal{A}'$. The setting of the theory is that of σ-finite von Neumann algebras, which is a non-commuttaive generalization of the notion of measure spaces. See the introduction of Section 2.5 of [22] for a discussion of this point.

In the present case of the algebra $\mathcal{A} = \mathcal{B}(\mathbb{C}^n)$ and any faithful state $\omega \in \mathbb{M}_n$ the Tomita-Takesaki theory can be developed in a simplified manner.

Consider the GNS-representation as constructed in the previous section. Define an anti-linear operator S by $S(A \otimes \mathbb{I}\,\Omega) = A^* \otimes \mathbb{I}\,\Omega$ for all A in $\mathcal{B}(\mathbb{C}^n)$. It is well defined because $(A \otimes \mathbb{I}\,\Omega) = 0$ implies $A = 0$. Indeed, one has

$$A \otimes \mathbb{I}\,\Omega \quad = \quad \sum_i \sqrt{p_i} A f_i \otimes f_i.$$

Because $\sqrt{p_i} \neq 0$ and because the f_i form an orthonormal basis this expression can only vanish if $Af_i = 0$ vanishes for all i. The latter implies $A = 0$.

The adjoint S^* of the operator S is denoted F. The use of 'S' and 'F' for these operators is tradition. Note that, because of the anti-linearity one has $(Sx, y) = (Fy, x)$ for all x, y in $\mathbb{C}^n$. The action of the operator F is given by the following result.

Proposition 11 $F(\mathbb{I} \otimes A\,\Omega) = \mathbb{I} \otimes A^*\,\Omega$

Proof

Calculate

$$(F(\mathbb{I} \otimes A\,\Omega), B \otimes \mathbb{I}\,\Omega) \quad = \quad (S(B \otimes \mathbb{I}\,\Omega), \mathbb{I} \otimes A\,\Omega)$$

$$\begin{aligned} &= (B^* \otimes \mathbb{I}\,\Omega, \mathbb{I} \otimes A\,\Omega) \\ &= (B^* \otimes A^*\,\Omega,\, \Omega) \\ &= (\mathbb{I} \otimes A^*\,\Omega,\, B \otimes \mathbb{I}\,\Omega)\,. \end{aligned}$$

Because Ω is cyclic for the representation any vector in $\mathbb{C}^n \otimes \mathbb{C}^n$ is of the form $B \otimes \mathbb{I}\,\Omega$. Hence, it follows that $F(\mathbb{I} \otimes A\,\Omega) = \mathbb{I} \otimes A^*\,\Omega$.

□

Definition 4 *The modular operator is defined by $\Delta = FS = S^*S$. The modular conjugation operator J is defined by the polar decomposition $S = J\Delta^{1/2}$ of the operator S. It is given by $J = S\Delta^{-1/2}$.*

Note that J, like S, is an anti-linear operator.

Proposition 12 *In the present context the modular operator is given by $\Delta = \rho \otimes \rho^{-1}$.*

Proof

First calculate for any A and B in $\mathcal{B}(\mathbb{C}^n)$

$$\begin{aligned} ((\rho \otimes \rho^{-1})\, A \otimes \mathbb{I}\,\Omega, B \otimes \mathbb{I}\,\Omega) &= (B^*\rho A \otimes \rho^{-1}\Omega, \Omega) \\ &= \sum_i p_i^{1/2} (B^*\rho A \otimes \rho^{-1}\, f_i \otimes f_i, \Omega) \\ &= \sum_i p_i^{-1/2} ((B^*\rho A\, f_i) \otimes f_i, \Omega) \\ &= \sum_i (B^*\rho A\, f_i, f_i) \\ &= \text{Tr}\,\rho A B^* = \omega(AB^*). \end{aligned} \tag{3.20}$$

On the other hand is

$$\begin{aligned} (\Delta\, A \otimes \mathbb{I}\,\Omega, B \otimes \mathbb{I}\,\Omega) &= (S^*S\, A \otimes \mathbb{I}\,\Omega, B \otimes \mathbb{I}\,\Omega) \\ &= (B^* \otimes \mathbb{I}\,\Omega, A^* \otimes \mathbb{I}\,\Omega) \\ &= \sum_i p_i\, (AB^* f_i, f_i) \\ &= \text{Tr}\,\rho A B^* = \omega(AB^*). \end{aligned} \tag{3.21}$$

Use the previous result to obtain

$$(\Delta\, A \otimes \mathbb{I}\,\Omega, B \otimes \mathbb{I}\,\Omega) = ((\rho \otimes \rho^{-1})\, A \otimes \mathbb{I}\,\Omega, B \otimes \mathbb{I}\,\Omega).$$

Because Ω is cyclic for the representation it follows that $\Delta = \rho \otimes \rho^{-1}$.

□

Proposition 13 *The modular conjugation operator J satisfies $J = J^*$, $\Delta^{1/2}J = J\Delta^{-1/2}$ and $J^2 = 1$.*

Proof

It is clear that $S^2 = \mathbb{I}$. This implies that S is invertible and that $S^{-1} = S$. From the polar decomposition $S = J\Delta^{1/2}$ one obtains

$$J\Delta^{1/2} = S = S^{-1} = \Delta^{-1/2}J^*.$$

Multiply from the left with J to obtain

$$J^2\Delta^{1/2} \quad = \quad J\Delta^{-1/2}J^*.$$

The r.h.s. is a positive operator. The l.h.s. is a polar decomposition because J^2 is an isometry. From the uniqueness of the polar decomposition it then follows that $J^2 = \mathbb{I}$.

From $\Delta^{1/2} = J\Delta^{-1/2}J^*$, by multiplication with J from the right one obtains $\Delta^{1/2}J = J\Delta^{-1/2}$.

Finally, $J^*J = \mathbb{I}$ and $J^2 = \mathbb{I}$ imply that $J = J^{-1}$ and $J^* = J^{-1} = J$

□

Proposition 14 $S\Omega = \Delta\Omega = J\Omega = \Omega$

Proof

$S\Omega = \Omega$ follows from the definition of S.

Use the explicit forms of Δ and Ω to obtain

$$\begin{aligned}\Delta\Omega &= \rho\otimes\rho^{-1}\sum_i \sqrt{p_i}\, f_i\otimes f_i \\ &= \sum_i \sqrt{p_i}\,(\rho, f_i)\otimes(\rho^{-1}f_i) \\ &= \Omega.\end{aligned}$$

Finally, note that $J\Omega = S\Delta^{-1}\Omega = \Omega$ follows from the two other results.

□

The following result shows that the isometry J maps the algebra $\mathcal{B}(\mathbb{C}^n)\otimes\mathbb{I}$ onto its commutant $\mathbb{I}\otimes\mathcal{B}(\mathbb{C}^n)$.

Theorem 4 *For any A in $\mathcal{B}(\mathbb{C}^n)\otimes\mathbb{I}$ there exists X in $\mathcal{B}(\mathbb{C}^n)\otimes\mathbb{I}$ such that $J(A\otimes\mathbb{I})J=\mathbb{I}\otimes X$.*

Proof

For any A, B in $\mathcal{B}(\mathbb{C}^n)$ is

$$\begin{aligned}
J(A\otimes\mathbb{I})J(B\otimes\mathbb{I})\Omega &= J(A\otimes\mathbb{I})JS(B^*\otimes\mathbb{I})\Omega \\
&= J(A\otimes\mathbb{I})(\rho^{1/2}\otimes\rho^{-1/2})(B^*\otimes\mathbb{I})\Omega \\
&= J(A\rho^{1/2}B^*\otimes\rho^{-1/2})\Omega \\
&= J(A\rho^{1/2}B^*\rho^{-1/2}\otimes\mathbb{I})\Omega \\
&= JS(\rho^{-1/2}B\rho^{1/2}A^*\otimes\mathbb{I})\Omega \\
&= (B\rho^{1/2}A^*\otimes\rho^{-1/2})\Omega \\
&= (B\otimes\mathbb{I})\,(\rho^{1/2}A^*\otimes\rho^{-1/2})\Omega. \qquad (3.22)
\end{aligned}$$

In the case $B=\mathbb{I}$ this gives

$$J(A\otimes\mathbb{I})J\Omega = (\rho^{1/2}A^*\otimes\rho^{-1/2})\Omega.$$

Hence, (3.22) becomes

$$J(A\otimes\mathbb{I})J(B\otimes\mathbb{I})\Omega = (B\otimes\mathbb{I})\,J(A\otimes\mathbb{I})J\Omega.$$

Take now an additional C in $\mathcal{B}(\mathbb{C}^n)$ to obtain

$$\begin{aligned}
J(A\otimes\mathbb{I})J(BC\otimes\mathbb{I})\Omega &= (BC\otimes\mathbb{I})\,J(A\otimes\mathbb{I})J\Omega \\
&= (B\otimes\mathbb{I})\,J(A\otimes\mathbb{I})J\,(C\otimes\mathbb{I})\Omega.
\end{aligned}$$

Because Ω is cyclic for the representation it follows that $J(A\otimes\mathbb{I})JB=B\,J(A\otimes\mathbb{I})J$ for all B. This shows that $J(A\otimes\mathbb{I})J$ belongs to the commutant of $\mathcal{B}(\mathbb{C}^n)\otimes\mathbb{I}$. Hence, it is of the form $\mathbb{I}\otimes X$.

$\square$

3.4.3 Generalized Radon-Nikodym derivatives

The non-commutative generalization of the Radon-Nikodym derivative is non-unique. Araki [15] introduced a family of Radon-Nikodym derivatives, a single member of which is presented here.

The operator X in the following Proposition can be considered to be a Radon-Nikodym derivative of the state $\phi\in\mathbb{M}_n$ w.r.t. the given state ω.

Theorem 5 *For each state ϕ in $\mathbb{M}_n$ there exists a unique operator X in $\mathcal{B}(\mathbb{C}^n)$ such that*

$$\phi(A) \quad = \quad (A \otimes X\Omega, \Omega), \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

The operator satisfies $X > 0$.

Note that $X = \mathbb{I}$ corresponds with $\phi = \omega$.

Proof

Calculate

$$\begin{array}{rcl} S(\rho^{-1}\sigma \otimes \mathbb{I})S & = & J\Delta^{1/2}(\rho^{-1}\sigma \otimes \mathbb{I})\Delta^{-1/2}J \\ & = & J(\rho^{-1/2}\sigma\rho^{-1/2} \otimes \mathbb{I})J \end{array}$$

By Theorem 5 there exists X such that

$$S(\rho^{-1}\sigma \otimes \mathbb{I}) = \mathbb{I} \otimes X^*.$$

Let us now verify that this operator X has the required properties.

Calculate

$$\begin{array}{rcl} (A \otimes X\Omega, \Omega) & = & (A \otimes \mathbb{I}\Omega, \mathbb{I} \otimes X^*\Omega) \\ & = & (A \otimes \mathbb{I}\Omega, S(\rho^{-1}\sigma \otimes \mathbb{I})S\Omega) \\ & = & (A \otimes \mathbb{I}\Omega, \sigma\rho^{-1} \otimes \mathbb{I}\Omega) \\ & = & (\rho^{-1}\sigma A \otimes \mathbb{I}\Omega, \Omega) \\ & = & \omega(\rho^{-1}\sigma A) \\ & = & \mathrm{Tr}\,\rho(\rho^{-1}\sigma A) = \phi(A). \end{array} \tag{3.23}$$

This verifies that X is a Radon-Nikodym derivative of ϕ w.r.t. ω.

Uniqueness Assume both X and Y are Radon-Nikodym derivatives of ϕ w.r.t. ω. Then one has for all A that $(A \otimes \mathbb{I}\Omega, \mathbb{I} \otimes (X - Y)\Omega) = 0$. Because Ω is cyclic for the representation it follows that $\mathbb{I} \otimes (X - Y)\Omega) = 0$. The latter implies that $\mathbb{I} \otimes (X - Y)\Omega = 0$.

Now $\mathbb{I} \otimes (X - Y) = 0$ follows because for all A one has

$$0 = [A \otimes \mathbb{I}]\,[\mathbb{I} \otimes (X - Y)\Omega] = [\mathbb{I} \otimes (X - Y)]\,[A \otimes \mathbb{I}\Omega]$$

and Ω is cyclic for the representation.

Self-adjointness: $\phi(A) = \overline{\phi(A^*)} = (\Omega, A^* \otimes X\Omega)$

$\Rightarrow \quad \phi(A) = (A \otimes X^*\Omega, \Omega) \quad$ plus uniqueness $\Rightarrow X = X^*$

Positivity: $0 \leq \phi(A^*A) = ([\mathbb{I} \otimes X]\,[A \otimes \mathbb{I}]\Omega, [A \otimes \mathbb{I}]\Omega)$

Strict Positivity: Assume $\mathbb{I} \otimes X\,\Psi = 0$

Then $\exists B$ such that $\Psi = B \otimes \mathbb{I}\,\Omega$

$\phi(B^*B) = (B^*B \otimes X\,\Omega, \Omega) = (\mathbb{I} \otimes X\Psi, \Psi) = 0$

ϕ is faithful $\quad \Rightarrow \quad B = 0 \quad \Rightarrow \quad \Psi = 0$

□

The above theorem is the basis for giving an alternative definition of exponential arcs. Given two states ω and ϕ an exponential arc $t \mapsto \phi_t$ connecting ϕ to ω is defined by

$$\phi_t(A) = \frac{(A \otimes X^t\Omega, \Omega)}{(\mathbb{I} \otimes X^t\Omega, \Omega)}, \qquad A \in \mathcal{B}(\mathbb{M}_n).$$

The tangent vector to this arc at $t = 0$ is given by

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi_t(A) \quad = \quad (A \otimes [\log X - \beta'(0)]\Omega, \Omega)$$

with normalization given by $\beta'(0) = (\mathbb{I} \otimes \log X\,\Omega, \Omega)$.

The following example shows that the geometry in which these exponential arcs are geodesics is really different from the one introduced in the first part of this chapter.

3.4.4 Example

Use again the Pauli matrices introduced in Section 2.4.2. Let

$$\rho = \frac{1}{2}(\mathbb{I} + \epsilon\sigma_3) \quad \text{ and } \sigma = \frac{1}{2}(\mathbb{I} + \delta\sigma_1).$$

Verify the following

$$\omega(\sigma_3) = \epsilon,\ \phi(\sigma_1) = \delta,\ \omega(\sigma_1) = \omega(\sigma_2) = \phi(\sigma_2) = \phi(\sigma_3) = 0.$$

After some calculation one finds that

$$X = \frac{1}{1-\epsilon^2}\left(1 - \epsilon\sigma_3 + \delta\sqrt{1-\epsilon^2}\,\sigma_1\right).$$

The tangent vector evaluated for the observable σ_1 is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \phi_t(\sigma_1) = (\sigma_1 \otimes [\log X - \beta'(0)]\Omega, \Omega) = \delta(1-\epsilon^2)\frac{x}{\tanh x} \tag{3.24}$$

with $\tanh^2 x = \epsilon^2 + (1-\epsilon^2)\delta^2$. Compare this result with

$$[\chi_\phi]_\omega)(\sigma_1) = \text{ Tr } c_\rho(\sigma)[\sigma_1]^{\mathrm{K}}_\rho = \zeta \text{ Tr } c_\rho(\sigma)\sigma_1$$

with

$$\zeta = \frac{\epsilon}{\log(1+\epsilon) - \log(1-\epsilon)}.$$

Here, $[\sigma_1]^{\mathrm{K}}_\rho = \zeta\sigma_1$ is being used.

Recall that $c_\rho(\sigma) = \log\sigma - \log\rho + D(\rho||\sigma)$. This gives $[\chi_\phi]_\omega(\sigma_1) = 2\zeta v$ with $\tanh v = \delta$, because Tr $c_\rho(\sigma)\sigma_1 =$ Tr $\sigma_1 \log\sigma_1 = 2v$.

In general is $2\zeta v \neq \delta(1-\epsilon^2)x/tanhx$. Hence one reaches the conclusion that the definition of an exponential family gives different results for the two definitions of exponential arcs.

3.5 Appendix

Klein's inequality See [12], Section 2.5.2, or [23], Section 2.1.7.. The version below is taken from [59], Section 11.7.

Lemma 6 *Let A, B and C be self-adjoint operators with discrete spectrum. Assume that $C \geq 0$ and $BC = CB$. Let $f(x)$ be a convex function. Then one has*

$$\mathit{Tr}\, C[f(A) - f(B) - (A-B)f'(B)] \geq 0. \tag{3.25}$$

Proof

Let $(\phi_n)_n$ be an orthonormal basis in which A is diagonal. Let $(\psi_m)_m$ be an orthonormal basis in which B and C are simultaneously diagonal. Let $A\phi_n = a_n\phi_n$, $B\psi_m = b_m\psi_m$, and $C\psi_m = c_m\psi_m$. Denote $\lambda_{nm} = \langle\phi_m|\psi_n\rangle$. Then the convexity of $f(x)$ implies that

$$\begin{aligned} & \langle\phi_m|C(f(A) - f(B) - (A-B)f'(B))|\phi_m\rangle \\ = & \sum_n c_n|\lambda_{mn}|^2[f(a_m) - f(b_n) - (a_m - b_n)f'(b_n)] \end{aligned}$$

$$\geq \quad 0. \tag{3.26}$$

To see the inequality, use that a tangent line to a convex function always lies below the function.

Klein's inequality now follows by summing over m.

$\square$

Apply the inequality with $f(x) = x \log x$, $x > 0$, and $C = 1$, $A = \rho$ and $B = \sigma$. This gives

$$D(\rho||\sigma) = \text{ Tr } \rho(\log \rho - \log \sigma) \geq 0.$$

Alternatively, choose $f(x) = -\log x$, $A = \sigma$ and $B = \rho$.

The inequality can be improved.

Lemma 7 *Let $f(x) = x \log x$. One has*

$$f(x) - f(y) - (x - y)f'(y) \geq \frac{1}{2}(x - y)^2 \quad \textit{if } 0 < x \leq 1 \textit{ and } 0 < y \leq 1.$$

Proof

Let

$$g_y(x) = f(x) - f(y) - (x - y)f'(y) - \frac{1}{2}(x - y)^2.$$

One has

$$\begin{aligned} g_y(x) &= x(\log x - \log y) + x - y - \frac{1}{2}(x - y)^2, \\ g_y'(x) &= \log x - \log y + y - x, \\ g_y''(x) &= \frac{1 - x}{x}. \end{aligned}$$

From $g_y''(x) \geq 0$ it follows that the function g_y is convex. From $g_y(y) = g_y'(y) = 0$ it follows that $g_y(x) \geq 0$.

$\square$

The eigenvalues of density matrices ρ and σ lie in the interval $(0, 1]$. The result is then

$$D(\rho||\sigma) = \text{ Tr } \rho(\log \rho - \log \sigma) \geq \frac{1}{2} \text{ Tr } (\rho - \sigma)^2.$$

Klein's inequality can be generalized. See Theorem 11.10 of Petz [57].

Proof of the identity (3.3) The following argument is taken from [45], p. 156.

Positive-definite matrices P and Q satisfy the identity

$$1 - Q^v P^{-v} \quad = \quad \int_0^v \mathrm{d}u\, Q^u (\log P - \log Q) P^{-u}. \tag{3.27}$$

Indeed, the identity is trivially satisfied for $v = 0$ and the derivative of the l.h.s. equals

$$= \quad -(\log Q) Q^v P^{-1} + Q^v P^{-1} \log P = Q^v [-\log Q + \log P] P^{-1}.$$

This is the integrand of the r.h.s. of (3.27) evaluated at $u = v$.

Next, multiply (3.27) with P^v from the right to obtain

$$P^v - Q^v \quad = \quad \int_0^v \mathrm{d}u\, Q^u (\log P - \log Q) P^{v-u}.$$

Take $v = 1$, $P = \exp(H + tA)$ and $Q = \exp(H)$. This gives

$$\exp(H + tA) - \exp(H) \quad = \quad t \int_0^1 \mathrm{d}u\, e^{uH}\, A\, e^{(1-u)(H+tA)}.$$

Divide by t and take the limit $t = 0$ to obtain (3.3).

Chapter 4

The dually flat geometry

The present chapter continues with the study of the manifold $\mathbb{M}_n$ of faithful states on the von Neumann algebra $\mathcal{B}(\mathbb{C}^n)$. The manifold is equipped with the Bogoliubov metric.

4.1 A flat geometry

4.1.1 Parallel transport

For a full description of the geometry of the manifold $\mathbb{M}_n$ one still needs to specify a connection. The choice is not unique. The intention is to find a connection the geodesics of which are the exponential arcs defined in Section 3.2.2.

A connection determines how to transport tangent vectors along a smooth curve. Conversely, given transport mappings

$$\Pi(\omega_1 \mapsto \omega_2) : T_{\omega_1}\mathbb{M}_n \mapsto T_{\omega_2}\mathbb{M}_n$$

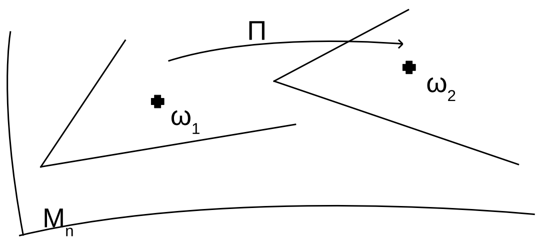

Figure 4.1: Parallel transport Π from $T_{\omega_1}\mathbb{M}_n$ to $T_{\omega_2}\mathbb{M}_n$

it is possible to reconstruct the connection [4]. It is the latter approach which is followed here.

4.1.2 Dual geometries

Every connection ∇ has a dual connection ∇^* w.r.t. the metric of the manifold. With the parallel transport Π corresponds a dual parallel transport Π^* The defining relation is

$$\left(\Pi(\omega_1 \mapsto \omega_2)V, \Pi^*(\omega_1 \mapsto \omega_2)W\right)_{\omega_2} = \left(V, W\right)_{\omega_1}.$$

In this expression V and W are vector fields.

Take for instance constant vector fields $V = \chi_\psi$ and $W = \chi_\phi$. Then

$$\begin{aligned} [\chi_\phi]_\omega(A) &= \text{Tr}\, [c_\rho(\sigma)]_\rho^{\mathrm{K}} A \\ [\chi_\psi]_\omega(A) &= \text{Tr}\, [c_\rho(\tau)]_\rho^{\mathrm{K}} A \\ (V, W)_\omega &= \text{Tr}\, [c_\rho(\tau)]_\rho^{\mathrm{K}} c_\rho(\sigma) \end{aligned} \tag{4.1}$$

Then the quantity

$$(V, W)_\omega = \text{Tr}\, [c_\rho(\tau)]_\rho^{\mathrm{K}} c_\rho(\sigma) \tag{4.2}$$

must remain constant when one vector is transported using Π and the other using Π^*.

A known property of the Levi-Civita connection, also called the metric connection, is that it is the only connection which preserves the metric: length of tangent vectors and angles between them are conserved under parallel transport. This means $\Pi^* = \Pi$. For all other connections is $\Pi^* \neq \Pi$.

The *m-connection* is the connection in which convex combinations of states are geodesics

$$t \mapsto (1-t)\omega + t\phi, \qquad t \in [0,1], \omega, \phi \in \mathbb{M}_n$$

The *e-connection* is the connection in which any exponential arc $t \mapsto \phi_t$ connecting a state ϕ to a state ω, $\phi, \omega \in \mathbb{M}_n$, is a geodesic.

In the next Section it is shown that the e-connection is the dual of the m-connection w.r.t. the Bogoliubov metric.

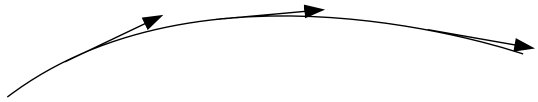

Figure 4.2: Example of parallel transport with $\Pi \neq \text{ Id}$

4.1.3 Theorem

First consider parallel transport of the m-connection.

Proposition 15 *The parallel transport map of the m-connection is the identity map* $\Pi = $ *Id.*

Proof

A smooth path γ is a geodesic when the vector field $\dot{\gamma}$ is invariant under parallel transport. This is

$$\Pi(\gamma_s \mapsto \gamma_t)[\dot{\gamma}]_{\gamma_s} = [\dot{\gamma}]_{\gamma_t}.$$

Consider the path γ given by $\gamma_t = (1-t)\omega + t\phi$, i.e. a geodesic of the m-connection. Its derivative is the vector field $\dot{\gamma} = \phi - \omega$. It is independent of t. Hence, it is a geodesic for the parallel transport $\Pi = \text{ Id}$.

□

Theorem 6 *Consider the manifold* $\mathbb{M}_n$ *of faithful states on the von Neumann algebra* $\mathcal{B}(\mathbb{C}^n)$*. Parallel transport of the e-connection is the dual* Π^* *of the parallel transport* $\Pi = $ *Id provided that the manifold is equipped with the Bogoliubov metric.*

Proof

Consider the exponential arc $t \mapsto \phi_t$ connecting a state ϕ to the state ω. Its density matrix equals

$$\sigma_t \& = \quad \exp(\log\rho + t(\log\sigma - \log\rho) - \alpha(t)). \tag{4.3}$$

Take the derivative to find

$$\begin{aligned}
\dot{\phi}_t(A) &= \frac{\mathrm{d}}{\mathrm{d}t}\phi_t(A) \\
&= \int_0^1 \mathrm{d}u \ \text{Tr}\, \sigma_t^u[\log\sigma - \log\rho - \ \text{Tr}\, \sigma_t(\log\sigma - \log\rho)]\sigma_t^{1-u} A \\
&= \text{Tr}\, [\log\sigma - \log\rho - \ \text{Tr}\, \sigma_t(\log\sigma - \log\rho)]^{\mathrm{K}}_{\sigma_t} A
\end{aligned} \tag{4.4}$$

In order to show that the exponential arc is a geodesic for the dual connection one should prove that $\Pi^*(\phi_s \mapsto \phi_t)\,\dot{\phi}_s = \dot{\phi}_t$. To do so the following result is needed.

Proposition 16 *The dual parallel transport map* $\Pi^*(\omega_1 \mapsto \omega_2)$ *maps* $\chi \in T_{\omega_1}\mathbb{M}_n$ *onto* $\xi \in T_{\omega_2}\mathbb{M}_n$ *with* $\chi(A) = \textit{Tr}\,[Y]^{\kappa}_{\rho_1} A$ *and* $\xi(A) = \textit{Tr}\,[X]^{\kappa}_{\rho_2} A$ *related by* $X = Y - \textit{Tr}\,\rho_2 Y$.

Proof

Take ζ of the form $\zeta(A) = \text{Tr}\,VA$ with $V = V^*$ and $\text{Tr}\,V = 0$. Apply the inverse Kubo transform: to find

$$\chi(A) = \text{Tr}\,[Y]^{\kappa}_{\rho_1} A \text{ and } \xi(A) = \text{Tr}\,[X]^{\kappa}_{\rho_2} A.$$

Because Π is the identity map the defining relation for Π^* simplifies to

$$(\zeta, \xi)_{\omega_2} = (\zeta, \Pi^*(\omega_1 \mapsto \omega_2)\chi)_{\omega_2} = (\zeta, \chi)_{\omega_1}.$$

From $\zeta(A) = \text{Tr}\,VA$ and $\xi(A) = \text{Tr}\,[X]^{\kappa}_{\rho_2} A$ it follows that $(\zeta, \xi)_{\omega_2} = TrVX$. Similarly is $(\zeta, \xi)_{\omega_2} = TrVX$. Hence, the equality $(\zeta, \xi)_{\omega_2} = (\zeta, \chi)_{\omega_1}$ implies that for all V one has $\text{Tr}\,VX = \text{Tr}\,VY$. Because V is arbitrary but traceless this implies that $X - Y$ is a multiple of the identity. The ω_2-expectation of X vanishes. Hence one obtains $X = Y - \text{Tr}\,\rho_2 Y$.

□

Continuation of the proof of the theorem Apply the Proposition with

$$\begin{aligned}\omega_1 = \phi_s, \chi = \dot{\phi}_s \text{ and } Y = \log\sigma - \log\rho - \text{Tr}\,\sigma_s(\log\sigma - \log\rho),\\ \omega_2 = \phi_t, \xi = \dot{\phi}_t \text{ and } X = \log\sigma - \log\rho - \text{Tr}\,\sigma_t(\log\sigma - \log\rho).\end{aligned}$$

This gives $\Pi^*(\phi_s \mapsto \phi_t)\,\dot{\phi}_s = \dot{\phi}_t$ because $X = Y - \text{Tr}\,\sigma_t Y$ is satisfied.

□

4.1.4 Coordinate representation

Let $(e_i)_i$ be the basis vectors of the tangent bundle, as given by (3.13). The following proposition shows that they are invariant under the dual parallel transport Π^*.

Proposition 17 *One has* $\Pi^*(\omega_1 \mapsto \omega_2)[e_i]_{\omega_1} = [e_i]_{\omega_2}$.

Proof

Apply the previous proposition with $Y = B_i - \text{ Tr } \rho_1 B_i$. This yields $X = B_i - \text{ Tr } \rho_2 B_i$ and hence

$$\xi(A) = \text{ Tr } [X]^{\kappa}_{\rho_2} A = [e_i]_{\omega_2}(A).$$

□

4.1.5 Covariant derivatives

The covariant derivative of a vector field V along a smooth curve γ is given by [4]

$$[\nabla^*_{\dot\gamma} V]_{\gamma_t} = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \Pi^*(\gamma_{t+s} \mapsto \gamma_t)\, V(\gamma_{t+s}).$$

Apply this expression to the basis vectors e_i

$$\begin{aligned} [\nabla^*_{\dot\gamma} e_i]_{\gamma_t} &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \Pi^*(\gamma_{t+s} \mapsto \gamma_t)\, [e_i]_{\gamma_{t+s}} \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} [e_i]_{\gamma_t} \\ &= 0. \end{aligned} \tag{4.5}$$

By definition the connection coefficients Γ^{*k}_{ij} are given by

$$\nabla^*_{e_i} e_j = \Gamma^{*k}_{ij} e_k.$$

One concludes that they vanish.

4.1.6 Flatness

A connection ∇ is flat if there exists an affine coordinate system, this is a coordinate system in which all connection coefficients Γ^k_{ij} vanish. Hence, the conclusion of the previous section is that $x^i(\phi) - x^i(\omega)$ are affine coordinates. Hence, the e-connection is flat.

This result is expected on the basis of the following result due to Chentsov: If a connection is flat then its dual is flat as well. For a proof see [45], Theorem 3.3 on p. 53.

4.2 The Legendre structure

4.2.1 A Hessian geometry

The manifold $\mathbb{M}_n$ equipped with the Bogoliubov metric and the e-connection has a Hessian geometry. There exist affine coordinates $x^i(\omega)$ and a potential $\Phi(\omega)$ such that the metric tensor $g_{ij}(\omega)$ is the Hessian of this potential $\Phi(\omega)$. This is proved in the following Theorem.

Theorem 7 *Let $(B_i)_i$ be the orthonormal basis in Hilbert-Schmidt space, $x^i(\omega)$ be corresponding coordinates and $g_{ij}(\omega)$ the metric tensor all defined in Section 3.3.2. There exists a potential $\Phi(x)$ such that*

i) $\partial_i\,\Phi(x(\omega)) = \omega(B_i)$

ii) $\partial_i\partial_j\,\Phi(x(\omega)) = g_{i,j}(\omega)$

Proof

Let

$$\Phi(x) = \log\ \mathrm{Tr}\ \exp(x^i B_i), \qquad x \in \mathbb{R}^{n^2-1}.$$

Take a derivative w.r.t. x^i to find

$$\begin{aligned} \partial_i\Phi(x) &= \frac{1}{\mathrm{Tr}\ \exp(x^i B_i)}\partial_i\ \mathrm{Tr}\ e^{x^i B_i} \\ &= \omega(B_i). \end{aligned}$$

Take a further derivative to obtain

$$\begin{aligned} \partial_i\partial_j\,\Phi(x(\omega)) &= \partial_j\ \mathrm{Tr}\ e^{x^i B_j} B_i\ -\ \omega(B_i)\omega(B_j) \\ &= \mathrm{Tr}\ ([B_j - \omega(B_j)]^{\mathrm{K}}_{\rho} B_i \\ &= \mathrm{Tr}\ [e^{(\mathrm{e})}_j]^{\mathrm{K}}_{\rho}\,(B_i - \omega(B_i)) \\ &= g_{ij}(\omega). \end{aligned}$$

□

4.2.2 Duality

The potential $\Phi(x)$ is a strictly convex function because its Hessian is a positive definite matrix. Hence it has a Legendre dual, which is given by

$$\Phi^*(y) = \sup\{x^i y_i - \Phi(x) :\ x \in \mathbb{R}^{n^2-1}\}, \quad y \in \mathbb{R}^{n^2-1}.$$

It is well-known that the supremum in this expression is actually a maximum that is reached at the unique solution x of the set of equations $\partial_i\Phi(x) = y_i$. A dual coordinate $x^*(\omega)$ is now defined by $x^*(\omega) = y$ when $x = x(\omega)$ solves the set of equations $\partial_i\Phi(x) = y_i$. It satisfies

$$\Phi(x^*(\omega)) \quad = \quad [x^i x_i^*]_\omega - \Phi(x(\omega)) \qquad \text{for all } \omega \in \mathbb{M}_n. \tag{4.6}$$

Proposition 18 *The dual coordinates $X_i^*(\omega)$ have the following properties*

(a) *The correspondence $x(\omega) \leftrightarrow x^*(\omega)$ is one-to-one;*

(b) *The dual of the relation $\partial_i\Phi(x) = x_i^*$ is*

$$x^j(\omega) = \frac{\partial}{\partial x_j^*}\Phi^*(x^*)\bigg|_{x^*=x^*(\omega)} \tag{4.7}$$

(c) *The inverse of the metric tensor is given by*

$$\frac{\partial^2}{\partial x_i^*\partial x_j^*}\Phi^*(x^*)\bigg|_{x^*=x^*(\omega)} = g^{ij}(\omega) \tag{4.8}$$

Proof

a) The one-to-one correspondence follows because the potential $\Phi(x)$ is strictly convex and hence it has a unique tangent plane at each point $\omega \in \mathbb{M}_n$.

b) Take the derivative from (4.6). This gives

$$\frac{\partial}{\partial x_j^*}\Phi^*(x^*) \quad = \quad x^j(\omega) + x_i^*(\omega)\frac{\partial x^i}{\partial x_j^*} - \frac{\partial}{\partial x_j^*}\Phi(x) \tag{4.9}$$

Use the chain rule to evaluate

$$\begin{aligned}\frac{\partial}{\partial x_j^*}\Phi(x) \quad &= \quad \left(\frac{\partial}{\partial x^i}\Phi(x)\right)\frac{\partial x^i}{\partial x_j^*} \\ &= \quad x_i^*(\omega)\frac{\partial x^i}{\partial x_j^*}.\end{aligned}$$

Hence, two terms of (4.9) cancel each other and (4.7) follows.

c) Take another derivative. This gives

$$\frac{\partial^2}{\partial x_i^* \partial x_j^*} \Phi^*(x^*) \quad = \quad \frac{\partial}{\partial x_i^*} x^j(\omega). \tag{4.10}$$

From

$$g_j^i = \frac{\partial x^i}{\partial x_k^*} \frac{\partial x_k^*}{\partial x^j} = \frac{\partial x^i}{\partial x_k^*} g_{kj}$$

it follows that

$$\frac{\partial x^i}{\partial x_k^*} \quad = \quad g^{ik}.$$

Hence, (4.10) becomes (4.7).

□

4.2.3 The tracial state

The tracial state ω^c is defined by

$$\omega^c(A) \quad = \quad \frac{1}{n} \operatorname{Tr} A, \qquad A \in \mathcal{B}(\mathbb{C}^n).$$

It satisfies $\omega^c(AB) = \omega^c(BA)$ for all A, B and for that reason it is called a central state.

All states of a commutative C^*-algebra are central states.

A short calculation gives

$$\begin{aligned} &x^i(\omega^c) = \omega^c(B_i) = \frac{1}{n} \operatorname{Tr} B_i = 0 \\ &[e_i]_{\omega^c}(A) = \frac{1}{n} \operatorname{Tr} B_i A \\ &g_{i,j}(\omega^c) = \frac{1}{n} \delta_{i,j}. \end{aligned}$$

These are the same results, up to a scaling factor, as in the case of the Bures metric. The two metrics coincide on the tangent space $T_{\omega^c}\mathbb{M}_n$ at ω^c. In the case of the Bures metric the m-connection has also a flat dual connection. However, it does not consist of exponential arcs as they have been defined in Section 3.2.2.

4.2.4 Generalization

Note that the coordinates x and the dual coordinates x^* vanish when the state ω is the central state ω^c. These coordinates are therefore said to be centered at $\omega = \omega^c$. Also the potential $\Phi(\omega)$ is minimal at the central state. A generalization of the Legendre structure so that it becomes centered at an arbitrary state ω in $\mathbb{M}_n$ follows below. In particular, the potential $\Phi(x)$ is generalized to $\Phi_\omega(A)$, $A = A^*$, centered at $\omega \in \mathbb{M}^n$.

In the parameterized case the Legendre structure of dually flat manifolds makes use of the self-duality of Euclidean space $\mathbb{R}^n$. Here, the manifold $\mathbb{M}_n$ consists of faithful states. They belong to the dual of the space of Hermitian matrices A. The dual of the map $\omega : A \mapsto \omega(A)$ is the map $A : \omega \mapsto \omega(A)$. Therefore the potential Φ_ω is a function of a Hermitian matrix A and the dual potential Φ^* should be labeled with a Hermitian matrix A.

Let the divergence between states ϕ and ω with density matrices σ, ρ be defined by $D(\phi||\omega) = D(\sigma||\rho)$. The potential Φ_ω centered at $\omega \in \mathbb{M}_n$ is now defined as the Legendre transform of the divergence $\phi \mapsto D(\phi||\omega)$

$$\Phi_\omega(A) \quad = \quad \sup\{\phi(A) - D(\phi||\omega) : \phi \in \mathbb{M}_n\}, \qquad A = A^*. \tag{4.11}$$

Proposition 19 *Assume $A = A^*$ satisfies $\omega(A) = 0$. Then one has*

(a) *The maximum in the r.h.s. of (4.11) is reached for $\phi = \psi_A$ with corresponding density matrix τ_A such that $c_\rho(\tau_A) = A$;*

(b) *The maximum value is given by*

$$\Phi_\omega(A) = D(\omega||\psi_A) = \log \textit{ Tr } \exp(\log\rho + A);$$

(c) *The inverse transform is given by*

$$D(\phi||\omega) = \sup\{\phi(A) - \Phi_\omega(A) : A = A^* \textit{ and } \omega(A) = 0\}.$$

Proof of (a) Let $\tau_A = \exp(\log\rho + A)/ \text{ Tr } \exp(\log\rho + A)$

Then it holds that $c_\rho(\tau_A) = \log\tau_A - \log\rho + D(\rho||\tau_A) = A$ and

$$\begin{aligned} \psi_A(A) - D(\tau_A||\rho) &= \text{Tr } \tau_A(A - \log\tau_A + \log\rho) \\ &= \log \text{ Tr } \exp(\log\rho + A) \end{aligned}$$

Let us show that ψ_A chosen in this way maximizes the quantity $\phi(A) - D(\phi||\omega)$. One calculates

$$0 \quad \leq \quad D(\sigma||\tau_A) = \text{ Tr } \sigma(\log\sigma - \log\tau_A)$$

$$
\begin{aligned}
&= D(\sigma||\rho) - \phi(A) + \log \text{ Tr } \exp(\log \rho + A) \\
&= [\psi_A(A) - D(\psi_A||\omega)] - [\phi(A) - D(\phi||\omega)]
\end{aligned}
$$

This shows the inequality

$$
\phi(A) - D(\phi||\omega) \quad \leq \quad \psi_A(A) - D(\psi_A||\omega).
$$

It implies (a).

Proof of remainder of (b) Note that

$$
\begin{aligned}
\log \text{ Tr } \exp(\log \rho + A) &= \log \text{ Tr } \exp(\log \tau_A + D(\rho||\tau_A)) \\
&= D(\rho||\tau_A) = D(\omega||\psi_A)
\end{aligned}
\tag{4.12}
$$

Proof of (c) $D(\phi||\omega) \geq \phi(A) - \Phi_\omega(A)$ follows from the definition of Φ_ω. Let $A = c_\rho(\sigma)$. Then the density matrix σ of ϕ satisfies $\sigma = \tau_A$ for all $A = A^*$ satisfying Tr $\rho A = 0$. Hence, equality is reached in (4.11).

□

4.2.5 Fréchet derivatives

A known property of the Legendre transform, used already in Section 4.2.2,

$$
f(x) \rightarrow f^*(x^*) = \sup_x \{xx^* - f(x)\}
$$

is that the maximum is reached when $f'(x) = x^*$. In the context of Banach spaces and their duals this property becomes the following: The maximum of $\phi(A) - \Phi_\omega(A)$ is reached when A is such that $\phi(B) = d_B\Phi_\omega(A)$ where d_B is the Fréchet derivative of $\Phi_\omega(A)$ in direction B.

Proposition 20 $d_B\Phi_\omega(A) = \psi_A(B)$

Proof

Recall that the definition of the Fréchet derivative $d_B\Phi_\omega(A)$ requires that the linear map $B \mapsto \phi(B)$ is such that

$$
|\Phi_\omega(A + \epsilon B) - \Phi_\omega(A) - \epsilon\phi(B)| = \text{ o } (\epsilon)
$$

A short calculation gives

$$
\begin{aligned}
\Phi_\omega(A + \epsilon B) &= \log \text{ Tr } \exp(\log \rho + A + \epsilon B) \\
&= \log \text{ Tr } \exp(\log \rho + A)[1 + \epsilon\psi_A(B) + \text{ O}(\epsilon^2)] \\
&= \Phi_\omega(A) + \epsilon\psi_A(B) + \text{ O}(\epsilon^2)
\end{aligned}
\tag{4.13}
$$

□

Chapter 5

Exponential families of density matrices

The study of exponential families of density matrices and corresponding exponential families of states on a C^*-algebra or a von Neumann algebra is well established in the present finite-dimensional case. The final Section of the Chapter discusses the problems arising when states on infinite-dimensional algebras are considered.

5.1 A family of states

5.1.1 Definitions

A *statistical model* is a set S of probability measures on a measure space X, μ parameterized by an injective map

$$\theta \in \Theta \mapsto \mu_\theta \in S, \qquad \Theta \subset \mathbb{R}^m$$

A *quantum statistical model* is a set S of states on a C^*-algebra $\mathcal{A}$ parameterized by an injective map

$$\theta \in \Theta \mapsto \omega_\theta \in S, \qquad \Theta \subset \mathbb{R}^m$$

The main advantage of working with models is that the number of parameters m can often be kept small compared to the dimension of the C^*-algebra $\mathcal{A}$. This is especially relevant when the dimension is infinite.

Definition 5 *The states ω_θ of a (quantum) statistical model $\theta \in \Theta \mapsto \omega_\theta$ form an* exponential family $\mathbb{M}_H$ *if $\exists$ self-adjoint operators H_k, $k = 1, 2, \cdots, m$, in $\mathcal{B}(\mathbb{C}^n)$*

such that

$$\rho_\theta \quad = \quad \exp(\theta^k H_k - \alpha(\theta)), \qquad \theta \in \Theta$$

with $\alpha(\theta) = \log$ *Tr* $\exp(\theta^k H_k)$. *The set of operators* $H_1, H_1, \cdots, H_k, \mathbb{I}$ *is assumed to be linearly independent. The* ρ_θ *are the density matrices of the states* ω_θ.

Notes In Statistical Physics the quantum exponential family is being used since long. Many quantum statistical models have been studied in great detail. An exponential family of probability distributions is called a Gibbs distribution.. A quantum exponential family is called a quantum Gibbs state. The operators H_k are pieces of the quantum Hamiltonian.

5.1.2 Properties

An important property of the normalization $\alpha(\theta)$ is

$$\begin{aligned}
\partial_i \alpha(\theta) &= \partial_i \log \text{ Tr } \exp(\theta^k H_k) \\
&= \frac{1}{\text{Tr } \exp(\theta^k H_k)} \partial_i \text{ Tr } \exp(\theta^k H_k) \\
&= \text{Tr } \rho_\theta H_i \\
&= \omega_\theta(H_i).
\end{aligned}$$

Notations Introduce the notation $U_i(\theta) = \omega_\theta(H_i)$. Later it is shown that the U_i are coordinates dual to the parameters θ^i. Let also $c_\theta(\rho_\eta) = c_{\rho_\theta}(\rho_\eta)$.

Further properties are

Proposition 21 *One has*

(a) $D(\rho_\theta||\rho_\eta) = (\theta^k - \eta^k)U_k(\theta) - \alpha(\theta) + \alpha(\eta)$

(b) $c_\theta(\rho_\eta) = (\eta^k - \theta^k)(H_k - U_k(\theta))$

Proof

One has

$$\begin{aligned}
D(\rho_\theta||\rho_\eta) &= \text{Tr } \rho_\theta(\log \rho_\theta - \log \rho_\eta) \\
&= \text{Tr } \rho_\theta \left[(\theta^k - \eta^k)H_k - \alpha(\theta) + \alpha(\eta)\right] \\
&= (\theta^k - \eta^k)U_k(\theta) - \alpha(\theta) + \alpha(\eta)
\end{aligned}$$

and

$$c_\theta(\rho_\eta) = c_{\rho_\theta}(\rho_\eta) \quad = \quad \log \rho_\eta - \log \rho_\theta + D(\rho_\theta||\rho_\eta)$$

$$= (\eta^k - \theta^k)(H_k - U_k(\theta)).$$

□

5.1.3 Tangent vectors

Tangent vectors $\chi_\eta \equiv \chi_{\omega_\eta}$ are given by

$$\begin{aligned}[\chi_\eta]_{\omega_\theta}(A) &= \text{Tr}\,[c_\theta(\rho_\eta)]^{\text{K}}_\theta A \\ &= (\eta^k - \theta^k)[e^{(\text{H})}_k]_\theta(A)\end{aligned}$$

with basis vectors given by

$$[e^{(\text{H})}_k]_\theta(A) = \text{Tr}\,[H_k - U_k(\theta)]^{\text{K}}_\theta A.$$

These tangent vectors belong to the subspace of $T_\theta\mathbb{M}_n$ spanned by the $[e^{(\text{H})}_k]_\theta$.

In this basis the metric tensor reads

$$\begin{aligned}g^{\text{H}}_{ij}(\theta) = (e^{(\text{H})}_i, e^{(\text{H})}_j)_\theta &= \text{Tr}\,[H_i - U_i(\theta))]^{\text{K}}_\theta(H_j - U_j(\theta)) \\ &= \text{Tr}\,(H_i - U_i(\theta)))[H_j - U_j(\theta)]^{\text{K}}_\theta \\ &= \text{Tr}\,[H_i]^{\text{K}}_\theta H_j - U_i(\theta)U_j(\theta). \qquad (5.1)\end{aligned}$$

The quantity

$$\langle\langle H_i - U_i, H_j - U_j\rangle\rangle_\theta$$

is called a 'Generalized covariance matrix' (see for instance Section 7.3 of [45]). It is 'generalized' because from one of the two entries a Kubo transformation is involved.

5.1.4 The Fisher information matrix

The definition of the Fisher information matrix in the commutative case reads

$$\begin{aligned}I_{jk}(\theta) &= \mathbb{E}_\theta(\partial_i \log p_\theta)(\partial_j \log p_\theta) \\ &= -\mathbb{E}_\theta\partial_i\partial_j \log p_\theta.\end{aligned}$$

A straightforward generalization is not obvious. Should one use

$$\omega_\theta\left((\partial_i \log \rho_\theta)(\partial_j \log \rho_\theta)\right)$$

or

$$-\omega_\theta\left(\partial_i\partial_j \log \rho_\theta\right)?$$

In general, these expressions differ because $\partial_i \log \rho_\theta$ does not commute with ρ_θ. This point has been stressed in particular by Petz. In [57], Section 10.3, he concludes that there exists "... a reasonable but still wide class of possible quantum Fisher informations."

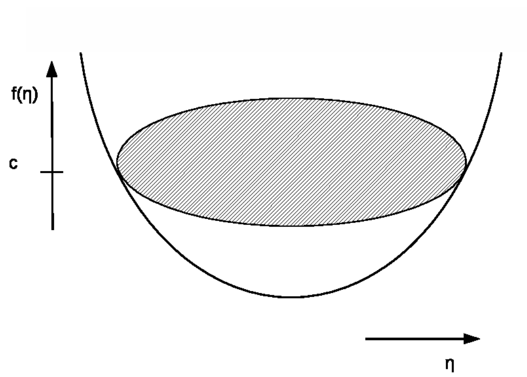

Figure 5.1: A level set of the parabolic function in $\mathbb{R}^2$. It consists of all points η satisfying $|\eta^2| \leq c$. This is the interior of a circle in the η-plane, including the circle itself.

5.1.5 Pythagorean relation

Lemma 8 *For any state $\phi \in \mathbb{M}_n$ the function*

$$\eta \in \mathbb{R}^m \mapsto f_\phi(\eta) = \alpha(\eta) - \eta^k \phi(H_k)$$

is strictly convex with compact level sets.

Proof Take $\eta \neq 0$ and $\lambda > 0$. ,The condition $f_\phi(\lambda\eta) \leq c$ is satisfied if and only if

$$\begin{aligned} &\log \text{ Tr } \exp(\lambda\eta^k H_k - \lambda\eta^k \phi(H_k) \leq c \\ &\qquad \leftrightarrow \quad \text{Tr } \exp(\lambda\eta^k H_k) \leq \exp(c + \lambda\eta^k \phi(H_k)) \end{aligned} \tag{5.2}$$

The operator $\eta^k H_k$ has at least one eigenvalue μ satisfying $\mu > \eta^k \phi(H_k)$ because ϕ is faithful and $\eta^k H_k$ is not a multiple of the identity $\mathbb{I}$. Hence the requirement $f_\phi(\lambda\eta) \leq c$ implies that

$$e^{\lambda\mu} < \text{ Tr } \exp(\lambda\eta^k H_k) \leq \exp(c + \lambda\eta^k \phi(H_k)).$$

The former inequality follows because the trace can be taken in a basis of eigenvectors. It decreases when only one of the positive terms is withhold. Take the logarithm to obtain

$$\lambda\mu \quad < \quad c + \lambda\eta^k \phi(H_k).$$

This can be written as

$$\lambda \quad \leq \quad \frac{c}{\mu - \eta^k \phi(H_k)}.$$

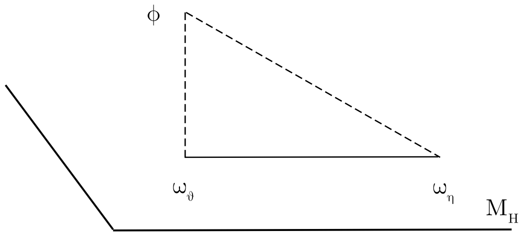

Figure 5.2: 'Orthogonal projection' of an arbitrary state ϕ onto the manifold $\mathbb{M}_{\text{H}}$ of the statistical model.

One reaches the conclusion that the level sets $\{\eta : f_\phi(\eta) \leq c\}$ are bounded and hence compact.

Strict convexity of the function $f_\phi(\eta)$follows from $\partial_i\partial_j f_\phi(\eta) = g^{\text{H}}_{ij}(\eta)$ and the observation that the metric tensor is a positive-definite matrix.

□

Theorem 8 *For any ϕ in $\mathbb{M}_n$ there exists a unique $\theta \in \mathbb{R}^m$ such that*

$$D(\phi||\omega_\eta) = D(\phi||\omega_\theta) + D(\omega_\theta||\omega_\eta), \qquad \forall \eta \in \mathbb{R}^m$$

It is fixed by the requirement that $U_k(\theta) = \phi(H_k)$.

Proof

Write $D(\phi||\omega_\eta) = \text{ Tr } \sigma \log \sigma + f_\phi(\eta)$. Because f_ϕ is strictly convex with compact level sets there exists a unique θ such that

$$D(\phi||\omega_\theta) \quad = \quad \min_\eta\{D(\phi||\omega_\eta)\}$$

It satisfies $\partial_i D(\phi||\omega_\eta)\big|_{\eta=\theta} = 0$. The latter implies that $U_k(\theta) = \phi(H_k)$. Now calculate

$$\begin{aligned} D(\phi||\omega_\eta) - D(\phi||\omega_\theta) &= \text{ Tr } \sigma(\log \rho_\theta - \log \rho_\eta) \\ &= (\theta^k - \eta^k)\phi(H_k) - \alpha(\theta) + \alpha(\eta) \\ &= (\theta^k - \eta^k)U_k(\theta) - \alpha(\theta) + \alpha(\eta) \\ &= D(\omega_\theta||\omega_\eta) \end{aligned} \tag{5.3}$$

This proves the Pythagorean relation.

□

5.2 The dual geometry

5.2.1 The e-connection

Proposition 22 *Assume that the domain Θ is a convex set. Take θ_0, θ_1 be points in Θ and let $\theta_t = (1-t)\theta_0 + t\theta_1$ Then $t \mapsto \omega_{\theta_t}$ is an exponential arc connecting ω_{θ_1} to ω_{θ_0}.*

The proof is straightforward.

See Section 1.8 of (Amari, Nagaoka, 2000) for the following definition.

Definition 6 *A submanifold S of $\mathbb{M}_n$ is said to be* **autoparallel** *w.r.t. ∇ if for all $\omega \in S$ and for any pair of vector fields V, W of the submanifold S the vector $[\nabla_V W]_\omega$ belongs to the tangent plane $T_\omega S$.*

Proposition 23 *Assume $\Theta \subset \mathbb{R}^m$ is an open convex set. Then the family of exponential states $\mathbb{M}_H$ is an autoparallel submanifold of $\mathbb{M}_n$*

Proof

(a) S **is a submanifold** The coordinates x^i are related to the parameters θ^k by

$$x^i(\omega_\theta) = (\log \rho_\theta, B^i)_{\text{HS}} = \theta^k \text{ Tr } H_k B^i.$$

- The relation is linear, hence it is C^∞;

- The derivatives Tr $H_k B^i$ are linearly independent vectors in $\mathbb{R}^{n^2-1}$;

- The intersection $U \cap \Theta$ with U open in $\mathbb{R}^n$ is open.

(b) S **is autoparallel in** $\mathbb{M}_n$ The geometry of $\mathbb{M}_n$. is flat and $\Theta \subset \mathbb{R}^n$ is an affine subset of the parameter space $\mathbb{R}^{n^2-1}$. Hence, by Theorem 1.1. of (Amari Nagaoka, 2000) S is autoparallel in $\mathbb{M}_n$.

□

5.2.2 The potential $\Phi_\theta(A)$

Introduce the notation $\Phi_\theta(A) \equiv \Phi_{\omega_\theta}(A)$, where Φ_{ω_θ} is the potential introduced in Section 4.2.4.

Proposition 24 *Assume $\Theta = \mathbb{R}^m$. If $A = A^*$ belongs to lin$\{\mathbb{I}, H_1, \cdots H_m\}$ and $\omega_\theta(A) = 0$ then there exists $\eta \in \Theta$ such that*

(a) $A = c_\theta(\rho_\eta)$;

(b) ω_η *maximizes* $\phi \mapsto \phi(A) - D(\phi||\omega_\theta)$*;*

(c) $\Phi_\theta(A) = (\theta^k - \eta^k)U_k(\theta) - \alpha(\theta) + \alpha(\eta)$.

Proof

By assumption the matrix A is a linear combination of the K_k and the identity. Hence one can write $A = a^0 + a^k H_k$. Let $\eta^k = \theta^k + a^k$. The assumption that $\omega_\theta(A)$ implies that $a^0 + a^k U_k(\theta) = 0$.

(a) Calculate

$$\begin{aligned} c_\theta(\rho_\eta) &= (\theta^k - \eta^k)(H_k - U_k(\theta)) \\ &= a^k(H_k - U_k(\theta)) \\ &= A - a^0 - a^k U_k(\theta) \\ = A. \end{aligned}$$

(b) The maximum is reached by ψ_A such that $c_\theta(\tau_A) = A$ holds. From $c_\theta(\rho_\eta) = A$ it then follows that $\tau_A = \rho_\eta$ and $\psi_A = \omega_\eta$.

(c) Section 4.2.4, Item (b) of Proposition 19, implies that

$$\Phi_\theta(A) = D(\omega_\theta||\omega_\eta).$$

From Section 5.1.2, Item (a) of Proposition 21 one obtains

$$D(\omega_\theta||\omega_\eta) = (\theta^k - \eta^k)U_k(\theta) - \alpha(\theta) + \alpha(\eta).$$

Both statements together prove (c).

□

Introduce the potential $\Phi_\theta^{\mathrm{H}}(\eta) = \Phi_\theta(c_\theta(\rho_\eta))$. It follows from Item (b) of Proposition 19 that $\Phi_\theta^{\mathrm{H}}(\eta) = D(\rho_\eta||\rho_\theta)$. The following corollary show that $U_i(\eta) - U_i(\theta)$ is the dual coordinate of η_i w.r.t. this potential and that the Hessian is the metric tensor g^{H}.

Corollary 4 *The first derivative of the potential* $\Phi_\theta(\eta)$ *satisfies*

$$\frac{\partial}{\partial\eta^i}\Phi_\theta^H(\eta)) = U_i(\eta) - U_i(\theta). \tag{5.4}$$

The second derivative satisfies

$$\frac{\partial^2}{\partial\eta^j\partial\eta^i}\Phi_\theta^H(\eta) = g_{ij}^H(\eta).$$

Proof

One calculates

$$\begin{aligned}\frac{\partial}{\partial\eta^i}\Phi^{\mathrm{H}}_\theta(\eta)) &= \frac{\partial}{\partial\eta^i}\left[(\theta^k-\eta^k)U_k(\theta)+\alpha(\eta)-\alpha(\theta)\right]\\ &= -U_i(\theta)+\frac{\partial}{\partial\eta^i}\alpha(\eta).\end{aligned}$$

Use $\partial_i\alpha(\eta)=U_i(\eta)$ (see Section 5.1.2) to find (5.4).

The second derivative follows from

$$\begin{aligned}\frac{\partial^2}{\partial\eta^j\partial\eta^i}\Phi_\theta(A) &= \frac{\partial}{\partial\eta^j}U_i(\eta)\\ &= \frac{\partial}{\partial\eta^j}\,\mathrm{Tr}\,\exp(\eta^kH_k-\alpha(\eta))H_i\\ &= \mathrm{Tr}\,[H_j]^{\mathrm{K}}_\eta H_i-\frac{\partial\alpha}{\partial\eta^j}\omega_\eta(H_i)\\ &= \langle\langle H_i-U_i(\eta),H_j-U_j(\eta)\rangle\rangle_\eta\\ &= g^{\mathrm{H}}_{ij}(\eta)\end{aligned} \tag{5.5}$$

□

5.3 Quantum estimation

5.3.1 Quantum measurements

The quantum counterpart of a partition of sample space is a sequence $(E_k)_k$ of orthogonal projection operators, 2-by-2 commuting and summing up to the identity $\sum_k E_k=\mathbb{I}$.

A *von Neumann type measurement* returns probabilities p_k^{emp}, one for each projection operator E_k. In other words, the result of the measurement is the *empirical density matrix* $\rho^{\mathrm{emp}}=\sum_k p_k^{\mathrm{emp}}E_k$ and the corresponding *empirical state* ϕ^{emp}. The meaning of p_k^{emp} is the frequency of the 'event' that the result is in the range $E_k\mathscr{H}$ of the projection operator E_k. Results of a von Neumann type measurement are never assigned to elements of the Hilbert space not belonging to the range of any of the E_k.

In Quantum Physics elements of the Hilbert space are called wave functions.. One encounters the phrase "Reconstruction of the wave function". The question one poses is what was the state of the system before measuring? The assumption is that the measurement disturbs the state of the system in such a way that the outcome of the measurement is always in the range of one of the projection operators E_k, never a linear combination of such outcomes.

That the measurement disturbs the state of the system has several consequences

- It is necessary to prepare N copies of system all prepared in the same way;
- N measurements are performed, each time a fresh copy is used;
- Eventually, the preceeding steps have to be repeated with different measurement setups because von Neumann type measurements return results only for a set of mutually commuting operators .

Minimal divergence criterion A geometrically motivated choice for the optimal reconstructed state uses the 'orthogonal projection' discussed in Section 5.1.5 on the Pythagorean relation. It selects the model state ω_θ minimizing the divergence map $\eta \mapsto D(\phi^{\text{emp}}||\omega_\eta)$.

Proposition 25 *Let θ be the value which minimizes the map $\eta \mapsto D(\phi^{emp}||\omega_\eta)$. Then the state ω_θ satisfies $\omega_\theta(H_k) = \phi^{emp}(H_k)$ for all k.*

Proof

It is proved in Theorem 8 that $U_k(\theta) = \omega_\theta(H_k) = \phi^{\text{emp}}(H_k)$.

□

It is not clear whether this criterion is the best one can do. If the operators H_k determining the manifold $\mathbb{M}_{\text{H}}$ coincide with the projection operators E_k then the expectation values $\omega_\theta(E_k)$ coincide with the empirical values $\phi^{\text{emp}}(E_k) = p_k^{\text{emp}}$. To realize this property in a more general context is desirable.

5.3.2 Estimators

A recent paper on the quantum estimation problem is [72].

The operators H_k form an *unbiased estimator* for the dual parameters $U_k(\theta)$ because $\omega_\theta(H_k) = U_k$ holds for all $\theta \in \Theta$.

The estimator H_k is said to be *efficient* if the inequality of Cramér-Rao holds as an equality.

In the non-commutative case the quantum Fisher information matrix is non-unique. Hence the choice of metric is non-unique. Note that the inequality of Cramér-Rao depends on the choice of metric. Here the metric of Bogoliubov is used. The H_k are an efficient estimator for this metric.

The quantum version of the Theorem of Cramér-Rao states that generalized covariance is minimal when the estimator equals the set of operators H_k. This is Theorem 9 of (Hasegawa, 1997) [40].

Theorem 9 *For any unbiased estimator X_k the matrix*

$$\langle\langle X_i - \omega_\theta(X_i), X_j - \omega_\theta(X_j)\rangle\rangle_\theta - [(g^{H})_{ij}]_\theta$$

is non-negative-definite and vanishes for $X_k = H_k$.

Proof

By assumption is $\omega_\theta(X_j) = U_j(\theta)$ for all θ. This implies $\partial_i\omega_\theta(X_j) = \partial_i U_j(\theta) = g^{\mathrm{H}}_{ij}(\theta)$. Choose u^i, v^j arbitrary in $\mathbb{R}^m$. Then one has

$$\begin{aligned}
u^i g^{\mathrm{H}}_{ij}(\theta)v^j &= u^i\partial_i U_j(\theta)v^j = u^i\partial_i\omega_\theta(v^j X_j)\\
&= \text{Tr}\,[u^i(H_i - U_i(\theta))]^{\mathrm{K}}_\theta[v^j X_j]\\
&= \text{Tr}\,[u^i(H_i - U_i(\theta))]^{\mathrm{K}}_\theta[v^j(X_j - \omega_\theta(X_j))]\\
&= \int_0^1 \mathrm{d}w\ \text{Tr}\,\rho_\theta^w[u^i(H_i - U_i(\theta))]\rho_\theta^{1-w}[v^j(X_j - \omega_\theta(X_j))]\\
&= \int_0^1 \mathrm{d}w\ \text{Tr}\,A^*_w B_w
\end{aligned}\tag{5.6}$$

with $A_w = \rho_\theta^{(1-w)/2}[u^i(H_i - U_i(\theta))]\rho_\theta^{w/2}$ and $B_w = \rho_\theta^{(1-w)/2}[v^j(X_j - \omega_\theta(X_j))]\rho_\theta^{w/2}$. Use Schwarz inequality to find

$$\begin{aligned}
\left(u^i g^{\mathrm{H}}_{ij}(\theta)v^j\right)^2 &= \left(\int_0^1 \mathrm{d}w\ \text{Tr}\,A^*_w B_w\right)^2\\
&\le \int_0^1 \mathrm{d}w\ \text{Tr}\,A^*_w A_w \times \int_0^1 \mathrm{d}w\ \text{Tr}\,B^*_w B_w\\
&= \left(\text{Tr}\,[u^i(H_i - U_i(\theta))]^{\mathrm{K}}[u^j(H_j - U_j(\theta))]\right)\\
&\quad\times\left(\text{Tr}\,[v^k(X_k - \omega_\theta(X_k))]^{\mathrm{K}}[v^l(X_l - \omega_\theta(X_l))]\right)\\
&= u^i g^{\mathrm{H}}_{ij}(\theta)u^j \times \left(v^k(X_k - \omega_\theta(X_k), X_l - \omega_\theta(X_l))_\theta v^l\right).
\end{aligned}\tag{5.7}$$

Take $v = u$and divide out one factor of the r.h.s. to obtain

$$u^i g^{\mathrm{H}}_{ij}(\theta)u^j \le u^i(X_i - \omega_\theta(X_i), X_j - \omega_\theta(X_j))_\theta u^j.$$

□

5.4 Examples

5.4.1 The Pauli spin

See Section 2.4, “The case n=2”.

The simplest exponential family contains the states ω_θ, $\theta \in \mathbb{R}$, with density matrices

$$\rho_\theta \sim \exp(\theta\sigma_3) = \cosh\theta + \sigma_3 \sinh\theta.$$

This the commutative two-state model.

The requirement that the matrices $\{\mathbb{I}, H_1, \cdots, H_m\}$ must be linearly independent implies that $m \leq 3$. Consider for instance, $\rho(\theta) \sim \exp(\theta_1\sigma_3 + \theta_2\sigma_1)$. This two-parameter model is an exponential family of quantum states. Explicit calculations are left as an exercise.

5.4.2 Two spins

See Section 1.4, “A historical experiment”.

Let H be the orthogonal projection onto a one-dimensional subspace of $\mathbb{C}^2 \otimes \mathbb{C}^2$. Consider the density matrix

$$\rho_\theta = \frac{e^\theta H + \mathbb{I} - H}{e^\theta + 3}.$$

with normalization function

$$\alpha(\theta) = \log \text{ Tr } \exp\theta H = \log(e^\theta + 3).$$

A short calculation gives

$$U(\theta) = \mathbb{E}_\theta H \omega_\theta(H) = \frac{e^\theta}{e^\theta + 3}.$$

Let an observable X_φ be defined by

$$X_\varphi = \sigma_3 \otimes (\cos(2\varphi)\sigma_3 + \sin(2\varphi)\sigma_1).$$

It satisfies $X_\varphi^2 = \mathbb{I} \otimes \mathbb{I}$. Now calculate

$$\begin{aligned} \mathbb{E}_\theta X_\varphi &= \omega_\theta(X_\varphi) \\ &= \cos(2\varphi)\frac{\text{Tr } (e^\theta H + \mathbb{I} - H)\sigma_3 \otimes \sigma_3}{e^\theta + 3} \end{aligned}$$

$$\begin{aligned}
&+ \quad \sin(2\varphi)\frac{\mathrm{Tr}\,(e^\theta H + \mathbb{I} - H)\sigma_3\otimes\sigma_1}{e^\theta+3} \\
&= \quad \cos(2\varphi)\frac{e^\theta-1}{e^\theta+1}\,\mathrm{Tr}\,H\sigma_3\otimes\sigma_3 \\
&+ \quad \sin(2\varphi)\frac{e^\theta-1}{e^\theta+1}\,\mathrm{Tr}\,H\sigma_3\otimes\sigma_1
\end{aligned} \tag{5.8}$$

To make this result tractable make use of the isomorphism $\mathbb{C}^2\otimes\mathbb{C}^2 \simeq \mathbb{C}^4$. In particular us in what follows that

$$\sigma_3\otimes\sigma_\alpha \simeq \left(\begin{array}{cc}\sigma_\alpha & 0\\ 0 & -\sigma_\alpha\end{array}\right).$$

Choose now H equal to the one-dimensional projection operator defined by $H\psi = \psi$ with $\psi = \frac{1}{2}(1,-1,-1,1)^{\mathrm{T}}$. It satisfies

$$\begin{aligned}
\mathrm{Tr}\,H\sigma_3\otimes\sigma_3 &= \frac{1}{4}(\sigma_3\otimes\sigma_3\psi,\psi) = 1,\\
\mathrm{Tr}\,H\sigma_3\otimes\sigma_1 &= \frac{1}{4}(\sigma_3\otimes\sigma_1\psi,\psi) = 0.
\end{aligned}$$

Use this to evaluate

$$\omega_\theta(X_\varphi) = \cos(2\varphi)\frac{e^\theta-1}{e^\theta+1}.$$

Compare this result with the result from Section 1.4

$$p^{\mathrm{emp}} = \frac{1}{2}(1+\phi^{\mathrm{emp}}(X_\varphi)) = \frac{1}{2} + \kappa\cos(2\varphi).$$

Both results agree with $\kappa = 1/2$ THis shows that the experiment discussed in Chapter 1 can be modelled by an exponential family of quantum states.

5.5 Infinite-dimensional case

5.5.1 Introduction

The extension of non-commutative Information Geometry to the infinite-dimensional context is treated in a number of publications. Let me mention [47, 49, 50, 51, 55], recent works of Ciaglia et al [69, 71] and of the present author [67, 73].

Some of the problems that arise when the Hilbert space $\mathscr{H}$ is allowed to be infinite-dimensional are discussed below.

Matrices become operators, i.e. linear maps $\mathscr{D} \subset \mathscr{H} \mapsto \mathscr{H}$ with domain of definition $\mathscr{D}$ that can be taken to be all of $\mathscr{H}$ in the case of bounded operators and that is assumed to be a dense subspace of $\mathscr{H}$ in the case of unbounded operators. Bounded operators have a finite supremum norm and, if densely-defined, extend by continuity to all of $\mathscr{H}$.

Note that even simple systems of Quantum Mechanics such as the quantum harmonic oscillator involve an infinite-dimensional Hilbert space $\mathscr{H}$. Hence, the study of the infinite-dimensional case is more than just an academic exercise.

- The algebra $\mathcal{A}$ is a von Neumann algebra of bounded operators on a separable Hilbert space $\mathscr{H}$.

- In the general case, not all states on $\mathcal{A}$ are normal states.. Note that a state ω of $\mathcal{A}$ is normal if and only if there exists a density matrix ρ on $\mathscr{H}$ such that $\omega(A) = \text{Tr } \rho A$ for all $A \in \mathcal{A}$. This is not the definition but a characterisation — See Theorem 2.4.21 of [22].

- The model states ω_θ with $\theta \in \Theta \subset \mathbb{R}^m$ are assumed to be normal states on $\mathcal{A}$ represented by density matrices ρ_θ. A density matrix (density operator) ρ is a trace-class positive operator satisfying $\text{Tr } \rho = 1$.

- The operators $H_k$defining the manifold $\mathbb{M}_{\text{H}}$ may be unbounded self-adjoint operators with domains $\mathscr{D}_k$.

- Note that $H = H^*$ requires that the domains of H and H^* coincide: $\text{dom }(H) = \text{dom }(H^*)$. In general, a symmetric operator satisfies $\text{dom }(H) \subset \text{dom }(H^*)$ and the restriction opf H^* to the domain of H coincides with H.

Let us now list some of the delicate problems that arise.

Problem The operators $\theta^k H_k$ are only defined on the intersection $\cap \mathscr{D}_k$. It can happen that this intersection is the null space $\{0\}$. It is clear that in such a case the sum $\theta^k H_k$ is rather meaningless.

An obvious solution to this problem is to require that the operators $H_2, \cdots H_m$ bounded relative to H_1. An operator A is *relatively bounded* w.r.t. the operator T if there exists constants $a \geq 0$ and $b \geq 0$ such that

$$||Au|| \leq a||u|| + b||Tu|| \qquad \text{for all } u \in \text{ dom }(T) \subset \text{ dom }(A).$$

See Theorem 4.3 in Sect V of [9]. If $b < 1$ then one has in addition that the domain of the operator T is unchanged by adding the 'small perturbation' A: $\text{dom }(T + A) = \text{dom }(T)$.

Problem The operator $\exp(\theta^k H_k)$ must be trace-class. If it is then it becomes a density operator by normalizing its trace.

Usually it is necessary to restrict the parameter domain Θ to the set of θ-values for which one can prove that the operator $\exp(\theta^k H_k)$ is trace-class. An assumption often made in Statistical Physics is that the operators H_k bounded from above or from below. They appear in an exponential. If all eigenvalues of $\theta^k H_k$ are bounded from above then $\exp(\theta^k H_k)$ is a bounded operator. Indeed, eigenvalues of $\theta^k H_k$ diverging to $-\infty$, after exponentiation, tend to 0.

Problem The divergence (Bures' divergence or Umegaki's relative entropy) $D(\phi||\omega)$ may diverge even for normal states ϕ and ω close in norm.

A possible solution is the restriction to states ϕ 'absolutely continuous' w.r.t. to the given state ω. To do so one needs of course a generalization of absolute continuity to the non-commutative case. Alternatively, Streater [49] argues that a stronger topology is needed. In fact, this problem occurs also in the commutative case and motivated the introduction of Orlicz spaces by Pistone and Sempi [37].

5.5.1 Examples

A simple example of a one-parameter exponential family of density matrices on separable Hilbert space is the quantum harmonic oscillator. The three-parameter Jaynes-Cummings model [6, 44] is more involved but still tractable.

Harmonic oscillator Choose an orthonormal basis $(f_n)_{,n=0,12,\cdots}$ in separable Hilbert space $\mathscr{H}$.

Proposition 26 *A self-adjoint operator H is defined by $Hf_n = nf_n$ for all n.*

Proof

Let E_n denote orthogonal projection onto $\mathbb{C}f_n$. The spectral theorem implies that $H = \sum_n nE_n$ is a self-adjoint operator with domain

$$\text{dom }(H) = \{f : \sum_n n^2|(f, f_n)|^2 < +\infty\}.$$

$\square$

Choose the domain Θ of the model parameter θ equal to $\Theta = (-\infty, 0)$. The quantum harmonic oscillator is the one parameter statistical model $\theta \in \Theta \mapsto \omega_\theta$

with

$$\begin{array}{rcl}\rho_\theta &=& \exp(\theta H - \alpha(\theta)) \quad \text{and} \\ \alpha(\theta) &=& \text{Tr } \exp(\theta H).\end{array}$$

The quantum harmonic oscillator is an exponential family of quantum states.

Jaynes-Cummings model Consider the Hilbert space $\mathscr{H} = \mathscr{H}^{\text{HO}} \otimes \mathbb{C}^2$ with $\mathscr{H}^{\text{HO}}$ the Hilbert space of the quantum harmonic oscillator and $\mathbb{C}^2$ the 2-dimensional Hilbert space of a Pauli spin. Choose $H_1 = H^{\text{HO}} \otimes \mathbb{I}$ and $H_2 = \mathbb{I} \otimes \sigma_3$. Introduce the notations $|\uparrow\rangle = (1 \quad 0)^{\text{T}}$ and $|\downarrow\rangle = (0 \quad 1)^{\text{T}}$.

An operator H_3 is defined by linear extension of

$$\begin{array}{rcl}H_3 f_n \otimes |\uparrow\rangle &=& \sqrt{n+1} f_{n+1} \otimes |\downarrow\rangle, \\ H_3 f_n \otimes |\downarrow\rangle &=& \sqrt{n-1} f_{n-1} \otimes |\uparrow\rangle.\end{array}$$

An arbitrary element f of the Hilbert space $\mathscr{H}$ can be expanded into basis vectors by

$$f = \sum_n \sum_{s=\uparrow\downarrow} \lambda_{n,s} f_n \otimes |s\rangle.$$

It belongs to the domain of the operator H_1 if the squared norm satisfies

$$||H_1 f||^2 = \sum_n \sum_s n^2 |\lambda_{n,s}|^2 < +\infty.$$

Make now the estimates

$$\begin{array}{rcl}||H_3 f||^2 &=& \left[\sum_{n,s} n|\lambda_{n,s}|^2 + \sum_n |\lambda_{n,\uparrow}|^2 - \sum_n |\lambda_{n,\downarrow}|^2 - |\lambda_{0,\downarrow}|^2\right] \\ &\leq& |||f||||H_1 f|| + ||f||^2 \\ &\leq& \left(||f|| + \dfrac{1}{2}||H_1 f||\right)^2.\end{array} \tag{5.9}$$

Here, to estimate $\sum_{n,s} n|\lambda_{n,s}|^2$ the concavity of the square root is used. This estimate shows that $\theta^3 H_3$ is relatively bounded by $\theta^1 H_1$ with $b = |\theta^3/2\theta^1|$. if $|\theta^3| < 2|\theta^1|$ then $\theta^k H_k$ is self-adjoint with domain equal to $\text{dom}\,(H_1)$.

Note that $\theta^1 < 0$ is needed to make the operator ρ_θ trace class.

Bibliography

[1] J. W. Gibbs, *Elementary principles in statistical mechanics,* Reprint (Dover, New York, 1960)

[2] W. Heisenberg, *Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen,* Z. Phys. **33**, 879–893 (1925).

[3] A. Einstein, B. Podolsky and N. Rosen, *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?,* Phys. Rev. **47** 777–780 (1935).

[4] M. S. Knebelman, *Spaces of relative parallelism,* Ann. Math. **53**, 387–399 (1951).

[5] Umegaki, H. Conditional Expectation in an Operator Algebra. IV. Entropy and Information. *Kodai Math. Sem. Rep.* **1962**, *14,* 59–85.

[6] E.T. Jaynes, F.W. Cummings, *Comparison of quantum and semiclassical radiation theories with application to the beam maser,* Proc. IEEE **51**, 89 (1963).

[7] J. Bell, *On the Einstein Podolsky Rosen Paradox,* Physics **1** (3), 195–200 (1964).

[8] J. Dixmier, *Les C^*-algèbres et leurs représentations* (Gauthier-Villars, 1964)

[9] T. Kato, *Perturbation theory for linear operators* (Springer-Verlag, 1966)

[10] J. Dixmier, *Les algèbres d'operateurs dans l'espace Hilbertien* (Gauthier-Villars, 1969)

[11] J.F. Clauser, M.A. Horne, A. Shimony and R.A. Holt, *Proposed Experiment to Test Local Hidden-Variable Theories,* Phys. Rev. Lett. **23**, 880 (1969); Erratum Phys. Rev. Lett. **24**, 549 (1970).

[12] D. Ruelle, *Statistical Mechanics* (W.A. Benjamin, Inc., 1969)

[13] D. Bures, *An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite W*-algebras,* Trans. Am. Math. Soc. **135**, 199–212 (1969).

[14] S. Sakai, *C*-algebras and W*-algebras* (Springer-Verlag, 1971)

[15] H. Araki, *Some properties of modular conjugation operator of von Neumann algebras and a non-commutative Radon-Nikodym theorem with a chain rule,* Pac. J. Math, **50**, 309–354 (1974).

[16] J. Naudts, A. Verbeure, R. Weder, *Linear response theory and the KMS condition.,* Commun. math. Phys. **44,** 87–99 (1975).

[17] Lindblad, G. Completely positive maps and entropy inequalities. *Commun. math. Phys.* **1975**, *40,* 147–151.

[18] Araki, H. Relative entropies for states of von Neumann algebras. *publ. RIMS Kyoto Univ.* **1976**, *11,* 809 – 833.

[19] A. Uhlmann, *The "transition probability" in the state space of a *-algebra,* Rep. Math. Phys. **9**, 273–279 (1976).

[20] J. Dixmier, *C^*-algebras* (North-Holland, 1977)

[21] J.F. Clauser and A. Shimony, *Bell's theorem: experimental tests and implications,* Rep. Prog. Phys. **41**, 1881 – 1927 (1978).

[22] O. Bratteli, D.W. Robinson, *Operator Algebras and Quantum Statistical Mechanics I and II* (Springer-Verlag, 1979)

[23] W. Thirring, *Lehrbuch der Mathematischen Physik, Vol. 4,* (Springer-Verlag, 1980)

[24] J. Dixmier, *Von Neumann Algebras* (North-Holland Publishing, 1981)

[25] A. Aspect, Ph. Grangier, and G. Roger, *Experimental Tests of Realistic Local Theories via Bell's Theorem,* Phys. Rev. Lett. **47**, 460 (1981).

[26] W. K. Wootters, *Statistical distance and Hilbert space,* Phys. Rev. D. **23**, 357–362 (1981).

[27] S. Amari, *Differential-Geometrical Methods in Statistics,* Lecture Notes in Statistics **28** (Springer-Verlag, 1985)

[28] S. Eguchi, *A differential geometric approach to statistical inference on the basis of contrast functionals,* Hiroshima Math. J. **15**, 341–391 (1985).

[29] W. Rudin, *Functional Analysis* (McGraw-Hill, 1991)

[30] Eguchi, S.: Geometry of minimum contrast. Hiroshima Mathematical Journal 22(3), 631–647 (1992).

[31] M. Hübner, *Explicit computation of the Bures distance for density matrices,* Phys. Lett. A **163**, 239–242 (1992).

[32] D. Petz, G. Toth, *The Bogoliubov inner product in quantum statistics,* Lett. Math. Phys. **27**, 205–216 (1993).

[33] H. Hasegawa, α*-divergence of the non-commutative information geometry,* Rep. Math. Phys. **33**, 87–93 (1993).

[34] D. Petz, *Geometry of canonical correlation on the state space of a quantum system,* J. Math. Phys. **35**, 780–795 (1994).

[35] R. Jozsa, *Fidelity for Mixed Quantum States,* J. Mod. Opt. **41**, 2315–2323 (1994).

[36] D. Aerts, S. Aerts, *Applications of quantum statistics in psychological studies of decision processes,* Found. Sc. **1**, 85–97 (1994).

[37] G. Pistone, C. Sempi, *An infinite-dimensional structure on the space of all the probability measures equivalent to a given one,* Ann. Stat. **23**, 1543–1561 (1995).

[38] J. Dittmann, *On the Riemannian metric on the space of density matrices,* Rep. Math. Phys. **36**, 309–315 (1995).

[39] A. Uhlmann, *Geometric phases and related structures,* Rep. Math. Phys. **36**, 461–481 (1995).

[40] Hasegawa, H. Exponential and mixture families in quantum statistics: Dual struc- ture and unbiased parameter estimation. *Rep. Math. Phys.* **1997**, *39,* 49–68.

[41] Hasegawa H., Petz, D. Non-commutative extension of information geometry II. In *Quantum Communication, Computing and Measurement*. Hirota et al., Eds. (Plenum Press, New York, 1997) pp. 109–118.

[42] D. Aerts, J. Broekaert, S. Smets, *The liar paradox in a quantum mechanical perspective,* Found. Sc. **4**, 115–132 (1999).

[43] J. Dittmann, *The scalar curvature of the Bures metric on the space of density matrices,* J.Geom.Phys. **31**, 16–24 (1999)

[44] A.K. Rajagopal, K.L. Jensen, F.W. Cummings, *Quantum entangled supercorrelated states in the Jaynes-Cummings model,* Phys. Lett. A259, 285-290 (1999).

[45] S. Amari, H. Nagaoka, *Methods of Information Geometry* (Oxford University Press, 2000) (Originally published in Japanese by Iwanami Shoten, Tokyo, Japan, 1993)

[46] P. M. Alberti, A. Uhlmann, *On Bures Distance and -Algebraic Transition Probability between Inner Derived Positive Linear Forms over W-Algebras,* Acta Appl. Math. **60**, 1–37 (2000).

[47] M. R. Grasselli, R. F. Streater, *On the uniqueness of the Chentsov metric in quantum information geometry,* Infin. Dim. Anal. Qu. **4**, 173–182 (2001).

[48] H. J. Sommers, K. Życzkowski, *Bures volume of the set of mixed quantum states,* J. Phys. A**36**, 10083–10100 (2003).

[49] R. F. Streater, *Quantum Orlicz Spaces in Information Geometry,* Open Sys. & Inf. Dyn. **11**, 359–375 (2004).

[50] R. F. Streater, *Duality in Quantum Information Geometry,* Open Syst. & Inf. Dyn. **11**, 71–77 (2004).

[51] R. F. Streater, *Quantum Orlicz Spaces in Information Geometry,* Open Syst. & Inf. Dyn. **11**, 359–375 (2004).

[52] Ch. Gerry and P. Knight, *Introductory Quantum Optics* (Cambridge University Press, 2004).

[53] D. Aerts, M. Czachor, *Quantum aspects of semantic analysis and symbolic artificial intelligence,* J. Phys. A **37**, L123–L132 (2004).

[54] D. R. Cox, *Principles of Statistical Inference* (Cambridge University Press, 2006)

[55] A. Jenčová, *A construction of a nonparametric quantum information manifold,* J. Funct. Anal. **239**, 1–20 (2006).

[56] A. Khrennikov, *Quantum-like brain: "Interference of minds",* Biosystems **84**, 225–241 (2006).

[57] D. Petz, *Quantum Information Theory and Quantum Statistics* (Springer, 2008)

[58] A. Uhlmann, *Transition Probability (Fidelity) and Its Relatives,* Found. Phys. **41**, 288–298 (2011).

[59] J. Naudts, *Generalised Thermostatistics* (Springer, 2011).

[60] D. Aerts, J. Broekaert, M. Czachor, B. D'Hooghe, *A Quantum-Conceptual Explanation of Violations of Expected Utility in Economics,* in: D. Song,

M. Melucci, I. Frommholz, P. Zhang, L. Wang, S. Arafat (Eds): Quantum Interaction 2011, LNCS **7052**, 192–198 (2011).

[61] D. Aerts, M. Czachor, M. Kuna, S. Sozzo, *Systems, environments, and soliton rate equations: A non-Kolmogorovian framework for population dynamics,* Ecological Modelling, **267**, 80–92 (2013).

[62] E. Ercolessi, M. Schiavina, *Symmetric logarithmic derivative for general n-level systems and the quantum Fisher information tensor for three-level systems,* Phys. Lett. A**377**, 1996–2002 (2013).

[63] D. Aerts, J. Broekaert, M. Czachor, M. Kuna, B, Sinervo, S. Sozzo, *Quantum structure in competing lizard communities,* Ecological Modelling **281**, 38–51 (2014).

[64] S. Amari, *Information Geometry and its Applications* (Springer, 2016)

[65] N. Ay, J. Jost, H. Vân Lê, L. Schwachhöfer, *Information Geometry* (Springer, 2017).

[66] J. Naudts, *Quantum Statistical Manifolds,* Entropy **20**, 472 (2018), correction Entropy **20**, 796 (2018).

[67] J. Naudts, *Quantum statistical manifolds: the linear growth case,* Rep. Math. Phys. **84**, 151–169 (2019).

[68] M. Berthier and E. Provenzi, *When Geometry Meets Psycho-Physics and Quantum Mechanics: Modern Perspectives on the Space of Perceived Colors,* in: F. Nielsen and F. Barbaresco (Eds.): GSI 2019, LNCS 11712, 621–630 (2019).

[69] F. M. Ciaglia, A. Ibort, J. Jost, G. Marmo, *Manifolds of classical probability distributions and quantum density operators in infinite dimensions,* Inf. Geom. **2**, 231–271 (2019).

[70] J. Naudts, *Quantum Statistical Manifolds: The Finite-Dimensional Case,* In *Geometric Science of Information,* Ed. F. Nielsen and F. Barbaresco (Springer, 2019), pp. 631–637.

[71] F. M. Ciaglia, J. Jost, L. Schwachhöfer, *From the Jordan Product to Riemannian Geometries on Classical and Quantum States,* Entropy **22**, 637 (2020).

[72] J. Liu, H. Yuan, X.-M. Lu, X. Wang, *Quantum Fisher information matrix and multiparameter estimation,* J. Phys. A Math. Theor. 53(2), 023001–69 (2020).

[73] J. Naudts, *Exponential arcs in the manifold of vector states on a σ-finite von Neumann algebra,* submitted to Information Geometry (2000).

[74] J. Naudts, *Parameter-free description of the manifold of non-degenerate density matrices,* to appear in Eur. Phys. J. Plus (2021).

Index