A dynamical approach to quantum optimal transport

by

Chantel Maré

Submitted in partial fulfillment of the requirements for the degree

Master of Science

in the Department of Physics
in the Faculty of Natural and Agricultural Sciences

University of Pretoria
Pretoria

August 2021

Declaration

I, Chantel Maré, declare that the dissertation, which I hereby submit for the degree Master of Science at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE:

Acknowledgements

I thank my supervisor Prof Rocco Duvenhage for his help with many of the technical aspects of this dissertation and clarification of mathematical obscurities. I also thank my friends and family who have supported me throughout my academic career, in particular my mother for her unending love and support.

Summary

We begin with a brief overview of measure theory and the theory of optimal transport. We then proceed to study a special class of quantum states represented by quantum Markov semi-groups (QMS) on a finite dimensional C^*-algebra. We show that these semi-groups are ergodic and have a unique stationary state. We then proceed to define a notion of quantum detailed balance and show that these semi-groups satisfy this detailed balance condition with respect to the unique stationary state. This condition characterises the form of the generator of the QMS. Starting from the form of this generator we proceed to show how one can construct the operators of multiplication, gradient and divergence acting on a direct sum of Hilbert spaces. These notions are then used to obtain a quantum mechanical analog of the continuity equation for probability densities. We define a Riemannian manifold of density matrices and proceed to show that for a given metric, the time evolution of our quantum states can be written as gradient flow for the relative entropy functional. This is a direct quantum analog to the time evolution of probability densities on $\mathbb{R}^n$, which can be written as gradient flow for the Wasserstein metric.

Contents

Introduction

In this dissertation we shall study a notion of quantum optimal transport pertaining to a special class of quantum systems that satisfy a form of quantum detailed balance. The mathematical theory of optimal transport is currently an active area of ongoing research in many different research fields including pure mathematics, theoretical physics, computer science, operations research and engineering. Optimal transport has been studied in both purely theoretical and applied contexts. Some engineering focused applications include the study of fluid dynamics [Ott01, BB00], computer graphics [LS18] and signals processing [CGT18, CGGT17].

The classical optimal transport problem was originally formulated in 1781, by civil engineer Gaspard Monge, who was concerned with the problem of transporting a heap of soil from one location to another. This became known as the Monge formulation. The problem was revisited in 1942 by Soviet economist Leonid Kantorovich who was concerned with the problem of cargo transportation [Vil03]. Kantorovich reformulated the problem by introducing a relaxation on Monge's original formulation. This relaxation involved allowing mass to be split up in a way that is not allowed in the Monge formulation [Tho].

Optimal transport theory was further studied by Benamou and Brenier in [BB00] and by Jordan, Kinderlehrer and Otto in [Ott01, JKO98] which led to connections with the study of fluid dynamics, geometry, partial differential equations and functional analysis. In particular the Benamou-Brenier formulation provides a dynamical approach that considers smooth paths in the space of probability densities which satisfy a continuity equation. It is from this approach that we can begin to construct a quantum or non-commutative formulation of optimal transport.

We shall look specifically at how the evolution equations of density matrices can be viewed as gradient flow with respect to a metric that is analogous to the 2-Wasserstein metric. It was first shown by Felix Otto in [Ott01] that the evolution of classical probability distributions on $\mathbb{R}^n$ can be viewed as gradient flow with respect to the 2-Wasserstein metric. This point of view offers quantitative insight into the behaviour of such equations [CM14]. In quantum mechanics the probability distributions are replaced by density matrices. A density matrix is simply a positive trace class operator ρ that acts on some Hilbert space $\mathcal{H}$, such that $\mathrm{Tr}(\rho) = 1$. In the finite dimensional case the density matrix is then simply a matrix of unit trace with strictly positive eigenvalues. These density matrices serve as analogues to probability densities within the context of non-commutative probability theory. This point of view is originally due to Irving Segal in his work on non-commutative extensions of abstract integration theory [Seg53]. Segal's work starts with the observation that the set of all complex valued, bounded functions on a given domain, are measurable with respect to some σ-algebra. These functions, equipped with the complex conjugate involution, then form a commutative von Neumann algebra, and any probability measure induces a state on the algebra. A generalization is then obtained by dropping the requirement that the algebra be commutative. One then obtains a non-commutative probability space. The structure of these spaces turns out to be useful in the study of quantum mechanics.

Consider $\mathcal{B}(\mathcal{H})$, the set of all bounded linear operators on a Hilbert space $\mathcal{H}$. The set $\mathcal{B}(\mathcal{H})$ is well known to be a von Neumann algebra. We run into two problems when we now want to come up with a non-commutative, or quantum, analogue of the 2-Wasserstein metric. The first is the problem of decoupling. In classical optimal transport theory the objects of study are the set of couplings of two probability densities ρ_0 and ρ_1, which are decoupled into transportation plans. The 2-Wasserstein distance is defined in terms of these couplings and an associated cost function, $|x-y|^2$, which is simply the Euclidian distance on $\mathbb{R}^n$, the underlying metric space. In our finite dimensional quantum setting, the notion of a coupling of probability distributions ρ_0 and ρ_1 now takes the form of a density matrix κ on $\mathcal{H}\otimes\mathcal{H}$, such that by taking the partial trace over the first and second factor, we obtain the density matrices ρ_0 and ρ_1 respectively. In the general case there it is somewhat difficult to decompose such coupling of tensor products into a transportation plan. The second problem we encounter is that there is no underlying metric space in the quantum mechanical setting, thus it is not clear how to define a cost function. Some work has been done to work with transport plans directly and to use a cost function analogous to the Euclidean distance squared [Duv20, BV01].

The first problem is solved by the previously mentioned Benamou-Brenier formulation, which uses a dynamical approach to define couplings in terms of smooth paths on a Riemannian manifold of probability densities, that satisfy a continuity equation. This approach also simultaneously solves the second problem. With a Riemannian manifold in hand, we have an underlying metric space and can construct a cost function. Benamou and Brenier's approach naturally leads to a quantum analogue of the classical *Fokker-Planck equation*, a partial differential equation used to describe the time evolution of the velocity probability density for a particle subject to random forces, see [Ott01] for example.

The *Fokker-Plank* equation describes the time evolution of a probability density for a random process, and pertains to a large variety of physical systems in which randomness plays a role. Such systems include the diffusion of gas particles through a porous medium [Ott01] as well as in biological systems [SLE09]. Recently the Fokker-Plank equation has also been used in machine learning application [RM17]. A variational formulation of the Fokker-Plank equation first appeared in [JKO98] due to Jordan, Kinderlehrer and Otto. The time step was determined from the Wasserstein distance on the underlying probability distribution. This formulation demonstrated how the dynamics of the evolution equation can be related to the gradient flow for the free energy of the system with respect to the Wasserstein metric. This in turn revealed a previously unexplored relationship between the Fokker-Plank equation and the free energy functional associated to the system. Moreover they were able to deduce relevant properties of the evolution equation that were previously unexplored. One such important property was that of hypercontractivity, a property that has been successfully applied to construct proofs of various results in quantum information theory [Mon12].

The principal example of a physically interesting evolution equation in non-commutative spaces, for which the Wasserstein metric point of view has proven useful, is the *fermionic Fokker-Plank equation*, introduced by Gross in [Gro75] to study Sobolev inequalities that arise naturally in the construction of quantum fields. The fermionic Fokker-Plank equation is a quantum mechanical analogue of the classical partial differential equation, which describes the time evolution of a fermionic quantum system. It shares the same hypercontractivity properties as the usual Fokker-Plank equation, making it useful to the study of quantum information theory.

In this dissertation we shall follow closely the work of Carlen and Maas in [CM14] to show how the time evolution of a certain class of states can be written as gradient flow of the relative entropy functional with respect to a steady state σ. We start with a basic review of the mathematical tools that are needed. Chapter 1 provides a brief overview of measure theory, from basic notation to product measures. Chapter 2 serves as an introduction to the study of optimal transport and the connection to the Wasserstein metric. Chapter 3 introduces the framework of C^*-algebras to describe the dynamics of open quantum systems and introduces much of the terminology that will be used later on. Chapter 4 introduces the concept of detailed balance and defines a specific notion of quantum detailed balance that will be used throughout the rest of the dissertation. Chapter 5 introduces and characterizes the specific structure of the generators of the Quantum Markov semigroups, that will be used to obtain many of the results in the next two chapters. In Chapter 6 we state the continuity equation and define the notions of gradient and divergence that will be used to construct a quantum mechanical analogue of this equation. Chapter 7 starts with a brief review of differential geometry and the 2-Wasserstein metric. We then use the results of Chapters 5 and 6 to show how the time evolution of our quantum states can be written as gradient flow of the relative entropy functional, analogous to the classical case.

Chapter 1

Measure theory background

In this dissertation we aim to study the connection between optimal transport theory and the dynamics of open quantum systems. Optimal transport theory is the domain of mathematical analysis and probability theory. To facilitate a study of optimal transport some knowledge of measure theoretic probability is required. Measure theory is the study of measures, mathematical objects that intuitively serve to generalize the concepts of length, area and volume of sets. Measure theory allows for integration over more general sets other than the usual integration over $\mathbb{R}^n$ or $\mathbb{C}^n$.

This Chapter is included to serve as a brief review measure theory, containing only the very basics of this broad field of study. The reader is referred to [Rud06] and [Coh13] for a more complete exposition on this topic. The following is adapted from Chapter 1 of [Rud06].

1.1 Concept of measurability

The main object of study in measure theory is that of so called measurable functions. These function which we will define and study more closely in the next few sections, share many of their properties with that of the class of continuous functions. Since the concepts encountered in the study of measure theory can be rather abstract, it is helpful to present them in a way that is analogous to more intuitive concepts encountered in the study of continuous functions. For example the concepts of topological space, open sets, and continuous functions are closely related to that of measurable spaces, measurable sets, and measurable functions respectively. The following definitions clearly show this relation.:

Definition 1.1.1.

1. A collection τ of subsets of a set X is called a *topology* in X if τ satisfies

   i $\emptyset \in \tau, \quad X \in \tau$.

   ii If $V_i \in \tau$, then finite intersections are in τ i.e., $\bigcap_{i=1}^{n} V_i$.

   iii If $\{V_\alpha\}$ is any arbitrary collection of elements of τ, then their union is in τ i.e., $\bigcup_\alpha V_\alpha \in \tau$.

2. If τ is a topology in X, then we call the set (X, τ) a *topological space* and the members of τ are called open sets.

3. If X, Y are topological spaces, then a function $f : X \to Y$ is said to be *continuous* if $f^{-1}(V)$ is an open set for all open sets $V \in Y$.

The above definition should be familiar to both mathematicians and physicists who have taken graduate analysis courses. Note that this definition of a continuous function given above is equivalent to the $\epsilon - \delta$ definition usually encountered in a first analysis course. The above definition is of course more general as it pertains to any topological space whereas the $\epsilon - \delta$ definition only pertains to metric spaces.

Definition 1.1.2.

1. A collection $\mathcal{M}$ of subsets of a set X is a *σ-algebra* in X if $\mathcal{M}$ satisfies

   i $X \in \mathcal{M}$.

   ii If $A \in \mathcal{M}$, then the complement is also in $\mathcal{M}$ i.e $A^C \in \mathcal{M}$.

   iii If A is the countable union of elements in $\mathcal{M}$, then A is also in $\mathcal{M}$ i.e If $A = \bigcup_{n=1}^{\infty} A_n$, $A_n \in \mathcal{M}$ for all n,, then $A \in \mathcal{M}$.

2. If $\mathcal{M}$ is a σ-algebra in X, then we call the pair $(X, \mathcal{M})$ a *measurable space* and the members of $\mathcal{M}$ are called *measurable sets.* We sometimes use X to denote the measurable space when the associated collection of subsets $\mathcal{M}$ is clear from context.

3. If X is a measurable space and Y is a topological space, then a function $f : X \mapsto Y$ is said to be *measurable* if $f^{-1}(V)$ is a measurable set in X for all open sets $V \in Y$.

To further illustrate the analogy between measurable and continuous functions we state a simple theorem about their composition.

Theorem 1.1.3. *Let Y, Z be topological spaces and let $g : Y \mapsto Z$ be a continuous function.*

1. *If X is a topological space and $f : X \mapsto Y$ is continuous, then $h : X \mapsto Z$ defined by $h = g \circ f$ is continuous.*

2. *If X is a measurable space and $f : X \mapsto Y$ is continuous, then $h : X \mapsto Z$ defined by $h = g \circ f$ is measurable.*

The first result is the familiar statement that continuous functions of continuous functions are continuous. The second result simply states that continuous functions of measurable functions are measurable.

The existence of σ-algebras is of course central to the study of measure theory. We now prove an important theorem about the existence of σ algebras which leads to the definition of Borel sets, a class of mathematical objects ubiquitous in the study of measures.

Theorem 1.1.4. *Let X be a topological space and let F be any collection of subsets of X. Now there exists a smallest σ-algebra $\mathcal{M}^*$ in X such that F is contained in $\mathcal{M}^*$.*

Proof. Let Ω be the family of all σ-algebras $\mathcal{M}$ in X that contain F. Clearly there must exist at least one such σ-algebra since by definition the collection of all subsets of X is a σ-algebra containing F. Thus Ω is not empty. Now define

$$\mathcal{M}^* = \bigcap_{M \in \Omega} M, \tag{1.1.1}$$

then clearly $F \subset M^*$ and M^* is in every σ-algebra containing F. All that is left to do is to now show that $\mathcal{M}^*$ is indeed a σ-algebra by checking that it satisfies the properties outlined in definition 1.1.2.

i By definition $X \in M$ for all $M \in \Omega$ hence $X \in \mathcal{M}^*$.

ii If $A \in \mathcal{M}^*$, then $A \in M$ for all $M \in \Omega$. By definition, then $A^C \in M$ for all $M \in \Omega$ thus $A^C \in \mathcal{M}^*$.

iii Let $A = \bigcup_{n=1}^{\infty} A_n$, $A_n \in \mathcal{M}^*$ for all n. Now for all n we have that $A_n \in M$ for all $M \in \Omega$. Now again by definition $A \in M$ for all $M \in \Omega$. Thus $A \in \mathcal{M}^*$

□

We call the σ-algebra given by equation (1.1.1) the σ-algebra generated by F. Now consider the collection of all open sets in X. By Theorem 1.1.4 there exists a smallest σ-algebra $\mathcal{B}$ such that every open set in X is contained in $\mathcal{B}$. The elements of $\mathcal{B}$ are known as the Borel sets of X. In other words the Borel sets of X are the measurable sets of the σ-algebra generated by the collection of all open sets of X. The Borel sets play a fundamental role in measure theory and intuitively allow us to connect measure theory and topology. Note that we started with a topological space X, and have now constructed a measurable space $(X, \mathcal{B})$. The Borel sets also allow us to establish a connection with continuous functions.

Corollary 1.1.5. *If Y is a topological space and $f : X \mapsto Y$ is a continuous function, then f is Borel measurable.*

Proof. f is continuous so $f^{-1}(V)$ is an open set in X for every open set $V \in Y$. $\mathcal{B}$ contains every open set in X, hence $f^{-1}(V) \in \mathcal{B}$. □

Functions that are Borel measurable are sometimes known as Borel functions.

1.2 Simple functions

In this section we introduce the concept of simple functions which will form the basis of abstract integration of functions with respect to some measure.

Definition 1.2.1. The *characteristic function* is defined by

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Definition 1.2.2. A real or complex valued function s whose range consists of a finite set of points is called a *simple function*. Non-negative simple functions are those functions whose range consist of a finite subset of $[0, \infty)$.

From the above two definitions we can obtain a formula to describe any simple function. Let $\{\alpha_1, ..., \alpha_N\}$ be the distinct values that are in the range of a simple function s, and let $A_i = \{x : s(x) = \alpha_i\}$, then clearly we have

$$s = \sum_{i=1}^{N} = \alpha_i \chi_{A_i}. \tag{1.2.1}$$

From equation (1.2.1) it is easy to see that S is measurable if and only if the sets A_i are themselves measurable. Next we state an important theorem relating simple functions to arbitrary positive valued measurable functions. The proof is somewhat technical and not particularly enlightening, it can be found in [Rud06].

Theorem 1.2.3. *Let X be a measurable space and let $f : X \mapsto \mathbb{R}^+$ be a measurable function. There exist simple measurable functions s_n on X such that*

1. $0 \leq s_1 \leq ... \leq s_n \leq f$.
2. $s_n(x) \to f$ *as* $n \to \infty$, *for every* $x \in X$. *That is to say* s_n *converges pointwise to* f.

1.3 Basic properties of measures

Before we can say anything about the integration of functions with respect to a measure, we must of course first define exactly what we mean by a measure.

Definition 1.3.1. Let $\mathcal{M}$ be a σ-algebra.

1. A function μ defined on $\mathcal{M}$ is called a *positive measure* if $\mathrm{Ran}(\mu) = [0, \infty]$, and μ is countably additive, i.e

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

   for any collection of pairwise disjoint sets $\{A_i\}$ in $\mathcal{M}$.

2. A measure space is a measurable space which has a positive measure defined on the σ-algebra of its measurable sets.

Throughout this report we shall simply use the term 'measure' to refer to a positive measure.

The following theorem lists some of the most important properties of a positive measure. These properties follow almost immediately from the definition, the proof can be found in [Rud06].

Theorem 1.3.2. *Let μ be a positive measure on a σ-algebra $\mathcal{M}$ Then*

1. $\mu(\emptyset) = 0$.
2. $\mu(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} \mu(A_i)$, *for any collection of pairwise disjoint sets* $\{A_i\}$ *in* $\mathcal{M}$.
3. *If* $A \subset B$ *then* $\mu(A) \leq \mu(B)$, *for* $A, B \in \mathcal{M}$.

4. *If* $A = \bigcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{M}$ *for all* i, *and* $A_1 \subset A_2 \subset \ldots$, *then* $\mu(A_n) \to \mu(A)$ *as* $n \to \infty$.

5. *If* $A = \bigcap_{i=1}^{N} A_i$, $A_i \in \mathcal{M}$ *for all* $i = 1, \ldots, N$, *and* $A_1 \supset A_2 \supset \ldots$, *then* $\mu(A_n) \to \mu(A)$ *as* $n \to \infty$, *provided* $\mu(A_1)$ *is finite.*

The properties above formalize the intuitive concept of a measure as being something that describes the "size" of a set. Note that some sets can have a measure of ∞, capturing the idea that some sets, like the real line, are infinitely large. Property 1 states that the empty set has measure zero which captures the idea that the empty set has no elements in it. Property 2 is called finitely additive and is simply the idea that we can add the sizes of disjoint sets together to find the size of their union. Property 3 states that if A is a subset of B, then A must be smaller or equal to B in size as determined by the measure μ. With the definition of a measure in place we can now give the definition of a measure space.

Definition 1.3.3. Let X be a set, $\mathcal{M}$ be a σ-algebra on X and μ be a measure on $\mathcal{M}$. We call the triple $(X, \mathcal{M}, \mu)$ a *measure space*. Note that if $\mathcal{M}$ is clear from context we simply refer to μ as the measure on X.

1.4 Integration of positive functions

The goal of this section is to connect the abstract concepts of measure theory to the familiar notion of integration, and to some important formulas regarding integration with respect to some measure. First we have to define the extended real number line. The concept of infinity is encountered throughout measure and integration theory. We can construct sets of infinite measure, for example the real number line which has infinite length. We may also encounter infinity when dealing with the limits of real-valued functions. To deal with this we construct arithmetic operations on the extended positive real line $[0, \infty]$, in the following way:
We define addition by

$$a + \infty = \infty + a \quad \text{if} \quad 0 \leq a \leq \infty,$$

and multiplication by

$$a \cdot \infty = \infty \cdot a = \begin{cases} 0 & a = 0 \\ \infty & 0 < a \leq \infty \end{cases}$$

With these rules of arithmetic in $[0, \infty]$ in place we can proceed to abstract integration of positive functions. We start with a definition of how to integrate simple functions.

Definition 1.4.1. Let (X, M) be a measurable space, μ be a measure on $\mathcal{M}$, and let $s : X \to [0, \infty]$ be a simple function of the form

$$s = \sum_{i=1}^{N} \alpha_i \chi_{A_i},$$

where α_i are the distinct values of s. If $E \subset \mathcal{M}$, then we define

$$\int_E s d\mu = \sum_{i=1}^{N} \alpha_i \mu(A_i \cap E). \tag{1.4.1}$$

It is important to note here that we use the convention that $0 \cdot \infty = 0$ since we may have the case where $\alpha_i = 0$ and $\mu(A_i \cap E) = \infty$. Equation (1.4.1) is analogous to the usual Riemann sum, where we take the "size" of a set and multiplying with an average function value on that set. In the case of equation (1.4.1) we use the size of a given set where s takes on a value of α_i and go over all points in E where s takes on some value α_i.

Next we make use of Theorem 1.2.3 to define the integration of any positive function f with respect to a measure μ.

Definition 1.4.2. Let (X, M) be a measure space, μ be a measure on $\mathcal{M}$, and let $f : X \to [0, \infty]$ be measurable. We define

$$\int_E f d\mu = \sup \int_E s d\mu, \tag{1.4.2}$$

where the supremum is taken over all possible simple functions s such that $0 \leq s \leq f$.

The left hand side of equation (1.4.2) is called the Lebesgue integral of f with respect to μ and can be viewed as a generalization of the familiar Riemann integral. The next theorem lists some important properties of the integral we defined in equation (1.4.2). These properties follow directly from the definition of the integral and intuitively resemble the familiar properties of the Riemann integral. The proof can be found in [Rud06].

Theorem 1.4.3. *Assume all functions and sets are measurable.*

1. *If* $0 \leq f \leq g$, *then* $\int_E f d\mu \leq \int_E g d\mu$.
2. *If* $f \geq 0$ *and* $A \subset B$, *then* $\int_A f d\mu \leq \int_B f d\mu$.

3. *If* $f \geq 0$ *and* c *is a constant with* $0 \leq c < \infty$*, then*

$$\int_E cf d\mu = c \int_E f d\mu.$$

4. *If* $f(x)$*=0 for all* $x \in E$*, then* $\int_E f d\mu = 0$*, even if* $\mu(E) = \infty$*.*

5. *If* $\mu(E) = 0$*, then* $\int_E f d\mu = 0$*, even if* $f(x) = \infty$ *for every* $x \in E$*.*

Property 4 states that if our function takes on a value of zero everywhere on the set which we are integrating over, then the integral itself is zero. This is the case even if the set has an infinitely large measure. Conversely Property 5 states that if the set we are integrating over has measure zero, then the integral is zero, regardless of if the function goes to infinity on that set. This leads directly to our next discussion on the role of sets of measure zero. First we should state an important theorem that will be used in the next Chapter.

Theorem 1.4.4 (**Lebesgue's Monotone Convergence Theorem**)**.** *Let* $\{f_n\}$ *be a sequence of measurable functions on* X *and suppose that*

1. $0 \leq f_1(x) \leq f_2(x) \leq ... \leq \infty$ *for every* $x \in X$*,*

2. $f_n(x) \to f(x)$ *as* $n \to \infty$*, for every* $x \in X$*.*

Now f *is measurable, and*

$$\int_X f_n d\mu \to \int_X f d\mu$$

as $n \to \infty$*.*

1.5 Measure zero and almost everywhere

Let μ be a measure on a σ-algebra $\mathcal{M}$ and $E \in \mathcal{M}$. Let P denote some property that a function f might have, e.g. $f(x) > 0$. We say that the property P holds almost everywhere (a.e.) on E if the following condition holds: There exists some $N \in E$ such that $\mu(N)$ and P holds at every point in $E - N$. That is to say the points where P does not hold are exactly those points which belong to some subset of measure zero. An example is given below showing how this property can be applied.

Example 1.5.1. Let f, g be measurable sets on some σ-algebra $\mathcal{M}$, $E \subset \mathcal{M}$, and define

$$N = \{x : f(x) \neq g(x), x \in E\}.$$

We say that $f(x) = g(x)$ a.e. Notice that we essentially disregard sets of measure zero when integrating so that, even though the function values differ at some poits we still obtain

$$\int_E f d\mu = \int_E g d\mu.$$

1.6 Product measures

We have covered the basics of abstract integration on single measure spaces, and with these properties in place, we can easily construct product measures. Intuitively a product measure is similar to a Cartesian products of sets.

Definition 1.6.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. Let $X \times Y$ denote the usual Cartesian product of sets X and Y. A substet of $X \times Y$ is called a *rectangle with measurable sides* if it can be written as $A \times B$, $A \in \mathcal{A}$ and $B \in \mathcal{B}$. The collection of all such rectangles with measurable sides generates a σ-algebra. We call this σ-algebra the product of σ-algebras $\mathcal{A}$ and $\mathcal{B}$, and we denote the product by $\mathcal{A} \times \mathcal{B}$.

1.7 Probability spaces and L^p spaces

Lastly we introduce the concept of probability spaces, probability measures and L^p spaces.

Definition 1.7.1. A *probability space*$(X, \mathcal{M}, \mu)$ is a measure space such that $\mu(X) = 1$. The measure μ is now called a *probability measure.*

Here we restrict the range of the measure μ to the unit interval $[0, 1]$. The set X is the set of possible outcomes of a random variable and the elements of $\mathcal{M}$ are called the events. If $M \in \mathcal{M}$, then we call $\mu(M)$ the probability of event M. In this way we can formalize the intuitive notions of probability using the constructs of measure theory.

Definition 1.7.2. The space $L^p(X, \mathcal{M}, \mu)$ is the space of functions f on X such that

$$\int_X |f|^p d\mu < \infty,$$

equipped with the L^p norm

$$||f||_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}}. \tag{1.7.1}$$

For a proof that equation (1.7.1) indeed defines a norm see [Kre78].

Chapter 2

Optimal transport theory

2.1 Monge formulation

The classical optimal transport problem was originally formulated in 1781, by civil engineer Gaspard Monge, who was concerned with the problem of transporting a heap of soil from one location to another. The problem consisted of finding the optimal way of transporting the soil by minimizing the cost of doing so. This problem can easily be recast in terms of probability distributions by defining a transport map and a suitable cost function. The following is a brief exploration of both the Monge and Kantorovich problems in optimal transport theory, adapted from [Tho, Vil03].

First let's be clear about what we mean by a transport map. Consider two probability measures μ and ν on measurable spaces X and Y respectively. We then define a transport map T as follows:

Theorem 2.1.1. *An function $T : X \to Y$ is called a transport map from μ to ν if*

$$\nu(B) = \mu(T^{-1}(B)) \tag{2.1.1}$$

for all measurable sets $B \in Y$.

Here the inverse of T should be treated in the general sense of set values, that is to say if $T(x) = y$ then $x \in T^{-1}(y)$. If T is injective then we can rewrite equation (2.1.1) as simply $\mu(A) = \nu(T(A))$. It should be noted that, for the sake of generality, we usually prefer to work with the inverse function T^{-1}. For any ν-measurable set B and μ-measurable set $A = \{x \in X : T(x) \in B\}$, we have that $\mu(A) = \nu(B)$, that is to say all the mass of μ is transported to ν. This concept is illustrated in Figure 2.1 below.

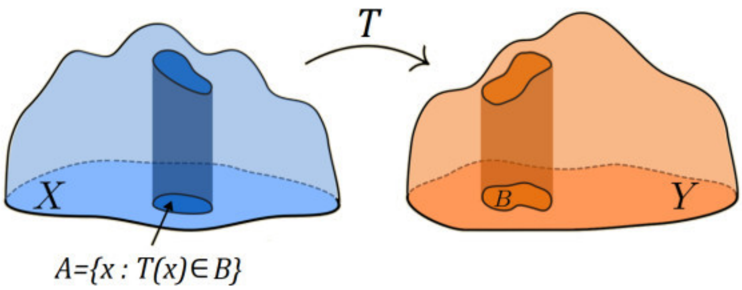

Figure 2.1: Monge transport map, from [Tho].

We write $\nu = T_{\#}\mu$ as shorthand to indicate a transport map from μ to ν, i.e. a mapping T which satisfies equation (2.1.1).
Concretely $T_{\#}\mu(A) = \mu(T^{-1}(A))$ for a μ-measurable set A.

With the definition of the transport map in hand we can prove some useful formulas pertaining to manipulating equations. The first is a change of variables formula:

Proposition 2.1.2. *Let $\mu \in \mathcal{P}(X), T : X \to Y$ and $f \in L^1$. Then*

$$\int_Y f(y)d(T_{\#}\mu)(y) = \int_X f(T(x))d\mu(x). \tag{2.1.2}$$

Proof. We start by considering any non-negative function $f : Y \to \mathbb{R}$. Now by Theorem 1.2.3 there exists simple functions $\{s_n\}$ such that

1. $0 \leq s_1 \leq s_2 \leq ... \leq f$,
2. $s_n \to f$ as $n \to \infty$.

Now since $f \in L^1$, we know that $f < \infty$ and thus we can apply Lebesgue's monotone convergence theorem (Theorem 1.4.4) to obtain

$$\int_Y s_n(y)d(T_{\#}\mu)(y) \to \int_Y f d(T_{\#}\mu)(y) \tag{2.1.3}$$

as $n \to \infty$. Recall from the definition of simple functions given by equation (1.2.1) that s_n can be written as

$$s_n(y) = \sum_{i=1}^{N} \alpha_i \chi_{U_i}$$

where $\{\alpha_i\}$ is the set of distinct values of s, $U_i = \{y \in Y : s(y) = \alpha_i\}$, and χ_{U_i} denotes the characteristic function of U_i. Now using equation (1.4.1) we have

$$\begin{aligned}\int_Y s(y)d(T_{\#}\mu)(y) &= \sum_{i=1}^{N} \alpha_i T_{\#}\mu(U_i) \\ &= \sum_{i=1}^{N} \alpha_i \mu(V_i) \\ &= \int_X s_n \circ T d\mu(x),\end{aligned}$$

where $V_i = T^{-1}(U_i) = \{x \in X : T(x) = y, y \in U_i\}$. Now if s_n satisfies the properties of Theorem 1.2.3 then so will $s_n \circ T$. Then we can again apply Theorem 1.4.4 to obtain

$$\int_X s_n \circ d\mu \to \int_X f(T(x))d\mu. \tag{2.1.4}$$

The result follows now from equations (2.1.3) and (2.1.4). Finally to complete the argument we generalize to signed functions by splitting the function into positive and negative parts, i.e. $f = f^+ - f^-$ thus proving the proposition for $f \in L^1$. □

Next we prove a composition rule for transport maps T and S:

Proposition 2.1.3. *Let $\mu \in \mathcal{P}(X), T : X \to Y, S : Y \to Z$.*

$$(S \circ T)_{\#}\mu = S_{\#}(T_{\#}\mu)). \tag{2.1.5}$$

Proof. Let $A \subset Z$. Now by successive application of equation (2.1.1) we have

$$\begin{aligned}S_{\#}(T_{\#}\mu))(A) &= T_{\#}\mu(S^{-1}(A)) \\ &= \mu(T^{-1}(S^{-1}(A))) \\ &= \mu(S \circ T)^{-1}(A) \\ &= (S \circ T)_{\#}\mu(A).\end{aligned}$$

□

Next we tackle the issue of existence properties for transport maps. It turns out that given probability measures μ and ν, a transport map T such that equation (2.1.1) is satisfied may not only be non-trivial but may also not even exist.

Example 2.1.4. Consider the discrete probability measures $\mu \in \mathcal{P}(X)$, $\mu = \delta_{x_1}$ and $\nu \in \mathcal{P}(Y)$, $\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$. Intuitively μ corresponds to having all the probability mass concentrated at a single point x_1, whereas ν corresponds to having the probability mass split up evenly between the points y_1 and y_2. Formally for any μ-measurable set A

$$\delta_{x_1}(A) = \chi_A(x_1) = \begin{cases} 0 & x_1 \notin A \\ 1 & x_1 \in A \end{cases}$$

From this we see that μ takes on discrete values $\{0, 1\}$ whereas ν takes on discrete values $\{0, \frac{1}{2}\}$. Clearly for any transport map $T : X \to Y$ we must have

$$\mu(T^{-1}(y_1)) \in \{0, 1\},$$

depending on whether $x_1 \in T^{-1}(y_1)$. However $\nu(y_1) \in \{0, \frac{1}{2}\}$ so no transport map satisfying equation (2.1.1) can exist.

There are two important cases where transport maps do exist [Vil03]:

1. When we have the discrete measures $\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$ and $\nu = \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i}$, and μ and ν have support of equal size, i.e supported on the same number of points.

2. When we have the absolutely continuous case $d\mu(x) = f(x)dx$ and $d\nu(y) = g(y)dy$.

Definition 2.1.5 (**The Monge Optimal Transport Problem**)**.** Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$.

$$\text{minimize } \mathbb{M}(T) = \int_X c(x, T(x))d\mu(x) \tag{2.1.6}$$

where the minimization is over all transport maps $T : X \to Y$ such that $\nu = T_{\#}\mu$.

The choice of cost function in equation (2.1.6) can significantly affect how difficult the problem is to be solved. Originally Monge considered the problem using the L^1 cost function $c(x, y) = |x - y|$. This turned out to be much more difficult to solve, and required stronger assumptions, than the case of the L^2 cost function $c(x, y) = |x - y|^2$, see [EG99]. The main difficulty in solving the Monge optimal transport problem is due to the constraint (2.1.1) being, in general, non-linear. As an illustration of this consider the case where we have μ and ν absolutely continuous with respect to the

Lebesgue measure, that is $d\mu(x) = f(x)dx$ and $d\nu(y) = g(y)dy$. Furthermore assume that the transport map $T : X \to Y$ is bijjective and that T, T^{-1} are differentiable. Now let's apply the change of variables formula (2.1.2) with the test function set to the constant function $h(y) = 1$. Then

$$\int_Y g(y)dy = \int_X f(x)dx.$$

If we treat the above as a simple change of coordinates then we see that the constraint (2.1.1) can be written as

$$f(x) = g(T(x)) \det |\nabla T(x)|. \tag{2.1.7}$$

where $\det |\nabla T(x)|$ is simply the Jacobian factor. Clearly this constraint is highly non-linear.

Example 2.1.4 servers to further illustrate another difficulty encountered in the Monge optimal transport problem, in that it does not allow the mass to be split up. In our example all the mass in X was concentrated at a single point x_1 whereas in Y the mass was split up between points y_1 and y_2. This splitting is not allowed for in the Monge formulation and hence no transport map exists. We shall investigate in the next section how the relaxation of this constraint leads to another formulation of optimal transport.

2.2 Kantorovich formulation

To allow for the probability mass to be split we consider a measure $\pi \in \mathcal{P}(X \times Y)$. Intuitively $d\pi(x, y)$ tells us how much of the mass at a point x is transferred to a point y. Of course we require that the total mass which is removed from any measurable set $A \subset X$ be equal to $\mu(A)$ whereas the total mass that is transferred to any measurable set $B \subset Y$ be equal to $\nu(B)$. Formally we have the constraints that

$$\pi(A \times Y) = \mu(A) \quad \pi(X \times B) = \nu(B). \tag{2.2.1}$$

Any measure $\pi \in \mathcal{P}(X \times Y)$ which satisfies the above is said to have first and second marginals μ and ν respectively. Notice that when we integrate over some test functions $f(x)$ and $g(y)$ we have

$$\int_{X \times Y} f(x)d\pi(x, y) = \int_X f(x)d\mu(x)$$

and

$$\int_{X\times Y} g(y)d\pi(x,y) = \int_Y g(y)d\nu(y)$$

Let $\Pi(\mu,\nu)$ denote the set of measures $\pi(x,y)$ which satisfy equation (2.2.1). We shall refer to $\Pi(\mu,\nu)$ as the set of transport plans between μ and ν. Note that some literature also refers to the transport plans $\pi \in \Pi(\mu,\nu)$ as couplings, see [DPT19, CM14, Ike20]. Given that X and Y are countable, it is easy to see that $\Pi(\mu,\nu)$ is non-empty, consider the trivial transport plan which transports every bit of probability mass around the point x to Y, proportional to $\nu(y)$.

To see how this formulation solves the problem presented by example 2.1.4 consider $\pi(x,y) = \frac{1}{2}\delta_{x_1}(\delta_{y_1} + \delta_{y_2})$. Clearly this measure will transport half of the mass concentrated at x_1 to y_1 and the other half to y_2 just as required.

We can formalize the Kantorovich optimal transport problem as follows:

Definition 2.2.1. (**The Kantorovich Optimal Transport Problem**)

$$\text{minimize } \mathbb{K}(\pi) = \int_{X\times Y} c(x,y)d\pi(x,y) \tag{2.2.2}$$

where the minimization is over all $\pi \in \Pi(\mu,\nu)$.

Example 2.1.4 clearly demonstrates that even if a transport map exists in the Kantorovich formulation, that does not imply a transport map exists in the Monge formulation. It is thus natural to ask what are the conditions necessary so that these two formulations do in fact coincide. It turns out that given a transport map T that is optimal for the Monge formulation we can construct a transport map π in the sense of the Kantorovich formulation.

Proposition 2.2.2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $A \subset X$, and $B \subset Y$. Let $T : X \to Y$ be an optimal transport map for Monge, that is $\nu = T_{\#}\mu$.*
Let $c(x,y)$ denote some cost function.
Define $d\pi = d\mu\delta_{y=T(x)}$. Then π is a transport map in the Kantorovich formulation and

$$\inf(\mathcal{K}) \leq \inf(\mathcal{M}).$$

Proof. To show that $\pi \in \Pi(\mu, \nu)$ we must show that the marginals given by equation (2.2.1) are achieved. For the first marginal we have

$$\begin{aligned}\pi(A \times Y) &= \int_A d\mu(x)\delta_{T(x)=y}(Y) \\ &= \mu(A)\delta_{T(x)\in Y} \\ &= \mu(A).\end{aligned}$$

The second line in the above calculation follows simply from the fact that $\mathrm{Ran}(T) \subset Y$, hence $\delta_{T(x)\in Y} = 1$ for all $x \in X$.

Now for the second marginal we have

$$\begin{aligned}\pi(X \times B) &= \int_X d\mu(x)\delta_{T(x)=y}(B) \\ &= \int_X d\mu(x)\delta_{T(x)\in B}.\end{aligned}$$

Now we will only have a non zero value in the integrand above for those values of $x \in X$ such that $T(x) \in B$. In other words the measure μ ultimately only acts on the inverse image of $T(x)$ on the subset B. Thus

$$\begin{aligned}\pi(X \times B) &= \mu(T^{-1}(B)) \\ &= T_{\#\mu}(B) \\ &= \nu(B),\end{aligned}$$

where the second line follows from the fact that T is an optimal transport map for Monge. Hence we have shown that $\pi \in \Pi(\mu, \nu)$.

Finally we note that

$$\begin{aligned}\int_{X\times Y} c(x,y)d\pi(x,y) &= \int_X \int_Y c(x,y)d\mu(x)\delta_{T(x)=y}dy \\ &= \int_X c(x,T(x))d\mu(x).\end{aligned}$$

The second line again follows from the fact that the integrand will only be nonzero for those points in Y which can be written as $y = T(x)$ for some $x \in X$. Recall that T was taken to be optimal for Monge so in fact we have

$$\inf(\mathcal{K}) \leq \int_X c(x,T(x))d\mu(x) = \inf(\mathcal{M}).$$

□

Kantorovich duality

The Kantorovich problem, being a linear minimization problem with convex constraints [Tho], admits a dual formulation. As usual let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$. Let $\mathcal{J} : L^1(\mu) \times L^1(\nu) \to \mathbb{R}$ be a mapping defined by:

$$\mathcal{J}(\varphi, \psi) = \int_X \varphi d\mu + \int_Y \psi d\nu. \tag{2.2.3}$$

Let

$$\Phi_c = \{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu) : \varphi(x) + \psi(y) \leq c(x, y)\} \tag{2.2.4}$$

for some specific cost function $c(x, y)$.

Then the minimization problem of equation (2.2.2) can be equivalently recast as:

$$\inf(\mathcal{K}(\pi)) = \sup(\mathcal{J}(\varphi, \psi)), \tag{2.2.5}$$

where the supremum on the right hand side is taken over all $(\varphi, \psi) \in \Phi_c$.

The Kantorovich duality can be intuitively understood using the following analogy by Villani [Vil03]:

Suppose you own a given number of coal mines, and a given number of coal processing factories. The amount of coal recieved from each mine is fixed, as is the amount of coal to be recieved at each factory. Transporting coal from mine x to factory y has an associated cost $c(x, y)$. You, being a clever industrialist, of course want to minimize this cost function, in other words solve the Kantorovich optimal transport problem. A shipper comes to you and offers to load the coal for a price $\varphi(x)$, and unload the coal for a price $\psi(y)$, while you pay no additional transport cost. Of course the shipper must ensure the deal is in your interest by fixing their prices such that $\varphi(x) + \psi(y) \leq c(x, y)$. Now the Kantorovich duality tells us that the shipper can always find a price scheme such that you pay them just as much as you would have paid to do the shipping yourself. Economically this can be considered a win-win situation.

2.3 The Wasserstein distance

So far in our discussion of optimal transport theory we have not paid any close attention to our specific choice of cost function $c(x, y)$. For most practical purposes we choose the cost function to be the L^2 Euclidean distance squared between points x and y. In some cases we might use the L^1 cost function or some other L^p metric [KPT+17]. The common characteristic of the L^p cost functions is that they define a metric based on pointwise differences between the initial and final distribution of our transport problem. This presents a particular difficulty in applications of gradient descent based optimization problems. If one is trying to fit some parameterised curve f_t to a function f, then one can run into the difficulty where starting from a bad initialization of the problem leads to a constant cost function, hence a zero valued derivative [Tho]. In this section we will see how to define a distance function based on the optimal transport problem.

Consider a linear space X and define

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(x) : \int_X |x|^p d\mu(x) < \infty \right\}. \tag{2.3.1}$$

We refer to a probability measure $\mu \in \mathcal{P}_p(x)$ as having a bounded pth moment. Next we can define the Wasserstein distance as follows:

Definition 2.3.1. (The Wasserstein distance) Let $\mu \in \mathcal{P}_p(X)$ and $\nu \in \mathcal{P}_p(Y)$. Define

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{X \times Y} |x - y|^p d\pi(x, y) \right)^{\frac{1}{p}}. \tag{2.3.2}$$

The Wasserstein distance is a true metric, in that it satisfies all the properties of symmetry, non-negativity and the triangle inequality [Vil03]. For $p = 2$, equation (2.3.2) defines the 2-Wasserstein metric. We can rewrite the definition in equation (2.3.2) above in a slightly more compact form by defining for any L^p cost function $c(x, y)$ and coupling $\pi \in \Pi(\mu, \nu)$ the associated cost:

$$C(\pi) = \int_{X \times Y} c(x, y) d\pi(x, y),$$

then the Wasserstein distance is the optimal transport cost given by

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} (C(\pi))^{\frac{1}{p}}. \tag{2.3.3}$$

Equation (2.3.3) now clearly characterizes the optimal transport problem as a linear optimization problem. The problem always admits solutions under regularity assumptions [DPT19]. The 2-Wasserstein distance has been studied extensively by Brenier [BB00] and has been shown to induce a Riemannian metric on the space of probability densities on $\mathbb{R}^n$. The geometric properties of this metric have been studied with regard to their applications to the study several fields including partial differential equations [Ott01], stochastic analysis [ASZ09], machine learning [PC+19], optimization [CLPS14] and synthetic Ricci curvature [LV09].

Chapter 3

C^*-algebras and the dynamics of open quantum systems

We provide here a brief introduction to the C^*-algebraic formulation of quantum mechanics that shall be used throughout the dissertation. The main reason to use this formulation is that, while it is mathematically equivalent to the conventional Hilbert space formulation frequently encountered in introductory physics texts, the C^*-algebra formulation provides a more powerful mathematical toolset to describe quantum systems. In particular the C^*-algebraic formulation allows us to look at and describe the dynamics of open quantum systems, that is systems that interact with their environment. Thus we can obtain a better description of the type of systems found in nature, which are usually not the isolated ideal systems studied in introductory textbooks [Sew02]. Of course these systems can be studied without making use of C^*-algebras. The main advantage of the C^*-algebra approach however, is that the mathematical structure of the theory provided us with some useful mathematical tools we can apply. Another advantage of this approach is that the correspondence of quantum mechanics with classical probability theory becomes much clearer, as classical probability theory can also be written in terms of C^*-algebras.

3.1 C^*-algebras

We describe here the basic properties of *-algebras and in particular C^*-algebras. The reader is referred to [Mur14] and [BR12] for a complete treatment of the topic. Basic familiarity with vector spaces and inner product spaces are assumed. All vector spaces are assumed to be over the field of complex numbers.

Definition 3.1.1. An *algebra* $\mathcal{A}$ over $\mathbb{C}$ is a vector space equipped with a bilinear map

$$\mathcal{A} \times \mathcal{A} \to \mathcal{A} : (a, b) \mapsto ab,$$

such that the following properties hold:

1. $a(bc) = (ab)c \quad a, b, c \in \mathcal{A}$
2. $a(b + c) = ab + ac \quad a, b, c \in \mathcal{A}$
3. $\alpha\beta(ab) = (\alpha a)(\beta b)$, for all $\alpha, \beta \in \mathbb{C}$.

If an algebra $\mathcal{A}$ admits an element $\mathbf{1}$ such that $\mathbf{1}A = A\mathbf{1} = A$ for all $A \in \mathcal{A}$, then $\mathcal{A}$ is called a unital algebra and the element $\mathbf{1}$ is called the unit.

A normed algebra is an algebra that admits a norm, in other words to each element $a \in \mathcal{A}$ there is an associated real number $||a||$ that satisfies the usual properties of a norm i.e., for all $a, b \in \mathcal{A}$ and $\alpha \in \mathbb{C}$ we have:

1. $||a|| \geq 0$
2. $||a|| = 0$ if and only if $a = 0$
3. $||\alpha a|| = \alpha||a|| \quad \alpha \in \mathbb{C}$
4. $||a + b|| \leq ||a|| + ||b|| \quad a, b \in \mathcal{A}$
5. $||ab|| \leq ||a|| \cdot ||b||$

An algebra $\mathcal{A}$ is said to be *complete* if every Cauchy sequence in $\mathcal{A}$ coverges to a limit that is also in $\mathcal{A}$. A Banach algebra is a complete normed algebra.

Definition 3.1.2. A **-algebra* is an algebra $\mathcal{A}$ together with a map

$$* : \mathcal{A} \to \mathcal{A} : a \mapsto a^*,$$

that satisfies the following properties:

1. $(a^*)^* = a \quad a \in \mathcal{A}$
2. $(ab)^* = b^*a^* \quad a, b \in \mathcal{A}$
3. $(\alpha a + \beta b)^* = \bar{\alpha}a + \bar{\beta}b \quad a, b \in \mathcal{A}, \alpha, \beta \in \mathbb{C},$

where $\bar{\alpha}$ denotes the complex conjugate of α. The operator $*$ is known as an involution.

An element A of a *-algebra $\mathcal{A}$ is said to be *self-adjoint* if $A^* = A$.

Definition 3.1.3. A *C^*-algebra* is a Banach *-algebra with the additional property that

$$||a^*a|| = ||a||^2.$$

The set of $n \times n$ complex valued matrices, denoted by $\mathcal{M}_n(\mathbb{C})$ is an example of a C^*-algebra with the identity matrix as the unit, the involution given by the taking the conjugate transpose, and using the operator norm. In fact any finite dimensional C^*-algebra is of the form $\mathcal{M}_n(\mathbb{C}) \bigoplus \mathcal{M}_n(\mathbb{C})$, where $\bigoplus$ is used to denote the direct sum. This is a direct consequence of the Gelfand-Naimark Theorem, see section 3.4 of [Mur14].

Definition 3.1.4. A *state* on a C^* algebra is a positive linear functional

$$\mu : \mathcal{A} \to \mathbb{C}$$

such that

1. $\mu(a^*a) \geq 0$ for all $a \in \mathcal{A}$
2. $||\mu||$=1

Next we state a definition regarding the spectrum of a C^*-algebra. Since we will only be concerned with the finite dimensional case, we may treat our C^*-algebra as a sub-algebra of $\mathcal{M}_n(\mathbb{C})$.

Definition 3.1.5. Let $\mathcal{A}$ be a C^*-algebra with unit $\mathbf{1}$. We define the *resolvent* of $a \in \mathcal{A}$ as follows:

$$r(a) = \{\lambda \in \mathbb{C} \mid A - \lambda\mathbf{1} \quad \text{is invertible in } \mathcal{A}\},$$

and correspondingly we define the spectrum $\sigma(a)$ as the complement of the resolvent set

$$\sigma(a) = \mathbb{C} - r(a).$$

For the case where our C^*-algebra is simply $\mathcal{M}_n(\mathbb{C})$ we note that $\sigma(a)$ is simply the set of eigenvalues of the matrix a.

Definition 3.1.6. Let $\mathcal{A}$ be a C^*-algebra. An element $a \in \mathcal{A}$ is said to be *positive* if it is self-adjoint and

$$\sigma(a) \in \mathbb{R}^+.$$

and is said to be *strictly positive* if

$$\sigma(a) \in \mathbb{R}^+/\{0\}.$$

Again thinking of the algebra $\mathcal{M}_n(\mathbb{C})$, we note that a strictly positive element simply means a matrix a that has strictly positive eigenvalues. Such matrices are also called *non–degenerate*. We denote strictly positive elements by $a > 0$.

Next we state some properties of positive elements that we will make use of.

1. Let a be a positive element of some C^*-algebra. Then a is self-adjoint, i.e $a^* = a$.

2. The set of positive elements of a C^*-algebra form a convex cone.

3. Any self-adjoint element of a C^* algebra can be decomposed into a unique sum of positive elements. In other words let $a \in \mathcal{A}$ then we can write
$$a = a_+ - a_-,$$
where $a_+ \geq 0$ and $a_- \geq 0$.

3.2 Algebraic formulation of quantum mechanics

In this section we will briefly review the standard formulation of quantum mechanics, that is, the state vector and density matrix formulation and discussed in most undergraduate texts. We then proceed to the limitations of this formulation and how these limitations are overcome by the more general algebraic formulation that will be used in the rest of the text.

A quantum system whose state is known exactly and is described by a state vector $|\psi\rangle$ is called a pure state. Some quantum systems, however, cannot be described by such a single state vector. If we have a quantum system that arises as the result of a random process or a system that is not isolated from its environment, so called open systems. Then we cannot describe the system as single state vector, since we have incomplete information about the system. Examples of such systems include the polarization of photons from a natural light source or the kinetic energy of atoms of a beam emitted by a furnace at some temperature [CTDL91].

For the rest of this dissertation we will make use of the bra-ket notation to denote vectors as is common practice in the study of quantum mechanics.

We can describe such a system as an ensemble of possible states where the system is in the state $|\psi_\alpha\rangle$ with probability p_α. In our finite dimensional setting we of course have $|\psi_\alpha\rangle \in \mathbb{C}^n$ and our underlying Hilbert space is $\mathcal{M}_n$.

Note that the $|\psi_\alpha\rangle$ need not form a basis and so are not necessarily orthogonal to each other. Such a system is said to be in a mixed state. Note that the uncertainty here is an underlying classical uncertainty and not a quantum effect [SW10]. To describe such systems we require the density operator formulation of quantum mechanics. Suppose we want to measure an observable A on the discrete ensemble $\{p_\alpha, |\psi_\alpha\rangle\}$. For every α we have $\langle A\rangle_\alpha = \langle\psi_\alpha|A|\psi_\alpha\rangle$. The expectation value of A over the whole ensemble is given by:

$$\begin{aligned}\langle A\rangle &= \sum_\alpha p_\alpha \langle A\rangle_\alpha \\ &= \sum_\alpha p_\alpha \operatorname{Tr}\big(|\psi_\alpha\rangle\langle\psi_\alpha|A\big)\end{aligned} \tag{3.2.1}$$

where we have used the fact that $\langle\psi_\alpha|A|\psi_\alpha\rangle = \operatorname{Tr}\big(|\psi_\alpha\rangle\langle\psi_\alpha|A\big)$.

If we now define

$$\rho = \sum_\alpha p_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| \tag{3.2.2}$$

and note that the trace is a linear operator then (3.2.1) becomes

$$\langle A\rangle = \operatorname{Tr}(\rho A). \tag{3.2.3}$$

The density operator ρ then allows us to find the expectation value of an observable A for an ensemble of quantum states.

Now recall that $\mathcal{M}_n$ is a C^*-algebra. Clearly ρ as defined by equation (3.2.2) is in $\mathcal{M}_n$. Formally we define the density operator as follows:

Definition 3.2.1. Let $\rho : \mathcal{H} \mapsto \mathcal{H}$ be a bounded linear operator over some Hilbert space $\mathcal{H}$. Then ρ is a called a *density operator* if the following two properties are satisfied

1. $\rho \geq 0$
2. $Tr(\rho) = 1$

In the context of our finite dimensional setting ρ simply corresponds to a self-adjoint matrix with strictly positive eigenvalues whose diagonal elements sum to one. The second property corresponds to the fact that probabilities should sum to one. To see how the notion of a density operator connects to that of a state on an algebra, we show there is in fact a one-to-one correspondence between a density matrix ρ and a state μ on the algebra $\mathcal{M}_n$.

Proposition 3.2.2. *The functional μ defined by $\mu(A) = \mathrm{Tr}(\rho A)$ is a state on $\mathcal{M}_n$.*

Proof. Clearly μ is linear, by the linearity of the trace. Using the fact that ρ is self-adjoint and positive we see that $\rho^{1/2}$ is well defined and self-adjoint. Then making use of the properties of the trace we obtain

$$\begin{aligned}
\mu(A^*A) &= \mathrm{Tr}(\rho A^*A) \\
&= \mathrm{Tr}(\rho^{1/2}\rho^{1/2}A^*A) \\
&= \mathrm{Tr}(\rho^{1/2}A^*A\rho^{1/2} \\
&= \mathrm{Tr}((A\rho^{1/2})^*A\rho^{1/2}) \\
&> 0.
\end{aligned}$$

Finally all that remains is to show that $||\mu|| = 1$. In our finite dimensional setting this is equivalent to showing that

$$\mu(\mathbf{1}) = 1.$$

This result immediately follows from the fact that $\mathrm{Tr}(\rho) = 1$:

$$\begin{aligned}
\mu(\mathbf{1}) &= \mathrm{Tr}(\rho\mathbf{1}) && (3.2.4) \\
&= \mathrm{Tr}(\rho) && (3.2.5) \\
&= 1. && (3.2.6)
\end{aligned}$$

□

Proposition 3.2.3. *Let μ be a state on $\mathcal{M}_n$. Then there exists a unique element $\rho \in \mathcal{M}_n$ such that ρ is a density matrix and*

$$\mu(A) = \mathrm{Tr}(\rho A),$$

for all $A \in \mathcal{M}_n$.

Proof. We will prove the proposition by constructing the relevant density matrix in a similar way to how it was constructed in the previous discussion of an ensemble of states. We start with the spectral decomposition of the matrix A:

$$A = \sum_{j,k=1}^{n} \langle V_j|AV_k\rangle |V_j\rangle\langle V_k|,$$

where $\{|V_j\rangle\}_{j=1,..,n}$ is the eigenbasis of A. Next we make use of the linearity of μ to obtain

$$\mu(A) = \sum_{j,k=1}^{n} \langle V_j|AV_k\rangle \mu\left(|V_j\rangle\langle V_k|\right). \tag{3.2.7}$$

Finally we make use of the fact that $\langle V_j|AV_k\rangle = \mathrm{Tr}\left(|V_j\rangle\langle V_k A|\right)$ and set

$$\rho = \sum_{j,k=1}^{n} \mu\left(|V_j\rangle\langle V_k|\right)|V_j\rangle\langle V_k|.$$

Then equation (3.2.7) becomes

$$\mu(A) = \mathrm{Tr}(\rho A),$$

using the linearity of the trace. Clearly ρ is unique as the eigenbasis of A is unique. □

The density operator formalism is mathematically equivalent to the state vector formalism of quantum mechanics, however it is more general as it allows us to describe a wider class of systems [SW10].

The density matrix can be viewed as the quantum mechanical analog of a classical probability distribution. This view is justified when we consider the idea of performing measurements on a quantum system. A measurement performed on a quantum system can be thought of as some physical process that generates from the state of the system, a probability distribution for the set of possible outcomes of the measurement. For most applications of quantum information processing we are interested only in the probability of obtaining a specific outcome, not in the actual values of these outcomes or even the state of the system post-measurement [FVDG99].

Now suppose we have a set of possible outcomes on a system, then for each possible outcome k we want to define a function that takes as its input the state of the system described by the density operator ρ and gives as output the probability of measuring the outcome k. This is the main idea behind the so called positive operator valued measure (POVM) formalism. We assign to each outcome k an operator E_k which can be applied to ρ so as to obtain the probability of measuring the outcome k. In particular, taking the trace of the product of E_k with ρ gives the probability of measuring outcome k given that the state was ρ. The probability of measuring the outcome k is then given by the expectation value of the operator E_k

$$p_k = \text{Tr}(E_k \rho). \tag{3.2.8}$$

The linearity of the trace operator plays an important role here. We expect the possible outcomes to be independent of each other. So the probabilities of obtaining different outcomes should add linearly. Note that k is only used to label the operator E_k, and that E_k does not depend on the value of k in any way. This also explains the terminology "density operator" since it is used to obtain a probability density function for a set of measurement outcomes. The reader is referred to [Bra99] for a further exposition on the theory of measurement operators in quantum mechanical systems.

3.3 Basic notation and the modular operator

Now that we have the algebraic formalism of quantum mechanics in place we can set up the basic notation and terminology that will be used throughout the rest of the dissertation. Let $\mathcal{A}$ denote a finite dimensional algebra C^*-algebra with unit $\mathbf{1}$. The Gelfand-Naimark theorem, a central result in the study of operator algebras, states that we may regard $\mathcal{A}$ as a subalgebra of a matrix algebra $\mathcal{M}_n$ [Mur14]. Let τ denote the normalized trace on $\mathcal{A}$ such that for $A \in \mathcal{A}$ we have

$$\tau(A) = \frac{\text{Tr}(A)}{\text{Tr}(\mathbf{1})}.$$

Let $\mathcal{H}_{\mathcal{A}}$ denote the Hilbert space with the inner product defined by

$$\langle A, B \rangle_{\mathcal{H}_{\mathcal{A}}} = \tau[A^* B].$$

The inner product defined above is known as the GNS inner product. A state φ on $\mathcal{A}$ is said to be a faithful state if $\varphi(A^*A) = 0 \implies A = 0$. Let $\mathcal{G}_+$ denote the set of strictly positive density matrices on $\mathcal{A}$. We shall consider only the faithful states on $\mathcal{A}$ that are given by these strictly positive density matrices. This sets up the basic notation we will make use of. Now we define and state the properties of the modular operator.

Definition 3.3.1. Let $\sigma \in \mathcal{G}_+$ and define the *modular operator* Δ_σ on $\mathcal{H}_\mathcal{A}$ by

$$\Delta_\sigma(A) = \sigma A \sigma^{-1},$$

for all $A \in \mathcal{A}$.

Associated to the modular operator is the modular generator defined by

$$h = \ln \sigma.$$

Note that h is well defined since σ is diagonalizable. Clearly h is self-adjoint since σ is strictly positive.

Definition 3.3.2. The *modular automorphism group* α_t on $\mathcal{A}$ is defined by

$$\alpha_t(A) = e^{ith} A e^{-ith},$$

for $A \in \mathcal{A}$ and $t \in \mathbb{R}$.

From simple algebra we see that $(\Delta_\sigma)^t = \alpha_{it}$. The characterization of quantum Markov semigroups using the modular automorphism group was first studied by Davies in [Dav74]. There it was shown that a class of quantum Markov semigroups arises naturally from a quantum system whose dynamics are governed by an internal Hamiltonian h and is coupled to a heat bath. It was found that this class of quantum Markov semigroups satisfies the property that the semigroup commutes with the modular operator, i.e commutes with the modular automorphism group. The density matrix σ is then given by the usual canonical ensemble

$$\sigma = \frac{e^{-h}}{\mathrm{Tr}(e^{-h})}.$$

It was further shown by Alicki in [Ali76] that the property of commutation with the modular operator plays an important role in characterizing a notion of quantum detailed balance. We shall make use of this particular notion of detailed balance in Chapter 4. We shall make use of the following property of linear operators:

Definition 3.3.3. Let $\mathcal{K}$ be a linear operator on $\mathcal{A}$. Then $\mathcal{K}$ is said to be *positivity preserving* if for all $A \in \mathcal{A}$, $A \geq 0 \implies \mathcal{K}A \geq 0$.

Definition 3.3.4. Let $\mathcal{K}$ be a linear operator on $\mathcal{A}$. Then $\mathcal{K}$ is said to be *self-adjointness preserving* if for all $A \in \mathcal{A}$, $(\mathcal{K}A)^* = \mathcal{K}A^*$, or equivalently $(\mathcal{K}A)^* = \mathcal{K}A$ whenever A is self-adjoint.

Proposition 3.3.5. *The modular operator is self-adjoint with respect to the inner product on $\mathcal{H}_{\mathcal{A}}$ and positive.*

Proof. First we show that $\Delta_\sigma = \Delta_\sigma^\dagger$.

$$\begin{aligned}\langle A, \Delta_\sigma B\rangle_{\mathcal{H}_{\mathcal{A}}} &= \mathrm{Tr}[A^*\Delta_\sigma B] \\ &= \mathrm{Tr}[A^*\sigma B\sigma^{-1}] \\ &= \mathrm{Tr}[\sigma^{-1}A^*\sigma B] \\ &= \mathrm{Tr}[\Delta_\sigma^{-1}A^*B] \\ &= \mathrm{Tr}[(\Delta_\sigma A)^*B] \\ &= \langle \Delta_\sigma A, B\rangle_{\mathcal{H}_{\mathcal{A}}},\end{aligned}$$

where we have made use of the fact that that

$$(\Delta_\sigma A)^* = (\sigma A\sigma^{-1})^* = \sigma^{-1}A^*\sigma = \Delta_\sigma^{-1}A^*. \tag{3.3.1}$$

Next we show that the modular operator is positive with a simple calculation:

$$\begin{aligned}\langle A, \Delta_\sigma A\rangle_{\mathcal{H}_{\mathcal{A}}} &= \mathrm{Tr}[A^*\Delta_\sigma A] \\ &= \mathrm{Tr}[A^*\sigma A\sigma^{-1}] \\ &= \mathrm{Tr}[\sigma^{1/2}\sigma^{1/2}A\sigma^{-1/2}\sigma^{-1/2}A^*] \\ &= \mathrm{Tr}[\sigma^{1/2}A\sigma^{-1/2}\sigma^{-1/2}A^*\sigma^{1/2}] \\ &= \mathrm{Tr}[(\sigma^{1/2}A\sigma^{-1/2})(\sigma^{1/2}A\sigma^{-1/2})^*] \\ &= \mathrm{Tr}[|\sigma^{1/2}A\sigma^{-1/2}|^2] \\ &> 0.\end{aligned}$$

□

Proposition 3.3.6. *Any positivity preserving operator is self-adjointness preserving.*

Proof. Let $A \in \mathcal{A}$ then we can decompose it as follows:

$$A = A_+ - A_-,$$

where $A_+ \geq 0$ and $A_- \geq 0$. Let $\mathcal{K}$ be positivity preserving then

$$\mathcal{K}A = \mathcal{K}A_+ - \mathcal{K}A_-.$$

Since $\mathcal{K}$ is positivity preserving $\mathcal{K}A_+ \geq 0$ and $\mathcal{K}A_- \geq 0$. It now follows that $\mathcal{K}A$ is self-adjoint since positive elements are self-adjoint by definition. □

Since $\Delta_\sigma = \Delta_\sigma^\dagger$ there must exist an orthonormal basis $\{E_1, ..., E_n\}$ consisting of the eigenvectors of Δ_σ. Of course $\mathbf{1}$ is an eigenvector of Δ_σ so we may set $E_1 = \mathbf{1}$. Since the eigenvectors are orthonormal we then have for all $\gamma > 1$ that $\mathrm{Tr}[E_\gamma]$=0. Moreover since Δ_σ is a strictly positive operator we then have that all the eigenvalues are positive and can be written in the form e^{ω_γ} for some $\omega_\gamma \in \mathbb{R}$.

Proposition 3.3.7. *For all $E \in \mathcal{H}_\mathcal{A}$ and $\omega \in \mathbb{R}$ we have that*

$$\Delta_\sigma E = e^{-\omega} E$$

if and only if

$$\Delta_\sigma E^* = e^{\omega} E^*.$$

Proof. Suppose $\Delta_\sigma E = e^{-\omega} E$, then $(\Delta_\sigma E)^* = e^{-\omega} E^*$. From equation (3.3.1) it follows that

$$\Delta_\sigma^{-1} E^* = e^{-\omega} E^*.$$

Now $\Delta_\sigma(\Delta_\sigma^{-1} E^*) = E^*$ thus

$$E^* = \Delta_\sigma(e^{-\omega} E^*) = e^{-\omega} \Delta_\sigma E^*.$$

Consequently we must have $\Delta_\sigma E^* = e^\omega E^*$. To prove the converse statement we can use the exact argument as above with the roles of E and E^* interchanged. □

From Proposition 3.3.7 we immediately see that if $e^{-\omega}$ is an eigenvalue of Δ_σ then so is e^ω.

Proposition 3.3.8. *The two eigenspaces $\{E_1, ..., En\}$ and $\{E_1^*, ..., E_n^*\}$ are orthogonal and have the same dimension.*

Proof. Let $\mathcal{V}_\lambda$ denote the eigenspace of Δ_σ that corresponds to the eigenvalue λ. We consider two cases:

1. For $\lambda \neq 1$. Consider any orthonormal basis $E_{\lambda,1}, ..., E_{\lambda,n_\lambda}$ for every $\mathcal{V}_\lambda$ with $\lambda > 1$. It now follows immediately from Proposition 3.3.7 that $E_{\lambda,1}, ..., E_{\lambda,n_\lambda}$ is an orthonormal basis for $\mathcal{V}_{\lambda^{-1}}$.

2. For $\lambda = 1$. We have $E_1 = 1 \in \mathcal{V}_1$. Of course all operators that commute with σ are in $\mathcal{V}_1$, so we could have $\dim(\mathcal{V}_1) > 1$. Let $F_2, F_3, ..., F_m$ be any basis for $\mathcal{V}_1 \perp (\mathbb{C}E_1)$, where the orthogonal complement is understood to be the orthogonal complement in $\mathcal{V}_1$.

By Proposition 3.3.7 that $F_2^*, F_3^*, ..., F_m^* \in \mathcal{V}_1$. Thus $\mathcal{V}_1 \perp (\mathbb{C}E_1)$ is spanned by

$$F_j + F_j^*, i(F_j - F_j^*) \quad \text{for j=2,...,m.}$$

Consider the real vector space $\mathcal{W}$ consisting of all linear combinations of $F_j + F_j^*$,$i(F_j - F_j^*)$ with real scalars. All the elements of $\mathcal{W}$ are self-adjoint. For any $A, B \in \mathcal{W}$ we have

$$\langle A, B\rangle_{\mathcal{H}_\mathcal{A}} = \tau(A^*B) = \tau(AB) = \tau(BA) = \tau(B^*A) = \langle B, A.\rangle_{\mathcal{H}_\mathcal{A}}$$

Now choose any orthonormal basis $E_2, ..., E_p$ for the real Hilbert space $\big(\mathcal{W}, \langle\cdot,\cdot\rangle_{\mathcal{H}_\mathcal{A}}\big)$. In particular all elements $F_j + F_j^*$,$i(F_j - F_j^*)$ are real linear combinations of $E_2, ..., E_p$, while complex linear combinations span $\mathcal{V}_1 \perp (\mathbb{C}E_1)$. Note that $E_2, ..., E_p$ are all self-adjoint. Because $E_2, ..., E_p$ are orthonormal in $\mathcal{W}$, they are orthonormal in $\mathcal{H}_\mathcal{A}$, since it is the same inner product. Thus $E_2, ..., E_p$ is an orthonormal basis for $\mathcal{V}_1$ consisting of self-adjoint operators. Note that $\dim(\mathcal{V}_1 \perp \mathbb{C}E_1)$ is equal to $\dim(\mathcal{W})$ thus $m = p$.

□

From Proposition 3.3.8 we see that we can create an orthonormal basis for $\mathcal{H}_\mathcal{A}$ by adding $E_1, ..., E_p$, $E_{\lambda,1}, ..., E_{\lambda,n_\lambda}$ and $E_{\lambda,1}^*, ..., E_{\lambda,n_\lambda}^*$. Propositions 3.3.7 and 3.3.8 show how we can construct a modular basis defined by the following properties:

Definition 3.3.9. Let $\mathcal{A}$ be a finite dimensional C^*-algebra and let $\sigma \in \mathcal{G}_+(\mathcal{A})$. Then there exists an orthonormal basis $\{E_1, ..., E_n\}$ such that

(i) $\{E_1, ..., E_n\}$ consists of the eigenvectors of Δ_σ.

(ii) $E_1 = \mathbf{1}$.

(iii) $\{E_1, ..., E_n\}$=$\{E_1^*, ..., E_n^*\}$.

The basis $\{E_1, ..., E_n\}$ is called the *modular basis.*

In the Heisenberg picture of quantum mechanics, the time evolution behaviour of a system is governed by a one parameter semigroup $\mathcal{P}_t$ of completely positive operators acting on the set of bounded linear operators on the associated Hilbert space $\mathcal{H}$. This is essentially a quantum analog of the usual Markov semigroups used to study dynamical systems, with the underlying state space replaced by a non-commutative operator algebra [Fag99]. Formally the quantum Markov semigroup is defined as follows:

Definition 3.3.10. Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{B}(\mathcal{H})$ be the set of bounded linear operators on $\mathcal{H}$. A *quantum Markov semigroup* (QMS) is a one-parameter family of operators $\{\mathcal{P}_t\}$ satisfying the following properties:

(i) $\mathcal{P}_0(X) = X$ for all $X \in \mathcal{B}(\mathcal{H})$.

(ii) $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s$ for all $t, s > 0$.

(iii) $\lim_{t\to 0} ||\mathcal{P}_t(X) - X|| = 0$ for all $X \in \mathcal{B}(\mathcal{H})$.

(iv) $\mathcal{P}_t(\mathbf{1}) = \mathbf{1}$, $||\mathcal{P}_t(X) - X|| = 0$.

(v) $\lim_{n\to\infty} \mathcal{P}_t(X_n) = \mathcal{P}_t(X)$ whenever we have that $\lim_{n\to\infty} X_n = X$

The limit in the above definition is understood to be the weak limit. Luckily in our finite dimensional setting we do not need to concern ourselves with this technicality as weak closure is the same as norm closure, our C^* algebra is also a Von Neumann algebra [CM17]. We also assume uniform continuity in which case there exists an element $\mathcal{L} \in \mathcal{B}(\mathcal{H})$ such that for all $X \in \mathcal{B}(\mathcal{H})$ we have [Par12]

$$\mathcal{L}(X) = \lim_{t\to 0} \frac{1}{t}(\mathcal{P}_t(X) - X), \tag{3.3.2}$$

where the limit is understood to be in the operator norm. The operator in $\mathcal{L}$ in equation (3.3.2) is known as the generator of $\mathcal{P}_t$ and the QMS can then be written in the following way

$$\mathcal{P}_t = e^{t\mathcal{L}}$$

When a quantum Markov semigroup $\mathcal{P}_t = e^{\mathcal{L}t}$ is a self-adjointess preserving semigroup then it follows that the generator $\mathcal{L}$ is also self-adjointess preserving. We can clearly see this property from equation (3.3.2). Of course any QMS $\mathcal{P}_t$ acts on $\mathcal{B}(\mathcal{H})$, which is a normed space, so we can associate to any $\mathcal{P}_t$ a dual semigroup $\mathcal{P}_t^\dagger$ acting on the set of faithful states $\mathcal{G}_+(\mathcal{A})$.

Definition 3.3.11. A QMS $\mathcal{P}_t$ is said to be *ergodic* if it is the case that $\mathbf{1}$ spans the eigenspace of $\mathcal{P}_t$.

In that case there will exist a unique invariant state $\sigma \in \mathcal{G}_+$ [CM17]. Intuitively this unique invariant state will essentially play the role of a stationary or target distribution when we apply the ideas of optimal transport theory in later Chapters.

Chapter 4

Detailed balance

4.1 Classical detailed balance

Consider a Markov chain on a finite state space with elements $\{x_1, ..., x_n\}$. Let σ be a stationary distribution and let P denote the transition matrix. Since σ is a stationary distribution we have by definition that $P\sigma = \sigma$ or equivalently $\sigma_j = \sum_{i=1}^{n} P_{ij}\sigma_j \forall j$. The transition matrix P is said to satisfy the detailed balance condition if the following condition holds:

$$\sigma_i P_{ij} = \sigma_j P_{ji}. \tag{4.1.1}$$

In other words the probability of switching from state i to state j is the same as the probability of switching from state j to state i. Let X_n be a Markov process started from σ and let $\Pr[X_n = i]$ denote the probability of X_n being in state i, then the detailed balance condition given by equation (4.1.1) is equivalent to

$$\Pr[X_n = i, X_{n+1} = j] = \Pr[X_n = j, X_{n+1} = i]$$

for all i, j and n. Now since (X_n, X_{n+1}) and (X_{n+1}, X_n) have the same joint distributions we can say that equation (4.1.1) characterizes time reversal invariance. We can express equation (4.1.1) in terms of self adjointness with respect to an inner product defined as follows:

$$\langle f, g \rangle_\sigma := \sum_{k=1}^{n} \sigma_k \bar{f}_k g_k \tag{4.1.2}$$

where $f, g \in \mathbb{C}^n$.

Theorem 4.1.1. *The matrix P is self adjoint with respect to the inner product defined by (4.1.2) if and only if the detailed balance condition (4.1.1) is satisfied.*

Proof. First suppose that P is self-adjoint with respect to $\langle\cdot,\cdot\rangle_\sigma$. That is

$$\langle Pf, g\rangle_\sigma = \langle f, Pg\rangle_\sigma . \tag{4.1.3}$$

Expanding equation (4.1.3) using (4.1.2) we obtain

$$\sum_{k=1}^{n}\sum_{j=1}^{n} \sigma_k P_{kj} \bar{f}_j g_k = \sum_{k=1}^{n}\sum_{j=1}^{n} \sigma_k \bar{f}_k P_{kj} g_j. \tag{4.1.4}$$

It then follows that equation (4.1.1) holds. Conversely suppose that equation (4.1.1) holds. Then we again obtain equation (4.1.4) from which equation (4.1.2) follows. □

The main advantage of writing the detailed balance condition in terms of an inner product is that it gives us some idea of how to generalize the notion of detailed balance to a quantum setting. It turns out that there are several ways to define a notion of quantum detailed balance.

4.2 Quantum detailed balance

Let $\mathcal{P}_t$ be a QMS on $\mathcal{A}$, a finite dimensional C^*-algebra. We say a state $\sigma \in \mathcal{G}_+$ is invariant under $\mathcal{P}_t^\dagger$ if $\mathcal{P}_t^\dagger \sigma = \sigma$ which is equivalent to $\mathcal{L}^\dagger \sigma = 0$. To see the above equivalence note that we can write $\mathcal{P}_t$ in terms of its generator $\mathcal{L}$ as follows $\mathcal{P}_t^\dagger = e^{t\mathcal{L}^\dagger}$. Now

$$\begin{aligned}
\mathcal{P}_t^\dagger \sigma &= e^{t\mathcal{L}^\dagger}\sigma \\
&= \sum_{k=0}^{\infty} \left(t\mathcal{L}^\dagger\right)^k \sigma \\
&= \left(\mathbf{1} + \sum_{k=1}^{\infty} \left(t\mathcal{L}^\dagger\right)^k\right) \sigma \\
&= \sigma + \sum_{k=1}^{\infty} \left(t\mathcal{L}^\dagger\right)^k \sigma.
\end{aligned}$$

Since $\mathcal{L}^\dagger \sigma = 0$ it follows that $\mathcal{P}_t^\dagger \sigma = \sigma$.

Definition 4.2.1. An inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{A}$ is said to be *compatible* with a state $\sigma \in \mathcal{G}_+$ if

$$\mathrm{Tr}[\sigma A] = \langle \mathbf{1}, A\rangle$$

for all $A \in \mathcal{A}$.

Theorem 4.2.2. *If $\mathcal{P}_t$ is self-adjoint with respect to an inner product that is compatible with σ then σ is invariant under $\mathcal{P}_t$.*

Proof.

$$\begin{aligned}\mathrm{Tr}[\sigma A] &= \langle \mathbf{1}, A\rangle \\ &= \langle \mathcal{P}_t \mathbf{1}, A\rangle \\ &= \langle \mathbf{1}, \mathcal{P}_t A\rangle \\ &= \mathrm{Tr}[\sigma \mathcal{P}_t A]\end{aligned}$$

Now keep in mind that the trace defines an inner product in the form $\langle \sigma, A\rangle = \mathrm{Tr}[\sigma A]$. From the above calculation we see that

$$\begin{aligned}\langle \sigma, A\rangle &= \mathrm{Tr}[\sigma A] \\ &= \mathrm{Tr}[\sigma \mathcal{P}_t A] \\ &= \langle \sigma, \mathcal{P}_t A\rangle\end{aligned}$$

Now finally by definition $\left\langle \mathcal{P}_t^\dagger \sigma, A\right\rangle = \langle \sigma, \mathcal{P}_t A\rangle$ so then we have

$$\left\langle \mathcal{P}_t^\dagger \sigma, A\right\rangle = \langle \sigma, A\rangle$$

Since this is true for any $A \in \mathcal{A}$ it follows that $\mathcal{P}_t^\dagger \sigma = \sigma$. □

Definition 4.2.3. Let $\sigma \in \mathcal{G}_+$ be a non-degenerate density matrix. For each $s \in \mathbb{R}$ and $A, B \in \mathcal{A}$ we define the following inner product:

$$\begin{aligned}\langle A, B\rangle_s &= \mathrm{Tr}\left[\left(\sigma^{(1-s)/2} A \sigma^{s/2}\right)^* \left(\sigma^{(1-s)/2} B \sigma^{s/2}\right)\right] \\ &= \mathrm{Tr}\left[\sigma^{s/2} A^* \sigma^{(1-s/2)} \sigma^{(1-s/2)} B \sigma^{s/2}\right] \\ &= \mathrm{Tr}\left[\sigma^s A^* B \sigma^{1-s}\right] \qquad (4.2.1)\end{aligned}$$

Note that we can rewrite the inner product in equation (4.2.1) as follows:

$$\begin{aligned}\langle A, B\rangle_s &= \mathrm{Tr}\left[\sigma \sigma^{(s-1)} A^* B \sigma^{(1-s)}\right] \\ &= \mathrm{Tr}\left[A^* (\Delta_\sigma^{1-s} B) \sigma\right] \qquad (4.2.2)\end{aligned}$$

where we have made use of the cyclic property of the trace.

As a generalization of inner product in definition 4.2.3 we can define for any function $f : (0, \infty) \mapsto (0, \infty)$ the following inner product:

$$\langle A, B\rangle_f := \mathrm{Tr}[A^* (f(\Delta_\sigma) B) \sigma], \qquad (4.2.3)$$

where the function f is understood to act on the components of Δ_σ represented as a diagonal matrix i.e:

$$f(\Delta_\sigma) = f \begin{bmatrix} \Delta_{11} & & \\ & \ddots & \\ & & \Delta_{nn} \end{bmatrix} = \begin{bmatrix} f(\Delta_{11}) & & \\ & \ddots & \\ & & f(\Delta_{nn}) \end{bmatrix}$$

Let R_A denote the operation of right multiplication by A and define the operator

$$\Omega_\sigma^f := R_\sigma \circ f(\Delta_\sigma).$$

Now we can rewrite equation (4.2.3) as

$$\langle A, B \rangle_f = \text{Tr}[A^* \Omega_\sigma^f B].$$

Theorem 4.2.4. *Let $\mathcal{K}$ be a linear operator on $\mathcal{A}$. $\mathcal{K}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_f$ if and only if $\Omega_\sigma^f \circ \mathcal{K} = \mathcal{K}^\dagger \circ \Omega_\sigma^f$, where $K^\dagger$ denotes the inner product on $\mathcal{H}_\mathcal{A}$, or what is the same the Hilbert-Schmidt inner product.*

Proof. Suppose $\Omega_\sigma^f \circ \mathcal{K} = \mathcal{K}^\dagger \circ \Omega_\sigma^f$ then

$$\begin{aligned} \langle A, \mathcal{K}B \rangle_f &= \text{Tr}[A^\dagger \Omega_\sigma^f \mathcal{K} B] \\ &= \text{Tr}[A^\dagger \mathcal{K}^\dagger \Omega_\sigma^f B] \\ &= \text{Tr}[(\mathcal{K}A)^\dagger \Omega_\sigma^f B] \\ &= \langle \mathcal{K}A, B \rangle_f \end{aligned}$$

Conversely suppose that $\mathcal{K}$ is self-adjoint w.r.t. the inner product $\langle \cdot, \cdot \rangle_f$ then we have that $\langle \mathbf{1}, KB \rangle_f = \langle K, B \rangle_f$ for all $B \in \mathcal{A}$.
Thus $\text{Tr}[\Omega_\sigma^f KB] = \text{Tr}[K^\dagger \Omega_\sigma^f B]$ for all $B \in \mathcal{A}$ and the result follows. □

The authors of [TKR+10] define quantum detailed balance in terms of self-adjointness with respect to $\langle \cdot, \cdot \rangle_f$ which then gives rise to several *a priori* notions of quantum detailed balance. These definitions will naturally depend on the choice of function f. It was shown in [TKR+10] that if a linear operator is self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\frac{1}{2}}$ then it is not necessarily self-adjoint with $\langle \cdot, \cdot \rangle_f$ for all functions f. Specifically the self-adjointness property then does not hold for $\langle \cdot, \cdot \rangle_{(1+t)/2}$ which corresponds to the Bures metric. The Bures metric, also known as the Helstrom metric, is generalization of the Fisher information metric from information geometry, and is used in quantum information geometry to define infinitesimal distances between density operators [Bur69]. From the example in [TKR+10] we see how different choices of the function f can yield different notions of detailed balance in the quantum case. Fortunately it turns out that self-adjointness with respect to $\langle\cdot,\cdot\rangle_s$ for any $s \neq \frac{1}{2}$ also implies self-adjointness with respect to $\langle\cdot,\cdot\rangle_f$ for any function f. We shall prove this result following closely the argument outlined in section 2 of [CM17].

First we show that the modular automorphism group α_{it} is self-adjoint with respect to $\langle\cdot,\cdot\rangle_s$ for all $s \in \mathbb{R}$.

Lemma 4.2.5. *Let $\alpha_{it} = (\Delta_\sigma)^t$ denote the modular automorphism group. For all $t \in \mathbb{R}$ and $A, B \in \mathcal{A}$*

$$\langle \alpha_{it} A, B\rangle_s = \langle A, \alpha_{it} B\rangle_s \,.$$

Proof. To see this equivalence simply note that

$$\begin{aligned}
\langle \alpha_{it} A, B\rangle_s &= \mathrm{Tr}[\sigma^s(\alpha_{it}A)^*\sigma^{1-s}B] \\
&= \mathrm{Tr}[\sigma^s(\sigma^t A\sigma^{-t})^*\sigma^{1-s}B] \\
&= \mathrm{Tr}[\sigma^{s-t}A^*\sigma^{1-(s-t)}B] \\
&= \mathrm{Tr}[\sigma^s A^*\sigma^{1-s}\sigma^t B\sigma^{-t}] \\
&= \mathrm{Tr}[\sigma^s A^*\sigma^{1-s}(\Delta_\sigma)^t B] \\
&= \langle A, \alpha_{it} B\rangle_s
\end{aligned}$$

□

Note that the expression in the third line is by definition $\langle A, B\rangle_{s-t}$. This leads to the formula

$$\langle \alpha_{it} A, B\rangle_s = \langle A, B\rangle_{s-t} \tag{4.2.4}$$

which we shall make use of to prove the next lemma.

Next we shall show how a certain class of operators commute with the modular automorphism group.

Lemma 4.2.6. *Let $\sigma \in \mathcal{G}_+$ be a non-degenerate density matrix, and let $s \in [0,1]$ with $s \neq \frac{1}{2}$. Let $\mathcal{K}$ be any operator on $\mathcal{A}$ that is self-adjoint with respect to $\langle\cdot,\cdot\rangle_s$, and is self-adjointness preserving. Then $\mathcal{K}$ commutes with α_{it} for all $t \in \mathbb{R}$. Equivalently $\mathcal{K}$ commutes with the modular operator Δ_σ.*

Proof. The proof is done in two steps. First we shall show that $\mathcal{K}$ commutes with $\alpha_{i(2s-1)}$.

$$\begin{aligned}\left\langle \mathcal{K}\alpha_{i(2s-1)}A, B\right\rangle_s &= \left\langle \alpha_{i(2s-1)}A, \mathcal{K}B\right\rangle_s \\ &= \mathrm{Tr}[\sigma^s(\sigma^{2s-1}A\sigma^{1-2s})^*\sigma^{1-s}\mathcal{K}B] \\ &= \mathrm{Tr}[\sigma^s\sigma^{1-2s}A^*\sigma^{2s-1}\sigma^{1-s}\mathcal{K}B] \\ &= \mathrm{Tr}[\sigma^{1-s}A^*\sigma^s\mathcal{K}B] \\ &= \mathrm{Tr}[\sigma^s\mathcal{K}B\sigma^{1-s}A^*] \\ &= \mathrm{Tr}[\sigma^s(\mathcal{K}(B^*))^*\sigma^{1-s}A^*].\end{aligned}$$

In the final line above we made use of the fact that $\mathcal{K}$ is self-adjointness preserving, i.e. $\mathcal{K}B = (\mathcal{K}(B^*))^*$. Now

$$\begin{aligned}\mathrm{Tr}[\sigma^s(\mathcal{K}(B^*))^*\sigma^{1-s}A^*] &= \langle \mathcal{K}(B^*), A\rangle_s \\ &= \langle B^*, \mathcal{K}A\rangle_s \\ &= \mathrm{Tr}[\sigma^s B\sigma^{1-s}(\mathcal{K}(A^*))^*] \\ &= \mathrm{Tr}[\sigma^{1-s}(\mathcal{K}(A^*))^*\sigma^s B] \\ &= \langle \mathcal{K}A, B\rangle_{1-s}\end{aligned}$$

where in the third line we have once again made use of the fact that $\mathcal{K}$ is self-adjointness preserving. Now by noting that $1-s = s-(2s-1)$ we can apply equation (4.2.4) to obtain

$$\langle \mathcal{K}A, B\rangle_{1-s} = \left\langle \alpha_{i(2s-1)}\mathcal{K}A, B\right\rangle_s .$$

Thus we have that $\left\langle \mathcal{K}\alpha_{i(2s-1)}A, B\right\rangle_s = \left\langle \alpha_{i(2s-1)}\mathcal{K}A, B\right\rangle_s$. Since this result is true for arbitrary $A, B \in \mathcal{A}$ it follows that $\mathcal{K}$ commutes with $\alpha_{i(2s-1)}$.

For the second step of the proof we shall argue why commutation with $\alpha_{i(2s-1)}$ leads to commutation with α_{it} for all $t \in \mathbb{R}$.

Since $\mathcal{K}$ commutes with $\alpha_{i(2s-1)}$, i.e. $\mathcal{K}$ commutes with the operator Δ_σ^{2s-1}, it also commutes with every polynomial of Δ_σ^{2s-1}. We have restricted our operators to a finite-dimensional setting so Δ_σ^{2s-1} has a matrix representation

$$\Delta_\sigma^{2s-1} = \begin{bmatrix}\delta_1 & & \\ & \ddots & \\ & & \delta_n\end{bmatrix}, \qquad \delta_j \geq 0 \forall j.$$

Next consider any function $f : (0, \infty) \to (0, \infty)$. We can then represent f acting on Δ_σ^{2s-1} as

$$f(\Delta_\sigma^{2s-1}) = \begin{bmatrix} f(\delta_1) & & \\ & \ddots & \\ & & f(\delta_n) \end{bmatrix}$$

so the function need only be defined on a finite set of values. Thus we have a set of n points $(\delta_j, f(\delta_j))$, $j = 1, ..., n$. We can always fit a polynomial p_f to these points such that the values at the points δ_j correspond to that of the function f, i.e $p_f(\delta_j) = f(\delta_j)$ for all $j = 1, ..., n$. Since $\mathcal{K}$ commutes with any fucntion f in particular then $\mathcal{K}$ commutes with α_t for all $t \in \mathbb{R}$. □

Finally the next lemma is used to show how self-adjointness with respect to $\langle \cdot, \cdot \rangle_s$ with $s \neq \frac{1}{2}$ leads to self-adjointness with respect to $\langle \cdot, \cdot \rangle_f$ for any function f.

Lemma 4.2.7. *Let $\sigma \in \mathcal{G}_+$ be a non-degenerate density matrix. Let $\mathcal{K}$ be any operator on $\mathcal{A}$. If $\mathcal{K}$ commutes with the modular automorphism group of σ and $\mathcal{K}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_f$ for some function $f : (0, \infty) \mapsto (0, \infty)$, then $\mathcal{K}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle_g$ for all $g : (0, \infty) \mapsto (0, \infty)$*

Proof. Let $h = g/f$. $\mathcal{K}$ commutes with the modular automorphism group of σ, i.e commutes with Δ_σ. Arguing similarly as in the proof of Lemma 4.2.6 we note that $\mathcal{K}$ commutes with any function of Δ_σ and in particular commutes with $h(\Delta_\sigma)$. Thus for all $A, B \in \mathcal{A}$ we have that

$$\begin{aligned} \langle A, \mathcal{K}(B) \rangle_g &= \mathrm{Tr}[\sigma A^* g(\Delta_\sigma) \mathcal{K}(B)] \\ &= \mathrm{Tr}[\sigma A^* f(\Delta_\sigma) h(\Delta_\sigma) \mathcal{K}(B)] \\ &= \mathrm{Tr}[\sigma A^* f(\Delta_\sigma) \mathcal{K} h(\Delta_\sigma)(B)] \\ &= \langle A, \mathcal{K} h(\Delta_\sigma) B \rangle_f \,. \end{aligned}$$

Since $\mathcal{K}$ is self-adjoint with respect to f it follows that

$$\begin{aligned} \langle A, \mathcal{K} h(\Delta_\sigma) B \rangle_f &= \mathrm{Tr}[\sigma (\mathcal{K}(A))^* f(\Delta_\sigma) h(\Delta_\sigma) B] \\ &= \mathrm{Tr}[\sigma (\mathcal{K}(A))^* g(\Delta_\sigma) B] \\ &= \langle \mathcal{K}(A), B \rangle_g \,. \end{aligned}$$

□

The results of Lemma 4.2.6 and Lemma 4.2.7 can be concisely summarized in the following theorem:

Theorem 4.2.8. *Let $\sigma \in \mathcal{G}_+$ be a non-degenerate density matrix and let $\mathcal{K}$ be any operator on $\mathcal{A}$. If $\mathcal{K}$ is self-adjoint with respect to the GNS inner product $\langle \cdot, \cdot \rangle_{s=1}$ and $\mathcal{K}$ is self-adjointness preserving, then $\mathcal{K}$ commutes with the modular automorphism group of σ and is self-adjoint with respect to $\langle \cdot, \cdot \rangle_f$ for any function $f : (0, \infty) \mapsto (0, \infty)$.*

Finally we are in a position to define a sensible notion of quantum detailed balance that makes use of the properties outlined above.

Definition 4.2.9 (Detailed balance)**.** A QMS $\mathcal{P}_t$ satisfies the *detailed balance condition* with respect to σ (henceforth referred to as the σ-DBC), $\sigma \in \mathcal{G}_+$, if for all $t > 0$, $\mathcal{P}_t$ is self-adjoint with respect to the GNS inner product $\langle \cdot, \cdot \rangle_1$.

Remark 4.2.10. Note that $\langle 1, A \rangle_1 = \mathrm{Tr}[\sigma A]$ for all $A \in \mathcal{A}$. In other words $\langle \cdot, \cdot \rangle_1$ is a compatible inner product with respect to σ in the sense of Definition 4.2.1, consequently σ is invariant under $\mathcal{P}_t$.

Remark 4.2.11. Note that any QMS $\mathcal{P}_t = e^{t\mathcal{L}}$ is self-adjointness preserving. For any $\mathcal{P}_t$ that satisfies the σ-DBC it then follows from Theorem 4.2.8 that

$$\alpha_{t'}(\mathcal{P}_t(A)) = \mathcal{P}_t(\alpha_{t'}(A)) \tag{4.2.5}$$

for all $t, t' \in \mathbb{R}$ and all $A \in \mathcal{A}$.

The previous remark is an observation first made by Robert Alicki in [Ali76]. The fact that $\mathcal{P}_t$ commutes with the modular automorphism group α_t shows that $\mathcal{P}_t$ commutes with time-translation of the Hamiltonian operator h corresponding to the state σ. Hence equation (4.2.5) may be viewed as analogous to (4.1.1) in that it characterizes a form of time-reversal invariance for the quantum case. Thus our definition of quantum detailed balance intuitively resembles the classical definition of detailed balance.

Chapter 5

Generators of quantum Markov semigroups

The requirement that a QMS $\mathcal{P}_t = e^{t\mathcal{L}}$ satisfies the σ-DBC imposes a certain structure on the generator $\mathcal{L}$. This was first studied by Alicki in [Ali76] for a non-degenerate $\sigma \in \mathcal{G}_+$. It was found that the generator of a QMS that satisfies the σ-DBC has the form of equation (5.0.1) given below. It turns out that the requirement that σ be strictly positive is in fact not necessary. Alicki's Theorem can be generalized to a matrix algebra setting. This was done by Carlen and Maas in [CM17] and the proof, which is quite technical, can be found in Appendix A of [CM17].

Theorem 5.0.1. *Let $\mathcal{P}_t = e^{t\mathcal{L}}$ be a QMS on a unital C^*-subalgebra $\mathcal{A}$ on $\mathcal{M}_n(\mathbb{C})$. Suppose that $\mathcal{P}_t$ satisfies the σ-DBC for $\sigma \in \mathcal{G}_+(\mathcal{A})$ and that $\mathcal{P}_t$ has an extension $\widehat{\mathcal{P}}_t$ to a QMS on $\mathcal{M}_n(\mathbb{C})$. Regard the modular operator Δ_σ as an operator on $\mathcal{M}_n(\mathbb{C})$ and let τ denote the normalized trace on $\mathcal{M}_n(\mathbb{C})$. Then the generator $\mathcal{L}$ of $\mathcal{P}_t$ has the following form:*

$$\mathcal{L}A = \sum_{j\in\mathcal{J}} \left(e^{-\omega_j/2} V_j^*[A, V_j] + e^{\omega_j/2}[V_j, A]V_j^*\right) \tag{5.0.1}$$

$$= \sum_{j\in\mathcal{J}} e^{-\omega_j/2} \left(V_j^*[A, V_j] + [V_j^*, A]V_j\right) \tag{5.0.2}$$

where $\omega_j \in \mathbb{R}$ for all $j \in \mathcal{J}$, and $\{V_j\}_{j\in\mathcal{J}}$ is a set in $\mathcal{M}_n(\mathbb{C})$ that satisfies the following properties:

(i) $\tau[V_j^*, V_k] = \delta_{j,k}$ *for all* $j, k \in \mathcal{J}$.

(ii) $\tau[V_j] = 0$ *or all* $j \in \mathcal{J}$

(iii) $\{V_j\}_{j\in\mathcal{J}} = \{V_j^*\}_{j\in\mathcal{J}}$

(iv) $\{V_j\}_{j\in\mathcal{J}}$ *consists of eigenvectors of the modular operator* Δ_σ *with*

$$\Delta_\sigma V_j = e^{-\omega_j} V_j. \tag{5.0.3}$$

Conversely given any $\sigma \in \mathcal{G}_+(\mathcal{A})$*, and any set* $\{V_j\}_{j\in\mathcal{J}}$ *which satisfies (iii) and (iv) for some*$\{\omega_j\}_{j\in\mathcal{J}} \subseteq \mathbb{R}$*, the operator* $\mathcal{L}$ *given in equation* (5.0.1) *is the generator of a QMS* $\mathcal{P}_t$ *that satisfies the* σ*-DBC.*

Theorem 5.0.1 will be extremely useful going forward. The form of the generator given in equation 5.0.1 will serve as our starting point to define the notions of gradient and divergence in Chapter 6. These notions will be required to set up a quantum analogue of the continuity equation for probability densities.

Proposition 5.0.2. *If* σ *is the normalized trace then equation* (5.0.2) *reduces to*

$$\mathcal{L}A = -\sum_{j\in\mathcal{J}} [V_j, [V_j, A]]. \tag{5.0.4}$$

Proof. Recall that for any V_j we have that

$$\Delta_\sigma V_j = e^{-\omega_j} V_j = V_j$$

and

$$\Delta_\sigma V_j^* = e^{\omega_j} V_j^* = V_j^*.$$

The eigenvectors of Δ_σ with eigenvalues other than 1 cannot be self-adjoint. If σ is the normalized trace however, then $\tau(A) = \mathrm{Tr}[\sigma A] \forall A \in \mathcal{A}$. Then σ must have the form $\begin{bmatrix} \frac{1}{n} & & \\ & \ddots & \\ & & \frac{1}{n} \end{bmatrix}$.

Now

$$\Delta_\sigma A = \sigma A \sigma^{-1} = A \quad \forall A \in \mathcal{A},$$

hence Δ_σ is the identity matrix. So each V_j is an eigenvector of Δ_σ with corresponding eigenvalue $e^{-\omega_j} = 1$, i.e. $\omega_j = 0$ for all $j \in \mathcal{J}$. Thus it is possible to take every V_j to be self-adjoint and consequently equation (5.0.2) reduces to equation (5.0.4) as required. □

The above formulation arises naturally in the study of Fermionic oscillators via the Fermi Ornstein-Uhlenbeck semigroup. The reader is referred to [CM14, CL02] for further details.

Next we shall derive an expression for the Hilbert-Schmidt adjoint of the generator $\mathcal{L}$.

Proposition 5.0.3. *The Hilbert-Schmidt adjoint of $\mathcal{L}$, denoted by $\mathcal{L}^\dagger$ is given by*

$$\mathcal{L}^\dagger \rho = \sum_{j\in\mathcal{J}} \left(e^{-\omega_j/2}[V_j\rho, V_j^*] + e^{\omega_j/2}[V_j^*, \rho V_j]\right) \tag{5.0.5}$$

$$= \sum_{j\in\mathcal{J}} e^{-\omega_j/2} \left([V_j\rho, V_j^*] + [V_j, \rho V_j^*]\right) \tag{5.0.6}$$

Proof. Recall that the Hilbert-Schmidt adjoint is defined by

$$\langle A, \mathcal{L}B\rangle_{HS} = \mathrm{Tr}(A^*\mathcal{L}B) \tag{5.0.7}$$

Now for the sake of compactness we drop the summation term over $\mathcal{J}$ and look at a single term j in the summation defined by equation (5.0.7). Applying equation (5.0.1) and expanding the commutator terms we find that a single term in the summation looks like

$$e^{-\omega_{j/2}}[\mathrm{Tr}(A^*V_j^*BV_j) - \mathrm{Tr}(A^*V_j^*V_jB)] + e^{\omega_{j/2}}[\mathrm{Tr}(A^*V_jBV_j^*) - \mathrm{Tr}(A^*BV_jV_j^*)].$$

Applying the cyclic property of the trace to the first, third and fourth terms above we have

$$e^{-\omega_{j/2}}[\mathrm{Tr}(V_jA^*V_j^*B) - \mathrm{Tr}(A^*V_j^*V_jB)] + e^{\omega_{j/2}}[\mathrm{Tr}(V_j^*A^*V_jB) - \mathrm{Tr}(V_jV_j^*A^*B)].$$

Next we regroup terms to find

$$e^{-\omega_{j/2}}\,\mathrm{Tr}([V_j, A^*V_j^*]B) + e^{\omega_{j/2}}\,\mathrm{Tr}([V_j^*A^*, V_j]B).$$

We can make use of the commutator identity $([A, B])^* = [B^*, A^*]$ to rewrite the above equation as

$$e^{-\omega_{j/2}}\,\mathrm{Tr}([V_jA, V_j^*]^*B) + e^{\omega_{j/2}}\,\mathrm{Tr}([V_j^*, AV_j]^*B). \tag{5.0.8}$$

Finally comparing equation (5.0.5) and equation (5.0.8), and remembering to account for the summation over all $j \in \mathcal{J}$ we see that

$$\left\langle \mathcal{L}^\dagger A, B\right\rangle_{HS} = \mathrm{Tr}\left(\left(\mathcal{L}^\dagger A\right)^* B\right) = \mathrm{Tr}\left(A^*, \mathcal{L}B\right) = \langle A, \mathcal{L}B\rangle_{HS}. \tag{5.0.9}$$

Thus equation (5.0.5) is indeed the Hilbert-Schmidt adjoint of the generator $\mathcal{L}$. To obtain equation (5.0.6), we start from equation (5.0.2) and follow the same procedure as we did to obtain equation (5.0.5) □

Remark 5.0.4. It is sensible to ask whether the set V_j is uniquely determined. In our finite dimensional setting we can view any V_j as an $n \times n$ matrix. Property (ii) in Theorem 5.0.1 imposes a restriction that we have $n^2 - 1$ independent terms in such a matrix, whereas Property (i) further restricts the number of independent terms. Thus the set $|\mathcal{J}|$ has cardinality of at most $n^2 - 1$. Now it turns out that while the set $\{V_j\}_{j \in J}$ is not in general uniquely determined, the cardinality of the set $|\mathcal{J}|$ is uniquely determined. This was shown by Carlen and Maas in Appendix A of [CM17]. It was shown there that given two sets $\{V_j\}_{j \in \mathcal{J}}$ and $\{\tilde{V}_j\}_{j \in \mathcal{J}}$ there exists a $|\mathcal{J}| \times |\mathcal{J}|$ unitary matrix $U_{j,k}$ such that for all $j \in \mathcal{J}$

$$\tilde{V}_j = \sum_{k \in \mathcal{J}} U_{j,k} V_k$$

where $U_{j,k} = 0$ if $\omega_j \neq \tilde{\omega_k}$. Thus there is a canonical association of the set $\{V_j\}_{j \in \mathcal{J}}$ to the generator $\mathcal{L}$. Moreover then once the state σ and the set $\{V_j\}_{j \in \mathcal{J}}$ have been chosen, the set of real numbers $\{\omega_j\}_{j \in \mathcal{J}}$ are fixed.

Chapter 6

Dirichlet form of a quantum Markov generator

We begin with a quick overview of the problem we are trying to solve in this chapter and the next. Our main goal is to obtain a quantum mechanical analogue of the continuity equation for probability distribution functions. We first need to lay the groundwork for this in sections 6.1, 6.2 and 6.3. We obtain the quantum mechanical analogue of our continuity equation in section 6.4. First consider the usual continuity equation

$$\frac{\partial}{\partial t}\rho(x,t) + \operatorname{div}[\bar{v}(x,t)\rho(x,t)] = 0. \tag{6.0.1}$$

For any $\rho(x,t) > 0$ that can be written in the form

$$\frac{\partial}{\partial t}\rho(x,t) = \operatorname{div}[\bar{a}(x,t)] \tag{6.0.2}$$

for some vector field $\bar{a}(x,t)$, we can obtain equation (6.0.1) by setting

$$\bar{v}(x,t) = -\frac{\bar{a}(x,t)}{\rho(x,t)}.$$

Conversely given equation (6.0.1) and defining

$$\bar{a}(x,t) = -\bar{v}(x,t)\rho(x,t)$$

then we have

$$\begin{aligned}\frac{\partial}{\partial t}\rho(x,t) &= -\operatorname{div}[\bar{v}(x,t)\rho(x,t)]\\ &= \operatorname{div}[\bar{a}(x,t)].\end{aligned}$$

There are several obstacles to obtaining a quantum analogue of equation (6.0.1). First off our probability density functions are replaced with density matrices ρ, positive trace class operators on some Hilbert space with $\operatorname{Tr}\rho = 1$. Our "vector fields" also look a little bit more complicated. For the quantum case we will work with the direct sum of Hilbert spaces. This leads to a further complication since we now also need to define what exactly "multiplication" looks like between a density matrix and a direct sum of Hilbert spaces. We will also need to define what a "divergence" and "gradient" of such structures would look like. This is where the form of the generator of a QMS given by equation (5.0.2) will come in handy.

6.1 Gradient and divergence

Let $\mathcal{P}_t$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for some $\sigma \in \mathcal{G}_+(\mathcal{A})$. We know from Chapter 5 that the generator of the QMS can be written in the canonical form given by equation (5.0.2). We now fix such a generator and the sets $\{V_j\}_{j\in\mathcal{J}}$. Remark 5.0.4 in the previous section shows that the numbers $\{\omega_j\}$ are now also fixed.

Next we define the following operators ∂_j on $\mathcal{A}$:

$$\partial_j A = [V_j, A] \quad \text{so that} \quad \partial_j^\dagger A = [V_j^*, A].$$

The operators ∂_j are derivations, that is to say a mapping which generalizes properties of the derivative. In this case it is easy to see as commutators are linear operators and satisfy a Leibniz rule. The operator $\partial_j^\dagger$ is the Hilbert-Schmidt adjoint of ∂_j. The following quick calculation shows this.

$$\begin{aligned}
\langle A, \partial_j B\rangle_{HS} &= \operatorname{Tr}(A^*[V_j, B)] \\
&= \operatorname{Tr}(A^* V_j B) - \operatorname{Tr}(A^* B V_j) \\
&= \operatorname{Tr}(A^* V_j B) - \operatorname{Tr}(V_j A^* B) \\
&= \operatorname{Tr}(A^* V_j B - V_j A^* B) \\
&= \operatorname{Tr}([A^*, V_j]B) \\
&= \operatorname{Tr}([V_j^*, A]^* B) \\
&= \left\langle \partial_j^\dagger A, B \right\rangle_{HS}.
\end{aligned}$$

We may now proceed to construct non-commutative analogues of the gradient, divergence and Laplace operator with respect to the Hilbert-Schmidt inner product. We start off with the Laplacian:

Definition 6.1.1. Non-commutative Laplacian: Given the set $\{V_j\}_{j\in\mathcal{J}}$, we define an operator $\mathcal{L}_0$ on the Hilbert space $\mathcal{H}_\mathcal{A}$ by:

$$\mathcal{L}_0 A = -\sum_{j\in\mathcal{J}} \partial_j^\dagger \partial_j A = -\sum_{j\in\mathcal{J}} [V_j^*, [V_j, A]]. \tag{6.1.1}$$

Proposition 6.1.2. *The operator $\mathcal{L}_0$ defined in equation (6.1.1) is self-adjoint with respect to the Hilbert-Schmidt inner product.*

Proof. Note that

$$\begin{aligned}
\mathrm{Tr}(A^*[V_j^*, [V_j, B]]) &= \mathrm{Tr}(A^*(V_j^*(V_jB - BV_j) - (V_jB - BV_j)V_j^*)) \\
&= \mathrm{Tr}(A^*V_j^*V_jB) - \mathrm{Tr}(A^*V_j^*BV_j) - \mathrm{Tr}(A^*V_jBV_j^*) + \mathrm{Tr}(A^*BV_jV_j^*) \\
&= \mathrm{Tr}(A^*V_j^*V_jB) - \mathrm{Tr}(V_jA^*V_j^*B) - \mathrm{Tr}(V_j^*A^*V_jB) + \mathrm{Tr}(V_jV_j^*A^*B) \\
&= \mathrm{Tr}([A^*, V_j^*]V_jB) - \mathrm{Tr}(V_j[A^*, V_j^*]B) \\
&= \mathrm{Tr}([[A^*, V_j^*], V_j]B)
\end{aligned}$$

Now if we apply the commutator property $[A, B]^* = [B^*, A^*]$ we see that

$$\mathrm{Tr}([[A^*, V_j^*], V_j]B) = \mathrm{Tr}([[V_j, A]^*, V_j]B) = \mathrm{Tr}([V_j^*, [V_j, A]]^*B).$$

It then follows that

$$\begin{aligned}
\langle A, \mathcal{L}_0 B\rangle_{HS} &= \sum_{j\in\mathcal{J}} \mathrm{Tr}(A^*[V_j^*, [V_j, B]]) \\
&= \sum_{j\in\mathcal{J}} \mathrm{Tr}([V_j^*, [V_j, A]]^*B) \\
&= \langle \mathcal{L}_0 A, B\rangle_{HS}.
\end{aligned}$$

Thus we find that $\mathcal{L}_0 = \mathcal{L}_0^\dagger$. □

Note that if we expand equation (6.1.1) we obtain

$$\begin{aligned}
\mathcal{L}_0 A &= \sum_{j\in\mathcal{J}} [[V_j, A], V_j^*] \\
&= \sum_{j\in\mathcal{J}} ([V_j, A]V_j^* - V_j^*[V_j, A]) \\
&= \sum_{j\in\mathcal{J}} (V_jAV_j^* - AV_jV_j^* - V_j^*V_jA + V_j^*AV_j) \\
&= \sum_{j\in\mathcal{J}} (V_j^*[A, V_j] + [V_j^*, A]V_j).
\end{aligned}$$

The above has the form of equation (5.0.2) with $\omega_j = 0$ for all $j \in \mathcal{J}$. It follows from Theorem 5.0.1 that $\mathcal{L}_0$ is the generator of a QMS $\mathcal{P}_{0,t} = e^{t\mathcal{L}_0}$ satisfying detailed balance with respect to the normalized trace. We call $\mathcal{P}_{0,t}$ the *heat semigroup* associated to the $\mathcal{P}_t$, and $\mathcal{L}_0$ the *Laplace operator associated to* $\mathcal{L}$ [CM17].

Next we will define exactly what our analogue of a "vector field" in equation (6.0.1) will look like. We define a Hilbert space $\mathcal{H}_{\mathcal{A},\mathcal{J}}$ as follows:

$$\mathcal{H}_{\mathcal{A},\mathcal{J}} = \bigoplus_{j\in\mathcal{J}} \mathcal{H}_{\mathcal{A}}^{(j)},$$

where each $\mathcal{H}_{\mathcal{A}}^{(j)}$ is a copy of the original Hilbert space $\mathcal{H}_{\mathcal{A}}$. Note that $\mathcal{H}_{\mathcal{A},\mathcal{J}} = \bigoplus_{j\in\mathcal{J}} \mathcal{H}_{\mathcal{A}}^{(j)}$ is a direct sum of Hilbert spaces, which is itself a Hilbert space of dimension $|\mathcal{J}| \times n$, where n is the dimension of $\mathcal{H}_{\mathcal{A}}$ and $|\mathcal{J}|$ is the cardinality of $\mathcal{J}$. Recall from the discussion in remark 5.0.4 that the $|\mathcal{J}|$ is at most $n^2 - 1$. Thus $\mathcal{H}_{\mathcal{A},\mathcal{J}}$ has dimension of at most $n(n^2 - 1)$. We can choose some linear ordering for the set $\mathcal{J}$ to define the following

$$\mathbf{A} = (A_1, ..., A_{|\mathcal{J}|}).$$

We equip $\mathcal{H}_{\mathcal{A}}$, with the standard inner product

$$\langle \mathbf{A}, \mathbf{B} \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} = \sum_{j\in\mathcal{J}} \langle A_j, B_j \rangle_{\mathcal{H}_{\mathcal{A}}}.$$

Next we define an analogue to the usual gradient operator that takes an element in $\mathcal{H}_{\mathcal{A}}$ and maps it to $\mathcal{H}_{\mathcal{A},\mathcal{J}}$. Formally we define $\nabla : \mathcal{H}_{\mathcal{A}} \to \mathcal{H}_{\mathcal{A},\mathcal{J}}$ by:

$$\nabla A = (\partial_1 A, ..., \partial_{|\mathcal{J}|} A). \tag{6.1.2}$$

If we consider the elements of $\mathcal{A}$ to be analogues of functions on some manifold, then it is natural to think of $\mathbf{A} = (A_1, ..., A_{|\mathcal{J}|})$ as a vector field.

Next we will define an analogue of the usual divergence operator $\operatorname{div} \mathbf{A} : \mathcal{H}_{\mathcal{A},\mathcal{J}} \to \mathcal{H}_{\mathcal{A}}$ by

$$\operatorname{div} \mathbf{A} = -\sum_{j\in\mathcal{J}} \partial^\dagger A_j = \sum_{j\in\mathcal{J}} [A_j, V_j^*]. \tag{6.1.3}$$

Proposition 6.1.3. *Let* div *be the operator defined in equation* (6.1.3) *and let* ∇ *be the operator defined in equation* (6.1.2)*. Furthermore let* $A \in \mathcal{A}$*, then* $\operatorname{div} \circ \nabla(A) = \mathcal{L}_0(A)$*.*

Proof.

$$\begin{aligned}
\operatorname{div}\circ\nabla(A) &= \operatorname{div}(\partial_1 A, ..., \partial_{|\mathcal{J}|}A)\\
&= \operatorname{div}([V_1, A], [V_2, A], ..., [V_{|\mathcal{J}|}, A])\\
&= -\sum_{j\in\mathcal{J}} [\partial_j^\dagger, [V_j, A]]\\
&= -\sum_{j\in\mathcal{J}} [V_j^*, [V_j, A]]\\
&= \mathcal{L}_0 A.
\end{aligned}$$

□

The elements of Null(div) are referred to as *divergence free vector fields.*

Proposition 6.1.4. *A vector field* $\mathbf{A}$ *is a gradient if and only if it is orthogonal in* $\mathcal{H}_{\mathcal{A},\mathcal{J}}$ *to every divergence free vector field, i.e* $(\mathrm{Null}(\mathrm{div}))^\perp = \mathrm{Ran}\nabla$.

Proof.

(i) Let $\mathbf{B} = (B_1, ..., B_{|\mathcal{J}|}) \in \mathrm{Ran}(\nabla)$ and $\mathbf{A} = (A_1, ..., A_{|\mathcal{J}|}) \in \mathrm{Null}(\mathrm{div})$. Since $\mathbf{B} \in \mathrm{Ran}(\nabla)$ there exists an element $B \in \mathcal{A}$ such that

$$\nabla B = (\partial_1 B, ..., \partial_{|\mathcal{J}|}B) = (B_1, ..., B_{|\mathcal{J}|}) = \mathbf{B}.$$

Then we have

$$\begin{aligned}
\langle \mathbf{A}, \mathbf{B}\rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} &= \sum_{j\in\mathcal{J}} \langle A_j, B_j\rangle_{\mathcal{H}_\mathcal{A}}\\
&= \sum_{j\in\mathcal{J}} \langle A_j, \partial_j B\rangle_{\mathcal{H}_\mathcal{A}}\\
&= \sum_{j\in\mathcal{J}} \left\langle \partial_j^\dagger A_j, B\right\rangle_{\mathcal{H}_\mathcal{A}}\\
&= \left\langle \sum_{j\in\mathcal{J}} \partial_j^\dagger A_j, B\right\rangle_{\mathcal{H}_\mathcal{A}}\\
&= \langle 0, B\rangle_{\mathcal{H}_\mathcal{A}}\\
&= 0.
\end{aligned}$$

Thus $\mathrm{Ran}(\nabla) \subseteq (\mathrm{Null}(\mathrm{div}))^\perp$.

(ii) To show the reverse inclusion $(\mathrm{Null}(\mathrm{div}))^{\perp} \subseteq \mathrm{Ran}(\nabla)$ we note that in our finite dimensional setting this is equivalent to showing that $(\mathrm{Ran}(\nabla))^{\perp} \subseteq \left((\mathrm{Null}(\mathrm{div}))^{\perp}\right)^{\perp} = \mathrm{Null}(\mathrm{div})$. Now let $\mathbf{B} \in \mathrm{Ran}(\nabla)$ as before and let $\mathbf{A} \in (\mathrm{Ran}(\nabla))^{\perp}$. Now

$$\begin{aligned} 0 &= \langle \mathbf{A}, \mathbf{B} \rangle_{\mathcal{H}_{\mathcal{A},|\mathcal{J}|}} \\ &= \sum_{j\in\mathcal{J}} \langle A_j, B_j \rangle_{\mathcal{H}_\mathcal{A}} \\ &= \sum_{j\in\mathcal{J}} \langle A_j, \partial_j B \rangle_{\mathcal{H}_\mathcal{A}} \\ &= \sum_{j\in\mathcal{J}} \left\langle \partial_j^\dagger A_j, B \right\rangle_{\mathcal{H}_\mathcal{A}} \\ &= \left\langle \sum_{j\in\mathcal{J}} \partial_j^\dagger A_j, B \right\rangle_{\mathcal{H}_\mathcal{A}} \end{aligned}$$

Since our choice of $\mathbf{B} \in \mathrm{Ran}(\nabla)$ was arbitrary it follows that $\sum_{j\in\mathcal{J}} \partial_j^\dagger A_j = 0$ i.e. $\mathbf{A} \in \mathrm{Null}(\mathrm{div})$. Thus we have shown that $(\mathrm{Ran}(\nabla))^{\perp} \subseteq \mathrm{Null}(\mathrm{div})$.

□

Proposition 6.1.5. $\mathrm{Null}(\nabla) = \mathrm{Null}(\mathcal{L}_0)$

Proof. Note that for $A, B \in \mathcal{A}$

$$\begin{aligned} \langle A, \mathcal{L}_0 B \rangle_{\mathcal{H}_\mathcal{A}} &= \left\langle A, \sum_{j\in\mathcal{J}} \partial_j^\dagger \partial_j B \right\rangle_{\mathcal{H}_\mathcal{A}} \\ &= \sum_{j\in\mathcal{J}} \left\langle A, \partial_j^\dagger \partial_j B \right\rangle_{\mathcal{H}_\mathcal{A}} \\ &= \sum_{j\in\mathcal{J}} \langle \partial_j A, \partial_j B \rangle_{\mathcal{H}_\mathcal{A}} \\ &= \left\langle \sum_{j\in\mathcal{J}} \partial_j A, \sum_{j\in\mathcal{J}} \partial_j B \right\rangle_{\mathcal{H}_\mathcal{A}} \\ &= - \langle \nabla A, \nabla B \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} . \end{aligned}$$

Now if $A \in \mathrm{Null}(\mathcal{L}_0)$ then

$$0 = \langle A, \mathcal{L}_0 A \rangle_{\mathcal{H}_\mathcal{A}} = - \langle \nabla A, \nabla A \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} .$$

Thus $\nabla A = 0$ and consequently $\mathrm{Null}(\mathcal{L}_0) \subseteq \mathrm{Null}(\nabla)$.
Conversely if $A \in \mathrm{Null}(\nabla)$ then for any $B \in \mathcal{H}_{\mathcal{A}}$ we have

$$0 = \langle \nabla B, \nabla A \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} = \langle B, \mathcal{L}_0 A \rangle_{\mathcal{H}_{\mathcal{A}}} .$$

Of course since our choice of B was arbitrary it follows that $\mathcal{L}_0 A = 0$. Thus $\mathrm{Null}(\nabla) \subseteq \mathrm{Null}(\mathcal{L}_0)$. □

6.2 Dirichlet form of the generator

The differential structure we have introduced in the above section now allows us to write the generator $\mathcal{L}$ of a QMS in terms of a Dirichlet form. A Dirichlet form can be thought of as a generalization of the Laplacian. It is used in the study of functional analysis and in particular the study of harmonic functions. Formally a Dirichlet form is a non-negative, definite, symmetric bilinear form on an L^2 space that is Markovian and closed. Extensive work has been done by Cipriani on Dirichlet forms on non-commutative spaces and how these forms relate to the study of quantum statistical mechanics, see [Cip97, Cip08].

Lemma 6.2.1. *For all $s \in [0,1]$, $j \in \mathcal{J}$, and all $A, B \in \mathcal{A}$ we have*

$$\langle \partial_j B, A \rangle_s = \left\langle B, e^{s\omega_j}(e^{-\omega_j} V_j^* A - A V_j^*) \right\rangle_s . \tag{6.2.1}$$

Proof. For any $A, B \in \mathcal{M}_n(\mathbb{C})$ we have

$$\begin{aligned}
\langle \partial_j B, A \rangle_s &= \mathrm{Tr}[\sigma^s (\partial_j B)^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s [V_j, B]^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s [B^*, V_j^*] \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s B^* V_j^* \sigma^{1-s} A] - \mathrm{Tr}[\sigma^s V_j^* B^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s B^* \sigma^{1-s} \sigma^{s-1} V_j^* \sigma^{1-s} A] - \mathrm{Tr}[\sigma^s V_j^* \sigma^{-s} \sigma^s B^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s B^* \sigma^{1-s} \Delta_\sigma^{s-1}(V_j^*) A] - \mathrm{Tr}[\Delta_\sigma^s(V_j^*) \sigma^s B^* \sigma^{1-s} A]
\end{aligned}$$

Now we make use of property (iv) of Theorem 5.0.1. By applying equation (5.0.3) $\Delta_\sigma V_j = e^{-\omega_j} V_j$ to the above we find

$$\begin{aligned}
&\mathrm{Tr}[\sigma^s B^* \sigma^{1-s} \Delta_\sigma^{s-1}(V_j^*) A] - \mathrm{Tr}[\Delta_\sigma^s(V_j^*) \sigma^s B^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s B^* \sigma^{1-s} e^{(s-1)\omega_j} V_j^* A] - \mathrm{Tr}[e^{s\omega_j} V_j * \sigma^s B^* \sigma^{1-s} A] \\
&= \mathrm{Tr}[\sigma^s B^* \sigma^{1-s} e^{s\omega_j}(e^{-\omega_j} V_j^* A - A V_j^*)] \\
&= \left\langle B, e^{s\omega_j}(e^{-\omega_j} V_j^* A - A V_j^*) \right\rangle_s .
\end{aligned}$$

□

Proposition 6.2.2. *For all $s \in [0,1]$ and for all $A, B \in \mathcal{A}$*

$$e^{(1/2-s)\omega_j} \langle \partial_j B, \partial_j A \rangle_s = -\left\langle B, e^{-\omega_j/2} V_j^* [A, V_j] + e^{\omega_j/2} [V_j, A] V_j^* \right\rangle_s . \quad (6.2.2)$$

Proof. We start by applying lemma 6.2.1 to obtain

$$\begin{aligned} e^{(1/2-s)\omega_j} \langle \partial_j B, \partial_j A \rangle_s &= e^{(1/2-s)\omega_j} \left\langle B, e^{s\omega_j} (e^{-\omega_j} V_j^* (\partial_j A) - (\partial_j A) V_j^*) \right\rangle_s \\ &= e^{(1/2-s)\omega_j} \left\langle B, e^{s\omega_j} (e^{-\omega_j} V_j^* [V_j, A] - [V_j, A] V_j^*) \right\rangle_s \\ &= \left\langle B, e^{1/2\omega_j} (e^{-\omega_j} V_j^* [V_j, A] - [V_j, A] V_j^*) \right\rangle_s \\ &= \left\langle B, e^{-1/2\omega_j} V_j^* [V_j, A] - e^{1/2\omega_j} [V_j, A] V_j^* \right\rangle_s \\ &= -\left\langle B, e^{-\omega_j/2} V_j^* [A, V_j] + e^{\omega_j/2} [V_j, A] V_j^* \right\rangle_s \end{aligned}$$

□

Now we define a mapping

$$\mathcal{E}_s(B, A) := \sum_{j \in \mathcal{J}} e^{(1/2-s)\omega_j} \langle \partial_j B, \partial_j A \rangle_s .$$

Recall the expression for the generator $\mathcal{L}$ given by equation (5.0.1)

$$\mathcal{L}A = \sum_{j \in \mathcal{J}} \left(e^{-\omega_j/2} V_j^* [A, V_j] + e^{\omega_j/2} [V_j, A] V_j^* \right)$$

then we have by Proposition 6.2.2 that

$$\mathcal{E}_s(A, B) = -\langle B, \mathcal{L}A \rangle_s .$$

In particular if we consider $s = \frac{1}{2}$ we obtain

$$\mathcal{E}_{1/2}(B, A) = -\langle B, \mathcal{L}A \rangle_{1/2} \quad \mathcal{E}_{1/2}(A, B) := \sum_{j \in \mathcal{J}} \langle \partial_j B, \partial_j A \rangle_{1/2} . \quad (6.2.3)$$

This expresses our generator $\mathcal{L}$ in terms of a Dirichlet form [Cip08].

6.3 Ergodicity and geometry

Theorem 6.3.1. *Let $\mathcal{P}_t = e^{t\mathcal{L}}$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for some non-degenerate $\sigma \in \mathcal{G}_+(\mathcal{A})$. Then the commutant of $\{V_j\}_{j \in \mathcal{J}}$ equals the null space of $\mathcal{L}$, in particular $\mathcal{P}_t$ is ergodic if and only if the commutant of $\{V_j\}_{j \in \mathcal{J}}$ is spanned by the identity.*

Proof. Let $(\{V_j\}_{j\in\mathcal{J}})'$ denote the commutant of $\{V_j\}_{j\in\mathcal{J}}$ and let $A \in (\{V_j\}_{j\in\mathcal{J}})'$. Then by definition $[V_j, A] = \partial_j A = 0$ for all $j \in \mathcal{J}$. Consider the form of the generator given by

$$\mathcal{L}A = \sum_{j\in\mathcal{J}} \left(e^{-\omega_j/2} V_j^*[A, V_j] + e^{\omega_j/2}[V_j, A]V_j^*\right)$$

then clearly $\mathcal{L}A = 0$. Thus $(\{V_j\}_{j\in\mathcal{J}})' \subseteq \text{Null}(\mathcal{L})$.

Conversely suppose that $\mathcal{L}A = 0$. From equation (6.2.3) we find

$$0 = -\langle A, \mathcal{L}A\rangle_{1/2} \tag{6.3.1}$$

so that $\sum_{j\in\mathcal{J}} \langle \partial_j A, \partial_j A\rangle_{1/2} = 0$.

Since $\langle \partial_j A, \partial_j A\rangle_{1/2} \geq 0$, equation (6.3.1) can only be true if $\partial_j A = 0$ for all $j \in \mathcal{J}$. Hence $[V_j, A] = 0$ for all $j \in \mathcal{J}$ and $\text{Null}(\mathcal{L}) \subseteq (\{V_j\}_{j\in\mathcal{J}})'$. □

The next theorem is a powerful geometric result that will be of great use when we consider gradient flows in the next Chapter.

Theorem 6.3.2. *et $\mathcal{P}_t = e^{t\mathcal{L}}$ be an ergodic QMS on $\mathcal{A}$ that satisfies the σ-DBC for some non-degenerate $\sigma \in \mathcal{G}_+(\mathcal{A})$. Let $\mathcal{L}_0$ be the associated Laplacian operator. For a given $B \in \mathcal{H}_\mathcal{A}$ the equation*

$$\mathcal{L}_0 X = B$$

has a solution if and only if $\tau[B] = 0$. Consequently when $\tau[B] = 0$, there exists a non-trivial affine subspace of $\mathcal{H}_{\mathcal{A},\mathcal{J}}$ consisting of the elements $\mathbf{A}$ such that $\text{div}\,\mathbf{A} = B$.

Proof. Note that we are working in a finite dimensional setting so we can therefore take the dimension of $\mathcal{H}_\mathcal{A}$ to be n^2 for some $n \in \mathbb{N}$.
Let $\mathcal{S} = \{X \in \mathcal{H}_\mathcal{A} : \tau[X] = 0\}$. We are required to show that $\text{Ran}(\mathcal{L}_0) = \mathcal{S}$.

(i) Suppose $B \in \text{Ran}(\mathcal{L}_0)$ then there exists an element $X \in \mathcal{H}_\mathcal{A}$ such that

$$\mathcal{L}_0 X = B.$$

Note that

$$\begin{aligned}\mathcal{L}_0 X &= \sum_{j\in\mathcal{J}} \partial_j^\dagger \partial_j X \\ &= \sum_{j\in\mathcal{J}} [V_j^*, [V_j, X]]\end{aligned}$$

$$\begin{aligned}\tau(\mathcal{L}_0 X) &= \sum_{j\in\mathcal{J}} \tau(V_j^* V_j X - V_j X V_j^* - V_j^* X V_j + X V_j V_j^*) \\ &= \sum_{j\in\mathcal{J}} \tau(V_j^* V_j X) - \tau(V_j X V_j^*) - \tau(V_j^* X V_j) + \tau(X V_j V_j^*) \\ &= \sum_{j\in\mathcal{J}} \tau(V_j^* V_j X) - \tau(V_j^* V_j X) - \tau(V_j^* X V_j) + \tau(V_j^* X V_j) \\ &= 0.\end{aligned}$$

Hence $\tau[B] = 0$ and consequently we have that $\mathrm{Ran}(\mathcal{L}_0) \subset \mathcal{S}$.

(ii) In our finite dimensional setting we think of the elements of $\mathcal{H}_\mathcal{A}$ as $n \times n$ matrices. The restriction $\tau[B] = 0$ on $B \in \mathcal{H}_\mathcal{A}$ implies that

$$\dim(\mathcal{S}) = n^2 - 1.$$

Now since $\mathcal{P}_t$ is ergodic it follows from Theorem 6.3.1 that $\mathrm{Null}(\nabla) = 0$. By proposition 6.1.5 we then have that

$$\mathrm{Null}(\mathcal{L}_0) = \mathrm{Null}(\nabla) = 1.$$

Simple application of the rank theorem shows that

$$\dim(\mathrm{Ran}(\mathcal{L}_0)) = \dim(\mathcal{H}_\mathcal{A}) - \dim(\mathrm{Null}(\mathcal{L}_0)) = n^2 - 1.$$

Hence $\dim(\mathrm{Ran}(\mathcal{L}_0)) = \dim(\mathcal{S})$ which in our finite dimensional setting implies that $\mathrm{Ran}(\mathcal{L}_0) = \mathcal{S}$.

□

6.4 Non-commutative multiplication

With the structure of our vector fields and the operations we can perform on them in place, we are well on our way to defining a quantum analogue of equation (6.0.1). All that remains is to define a sensible notion of multiplication between a density operator ρ and a vector field $\mathbf{A} \in \mathcal{H}_{\mathcal{A},\mathcal{J}}$. To that end we will start with defining a notion of multiplication between ρ and a single element A_j of $\mathbf{A}$.

First we prove a useful chain rule type identity.

Lemma 6.4.1. *For all $V \in \mathcal{M}_n(\mathbb{C}), \rho \in \mathcal{G}_+$ and $\omega \in \mathbb{R}$, we have*

$$\int_0^1 e^{\omega(s-1/2)} R_\rho \Delta_\rho^s \big(V \ln(e^{-\omega/2}\rho) - \ln(e^{\omega/2}\rho)V\big) ds = e^{-\omega/2} V\rho - e^{\omega/2}\rho V, \tag{6.4.1}$$

where R_ρ denotes right multiplication by ρ.

Proof. Define a function $f(s) = e^{\omega(1/2-s)}\rho^{1-s}V\rho^s$. Note that the right hand side of equation (6.4.1) is then given by $f(1) - f(0)$.

Now we take the derivative of f with respect to s to obtain

$$\begin{aligned} f'(s) &= e^{\omega(1/2-s)}\rho^{1-s}\big(-\omega V - \ln(\rho)V + V\ln(\rho)\big)\rho^s \\ &= e^{\omega(1/2-s)}\rho^{1-s}\big(\ln(e^{-\omega})V - \ln(\rho)V + V\ln(\rho)\big)\rho^s \\ &= e^{\omega(1/2-s)}\rho^{1-s}\big(\ln(e^{-\omega/2})V - \ln(e^{\omega/2})V - \ln(\rho)V + V\ln(\rho)\big)\rho^s \\ &= e^{\omega(1/2-s)}\rho^{1-s}\big(V\ln(e^{-\omega/2}\rho) - \ln(e^{\omega/2}\rho)V\big)\rho^s. \end{aligned}$$

Note that by substituting $1-s$ in the above equation we have

$$\begin{aligned} f'(1-s) &= e^{\omega(s-1/2)}\rho^s\big(V\ln(e^{-\omega/2}\rho) - \ln(e^{\omega/2}\rho)V\big)\rho^{1-s} \\ &= e^{\omega(s-1/2)}R_\rho\Delta_\rho^s\big(V\ln(e^{-\omega/2}\rho) - \ln(e^{\omega/2}\rho)V\big) \end{aligned}$$

We have shown that the left hand side of equation (6.4.1) is

$$\int_0^1 f'(1-s)ds. \tag{6.4.2}$$

The result then follows from the fundamental theorem of calculus. □

Now consider a function

$$f_\omega := \int_0^1 e^{\omega(s-1/2)} t^s ds = e^{\omega/2}\frac{t - e^\omega}{\ln(t) + \omega}. \tag{6.4.3}$$

The following is a simple calculation to show the equality in equation (6.4.3)

$$\begin{aligned}\int_0^1 e^{\omega(s-1/2)}t^s ds &= \int_0^1 e^{\omega(s-1/2)}e^{s\ln t}ds\\ &= \int_0^1 e^{\omega s-1/2\omega+s\ln(t)}ds\\ &= [\omega+\ln(t)]^{-1}e^{\omega s-1/2\omega+s\ln(t)}|_0^1\\ &= [\omega+\ln(t)]^{-1}[e^{\omega/2+s\ln(t)}-e^{-\omega/2}]\\ &= \frac{e^{\omega/2}t-e^{-\omega/2}}{\omega+\ln(t)}\\ &= \frac{e^{\omega/2}(t-e^{-\omega})}{\omega+\ln(t)}.\end{aligned}$$

Notice that $f_\omega(\Delta_\rho)$ becomes $\int_0^1 e^{\omega(s-1/2)}\Delta_\rho^s ds$ then the right hand side of equation (6.4.1) can be rewritten as

$$R_\rho f_\omega(\Delta_\rho)\big(V\ln(e^{-\omega/2}\rho)-\ln(e^{\omega/2}\rho)V\big)$$

so our chain rule identity becomes

$$R_\rho f_\omega(\Delta_\rho)\big(V\ln(e^{-\omega/2}\rho)-\ln(e^{\omega/2}\rho)V\big)=e^{-\omega/2}V\rho-e^{\omega/2}\rho V. \tag{6.4.4}$$

Note that setting $\omega=0$ in equation (6.4.4) we obtain the commutator identity

$$R_\rho f_\omega(\Delta_\rho)[V,\ln(\rho)]=[V,\rho]. \tag{6.4.5}$$

For $A\in\mathcal{H}_\mathcal{A}$, $\rho\in\mathcal{G}_+$, we can think of the operation $A\mapsto R_\rho f_\omega(\Delta_\rho)A$ as a form of non-commutative multiplication of A by ρ. The following proposition serves to justify this viewpoint.

Proposition 6.4.2. *For $A\in\mathcal{H}_\mathcal{A}$ and $\rho\in\mathcal{G}_+$, if A commutes with ρ then*

$$f_\omega(\Delta_\rho)A=\alpha_\omega A,$$

where α_ω is defined as follows:

$$\alpha_\omega=\begin{cases}1 & \omega=0\\ \frac{\omega}{e^{\omega/2}-e^{-\omega/2}} & \omega\neq 0\end{cases}$$

Proof.

i If $\omega = 0$ then

$$\begin{aligned} f_0(\Delta_\rho)A &= \int_0^1 \Delta_\rho^s A ds \\ &= \int_0^1 \rho^s A \rho^{-s} ds \\ &= A \int_0^1 \rho^s \rho^{-s} ds \\ &= A. \end{aligned}$$

ii If $\omega \neq 0$ then

$$\begin{aligned} f_\omega(\Delta_\rho)A &= \int_0^1 e^{\omega(s-1/2)} \Delta_\rho^s A ds \\ &= \int_0^1 e^{\omega(s-1/2)} \rho^s A \rho^{-s} ds \\ &= \int_0^1 e^{\omega(s-1/2)} A \rho^s \rho^{-s} ds \\ &= \int_0^1 e^{\omega(s-1/2)} A ds \\ &= \frac{e^{\omega/2} - e^{-\omega/2}}{\omega} A \end{aligned}$$

□

In light of the above we can view $R_\rho f_\omega$ as a non-commutative multiplication operation. Observe that in the special case of $\omega = 0$ and A commutes with ρ, we recover the usual commutative multiplication by ρ. Since α_ω is simply a linear constant we can still think of this as being the usual multiplication even in the case where we do not have $\omega = 0$.

Definition 6.4.3. For $\rho \in \mathcal{G}_+$, and $\omega \in \mathbb{R}$, define the operator $[\rho]_\omega : \mathcal{M}_n(\mathbb{C}) \mapsto \mathcal{M}_n(\mathbb{C})$ by:

$$[\rho]_\omega = R_\rho \circ \alpha_\omega f_\omega(\Delta_\rho). \tag{6.4.6}$$

We then view the operator $[\rho]_\omega$ as a form of non-commutative multiplication by ρ, parameterized by ω. Next we prove a statement about the invertibility of this operator. By inspection the inverse of $[\rho]_\omega$ is simply given by $[\rho]_\omega^{-1} = (1/f_\omega)(\Delta_\rho) \circ R_{\rho-1}$. This expression is not particularly useful to us as it requires finding the inverse of $f_\omega(t)$. We can derive an equivalent expression for $[\rho]_\omega^{-1}$. The following two identities which hold for $\lambda, \mu > 0$ will be used in the proof:

$$\int_0^1 \lambda^{1-s}\mu^s ds = \frac{\lambda - \mu}{\ln(\lambda) - \ln(\mu)}, \tag{6.4.7}$$

and

$$\int_0^\infty \frac{1}{(t+\lambda)(t+\mu)} dt = \frac{\ln(\lambda) - \ln(\mu)}{\lambda - \mu}. \tag{6.4.8}$$

Note that the right hand side of equation (6.4.7) and equation (6.4.8) are inverses of each other.

Proposition 6.4.4. *For all $\omega \in \mathbb{R}$, the operator $[\rho]_\omega$ is invertible with*

$$[\rho]_\omega^{-1} = \int_0^\infty \frac{1}{(t + e^{\omega/2}L_\rho) + (t + e^{-\omega/2}R_\rho)}$$

Proof. We start by writing out the operator $[\rho]_\omega$ using the definition of f_ω given by equation (6.4.3). We use the notation L_ρ to denote left multiplication by ρ here for the sake of clarity. Then we have

$$\begin{aligned}
[\rho]_\omega &= R_\rho f_\omega(\Delta_\rho) \\
&= R_\rho \int_0^1 e^{\omega(s-1/2)} \Delta_\rho^s ds \\
&= \int_0^1 e^{\omega(s-1/2)} R_\rho L_\rho^s R_\rho^{-s} ds \\
&= \int_0^1 e^{\omega(s-1/2)} L_\rho^s R_\rho^{1-s} ds \\
&= \int_0^1 (e^{\omega s} L_\rho^s)(e^{-\omega/2} R_\rho^{1-s}) \\
&= \int_0^1 (e^{\omega/2} L_\rho)^s (e^{\omega s/2} e^{-\omega/2} R_\rho^{1-s}) ds \\
&= \int_0^1 (e^{-\omega/2} L_\rho)^s (e^{\omega/2} R_\rho)^{1-s} ds
\end{aligned} \tag{6.4.9}$$

From the above calculation it is clear that by setting $\lambda = e^{\omega/2}R_\rho$ and $\mu = e^{-\omega/2}L_\rho$ we have an equation of the form of (6.4.7). The inverse of this equation is given by the left hand side of equation (6.4.8). After substituting back for λ and μ we obtain

$$[\rho]_\omega^{-1} = \int_0^\infty \frac{1}{(t + e^{\omega/2}L_\rho) + (t + e^{-\omega/2}R_\rho)}. \tag{6.4.10}$$

□

From the integral formulations of the operators $[\rho]_\omega$ and $[\rho]_\omega^{-1}$ it is easy to see that the mappings $\rho \mapsto [\rho]_\omega$ and $\rho \mapsto [\rho]_\omega^{-1}$ are infinitely differentiable with respect to ρ.

Corollary 6.4.5. *For all $A \in \mathcal{H}_\mathcal{A}$*

$$([\rho]_\omega A)^* = [\rho]_{-\omega} A^* \tag{6.4.11}$$

and consequently

$$([\rho]_\omega^{-1} A)^* = [\rho]_{-\omega}^{-1} A^* \tag{6.4.12}$$

Proof. We start from equation (6.4.9) to obtain

$$\begin{aligned}
([\rho]_\omega A)^* &= A^* \int_0^1 \left((e^{-\omega/2} R_\rho)^{1-s}\right)^* \left((e^{\omega/2} L_\rho)^s\right)^* ds \\
&= A^* \int_0^1 (e^{-\omega/2} R_\rho)^{1-s} (e^{\omega/2} L_\rho)^s ds \\
&= [\rho]_{-\omega} A^*
\end{aligned}$$

where we have made use of the fact that $\rho^* = \rho$. □

We now have constructed a way to both multiply and divide our vector fields by a density matrix ρ.

By applying equation (6.4.5) to a density matrix ρ and setting $V = V_j$ for some $V_j \in \{V_j\}_{j \in \mathcal{J}}$ we obtain

$$[\rho]_0(\partial_j \ln \rho) = R_\rho f_0(\Delta_\rho)(\partial_j \ln \rho) = \partial_j \rho. \tag{6.4.13}$$

The above equation can be viewed as the quantum analogue of the classical identity that holds for all positive, smooth functions $p(x)$

$$p(x) \nabla \ln p(x) = \nabla p(x). \tag{6.4.14}$$

Proposition 6.4.6.

Proof. Note that from the definition of ∇ we have that

$$[\rho]_0 \nabla \ln \rho = [\rho]_0 (\partial_1 \ln \rho, ..., \partial_\mathcal{J} \ln \rho).$$

From equation (6.4.13) we then have

$$[\rho]_0 \nabla \ln \rho = (\partial_1 \rho, ..., \partial_\mathcal{J} \rho) = \nabla \rho.$$

Recall that $\mathcal{L}_0 = \operatorname{div} \circ \nabla$ and the result follows. □

6.5 Relative entropy

Our main goal in the next chapter is to show that there is a Riemannian metric on $\mathcal{G}_+$ such that the quantum heat flow equation can be written as gradient flow for the relative entropy functional of ρ with respect to the normalized trace. First we will need to rewrite $\mathcal{L}_0^\dagger\rho$ in such a way that the connection to the relative entropy becomes clear.

Recall the definition of the relative entropy of a ρ with respect to σ is given by

$$D(\rho||\sigma) = \mathrm{Tr}[\rho(\ln\rho - \ln\sigma)].$$

Lemma 6.5.1. *Let $\mathcal{P}_t = e^{t\mathcal{L}}$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for $\sigma \in \mathcal{G}_+(\mathcal{A})$, and let $\mathcal{L}$ be given in the form of equation (5.0.2). Then for all $\rho \in \mathcal{G}_+$ and all $j \in \mathcal{J}$, we have that*

$$\partial_j(\ln\rho - \ln\sigma) = V_j \ln(e^{-\omega_j/2}\rho) - \ln(e^{\omega_j/2}\rho)V_j.$$

Proof. Recall that σ is by definition a positive definite matrix of unit trace. Thus it is always possible to choose a basis where σ is diagonal. Choosing such a basis we can write

$$\sigma = e^{\kappa}$$

for some $\kappa \in \mathcal{M}_n(\mathbb{C})$. Now

$$[V_j, \ln\sigma] = [V_j, \kappa] = V_j\kappa - \kappa V_j,$$

and

$$\begin{aligned}\Delta_\rho^s(V_j) &= \sigma^s V_j \sigma^{-s}\\ &= e^{s\kappa}V_j e^{-s\kappa}.\end{aligned}$$

Taking the partial derivative of the above expression with respect to s we obtain

$$\partial_s \Delta_\rho^s V_j = \kappa e^{s\kappa}V_j e^{-s\kappa} - e^{s\kappa}V_j e^{-s\kappa}\kappa.$$

The above expression evaluated at zero simply gives back the commutator identity thus we conclude that

$$\partial_s|_{s=0}\Delta_\rho^s V_j = \kappa V_j - V_j\kappa = -[V_j, \kappa].$$

Recall that $\Delta_\rho^s V_j = e^{-s\omega_j}V_j$ by definition of our elements V_j. Now

$$\begin{aligned}\frac{\partial}{\partial s}\Delta_\rho^s V_j &= \frac{\partial}{\partial s}e^{-s\omega_j}V_j\\ &= -\omega_j e^{-s\omega_j}V_j.\end{aligned}$$

Evaluating this partial derivative at zero we get

$$\frac{\partial}{\partial s}|_{s=0}\Delta_\rho^s V_j = -\omega_j V_j.$$

So we conclude that

$$[V_j, \ln\sigma] = -\partial_s|_{s=0}\Delta_\rho^s V_j = \omega_j V_j.$$

Now consider

$$\begin{aligned}
\partial_j(\ln\rho - \ln\sigma) &= [V_j, (\ln\rho - \ln\sigma)] \\
&= [V_j, \ln\rho] - [V_j, \ln\sigma] \\
&= [V_j, \ln\rho] - \omega_j V_j \\
&= V_j(\ln\rho) - (\ln\rho)V_j - (\ln e^{\omega_j})V_j \\
&= V_j(\ln\rho) - (\ln\rho)V_j - (\ln e^{\omega_j/2} + \ln e^{\omega_j/2})V_j \\
&= V_j(\ln\rho) - (\ln\rho)V_j + (\ln e^{-\omega_j/2})V_j - (\ln e^{\omega_j/2})V_j \\
&= V_j \ln(e^{-\omega_j/2}\rho) - \ln(e^{\omega_j/2}\rho)V_j,
\end{aligned}$$

where in the final line we grouped the first and second terms with the third and fourth terms respectively. □

The next theorem follows easily from Lemma 6.5.1 and the chain rule identity we derived in equation (6.4.4).

Theorem 6.5.2. *Let $\mathcal{P}_t$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for a $\sigma \in \mathcal{G}_+(\mathcal{A})$, and let $\mathcal{L}$ be given in the form of equation (5.0.2). Then, for all $\rho \in \mathcal{G}_+$,*

$$-\mathcal{L}^\dagger\rho = \sum_{j\in\mathcal{J}} \partial_j^\dagger\left([\rho]_{\omega_j}\partial_j(\ln\rho - \ln\sigma)\right). \tag{6.5.1}$$

Proof. We start from the right hand side of equation (6.5.1) and first apply Lemma 6.5.1

$$\begin{aligned}
\sum_{j\in\mathcal{J}} \partial_j^\dagger\left([\rho]_{\omega_j}\partial_j(\ln\rho - \ln\sigma)\right) &= \sum_{j\in\mathcal{J}} \partial_j^\dagger\left([\rho]_{\omega_j}\left(V_j\ln(e^{-\omega_j/2}\rho) - \ln(e^{\omega_j/2}\rho)V_j\right)\right) \\
&= \sum_{j\in\mathcal{J}} \partial_j^\dagger\left(e^{-\omega_j/2}V_j\rho - e^{\omega_j/2}\rho V_j\right) \\
&= -\sum_{j\in\mathcal{J}}\left(e^{-\omega_j/2}[V_j\rho, V_j^*] + e^{\omega_j/2}[V_j^*, \rho V_j]\right) \\
&= -\mathcal{L}^\dagger\rho,
\end{aligned}$$

where in the second line we applied equation (6.4.4), in the third line we simply applied the definition of $\partial_j^\dagger$. □

Equation (6.5.1) is the main result of this Chapter and will serve as our springboard to the next Chapter where we will construct a Riemannian metric such that the quantum heat flow equation can be written as gradient flow for the relative entropy functional.

Chapter 7

Gradient flow on a Riemannian metric

In the previous chapter we derived a quantum mechanical analog of the classical continuity equation. In this chapter we shall discuss a few concepts related to differential geometry and gradient flows on Riemannian metric. Finally in sections 7.5 and 7.6 we will combine these concepts with the structures defined in Chapter 6 to arrive at our final result, which is a the time evolution of our density matrices written as gradient flow with respect to the relative entropy functional. This is a non-commutative analogue of the time evolution of probability densities written as gradient flow with repsect to the 2-Wasserstein metric, discussed in the section below.

7.1 The 2-Wasserstein metric

It was shown by Felix Otto that a large class of partial differential equations describing the time evolution of probability densities on $p(x,t)$ on $\mathbb{R}^n$, can be viewed as gradient flow with respect to the 2-Wasserstein metric [Ott01]. First let's define exactly what we mean by gradient flow.

Definition 7.1.1. (**Gradient flow**) Let X be a linear manifold and let $x : \mathbb{R} \to X$ be a parameterised curve on X. Let $F(x,t)$ denote some functional. The curve $x(t)$ is said to be *gradient flow* with respect to $F(x,t)$ if

$$x'(t) = -\nabla F(x,t). \tag{7.1.1}$$

Thus the rate of change of the curve is given by the direction of steepest descent for some functional $F(x,t)$. To make sense of gradient flow we require three ingredients: A differentiable manifold M, a metric tensor g

such that (M, g) becomes a Riemannian manifold and a functional F on the manifold. In the physically motivated view of Otto this functional is usually the energy of a system, or some other thermodynamic quantity.

Notice that the left hand side of equation (7.1.1) is a statement about the dynamics of the system, i.e. how the input and output spaces are related, whereas the right hand side is usually a statement about the statistical behaviour of the system.

It was shown by Benamou and Brenier in [BB00] that the optimal transport problem can be expressed in the language of Lagrangian and Hamiltonian dynamics by minimizing a type of action integral. More specifically suppose we have an initial mass distribution on $\mathbb{R}$ of ρ_0 and a target mass distribution of ρ_1. We want to transport our initial distribution ρ_0 to our target distribution ρ_1 via a transport plan $T : x \to T(x)$ that is optimal, in the Monge formulation. Recall the definition of the 2-Wasserstein distance from equation (2.3.2):

$$W_2(\rho_0, \rho_1) = \inf \left\{ \int ||x - T(x)||^2 \rho_0(x) \right\}. \tag{7.1.2}$$

From the discussion on the Monge formulation in section 2.1 we know we can rewrite the above using equation (2.1.7) to obtain:

$$W_2(\rho_0, \rho_1) = \inf \left\{ \rho_1(T(x)) \det |\nabla T(x)| \right\}. \tag{7.1.3}$$

Benamou and Brenier introduced the idea of time-varying distributions $\rho(x, t)$, $t \in [0, 1]$ and velocity fields $v(x, t) \in \mathbb{R}^n$ that satisfy a continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0,$$

and is subject to the following boundary conditions

$$\rho(x, 0) = \rho_0, \quad \rho(x, 1) = \rho_1.$$

The integral in equation (7.1.3) can then be conveniently rewritten as a type of action integral in the form

$$W_2(\rho_0, \rho_1) = \inf \left\{ \int_{-\infty}^{\infty} \int_0^1 (\rho(x, t))^{-1} ||u(x, t)||^2 dt dx \right\}, \tag{7.1.4}$$

where $u(x, t) = \rho(x, t)v(x, t)$ is called the momentum field. This work provided a Lagrangian formulation of the optimal transport problem, motivated by physical systems. It was this physical insight that led to the development of the geometry induced by the Wasserstein distance and its connection to entropy functionals. The key idea first originated from work by Jordan, Kinderleher and Otto in [JKO98] and was further developed by Otto in [Ott01]. We briefly frame the key ideas here before proceeding to the non-commutative case.

7.2 Differential geometry

We briefly review some elementary aspects of differential geometry. The reader is referred to [Hit12] and [Car19] for a more in depth study of the topics discussed here. The principal object of study in differential geometry are so called differentiable manifolds. Intuitively a manifold is an n-dimensional space that locally resembles $\mathbb{R}^n$. We can make this concept mathematically precise with the help of the concepts of coordinate charts and atlases. We start with the concept of a coordinate chart.

Definition 7.2.1. A *coordinate chart* on a set M is a subset $U \in M$ together with a bijection ϕ such that

$$\phi : U \to \phi(U) \in \mathbb{R}^n,$$

where $\phi(U)$ is an open set in $\mathbb{R}^n$.

The above definition allows us to parameterize a point $x \in U$ using n coordinates. The coordinates are given by $\phi(x) = (x_1, x_2, ..., x_n)$. The coordinate chart is the ordered pair (U, ϕ). Next we can define the notion of an atlas.

Definition 7.2.2. An *atlas* is an indexed collection of coordinate charts $\{U_\alpha, \phi_\alpha\}_{\alpha \in I}$ that satisfies the following properties:

1. The union of U_α covers M.
2. $\phi_\alpha(U_\alpha \cap U_\beta)$ is open in $\mathbb{R}^n$ for all $\alpha, \beta \in I$.
3. The map $\phi_\beta \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$ is C^∞.

It turns out that the existence of an atlas is sufficient to define a manifold. Property 3 above ensures that different charts are smoothly sewn together. If we think of a chart as similar to a coordinate system on some open set, then an atlas is simply a collection of coordinate systems that are smoothly related where they overlap. Of course we want these objects to be independent of the basis of $\mathbb{R}^n$ that we are using. So we define the notion of compatible atlases.

Definition 7.2.3. Two atlases $\{(U_\alpha, \psi_\alpha)\}$ and $\{(V_\beta, \psi_\beta)\}$ are said to be *compatible* if their union is also an atlas.

Clearly compatibility defines an equivalence relation and we define the following:

Definition 7.2.4. A *differentiable structure* is an equivalence class of atlases.

Definition 7.2.5. An *n-dimensional manifold* M is a a space M along with a differentiable structure.

Put slightly differently an n-dimensional manifold is simply a set M along with a maximal atlas, that contains every other compatible atlas. This definition ensures that two equivalent spaces equipped with different atlases are not regarded as different manifolds.

Next we will how to describe to move from one manifold to another using the idea of a *smooth map.*

Definition 7.2.6. A *smooth map* between manifolds is a map $F : M \to N$ such that for every chart (U_α, ϕ_α) of M and $x \in U_\alpha$ and chart (V_β, ψ_β) of N with $F(x) \in V_\beta$, the set $F^{-1}(V_\beta)$ is open in M and the composite map $\psi_\beta F \phi_\alpha^{-1}$ is C^∞. A smooth map F is called a *diffeomorphism* if it's inverse F^{-1} is also a smooth map .

Our next task is to define the concepts that will allow us to define the derivative of a function in such a way that it is independent of the chosen coordinate system. First we define a cotangent space:

Definition 7.2.7. Let $C^\infty(M)$ denote the set of smooth maps on a manifold M and let Z_a denote the subset of those functions whose derivative vanishes at a point $a \in M$. We define the *cotangent space* T_a^* as the quotient space

$$T_a^* = C^\infty(M)/Z_a.$$

Next we list some important properties of the cotangent space T_a^*:

Proposition 7.2.8. *Le M be an n-dimensional manifold and let $a \in M$ then:*

1. *T_a^* forms an n-dimensional vector space.*
2. *If (U, ϕ) is a coordinate chart around $x \in U$ with coordinates $(x_1, ..., x_n)$, then the elements $(d(x_1)_a, ..., d(x_n)_a$ form a basis of T_a^*.*

3. *If* $f \in C^\infty(M)$ *then*

$$(df)_a = \sum_i \frac{\partial f}{\partial x_i} d(x_i).$$

Now we can define the tangent space:

Definition 7.2.9. The *tangent space* T_a at $a \in M$ is the dual of the cotangent space T_a^*.

Note that if $(x_1, ..., x_n)$ is a local coordinate system at a then $\left(\left(\frac{\partial}{\partial x_1}\right), ..., \left(\frac{\partial}{\partial x_n}\right)_a\right)$ is a basis for T_a. Finally we define the concept of a Riemannian metric on a manifold M.

Definition 7.2.10. A *Riemannian metric* on a manifold M is a smooth strictly positive inner product on the tangent space T_x at a point $x \in M$. A manifold that admits such a metric is called a *Riemannian manifold.*

With this rudimentary understanding of differential geometry we can now apply these concepts to our understanding of manifolds and gradient flows.

7.3 Riemannian manifold on probability densities

We now briefly consider the connections between the geometry induced by the Wasserstein distance on a manifold of probability densities, entropy functionals and the heat equation. This will serve as a roadmap for when we move on to case of a manifold of density matrices. First we consider a manifold $\mathcal{D}$ of probability densities on $\mathbb{R}^n$.

$$\mathcal{D} = \left\{ \rho \geq 0 : \int_{\mathbb{R}^n} \rho = 1 \right\}.$$

We identify the tangent space with the set of functions that integrate to zero.

$$T_\rho \cong \left\{ \delta : \int_{\mathbb{R}} \delta = 0 \right\}.$$

The idea that a manifold of scalar probability densities admits a Riemannian structure originally came from the work of Jordan *et al* in [JKO98]. The geometric approach was further developed and refined by Otto in [Ott01] to study the dynamics of porous medium systems. The essential idea is that under physically motivated assumptions of differentiability of the functions ρ and δ, one can solve the Poisson equation

$$\delta = -\nabla \cdot (\rho \nabla u), \tag{7.3.1}$$

Thus we can identify an element of the tangent space with a unique function u, up to an additative constant of course. Let u_δ denote the solution to equation (7.3.1) with regard to a fixed δ. The vector field corresponding to u_δ is then denoted by $v_\delta = \nabla u_\delta$. The Riemannian metric is defined by the following inner product for δ_1,$\delta_2 \in T_\rho$

$$\langle \delta_1, \delta_2 \rangle_\rho = \int \rho \langle v_{\delta_1}, v_{\delta_2} \rangle , \tag{7.3.2}$$

where the inner product on the right hand side is the usual dot product between vectors in $\mathbb{R}^n$. Rewriting equation (7.3.2) in terms of the functions u_δ and applying simple integration by parts we obtain the norm of a given function δ as follows:

$$\begin{aligned} ||\delta||^2 &= \langle \delta, \delta \rangle_\rho \\ &= \int \rho \langle \nabla u_\delta, \nabla u_\delta \rangle \\ &= -\int u_\delta \nabla \cdot (\rho \nabla u_\delta) \\ &= \int u_\delta \delta, \end{aligned}$$

where in the third line we applied integration by parts. We follow the example of Otto in [Ott01] and assume suitable conditions such that the boundary terms vanish. It was further shown by Otto in [Ott01] that using this characterization one can show that the metric induces exactly the Wasserstein distance as given by equation (2.3.2). Given an initial state ρ_0 and a final state ρ_1, the path that minimizes the formulation was found to coincide with a displacement interpolating curve first studied by Robert McCann in [McC97]. Here it was shown that in the study of interacting gas models an interpolation curve can be found to obtain the minimum energy state of such systems. The displacement curve that solves this minimum was found to coincide exactly with the minimizer shown by Otto. The path $\rho(t)$ that is such a minimizer is also referred to as a Wasserstein geodesic.

7.4 Gradient flow of Shannon entropy

The connection between the Wasserstein geometry induced on probability densities and entropy functionals was also studied by Otto in [Ott01]. Let $\rho \in \mathcal{D}$ be as before and consider the relative Shannon entropy of ρ defined by

$$S(\rho) = -\int_{\mathbb{R}^n} \rho \ln \rho.$$

Evaluating this entropy functional along the minimizing path $\rho(t)$ and taking the derivative with respect to time we obtain

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{d}{dt}\int_{\mathbb{R}^n} \rho(t) \ln \rho(t) \\
&= -\int_{\mathbb{R}^n} \frac{\partial \rho}{\partial t} \ln \rho(t) + \rho(t) \frac{\partial \ln \rho(t)}{\partial t} \\
&= -\int_{\mathbb{R}^n} \frac{\partial \rho}{\partial t} \ln \rho(t) + \frac{\partial \rho}{\partial t} \\
&= -\int_{\mathbb{R}^n} \frac{\partial \rho}{\partial t} \ln \rho(t) - \frac{d}{dt}\int_{\mathbb{R}^n} \rho(t) \\
&= -\int_{\mathbb{R}^n} \frac{\partial \rho}{\partial t} \ln \rho(t),
\end{aligned}$$

From the characterization of the norm $||\delta||^2 = \int u_\delta \delta$ we see that the direction of steepest descent with respect to the Wasserstein metric is given by $u = -\ln \rho$. Thus we have

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \nabla \ln \rho).$$

The above is known as the linear heat equation and we have shown it corresponds to gradient flow of the relative entropy, of course with respect to the Wasserstein geometry, where we have made use of the fact that $\int_{\mathbb{R}^n} \rho(t) = 1$.

7.5 Riemannian metrics on density matrices

We now proceed to apply the ideas of the previous discussion on Riemannian manifolds of probability densities to that of our quantum Markov semigroups. Let $\mathcal{P}_t$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for some

$\sigma \in \mathcal{G}_+(\mathcal{A})$. We shall make use of the standard form of the generator $\mathcal{L}$ given by equation (5.0.2) to construct a metric on $\mathcal{G}_+$ such that the flow given by $\mathcal{P}_t^\dagger$ is gradient flow for the relative entropy with respect to σ. The notions of gradient and divergence constructed in section 6 will be used to construct this metric. We shall assume that $\mathcal{P}_t$ is ergodic.

Let $\rho(t)$, $t \in [0, 1]$, be any differentiable path in $\mathcal{G}_+$. Since we are working in a finite dimensional setting we can view the elements of $\mathcal{G}_+$ as strictly positive matrices of unit trace. We denote the time derivative of $\rho(t)$ by $\dot{\rho}(t)$, where $\rho(t) \in \mathcal{A}$. If $\rho(t)$ is any differential path on $\mathcal{G}_+$ defined on an interval $(-\epsilon, \epsilon)$ for some $\epsilon > 0$, then

$$\mathrm{Tr}[\dot{\rho}(0)] = \frac{d}{dt}\,\mathrm{Tr}[\rho(0)] = 0.$$

Thus we can apply Theorem 6.3.2 to conclude that there exists an affine subspace of $\mathcal{H}_{A,\mathcal{J}}$ consisting of elements of $\mathbf{A}$ such that

$$\dot{\rho}(0) = \mathrm{div}\,\mathbf{A}. \tag{7.5.1}$$

Recall from the discussion at the beginning of the previous section that we need to obtain a quantum analogue of the continuity equation

$$\frac{\partial}{\partial t}\rho(x,t) + \mathrm{div}[\bar{v}(x,t)\rho(x,t)] = 0, \tag{7.5.2}$$

which arises when we have

$$\frac{\partial}{\partial t}\rho(x,t) = \mathrm{div}[\bar{a}(x,t)], \tag{7.5.3}$$

and set

$$\bar{v}(x,t) = -\frac{\bar{a}(x,t)}{\rho(x,t)}.$$

Thus our goal in this section is to define a suitable structure on the manifold of $\mathcal{G}_+$ such that the notion of gradient flow makes sense. We also need to define a Riemannian metric such that the flow given by $\mathcal{P}_t$, which is to say the set of curves defined by $\frac{\partial}{\partial t}\rho = \mathcal{L}^\dagger\rho$ can be written as

$$\mathcal{L}^\dagger\rho = \mathrm{grad}\,D(\rho||\sigma), \tag{7.5.4}$$

where the gradient is now w.r.t the metric defined on the manifold. In the previous chapter we went through some considerable effort to obtain an explicit form of $\mathcal{L}^\dagger \rho$ that we will make use of here. This form is given by equation (6.5.1) which we restate here for convenience

$$-\mathcal{L}^\dagger \rho = \sum_{j \in \mathcal{J}} \partial_j^\dagger \left([\rho]_{\omega_j} \partial_j (\ln \rho - \ln \sigma) \right).$$

In Chapter 6 we defined the notions of gradient and divergence. Equation (6.4.6) defines a one-parameter way of multiplying a single element $A \in \mathcal{H}_{\mathcal{A}}$ by a density matrix $\rho \in \mathcal{G}_+$. We wish to multiply vector fields $\mathbf{A} = (A_1, ..., A_{\mathcal{J}})$ by ρ so we simply extend the definition of multiplication with a single element to element wise multiplication as follows:

Definition 7.5.1. Let $\vec{\omega} \in \mathbb{R}^{|\mathcal{J}|}$ and for $\rho \in \mathcal{G}_+$ define $[\rho]_{\vec{\omega}}$ on $\mathcal{H}_{\mathcal{A}}$ by

$$[\rho]_{\vec{\omega}}(A_1, ..., A_{|\mathcal{J}|}) = ([\rho]_{\omega_1} A_1, ..., [\rho]_{\omega_{|\mathcal{J}|}} A_{|\mathcal{J}|}). \tag{7.5.5}$$

Note that $[\rho]_\omega$ is invertible so it follows that $[\rho]_{\vec{\omega}}$ is invertible with $[\rho]_{\vec{\omega}}^{-1}$ simply given by

$$[\rho]_{\vec{\omega}}^{-1}(A_1, ..., A_{|\mathcal{J}|}) = ([\rho]_{\omega_1}^{-1} A_1, ..., [\rho]_{\omega_{|\mathcal{J}|}}^{-1} A_{|\mathcal{J}|}). \tag{7.5.6}$$

Now we are in a position to rewrite equation (7.5.3) as a quantum analogue of the classical continuity equation. Given $\rho \in \mathcal{H}_{\mathcal{A}}$, $\mathbf{A} \in \mathcal{H}_{\mathcal{A},\mathcal{J}}$ and $\vec{\omega} \in \mathbb{R}^n$ we can construct a vector field $\mathbf{V}$ by defining

$$\mathbf{V} = -[\rho]_{\vec{\omega}}^{-1} \mathbf{A}.$$

Then equation (7.5.3) becomes

$$\dot{\rho}(0) + \operatorname{div}([\rho]_{\vec{\omega}} \mathbf{V}) = 0. \tag{7.5.7}$$

Equation (7.5.7) will be referred to as the non-commutative continuity equation. Of course one must naturally ask if the vector field $\mathbf{A}$ that is relevant in equation (7.5.7) is unique. Recall that $\mathbf{A}$ is any vector field that satisfies $\dot{\rho}(0) = \operatorname{div} \mathbf{A}$, and Theorem 6.3.2 shows that there exists a whole affine subspace of such fields. Luckily we are in a finite dimensional setting so there exists a smallest element in such an affine subspace, which is just a closed convex set [CM17]. Next we define an inner product that will give rise to a norm relevant to our purposes of obtaining a suitable Riemannian metric.

Definition 7.5.2. For a given $\rho \in \mathcal{G}_+$ and a generator $\mathcal{L}$ of a QMS, we define the inner product $\langle \cdot, \cdot \rangle_{\mathcal{L},\rho}$ between $\mathbf{V}, \mathbf{W} \in \mathcal{H}_{\mathcal{A},\mathcal{J}}$ by:

$$\langle \mathbf{W}, \mathbf{V} \rangle_{\mathcal{L},\rho} = \sum_{j \in \mathcal{J}} \left\langle W_j, [\rho]_{\omega_j} V_j \right\rangle_{\mathcal{H}_{\mathcal{A}}}. \tag{7.5.8}$$

The corresponding norm is then given by

$$||\mathbf{V}||_{\mathcal{L},\rho} = \langle \mathbf{V}, \mathbf{V} \rangle_{\mathcal{L},\rho} .$$

The next theorem is a statement about the uniqueness of the vector field $\mathbf{V}$ in equation (7.5.7) and the existence of a minimal norm.

Theorem 7.5.3. *Let $\rho(t)$ be a differentiable path on $\mathcal{G}_+$ that is defined on some interval $(-\epsilon, \epsilon)$, $\epsilon > 0$. Let $\rho(0) = \rho_0$. There exists a unique, traceless, selfadjoint element $\mathbf{V} \in \mathcal{H}_{\mathcal{A},\mathcal{J}}$ of the form $\mathbf{V} = \nabla U$ for some $U \in \mathcal{H}_{\mathcal{A}}$, for which the non-commutative continuity equation holds.*

Proof. In light of the discussion on finite dimensions we may take $\mathbf{V}$ to be the smallest element in our affine subspace such that for any other $\mathbf{W}$ such that $\rho(0) + \operatorname{div} \mathbf{W} = 0$, we have that

$$||\mathbf{V}||_{\mathcal{L},\rho_0} < ||\mathbf{W}||_{\mathcal{L},\rho_0}.$$

Now let $\mathbf{A}$ be an arbitrary divergence-free field and define $\mathbf{W} = [\rho_0]^{-1}\mathbf{A}$. Now consider the vector field $\mathbf{V}_\epsilon = \mathbf{V} + \epsilon\mathbf{W}$.

$$\begin{aligned}
\dot{\rho}(0) + \operatorname{div}([\rho_0]_{\vec{\omega}}^{-1}\mathbf{V}_\epsilon) &= \dot{\rho}(0) + \operatorname{div}([\rho_0]_{\vec{\omega}}^{-1}\mathbf{V}) + \operatorname{div}(\epsilon[\rho_0]_{\vec{\omega}}^{-1}\mathbf{A}) \\
&= \dot{\rho}(0) + \operatorname{div}([\rho_0]_{\vec{\omega}}^{-1}\mathbf{V}) + \epsilon[\rho_0]_{\vec{\omega}}^{-1} \operatorname{div}(\mathbf{A}) \\
&= \dot{\rho}(0) + \operatorname{div}([\rho_0]_{\vec{\omega}}^{-1}\mathbf{V}) + 0 \\
&= 0.
\end{aligned}$$

Thus $||\mathbf{V}||_{\mathcal{L},\rho_0} < ||\mathbf{V}_\epsilon||_{\mathcal{L},\rho_0}$ for all $\epsilon > 0$. Now

$$\begin{aligned}
||\mathbf{V}_\epsilon||_{\mathcal{L},\rho_0} &= ||\mathbf{V} + \epsilon\mathbf{W}||_{\mathcal{L},\rho_0} \\
&= \sqrt{\langle \mathbf{V} + \epsilon\mathbf{W}, \mathbf{V} + \epsilon\mathbf{W} \rangle_{\mathcal{L},\rho_0}} \\
&= \sqrt{\langle \mathbf{V}, \mathbf{V} \rangle_{\mathcal{L},\rho_0} + 2\epsilon \langle \mathbf{V}, \mathbf{W} \rangle_{\mathcal{L},\rho_0} + \langle \mathbf{W}, \mathbf{W} \rangle_{\mathcal{L},\rho_0}}.
\end{aligned}$$

Since $||\mathbf{V}||_{\mathcal{L},\rho_0} < ||\mathbf{V}_\epsilon||_{\mathcal{L},\rho_0}$ must hold for all $\epsilon > 0$, it follows from the above that $\langle \mathbf{V}, \mathbf{W} \rangle_{\mathcal{L},\rho_0} = 0$. Note that

$$\begin{aligned}
\langle \mathbf{V}, \mathbf{W} \rangle_{\mathcal{L},\rho_0} &= \left\langle \mathbf{V}, [\boldsymbol{\rho_0}]_{\vec{\omega}}^{-\mathbf{1}}\mathbf{A} \right\rangle_{\mathcal{L},\rho_0} \\
&= \sum_{j\in\mathcal{J}} \left\langle V_j, [\rho_0]_{\omega_j} [\rho_0]_{\omega_j}^{-1} A_j \right\rangle_{\mathcal{H}_\mathcal{A}} \\
&= \sum_{j\in\mathcal{J}} \langle V_j, A_j \rangle_{\mathcal{H}_\mathcal{A}} \\
&= \langle \mathbf{V}, \mathbf{A} \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} .
\end{aligned}$$

Hence $\langle \mathbf{V}, \mathbf{A} \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} = 0$.

We showed previously that $\mathbf{W}$ was a divergence-free vector field, i.e. $\mathbf{W} \in \text{Null}(\text{div})$. We have also showed that $\mathbf{V}$ is orthogonal to $\mathbf{W}$, i.e. $\mathbf{V} \in (\text{Null}(\text{div}))^{\perp}$. It follows from Proposition 6.1.4 then that $\mathbf{V} \in \text{Range}(\nabla)$. Thus we can write $\mathbf{V} = \nabla U$ for some $U \in \mathcal{H}_{\mathcal{A}}$. By subtracting a suitable multiple of the identity we may take U to be traceless. Since $\mathcal{P}_t$ is ergodic it now follows from Theorem 6.3.1 that U is uniquely determined. Finally all that is left to show is that this U is self adjoint. To that end we define an operator

$$\mathcal{L}_\rho A = \text{div}([\rho]_{\vec{\omega}} \nabla A).$$

Expanding the above we obtain

$$\begin{aligned}
\mathcal{L}_\rho A &= \text{div}([\rho]_{\vec{\omega}} \nabla A) \\
&= \text{div}([\rho]_{\omega_1} \partial_1 A, ..., [\rho]_{\omega_{|\mathcal{J}|}}) \partial_{|\mathcal{J}|} A) \\
&= -\sum_{j \in \mathcal{J}} \partial_j^\dagger [\rho]_{\omega_j} \partial_j A \\
&= -\sum_{j \in \mathcal{J}} [\rho]_{\omega_j} \partial_j^\dagger \partial_j A \\
&= \sum_{j \in \mathcal{J}} ([\rho]_{\omega_j} (V_j A - A V_j)) V_j^* - \sum_{j \in \mathcal{J}} V_j^* ([\rho]_{\omega_j} (V_j A - A V_j)).
\end{aligned}$$

Next we apply the result of Corollary 6.4.5 to obtain

$$(\mathcal{L}A)^* = \sum_{j \in \mathcal{J}} ([\rho]_{-\omega_j} (V_j^* A^* - A^* V_j^*)) V_j - \sum_{j \in \mathcal{J}} V_j ([\rho]_{-\omega_j} (V_j^* A^* - A^* V_j^*)).$$

Recall that $\{V_j\} = \{V_j^*\}$, since we sum over all $j \in \mathcal{J}$ in the above equations it follows that $(\mathcal{L}A)^* = \mathcal{L}A^*$.

Now we can write

$$\dot{\rho}(0) = -\text{div}([\rho_0]_{\vec{\omega}} \nabla U),$$

and apply the definition of $\mathcal{L}_\rho$ to obtain

$$\dot{\rho}(0) = -\mathcal{L}_{\rho_0} U.$$

Since ρ_0 is self-adjoint it follows that

$$\mathcal{L}_{\rho_0} U = (\mathcal{L}_{\rho_0} U)^* = \mathcal{L}_{\rho_0} U^*.$$

Since U is unique it follows from the above that U is self-adjoint. □

Finally we are in a position to use the tools from differential geometry to construct our Riemannian metric. For the remainder of this section we shall consider a curve $t \mapsto \rho + tA_k$ in our manifold $\mathcal{G}_+$, $A \in \mathcal{H}_{\mathcal{A}}$ is self-adjoint and traceless. Of course we need to show that the element $\rho + tA \in \mathcal{G}_+$. To that end we provide here a proof that the set of density matrices is an open set.

Proposition 7.5.4. *The set of strictly positive square matrices is an open set.*

Proof. Let ρ be a strictly positive $n \times n$ matrix, and B be any other $n \times n$ matrix. Suppose we have $||\rho - B|| < \epsilon$ for some $\epsilon > 0$. Since $\rho > 0$ it follows that for any $x \in \mathcal{H}_{\mathcal{A}}$ we have

$$\langle \rho x, x \rangle > 0,$$

where $\mathcal{H}_{\mathcal{A}}$ is the associated Hilbert space. Now

$$\begin{aligned} \langle \rho x, x \rangle - \langle Bx, x \rangle &< |\, \langle \rho x, x \rangle - \langle Bx, x \rangle \,| \\ &= |\, \langle (\rho - B)x, x \rangle \,| \\ &\leq ||(\rho - B)x|| \cdot ||x|| \\ &\leq ||\rho - B|| \cdot ||x||^2 \\ &< \epsilon. \end{aligned}$$

In the third line we used the Cauchy-Schwartz inequality and in the fourth line we make use of the fact that $||\rho - B|| = \sup_{||x|| \leq 1} ||(\rho - B)x||$.

Choose $\epsilon = \frac{1}{2} \inf_{x \in \mathcal{H}_{\mathcal{A}}} \langle \rho x, x \rangle$ then

$$\begin{aligned} \langle Bx, x \rangle &> \langle \rho x, x \rangle - \epsilon \\ &> 0. \end{aligned}$$

In other words B is a strictly positive matrix so the set of strictly positive matrices is open. □

Next we show that for sufficiently small t we have that $\rho + tA$ is a density matrix, with A a self-adjoint, traceless element of $\mathcal{H}_{\mathcal{A}}$.

Proposition 7.5.5. *Let $\rho \in \mathcal{G}_+$ and A be a traceless self-adjoint element in $\mathcal{H}_{\mathcal{A}}$. Then there exists $t \in \mathbb{R}$ such that $\rho + tA \in \mathcal{G}_+$.*

Proof. First note that for any $t \in \mathrm{R}$

$$\begin{aligned} \mathrm{Tr}(\rho + tA) &= \mathrm{Tr}(\rho) + t\,\mathrm{Tr}(A) \\ &= \mathrm{Tr}(\rho) + 0 \\ &= 1. \end{aligned}$$

Now all that is left to show is that for sufficiently small t, $\rho + tA$ is strictly positive. Since we are working in a finite dimensional setting all norms are equivalent. We consider the Hilbert-Schmidt norm:

$$\begin{aligned} ||\rho - (\rho + tA)||_{H.S} &= ||tA||_{HS} \\ &= \mathrm{Tr}((tA^*)(tA)) \\ &= t^2\,\mathrm{Tr}(A^2). \end{aligned}$$

Now choose ϵ as in Proposition 7.5.4. If $\mathrm{Tr}(A^2) = 0$ then we are done and clearly $\rho + tA > 0$. If $\mathrm{Tr}(A^2) \neq 0$, choose $t = \sqrt{\frac{\epsilon}{2\,\mathrm{Tr}(A^2)}}$, then

$$\begin{aligned} ||\rho - (\rho + tA)||_{H.S} &= t^2\,\mathrm{Tr}(A^2) \\ &= \frac{\epsilon}{2\,\mathrm{Tr}(A^2)}\,\mathrm{Tr}(A^2) \\ &= \frac{\epsilon}{2} \\ &< \epsilon. \end{aligned}$$

By Propostion 7.5.4 and the equivalence of norms in finite dimensions it follows that $t + \rho A > 0$. □

Now we move on to the first step of constructing a Riemannian metric, which is of course to identify a suitable tangent space.

Definition 7.5.6. For each $\rho \in \mathcal{G}_+$, we identify the tangent space T_ρ at $\rho(0) = \rho_0$ with the set of vector field gradients

$$\{\nabla U : U \in \mathcal{H}_{\mathcal{A}}, U = U^*\}.$$

Equation (7.5.7) together with the result of Theorem 7.5.3 provide a one-to-one correspondence between ρ and U, i.e. a one-to-one correspondence between points on our manifold and elements of the tangent space. We define a Riemannian metric by

$$||\dot{\rho}(0)||^2_{g_{\mathcal{L},\rho}} = ||\mathbf{V}||^2_{\mathcal{L},\rho}, \tag{7.5.9}$$

where $\mathbf{V}$ and $\dot{\rho}(0)$ are related by equation (7.5.7).

Let n be the dimension of $\mathcal{A}$, and let $A_1, ..., A_{n-1}$ be an orthonormal set of traceless self-adjoint elements of $\mathcal{H}_{\mathcal{A}}$. Then we can define a chart, or coordinate mapping, $u : \mathcal{G}_+ \to \mathbb{R}^{n-1}$ by

$$u(\rho) = (\mathrm{Tr}[A_1\rho], ..., \mathrm{Tr}[A_{n-1}\rho]). \tag{7.5.10}$$

Given that $A_1, ..., A_{n-1}$ is an orthonormal set we now have a one-to-one mapping of $\mathcal{G}_+$ onto a convex subset of $\mathbb{R}^{n-1}$. We can verify this with a simple calculation. Suppose $u(\rho_1) = u(\rho_2)$ then

$$\mathrm{Tr}[A_j\rho_1] = \mathrm{Tr}[A_j\rho_2],$$

for all $j = 1, ..., n-1$. This in turn implies that

$$\mathrm{Tr}[A_j(\rho_1 - \rho_2)] = 0,$$

Since this result must hold for all $j = 1, ..., n-1$ it follows that $\rho_1 = \rho_2$.

Now we move on to the structure of our metric tensor. We define the kth coordinate function by

$$u^k(\rho) = \mathrm{Tr}[A_k\rho]. \tag{7.5.11}$$

The kth coordinate vector field will be tangent to the curve $\rho + tA_k$ for t in a sufficiently small interval such that $\rho + tA_k \in \mathcal{G}_+$. We showed such an interval exists in Proposition 7.5.5. We define the kth potential function $X_k(\rho)$ then as the unique traceless element X that satisfies

$$\mathrm{div}([\rho]_{\vec{\omega}}\nabla X) = A_k. \tag{7.5.12}$$

For the particular curve $t \mapsto \rho + tA_k$ we are moving only in the direction of A_k and thus we have

$$\mathbf{A} = (0, ..., A_k, ..., 0).$$

Thus $\dot{\rho}(0) = A_k$ which in turn implies that

$$\dot{\rho}(0) = \mathrm{div}([\rho]_{\vec{\omega}}]\nabla X).$$

From this we see that the kth coordinate tangent vector field is given by

$$\frac{\partial}{\partial u^k} = \nabla X_k(\rho).$$

Applying the to the usual definition of a Riemannian metric we then obtain the specific form of the k,l components of our metric tensor from equation (7.5.8) as

$$\begin{aligned}[g_{\mathcal{L}}(\rho)]_{k,l} &= \langle \nabla X_k, [\rho]_{\vec{\omega}} \nabla X_l \rangle_{\mathcal{H}_{\mathcal{A},\mathcal{J}}} \\ &= \sum_{j \in \mathcal{J}} \left\langle X_k, [\rho]_{\omega_j} X_l \right\rangle_{\mathcal{H}_{\mathcal{A}}}\end{aligned}$$

We argued previously that $\rho \mapsto [\rho]_{\omega_j}$ is C^∞ and it follows that $\rho \mapsto (\mathrm{div}([\rho]_{\vec{\omega}} \nabla))^{-1}$ is also C^∞. Consequently for our metric tensor the map $\rho \mapsto [g_{\mathcal{L}}(\rho)]_{k,l}$ is C^∞. Thus we have a Riemannian metric.

7.6 Steepest descent and gradient flow

In this last section we will relate the differential structure of our manifold to the relative entropy functional and show that the generator $\mathcal{L}$ of our QMS is in fact gradient flow of the relative entropy.

Consider a function $F : \mathcal{G} \to \mathbb{R}$ which is differentiable with respect to $\rho \in \mathcal{G}$.

We denote by F by $\frac{\delta F}{\delta \rho}(\rho)$, the differential of F which is the unique traceless self-adjoint element in $\mathcal{A}$ that satisfies

$$\lim_{t \to 0} \frac{1}{t}(F(\rho + tA) - F(\rho) = \mathrm{Tr}\left[\frac{\delta F}{\delta \rho}(\rho) A\right], \tag{7.6.1}$$

for all traceless, selfadjoint $A \in \mathcal{A}$.

We associate to any such function a corresponding gradient field, which we denote by $\mathrm{grad}_{g_{\mathcal{L}}} F(\rho)$. Such a gradient field is identified with the tangent space defined in Definition 7.5.6.

Definition 7.6.1. Let $\rho(t)$ be a differentiable path defined on an interval $(-\epsilon, \epsilon)$, with $\epsilon > 0$. Furthermore let $\rho(0) = \rho$ and let $\rho(t)$ be such that the continuity equation is satisfied,

$$\dot{\rho}(0) + \mathrm{div}([\rho]_{\vec{\omega}} \nabla U) = 0,$$

for some self-adjoint element U. Then we define $\mathrm{grad}_{g_{\mathcal{L}}} F(\rho)$ as the unique element satisfying

$$\frac{d}{dt} F(\rho(t))|_{t=0} = \left\langle \mathrm{grad}_{g_{\mathcal{L}}} F(\rho), \nabla U \right\rangle_{\mathcal{L},\rho}. \tag{7.6.2}$$

Recall the trace is simply the inner product

$$\mathrm{Tr}\left[\frac{\delta F}{\delta \rho}\right] = \left\langle \frac{\delta F}{\delta \rho}, A \right\rangle_{\mathcal{H},\mathcal{A}}.$$

It then easily follows from equations (7.6.1) and (7.6.2) that

$$\left\langle \frac{\delta F}{\delta \rho}, \mathrm{div}([\rho]_{\vec{\omega}} \nabla U) \right\rangle_{\mathcal{H},\mathcal{A}} = -\left\langle \mathrm{grad}_{g_{\mathcal{L}}} F(\rho), [\rho]_{\vec{\omega}} \nabla U \right\rangle_{\mathcal{L},\rho}.$$

Recall that the div operator is negative the adjoint of the ∇ operator. So if we move the div operator in the second argument of the left hand side of the above equation to the first argument we obtain

$$\left\langle \frac{\delta F}{\delta \rho}, [\rho]_{\vec{\omega}} \nabla U \right\rangle_{\mathcal{H},\mathcal{A}} = -\left\langle \mathrm{grad}_{g_{\mathcal{L}}} F(\rho), [\rho]_{\vec{\omega}} \nabla U \right\rangle_{\nabla\mathcal{L},\rho}.$$

Since this must be true for all paths $\rho(t)$ it follows from Theorem 7.5.3 that it holds for arbitrary U. Thus we have the following result:

Theorem 7.6.2. *Let $F : \to \mathbb{R}$ be a differentiable function. Then the gradient with respect to the Riemannian metric $g_{\mathcal{L}}$ is given by*

$$\mathrm{grad}_{g_{\mathcal{L}}} f(\rho) = \nabla \frac{\delta F}{\delta \rho}(\rho).$$

The corresponding equation for gradient flow of steepest decent is then given by

$$\dot{\rho}(t) = \mathrm{div}\left([\rho(t)]_{\vec{\omega}} \nabla \frac{\delta F}{\delta \rho}(\rho(t)) \right).$$

Now consider the relative entropy functional $F(\rho) = D(\rho||\sigma)$ where

$$D(\rho||\sigma) = \mathrm{Tr}[\rho(\ln\rho - \ln\sigma)].$$

For this functional we can calculate the differential as follows:

$$\begin{aligned} F(\rho + tA) - F(\rho) &= \mathrm{Tr}[(\rho + tA)(\ln(\rho + tA)) - \ln\sigma)] - \mathrm{Tr}[\rho(\ln\rho - \ln\sigma) \\ &= \mathrm{Tr}[\rho(\ln(\rho + tA) - \ln\sigma) + tA(\ln(\rho + tA) - \ln\sigma) - \rho(\ln(\rho) - \ln\sigma)] \\ &= \mathrm{Tr}[\rho\ln(\rho + tA) + tA\ln(\rho + tA) - tA\ln\sigma - \rho\ln\rho]. \end{aligned}$$

Multiplying the last line of the above calculation by $\frac{1}{t}$ we obtain

$$\mathrm{Tr}\left[\frac{\rho}{t}\ln(\rho + tA) - \frac{\rho}{t}\ln\rho\right] + \mathrm{Tr}\left[A\ln(\rho + tA) - A\ln\sigma\right].$$

If we now take the limit as $t \to \infty$ it is clear that the first two terms go to zero and we can conclude that

$$\frac{\delta F}{\delta \rho} = \ln(\rho) - \ln \sigma. \tag{7.6.3}$$

Finally we recall the result of Theorem 6.5.2:

$$-\mathcal{L}^\dagger \rho = \sum_{j \in \mathcal{J}} \partial_j^\dagger \left([\rho]_{\omega_j} \partial_j (\ln \rho - \ln \sigma) \right).$$

Together with equation (7.6.3) we then have

$$\begin{aligned} \mathcal{L}^\dagger(t)\rho &= -\sum_{j \in \mathcal{J}} \partial_j^\dagger \left([\rho(t)]_{\omega_j} \partial_j \frac{\delta F}{\delta \rho} \right) \\ &= \operatorname{div}([\rho(t)]_{\vec{\omega}} \nabla \frac{\delta F}{\delta \rho}) \\ &= \dot{\rho}(t). \end{aligned}$$

Thus we have proven the following theorem:

Theorem 7.6.3. *Let $\mathcal{P}_t = e^{t\mathcal{L}}$ be a QMS on $\mathcal{A}$ that satisfies the σ-DBC for some $\sigma \in \mathcal{G}_+$. Then the curve $\rho(t)$ is gradient flow for the relative entropy functional $D(\cdot||\sigma)$ in the metric $g_{\rho,\mathcal{L}}$ that is canonically associated to the generator $\mathcal{L}$ via its representation given by equation (5.0.2).*

Conclusion and further research

We have shown how the time evolution of a certain class of quantum states can be written as gradient flow of the relative entropy functional. This formulation of the time evolution is useful for both a conceptual understanding of the entropy of quantum systems, as well as more practical applications such as deriving new results. In particular we could apply this formulation to the study of quantum information theory to obtain new bounds on the entropy such as was done in [CM20] and [CM17].

In this dissertation we have considered only those quantum states that satisfy a very particular detailed balance condition. This was done so that we can use a suitable form of the generator for the quantum Markov semigroup that would allow us to construct the necessary formulations of gradient and divergence required in Chapter 7. There are, however, many different ways to define quantum detailed balance. The formulation of detailed balance studied in [Duv18] and [DS14] could provide some useful connections to entanglement and the geometric interpretation of the relative entropy functional.

Finally we have only considered the finite dimensional case in this dissertation. In principle one could apply these ideas to the infinite dimensional case, however this does seem to be a non-trivial problem. Thus there are many potential areas of future research where similar ideas to those in this dissertation could be applied.

Bibliography

[Ali76] Robert Alicki. On the detailed balance condition for non-hamiltonian systems. *Reports on Mathematical Physics*, 10(2):249–258, 1976.

[ASZ09] Luigi Ambrosio, Giuseppe Savaré, and Lorenzo Zambotti. Existence and stability for fokker–planck equations with log-concave reference measure. *Probability theory and related fields*, 145(3-4):517–564, 2009.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[BR12] Ola Bratteli and Derek William Robinson. *Operator Algebras and Quantum Statistical Mechanics: Volume 1: C*-and W*-Algebras. Symmetry Groups. Decomposition of States.* Springer Science & Business Media, 2012.

[Bra99] Howard E Brandt. Positive operator valued measure in quantum information processing. *American Journal of Physics*, 67(5):434–439, 1999.

[Bur69] Donald Bures. An extension of kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebras. *Transactions of the American Mathematical Society*, 135:199–212, 1969.

[BV01] Philippe Biane and Dan Voiculescu. A free probability analogue of the wasserstein metric on the trace-state space. *Geometric & Functional Analysis GAFA*, 11(6):1125–1138, 2001.

[Car19] Sean M Carroll. *Spacetime and geometry.* Cambridge University Press, 2019.

[CGGT17] Yongxin Chen, Wilfrid Gangbo, Tryphon T Georgiou, and Allen Tannenbaum. On the matrix monge–kantorovich problem. *European Journal of Applied Mathematics*, pages 1–27, 2017.

[CGT18] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Wasserstein geometry of quantum states and optimal transport of matrix-valued measures. In *Emerging Applications of Control and Systems Theory*, pages 139–150. Springer, 2018.

[Cip97] Fabio Cipriani. Dirichlet forms and markovian semigroups on standard forms of von neumann algebras. *journal of functional analysis*, 147(2):259–300, 1997.

[Cip08] Fabio Cipriani. Dirichlet forms on noncommutative spaces. In *Quantum potential theory*, pages 161–276. Springer, 2008.

[CL02] Eric A Carlen and Elliott H Lieb. Optimal hypercontractivity for fermi fields and related non-commutative integration inequalities. In *Inequalities*, pages 151–170. Springer, 2002.

[CLPS14] Sergio Caracciolo, Carlo Lucibello, Giorgio Parisi, and Gabriele Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Physical Review E*, 90(1):012118, 2014.

[CM14] Eric A Carlen and Jan Maas. An analog of the 2-wasserstein metric in non-commutative probability under which the fermionic fokker–planck equation is gradient flow for the entropy. *Communications in mathematical physics*, 331(3):887–926, 2014.

[CM17] Eric A Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017.

[CM20] Eric A Carlen and Jan Maas. Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. *Journal of Statistical Physics*, 178(2):319–378, 2020.

[Coh13] Donald L Cohn. *Measure theory*. Springer, 2013.

[CTDL91] C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Quantum Mechanics*. Quantum Mechanics. Wiley, 1991.

[Dav74] E Brian Davies. Markovian master equations. *Communications in mathematical Physics*, 39(2):91–110, 1974.

[DPT19] Giacomo De Palma and Dario Trevisan. Quantum optimal transport with quantum channels. *arXiv preprint arXiv:1911.00803*, 2019.

[DS14] Rocco Duvenhage and Machiel Snyman. Detailed balance and entanglement. *arXiv preprint arXiv:1407.0520*, 2014.

[Duv18] Rocco Duvenhage. Fermionic quantum detailed balance and entanglement. *Journal of Physics A: Mathematical and Theoretical*, 51(36):365305, 2018.

[Duv20] Rocco Duvenhage. Quadratic wasserstein metrics for von neumann algebras via transport plans. *arXiv preprint arXiv:2012.03564*, 2020.

[EG99] Lawrence C Evans and Wilfrid Gangbo. *Differential equations methods for the Monge-Kantorovich mass transfer problem*. Number 653. American Mathematical Soc., 1999.

[Fag99] Franco Fagnola. Quantum markov semigroups and quantum flows. *Proyecciones*, 18(3):1–144, 1999.

[FVDG99] Christopher A Fuchs and Jeroen Van De Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.

[Gro75] Leonard Gross. Hypercontractivity and logarithmic sobolev inequalities for the clifford-dirichlet form. *Duke Mathematical Journal*, 42(3):383–396, 1975.

[Hit12] Nigel Hitchin. Differentiable manifolds. *Lecture Notes, Oxford University*, 2012.

[Ike20] Kazuki Ikeda. Foundation of quantum optimal transport and applications. *Quantum Information Processing*, 19(1):25, 2020.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[KPT+17] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4):43–59, 2017.

[Kre78] Erwin Kreyszig. *Introductory functional analysis with applications*, volume 1. wiley New York, 1978.

[LS18] Bruno Lévy and Erica L Schwindt. Notions of optimal transport theory and how to implement them on a computer. *Computers & Graphics*, 72:135–148, 2018.

[LV09] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[McC97] Robert J McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.

[Mon12] Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12):122206, 2012.

[Mur14] Gerald J Murphy. *C*-algebras and operator theory.* Academic press, 2014.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.

[Par12] Kalyanapuram R Parthasarathy. *An introduction to quantum stochastic calculus*, volume 85. Birkhäuser, 2012.

[PC+19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[RM17] Pierre H Richemond and Brendan Maginnis. On wasserstein reinforcement learning and the fokker-planck equation. *arXiv preprint arXiv:1712.07185*, 2017.

[Rud06] Walter Rudin. *Real and complex analysis.* Tata McGraw-hill education, 2006.

[Seg53] Irving E Segal. A non-commutative extension of abstract integration. *Annals of mathematics*, pages 401–457, 1953.

[Sew02] Geoffrey Sewell. *Quantum mechanics and its emergent macrophysics.* Princeton University Press, 2002.

[SLE09] Paul Sjöberg, Per Lötstedt, and Johan Elf. Fokker–planck approximation of the master equation in molecular biology. *Computing and Visualization in Science*, 12(1):37–50, 2009.

[SW10] Benjamin Schumacher and Michael D. Westmoreland. *Quantum Processes Systems, and Information*. Cambridge University Press, New York, 2010.

[Tho] Matthew Thorpe. Introduction to optimal transport.

[TKR+10] Kristan Temme, Michael James Kastoryano, Mary Beth Ruskai, Michael Marc Wolf, and Frank Verstraete. The χ 2-divergence and mixing times of quantum markov processes. *Journal of Mathematical Physics*, 51(12):122201, 2010.

[Vil03] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.