Classical and Quantum Information Geometry

Matheus da Rocha Grasselli

A dissertation submitted to the University of
London in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Mathematics

Department of Mathematics
King's College London

October 2001

To Leonardo, ultimate reason of my being.

Todos os dias quando acordo

Não tenho mais o tempo que passou

Mas tenho muito tempo

Temos todo o tempo do mundo...

Renato Russo in *Tempo Perdido*

Acknowledgements

My warmest and unreserved thanks go to Ray Streater. From my very first interview with him, his enthusiasm for talking about mathematical physics (literally for many hours), constant encouragement and unbounded patience were vital for my PhD. What else to ask from a supervisor ? His deep knowledge of the field and truly idiosyncratic views about research in it, plus a seemingly infinite collection of anecdotes were added bonuses. Not to mention his habit of replying to e-mails almost instantly, which was particularly instrumental in the final stages of revision for this thesis, when we were thousands of miles apart.

Still at King's College, I thank the other members of staff, especially those at the Analysis group. Through seminars, colloquia and coffee conversations they have shown me what doing world class mathematics is about. Also my fellow PhD students, who through lunches at LSE and G01, beers at The Lyceum, countless 'conventions' all over the town and coffee conversations have shown me what doing a PhD in a place like London is all about.

I would like to acknowledge the financial support of a grant from Capes, Brazil, and of an Overseas Research Scholarship from the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom.

As for my academic trajectory before hitting London, I should thank two special people: Elza Gomide from São Paulo, for shaping my mathematical taste (chilled coffee notwithstanding) and Silvio dos Santos from Porto Alegre, for putting me in the right track.

My mom Bety and my dad Paulo deserve not only thanks for unconditional support during all phases of my life, but also a sort of apology for my pursuing of a career so far away from home for so many years.

Lastly, but also from my deepest levels, I thank Ana for her love and for giving me the most wonderful son a man could dream of, to whom this thesis is dedicated.

Abstract

We begin with the construction of an infinite dimensional Banach manifold of probability measures using the completion of the set of bounded random variables in the appropriate Orlicz norm as coordinate spaces. The infinite dimensional version of the Fisher metric as well as the exponential and mixture connections are introduced. It is then proved that they form a dualistic structure in the sense of Amari. The interpolating α-connections are defined, at the level of covariant derivatives, via embeddings into L^r-spaces and then found to be convex mixtures of the ± 1-connections. Several well known parametric results are obtained as finite dimensional restrictions of the nonparametric case.

Next, for finite dimensional quantum systems, we study a manifold of density matrices and explore the concepts of monotone metrics and duality in order to establish that the only monotone metrics with respect to which the exponential and mixture connections are mutually dual are the scalar multiples of the Bogoliubov-Kubo-Mori inner product of quantum statistical mechanics.

For infinite dimensional quantum systems, we present a general construction of a Banach manifold of density operators using the technique of ε-bounded perturbations, which contains small perturbations of forms and operators in the sense of Kato as special cases. We then describe how to obtain an affine structure in such a manifold, together with the corresponding exponential connection. The free energy functional is proved to be analytic on small neighbourhoods in the manifold.

We conclude with an application of the methods of Information Geometry and Statistical Dynamics to a concrete problem in fluid dynamics: the derivation of the time evolution equations for the density, energy and momentum fields of a fluid under an external field.

Contents

Chapter 1

Introduction

It was just over half a century ago that the Fisher information

$$g_{ij} = \int \frac{\partial \log p(x,\theta)}{\partial \theta^i} \frac{\partial \log p(x,\theta)}{\partial \theta^j} p(x,\theta) dx \tag{1.1}$$

was independently suggested by Rao and Jeffreys as a Riemannian metric for a parametric statistical model $\{p(x,\theta), \theta = (\theta^1, \dots, \theta^n)\}$ [51, 28]. The Riemannian geometry of statistical models was then studied as a mathematical curiosity for some years, with an emphasis in the geodesic distances associated with the Levi-Civita connection for this metric. A greater amount of attention was devoted to the subject after Efron introduced the concept of statistical curvature, pointing out its importance to statistical inference, as well as implicitly using a new affine connection, which would be known as the exponential connection [11]. This exponential connection, together with another connection, later to be called the mixture connection, were further investigated by Dawid [10].

The work of several years on the geometric aspects of parametric statistical models culminated with Amari's masterful account [1], where the whole finite dimensional differential-geometric machinery is employed, including a one-parameter family of α-connections, the essential concept of duality and the notions of statistical divergence, projections and minimisation procedures. Among the successes of the research at these early stages one could single out the rigidity of the geometric structures, like in Chentsov's result concerning the uniqueness of the Fisher metric with respect to monotonicity [8] and Amari's result concerning the uniqueness of the α-connections as introduced by invariant statistical divergences. The ideas were then extensively used in statistics, in particular higher order asymptotic inference and curved exponential models [30]. Its links with information theory through entropy functions, which appear as special cases of divergences, and its natural setting as part of probability theory in general, rendered this theory the denomination of Information Geometry.

Two different lines of investigation in Information Geometry took off in the nineties: infinite dimensional manifolds of classical probabilities and finite dimensional manifolds of density matrices, representing quantum probabilities.

As for the first, one had, on the practical side, the need to deal with nonparametric models in statistics, where the shape of the underlying distribution is not assumed to be known. On a more fundamental level, there was the desire of having parametric statistical manifolds defined simply as finite dimensional submanifolds of a well defined manifold of all probability measures on a sample space. The motivating idea was already in Dawid's paper [10] and was also addressed by Amari [1]. The first sound mathematical construction, however, is due to Pistone and Sempi [50]. Given a measure space $(\Omega, \mathcal{F}, \mu)$, they showed how to construct a Banach manifold $\mathcal{M}_\mu$ of all (strictly positive) probability measures equivalent to μ. The Banach space used as generalised coordinates was the Orlicz space L^{Φ_1}, where Φ_1 is an exponential Young function. In order to prove that their construction leads to a manifold, they introduced the notion of exponential convergence, whose topology is equivalent to the one induced by L^{Φ_1}. In a subsequent work [49], further properties of this manifold were analysed, in particular the concepts of orthogonality and submanifolds. The next step in this development was the definition of the exponential connection as the natural connection induced by the use of L^{Φ_1} [15]. The authors then propose a mixture connection acting on the pretanget bundle ${}^*T\mathcal{M}_\mu$ and prove that it is dual to the exponential connection, in the sense of duality for Banach spaces. They further define the α-connections through generalised α-embeddings and show that the formal relation between the exponential, mixture and α-connections are the same as in the parametric case, that is

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla^{(e)} + \frac{1-\alpha}{2}\nabla^{(m)}. \tag{1.2}$$

We argue, however, that the neither of these two results (duality for the exponential and mixture connection and α-connections as convex mixture of them) is a proper generalisation of the corresponding parametric ones, the reason being twofold. First, Banach space duality is not Amari's duality. The latter refers to a metric being preseverd by the joint action of two parallel transports, which are then said to be dual (see (2.23)). Secondly, all the α-connections in the parametric case all act on the tangent bundle, whereas in [15] each of them acts on its own bundle-connection pair, making a formula like (1.2) at least difficult to interpret.

In order to solve these problems, in chapter 2 we start with the smaller space M^{Φ_1} as generalised coordinates. The effect of this is that the mixture connection can be meaningfully introduced on the tangent bundle $T\mathcal{M}_\mu$. It is then possible to prove that it is dual to the exponential connection in the sense of Amari. We also define the putative α-covariant derivatives through the α-embeddings, but in a way that makes them all act on $T\mathcal{M}_\mu$ as well. Equation (2.42) is then proved and shows that they are well defined as covariant derivatives on $T\mathcal{M}_\mu$, being the convex mixture of well defined connections. We also show that all the definitions and constructions reproduce the parametric results when one considers finite dimensional submanifolds of $\mathcal{M}_\mu$.

The second line of research, that is, manifolds of density matrices, has been inspired by applications in quantum statistical mechanics and quantum computing [25, 5, 67]. The first attempts to extend Amari's theory to density matrices acting on finite dimensional Hilbert spaces are found in the work of Nagaoka [40] and Hasegawa [22, 23]. In comparison with the classical case, operators on finite dimensional Hilbert spaces are analogue to probabilities on a finite sample space. It was already known by then that Chentsov's uniqueness does not hold in the noncommutative setup

[37], that is, that there are infinitely many inequivalent monotone metrics on matrix spaces. Arriving at the subject through the path of his early work in entropies for quantum systems [41, 46, 42], Petz set himself the task of characterising them all. The result is contained in a series of papers [43, 45, 44] and states a bijection between monotone metrics on matrix spaces and operator monotone functions. Different versions of Petz's characterisation continued to appear in the literature [35, 14].

In chapter 3, we ask for possible Riemannian metrics that make the quantum exponetial and mixture connection dual. We find that they are matrix multiples of the Bogoliubov-Kubo-Mori metric. When we combine this with Petz's theorem, that is, when we require further that the metric should be monotone, we get the improved result of scalar multiples of the *BKM* metric as the only possibilities.

The research in quantum Information Geometry for infinite dimensional Hilbert spaces was started by Streater, motivated by his work in Statistical Dynamics. An ancestor of this work, but limited to bounded operators in the context of Tomita-Takesaki theory is the paper [4] by Araki. Streater begins with the parametric case [57], that is, the Hilbert space $\mathcal{H}$ is infinite dimensional but we consider finite dimensional manifolds of density operators, whose tangent space consists of a finite dimensional subspace of (possible unbounded) operators on $\mathcal{H}$. This is the quantum analogue of the classical parametric case on general sample spaces. The starting point is a density operator ρ_0 that can be written as a canonical state for an 'unperturbed' Hamiltonian H_0. The method to obtain the parametric family is to take a finite dimensional set $\mathcal{X}$ spanned by linearly independent operators X_j and to consider the family of canonical states with 'perturbed Hamiltonian' $H_X = H_0 + X$. He then puts forward three axioms to be satisfied by H_0 and the operators X_j and uses them to prove that H_X is well defined and to obtain the exponential and mixture connections, as well as the *BKM* metric, for the finite dimensional manifold considered. The most relevant of the axioms says that the X_j must be operator-small perturbations of H_0, thus including the bounded operators (which are operator-tiny perturbations) of Araki [4] as special cases.

Streater then moved to the harder case of infinite dimensional manifolds of density operators on infinite dimensional Hilbert spaces [64], which is in analogy with the fully nonparametric classical case. The method is still to consider perturbations of a given density operator expressed as a canonical state with Hamiltonian H_0 satisfying certain conditions. This time, the set of all form-small perturbations of H_0 was considered as a generalised coordinate space, which was shown to be a Banach space with an appropriate norm. The manifold obtained is provided with the exponential connection and has enough regularity so that the partition function is a Lipschitz continuous funtional on it. It lacks regularity, however, to be endowed with the *BKM* scalar product. This problem is solved in [63], where the set of all operator-small perturbations of H_0 is taken as the coordinate Banach space. The manifold obtained is regular enough so that the free energy functional is analytic, and the *BKM* scalar product can then certainly be defined as its second derivative. Incidentally, one of the axioms of [57] is found to be unnecessary. In chapter 4 we reproduce all these results but for the interpolating class of ε-bounded perturbations, which contains all the previously used perturbations as special cases.

Our last chapter reviews the applications of Information Geometry to Statistical Dynamics, in particular as a tool for deriving equations of motion for conserved quantities of macroscopic physical systems. As a new application, we obtain hydrodynamical equations for the mass-density, energy and the three componets of momentum of a fluid moving under the influence of an external field [19].

Chapter 2

Classical Information Manifolds

It is customary in books introducing the subject of parametric statistical manifolds [1, 2, 38] to enumerate a series of technical conditions to be satisfied by the set of probability densities one wants to work with, such as linear independence of the scores and smoothness as functions of the paremeters. It is then said that, if the set of all probabilities on a measurable space were a *bona fide* manifold $\mathcal{M}$, then all these conditions would be reduce to saying that a parametric statistical manifold is a finite dimensional submanifold of $\mathcal{M}$. On the other hand, in papers dealing with the construction of nonparametric statistical manifolds, the subject of explicitly recovering the parametric case as a special submanifold is barely mentioned. In this chapter, we review the construction of such manifolds as proposed by Pistone and Sempi [50] and show how to obtain the parametric results as by-products. The exposition presented here, as well as some of the results, follows that of our paper [16].

The novelty of our approach is in fully exploring the consequences of the use of the Banach space M^{Φ_1} (following a suggestion in [48]), the completion of the bounded random variables in the norm of L^{Φ_1}, instead of the whole of L^{Φ_1} as used in [50, 49]. We also present a direct proof that the construction yields a C^∞ Banach manifold without using the concept of exponential convergence.

Having defined the information manifold, the next step in the programme for nonparametric Information Geometry is to define the infinite dimensional analogues of a metric and dual connections, ideas that play a leading role in the parametric version of the theory. A proposal for exponential and mixture, as well as for the intermediate α-connection, has been advocated by Gibilisco and Pistone [15]. However, we argue that their elegant definition does not properly generalise the original ideas from the parametric case. Their connections each act on a different vector bundle instead of all acting on the tangent bundle as in the finite dimensional case. The duality observed between them does not involve any metric, while in parametric Information Geometry dual connections with respect to one metric can fail to be dual with respect to an arbitrarily different metric.

We present in this chapter our proposal for the infinite dimensional exponential and mixture connections, together with the appropriate concept of duality, as well as the generalised metric that makes them dual to each other. We also show that these definitions reduce to the familiar ones for finite dimensional submanifolds and that exponential and mixture families are geodesic for the exponential and mixture connections, respectively.

We then move to the subject of α-connections, where we again rearrange the definitions of [15] in order to have them all acting on the same bundle and with the desired relation between them, the exponential and the mixture connections still holding. To carry on with the comparison with the finite dimensional case, we also recall the ideas concerning the α-representations of the tangent space at a point in $\mathcal{M}$, which will turn out to be instrumental in the next chapter as well, when we venture into the quantum arena.

The main analytical tool used for these purposes is the theory of Orlicz spaces, an introduction to which is given below.

2.1 Orlicz Spaces

We present here the aspects of the theory of Orlicz spaces that will be relevant for the construction of the information manifold. Similarly oriented short introductions to the subject can be found in [50, 49, 15]. For more comprehensive accounts the reader is refered to the monographs [52] and [32].

Recall first that a *Young* function is a convex function $\Phi : \mathbb{R} \mapsto \overline{\mathbb{R}}^+$ satisfying

i. $\Phi(x) = \Phi(-x), \quad x \in \mathbf{R}$,

ii. $\Phi(0) = 0$,

iii. $\lim\limits_{x \mapsto \infty} \Phi(x) = +\infty$.

Note that, in this generality, Φ can vanish on an interval around the origin (as opposed to vanishing if and only if $x = 0$) and it can also happen that $\Phi(x) = +\infty$, for $0 < x_1 \leq x$, for some positive x_1, although it must be continuous where it is finite (due to convexity). In the absence of these annoyances, most of the theorems have stronger conclusions. This will be the case for the following three Young functions used in Information Geometry:

$$\Phi_1(x) = \cosh x - 1, \tag{2.1}$$

$$\Phi_2(x) = e^{|x|} - |x| - 1, \tag{2.2}$$

$$\Phi_3(x) = (1 + |x|)\log(1 + |x|) - |x| \tag{2.3}$$

(in the sequel, Φ_1,Φ_2 and Φ_3 will always refer to these particular functions, with other symbols being used to denote generic Young functions). Any Young function Φ (including those with a jump to infinity) admits an integral representation

$$\Phi(x) = \int_0^x \phi(t)dt, \quad x \geq 0, \tag{2.4}$$

where $\phi : \mathbb{R}^+ \mapsto \overline{\mathbb{R}}^+$ is nondecreasing, left continuous, $\phi(0) = 0$ and $\phi(x) = +\infty$ for $x \geq a$ if $\Phi(x) = +\infty, x \geq a > 0$.

We define the complementary (conjugate) function to Φ as the Young function Ψ given by

$$\Psi(y) = \int_0^y \psi(t)dt, \quad y \geq 0, \tag{2.5}$$

where ψ is the generalised inverse of ϕ, that is

$$\psi(s) = \inf\{t : \phi(t) > s\}, \quad s \geq 0. \tag{2.6}$$

One can verify that Φ_2 and Φ_3 are a complementary pair.

Two Young functions Ψ_1 and Ψ_2 are said to be *equivalent* if there exist real numbers $0 < c_1 \leq c_2 < \infty$ and $x_0 \geq 0$ such that

$$\Psi_1(c_1 x) \leq \Psi_2(x) \leq \Psi_1(c_2 x), \quad x \geq x_0. \tag{2.7}$$

For example, the functions Φ_1 and Φ_2 are equivalent.

There are several classifications of Young functions according to their growth properties. The only one we are going to need for the construction of the information manifold is the so called Δ_2-class. A Young function $\Phi : \mathbb{R} \mapsto \mathbb{R}^+$ satisfies the Δ_2-condition if

$$\Phi(2x) \leq K\Phi(x), \quad x \geq x_0 \geq 0, \tag{2.8}$$

for some constant $K > 0$. Examples of functions in this class are $\Phi(x) = |x|^p, p \geq 1$ and the function Φ_3.

Now let $(\Omega, \mathcal{F}, \mu)$ be a measure space. The theory of Orlicz spaces can be developed using a general measure μ. However, in several important theorems, to get necessary and sufficient conditions, instead of just sufficient ones, one needs to impose a couple of technical restrictions on the measure. In this thesis, we are going to assume without further mention that all our measures have the *finite subset* property and are *diffuse* on a set of positive measure [52, p 46]. The reader must be aware that some of the results we are going to state do not hold if these conditions are not assumed and is refered to [52] for the full version of the theorems when unrestricted measures are considered. The finite subset condition only excludes pathological cases like $\mu(A) = 0$ if $A = \emptyset$ and $\mu(A) = \infty$ otherwise. It is satisfied, for instance, by all σ-finite measures. We also mention that the Lebesgue measure on the Borel σ-algebra of $\mathbb{R}^n$ is diffuse on a set of positive measure, as are many other measures likely to appear in applications of Information Geometry.

The *Orlicz class* associated with a Young function Φ is defined as

$$\tilde{L}^{\Phi}(\mu) = \left\{ f : \Omega \mapsto \overline{\mathbb{R}}, \text{measurable} : \int_{\Omega} \Phi(f) < \infty \right\} \tag{2.9}$$

It is a convex set. However, it is a vector space if and only if the function Φ satisfies the Δ_2-condition.

The *Orlicz space* associated with a Young function Φ is defined as

$$L^{\Phi}(\mu) = \left\{ f : \Omega \mapsto \overline{\mathbb{R}}, \text{measurable} : \int_{\Omega} \Phi(\alpha f) < \infty, \text{ for some } \alpha > 0 \right\} \tag{2.10}$$

It is easy to prove that this is a vector space and that it coincides with $\tilde{L}^{\Phi}$ iff Φ satisfies the Δ_2-condition. Moreover, if we identify functions which differ only on sets of measure zero, then L^{Φ} is a Banach space when furnished with the *Luxemburg* norm

$$N_{\Phi}(f) = \inf \left\{ k > 0 : \int_{\Omega} \Phi(\frac{f}{k}) d\mu \leq 1 \right\}, \tag{2.11}$$

or with the equivalent *Orlicz* norm

$$\|f\|_{\Phi} = \sup \left\{ \int_{\Omega} |fg| d\mu : g \in L^{\Psi}(\mu), \int_{\Omega} \Psi(g) d\mu \leq 1 \right\}, \tag{2.12}$$

where Ψ is the complementary Young function to Φ.

If two Young functions are equivalent, the Banach spaces associated with them coincide as sets and have equivalent norms. For example, $L^{\Phi_1}(\mu) = L^{\Phi_2}(\mu)$.

A key ingredient in the analysis of Orlicz spaces is the generalised Hölder inequality. If Φ and Ψ are complementary Young functions, $f \in L^{\Phi}(\mu)$, $g \in L^{\Psi}(\mu)$, then

$$\int_{\Omega} |fg| d\mu \leq 2 N_{\Phi}(f) N_{\Psi}(g). \tag{2.13}$$

It follows that $L^{\Phi} \subset \left(L^{\Psi}\right)^*$ for any pair of complementary Young functions, the inclusion being strict in general.

Suppose now that the measure space is finite. Then it is clear that $L^\infty(\mu) \subset L^\Phi(\mu)$. Let E^Φ denote the closure of L^∞ in the L^Φ-norm and define also

$$M^\Phi = \left\{ f \in L^\Phi \, : \, \int_\Omega \Phi(kf) < \infty, \text{ for all } k > 0 \right\}. \tag{2.14}$$

In general, we have that $M^\Phi \subset E^\Phi$. In the next lemma, we collect for later use the results for the case of a continuous Young function vanishing only at the origin. We need the following definition first. We say that $f \in L^\Phi(\mu)$ has an *absolutely continuous* norm $N_\Phi(f)$ if $N_\Phi(f\chi_{A_n}) \to 0$ for each sequence of measurable sets $A_n \downarrow \emptyset$. In terms of the Orlicz norm, this is equivalent to the statement that for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|f\chi_A\|_\Phi = \sup \left\{ \int_A |fg| d\mu : g \in L^\Psi(\mu), \int_\Omega \Psi(g) d\mu \leq 1 \right\} < \varepsilon \tag{2.15}$$

provided $A \in \mathcal{F}$ and $\mu(A) < \delta$.

Lemma 2.1.1 *Suppose that $\mu(\Omega) < \infty$ and let (Φ, Ψ) be a complementary pair of Young functions, Φ continuous, $\Phi(x) = 0$ iff $x = 0$. Then:*

i. $M^\Phi = E^\Phi$.

ii. $\left(M^\Phi\right)^* = L^\Psi$.

iii. $f \in M^\Phi$ *iff f has an absolutely continuous norm.*

Furthermore, M^Φ is separable as a topological space iff $(\Omega, \mathcal{F}, \mu)$ is separable as a measure space. If, moreover, Φ satisfies the Δ_2-condition, then $M^\Phi = L^\Phi$.

As consequences of this lemma, we obtain $\left(L^{\Phi_3}\right)^* = L^{\Phi_1}$ and $\left(M^{\Phi_1}\right)^* = L^{\Phi_3}$.

2.2 The Pistone-Sempi Information Manifold

Consider the set $\mathcal{M}$ of all densities of probability measures equivalent to the measure μ, that is,

$$\mathcal{M} \equiv \mathcal{M}(\Omega, \mathcal{F}, \mu) = \{f : \Omega \mapsto \mathbb{R}, \text{measurable} : \ f > 0 \text{ a.e. and } \int_\Omega f d\mu = 1\}.$$

For each point $p \in \mathcal{M}$, let $L^{\Phi_1}(p)$ be the exponential Orlicz space over the measure space $(\Omega, \mathcal{F}, pd\mu)$. The measure $pd\mu$ inherits all the good properties assumed for μ (finite subset property and diffusiveness) in addition to being finite, so that all the statements from the last section hold for $L^{\Phi_1}(p)$. Instead of using the whole of $L^{\Phi_1}(p)$ as the model Banach space for the manifold to be constructed, we restrict ourselves to $M^{\Phi_1}(p) \subset L^{\Phi_1}(p)$ and take its closed subspace of p-centred random variables

$$B_p = \{u \in M^{\Phi_1}(p) : \int_\Omega upd\mu = 0\} \tag{2.16}$$

as the coordinate Banach space.

For definiteness, we choose to work with the Orlicz norm $\|\cdot\|_{\Phi_1}$, although everything could be done with the equivalent Luxemburg norm N_{Φ_1}, and use the notation $\|\cdot\|_{\Phi_1,p}$ when it is necessary to specify the base point p.

In probabilistic terms, the set $M^{\Phi_1}(p)$ has the characterisation given in the following proposition, whose proof is a simple adaptation of the one given in [50] for the case of $L^{\Phi_1}(p)$.

Proposition 2.2.1 *$M^{\Phi_1}(p)$ coincides with the set of random variables for which the moment generating function $m_u(t)$ is finite for all $t \in \mathbb{R}$.*

Proof: If $u \in M^{\Phi_1}(p)$, then

$$\int_\Omega \Phi_1(tu)pd\mu = \int_\Omega \left(\frac{e^{tu} + e^{-tu}}{2} - 1\right) pd\mu < \infty, \quad \text{for all } t > 0,$$

which implies

$$\int_\Omega e^{tu}pd\mu < \infty, \quad \text{for all } t \in \mathbb{R}.$$

Conversely, if $m_u(t) < \infty$ for all $t \in \mathbb{R}$, then both $\int_\Omega e^{tu}pd\mu$ and $\int_\Omega e^{-tu}pd\mu$ are finite, so $\int_\Omega \Phi_1(tu)pd\mu < \infty$ for all $t > 0$, which means that $u \in M^{\Phi_1}(p)$.

In particular, the moment generating functional $Z_p(u) = \int_\Omega e^u pd\mu$ (otherwise known as the *partition function*) is finite on the whole of $M^{\Phi_1}(p)$.

We now define the inverse of a chart for $\mathcal{M}$. It is the same as the one used by Pistone and Sempi [50], except that they apply it to elements of L^{Φ_1}, as opposed to M^{Φ_1} as we do. Let $\mathcal{V}_p$ be the open unit ball of B_p and consider map,

$$\begin{aligned} e_p : \mathcal{V}_p &\rightarrow \mathcal{M} \\ u &\mapsto \frac{e^u}{Z_p(u)}p. \end{aligned} \tag{2.17}$$

Denote by $\mathcal{U}_p$ the image of $\mathcal{V}_p$ under e_p. We verify that e_p is a bijection from $\mathcal{V}_p$ to $\mathcal{U}_p$, since

$$\frac{e^u}{Z_p(u)}p = \frac{e^v}{Z_p(v)}p$$

implies that $(u - v)$ is a constant random variable. But since $(u - v) \in B_p$, we must have $u = v$. Then let e_p^{-1} be the inverse of e_p on $\mathcal{U}_p$. One can check that

$$\begin{aligned} e_p^{-1} : \mathcal{U}_p &\rightarrow B_p \\ q &\mapsto \log\left(\frac{q}{p}\right) - \int_\Omega \log\left(\frac{q}{p}\right) p d\mu. \end{aligned} \tag{2.18}$$

and also that, for any $p_1, p_2 \in \mathcal{M}$,

$$\begin{aligned} e_{p_2}^{-1} e_{p_1} : e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}) &\rightarrow e_{p_2}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}) \\ u &\mapsto u + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(u + \log\frac{p_1}{p_2}\right) p_2 d\mu. \end{aligned} \tag{2.19}$$

The next lemma is our contribution in proving that Pistone and Sempi's construction yields a manifold without using their notion of exponential convergence.

Lemma 2.2.2 *For any $p_1, p_2 \in \mathcal{M}$, the set $e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$ is open in the topology of B_{p_1}.*

Proof: Suppose that $q \in \mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}$ for some $p_1, p_2 \in \mathcal{M}$. Then we can write it as

$$q = \frac{e^u}{Z_p(u)}p_1,$$

for some $u \in \mathcal{V}_{p_1}$. Using (2.19), we find

$$e_{p_2}^{-1}(q) = u + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(u + \log\frac{p_1}{p_2}\right) p_2 d\mu.$$

Since $e_{p_2}^{-1}(q) \in \mathcal{V}_{p_2}$, we have that

$$\|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2} = \left\| u + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(u + \log\frac{p_1}{p_2}\right) p_2 d\mu \right\|_{\Phi_1,p_2} < 1.$$

Consider an open ball of radius r around $u = e_{p_1}^{-1}(q) \in e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$ in the topology of B_{p_1}, that is, consider the set

$$A_r = \{v \in B_{p_1} : \|v - u\|_{\Phi_1,p_1} < r\}$$

and let r be small enough so that $A_r \in \mathcal{V}_{p_1}$. Then the image in $\mathcal{M}$ of each point $v \in A_r$ under e_{p_1} is

$$\tilde{q} = e_{p_1}(v) = \frac{e^v}{Z_{p_1}(v)} p_1.$$

We claim that $\tilde{q} \in \mathcal{U}_{p_1} \cap \mathcal{U}_{p_2}$. Indeed, applying $e_{p_2}^{-1}$ to it we find

$$e_{p_2}^{-1}(\tilde{q}) = v + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(v + \log\frac{p_1}{p_2}\right) p_2 d\mu,$$

so

$$\begin{aligned}
\|e_{p_2}^{-1}(\tilde{q})\|_{\Phi_1,p_2} &\leq \|v-u\|_{\Phi_1,p_2} + \left\| u + \log\left(\frac{p_1}{p_2}\right) - \int_\Omega \left(u + \log\frac{p_1}{p_2}\right) p_2 d\mu \right\|_{\Phi_1,p_2} \\
&\quad + \left\| \int_\Omega (v-u) p_2 d\mu \right\|_{\Phi_1,p_2} \\
&\leq \|v-u\|_{\Phi_1,p_2} + \|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2} + \int_\Omega |v-u| p_2 d\mu \|1\|_{\Phi_1,p_2} \\
&= \|v-u\|_{\Phi_1,p_2} + \|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2} + \|v-u\|_{1,p_2} K, \qquad (2.20)
\end{aligned}$$

where $K = \|1\|_{\Phi_1,p_2}$ and we use the notation $\|\cdot\|_{1,p_2}$ for the $L^1(p_2)$-norm. It follows from the growth properties of Φ_1 that there exists $c_1 > 0$ such that $\|f\|_{1,p_2} \leq c_1 \|f\|_{\Phi_1,p_2}$. Moreover, it was found in [50, 49] that $L^{\Phi_1}(p_1) = L^{\Phi_1}(p_2)$, so there exists a constant $c_2 > 0$ such that $\|f\|_{\Phi_1,p_2} \leq c_2 \|f\|_{\Phi_1,p_1}$. Therefore, the inequality (2.20) becomes

$$\begin{aligned}
\|e_{p_2}^{-1}(\tilde{q})\|_{\Phi_1,p_2} &\leq c_2 \|v-u\|_{\Phi_1,p_1} + \|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2} + c_1 c_2 K \|v-u\|_{\Phi_1,p_1} \\
&= c_2(1 + c_1 K)\|v-u\|_{\Phi_1,p_1} + \|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2}.
\end{aligned}$$

Thus, is we choose

$$r < \frac{1 - \|e_{p_2}^{-1}(q)\|_{\Phi_1,p_2}}{c_2(1+c_1K)},$$

we will have that

$$\|e_{p_2}^{-1}(\tilde{q})\|_{\Phi_1,p_2} < 1$$

which proves the claim. What we just have proved is that $e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$ consists entirely of interior points in the topology of B_{p_1}, so $e_{p_1}^{-1}(\mathcal{U}_{p_1} \cap \mathcal{U}_{p_2})$ is open in B_{p_1}.

We then have that the collection $\{(\mathcal{U}_p, e_p^{-1}), p \in \mathcal{M}\}$ satisfies the three axioms for being a C^∞–atlas for $\mathcal{M}$ (see [34, p 20]), that is,

i. each $\mathcal{U}_p$ is a subset of $\mathcal{M}$ and the U_p's cover $\mathcal{M}$;

ii. each $e_p^{(-1)}$ is a bijection of $\mathcal{U}_p$ onto an open subset $e_p^{(-1)}(\mathcal{U}_p)$ of a Banach space B_p and for any i, j, $e_{p_i}(\mathcal{U}_{p_i} \cap \mathcal{U}_{p_j})$ is open in B_{p_i};

iii. the map

$$e_{p_j}^{-1} e_{p_i} : e_{p_j}^{-1}(\mathcal{U}_{p_i} \cap \mathcal{U}_{p_j}) \to e_{p_j}^{-1}(\mathcal{U}_{p_i} \cap \mathcal{U}_{p_j})$$

is a C^∞–isomorphism for each pair i, j.

Moreover, since all the spaces B_p are toplinear isomorphic, we can say that $\mathcal{M}$ is a C^∞–manifold modelled on B_p.

As usual, the tangent space at each point $p \in \mathcal{M}$ can be abstractly identified with B_p. A concrete realisation has been given in [49, proposition 21], namely each curve through $p \in \mathcal{M}$ is tangent to a one-dimensional exponential model $\frac{e^{tu}}{Z_p(tu)}p$, so we take u as the tangent vector representing the equivalence class of such a curve.

Since we are using M^{Φ_1} instead of L^{Φ_1} to construct the manifold, we need the following corresponding definition for the *maximal exponential model* at each $p \in \mathcal{M}$:

$$\mathcal{E}(p) = \left\{ \frac{e^u}{Z_p(u)} p, u \in B_p \right\}. \tag{2.21}$$

The following proposition is included here for completeness.

Proposition 2.2.3 *$\mathcal{E}(p)$ is the connected component of $\mathcal{M}$ containing p.*

Proof: We first need to prove that $\mathcal{E}(p)$ is indeed connected. Let $q_1, q_2 \in \mathcal{E}(p)$. Then we can write

$$\begin{aligned} q_1 &= \frac{e^{u_1}}{Z_p(u_1)} p \\ q_2 &= \frac{e^{u_2}}{Z_p(u_2)} p \end{aligned}$$

for some $u_1, u_2 \in B_p$. Therefore

$$q_2 = \frac{e^{u_2 - u_1} Z_p(u_1)}{Z_p(u_2)} q_1.$$

Since $M^{\Phi_1}(q_1) = M^{\Phi_1}(p)$, we can define an element of B_{q_1} by

$$u = (u_2 - u_1) - \int_\Omega (u_2 - u_1) q_1 d\mu$$

and write $q_2 = \frac{e^u}{Z_{q_1}(u)} q_1$.

It is then clear that q_2 and q_1 can be joined by a finite sequence $q_1 = f_0, f_1, \dots, f_n = q_2$ where each f_j belongs to $\mathcal{U}_{f_{j-1}}$, for $j = 1, \dots, n$. We just need to put, for instance,

$$f_1 = \frac{e^{\frac{u}{2\|u\|}}}{Z_{q_1}\left(\frac{u}{2\|u\|}\right)} q_1$$

to obtain $f_1 \in \mathcal{U}_{q_1}$, since $\left\| \frac{u}{2\|u\|} \right\| < 1$ and u is centred around q_1. For the next element in the sequence we can write

$$f_2 = \frac{e^{v_1}}{Z_{q_1}(v_1)} f_1$$

where v_1 can be chosen as

$$v_1 = \frac{u - \int_\Omega u f_1 d\mu}{2\|u - \int_\Omega u f_1 d\mu\|}$$

so that $\|v_1\| < 1$ and v_1 is centred around f_1. We proceed in this fashion, using only multiples of u and constants in order to have all the elements of the sequence in the right neighbourhoods until we finally reach q_2.

We now need to prove maximality of $\mathcal{E}(p)$, that is, if we have a sequence $p = f_0, f_1, \ldots, f_n$ such that each f_j belongs to $\mathcal{U}_{f_{j-1}}$, for $j = 1, \ldots, n$, then $f_n \in \mathcal{E}(p)$. It suffices to show that given $f_1 \in \mathcal{E}(p)$ and $f_2 \in \mathcal{U}_{f_1}$, then $f_2 \in \mathcal{E}(p)$. But this follows clearly from just writing

$$\begin{aligned} f_1 &= \frac{e^u}{Z_p(u)} p, \qquad \text{for some } u \in B_p, \\ f_2 &= \frac{e^v}{Z_{f_1}(v)} f_1, \qquad \text{for some } v \in \mathcal{V}_{f_1}, \end{aligned}$$

so that

$$f_2 = \frac{e^{v+u}}{Z_p(u) Z_{f_1}(v)} q = \frac{e^{\tilde{u}}}{Z_p(\tilde{u})} q$$

where $\tilde{u} = (v + u) - \int_\Omega (v + u) q d\mu$.

2.3 The Fisher Information and Dual Connections

In the parametric version of Information Geometry, Amari and Nagaoka have introduced the concept of dual connections with respect to a Riemannian metric [2]. For finite dimensional manifolds, any continuous assignment of a positive definite symmetric bilinear form to each tangent space determines a Riemannian metric. In infinite dimensions, we need to impose that the tangent space is self-dual and that the bilinear form is bounded. Since our tangent spaces B_p are not even reflexive, let alone self-dual, we abandon the idea of having a Riemannian structure on $\mathcal{M}$ and propose a weaker version of duality, the duality with respect to a continuous scalar product. When restricted to finite dimensional submanifolds, the scalar product becomes a Riemannian metric and the original definition of duality is recovered.

Let $\langle \cdot, \cdot \rangle_p$ be a continuous positive definite symmetric bilinear form assigned continuously to each $B_p \simeq T_p\mathcal{M}$. A pair of connections (∇, ∇^*) are said to be dual with respect to $\langle \cdot, \cdot \rangle_p$ if

$$\langle \tau u, \tau^* v \rangle_q = \langle u, v \rangle_p \tag{2.22}$$

for all $u, v \in T_p\mathcal{M}$ and all smooth curves $\gamma : [0,1] \to \mathcal{M}$ such that $\gamma(0) = p$,$\gamma(1) = q$, where τ and τ^* denote the parallel transports associated with ∇ and ∇^*, respectively. Equivalently, (∇, ∇^*) are dual with respect to $\langle \cdot, \cdot \rangle_p$ if

$$v\left(\langle s_1, s_2 \rangle_p\right) = \langle \nabla_v s_1, s_2 \rangle_p + \langle s_1, \nabla_v^* s_2 \rangle_p \tag{2.23}$$

for all $v \in T_p\mathcal{M}$ and all smooth vector fields s_1 and s_2.

We stress that this *is not* the kind of duality obtained when a connection ∇ on a vector bundle $\mathcal{B}$ is used to construct another connection ∇' on the dual vector bundle $\mathcal{B}'$ as defined, for instance, in [15, definiton 6]. The latter is a construction that does not involve any metric or scalar product and the two connections act on different bundles, while Amari's duality is a duality with respect to a specific scalar product (or metric, in the finite dimensional case) and the dual connections act on the same bundle, the tangent bundle.

The infinite dimensional generalisation of the Fisher information is given by

$$\langle u, v \rangle_p = \int_\Omega (uv) p d\mu, \quad \forall u, v \in B_p. \tag{2.24}$$

This is clearly bilinear, symmetric and positive definite. Also, since $L^{\Phi_1} \subset L^{\Phi_3}$, the generalised Hölder inequality gives

$$|\langle u, v \rangle_p| \leq K \|u\|_{\Phi_1,p} \|v\|_{\Phi_1,p}, \quad \forall u, v \in B_p, \tag{2.25}$$

which implies the continuity of $\langle \cdot, \cdot \rangle_p$.

The use of exponential Orlicz space to model the manifold induces naturally a globally flat affine connection on the tangent bundle $T\mathcal{M}$, called the *exponential* connection and denoted by $\nabla^{(1)}$. It is defined on each connected component of the manifold $\mathcal{M}$, which is equivalent to saying that its parallel transport is defined between points connected by an exponential model [50, theorem 4.1]. If p and q

are two such points, then $L^{\Phi_1}(p) = L^{\Phi_1}(q)$ and the exponential parallel transport is given by

$$\begin{aligned} \tau_{pq}^{(1)} : T_p\mathcal{M} &\rightarrow T_q\mathcal{M} \\ u &\mapsto u - \int_\Omega uqd\mu. \end{aligned} \tag{2.26}$$

It is well defined, since $T_p\mathcal{M} = B_p$ and $T_q\mathcal{M} = B_q$ are subsets of the same set $M^{\Phi_1}(p) = M^{\Phi_1}(q)$, so the exponential parallel transport just subtracts a constant from u to make it centred around the right point.

We now want to define the dual connection to $\nabla^{(1)}$ with respect to the Fisher information. We begin by proving the following lemma.

Lemma 2.3.1 *Let p and q be two points in the same connected component of $\mathcal{M}$. Then $\frac{p}{q}u \in B_q$, for all $u \in B_p$.*

Proof: From the hypothesis, u has absolutely continuous norm in $L^{\Phi_1}(p)$, so for every $\varepsilon > 0$, there exists $\delta > 0$ such that $A \in \mathcal{F}$ and $\mu(A) < \delta$ implies

$$\|u\chi_A\|_{\Phi_1,p} = \sup\left\{\int_A |uv|pd\mu : v \in L^{\Phi_3}(p), \int_\Omega \Phi_3(v)pd\mu \leq 1\right\} < \varepsilon. \tag{2.27}$$

But since $M^{\Phi_1}(p) = M^{\Phi_1}(q)$, as they are the completion of the same set L^∞ under equivalent norms (recall that p and q are supposed to be connected by an exponential model), we have that $L^{\Phi_3}(p) = (M^{\Phi_1}(p))^* = (M^{\Phi_1}(q))^* = L^{\Phi_3}(q)$, in the sense that they are the same set furnished with equivalent norms. We then use (2.27) to conclude that

$$\begin{aligned} \varepsilon &> \sup\left\{\int_A |uv|pd\mu : v \in L^{\Phi_3}(p), \int_\Omega \Phi_3(v)pd\mu \leq 1\right\} \\ &= \sup\left\{\int_A |uv|pd\mu : v \in L^{\Phi_3}(p), N_{\Phi_3,p}(v) \leq 1\right\} \\ &= \frac{1}{k}\sup\left\{\int_A |\frac{p}{q}u(kv)|qd\mu : v \in L^{\Phi_3}(p), N_{\Phi_3,p}(v) \leq 1\right\} \\ &\geq \frac{1}{k}\sup\left\{\int_A |\frac{p}{q}u(v')|qd\mu : v' \in L^{\Phi_3}(q), N_{\Phi_3,q}(v') \leq 1\right\} \\ &= \frac{1}{k}\left\|\frac{p}{q}u\chi_A\right\|_{\Phi_1,q}, \end{aligned}$$

where we used the facts that $\int_\Omega \Phi_3(v)pd\mu \le 1$ iff $N_{\Phi_3,p}(v) \le 1$ [52, theorem 3.2.3] and that there exists a constant k such that $N_{\Phi_3,p}(\cdot) \le kN_{\Phi_3,q}(\cdot)$. Since ε was arbitrary, this proves that $\frac{p}{q}u$ has absolutely continuous norm in $L^{\Phi_1}(q)$. The lemma then follows from lemma 2.1.1 and the fact that $\frac{p}{q}u$ is centred around q.

We can then define the *mixture* connection on $T\mathcal{M}$, as

$$\begin{aligned} \tau^{(-1)}_{pq} : T_p\mathcal{M} &\rightarrow T_q\mathcal{M} \\ u &\mapsto \frac{p}{q}u, \end{aligned} \tag{2.28}$$

for p and q in the same connected component of $\mathcal{M}$. We notice that it is also globally flat and prove the following result.

Theorem 2.3.2 *The connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ are dual with respect to the Fisher information.*

Proof: We have that

$$\begin{aligned} \langle \tau^{(1)}u, \tau^{(-1)}v\rangle_q &= \left\langle u - \int_\Omega uqd\mu, \frac{p}{q}v \right\rangle_q \\ &= \int_\Omega u(p/q)vqd\mu - \left(\int_\Omega uqd\mu\right)\int_\Omega (p/q)vqd\mu \\ &= \int_\Omega uvpd\mu \\ &= \langle u, v\rangle_p, \quad \forall u, v \in B_p, \end{aligned}$$

where, to go from the second to the third line above, we used that v is centred around p.

2.4 Covariant Derivatives and Geodesics

We begin this section recalling that the covariant derivative for the exponential connection has been computed in [15, proposition 25] and found to be

$$\left(\nabla^{(1)}_v s\right)(p) = (d_v s)(p) - E_p\left((d_v s)(p)\right), \tag{2.29}$$

where $s \in S(T\mathcal{M})$ is a differentiable vector field, $v \in T_p\mathcal{M}$ is a tangent vector at p, $E_p(\cdot)$ denotes the expected value with respect to the measure $pd\mu$ and $d_v s$ denotes the directional derivative in M^{Φ_1} of s composed with some patch e_p as a map between Banach spaces.

In section 2.6 we shall see that this gives the usual covariant derivative for the exponential connection in parametric information geometry [38, pp 117-118].

We can also verify that one-dimensional exponential models of the form

$$q(t) = \frac{e^{tu}}{Z_p(tu)} p \tag{2.30}$$

are geodesics for $\nabla^{(1)}$, since if $s(t) = \frac{d}{dt}\left(\log \frac{q(t)}{p}\right)$ is the vector field tangent to $q(t)$ at each point t [49, proposition 21], then (2.29) gives

$$\begin{aligned}
\left(\nabla^{(1)}_{\dot{q}(t)} s(t)\right)(q(t)) &= \frac{d^2}{dt^2}\left(tu - \log Z_p(tu)\right) - E_{q(t)}\left(\frac{d^2}{dt^2}\left(tu - \log Z_p(tu)\right)\right) \\
&= -\frac{d^2}{dt^2}\log Z_p(tu) + E_{q(t)}\left(\frac{d^2}{dt^2}\log Z_p(tu)\right) \\
&= 0.
\end{aligned}$$

As we emphasised in the previous section, the definition given in [15, definition 22] for the mixture connection differs from ours (due to the different concepts of duality employed), so we have to compute its covariant derivative according to the definition given here, at least to have the notation right.

Proposition 2.4.1 *Let $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{M}$ be a smooth curve such that $p = \gamma(0)$ and $v = (e_p^{-1} \circ \gamma)'(0)$ and let $s \in S(T\mathcal{M})$ be a differentiable vector field. Then*

$$\left(\nabla^{(-1)}_v s\right)(p) = (d_v s)(p) + s(p)\ell'(0), \tag{2.31}$$

where $\ell(t) = \log(\gamma(t))$.

Proof:

$$\begin{aligned}
\left(\nabla_v^{(-1)} s\right)(p) &= \left(\nabla^{(-1)}_{(e_p^{-1}\circ\gamma)'(0)} s\right)(\gamma(0)) \\
&= \lim_{h\to 0}\frac{1}{h}\left[\tau^{(-1)}_{\gamma(h)\gamma(0)} s(\gamma(h)) - s(\gamma(0))\right] \\
&= \lim_{h\to 0}\frac{1}{h}\left[\frac{\gamma(h)}{\gamma(0)} s(\gamma(h)) - s(\gamma(0))\right] \\
&= \lim_{h\to 0}\frac{1}{h}\left[s(\gamma(h)) - s(\gamma(0))\right] + \lim_{h\to 0}\frac{1}{h}\left[\frac{\gamma(h)-\gamma(0)}{\gamma(0)} s(\gamma(h))\right] \\
&= (d_v s)(p) + s(p)\ell'(0).
\end{aligned}$$

Again this reduces to the parametric result for the case of finite dimensional submanifolds of $\mathcal{M}$, as we shall see in section 2.6.

The mixture connection owes its name to the fact that in the parametric version of Information Geometry a convex mixture of two densities describes a geodesic with respect to $\nabla^{(-1)}$. To verify the same statement in the nonparametric case, we first need to check that a convex mixture of two points in a connected component of $\mathcal{M}$ remains in the same connected component.

Proposition 2.4.2 *If q_1 and q_2 are two points in $\mathcal{E}(p)$ for some $p \in \mathcal{M}$, then*

$$q(t) = tq_1 + (1-t)q_2 \tag{2.32}$$

belongs to $\mathcal{E}(p)$ for all $t \in (0,1)$.

Proof: We begin by writing

$$q_1 = \frac{e^{u_1}}{Z_p(u_1)}p \quad \text{and} \quad q_2 = \frac{e^{u_2}}{Z_p(u_2)}p,$$

for some $u_1, u_2 \in B_p \subset M^{\Phi_1}(p)$. To simplify the notation, let us define

$$\tilde{u}_1 = u_1 - \log Z_p(u_1) \quad \text{and} \quad \tilde{u}_2 = u_2 - \log Z_p(u_2).$$

We want to show that, if we write

$$e^{\tilde{u}}p = q(t) = te^{\tilde{u}_1}p + (1-t)e^{\tilde{u}_2}p,$$

then $\tilde{u}$ is an element of $M^{\Phi_1}(p)$, so that

$$u = \tilde{u} - \int_\Omega \tilde{u} p d\mu \in B_p$$

and

$$q(t) = \frac{e^u}{Z_p(u)} \in \mathcal{E}(p).$$

All we need to prove is that both $\int_\Omega e^{k\tilde{u}} p d\mu$ and $\int_\Omega e^{-k\tilde{u}} p d\mu$ are finite for all $k > 0$. We have that

$$e^{\tilde{u}} = t e^{\tilde{u}_1} + (1-t) e^{\tilde{u}_2}$$

which implies

$$\begin{aligned} e^{k\tilde{u}} &= \left(t e^{\tilde{u}_1} + (1-t) e^{\tilde{u}_2}\right)^k \\ &\leq 2^k \left(t^k e^{k\tilde{u}_1} + (1-t)^k e^{k\tilde{u}_2}\right). \end{aligned}$$

Thus

$$\int_\Omega e^{k\tilde{u}} p d\mu \leq 2^k t^k \int_\Omega e^{k\tilde{u}_1} p d\mu + 2^k (1-t)^k \int_\Omega e^{k\tilde{u}_2} p d\mu < \infty$$

since both $\tilde{u}_1$ and $\tilde{u}_2$ are in $M^{\Phi_1}(p)$. As for the other integral, observe that

$$\begin{aligned} e^{-k\tilde{u}} &= \left(t e^{\tilde{u}_1} + (1-t) e^{\tilde{u}_2}\right)^{-k} \\ &= \frac{1}{\left(t e^{\tilde{u}_1} + (1-t) e^{\tilde{u}_2}\right)^k} \\ &\leq \frac{1}{t^k e^{k\tilde{u}_1}} = t^{-k} e^{-k\tilde{u}_1}. \end{aligned}$$

Therefore

$$\int_\Omega e^{-k\tilde{u}} p d\mu \leq t^{-k} \int_\Omega e^{-k\tilde{u}_1} p d\mu < \infty,$$

since $\tilde{u}_1 \in M^{\Phi_1}(p)$.

We can now verify that a family of the form

$$q(t) = t q_1 + (1-t) q_2, \qquad t \in (0,1)$$

is a geodesic for $\nabla^{(-1)}$. Let $s(t) = \frac{d}{dt}\left(\log \frac{q(t)}{p}\right)$ be the vector field tangent to $q(t)$ at each point t, then (2.31) gives

$$
\begin{aligned}
\left(\nabla^{(-1)}_{(e_p^{-1}\circ q)'(t)} s(t)\right)(q(t)) &= \frac{d^2}{dt^2}\left[\log \frac{tq_1 + (1-t)q_2}{p}\right] \\
&\quad + \frac{d}{dt}\left[\log \frac{tq_1 + (1-t)q_2}{p}\right] \frac{d}{dt}[\log tq_1 + (1-t)q_2] \\
&= \frac{d}{dt}\left[\frac{p}{tq_1 + (1-t)q_2} \frac{(q_1 - q_2)}{p}\right] \\
&\quad + \left(\frac{p}{tq_1 + (1-t)q_2} \frac{q_1 - q_2}{p}\right) \frac{(q_1 - q_2)}{tq_1 + (1-t)q_2} \\
&= -\left(\frac{(q_1 - q_2)}{tq_1 + (1-t)q_2}\right)^2 + \left(\frac{(q_1 - q_2)}{tq_1 + (1-t)q_2}\right)^2 \\
&= 0.
\end{aligned}
$$

2.5 α-connections

In this section, we address the definition of the infinite dimensional analogue of the α-connections introduced in the parametric case independently by Chentsov [8] and Amari [1]. We use the same technique proposed by Gibilisco and Pistone [15], namely exploring the geometry of spheres in the Lebesgue spaces L^r, but modified in such a way that the resulting connections all act on the tangent bundle $T\mathcal{M}$.

We begin with the generalised Amari α-embeddings

$$
\begin{aligned}
\ell_\alpha : \mathcal{M} &\to L^r(\mu) \\
p &\mapsto \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}, \quad \alpha \in (-1, 1),
\end{aligned}
\tag{2.33}
$$

where $r = \frac{2}{1-\alpha}$.

Observe that

$$
\|\ell_\alpha(p)\|_r = \left[\int_\Omega \ell_\alpha(p)^r d\mu\right] = \left[\int_\Omega \left(\frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}\right)^r d\mu\right]^{1/r} = r,
$$

so $\ell_\alpha(p) \in S^r(\mu)$, the sphere of radius r in $L^r(\mu)$ (we warn the reader that throughout this chapter, the r in S^r refers to the fact that this is a sphere of radius r, while the fact that it is a subset of L^r is judiciously omitted from the notation).

According to Gibilisco and Pistone [15], the tangent space to $S^r(\mu)$ at a point f is

$$T_f S^r(\mu) = \left\{ g \in L^r(\mu) : \int_\Omega g f^* d\mu = 0 \right\}, \tag{2.34}$$

where $f^* = \mathrm{sgn}(f)|f|^{r-1}$. In our case,

$$f = \ell_\alpha(p) = rp^{1/r} \tag{2.35}$$

so that

$$f^* = \left(rp^{1/r}\right)^{r-1} = r^{r-1}p^{1-1/r}. \tag{2.36}$$

Therefore, the tangent space to $S^r(\mu)$ at $rp^{1/r}$ is

$$T_{rp^{1/r}} S^r(\mu) = \left\{ g \in L^r(\mu) : \int_\Omega g p^{1-1/r} d\mu = 0 \right\}. \tag{2.37}$$

We now look for a concrete realisation of the push-forward of the map ℓ_α when the tangent space $T_p\mathcal{M}$ is identified with B_p as in the previous sections. Since

$$\frac{d}{dt}\left(\frac{2}{1-\alpha}p^{\frac{1-\alpha}{2}}\right) = p^{\frac{1-\alpha}{2}}\frac{d\log p}{dt},$$

the push-forward can be formally implemented as

$$\begin{aligned} (\ell_\alpha)_{*(p)} : T_p\mathcal{M} = B_p &\rightarrow T_{rp^{1/r}} S^r(\mu) \\ u &\mapsto p^{\frac{1-\alpha}{2}} u. \end{aligned} \tag{2.38}$$

For this to be well defined, we need to check that $p^{\frac{1-\alpha}{2}}u$ is an element of $T_{rp^{1/r}}S^r(\mu)$. Indeed, since $L^{\Phi_1}(p) \subset L^s(p)$ for all $s > 1$, we have that

$$\int_\Omega \left(p^{1/r}u\right)^r d\mu = \int_\Omega u^r p d\mu < \infty,$$

so $p^{\frac{1-\alpha}{2}}u \in L^r(\mu)$. Moreover

$$\int_\Omega p^{1/r} u p^{1-1/r} d\mu = \int_\Omega u p d\mu = 0,$$

which verifies that $p^{1/r}u \in T_{rp^{1/r}}S^r(\mu)$.

The sphere $S^r(\mu)$ inherits a natural connection obtained by projecting the trivial connection on $L^r(\mu)$ (the one where parallel transport is just the identity map) onto its tangent space at each point. For each $f \in S^r(\mu)$, a canonical projection from the tangent space $T_f L^r(\mu)$ onto the tangent space $T_f S^r(\mu)$ can be uniquely defined, since the spaces $L^r(\mu)$ are uniformly convex [13], and is given by

$$\begin{aligned} \Pi_f : T_f L^r(\mu) &\rightarrow T_f S^r(\mu) \\ g &\mapsto g - \left(r^{-r} \int_\Omega g f^* d\mu \right) f. \end{aligned} \tag{2.39}$$

When $f = rp^{1/r}$ and $f^* = r^{r-1}p^{1-1/r}$, the formula above gives

$$\begin{aligned} \Pi_{rp^{1/r}} : T_{rp^{1/r}} L^r(\mu) &\rightarrow T_{rp^{1/r}} S^r(\mu) \\ g &\mapsto g - \left(\int_\Omega g p^{1-1/r} d\mu \right) p^{1/r}. \end{aligned} \tag{2.40}$$

We are now ready to formally define the covariant derivatives for the α-connections. In what follows, $\widetilde{\nabla}$ is used to denote the trivial connection on $L^r(\mu)$.

Definition 2.5.1 For $\alpha \in (-1,1)$, let $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{M}$ be a smooth curve such that $p = \gamma(0)$ and $v = \dot{\gamma}(0)$ and let $s \in S(T\mathcal{M})$ be a differentiable vector field. The α-connection on $T\mathcal{M}$ is given by

$$\left(\nabla_v^{(\alpha)} s \right)(p) = (\ell_\alpha)^{-1}_{*(p)} \left[\Pi_{rp^{1/r}} \widetilde{\nabla}_{(\ell_\alpha)_{*(p)} v} (\ell_\alpha)_{*(\gamma(t))} s \right]. \tag{2.41}$$

A formula like (2.41) deserves a more wordy explanation. We take the vector field s and push it forward along the curve γ to obtain $(\ell_\alpha)_{*(\gamma(t))} s$. Then we take its covariant derivative with respect to the trival connection $\widetilde{\nabla}$ in the direction of $(\ell_\alpha)_{*(p)} v$, the push-forward of the tangent vector v. The result is a vector in $T_{rp^{1/r}} L^r(\mu)$, so we use the canonical projection Π to obtain a vector in $T_{rp^{1/r}} S^r(\mu)$. Finally, we pull it back to $T_p\mathcal{M}$ using $(\ell_\alpha)^{-1}_{*(p)}$.

The next theorem shows that the relation between the exponential, the mixture and the α-connections just defined is the same as in the parametric case. Its proof resembles the calculation in the last pages of [15], except that all our connections act on the same bundle, whereas in [15] each one is defined on its own bundle-connection pair. It shows that the α-connections are well defined objects at the level of covariant derivatives.

Theorem 2.5.2 *The exponential, mixture and α-covariant derivatives on $T\mathcal{M}$ satisfy*

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}. \tag{2.42}$$

Proof: Let $\ell(t) = \log(\gamma(t))$ with γ, s, p and v as in definition 2.5.1. Before explicitly computing the derivatives in (2.41), observe that since $s(\gamma(t)) \in B_{\gamma(t)}$ for each $t \in (-\varepsilon, \varepsilon)$, we have

$$\begin{aligned}
\frac{d}{dt}\int_\Omega s(\gamma(t))\gamma(t)d\mu &= 0 \\
\int_\Omega \frac{ds(\gamma(t))}{dt}\gamma(t)d\mu &= -\int_\Omega s(\gamma(t))\frac{d\gamma(t)}{dt}d\mu \\
\int_\Omega \frac{ds(\gamma(t))}{dt}\gamma(t)d\mu &= -\int_\Omega s(\gamma(t))\frac{d\log(\gamma(t))}{dt}\gamma(t)d\mu.
\end{aligned}$$

In particular, for $t = 0$, we get

$$\int_\Omega (d_v s)(p)pd\mu = -\int_\Omega s(p)\dot{\ell}(0)pd\mu \tag{2.43}$$

We can now look more closely at (2.41). It reads

$$\begin{aligned}
\left(\nabla_v^{(\alpha)} s\right)(p) &= p^{-1/r}\left[\Pi_{rp^{1/r}}\widetilde{\nabla}_{p^{1/r}v}\gamma(t)^{1/r}s(\gamma(t))\right] \\
&= p^{-1/r}\left[\Pi_{rp^{1/r}}\frac{d}{dt}\left(\gamma(t)^{1/r}s(\gamma(t))\right)\Big|_{t=0}\right] \\
&= p^{-1/r}\left[\Pi_{rp^{1/r}}\left(\frac{1}{r}p^{1/r}\left.\frac{d\log(\gamma(t))}{dt}\right|_{t=0} s(p) + p^{1/r}\left.\frac{ds(\gamma(t))}{dt}\right|_{t=0}\right)\right] \\
&= p^{-1/r}\left[\Pi_{rp^{1/r}}\left(\frac{1}{r}p^{1/r}\ell'(0)s(p) + p^{1/r}(d_v s)(p)\right)\right] \\
&= \frac{1}{r}\ell'(0)s(p) + (d_v s)(p) - \int_\Omega\left(\frac{1}{r}\ell'(0)s(p) + (d_v s)(p)\right)pd\mu.
\end{aligned}$$

At this point we make use of (2.43) in the integrand above to obtain

$$\begin{aligned}
\left(\nabla_v^{(\alpha)} s\right)(p) &= \frac{1}{r}\ell'(0)s(p) + (d_v s)(p) + \left(\frac{1}{r} - 1\right)\int_\Omega (d_v s)(p)pd\mu \\
&= \left(\frac{1+\alpha}{2}\right)\left[(d_v s)(p) - E_p\left((d_v s)(p)\right)\right] \\
&\qquad + \left(\frac{1-\alpha}{2}\right)\left[(d_v s)(p) + s(p)\ell'(0)\right] \\
&= \frac{1+\alpha}{2}\left(\nabla_v^{(1)} s\right)(p) + \frac{1-\alpha}{2}\left(\nabla_v^{(-1)} s\right)(p).
\end{aligned}$$

Corollary 2.5.3 *The connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are dual with respect to the Fisher information $\langle\cdot,\cdot\rangle_p$, for all $\alpha \in (-1,1)$.*

Proof: Let $v \in T_p\mathcal{M}$ be a tangent vector at p and s_1 and s_2 be smooth vector fields. Since the exponential and mixture connections are dual with respect to the Fisher information, we can use first 2.42 and then 2.23 to obtain

$$\begin{aligned}
\langle\nabla^{(\alpha)}_v s_1, s_2\rangle_p + \langle s_1, \nabla^{(-\alpha)}_v s_2\rangle_p &= \frac{1+\alpha}{2}\langle\nabla^{(1)}_v s_1, s_2\rangle_p + \frac{1-\alpha}{2}\langle\nabla^{(-1)}_v s_1, s_2\rangle_p \\
&\quad + \frac{1-\alpha}{2}\langle s_1, \nabla^{(1)}_v s_2\rangle_p + \frac{1+\alpha}{2}\langle s_1, \nabla^{(-1)}_v s_2\rangle_p \\
&= \frac{1+\alpha}{2}\left(\langle\nabla^{(1)}_v s_1, s_2\rangle_p + \langle s_1, \nabla^{(-1)}_v s_2\rangle_p\right) \\
&\quad + \frac{1-\alpha}{2}\left(\langle\nabla^{(-1)}_v s_1, s_2\rangle_p + \langle s_1, \nabla^{(1)}_v s_2\rangle_p\right) \\
&= v\left(\langle s_1, s_2\rangle_p\right)
\end{aligned}$$

2.6 Parametric Classical Statistical Manifolds

We start this section with the definition of submanifolds of infinite dimensional manifolds, conveniently phrased for the case we have in mind [34]. A subset $\mathcal{N}$ of our manifold $\mathcal{M}$ is said to be a submanifold of $\mathcal{M}$ if for each point $p \in \mathcal{N}$ there exist a chart $(\mathcal{W}_p, \sigma_p)$ such that

i. σ_p is an isomorphism of $\mathcal{W}_p$ with a product $\mathcal{V}_1 \times \mathcal{V}_2$ where $\mathcal{V}_1$ is open in some space V_1, $\mathcal{V}_2$ is open in some space V_2 and $V_1 \times V_2 \simeq B_p$;

ii. $\sigma(\mathcal{N} \cap \mathcal{W}_p) = \mathcal{V}_1 \times \{0\}$.

Furthermore, the map σ_p induces a bijection

$$\tilde{\sigma}_p : \mathcal{N} \cap \mathcal{W}_p \to \mathcal{V}_1. \tag{2.44}$$

As previously anticipated, we adopt the following definition.

Definition 2.6.1 A parametric classical statistical manifold $\mathcal{S}$ is a finite dimensional submanifold of the Pistone-Sempi manifold $\mathcal{M}$.

This means that, if $\mathcal{S}$ is such a manifold, then V_1 is (homeomorphic to) a finite dimensional subspace of B_p and, given a basis for V_1, we can take as coordinates for $q \in \mathcal{S} \cap \mathcal{W}_p$ the coefficients of the expansion of $\tilde{\sigma}_p(q) \in V_1$ in such basis. These coordinates will be called the parameters of q relative to the chart $\mathcal{W}_p$ or, when the whole of $\mathcal{S}$ is covered by a single chart, simply the parameters of q.

We want to verify that this definition does reproduce all the technical assumptions usually required for a set of probability distributions to be a parametric statistical manifold. We take the recent book by Amari and Nagaoka [2, pp 26-29] as a reference for such conditions.

Let $\mathcal{S}$ be an n-dimensional parametric statistical manifold according to definition 2.6.1 and assume for simplicity (as it is often done in most of the literature concerning them) that $\mathcal{S}$ is covered by a single chart. Let $\theta = (\theta^1, \dots, \theta^n)$ be a coordinate system for it, obtained as above, and denote by $p(x, \theta)$ a point in $\mathcal{S}$, where $x \in \Omega$. Then first of all the domain where θ ranges is an open subset Ξ of $\mathbb{R}^n$ and for each $x \in \Omega$ the function $\theta \mapsto p(x, \theta)$ is C^∞ (being the inverse of a chart for a C^∞ manifold).

We now want to prove that we can differentiate with respect to the parameters θ under the integral sign in Ω. For example, we want formulae like the following to hold

$$\int_\Omega \frac{\partial p(x,\theta)}{\partial \theta^i} d\mu = \frac{\partial}{\partial \theta^i} \int_\Omega p(x,\theta) d\mu = \frac{\partial}{\partial \theta^i} 1 = 0. \tag{2.45}$$

For the case where Ω is a topological space in its own right, to guarantee that we can interchange integration and differentiantion it is enough to prove [3, theorem 10.39] that for each $i = 1, \dots, n$ there exists a nonnegative integrable function G_i such that

$$\left| \frac{\partial p(\theta, x)}{\partial \theta^i} \right| \leq G_i(x), \tag{2.46}$$

for all interior points of $\Xi \times \Omega$. Let us consider each θ^i separately, fixing all the other parameters. We are then dealing with a regular curve $p(x, t) \in \mathcal{M}$, $t \equiv \theta^i$, and [49, propositon 21-1] tells us that

$$\frac{d}{dt}\log\frac{p(t)}{p}\bigg|_{t=0} \in L^{\Phi_1}(p), \tag{2.47}$$

where $p = p(0)$. But we also have that

$$\frac{d}{dt}\log\frac{p(t)}{p} = \frac{p}{p(t)}\frac{d}{dt}\frac{p(t)}{p}$$

so that

$$\frac{d}{dt}p(t)\bigg|_{t=0} = p\,\frac{d}{dt}\log\frac{p(t)}{p}\bigg|_{t=0}.$$

Now we can use the fact that $L^{\Phi_1}(p) \subset L^1(p)$ to conclude that

$$\int_\Omega \frac{d}{dt}p(t)\bigg|_{t=0} d\mu = \int_\Omega \frac{d}{dt}\log\frac{p(t)}{p}\bigg|_{t=0} p d\mu < \infty, \tag{2.48}$$

which certainly ensures (2.46) if we choose trivially $G_i(x) = \left|\frac{\partial p(\theta,x)}{\partial\theta^i}\right|$.

We now move to the subject of the α-representations of the tangent space $T_p\mathcal{S}$. The goal is to find isomorphisms between the abstract tangent space at a point in $\mathcal{S}$ and some concrete vector spaces of random variables. It is deeply rooted in the very origins of information geometry, where already in the pioneering work of Rao [51], a basis vector of the form $\frac{\partial}{\partial\theta^i}\big|_p$ is mapped to the random variable $\frac{\partial\log p}{\partial\theta^i}$ in order to compute the Fisher metric

$$g_{ij} = E_p\left[\frac{\partial\log p}{\partial\theta^i}\frac{\partial\log p}{\partial\theta^j}\right] \tag{2.49}$$

In his earlier book [1], Amari achieves these isomorphisms by hand, in a coordinate based approach. For example, he proves that if we take the linear span of $\left\{\frac{\partial\log p}{\partial\theta^i}\right\}$, for $i = 1, \dots, n$ (which are supposed to be linearly independent to begin with, a result that we aim to prove), we obtain a vector space in which the components of the vectors transform according to the same rule as the components of the tangent vectors spanned by $\left\{\frac{\partial}{\partial\theta^1}\big|_p, \dots \frac{\partial}{\partial\theta^n}\big|_p\right\}$ when we change from one coordinate system to another. The same sort of argument is carried over for all the other α-representations.

In his more recent book with Nagaoka [2], the α-representations are briefly introduced in equation (2.59), followed shortly by the comment that the α-connections can be seen as being induced from the affine structure of the space of *all* functions on the sample space Ω. This is more intrinsic than the previous construction and is perfectly acceptable for finite sample spaces, where the topological and measurable structures of the set of all random variables are trivial. In dealing with general sample spaces however, the target spaces for the α-embeddings ought to be more carefully restricted.

The affine structure of the space of all measures equivalent to a given one is analysed in some detailed by Murray and Rice [38], and it is then used to defined the 1-representation of tangent vectors. They go through all this effort because they do not have the concept of the Pistone-Sempi manifold to use.

Our own view on the subject is that, for $\alpha \in [-1,1)$, the α-representations are a consequence of the α-embeddings, but into appropriate L^p-spaces, whereas the 1-representation is a special case of the general identification between tangent vectors and elements in B_p done previously. The details are as follows.

Let $v \in T_p\mathcal{S}$. Then v is tangent to a curve $p(t) \in \mathcal{M}$ and is therefore associated with the element $\frac{d}{dt} \log \frac{p(t)}{p}\Big|_{t=0}$ in B_p. In the case where $v = \frac{\partial}{\partial\theta^i}\big|_p$, the curve is of the form $p(\theta^1, \ldots, \theta^i + t, \ldots, \theta^n)$ and v is identified with the random variable $\frac{\partial \log p}{\partial\theta^i}$. Since the vectors $\left\{\frac{\partial}{\partial\theta^1}, \ldots, \frac{\partial}{\partial\theta^n}\right\}$ are a basis for $T_p\mathcal{S}$ and the correspondence is linear, we find that $\left\{\frac{\partial \log p}{\partial\theta^1}, \ldots, \frac{\partial \log p}{\partial\theta^n}\right\}$ are linearly independent and span a n-dimensional subspace of B_p, which is the 1-representation of $T_p\mathcal{S}$.

For $\alpha \in [-1,1)$, the generalised Amari embedding

$$\begin{aligned} \ell_\alpha : \mathcal{M} &\rightarrow L^r(\mu) \\ p &\mapsto \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}, \quad \alpha \in [-1,1), \end{aligned} \tag{2.50}$$

where $r = \frac{2}{1-\alpha}$, maps $\mathcal{M}$ into vector spaces, whose tangent vectors are obtained by differentiating curves on the vector space itself. The α-representations are then simply the push-forwards of the α-embeddings. If $v \in T_p\mathcal{S}$ and $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{S}$

with $\gamma(0) = p$ is a curve in the equivalence class of the tangent vector v, then the α-representation of v is given by $(\ell_\alpha \circ \gamma)'(0)$. In particular, if $v = \frac{\partial}{\partial \theta^i}\big|_p$, then its α-representation is

$$\frac{\partial \ell_\alpha(p)}{\partial \theta^i}. \tag{2.51}$$

But since

$$\frac{d}{dt}\left(\frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}\right) = p^{\frac{1-\alpha}{2}} \frac{d \log p}{dt},$$

it can be written in the more familiar form

$$p^{\frac{1-\alpha}{2}} \frac{\partial \log p}{\partial \theta^i}. \tag{2.52}$$

With these concepts in mind, we can now show that the definitions and constructions of the previous sections reproduce the parametric results when restricted to finite dimensional submanifolds.

For if $\{\theta^1, \ldots, \theta^n\}$ is a coordinate system in a finite dimensional submanifold of $\mathcal{M}$ we can put $v = \frac{\partial}{\partial \theta^i}$ and $s = \frac{\partial \log p}{\partial \theta^j}$ as the 1-representation of the vector field $\frac{\partial}{\partial \theta^j}$, then (2.29) reduces to

$$\left(\nabla^{(1)}_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^j}\right)(p) = \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} - E_p\left(\frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j}\right), \tag{2.53}$$

which is the classical finite dimensional result for the exponential connection. Accordingly, parametric exponential models are $\nabla^{(1)}$-flat submanifolds of $\mathcal{M}$.

For the case of the mixture connection, we find that putting $v = \frac{\partial}{\partial \theta^i}$ and $s = \frac{\partial \log p}{\partial \theta^j}$ in (2.31) gives

$$\left(\nabla^{(-1)}_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^j}\right)(p) = \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} + \frac{\partial \log p}{\partial \theta^i}\frac{\partial \log p}{\partial \theta^j}, \tag{2.54}$$

which is again the parametric result. Mixture families, which according to propositon 2.4.2 are well defined in $\mathcal{M}$, are then $\nabla^{(-1)}$-flat submanifolds of $\mathcal{M}$.

Chapter 3

Parametric Quantum Systems

The first task of quantum information geometry is to extend to non-commutative probability spaces the uniqueness and rigidity of the geometrical structures used in the classical version of the theory. This includes, for instance, the uniqueness result of Chentsov concerning the Fisher metric [8] and the two equivalent definitions of α-connections given by Amari [1]: either using the α-embeddings introduced in the previous chapter or as the convex mixture of the exponential and mixture connections.

As for the first problem, Chentsov himself made the first attempt to find the possible Riemannian metrics on a quantum information manifold with the property of having its line element reduced under stochastic maps [37] (cf [43]). This monotonicity property, which is the quantum analogue of being reduced under Markov morphisms, was later investigated by Petz [43]. Unlike the classical case, he found that there are infinitely many Riemannian metrics satisfying it. Concerning the second problem, several definitions of α-connections have been proposed [40, 22, 13], both for finite and infinite dimensional quantum systems. Since some of these definitions involve finding dual connections with respect to some chosen Fisher metric, it is clear that the multitude of possible candidates for the metric encourages the appearance of non-equivalent definitions for the α-connections.

We take the position that quantum information manifolds are equipped with two natural flat connections: the mixture connection, obtained from the linear structure of trace class operators themselves, and the exponential connection, obtained when combinations of states are performed by adding their logarithms [18, 64]. In section 3.1, we present our preferred definition of the ± 1-connections in finite dimensional quantum information manifolds. It is inspired by the definition given in the preceding chapter for classical systems, but we take advantage of the fact that, in this chapter, we will be working with finite dimensional spaces to define them in a slightly more concrete manner.

Following Amari [1, 2], we consider duality as a fundamental structure to be explored. Thus, given these two connections, we should ask what are the Riemannian metrics that make them dual. We give our answer to this question in section 3.2.

In section 3.3, we combine this result with Petz's characterisation of monotone metrics in finite dimensional quantum systems to find that the *BKM* metric [33, 36] is, up to a factor, the unique monotone Riemannian metric with respect to which the exponential and mixture connections are dual. These results have been described in two of our previous publications [20, 21].

Towards the end of the chapter we present some physically motivated arguments to justify the requirement of both monotonicity and duality upon the metrics on $\mathcal{S}$. We also comment on an open problem regarding the quantum analogue of the α-connections.

3.1 The Exponential and Mixture Connections

Let $\mathcal{H}^N$ be a finite dimensional complex Hilbert space, $\mathcal{A}$ the subspace of self-adjoint operators and $\mathcal{S}$ the set of all invertible density operators on $\mathcal{H}^N$. Then $\mathcal{A}$ is an N^2-dimensional real vector space and $\mathcal{S}$ is an n-dimensional submanifold with $n = N^2 - 1$. Defining the 1-embedding of $\mathcal{S}$ into $\mathcal{A}$ as

$$\begin{aligned} \ell_1 : \mathcal{S} &\rightarrow \mathcal{A} \\ \rho &\mapsto \log\rho, \end{aligned} \tag{3.1}$$

we can use the linear space structure of $\mathcal{A}$ to obtain a representation of the tangent bundle of $\mathcal{S}$ in terms of operators in $\mathcal{A}$ even though $\mathcal{S}$ is not a vector space itself. At each point $\rho \in \mathcal{S}$, consider the subspace $\mathcal{A}_\rho = \{A \in \mathcal{A} : \mathrm{Tr}(\rho A) = 0\}$ of $\mathcal{A}$. We define the isomorphism

$$\begin{aligned} (\ell_1)_{*(\rho)} : T_\rho\mathcal{S} &\rightarrow \mathcal{A}_\rho \\ v &\mapsto (\ell_1 \circ \gamma)'(0), \end{aligned} \tag{3.2}$$

where $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{S}$ is a curve in the equivalence class of the tangent vector v. We call this isomorphism the 1-representation of the tangent space $T_\rho\mathcal{S}$. If $(\theta^1, \ldots, \theta^n)$ is a coordinate system for $\mathcal{S}$, then the 1-representation of the basis $\left\{ \frac{\partial}{\partial\theta^1}\Big|_\rho, \ldots, \frac{\partial}{\partial\theta^n}\Big|_\rho \right\}$ of $T_\rho\mathcal{S}$ is $\left\{ \frac{\partial \log\rho}{\partial\theta^1}, \ldots, \frac{\partial\log\rho}{\partial\theta^n} \right\}$. The 1-representation of a vector field X on $\mathcal{S}$ is therefore the $\mathcal{A}$-valued function $(X)^{(1)}$ given by $(X)^{(1)}(\rho) = (\ell_1)_{*(\rho)}X_\rho$, which we sometimes simply denote by X^+.

The exponential or 1-connection is the connection obtained from the 1-embedding through the following parallel transport [64]

$$\begin{aligned} \tau^{(1)}_{\rho_0,\rho_1} : T_{\rho_0}\mathcal{S} &\rightarrow T_{\rho_1}\mathcal{S} \\ v &\mapsto (\ell_1)^{-1}_{*(\rho_1)} \left((\ell_1)_{*(\rho_0)}v - \mathrm{Tr}[\rho_1(\ell_1)_{*(\rho_0)}v] \right). \end{aligned} \tag{3.3}$$

Giving the parallel transport in a neighbourhood of ρ is equivalent to specifying the covariant derivative. It is readily verified that the 1-representation of the 1-covariant derivative, applied to the vector field $\frac{\partial}{\partial\theta^j}$, is

$$\left(\nabla^{(1)}_{\partial_i} \frac{\partial}{\partial\theta^j} \right)^{(1)} = \frac{\partial^2 \log\rho}{\partial\theta^i\partial\theta^j} - \mathrm{Tr}\left(\rho \frac{\partial^2\log\rho}{\partial\theta^i\partial\theta^j} \right). \tag{3.4}$$

At a more abstract level, the construction above corresponds to making $\mathcal{S}$ into an affine space and endowing it with the *natural* flat connection [38]. Rather than exploring this concept further, we benefit from dealing with finite dimensional spaces and prove that the 1-connection is flat by exhibiting an affine coordinate system for it. Let $\{\mathbf{1}, X_1, \ldots, X_n\}$ be a basis for $\mathcal{A}$. Since $\log\rho \in \mathcal{A}$, there exist real numbers $\{\theta^1, \ldots, \theta^n, \Psi\}$ such that

$$\log\rho = \theta^1 X_1 + \cdots + \theta^n X_n - \Psi\mathbf{1},$$

that is,

$$\rho = \exp\left(\theta^1 X_1 + \cdots + \theta^n X_n - \Psi \mathbf{1}\right). \tag{3.5}$$

The normalisation condition $\mathrm{Tr}\rho = 1$ means, however, that only n among these numbers are independent, so we choose $\Psi \equiv \Psi(\theta)$ to be the one determined by the others. Then $\theta = (\theta^1, \ldots, \theta^n)$ form a 1-affine coordinate system, as can be seen from the following calculation

$$\begin{aligned}
\left(\nabla^{(1)}_{\partial_i} \frac{\partial}{\partial\theta^j}\right)^{(1)} &= \frac{\partial^2 \log\rho}{\partial\theta^i \partial\theta^j} - \mathrm{Tr}\left(\rho \frac{\partial^2 \log\rho}{\partial\theta^i \partial\theta^j}\right) \\
&= -\frac{\partial^2 \Psi}{\partial\theta^i \partial\theta^j}(\theta) + \mathrm{Tr}\left(\rho \frac{\partial^2 \Psi}{\partial\theta^i \partial\theta^j}(\theta)\right) \\
&= 0.
\end{aligned}$$

Now let $\mathcal{A}_0$ be the subspace of traceless operators in $\mathcal{A}$. Consider the -1-embedding

$$\begin{aligned}
\ell_{-1} : \mathcal{S} &\rightarrow \mathcal{A} \\
\rho &\mapsto \rho,
\end{aligned} \tag{3.6}$$

and define, at each $\rho \in \mathcal{S}$, the -1-representation of tangent vectors as

$$\begin{aligned}
(\ell_{-1})_{*(\rho)} : T_\rho\mathcal{S} &\rightarrow \mathcal{A}_0 \\
v &\mapsto (\ell_{-1} \circ \gamma)'(0),
\end{aligned} \tag{3.7}$$

where $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{S}$ is again a curve in the equivalence class of the tangent vector v. In coordinates, the -1-representation of the basis $\left\{ \frac{\partial}{\partial\theta^1}\big|_\rho, \ldots, \frac{\partial}{\partial\theta^n}\big|_\rho \right\}$ of $T_\rho\mathcal{S}$ is $\left\{\frac{\partial\rho}{\partial\theta^1}, \ldots, \frac{\partial\rho}{\partial\theta^n}\right\}$. As before, the -1-representation of a vector field X on $\mathcal{S}$ is an $\mathcal{A}_0$-valued function denoted by $(X)^{(-1)}$ or X^-.

We obtain the mixture or -1-connection by defining the parallel transport

$$\begin{aligned}
\tau^{(-1)}_{\rho_0,\rho_1} : T_{\rho_0}\mathcal{S} &\rightarrow T_{\rho_1}\mathcal{S} \\
v &\mapsto (\ell_{-1})^{-1}_{*(\rho_1)}\left((\ell_{-1})_{*(\rho_0)} v\right),
\end{aligned} \tag{3.8}$$

and we find that its covariant derivative in the direction ∂_i is

$$\left(\nabla^{(-1)}_{\partial_i} \frac{\partial}{\partial\theta^j}\right)^{(-1)} = \frac{\partial^2 \rho}{\partial\theta^i \partial\theta^j}. \tag{3.9}$$

If we equip $\mathcal{A}$ with the trace norm, then the -1-embedding maps $\mathcal{S}$ into the unit sphere $\mathcal{S}^1$ of $\mathcal{A}$, and the -1-connection given here is nothing but the projection onto $\mathcal{S}^1$ of the *natural* flat connection in this space. It turns out that the unit sphere with respect to the trace norm is flat in $\mathcal{A}$, hence the -1-connection is flat on $\mathcal{S}$. Again, it is a convenience of the finite dimensional case that we can prove flatness of the -1-connection by direct construction of a -1-affine coordinate system. Suppose $\{X_1, \ldots, X_{n+1}\}$ is a normalised basis for $\mathcal{A}$, then there exist real numbers $(\xi_1, \ldots, \xi_{n+1})$ such that

$$\rho = \xi_1 X_1 + \cdots + \xi_{n+1} X_{n+1}. \tag{3.10}$$

Since $\mathrm{Tr}\rho = 1$, we can take $\xi_{n+1} = 1 - (\xi_1 + \cdots + \xi_n)$ as a function of the first n independent parameters $\xi_1, \ldots, \xi_n$. Then $\xi = (\xi_1, \ldots, \xi_n)$ is a -1-affine coordinate system for $\mathcal{S}$, because

$$\begin{aligned}
\left(\nabla^{(-1)}_{\frac{\partial}{\partial \xi_i}} \frac{\partial}{\partial \xi_j}\right)^{(-1)} &= \frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j} \\
&= \frac{\partial}{\partial \xi_i}(X_j - X_{n+1}) \\
&= 0.
\end{aligned}$$

3.2 Duality and the *BKM* Metric

The generalised concept of duality for connections on a infinite dimensional manifold was introduced in the previous chapter. For finite dimensional manifolds it reduces to the following well known definition. Two connections ∇ and ∇^* on a Riemannian manifold $(\mathcal{S}, g)$ are dual with respect to g if and only if

$$Xg(Y,Z) = g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla^*_X Z\right), \tag{3.11}$$

for any vector fields X, Y, Z on $\mathcal{S}$ [1, 38]. Equivalently, if $\tau_{\gamma(t)}$ and $\tau^*_{\gamma(t)}$ are the respective parallel transports along a curve $\{\gamma(t)\}_{0\leq t\leq 1}$ on $\mathcal{S}$, with $\gamma(0) = \rho$, then ∇ and ∇^* are dual with respect to g if and only if for all $t \in [0,1]$,

$$g_\rho(Y,Z) = g_{\gamma(t)}\left(\tau_{\gamma(t)} Y, \tau^*_{\gamma(t)} Z\right). \tag{3.12}$$

Given any connection ∇ on $(\mathcal{S}, g)$, we can always find a unique connection ∇^* such that ∇ and ∇^* are dual with respect to g. On the other hand, given two connections ∇ and ∇^*, we can ask what are the possible Riemannian metrics g with respect to which they are dual. In particular, we want to explore this question for the case of the exponential and mixture connections on a manifold of density matrices.

A different concept of duality also used by Amari [1] is that of dual coordinate systems, regardless of any connection. Two coordinate systems $\theta = (\theta^i)$ and $\eta = (\eta_i)$ on a Riemannian manifold $(\mathcal{S}, g)$ are dual with respect to g if and only if their natural bases for $T_\rho \mathcal{S}$ are *biorthogonal* at every point $\rho \in \mathcal{S}$, that is,

$$g\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = \delta^i_j. \tag{3.13}$$

Equivalently, $\theta = (\theta^i)$ and $\eta = (\eta_i)$ are dual with respect to g if and only if

$$g_{ij} = \frac{\partial \eta_i}{\partial \theta^j} \quad \text{and} \quad g^{ij} = \frac{\partial \theta_i}{\partial \eta^j} \tag{3.14}$$

at every point $\rho \in \mathcal{S}$, where, as usual, $g^{ij} = (g_{ij})^{-1}$.

The following theorem [1, theorem 3.4] gives a characterisation of dual coordinate systems in terms of potential functions, thus bringing convexity theory and the related duality with respect to Legendre transforms into the discussion.

Theorem 3.2.1 *When a Riemannian manifold $(\mathcal{S}, g)$ has a pair of dual coordinate systems (θ, η), there exist potential functions $\Psi(\theta)$ and $\Phi(\eta)$ such that*

$$g_{ij}(\theta) = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j} \quad \textit{and} \quad g^{ij} = \frac{\partial^2 \Phi(\eta)}{\partial \eta_i \partial \eta_j}. \tag{3.15}$$

Conversely, when either potential function Ψ or Φ exists from which the metric is derived by differentiating it twice, there exist a pair of dual coordinate systems. The dual coordinate systems and the potential functions are related by the following Legendre transforms

$$\theta^i = \frac{\partial \Phi(\eta)}{\partial \eta_i}, \quad \eta_i = \frac{\partial \Psi(\theta)}{\partial \theta^i} \tag{3.16}$$

and

$$\Psi(\theta) + \Phi(\eta) - \theta^i \eta_i = 0 \tag{3.17}$$

In contrast to the case of dual connections, dual coordinate systems do not necessarily exist on every Riemannian manifold [1]. When the additional property of flatness is required, the following theorem [1, theorem 3.5] provides a link between the two concepts of duality. In the sense used in this thesis, a connection ∇ on manifold $\mathcal{S}$ is said to be flat if $\mathcal{S}$ admits a global ∇-affine coordinate system. This is equivalent to its curvature and torsion both being zero.

Theorem 3.2.2 *Suppose that ∇ and ∇^* are two flat connections on a manifold $\mathcal{S}$. If they are dual with respect to a Riemannian metric g on $\mathcal{S}$, then there exists a pair (θ, η) of dual coordinate systems such that θ is ∇-affine and η is a ∇^*-affine.*

We now return to our manifold $\mathcal{S}$ of density matrices and consider the problem of finding a unique Riemannian metric for it. Using either of the ± 1-representations of the tangent bundle $T\mathcal{S}$, we define a Riemannian metric on $\mathcal{S}$ by a smooth assignment of an inner product $\langle \cdot, \cdot \rangle_\rho$ in $\mathcal{A} \subset B(\mathcal{H}^N)$ for each point $\rho \in \mathcal{S}$. The *BKM* (Bogoliubov-Kubo-Mori) metric is the Riemannian metric on $\mathcal{S}$ obtained from the *BKM* inner product [42]. If $A^{(1)}, B^{(1)}$ and $A^{(-1)}, B^{(-1)}$ are, respectively, the 1 and -1-representations of $A, B \in T_\rho\mathcal{S}$, then the *BKM* metric has the following equivalent expressions:

$$\begin{aligned} g^B_\rho(A,B) &= \mathrm{Tr}\left(A^{(-1)}B^{(1)}\right) \\ &= \int_0^1 \mathrm{Tr}\left(\rho^\lambda A^{(1)} \rho^{1-\lambda} B^{(1)}\right) d\lambda \\ &= \int_0^\infty \mathrm{Tr}\left(\frac{1}{t+\rho} A^{(-1)} \frac{1}{t+\rho} B^{(-1)}\right) dt \end{aligned} \tag{3.18}$$

It is well known that the exponential and mixture connections are dual with respect to the *BKM* metric [40, 23]. A natural question is whether it is the unique metric with this property. The next theorem tells us how much uniqueness can be achieved from duality alone.

Theorem 3.2.3 *If the connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ are dual with respect to a Riemannian metric g on $\mathcal{S}$, then there exist a constant (independent of ρ) $n \times n$ matrix M, such that $(g_\rho)_{ij} = \sum_{k=1}^{n} M_i{}^k (g_\rho^B)_{kj}$, in some 1-affine coordinate system.*

Proof: Since the two connections are flat, by theorem 3.2.2, there exist dual coordinate systems (θ, η) such that θ is $\nabla^{(1)}$-affine and η is $\nabla^{(-1)}$-affine. Thus, applying theorem 3.2.1, there exist a potential function $\Psi(\theta)$ such that

$$g_{ij}(\theta) = \frac{\partial^2 \Psi(\theta)}{\partial \theta^i \partial \theta^j}$$

and

$$\eta_i = \frac{\partial \Psi(\theta)}{\partial \theta^i}.$$

On the other hand, since θ is $\nabla^{(1)}$-affine, there exist linearly independent operators $\{\mathbf{1}, X_1, \ldots, X_n\}$ such that

$$\rho = \exp\left(\theta^1 X_1 + \cdots + \theta^n X_n - \tilde{\Psi}(\theta)\mathbf{1}\right), \tag{3.19}$$

where

$$\tilde{\Psi}(\theta) = \log \mathrm{Tr}\left[\exp\left(\theta^1 X_1 + \cdots + \theta^n X_n\right)\right]. \tag{3.20}$$

Without loss of generality, we can assume that the operators $X_1, \ldots, X_n$ are traceless, since if we add multiples of the identity to any X_j in (3.19), we can still have the same parameters θ as coordinates for the same point ρ just by modifying the function $\tilde{\Psi}$. But any such set of operators define a $\nabla^{(-1)}$-affine coordinate system through the formula

$$\tilde{\eta}_i = \mathrm{Tr}(\rho X_i), \tag{3.21}$$

because the latter are affinely related to the ξ coordinates defined in section 3.1 (with $X_{n+1} = \mathbf{1}/n$). Differentiating $\tilde{\Psi}$ with respect to θ^i we obtain

$$\frac{\partial \tilde{\Psi}(\theta)}{\partial \theta^i} = \mathrm{Tr}(\rho X_i) = \tilde{\eta}_i.$$

Thus $\tilde{\eta}_i = \frac{\partial\tilde{\Psi}(\theta)}{\partial\theta^i}$ and $\eta_i = \frac{\partial\Psi(\theta)}{\partial\theta^i}$ are two $\nabla^{(-1)}$-affine coordinate systems, so they must be related by an affine transformation [38]. So there exist an $n \times n$ matrix M and numbers $(a_1, \dots, a_n)$ such that

$$\eta_i = \sum_{k=1}^{n} M_i{}^k \tilde{\eta}_k + a_i, \tag{3.22}$$

that is,

$$\frac{\partial\Psi(\theta)}{\partial\theta^i} = \sum_{k=1}^{n} M_i{}^k \frac{\partial\Psi(\theta)}{\partial\theta^k} + a_i, \tag{3.23}$$

and differentiating this equation with respect to θ^j gives

$$g_{ij}(\theta) = \frac{\partial^2\Psi(\theta)}{\partial\theta^i\partial\theta^j} = \sum_{k=1}^{n} M_i{}^k \frac{\partial^2\tilde{\Psi}(\theta)}{\partial\theta^j\partial\theta^k}. \tag{3.24}$$

But we can calculate the second derivative of $\tilde{\Psi}$ directly from (3.20), obtaining

$$\begin{aligned}
\frac{\partial^2\tilde{\Psi}(\theta)}{\partial\theta^j\partial\theta^k} &= \int_0^1 \mathrm{Tr}\left(\rho^\lambda \frac{\partial \log\rho}{\partial\theta^j} \rho^{1-\lambda} \frac{\partial\log\rho}{\partial\theta^k}\right) d\lambda \\
&= g_\rho^{B}\left(\frac{\partial}{\partial\theta^j}, \frac{\partial}{\partial\theta^k}\right) \\
&= g_{jk}^{B}(\theta).
\end{aligned}$$

Inserting this in (3.24) completes the proof.

3.3 The Condition of Monotonicity

We have seen in the previous section that requiring duality between the exponential and mixture connections reduces the set of possible Riemannian metrics on $\mathcal{S}$ to matrix multiples of the *BKM* metric. We now investigate the effect of imposing a monotonicity property on this set.

If we use the -1-representation to define a Riemannian metric g on $\mathcal{S}$ by means of the inner product $\langle\cdot,\cdot\rangle_\rho$ in $\mathcal{A} \subset B(\mathcal{H}^N)$, then we say that g is *monotone* if and only if

$$\left\langle P(A^{(-1)}), P(A^{(-1)})\right\rangle_{P(\rho)} \leq \left\langle A^{(-1)}, A^{(-1)}\right\rangle_\rho, \tag{3.25}$$

for every $\rho \in \mathcal{S}$, $A \in T_\rho\mathcal{S}$, and every trace preserving, completely positive linear map $P : \mathcal{A} \to \mathcal{A}_m$, for all integers n, m, where $\mathcal{A}_m$ is a matrix space of arbitrary dimension m.

In a series of papers [41, 43, 45], Petz has given a complete characterisation of monotone metrics on full matrix spaces in terms of operator monotone functions. The monotonicity condition (3.25), however, is defined for metrics on the space $\mathcal{S}$ of faithful *states*, and must first be extended to $\mathcal{A}$ before we can use Petz's results. Let $\widehat{\mathcal{S}}$ be the manifold of faithful weights (the positive definite matrices). Let g be a metric on $T\mathcal{S}$. We can extend g to $\mathcal{A} \simeq T\widehat{\mathcal{S}}$ as follows. At $\rho \in \mathcal{S}$ and $\hat{A}, \hat{B} \in T\widehat{\mathcal{S}}_\rho$, put $\hat{A}^{(-1)} = A_0\rho + A^-$, where $A_0 = \operatorname{Tr} \hat{A}^{(-1)} \in \mathbb{R}$ and $\operatorname{Tr} A^- = 0$, and similarly for B. Then put

$$\hat{g}_\rho(\hat{A}, \hat{B}) := \alpha A_0 B_0 + g_\rho(A^-, B^-), \tag{3.26}$$

where $\alpha > 0$ is arbitrary. For $\hat{g}^B$, with $\alpha = 1$, this extension acquires the same form as in equation (3.18). More generally, if g is monotone on $T\mathcal{S}$, then $\hat{g}$ is monotone on $T\widehat{\mathcal{S}}$. For, let P be a trace preserving, completely positive linear map on $T\widehat{\mathcal{S}}$, and $\rho \in \mathcal{S}$. Then P maps $T\mathcal{S}$ to itself, and

$$\hat{g}_{P\rho}(P\hat{A}, P\hat{A}) = A_0^2 + g_{P\rho}(PA^-, PA^-) \leq A_0^2 + g_\rho(A^-, A^-) = \hat{g}_\rho(A, A).$$

For any metric $\hat{g}$ on $T\widehat{\mathcal{S}}$, and putting $\hat{A}^{(-1)} = A_0\rho + A^-$, we define the positive (super) operator K_ρ on $\mathcal{A}$ by

$$\hat{g}_\rho(\hat{A}, \hat{B}) = \left\langle \hat{A}^{(-1)}, K_\rho\left(\hat{B}^{(-1)}\right)\right\rangle_{HS} = \operatorname{Tr}\left(\hat{A}^{(-1)} K_\rho\left(\hat{B}^{(-1)}\right)\right). \tag{3.27}$$

Note that our K is denoted K^{-1} by Petz. He also defines the (super) operators, $L_\rho X := \rho X$ and $R_\rho X := X\rho$, for $X \in \mathcal{A}$, which are also positive. Then he proved the following [43].

Theorem 3.3.1 (Petz) *A Riemannian metric g on $\mathcal{A}$ is monotone if and only if*

$$K_\rho = \left(R_\rho^{1/2} f(L_\rho R_\rho^{-1}) R_\rho^{1/2}\right)^{-1}, \tag{3.28}$$

where K_ρ is defined in (3.27) and $f : R^+ \to R^+$ is an operator monotone function satisfying $f(t) = tf(t^{-1})$.

In particular, the *BKM* metric is monotone and its corresponding operator monotone function is

$$f^{B}(t) = \frac{t-1}{\log t}. \tag{3.29}$$

Combining this characterisation with our theorem (3.2.3), we obtain the following improved uniqueness result.

Theorem 3.3.2 *If the connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ are dual with respect to a monotone Riemannian metric g on $\mathcal{S}$, then g is a* scalar *multiple of the BKM metric.*

Proof: Let K_ρ^g and K_ρ^B be the (super)operators associated with the monotone metrics $\hat{g}$ and $\hat{g}^B$ as in equation (3.27). We want to translate the result of theorem 3.2.3 in terms of these operators. We first extend the $(N^2-1)\times(N^2-1)$ matrix M by one row and column, to a matrix $\hat{M}$, where $\hat{M}_{ij} = M_{ij}$, $(1 \leq i, j \leq N^2-1)$, and

$$\hat{M}_{i,N^2} = \hat{M}_{N^2,i} = \alpha\delta_{i,N^2}, (i = 1, \dots, N^2). \tag{3.30}$$

We claim that K^g and K^B satisfy the following relation:

$$K_\rho^g = \hat{M}^t K_\rho^B. \tag{3.31}$$

To verify the claim, we want to show that the corresponding scalar product relation

$$\left\langle \hat{A}^{(-1)}, K_\rho^g \hat{B}^{(-1)} \right\rangle_{HS} = \left\langle \hat{A}^{(-1)}, \hat{M}^t K_\rho^B \hat{B}^{(-1)} \right\rangle_{HS} \tag{3.32}$$

holds for all $\hat{A}^{(-1)}, \hat{B}^{(-1)} \in \mathcal{A}$. Now, in the representation $\hat{A}^{(-1)} = A_0\rho + A^-$ of equation (3.26), the extension of g to $\hat{g}$ in the orthogonal direction ρ is the same for all g. In particular,

$$\hat{g}(A_0\rho, B^-) = \hat{g}^B(A_0\rho, B^-) = 0 \tag{3.33}$$

and

$$\hat{g}(A_0\rho, B_0\rho) = \alpha\hat{g}^B(A_0\rho, B_0\rho) = \alpha A_0 B_0. \tag{3.34}$$

So we only need to prove (3.32) for traceless operators, which are isomorphic to $T_\rho\mathcal{S}$ through the -1-representation.

Now if $(\theta^1, \ldots, \theta^n)$ is a coordinate system for $\mathcal{S}$, then, from theorem 3.2.3, we have

$$g_\rho\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\theta^j}\right) = \sum_{k=1}^{n} {M_i}^k g_\rho^{B}\left(\frac{\partial}{\partial\theta^k}, \frac{\partial}{\partial\theta^j}\right), \tag{3.35}$$

which, when expressed in the -1-affine coordinates (by taking inverses and using theorem 3.2.1), gives

$$g_\rho\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right) = \sum_{k=1}^{n} g_\rho^{B}\left(\frac{\partial}{\partial\tilde{\eta}_j}, \frac{\partial}{\partial\tilde{\eta}_k}\right)({M_i}^k)^{-1}, \tag{3.36}$$

where η and $\tilde{\eta}$ are the same as in theorem 3.2.3. In particular they are related through equation (3.22), which means that

$$\frac{\partial}{\partial\tilde{\eta}_k} = \sum_{l=1}^{n} \frac{\partial\eta_l}{\partial\tilde{\eta}_k}\frac{\partial}{\partial\eta_l} = \sum_{l=1}^{n} {M_l}^k \frac{\partial}{\partial\eta_l}. \tag{3.37}$$

Inserting this into (3.36) gives

$$\begin{aligned}
g_\rho\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_j}\right) &= \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} {M_l}^k {M_m}^j g_\rho^{B}\left(\frac{\partial}{\partial\eta_l}, \frac{\partial}{\partial\eta_m}\right)({M_i}^k)^{-1} \\
&= \sum_{l=1}^{n}\sum_{m=1}^{n} {\delta_l}^i {M_m}^j g_\rho^{B}\left(\frac{\partial}{\partial\eta_l}, \frac{\partial}{\partial\eta_m}\right) \\
&= \sum_{m=1}^{n} {M_m}^j g_\rho^{B}\left(\frac{\partial}{\partial\eta_i}, \frac{\partial}{\partial\eta_m}\right),
\end{aligned}$$

which proves the claim.

Therefore, if f^g and f^B are the operator monotone functions corresponding respectively to g and g^B, from theorem 3.3.1, we have

$$\begin{aligned}
\left(R_\rho^{1/2} f^g(L_\rho R_\rho^{-1}) R_\rho^{1/2}\right)^{-1} &= \hat{M}^t \left(R_\rho^{1/2} f^B(L_\rho R_\rho^{-1}) R_\rho^{1/2}\right)^{-1} \\
\left(R_\rho^{1/2} f^g(L_\rho R_\rho^{-1}) R_\rho^{1/2}\right)\hat{M}^t &= \left(R_\rho^{1/2} f^B(L_\rho R_\rho^{-1}) R_\rho^{1/2}\right) \\
\hat{M}^t &= f^g(L_\rho R_\rho^{-1})^{-1} f^B(L_\rho R_\rho^{-1}),
\end{aligned}$$

as everything commutes. Thus, the operator $\hat{M}$ is given as a function of the operator $L_\rho R_\rho^{-1}$, but it is itself independent of the point ρ, so we conclude that it must be a scalar multiple of the identity operator.

3.4 The Hasegawa-Petz Duality

As part of the programme of characterising monotone metrics by means of operator monotone functions, Hasegawa and Petz [24] studied the properties of a family of metrics called *WYD* (Wigner-Yanase-Dyson) metrics. They proved that these metrics are monotone by explicitly finding that the operator monotone functions associated with them as in theorem 3.3.1 are

$$f_p(x) = \frac{p(1-p)(x-1)^2}{(x^p-1)(x^{1-p}-1)}, \tag{3.38}$$

for $-1 < p < 1$. They then proceeded to investigate these metrics in terms of *duality*. To fully compare their result with ours, we need to quote from their paper a bit more lengthily.

Since all monotone metrics have the same form on the subspace

$$\mathcal{C}_\rho = \{X \in \mathcal{A} : \rho X - X\rho = 0\} \tag{3.39}$$

(they all reduce to the Fisher metric of classical probability) one only needs to look for different forms on the complement of $\mathcal{C}_\rho$. Consider the (super)operator

$$\begin{aligned} \mathcal{L}_\rho \;\; &: \;\; \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}) \\ & \quad\;\; X \mapsto i[\rho, X]. \end{aligned}$$

Notice that the range of $\mathcal{L}_\rho$ is invariant under the replacement of ρ by $h(\rho)$, where h is a monotone function. Hasegawa and Petz then say that a Riemannian $\hat{g}$ metric on the full matrix space *admits* dual affine connections if, for a pair of monotone functions h and h^* we have

$$\hat{g}_\rho(\mathcal{L}_\rho(X), \mathcal{L}_\rho(Y)) = \langle \mathcal{L}_{h(\rho)}(X), \mathcal{L}_{h^*(\rho)}(Y)\rangle_{HS}. \tag{3.40}$$

This implies that (using their notation)

$$\delta_C \hat{g}_\rho(A, B) = \langle \delta_C\delta_A h(\rho), \delta_B h^*(\rho)\rangle_{HS} + \langle \delta_A h(\rho), \delta_C\delta_B h^*(\rho)\rangle_{HS}, \tag{3.41}$$

for $A = i[\rho, X]$ and $B = i[\rho, Y]$, which they claim to be "the quantum version of Amari's duality concept for affine connections". If one further imposes that $\hat{g}_\rho$

should be monotone, then condition (3.40) leads to the conclusion that $h(x) = ax^p$ and $h^*(x) = bx^{1-p}$, for $ab = c = 1/p(1-p)$ and with the limit $p \to 0$ or 1 giving x and $\log x$. Since with these functions one recovers the *WYD* metrics, their theorem has then been proved:

Theorem 3.4.1 (Hasegawa/Petz, 1997) *In the class of symmetric monotone metrics, the* WYD *skew information is characterised by the property that it admits dual affine connection (in the sense of equation (3.40)).*

Since the *WYD* metrics include the *BKM* metric as a special case, we need to explain the difference between this result and our main result in this chapter (theorem 3.3.2). First of all, Amari's concept of duality for affine connections is a purely geometrical concept, valid for any Riemannian manifold, regardless of whether its points are classical probability densities or density matrices. As we mentioned earlier in this chapter, we can always find the dual connection to a given one with respect to any metric. Therefore *all* metrics admit dual affine connections. Equation (3.41) is a special instance of Amari's duality when the connections are obtained through embeddings into affine spaces. This is indeed the case for the ± 1- connections, which do satisfy equation (3.41) if the metric is the *BKM* metric. However, there could conceivably exist other monotone metrics with respect to which the ± 1-connections would be Amari dual (therefore satisfying (3.41)) but not Hasegawa-Petz dual (not satisfying (3.40)). That this is not the case is the content of our result.

3.5 The Case for Duality

The monotonicity condition (3.25) has an appealing motivation coming from estimation theory. If we interpret the geodesic distance between two density matrices as a measure of their statistical distinguishability, then (3.25) tells us that they will become less distinguishable if we introduce randomness into the system under consideration. In other words, their distance decreases under coarse-graining.

As it is, estimation theory is more basic than physics itself, since it does not assume any particular underlying physical process, being just a tool to help analyse statistical data. Nevertheless, the interpretation above carries over to statistical mechanical systems as well, where stochastic (i.e completely positive, trace-preserving) maps appear as a mathematical implementation of the time evolution of a system whose states are described by density matrices [56]. In this case, monotonicity means that the distance between different states decreases under the same time evolution. If it decreases asymptotically to zero for any two points in a certain set of 'initial' states, then we are in the presence of a fixed point for the dynamics, or in other words, an equilibrium state. From all this, it seems that imposing a monotonicity condition on the possible Riemannian metrics on a statistical manifold is not at all an artificial technicality.

Our motivation behind Amari's duality for this chapter is less general and ultimately rests upon quantum statistical mechanics alone. In chapter 5, we are going to describe in more detail, for classical systems, a general framework for nonequlibrium statistical mechanics called *Statistical Dynamics* [55, 56], which makes use of the geometrical ideas discussed in this thesis. For now, let us recall that the von Neumann entropy for a state $\rho \in \mathcal{S}$ is defined as [69]

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho) \tag{3.42}$$

and that the relative (Kullback-Leibler) entropy of the state ρ given the state σ is

$$S(\rho|\sigma) = \mathrm{Tr}[\rho(\log \rho - \log \sigma)] \tag{3.43}$$

Now let us choose a set of $m \leq n$ observables $Y_1, \dots, Y_m$ such that $\{\mathbf{1}, Y_1, \dots, Y_m\}$ is a linearly independent subset of $\mathcal{A}$. Among all possible observables in $\mathcal{A}$, these ones represent the *slow variables* of the theory, that is, those whose means we can measure at any given time. Then it is an easy exercise, using the Lagrange multipliers technique, to show that the states which maximise the von Neumann entropy subject to keeping the means of all $\{Y_i\}$, $i = 1, \dots, m$, constant are the

Gibbs states of the form

$$\rho = \exp\left(\theta^1 Y_1 + \cdots + \theta^m Y_m - \Psi \mathbf{1}\right), \tag{3.44}$$

where $\Psi(\theta)$ is determined by the normalisation condition $\text{Tr}\rho = 1$. For example, if $Y_1 = H$ is the energy operator, then we obtain the so called canonical ensemble, whereas if we have $Y_1 = H, Y_2 = N$ where N is the number of particles, we get the grand canonical ensemble. We immediately recognise these states as constituting a $\nabla^{(1)}$-flat, m-dimensional, submanifold $\mathcal{S}_m \subset \mathcal{S}$, which is determined by our choice of $Y_1, \ldots, Y_m$, that is, by our choice of the level of description adopted.

Inasmuch as entropy is negative information, the principle of *maximum entropy*, advocated in information theory and statistical physics by Jaynes [26, 27], tells us that, if the only information available about the system under consideration are the means of the random variables $Y_1, \ldots, Y_m$, then we should take as the state of the system the element in $\mathcal{S}_m$ with these means. The replacement of the *true* state $\rho \in \mathcal{S}$ by the one in $\mathcal{S}_m$ with the same means for $Y_1, \ldots, Y_m$ is a reflection of our ignorance of what really goes on with the system. It is the least biased choice of state given the information available.

The point of view in statistical dynamics [56] is somewhat different, in the sense that it regards the same *replacement* as part of the true dynamics of the system. For instance, the heat transfer in a local region of a fluid happens 10^8 times faster then most chemical reactions [12], so we can choose to regard the concentrations of the chemicals reacting as the slow variables while all other observables are thermalised (maximum entropy) along each time step in the dynamics. The skill of the scientist using statistical dynamics thus resides in correctly identifying what are the slow variables of the problem at hands and then following the time evolution of the system, which involves, apart from a stochastic dynamics particular to each problem, sucessive *projections* onto $\mathcal{S}_m$.

Information geometry provides a mathematical meaning for this projection [5, 61]. It is well known that the relative entropy (3.43) is the statistical divergence associated with the dualistic triple $(g^B, \nabla^{(1)}, \nabla^{(-1)})$ [40]. It then follows from the general theory [2] that, given an arbitrary point $\rho \in \mathcal{S}$, the point in $\mathcal{S}_m$ (which is $\nabla^{(1)}$-flat) that minimises $S(\rho|\sigma)$ is obtained uniquely by following a -1-geodesic from ρ that intercepts $\mathcal{S}$ orthogonally with respect to the *BKM* metric g^B. This is equivalent to the projection described above (maximum entropy subject to constant means) precisely because a path preserving the mean parameters (or mixture coordinates) is a -1-geodesic, that is, a straight line for the mixture connection.

However, if g is a general monotone metric, with respect to which $\nabla^{(1)}$ and $\nabla^{(-1)}$ are not necessarily dual, then the relative entropy might fail to be a divergence for $(g, \nabla^{(1)}, \nabla^{(-1)})$ and nothing guarantees that minimising $S(\rho|\sigma)$ will produce a point in $\mathcal{S}_m$ connected to ρ by a -1-geodesic intersecting $\mathcal{S}_m$ perpendicularly with respect to g. Information geometry no longer provides a mathematical implementation for statistical dynamics anymore.

As a final word in this chapter, let us mention that, having fixed the metric, the next step in the development of the theory for finite quantum systems is to define the α-connections, for $\alpha \in (-1, 1)$. We look for a definition that makes $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ dual and such that the extended manifold $\widehat{\mathcal{S}}$ is α-flat. Only then we can try to find the quantum analogue of the α-divergence.

Duality is easily achieved if we define the α-connections as the convex mixture

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}, \tag{3.45}$$

provided that the extreme connections $\nabla^{(\pm 1)}$ are themselves dual.

Flatness, on the other hand, is more apparent if we use the α-embedding, which is the approach hinted by Hasegawa [22], favoured by Amari and Nagaoka [2] and put into more general context by Jenčová [29]. None of the above authors, however, has attempted to prove that the two definitions are equivalent. Therefore, the α-connections obtained using the α-embeddings are not known to be dual with respect to a fixed metric in the quantum case. The main technical difficulty, as far as our attempts are concerned, stems from the fact that the classical equivalence result is proved essentially as an application of the chain rule for derivatives, which does not hold in the quantum case: ρ and $\delta\rho$ do not necesarily commute.

Chapter 4

Infinite Dimensional Quantum Information Geometry

In this chapter we introduce the quantum version of the construction of a infinite dimensional statistical manifold. Instead of attempting to carry the construction in the full generality of quantum integration theory, that is, using noncommutative L^p-spaces as proposed in [13], we adopt a more concrete Hilbert space version. Then the analogue of the measure space Ω of chapter 2 is a Hilbert space $\mathcal{H}$, where its trace functional plays the role of the measure μ. Some of the results of this chapter have been previously reported in [18, 17].

The idea behind the construction is essentially the same as the one used in the classical case, namely that the points in the manifold are obtained by exponentiating from the Banach space chosen for generalised coordinates, so that the states in $\mathcal{M}$ are parametrised by their logarithms. As we have seen in chapter 2, the generalised parameters in the classical case belong to the subspace B_p of the well defined Orlicz space $L^{\Phi_1}(p)$, for some $p \in \mathcal{M}$. For the quantum version, in the absence of a sound theory of *quantum* Orlicz spaces to rely on, the very definition of the Banach space of generalised coordinates is part of the effort in our construction of $\mathcal{M}$. The general strategy is to start at an arbitrary point $\rho_0 \in \mathcal{M}$, take its logarithm and then perturb it in such a way that the set of all perturbations constitute the desired Banach space. The program has been carried out by Streater using form-bounded [64] and operator-bounded [63] perturbations. We will review these definitions in section 4.1 and generalise them to an interpolating class of potentials, called ε-bounded perturbations of a given Hamiltonian, which included them as extreme cases.

4.1 ε-Bounded Perturbations

We recall the concepts of operator-bounded and form-bounded perturbations [31]. Given operators H and X defined on dense domains $\mathcal{D}(H)$ and $\mathcal{D}(X)$ in a Hilbert space $\mathcal{H}$, we say that X is *H-bounded* if

i. $\mathcal{D}(H) \subset \mathcal{D}(X)$ and

ii. there exist positive constants a and b such that

$$\|X\psi\| \leq a\,\|H\psi\| + b\,\|\psi\|\,, \text{ for all } \psi \in \mathcal{D}(H).$$

Analogously, given a positive self-adjoint operator H and its associated form q_H, with form domain $Q(H) = \mathcal{D}(H^{1/2})$, given by

$$q_H(\psi, \phi) = \langle \psi, H\phi \rangle, \text{ for all } \psi \in Q(H), \tag{4.1}$$

we say that a symmetric quadratic form X (or the symmetric sesquiform obtained from it by polarization) is *q_H-bounded* if

i. $Q(H) \subset Q(X)$ and

ii. there exist positive constants a and b such that

$$|X(\psi, \psi)| \leq a q_H(\psi, \psi) + b\langle \psi, \psi \rangle, \text{ for all } \psi \in Q(H).$$

In both cases, the infimum of such a is called the relative bound of X (with respect to H or with respect to q_H, accordingly). If the relative bound is less than 1, then we say that X is a *small* perturbation. If it happens to be zero, then X is said to be an *infinitesimally small* perturbation.

We now want to give a meaning for the left and right product of a quadratic form by operators. Suppose that X is a quadratic form with domain $Q(X)$ and A, B are operators on $\mathcal{H}$ such that A^* and B are densely defined. Suppose further that $A^* : \mathcal{D}(A^*) \to Q(X)$ and $B : \mathcal{D}(B) \to Q(X)$. Then the expression AXB means the form defined by

$$\begin{aligned} AXB : \mathcal{D}(A^*) \times \mathcal{D}(B) &\to \mathbb{C} \\ (\phi, \psi) &\mapsto X(A^*\phi, B\psi) \end{aligned}$$

With this definition in mind, let us specialise to the case where $H_0 \geq I$ is a self-adjoint operator with domain $\mathcal{D}(H_0)$, quadratic form q_0 and form domain $Q_0 = \mathcal{D}(H_0^{1/2})$, and let $R_0 = H_0^{-1}$ be its resolvent at the origin. Then it is easy to show that a symmetric operator $X : \mathcal{D}(H_0) \to \mathcal{H}$ is H_0-bounded if and only if $\|XR_0\| < \infty$. The set $\mathcal{T}_{1/2}(0)$ of all H_0-bounded symmetric operators X is a Banach space when we furnish it with norm $\|X\|_{1/2}(0) := \|XR_0\|$, since the map $A \mapsto AH_0$ from $\mathcal{B}(\mathcal{H})$ to the set of all H_0-bounded operators is an isometry and $\mathcal{T}_{1/2}(0)$ is a closed subset of it (the reason for the subscript 1/2 will be made clear below; the zero in brackets is meant to indicate that we are taking perturbations of H_0).

A similar result holds for forms [64, lemma 2]:

Lemma 4.1.1 *A symmetric quadratic form X defined on Q_0 is q_0-bounded if and only if $R_0^{1/2}XR_0^{1/2}$ is a bounded symmetric form defined everywhere. Moreover, if $\left\|R_0^{1/2}XR_0^{1/2}\right\| < \infty$ then the relative bound a of X with respect to q_0 satisfies $a \leq \left\|R_0^{1/2}XR_0^{1/2}\right\|$.*

The set $\mathcal{T}_0(0)$ of all q_0-bounded symmetric forms X is also a Banach space with norm $\|X\|_0(0) := \left\|R_0^{1/2}XR_0^{1/2}\right\|$, since the map $A \mapsto H_0^{1/2}AH_0^{1/2}$ from the set of all bounded self-adjoint operators on $\mathcal{H}$ onto $\mathcal{T}_0(0)$ is again an isometry.

It is known that every operator-bounded self-adjoint perturbation of H_0 is also a form-bounded perturbation of q_0 [53, theorem X.18]. We want to define an interpolating class between these two extremes. For $\varepsilon \in [0, 1/2]$, let $\mathcal{T}_\varepsilon(0)$ be the set of all symmetric forms X defined on Q_0 and such that $\|X\|_\varepsilon(0) := \left\| R_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}+\varepsilon} \right\|$ is finite. We note that $\mathcal{D}(H_0^{\frac{1}{2}+\varepsilon}) \subset \mathcal{D}(H_0^{\frac{1}{2}})$, for all $0 \leq \varepsilon \leq 1/2$, so $R_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}+\varepsilon}$ is well defined. The map $A \mapsto H_0^{\frac{1}{2}-\varepsilon} A H_0^{\frac{1}{2}+\varepsilon}$ is an isometry from the set of all bounded self-adjoint operators on $\mathcal{H}$ onto $\mathcal{T}_\varepsilon(0)$. Hence $\mathcal{T}_\varepsilon(0)$ is a Banach space with the ε-norm $\| \cdot \|_\varepsilon(0)$.

To prove the next lemma, we need to use a Banach space-valued version of the Hadamard three lines theorem [53, p 33], which reads as follows.

Theorem 4.1.2 (Hadamard) *Let $\phi(z)$ be a Banach space-valued function which is bounded and continuous on the closed strip $\{z \,|\, 0 \leq Re\ z \leq 1\}$, analytic in its interior and satisfies*

$$\|\phi(z)\| \leq M_0, \qquad \text{if } Re\ z = 0$$

and

$$\|\phi(z)\| \leq M_1, \qquad \text{if } Re\ z = 1.$$

Then $\|\phi(z)\| \leq M_0^{1-Re z} M_1^{Re z}$ for all z in the strip.

Lemma 4.1.3 *For fixed symmetric X defined on Q_0, $\|X\|_\varepsilon$ is a monotonically increasing function of $\varepsilon \in [0, 1/2]$.*

Proof: Suppose that $\left\| R_0^{\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta} \right\| < \infty$ for some $\delta \in [0, 1/2]$. For $0 \leq x \leq 1$ define the function

$$F(x) = R_0^{2\delta x + \frac{1}{2} - \delta} X R_0^{\frac{1}{2} + \delta - 2\delta x},$$

which has an analytic continuation to the strip $\{z \,|\, 0 \leq \text{Re } z \leq 1\}$ given by

$$F(0) = R_0^{2\delta z + \frac{1}{2} - \delta} X R_0^{\frac{1}{2} + \delta - 2\delta z}.$$

Since $e^{2\delta iy\log R_0}$ is a unitary operator for all real y, we notice that,

$$\|F(z)\| = \left\| R_0^{\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta} \right\|,$$

if Re $z = 0$, whereas

$$\|F(z)\| = \left\| R_0^{\frac{1}{2}+\delta} X R_0^{\frac{1}{2}-\delta} \right\|,$$

if Re $z = 1$. Using Hadamard's three line theorem and the fact that

$$\left\| R_0^{\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta} \right\| = \left\| R_0^{\frac{1}{2}+\delta} X R_0^{\frac{1}{2}-\delta} \right\|$$

we conclude that

$$\left\| R_0^{2\delta x+\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta-2\delta x} \right\| \leq \left\| R_0^{\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta} \right\|, \tag{4.2}$$

for all $0 \leq x \leq 1$. In particular, for $0 \leq x \leq 1/2$, define the variable $y = \delta(1-2x)$, so that $0 \leq y \leq \delta$. Then from the inequality (4.2) we obtain

$$\left\| R_0^{\frac{1}{2}-y} X R_0^{\frac{1}{2}+y} \right\| \leq \left\| R_0^{\frac{1}{2}-\delta} X R_0^{\frac{1}{2}+\delta} \right\|, \tag{4.3}$$

for $0 \leq y \leq \delta \leq 1/2$, which proves the result.

4.2 Construction of the Manifold

4.2.1 The Underlying Set and Other Preliminaries

In the classical case, the underlying set for the statistical manifold is just the set of all densities of probability measures equivalent to a given measure μ. For finite dimensional quantum systems, we just took the set of all invertible density matrices. For infinite dimensional quantum systems, we need to be more specific than in both previous cases. This is because we want that all the perturbed states obtained from an arbitrary starting point $\rho_0 \in \mathcal{M}$ have roughly the same properties as ρ_0. The set of all density operators on $\mathcal{H}$ proves to be way too big for this purpose. Instead, let $\mathcal{C}_p, 0 < p < \infty$, denote the set of compact operators

$A : \mathcal{H} \mapsto \mathcal{H}$ such that $|A|^p \in \mathcal{C}_1$, where $\mathcal{C}_1$ is the set of trace-class operators on $\mathcal{H}$. Define

$$\mathcal{C}_{<1} := \bigcup_{0<p<1} \mathcal{C}_p.$$

We take the underlying set of the quantum information manifold to be

$$\mathcal{M} = \mathcal{C}_{<1} \cap \Sigma \tag{4.4}$$

where $\Sigma \subseteq \mathcal{C}_1$ denotes the set of density operators. Thus since all elements in $\mathcal{M}$ are positive and since $\mathcal{C}_p \subset \mathcal{C}_q$ for all $p \leq q$ [47], we see that if $\rho_0 \in \mathcal{M}$ then there exists $\beta_0 < 1$ such that ρ_0^β is of trace class for all $\beta \geq \beta_0$.

At this level, $\mathcal{M}$ has a natural affine structure coming from the convex structure of the set $\mathcal{C}_{<1}$: if $\rho_1 \in \mathcal{C}_{p_1} \cap \Sigma$ and $\rho_2 \in \mathcal{C}_{p_2} \cap \Sigma$ then $\rho_1, \rho_2 \in \mathcal{C}_p \cap \Sigma$, where $p = \max\{p_1, p_2\}$. We can then define "$\lambda\rho_1 + (1-\lambda)\rho_2, 0 \leq \lambda \leq 1$" on $\mathcal{M}$ as the usual sum of operators in $\mathcal{C}_p$. This is called the (-1)-affine structure.

We want to cover $\mathcal{M}$ by a Banach manifold. In [64] this is achieved for neigbourhoods of $\rho \in \mathcal{M}$ defined using form-bounded perturbations. The manifold obtained there is shown to have a Lipschitz structure. This is not enough, however, to define a metric on $\mathcal{M}$ as the second derivative of the free energy, so more regularity needs to be imposed. In [63] the construction is done with the more restrictive class of operator-bounded perturbations. The result then is that the manifold has an analytic structure. We now proceed using ε-bounded perturbations, with a similar result.

The recurrent tool we are going to be using in the what follows is the *KLMN* theorem (named after Kato, Lions, Lax, Milgram and Nelson) [53, theorem X.17].

Theorem 4.2.1 (KLMN) *Let $H_0 \geq 0$ be a self-adjoint operator with quadratic form q_0 and form domain Q_0 and suppose that X is a q_0-small symmetric quadratic form. Then there exists a unique self-adjoint operator H_X with form domain Q_0 such that*

$$\langle H_X^{1/2}\phi, H_X^{1/2}\psi\rangle = q_0(\phi,\psi) + X(\phi,\psi), \tag{4.5}$$

for all $\phi, \psi \in Q_0$. Moreover, H_X is bounded below by $-b$.

4.2.2 The First Chart

Let $\varepsilon \in (0, 1/2]$ be fixed. To each $\rho_0 \in \mathcal{C}_{\beta_0} \cap \Sigma$, $\beta_0 < 1$, let $H_0 = -\log \rho_0 + cI \geq I$ be a self-adjoint operator with domain $\mathcal{D}(H_0)$ such that

$$\rho_0 = \frac{e^{-H_0}}{Z_0} = e^{-(H_0+\Psi_0)}, \tag{4.6}$$

where

$$Z_0 = \mathrm{Tr}\ e^{-H_0} \tag{4.7}$$

is the quantum partition function and $\Psi_0 = \log Z_0$ is the free energy.

In $\mathcal{T}_\varepsilon(0)$, take X such that $\|X\|_\varepsilon(0) < 1 - \beta_0$. From lemma 4.1.3, we have that

$$\|X\|_0(0) \leq \|X\|_\varepsilon(0) < 1 - \beta_0,$$

so X is also q_0-bounded with bound a_0 less than $1 - \beta_0$. The *KLMN* theorem then tells us that there exists a unique semi-bounded self-adjoint operator H_X with form $q_X = q_0 + X$ and form domain $Q_X = Q_0$. Following an unavoidable abuse of notation, we write $H_X = H_0 + X$ and consider the operator

$$\rho_X = \frac{e^{-(H_0+X)}}{Z_X} = e^{-(H_0+X+\Psi_X)}. \tag{4.8}$$

Then the following result applies [64, lemma 4].

Lemma 4.2.2 *Let X be a q_0-small symmetric quadratic form with relative bound $a_0 < 1 - \beta_0$. Denote by H_X the unique operator obtained from X by the* KLMN *theorem. Then* $\exp(-\beta H_X)$ *is of trace class for all $\beta \geq \beta_X = \beta_0/(1 - a_0)$.*

Therefore, $\rho_X \in \mathcal{C}_{\beta_X} \cap \Sigma$, where $\beta_X = \frac{\beta_0}{1-a_0} < 1$. We take as a neighbourhood $\mathcal{M}_0$ of ρ_0 the set of all such states, that is,

$$\mathcal{M}_0 = \{\rho_X : \|X\|_\varepsilon(0) < 1 - \beta_0\}. \tag{4.9}$$

The state ρ_X does not change if we add to H_X a multiple of the identity in such a way that $H_X + cI \geq I$, since this is balanced by a similar change in the partition function $Z_X = \mathrm{Tr}\ e^{-H_X}$. Without further comments, we will always assume that, for the perturbed state, we have $H_X \geq I$. Precisely because $\rho_X = \rho_{X+\alpha I}$, we introduce in $\mathcal{T}_\varepsilon(0)$ the equivalence relation $X \sim Y$ if and only if $X - Y = \alpha I$ for some $\alpha \in \mathbb{R}$. We then identify ρ_X in $\mathcal{M}_0$ with the line

$$\{Y \in \mathcal{T}_\varepsilon(0) : Y = X + \alpha I, \alpha \in \mathbb{R}\} \tag{4.10}$$

in $\mathcal{T}_\varepsilon(0)/\sim$. This is a bijection from $\mathcal{M}_0$ onto the subset of $\mathcal{T}_\varepsilon(0)/\sim$ defined by

$$\{\{X + \alpha I\}_{\alpha\in\mathbb{R}} : \|X\|_\varepsilon(0) < 1 - \beta_0\} \tag{4.11}$$

and $\mathcal{M}_0$ becomes topologised by transfer of structure.

Now that $\mathcal{M}_0$ is a (Hausdorff) topological space, we want to parametrise it by an open set in a Banach space. By analogy with the classical case of chapter 2, we want to use the Banach subspace of centred variables in $\mathcal{T}_\varepsilon(0)$; in our terms, perturbations with zero mean. The problem is that, although it is known that $\rho_0 X$ is an operator of trace class for the case where X is H_0-bounded [63, corollary to theorem 2.2], this is not known to be true when X is a general ε-bounded perturbation. Instead, following [64], define the *regularised* mean of $X \in \mathcal{T}_\varepsilon(0)$ in the state ρ_0 as

$$\rho_0 \cdot X := \mathrm{Tr}(\rho_0^\lambda X \rho_0^{1-\lambda}), \qquad \text{for } 0 < \lambda < 1. \tag{4.12}$$

The following result concerning the regularised mean was obtained by Streater [64, lemma 5].

Lemma 4.2.3 *Let $\rho_0 \in \mathcal{M}$ and suppose that X is a q_0-bounded perturbation. Then $\rho_0^\lambda X \rho_0^{1-\lambda}$ is of trace class for all $0 < \lambda < 1$ and its trace is independent of λ. Moreover, $\rho_0 \cdot X$ is continuous as a map from $\mathcal{T}_0(0)$ to $\mathbb{R}$.*

The continuity claim in the above lemma follows from the fact that in the course of its proof one finds that

$$|\rho_0 \cdot X| \leq C\|X\|_0(0) \tag{4.13}$$

Now suppose that X is an ε-bounded perturbation of H_0. Then from lemma 4.1.3 it is also q_0-bounded, so the result of the previous lemma holds unchanged. Moreover, from (4.13) we get that

$$|\rho_0 \cdot X| \leq C\|X\|_\varepsilon(0), \tag{4.14}$$

which implies that $\rho_0 \cdot X$ is continuous as a map from $\mathcal{T}_\varepsilon(0)$ to $\mathbb{R}$ as well. Therefore, the set

$$\widehat{\mathcal{T}}_\varepsilon(0) := \{X \in \mathcal{T}_\varepsilon(0) : \rho_0 \cdot X = 0\} \tag{4.15}$$

is a closed subspace of $\mathcal{T}_\varepsilon(0)$ and so is a Banach space with the norm $\|\cdot\|_\varepsilon$ restricted to it.

To each $\rho_X \in \mathcal{M}_0$, consider the unique intersection of the equivalence class of X in $\mathcal{T}_\varepsilon(0)/\sim$ with the set $\widehat{\mathcal{T}}_\varepsilon(0)$, that is, the point in the line $\{X + \alpha I\}_{\alpha \in \mathbb{R}}$ with $\alpha = -\rho_0 \cdot X$. Write

$$\widehat{X} = X - \rho_0 \cdot X \tag{4.16}$$

for this point. The map $\rho_X \mapsto \widehat{X}$ is a homeomorphism between $\mathcal{M}_0$ and the open subset of $\widehat{\mathcal{T}}_\varepsilon(0)$ defined by

$$\mathcal{V}_0 = \left\{\widehat{X} := \widehat{X} = X - \rho_0 \cdot X, \|X\|_\varepsilon < 1 - \beta_0\right\}. \tag{4.17}$$

The map

$$\begin{aligned} e_0^{-1} : \mathcal{M}_0 &\rightarrow \mathcal{V}_0 \\ \rho &\mapsto -(\log \rho + H_0) + \rho_0 \cdot (\log \rho + H_0) \end{aligned} \tag{4.18}$$

is then a global chart for the Banach manifold $\mathcal{M}_0$ modelled by $\widehat{\mathcal{T}}_\varepsilon(0)$. Its inverse is simply

$$\begin{aligned} e_0 : \mathcal{V}_0 &\rightarrow \mathcal{M}_0 \\ V &\mapsto \frac{e^{-(H_0+V)}}{Z_V}. \end{aligned} \tag{4.19}$$

By analogy with the classical case, we identify the tangent space at ρ_0 with $\widehat{\mathcal{T}_\varepsilon}(0)$. More explicitly, each curve through $\rho_0 \in \mathcal{M}_0$ is tangent to a one-dimensional exponential model

$$\rho(\lambda) = \frac{e^{-(H_0+\lambda X)}}{Z_{\lambda X}}, \qquad \lambda \in [-\delta, \delta] \tag{4.20}$$

and we take $\widehat{X} = X - \rho_0 \cdot X$ as the tangent vector representing the equivalence class of such a curve.

4.2.3 Enlarging the Manifold

We extend our manifold by adding new patches compatible with $\mathcal{M}_0$. The idea is to construct a chart around each perturbed state ρ_X as we did around ρ_0. Let $\rho_X \in \mathcal{M}_0$ with Hamiltonian $H_X \geq I$ and consider the Banach space $\mathcal{T}_\varepsilon(X)$ of all symmetric forms Y on Q_0 such that the norm $\|Y\|_\varepsilon(X) := \left\| R_X^{\frac{1}{2}-\varepsilon} Y R_X^{\frac{1}{2}+\varepsilon} \right\|$ is finite, where $R_X = H_X^{-1}$ denotes the resolvent of H_X at the origin. In $\mathcal{T}_\varepsilon(X)$, take Y such that $\|Y\|_\varepsilon(X) < 1 - \beta_X$. From lemma 4.1.3 we know that Y is q_X-bounded with bound a_X less than $1 - \beta_X$. Let H_{X+Y} be the unique semi-bounded self-adjoint operator, given by the *KLMN* theorem, with form

$$q_{X+Y} = q_X + Y = q_0 + X + Y \tag{4.21}$$

and form domain $Q_{X+Y} = Q_X = Q_0$. Then the operator

$$\rho_{X+Y} = \frac{e^{-H_{X+Y}}}{Z_{X+Y}} = \frac{e^{-(H_0+X+Y)}}{Z_{X+Y}} \tag{4.22}$$

is in $\mathcal{C}_{\beta_Y} \cap \Sigma$, where $\beta_Y = \frac{\beta_X}{1-a_X}$.

We take as a neighbourhood of ρ_X the set $\mathcal{M}_X$ of all such states. Again $\rho_{X+Y} = \rho_{X+Y+\alpha I}$, so we furnish $\mathcal{T}_\varepsilon(X)$ with the equivalence relation $Z \sim Y$ if and only if $Z - Y = \alpha I$ for some $\alpha \in \mathbb{R}$ and we see that $\mathcal{M}_X$ is mapped bijectively onto the set of lines

$$\{\{Z = Y + \alpha I\}_{\alpha \in \mathbb{R}}, \|Y\|_\varepsilon(X) < 1 - \beta_X\} \tag{4.23}$$

in $\mathcal{T}_\varepsilon(X)/\sim$. In this way we topologise $\mathcal{M}_X$, by transfer of structure, with the quotient topology of $\mathcal{T}_\varepsilon(X)/\sim$.

Again we can define the mean of Y in the state ρ_X by

$$\rho_X \cdot Y := \mathrm{Tr}(\rho_X^\lambda Y \rho_X^{1-\lambda}), \qquad \text{for } 0 < \lambda < 1. \tag{4.24}$$

and notice that it is finite and independent of λ. This is a continuous function of Y with respect to the norm $\|\cdot\|_\varepsilon(X)$, hence

$$\widehat{\mathcal{T}}_\varepsilon(X) = \{Y \in \mathcal{T}_\varepsilon(X) : \rho_X \cdot Y = 0\} \tag{4.25}$$

is closed and so is a Banach space with the norm $\|\cdot\|_\varepsilon(X)$ restricted to it. Finally, let $\widehat{Y}$ be the unique intersection of the line $\{Z = Y + \alpha I\}_{\alpha\in\mathbb{R}}$ with the hyperplane $\widehat{\mathcal{T}}_\varepsilon(X)$, given by $\alpha = -\rho_X \cdot Y$. Then $\rho_{X+Y} \mapsto \widehat{Y}$ is a homeomorphism between $\mathcal{M}_X$ and the open subset of $\widehat{\mathcal{T}}_\varepsilon(X)$ defined by

$$\mathcal{V}_X = \left\{\widehat{Y} \in \widehat{\mathcal{T}}_\varepsilon(X) : \widehat{Y} = Y - \rho_X \cdot Y, \|Y\|_\varepsilon(X) < 1 - \beta_X\right\}.$$

We obtain that

$$\begin{aligned} e_X^{-1} : \mathcal{M}_X &\to \mathcal{V}_X \\ \rho &\mapsto -(\log\rho + H_X) + \rho_X \cdot (\log\rho + H_X) \end{aligned} \tag{4.26}$$

is a chart for the manifold $\mathcal{M}_X$ modelled by $\widehat{\mathcal{T}}_\varepsilon(X)$, with inverse given by

$$\begin{aligned} e_X : \mathcal{V}_X &\to \mathcal{M}_X \\ V &\mapsto \frac{e^{-(H_X+V)}}{Z_{X+V}}. \end{aligned} \tag{4.27}$$

The tangent space at ρ_X is identified with $\widehat{\mathcal{T}}_\varepsilon(X)$ itself.

We now look to the union of $\mathcal{M}_0$ and $\mathcal{M}_X$. We need to show that our two previous charts are compatible in the overlapping region $\mathcal{M}_0 \cap \mathcal{M}_X$. But first we prove the following series of technical lemmas.

Lemma 4.2.4 *Let X be a symmetric form defined on Q_0 such that $\|X\|_0(0) < 1$. Then $\mathcal{D}(H_0^{\frac{1}{2}-\varepsilon}) = \mathcal{D}(H_X^{\frac{1}{2}-\varepsilon})$, for any $\varepsilon \in (0, 1/2)$.*

Proof: We know that $\mathcal{D}(H_0^{1/2}) = \mathcal{D}(H_X^{1/2})$, since X is q_0-small. Moreover, H_X and H_0 are comparable as forms, that is, there exists $c > 0$ such that

$$c^{-1}q_0(\psi) \leq q_X(\psi) \leq cq_0(\psi), \qquad \text{for all } \psi \in Q_0.$$

Using the fact that $x \mapsto x^\alpha \quad (0 < \alpha < 1)$ is an operator monotone function [9, lemma 4.20], we conclude that

$$c^{-(1-2\varepsilon)}H_0^{1-2\varepsilon} \leq H_X^{1-2\varepsilon} \leq c^{1-2\varepsilon}H_0^{1-2\varepsilon},$$

which implies that $\mathcal{D}(H_0^{\frac{1}{2}-\varepsilon}) = \mathcal{D}(H_X^{\frac{1}{2}-\varepsilon})$.

The conclusion remains true if we now replace H_X by $H_X + I$, if necessary in order to have $H_X \geq I$. This is assumed in the next corollary.

Corollary 4.2.5 *The operator $H_X^{\frac{1}{2}-\varepsilon}R_0^{\frac{1}{2}-\varepsilon}$ is bounded and has bounded inverse $H_0^{\frac{1}{2}-\varepsilon}R_X^{\frac{1}{2}-\varepsilon}$.*

Proof: $R_0^{\frac{1}{2}-\varepsilon}$ is bounded and maps $\mathcal{H}$ into $\mathcal{D}(H_0^{\frac{1}{2}-\varepsilon}) = \mathcal{D}(H_X^{\frac{1}{2}-\varepsilon})$. Then $H_X^{\frac{1}{2}-\varepsilon}R_0^{\frac{1}{2}-\varepsilon}$ is bounded, since $H_X^{\frac{1}{2}-\varepsilon}$ is closed. By exactly the same argument, we obtain that $H_0^{\frac{1}{2}-\varepsilon}R_X^{\frac{1}{2}-\varepsilon}$ is bounded. Finally

$$(H_0^{\frac{1}{2}-\varepsilon}R_X^{\frac{1}{2}-\varepsilon})(H_X^{\frac{1}{2}-\varepsilon}R_0^{\frac{1}{2}-\varepsilon}) = (H_X^{\frac{1}{2}-\varepsilon}R_0^{\frac{1}{2}-\varepsilon})(H_0^{\frac{1}{2}-\varepsilon}R_X^{\frac{1}{2}-\varepsilon}) = I.$$

Lemma 4.2.6 *For $\varepsilon \in (0, 1/2)$, let X be a symmetric form defined on Q_0 such that $\left\| R_0^{\frac{1}{2}+\varepsilon}XR_0^{\frac{1}{2}-\varepsilon} \right\| < 1$. Then $R_0^{\frac{1}{2}+\varepsilon}H_X^{\frac{1}{2}+\varepsilon}$ is bounded and has bounded inverse $R_X^{\frac{1}{2}+\varepsilon}H_0^{\frac{1}{2}+\varepsilon}$. Moreover, $\mathcal{D}(H_0^{\frac{1}{2}+\varepsilon}) = \mathcal{D}(H_X^{\frac{1}{2}+\varepsilon})$*

Proof: From lemma 4.1.3, we know that $\left\| R_0^{1/2}XR_0^{1/2} \right\| < 1$, so lemma 4.2.4 and its corollary apply. We have that

$$\begin{aligned} 1 &> \left\| R_0^{\frac{1}{2}+\varepsilon}XR_0^{\frac{1}{2}-\varepsilon} \right\| \\ &= \left\| R_0^{\frac{1}{2}+\varepsilon}(H_X - H_0)R_0^{\frac{1}{2}-\varepsilon} \right\| \\ &= \left\| R_0^{\frac{1}{2}+\varepsilon}H_XR_0^{\frac{1}{2}-\varepsilon} - I \right\|, \end{aligned}$$

thus $\left\|R_0^{\frac{1}{2}+\varepsilon} H_X R_0^{\frac{1}{2}-\varepsilon}\right\| < \infty$. We write this as

$$\left\|R_0^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}-\varepsilon} R_0^{\frac{1}{2}-\varepsilon}\right\| < \infty.$$

Since $H_X^{\frac{1}{2}-\varepsilon} R_0^{\frac{1}{2}-\varepsilon}$ is bounded and invertible, so is $R_0^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}+\varepsilon}$. Finally, the fact that $\left\|R_0^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}+\varepsilon}\right\| < \infty$ and $\left\|R_X^{\frac{1}{2}+\varepsilon} H_0^{\frac{1}{2}+\varepsilon}\right\| < \infty$ implies that $H_X^{\frac{1}{2}+\varepsilon}$ and $H_0^{\frac{1}{2}+\varepsilon}$ are comparable, hence $\mathcal{D}(H_0^{\frac{1}{2}+\varepsilon}) = \mathcal{D}(H_X^{\frac{1}{2}+\varepsilon})$.

The next theorem ensures the compatibility between the two charts in the overlapping region $\mathcal{M}_0 \cap \mathcal{M}_X$.

Theorem 4.2.7 *$\|\cdot\|_\varepsilon(X)$ and $\|\cdot\|_\varepsilon(0)$ are equivalent norms.*

Proof: We need to show that there exist positive constants m and M such that $m\|Y\|_\varepsilon(0) \le \|Y\|_\varepsilon(X) \le M\|Y\|_\varepsilon(0)$. We just write

$$\begin{aligned} \|Y\|_\varepsilon(X) &= \left\|R_X^{\frac{1}{2}-\varepsilon} H_0^{\frac{1}{2}-\varepsilon} R_0^{\frac{1}{2}-\varepsilon} Y R_0^{\frac{1}{2}+\varepsilon} H_0^{\frac{1}{2}+\varepsilon} R_X^{\frac{1}{2}+\varepsilon}\right\| \\ &\le \left\|R_X^{\frac{1}{2}+\varepsilon} H_0^{\frac{1}{2}+\varepsilon}\right\| \left\|H_0^{\frac{1}{2}-\varepsilon} R_X^{\frac{1}{2}-\varepsilon}\right\| \|Y\|_\varepsilon(0) \\ &= M\|Y\|_\varepsilon(0) \end{aligned}$$

and, for the inequality in the other direction, we write

$$\begin{aligned} \|Y\|_\varepsilon(0) &= \left\|R_0^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}+\varepsilon} R_X^{\frac{1}{2}+\varepsilon} Y R_X^{\frac{1}{2}-\varepsilon} H_X^{\frac{1}{2}-\varepsilon} R_0^{\frac{1}{2}-\varepsilon}\right\| \\ &\le \left\|R_0^{\frac{1}{2}+\varepsilon} H_X^{\frac{1}{2}+\varepsilon}\right\| \left\|H_X^{\frac{1}{2}-\varepsilon} R_0^{\frac{1}{2}-\varepsilon}\right\| \|Y\|_\varepsilon(X) \\ &= m^{-1}\|Y\|_\varepsilon(X). \end{aligned}$$

We see that $\mathcal{T}_\varepsilon(0)$ and $\mathcal{T}_\varepsilon(X)$ are, in fact, the same Banach space furnished with two equivalent norms, and observe that the quotient spaces $\mathcal{T}_\varepsilon(0)/\sim$ and $\mathcal{T}_\varepsilon(X)/\sim$ are exactly the same set. We can now verify that the change of coordinates formulae on the overlapping region are

$$\begin{aligned} e_X^{-1} e_0 : e_0^{-1}(\mathcal{M}_0 \cap \mathcal{M}_X) &\rightarrow e_X^{-1}(\mathcal{M}_0 \cap \mathcal{M}_X) \\ V &\mapsto V + (H_0 - H_X) - \rho_X \cdot [V + (H_0 - H_X)] \end{aligned} \tag{4.28}$$

and

$$\begin{aligned} e_0^{-1}e_X : e_X^{-1}(\mathcal{M}_0 \cap \mathcal{M}_X) &\rightarrow e_0^{-1}(\mathcal{M}_0 \cap \mathcal{M}_X) \\ V &\mapsto V + (H_X - H_0) - \rho_0 \cdot [V + (H_X - H_0)], \end{aligned} \tag{4.29}$$

from which we conclude that $\mathcal{M}_0 \cup \mathcal{M}_X$ has the structure of a C^∞-manifold modelled by $\widehat{\mathcal{T}}_\varepsilon$.

We continue in this way, adding a new patch around another point $\rho_{X'}$ in $\mathcal{M}_0$ or around some other point in $\mathcal{M}_X$ but outside $\mathcal{M}_0$. Whichever point we start from, we get a third piece $\mathcal{M}_{X'}$ with chart into an open subset of the Banach space $\{Y \in \mathcal{T}_\varepsilon(X') : \rho_{X'} \cdot Y = 0\}$, with norm $\|Y\|_\varepsilon(X') := \left\| R_{X'}^{\frac{1}{2}-\varepsilon} Y R_{X'}^{\frac{1}{2}+\varepsilon} \right\|$ equivalent to the previously defined norms. We can go on inductively, and all the norms of any overlapping regions will be equivalent.

Definition 4.2.8 The information manifold $\mathcal{M}(H_0)$ defined by H_0 consists of all states obtainable in a finite number of steps, by extending $\mathcal{M}_0$ as explained above.

These states are well defined in the following sense. If, for two different sets of perturbations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, we have $X_1 + \cdots + X_n = Y_1 + \cdots + Y_m$ as forms on Q_0, then we arrive at the same state either taking the route $X_1, \ldots, X_n$ or taking the route $Y_1, \ldots, Y_m$, since the self-adjoint operator associated with the form $q_0 + X_1 + \cdots + X_n = q_0 + Y_1 + \cdots + Y_m$ is unique.

4.2.4 Affine Geometry in $\mathcal{M}(H_0)$

The set $\mathcal{V}_0 = \left\{ \widehat{X} \in \widehat{\mathcal{T}}_\varepsilon(0) : \widehat{X} = X - \rho_0 \cdot X, \|X\|_\varepsilon(0) < 1 - \beta_0 \right\}$ is a convex subset of the Banach space $\widehat{\mathcal{T}}_\varepsilon(0)$ and so has an affine structure coming from its linear structure. We provide $\mathcal{M}_0$ with the affine structure induced from $\mathcal{V}_0$ using the patch e_0^{-1} and call this the canonical or $(+1)$-affine structure. The $(+1)$-convex mixture of ρ_X and ρ_Y in $\mathcal{M}_0$ is then $\rho_{\lambda X + (1-\lambda)Y}$, $(0 \leq \lambda \leq 1)$, which differs from the previously defined (-1)-convex mixture $\lambda\rho_X + (1-\lambda)\rho_Y$.

As in the classical case, this affine structure induces the *exponential* connection on $T\mathcal{M}_0$ as follows. Given two points ρ_X and ρ_Y in $\mathcal{M}_0$ and their tangent spaces $\widehat{\mathcal{T}}_\varepsilon(X)$ and $\widehat{\mathcal{T}}_\varepsilon(Y)$, the exponential parallel transport is given by

$$\begin{aligned} \tau^{(1)}_{\rho_X\rho_Y} : \widehat{\mathcal{T}}_\varepsilon(X) &\to \widehat{\mathcal{T}}_\varepsilon(Y) \\ Z &\mapsto Z - \rho_Y \cdot Z. \end{aligned} \tag{4.30}$$

We see that the exponential parallel transport just moves the representative point in the line $\{Z + \alpha I\}_{\alpha\in\mathbb{R}}$ from one hyperplane to another. It is manifestly path-independent, therefore the exponential connection on $T\mathcal{M}_0$ is flat.

Now consider a second piece of the manifold, say $\mathcal{M}_X$. We have the (+1)-affine structure on it again by transfer of structure from $\widehat{\mathcal{T}}_\varepsilon(X)$. Since both $\widehat{\mathcal{T}}_\varepsilon(0)$ and $\widehat{\mathcal{T}}_\varepsilon(X)$ inherit their affine structures from the linear structure of the same set (either $\mathcal{T}_\varepsilon(0)$ or $\mathcal{T}_\varepsilon(X)$), we see that the (+1)-affine structures of $\mathcal{M}_0$ and $\mathcal{M}_X$ are the same on their overlap. We define the parallel transport in $\mathcal{M}_X$ again by moving representative points around. To parallel transport a point between any two tangent spaces in the union of the two pieces, we proceed by stages. For instance, if $\tau^{(1)}_{\rho_0\rho_X}$ denotes the parallel transport from ρ_0 to ρ_X, it is straightforward to check that it takes a convex mixture in $\widehat{\mathcal{T}}_\varepsilon(0)$ to a convex mixture in $\widehat{\mathcal{T}}_\varepsilon(X)$. So, if $\rho_Y \in \mathcal{M}_0$ and $\rho_{Y'} \in \mathcal{M}_X$ are points outside the overlap, we parallel transport from ρ_Y to $\rho_{Y'}$ following the route $\rho_Y \to \rho_0 \to \rho_X \to \rho_{Y'}$. Continuing in this way, we furnish the whole $\mathcal{M}(H_0)$ with a (+1)-affine structure and a flat, torsion free, (+1)-affine connection.

4.3 Analyticity of the Free Energy

4.3.1 Differentiability of the Free Energy

We begin to address in this section the problem of defining an infinite dimensional version of the *BKM* metric. In the finite dimensional case it is obtained by twice differentiating the free energy. We pursue the same line here.

The free energy of the state $\rho_X = Z_X^{-1}e^{-H_X} \in \mathcal{C}_{\beta_X} \subset \mathcal{M}, \beta_X < 1$, is the function $\Psi : \mathcal{M} \to \mathbb{R}$ given by

$$\Psi(\rho_X) := \log Z_X. \tag{4.31}$$

When we compose it with the inverse of any of the local charts it becomes a map between Banach spaces

$$\Psi_X \equiv \Psi(\rho_X) : \widehat{\mathcal{T}}_\varepsilon \to \mathbb{R}. \tag{4.32}$$

We therefore need to use the concepts of calculus in Banach spaces [70, 68]. The relevant definitions of the different kinds of derivatives of maps between Banach spaces are given below. In what follows, $\mathcal{X}$ and $\mathcal{Y}$ are real Banach spaces and $F : U(x) \subset \mathcal{X} \to \mathcal{Y}$ is any map, not necessarily linear, defined in a neighbourhood $U(x)$ of $x \in \mathcal{X}$. We denote by $L(\mathcal{X}, \mathcal{Y})$ the set of all continuous linear operators from $\mathcal{X}$ to $\mathcal{Y}$.

Definition 4.3.1 We say that F is *Gatêaux differentiable* at $x \in \mathcal{X}$ if and only if there exist a continuous linear operator $F'(x) : \mathcal{X} \to \mathcal{Y}$ such that

$$F(x + tk) - F(x) = tF'(x)k + o(t), \qquad t \to 0, \tag{4.33}$$

for all $k \in \mathcal{X}$ with $\|k\| = 1$. The map $F'(x) \in L(\mathcal{X}, \mathcal{Y})$ is called the *Gatêaux derivative* of F at x and its value at k is called the *Gatêaux differential* of F in the direction k and is denoted by $DF(x; k) = F'(x)k$. When the Gatêaux derivative of F exists for all $x \in A \subset \mathcal{X}$ then the map

$$\begin{aligned} F' : A \subset \mathcal{X} &\to L(\mathcal{X}, \mathcal{Y}) \\ x &\mapsto F'(x) \end{aligned} \tag{4.34}$$

is called the Gatêaux derivative of F on A.

Definition 4.3.2 We say that F is *Fréchet differentiable* at $x \in \mathcal{X}$ if and only if there exist a continuous linear operator $F'(x) : \mathcal{X} \to \mathcal{Y}$ such that

$$F(x + h) - F(x) = F'(x)h + o(\|h\|), \qquad \|h\| \to 0, \tag{4.35}$$

for all $h \in \mathcal{X}$. The map $F'(x) \in L(\mathcal{X}, \mathcal{Y})$ is called the *Fréchet derivative* of F at x and its value at h is called the *Fréchet differential* of F in the direction h and is denoted by $DF(x; h) = F'(x)h$. When the Fréchet derivative of F exists for all $x \in A \subset \mathcal{X}$ then the map

$$\begin{aligned} F' : A \subset \mathcal{X} &\to L(\mathcal{X}, \mathcal{Y}) \\ x &\mapsto F'(x) \end{aligned} \tag{4.36}$$

is called the Fréchet derivative of F on A.

We can only share the reader's discomfort if she feels bewildered by the fact that both types of derivatives and differentials are denoted by exactly the same symbols. This is the unfortunate notation found in the nonlinear funtional analysis literature.

It is clear from the definitions that every Fréchet derivative is also a Gatêaux derivative. For the converse, however, we have the following useful criteria [70, proposition 4.8]. Let us remark that an equivalent way of writing (4.33) is saying that

$$\lim_{t \to 0} \left\| \frac{1}{t} [F(x + tk) - F(x) - tF'(x)k] \right\| = 0, \tag{4.37}$$

for all $k \in \mathcal{X}$ with $\|k\| = 1$.

Lemma 4.3.3 *A Gatêaux derivative at x for which the limit in (4.37) is* uniform *for all $k \in \mathcal{X}$ with $\|k\| = 1$ is also a Fréchet derivative at x.*

Lemma 4.3.4 *If F' exists as a Gatêaux derivative in some neigbourhood of x and is continuous at x, then $F'(x)$ is also a Fréchet derivative at x.*

Higher derivatives and differentials are defined successively. For instance, suppose that F' is defined on the whole $\mathcal{X}$. Then $F''(x)$ is the (Gatêaux or Fréchet) derivative of F' at x, where $F' : \mathcal{X} \to L(\mathcal{X}, \mathcal{Y})$ is itself regarded as a map between Banach spaces. Therefore $F''(x) \in L(\mathcal{X}, L(\mathcal{X}, \mathcal{Y}))$ and it is a standard exercise to show that $L(\mathcal{X}, L(\mathcal{X}, \mathcal{Y}))$ is isomorphic to $L(\mathcal{X} \times \mathcal{X}, \mathcal{Y})$, that is, the set of all bounded bilinear maps from $\mathcal{X} \times \mathcal{X}$ to $\mathcal{Y}$. We can then denote

$$D^2 F(x; h, k) = F''(x)(h, k) := (F''(x)h)\, k, \quad h, k \in X \tag{4.38}$$

The same is true for all higher order derivatives. That is, the n-th differential of F at x in the directions $v_1, v_2, \ldots, v_n \in \mathcal{X}$ is identified with a continuous n-linear map from $\mathcal{X} \times \mathcal{X} \cdots \times \mathcal{X}$ to $\mathcal{Y}$ and denoted by

$$D^n F(x; v_1, v_2, \ldots, v_n) = F^n(x)(v_1, v_2, \ldots, v_n). \tag{4.39}$$

The key tool for differentiating the free energy is the Duhamel formula for forms [64, theorem 9].

Theorem 4.3.5 (Duhamel's formula) *Let X be q_0-small symmetric form and let H_X be the self-adjoint operator with form $q_0 + X$. Then*

$$e^{-H_0} - e^{-H_X} = \int_0^1 e^{-\lambda H_0} X e^{-(1-\lambda)H_X} d\lambda, \tag{4.40}$$

as bounded sesquiforms on $\mathcal{H} \times \mathcal{H}$.

Using (4.40) Streater was able to show that, for the manifold contructed using form-bounded perturbation, the partition function is Lipschitz continuous [64, lemma 10, corollary 12]. For ε-bounded perturbations, a generalised version of his result goes as follows.

Theorem 4.3.6 *Suppose that X, Y are ε-bounded perturbations of H_0 and that $\|Y\|_\varepsilon(0) < 1 - \beta_0$. Then $e^{-\lambda H_0} X e^{-(1-\lambda)H_Y}$ is of trace class for $0 < \lambda < 1$ and its trace can be bounded independently of λ.*

Proof: The proof mimics the one used for lemma 4.2.3, namely choose a β such that

$$\beta_0 < \beta_Y < \beta < 1 \tag{4.41}$$

and write

$$e^{-\lambda H_0}Xe^{-(1-\lambda)H_Y} = e^{-\lambda\beta H_0}\left(e^{-\lambda(1-\beta)H_0}H_0^{\frac{1}{2}-\varepsilon}\right)\left(R_0^{\frac{1}{2}-\varepsilon}XR_0^{\frac{1}{2}+\varepsilon}\right) \left(H_0^{\frac{1}{2}+\varepsilon}R_Y^{\frac{1}{2}+\varepsilon}\right)\left(H_Y^{\frac{1}{2}+\varepsilon}e^{-(1-\lambda)(1-\beta)H_Y}\right)e^{-(1-\lambda)\beta H_0} \tag{4.42}$$

The operators in brackets are all bounded and using Hölder's inequality we obtain

$$\left\|e^{-\lambda H_0}Xe^{-(1-\lambda)H_Y}\right\|_1 \leq C_{\lambda,\beta}\|X\|_\varepsilon(0)\left\|e^{-\lambda\beta H_0}\right\|_{1/\lambda}\left\|e^{-(1-\lambda)\beta H_Y}\right\|_{1/(1-\lambda)} < \infty \tag{4.43}$$

Therefore $e^{-\lambda H_0}Xe^{-(1-\lambda)H_Y}$ is of trace class, although its trace norm is not uniformly bounded for $\lambda \in (0,1)$, due to the presence of the constant $C_{\lambda,\beta}$ in the bound (4.43).

To show that the *trace* of $e^{-\lambda H_0}Xe^{-(1-\lambda)H_Y}$ can have a bound independent of λ, suppose that $\lambda < 1/2$ (the case $\lambda \geq 1/2$ can be treated similarly). Let us first note that if we write

$$e^{-\lambda H_0}Xe^{-(1-\lambda)H_Y} = e^{-\lambda H_0}Xe^{-(1-\lambda-\delta)H_Y}e^{-\delta H_0}$$

for some $\delta > 0$ yet to be specified, then $e^{-\lambda H_0}Xe^{-(1-\lambda-\delta)H_Y}$ is still of trace class. This is because we can write it as

$$e^{-\lambda H_0}Xe^{-(1-\lambda-\delta)H_Y} = e^{-(\lambda-\delta_1)H_0}\left(e^{-\delta_1 H_0}H_0^{\frac{1}{2}-\varepsilon}\right)\left(R_0^{\frac{1}{2}-\varepsilon}XR_0^{\frac{1}{2}+\varepsilon}\right) \left(H_0^{\frac{1}{2}+\varepsilon}R_Y^{\frac{1}{2}+\varepsilon}\right)\left(H_Y^{\frac{1}{2}+\varepsilon}e^{-\delta_2 H_Y}\right)e^{-(1-\lambda-\delta_2-\delta)H_Y}, \tag{4.44}$$

for some $\delta_1, \delta_2 > 0$ such that $\delta + \delta_2 < (1-\lambda)$ and $\delta_1 < \lambda$. Then the operators in brackets are bounded and Hölder's inequality gives

$$\left\|e^{-\lambda H_0}Xe^{-(1-\lambda-\delta)H_Y}\right\|_1 \leq C_{\delta_1,\delta_2}\left\|e^{-(\lambda-\delta_1)H_0}\right\|_p\left\|e^{-(1-\lambda-\delta_2-\delta)H_Y}\right\|_q < \infty, \tag{4.45}$$

provided that we take

$$p = \frac{\beta_Y}{\lambda-\delta_1}, \qquad q = \frac{\beta_Y}{(1-\lambda)-\delta_2-\delta} \tag{4.46}$$

subject to $p^{-1}+q^{-1} = 1$, that is,

$$1-(\delta+\delta_1+\delta_2) = \beta_Y. \tag{4.47}$$

This can be satisfied for small enough $\delta, \delta_1, \delta_2$. In particular, we must have

$$\delta < 1 - \beta_Y. \tag{4.48}$$

Now suppose that $\alpha < 1-\beta_Y$, then we certainly have $\frac{(1-\lambda)\alpha}{2} < 1-\beta_Y$ and it follows cyclicity of the trace that

$$\begin{aligned}\left|\mathrm{Tr}\left(e^{-\lambda H_0} X e^{-(1-\lambda)H_Y}\right)\right| = \Big|\mathrm{Tr}\Big\{&\left(e^{-\frac{(1-\lambda)\alpha}{2}H_Y} H_Y^{\frac{1}{2}-\varepsilon}\right)\left(R_Y^{\frac{1}{2}-\varepsilon} H_0^{\frac{1}{2}-\varepsilon}\right) e^{-\lambda H_0} \\ &\left(R_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}+\varepsilon}\right)\left(H_0^{\frac{1}{2}+\varepsilon} R_Y^{\frac{1}{2}+\varepsilon}\right)\left(H_Y^{\frac{1}{2}+\varepsilon} e^{-\frac{(1-\lambda)\alpha}{2}H_Y}\right) e^{-(1-\lambda)(1-\alpha)H_Y}\Big\}\Big|.\end{aligned} \tag{4.49}$$

From the spectral theorem, it follows that

$$\left\| e^{-\frac{(1-\lambda)\alpha}{2}H_Y} H_Y^{\frac{1}{2}-\varepsilon}\right\| \le \sup_x \left\{x^{\frac{1}{2}-\varepsilon} e^{-\alpha x/4}\right\} \le \frac{c_1}{\alpha^{\frac{1}{2}-\varepsilon}}, \tag{4.50}$$

and that

$$\left\| e^{-\frac{(1-\lambda)\alpha}{2}H_Y} H_Y^{\frac{1}{2}+\varepsilon}\right\| \le \sup_x \left\{x^{\frac{1}{2}+\varepsilon} e^{-\alpha x/4}\right\} \le \frac{c_2}{\alpha^{\frac{1}{2}+\varepsilon}}, \tag{4.51}$$

for some real numbers c_1, c_2. The other factors in (4.49) are operators whose norm is bounded by a constant c, independent of λ and α. Applying Hölder's inequality once more we find

$$\begin{aligned}\left|\mathrm{Tr}\left(e^{-\lambda H_0} X e^{-(1-\lambda)H_Y}\right)\right| &\le cc_1c_2\delta^{-1}\left\|e^{-\lambda H_0}\right\|_{1/\lambda}\left\|e^{-(1-\lambda)(1-\alpha)H_Y}\right\|_{1/(1-\lambda)} \\ &\le K\alpha^{-1} Z_0^{\lambda}\left\|e^{-(1-\alpha)H_Y}\right\|_1^{1-\lambda},\end{aligned} \tag{4.52}$$

which is finite, since $1-\alpha > \beta_Y$, and can be made into a bound independent of λ.

Corollary 4.3.7 *The partition function $Z_X = Tre^{-(H_0+X)}$ is Lipschitz continuous on $\mathcal{M}_0$.*

Proof: From Duhamel's formula we have

$$\mathrm{Tr}e^{-H_0} - \mathrm{Tr}e^{-H_X} = \mathrm{Tr}\int_0^1 e^{-\lambda H_0} X e^{-(1-\lambda)H_X} d\lambda \tag{4.53}$$

Using lemma 4.3.6 with $X = Y$, we see that the trace of the integrand is a bounded function of $\lambda \in (0,1)$ so the integral of the trace is absolutely convergent. Since the trace is the sum of the diagonal elements in any orthogonal basis we conclude, from Fubini's theorem, that the trace and the integral can be exchanged. Therefore

$$\mathrm{Tr}e^{-H_0} - \mathrm{Tr}e^{-H_X} = \int_0^1 \mathrm{Tr}\left(e^{-\lambda H_0} X e^{-(1-\lambda)H_X}\right) d\lambda, \tag{4.54}$$

which gives that

$$\begin{aligned} |Z_0 - Z_X| &\leq \int_0^1 \left|\mathrm{Tr}\left(e^{-\lambda H_0} X e^{-(1-\lambda)H_X}\right)\right| d\lambda \\ &\leq C\|X\|_\varepsilon(0) \end{aligned} \tag{4.55}$$

In order to obtain more regularity for the partition function, and thence for the free energy, Streater was led to consider operator bounded perturbations [63]. The following result is the ε-bounded version of [63, theorem 2.3], his first success in this direction. We say that Y is an ε-bounded direction in $\mathcal{M}_X$ if $\|Y\|_\varepsilon(X) < \infty$.

Theorem 4.3.8 *The free energy Ψ_X is Fréchet differentiable and its differential in an ε-bounded direction Y is*

$$D\Psi_X(Y) = -\rho_X \cdot Y \tag{4.56}$$

Proof: Since $\Psi = \log Z$, all we need to prove is that

$$Z'(X)Y = -\mathrm{Tr}\left(e^{-\alpha H_X} Y e^{-(1-\alpha)H_X}\right). \tag{4.57}$$

From the definition of the Gatêaux derivative, consider the difference

$$\mathrm{Tr}e^{-(H_X+tY)} - \mathrm{Tr}e^{-H_X} + t\mathrm{Tr}\left(e^{-\alpha H_X} Y e^{-(1-\alpha)H_X}\right). \tag{4.58}$$

If t is small enough so that $\|tY\|_\varepsilon(X) < 1 - \beta_X$, we can use the same argument as in corollary 4.3.7 to obtain

$$-\int_0^1 \mathrm{Tr}\left(e^{-\alpha H_X} tY e^{-(1-\alpha)(H_X+tY)}\right) d\alpha + \mathrm{Tr}\left(e^{-\alpha H_X} tY e^{-(1-\alpha)H_X}\right). \tag{4.59}$$

Now since the second term above is independent of α, we can integrate it with respect to $0 < \alpha < 1$, so that the two terms together lead to the difference

$$
\begin{aligned}
& \int_0^1 \mathrm{Tr}\left[e^{-\alpha H_X} tY\left(e^{-(1-\alpha)H_X} - e^{-(1-\alpha)(H_X+tY)}\right)\right] d\alpha \\
= & \int_0^1 \mathrm{Tr}\left[e^{-\alpha H_X} tY \int_0^1 e^{-(1-\alpha)\lambda H_X}(1-\alpha)tYe^{-(1-\alpha)(1-\lambda)(H_X+tY)} d\lambda\right] d\alpha \\
= & \int_0^1\int_0^1 (1-\alpha)\mathrm{Tr}\left[e^{-\alpha H_X} tYe^{-(1-\alpha)\lambda H_X} tYe^{-(1-\alpha)(1-\lambda)(H_X+tY)}\right] d\lambda d\alpha \quad (4.60)
\end{aligned}
$$

by a second application of Duhamel's formula. Since this is clearly $o(t)$ and because $\rho_X \cdot Y$ is linear and continuous in Y, we conclude that Ψ is Gatêaux differentiable with differential at X in the ε-bounded direction Y given by 4.56. The final argument, proving that it is in fact Fréchet differentiable, will be given in the section 4.3.3.

4.3.2 The BKM Scalar Product

Before computing the second derivative of the free energy we prove the following result, which is a generalised version of [63, corollary to theorem 2.2] valid for ε-bounded perturbations

Theorem 4.3.9 *Suppose that $\rho_0 \in \mathcal{M}$ and that X is an ε-bounded perturbation of H_0. Then $\rho_0 H_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}$ is an operator of trace class and*

$$
Tr\left(\rho_0 H_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}\right) = Tr\left(\rho_0^{\lambda} X \rho_0^{1-\lambda}\right), \quad (4.61)
$$

for all $\lambda \in (0,1)$.

Proof: Let us choose $\lambda \in (0,1)$ such that $1 - 2\lambda > \beta_0$. Then we have that

$$
\begin{aligned}
\left\|\rho_0^{1-\lambda} X R_0^{\frac{1}{2}-\varepsilon}\right\|_1 &= \left\|\rho_0^{1-2\lambda}\rho_0^{\lambda} H_0^{\frac{1}{2}+\varepsilon} R_0^{\frac{1}{2}+\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}\right\|_1 \\
&\leq \left\|\rho_0^{1-2\lambda}\right\|_1 \left\|\rho_0^{\lambda} H_0^{\frac{1}{2}+\varepsilon}\right\| \left\|R_0^{\frac{1}{2}+\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}\right\| \\
&< \infty. \quad (4.62)
\end{aligned}
$$

Therefore, $\rho_0^{1-\lambda} X R_0^{\frac{1}{2}-\varepsilon}$ is of trace class. If we now write

$$
\rho_0 H_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}-\varepsilon} = \rho_0^{\lambda} H_0^{\frac{1}{2}-\varepsilon} \rho_0^{1-\lambda} X R_0^{\frac{1}{2}-\varepsilon}, \quad (4.63)
$$

we obtain that $\rho_0 H_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}$ is also of trace class, since $\rho_0^\lambda H_0^{\frac{1}{2}-\varepsilon}$ is bounded. Moreover, from cyclicity of the trace, it follows that

$$\begin{aligned} \mathrm{Tr}\left(\rho_0 H_0^{\frac{1}{2}-\varepsilon} X R_0^{\frac{1}{2}-\varepsilon}\right) &= \mathrm{Tr}\left(\rho^\lambda H_0^{\frac{1}{2}-\varepsilon} \rho_0^{1-\lambda} X R_0^{\frac{1}{2}-\varepsilon}\right) \\ &= \mathrm{Tr}\left(\rho_0^{1-\lambda} X \rho_0^\lambda\right), \end{aligned} \tag{4.64}$$

for one and hence for all $\lambda \in (0,1)$, due to lemma 4.2.3.

In what follows, a centred ε-bounded direction Y is an ε-bounded direction for which $\rho_X \cdot Y = 0$.

Theorem 4.3.10 *The free energy Ψ_X is twice Fréchet differentiable and its differential in the centred ε-bounded directions V_1, V_2 is*

$$D^2\Psi_X(V_1, V_2) = \int_0^1 Tr\left(H_X^{\frac{1}{2}-\varepsilon} \rho_X^\alpha V_2 \rho_X^{1-\alpha} V_1 R_X^{\frac{1}{2}-\varepsilon}\right) d\alpha \tag{4.65}$$

Proof: Since $Z = e^\Psi$, we have that $Z'' = \Psi'' Z + (\Psi')^2 Z$ and since V_1, V_2 are centred, all we need to prove is that

$$Z''(X)(V_1, V_2) = \int_0^1 \mathrm{Tr}\left(H_X^{\frac{1}{2}-\varepsilon} e^{-\alpha H_X} V_2 e^{-(1-\alpha)H_X} V_1 R_X^{\frac{1}{2}-\varepsilon}\right) d\alpha. \tag{4.66}$$

We now use theorem 4.3.9 to rewrite (4.57) as

$$Z'(X)V_1 = -\mathrm{Tr}\left(e^{-\alpha H_X} V_1 e^{-(1-\alpha)H_X}\right) = -\mathrm{Tr}\left(H_X^{\frac{1}{2}-\varepsilon} e^{-H_X} V_1 R_X^{\frac{1}{2}-\varepsilon}\right). \tag{4.67}$$

In order to lighten the notation, in the rest of this proof let us use $A = H_X^{\frac{1}{2}-\varepsilon}$. From the definition of the second derivative, we have to evaluate the difference

$$\mathrm{Tr}\left[A\left(e^{-H_X} - e^{-(H_X + tV_2)}\right) V_1 A^{-1}\right] - t\int_0^1 \mathrm{Tr}\left(A e^{-\alpha H_X} V_2 e^{-(1-\alpha)H_X} V_1 A^{-1}\right) d\alpha.$$

Using Duhamel's formula twice, this leads to

$$\int_0^1 \operatorname{Tr}\left(Ae^{-\alpha H_X}tV_2e^{-(1-\alpha)(H_X+tV_2)}V_1A^{-1}\right)d\alpha \tag{4.68}$$

$$-\int_0^1 \operatorname{Tr}\left(Ae^{-\alpha H_X}tV_2e^{-(1-\alpha)H_X}V_1A^{-1}\right)d\alpha$$

$$= -\int_0^1 \operatorname{Tr}\left[Ae^{-\alpha H_X}tV_2\left(e^{-(1-\alpha)H_X}-e^{-(1-\alpha)(H_X+tV_2)}\right)V_1A^{-1}\right]d\alpha$$

$$= \int_0^1 \operatorname{Tr}\left[Ae^{-\alpha H_X}tV_2\int_0^1 e^{-(1-\alpha)\lambda H_X}(\alpha-1)tV_2e^{-(1-\alpha)(1-\lambda)(H_X+tV_2)}V_1A^{-1}d\lambda\right]d\alpha$$

$$= \int_0^1\int_0^1(\alpha-1)\operatorname{Tr}\left[Ae^{-\alpha H_X}tV_2e^{-(1-\alpha)\lambda H_X}tV_2e^{-(1-\alpha)(1-\lambda)(H_X+tV_2)}V_1A^{-1}\right]d\lambda d\alpha.$$

which is clearly $o(t)$. The fact that $\int_0^1 \operatorname{Tr}\left(A\rho_X^{\alpha}V_2\rho_X^{1-\alpha}V_1A^{-1}\right)d\alpha$ is continuous in V_1, V_2 will be established in the next section. Since this is bilinear we can conclude that Ψ is twice Gatêaux differentiable with differential at X in the centred ε-bounded directions V_1, V_2 given by 4.65. The conclusion of the proof, showing that this is in fact a Fréchet derivative, will also follow from the main theorem of the next section.

Definition 4.3.11 The generalised *BKM* scalar product in the tangent space $\widehat{\mathcal{T}}_\varepsilon(X)$ is given by

$$\langle V_1, V_2\rangle_{\rho_X} := D^2\Psi_X(V_1,V_2) = \int_0^1 \operatorname{Tr}\left(H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha}V_2\rho_X^{1-\alpha}V_1R_X^{\frac{1}{2}-\varepsilon}\right)d\alpha, \tag{4.69}$$

for all $V_1, V_2 \in \widehat{\mathcal{T}}_\varepsilon(X)$.

We see that it reduces to the usual *BKM* formula when $\varepsilon = 1/2$, that is, when we are dealing with operator-bounded perturbations. From the previous theorem, we have that the generalised *BKM* scalar product is a continuous bilinear form on the tangent space. It also follows from the general theory of calculus on Banach spaces that it is symmetric, being a second Fréchet derivative [70, problem 4.3].

4.3.3 Higher Order Derivatives and the Taylor Series

In this section we show that $\Psi_X \equiv \Psi(\rho_X)$ has Fréchet derivative of all orders and that its Taylor series converges for sufficiently small neighbourhoods of ρ_X in $\mathcal{M}$.

The main result is the following.

Theorem 4.3.12 *Let $\alpha_n = 1 - \sum_{i=1}^{n-1} \alpha_i$. Then the map*

$$\int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \cdots \int_0^1 d\alpha_{n-1}\, Tr\left[H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha_1} V_1 \rho_X^{\alpha_2} V_2 \cdots \rho_X^{\alpha_n} V_n R_X^{\frac{1}{2}-\varepsilon}\right] \tag{4.70}$$

is a bounded multilinear functional on $\mathcal{T}_\varepsilon(X) \times \mathcal{T}_\varepsilon(X) \times \cdots \times \mathcal{T}_\varepsilon(X)$.

Proof: We begin by estimating the trace norm of $\left[H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha_1} V_1 \rho_X^{\alpha_2} V_2 \cdots \rho_X^{\alpha_n} V_n R_X^{\frac{1}{2}-\varepsilon}\right]$ as written as

$$[\rho_X^{\alpha_1\beta_X}][H_X^{1-\delta_n+\delta_1}\rho_X^{(1-\beta_X)\alpha_1}][R_X^{\delta_1}V_1R_X^{1-\delta_1}][\rho_X^{\alpha_2\beta_X}][H_X^{1-\delta_1+\delta_2}\rho_X^{(1-\beta_X)\alpha_2}]$$
$$[R_X^{\delta_2}V_2R_X^{1-\delta_2}]\cdots[\rho_X^{\alpha_n\beta_X}][H_X^{1-\delta_{n-1}+\delta_n}\rho_X^{(1-\beta_X)\alpha_n}][R_X^{\delta_n}V_nR_X^{1-\delta_n}],$$

with $\delta_n = \frac{1}{2}+\varepsilon$ and $\delta_j \in \left[\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon\right]$ for $j = 1, \ldots, n-1$ to be specified soon. In this product, we have n factors of the form $[\rho_X^{\alpha_j\beta_X}]$, n factors of the form $[R_X^{\delta_j}V_jR_X^{1-\delta_j}]$, and n factors of the form $[H_X^{1-\delta_{j-1}+\delta_j}\rho_X^{(1-\beta_X)\alpha_j}]$, with δ_0 standing for δ_n.

For the factors $[\rho_X^{\alpha_j\beta_X}]$, putting $p_j = 1/\alpha_j$, Hölder's inequality leads to the trace norm bound

$$\left\|[\rho_X^{\alpha_1\beta_X}]\cdots[\rho_X^{\alpha_n\beta_X}]\right\|_1 \leq \left\|\rho_X^{\beta_X}\right\|_1^{\alpha_1}\cdots\left\|\rho_X^{\beta_X}\right\|_1^{\alpha_n} = \left\|\rho_X^{\beta_X}\right\|_1 < \infty. \tag{4.71}$$

By virtue of lemma 4.1.3, we know that the factors $[R_X^{\delta_j}V_jR_X^{1-\delta_j}]$ are bounded in operator norm by

$$\left\|R_X^{\delta_j}V_jR_X^{1-\delta_j}\right\| \leq \left\|R_X^{\frac{1}{2}-\varepsilon}V_jR_X^{\frac{1}{2}+\varepsilon}\right\| = \|V_j\|_\varepsilon(X) < \infty. \tag{4.72}$$

In both these cases, the bounds are independent of α. The hardest case turns out to be the factors $[H_X^{1-\delta_{j-1}+\delta_j}\rho_X^{(1-\beta_X)\alpha_j}]$, where the estimate, as we will see, does depend on α and we have to worry about integrability. For them, the spectral theorem gives the operator norm bound

$$\begin{aligned}\left\| H_X^{1-\delta_{j-1}+\delta_j} \rho_X^{(1-\beta_X)\alpha_j} \right\| &= Z_X^{-\alpha_j(1-\beta_X)} \sup_{x\geq 1} \left\{ x^{1-\delta_{j-1}+\delta_j} e^{-(1-\beta_X)\alpha_j x} \right\} \\ &\leq \; Z_X^{-\alpha_j(1-\beta_X)} \left(\frac{1-\delta_{j-1}+\delta_j}{(1-\beta_X)\alpha_j} \right)^{1-\delta_{j-1}+\delta_j} e^{-(1-\delta_{j-1}+\delta_j)}. \end{aligned} \tag{4.73}$$

Apart from $\alpha_j^{-(1-\delta_{j-1}+\delta_j)}$, the other terms in (4.73) will be bounded independently of α. To deal with the integral of $\alpha_j^{-(1-\delta_{j-1}+\delta_j)} d\alpha_j$, we divide the region of integration in n (overlapping) regions $S_j := \{\alpha : \alpha_j \geq 1/n\}$ (since $\sum \alpha_j = 1$). For the region S_n, for instance, the integrability at $\alpha_j = 0$ is guaranteed if we choose δ_j such that $\delta_j < \delta_{j-1}$. So we take $\delta_n = \delta_0 > \delta_1 > \cdots > \delta_{n-1}$. We must have $\delta_j \in \left[\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right]$, so we choose

$$\begin{aligned} \delta_n &= \frac{1}{2} + \varepsilon, \\ \delta_1 &= \frac{1}{2} + \varepsilon - \frac{2\varepsilon}{n}, \\ \delta_2 &= \frac{1}{2} + \varepsilon - \frac{4\varepsilon}{n}, \\ &\vdots \\ \delta_{n-1} &= \frac{1}{2} - \varepsilon + \frac{2\varepsilon}{n}. \end{aligned}$$

Then each of the $(n-1)$ integrals, for $j = 1, \ldots, n-1$, is

$$\int_0^1 \alpha_j^{-(1-\delta_{j-1}+\delta_j)} d\alpha_j = (\delta_{j-1} - \delta_j)^{-1} = \frac{n}{2\varepsilon} \tag{4.74}$$

resulting in a contribution of $\left(\frac{n}{2\varepsilon}\right)^{n-1}$. The last integrand in S_n is

$$\alpha_n^{-(1-\delta_{n-1}+\delta_n)} \leq n^2. \tag{4.75}$$

The same bound holds for the other regions $S_j, j = 1, \ldots, n-1$, giving a total bound

$$\prod_{j=1}^{n} \int_0^1 \alpha_j^{-(1-\delta_{j-1}+\delta_j)} d\alpha_j \leq n \left[\frac{n^2 n^{n-1}}{(2\varepsilon)^{n-1}} \right] = \frac{n^2 n^n}{(2\varepsilon)^{n-1}}. \tag{4.76}$$

Now that we have fixed δ_j, the promised bound for the other terms in (4.73) is

$$\begin{aligned} \prod_{j=1}^{n} \quad & Z_X^{-\alpha_j(1-\beta_X)} \left(\frac{1-\delta_{j-1}+\delta_j}{1-\beta_X} \right)^{1-\delta_{j-1}+\delta_j} e^{-(1-\delta_{j-1}+\delta_j)} \\ \leq \; & 4 Z_X^{-(1-\beta_X)} (1-\beta_X)^{-n} e^{-n} \end{aligned} \tag{4.77}$$

since $(1-\delta_{j-1}+\delta_j)<1$ except for one term, when it is less than 2.

Collecting the estimates (4.71),(4.72),(4.76) and (4.77), we get the following bound for the n-point function (4.70)

$$\begin{aligned}
&\left|\int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \cdots \int_0^1 d\alpha_{n-1} \mathrm{Tr}\left(H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha_1}V_1\rho_X^{\alpha_2}V_2\cdots\rho_X^{\alpha_n}V_n R_X^{\frac{1}{2}-\varepsilon}\right)\right| \\
&\quad\le \int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \cdots \int_0^1 d\alpha_{n-1}\left|\mathrm{Tr}\left(H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha_1}V_1\rho_X^{\alpha_2}V_2\cdots\rho_X^{\alpha_n}V_n R_X^{\frac{1}{2}-\varepsilon}\right)\right| \\
&\quad\le \int_0^1 d\alpha_1 \int_0^1 d\alpha_2 \cdots \int_0^1 d\alpha_{n-1}\left\|H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha_1}V_1\rho_X^{\alpha_2}V_2\cdots\rho_X^{\alpha_n}V_n R_X^{\frac{1}{2}-\varepsilon}\right\|_1 \\
&\quad\le 4\left\|\rho_X^{\beta_X}\right\|_1 Z_X^{-(1-\beta_X)}(2\varepsilon)n^2 n^n e^{-n}\prod_j\left[\frac{\|V_j\|_\varepsilon(X)}{2\varepsilon(1-\beta_X)}\right]. \qquad (4.78)
\end{aligned}$$

Corollary 4.3.13 *The free energy Ψ_X is infinitely often Fréchet differentiable and has a convergent Taylor expansion for sufficiently small neighbourhoods of ρ_X is $\mathcal{M}$.*

Proof: We begin by completing the proofs from the last two sections. Applying theorem 4.3.12 to the case $n=2$ tells us that

$$\int_0^1 \mathrm{Tr}\left(H_X^{\frac{1}{2}-\varepsilon}\rho_X^{\alpha}V_2\rho_X^{1-\alpha}V_1 R_X^{\frac{1}{2}-\varepsilon}\right)d\alpha \qquad (4.79)$$

is a bounded bilinear functional on $\mathcal{T}_\varepsilon(X)\times\mathcal{T}_\varepsilon(X)$. This in turn was the only argument missing in the proof that Ψ had a second Gatêaux derivative given by (4.65). Moreover, a simple modification of the proof used for theorem 4.3.12 shows that the limit as $t\to 0$ in (4.68) is uniform in V_2. Using lemma 4.3.3 we obtain that (4.65) is indeed the Fréchet derivative of Ψ' at X. But this means that Ψ' is continuous at X [70, proposition 4.8] and lemma 4.3.3 allows us to conclude that Ψ has actually a Fréchet derivative at X given by (4.56).

Successive applications of Duhamel's formula show that the n-th variation of the partition function Z_X in the centred ε-bounded directions $V_1, V_2, \ldots, V_n$ is given by (4.70). The main theorem of this section then shows that it is a multilinear continuous functional. Hence Z has an n-th Gatêaux derivative at X. Since this holds for any n, we see that Z is infinitely often Gatêaux differentiable at X.

Moreover, we actually find that the limit procedure is uniform in V_i, hence the Gatêaux derivatives of Z at X are, in fact, Fréchet derivatives.

Therefore, Z is infinitely often Fréchet differentiable and it follows from the bound (4.78) that it has a convergent Taylor expansion for $Z_{(X+V)}$ provided that

$$\|V\|_{\varepsilon}(X) < (1 - \beta_X)2\varepsilon \tag{4.80}$$

Notice that this condition is stronger than to require that ρ_{V+X} lie in an ε-hood of ρ_X

Since Z_X is positive, the same is true for its logarithm, the free energy Ψ_X.

Chapter 5

Applications to Fluid Dynamics

Our interest in the mathematical development of Information Geometry in its most diverse forms - parametric, nonparametric, classical or quantum - is inspired by applications to Statistical Physics, as hinted in section 3.5. To be more specific, we believe that Information Geometry provides a mathematically sound foundation for some of the methods used in *Statistical Dynamics* , a general framework for obtaining models for nonequilibrium thermodynamics [61, 57]. A full account of this circle of ideas has already been enough to fill a book [56], so we will not describe them comprehensively here. What we can do now is quote its most significant features, loosely explaining their relations with Information Geometry. We follow it by a more illumitating description of one particular case and then apply the methods to a concrete problem [19].

In the classical version, the theory starts with the specification of a sample space Ω with a measurability structure, viewed as the possible configurations for the physical system, therefore being as much a part of the model as anything else that follows. The states are the probability measures on Ω, denoted by $\Sigma(\Omega)$, not the sample points themselves. For each measure μ satisfying the conditions described in chapter 2, the subset $\mathcal{M} \subset \Sigma(\Omega)$ of all μ-almost everywhere strictly positive probability densities relative to μ has the structure of an exponential Banach manifold.

An entropy functional (the von Neumann entropy) is defined on $\Sigma(\Omega)$. The observables of the theory are the commutative algebra of random variables on Ω. For each $\mathcal{M}$, we learn from chapter 2 that those observables belonging to $M^{\Phi_1}(p)$ are tangent vectors for the connected component of $\mathcal{M}$ containing p. Any n-dimensional set $\mathcal{X}$ of linearly independent observables in $M^{\Phi_1}(p)$ gives rise to a parametric exponential family $\mathcal{S}_n \subset \mathcal{M}$, the Gibbs state associated with $\mathcal{X}$, that is, a $+1$-flat submanifold of $\mathcal{M}$. Under certain assumptions (a sufficient one being a finite sample space), it can be shown that those are the states that maximise the von Neumann entropy subject to the constraint that the means of all random variables in $\mathcal{X}$ are kept constant. The dynamics is also probabilistic. One part of it is given by a one-parameter semigroup of bistochastic maps T^t (or, in the discrete time case, by an iterated bistochastic map T) from the algebra of random variables to itself (or by its dual action on the set of states), instead of the classical Hamiltonian dynamics that would be followed by, say, particles occupying well defined positions in the sample space. The map T is particular to each model and generally takes a point $p \in \mathcal{S}_n$ to a point $Tp \in \mathcal{M}$ not necessarily in $\mathcal{S}_n$. The principle of maximum entropy is then introduced into the theory by requiring that a thermalising map $Q : \mathcal{M} \to S_n$ should follow T at each time step of the dynamics.

It is here that Information Geometry makes its full appearance, because Q is defined to be the projection of Tp onto $\mathcal{S}_n$ following a -1-geodesic in $\mathcal{M}$ which intercepts $\mathcal{S}_n$ at right angles with respect to the Fisher metric. As anticipated in section 3.5, that this is an implementation of the principle of maximum entropy rests upon the fact that the relative entropy, or Kullback-Leibler information, is the statistical divergence associated with the dualistic triple $(g, \nabla^{(1)}, \nabla^{(-1)})$, a well established result in the parametric case (both classical and quantum) but yet to be proved in the fully nonparametric context. The time evolution then performs an orbit in $\mathcal{S}_n$ and one can try to use techniques such as the Lyapunov method to show that the system approaches an equilibrium state given by a fixed point of the dynamics. Unsurprisingly, a proof for this kind of result in full generality is hopelessly out of reach, but Information Geometry provides the geometrical framework for the cases where it can be done.

Most of the same concepts have parallels in the quantum version of the theory. We then start with a noncommutative local algebra of observables, the states on it (positive normalised functionals) and the relevant definitions of quantum entropy and quantum bistochastic maps. As we can see, dealing with different sample spaces in the classical case - finite, countable, uncountable but with a topological structure, etc - or with different noncommutaive algebras of observables in the quantum case - matrices, unbounded operators on Hilbert spaces, special types of C^*-algebras, etc - provides a driving force for the development of Information Geometry in its several variants, some of them discussed in this thesis.

5.1 Finite Sample Spaces

For finite sample spaces, all the necessary results in both Statistical Dynamics and Information Geometry have been duly proved, and the links between the two theories can be rigorously established. We review them in this section.

Let Λ be a finite discrete set, deemed to represent positions in space (a finite subset of a lattice, or of a tree, or of a graph, etc). To each point in $x \in \Lambda$, we associate a finite sample space Ω_x, representing the possible configurations of the system at x. The local structure of the theory is obtained by taking the total sample space to be the product space

$$\Omega = \prod_{x\in\Lambda} \Omega_x. \tag{5.1}$$

We then have $|\Omega| = N$, say. This definition implies, for instance, that if Λ_1, Λ_2 are disjoint subsets of Λ, then

$$\Omega(\Lambda_1 \cup \Lambda_2) = \Omega(\Lambda_1) \times \Omega(\Lambda_2). \tag{5.2}$$

We can provide Ω with a trivial measurability and topological structure, whereby all its subsets are measurable and open. The probability measures on Ω form then the $(N-1)$–simplex

$$\Sigma(\Omega) = \left\{ p : \Omega \to \mathbb{R} : p(\omega) \geq 0 \text{ and } \sum_{\omega\in\Omega} p(\omega) = 1 \right\}. \tag{5.3}$$

If we choose μ to be the counting measure on Ω, then the set $\mathcal{M}$ of chapter 2 reduces to the n-dimensional manifold

$$\mathcal{S} = \left\{ p : \Omega \to \mathbb{R} : p(\omega) > 0 \text{ and } \sum_{\omega\in\Omega} p(\omega) = 1 \right\}. \tag{5.4}$$

The random variables in this case are just the N-dimensional algebra of real valued functions on Ω, denoted by $\mathcal{A}$. It follows that, for each $p \in \mathcal{S}$, we have $\log p \in \mathcal{A}$, so there exist real numbers $\{\theta^1, \ldots, \theta^n, \Psi\}$ such that

$$\log p = \theta^1 X_1 + \cdots + \theta^n X_n - \Psi \mathbf{1}, \tag{5.5}$$

that is,

$$p = \exp\left(\theta^1 X_1 + \cdots + \theta^n X_n - \Psi \mathbf{1}\right), \tag{5.6}$$

where Ψ is determined by the normalisation condition $\sum_\omega p(\omega) = 1$. Therefore, $\theta = (\theta^1, \ldots, \theta^n)$ form a $+1$-affine coordinate system for $\mathcal{S}$, so that $\mathcal{S}$ is $\nabla^{(1)}$-flat. The tangent space $T_p\mathcal{S}$ is then identified with the n-dimensional subspace of $\mathcal{A}$ spanned by the scores $\left\{\frac{\partial \log p}{\partial\theta^1}, \ldots, \frac{\partial \log p}{\partial\theta^n}\right\}$. A similar argument to that given in chapter 3 shows that

$$\eta_i = \sum_\omega p(\omega) X_i(\omega) \tag{5.7}$$

defines a -1-affine coordinate system for $\mathcal{S}$ and we find that $\mathcal{S}$ is $\nabla^{(-1)}$-flat as well.

The von Neumann entropy is the functional on $\Sigma(\Omega)$ defined by

$$S(p) := -\sum_\omega p(\omega) \log p(\omega), \tag{5.8}$$

and the relative (Kullback-Leibler) entropy of the state p given the state q is

$$S(p|q) = \sum_\omega p \log \frac{p}{q}. \tag{5.9}$$

The same discussion presented in section 3.5 applies here. Namely, choosing a set of $m \leq n$ observables $Y_1, \dots, Y_m$ such that $\{\mathbf{1}, Y_1, \dots, Y_m\}$ is a linearly independent subset of $\mathcal{A}$, the states which maximise the von Neumann entropy subject to keeping the means of all $\{Y_i\}$, $i = 1, \dots, m$, constant are the Gibbs states of the form

$$p = \exp\left(\theta^1 Y_1 + \cdots + \theta^m Y_m - \psi \mathbf{1}\right), \tag{5.10}$$

where $\psi(\theta)$ is again determined by the normalisation condition $\sum_\omega p(\omega) = 1$. The states (5.10) form a $\nabla^{(1)}$-flat, m-dimensional, submanifold $\mathcal{S}_m \subset \mathcal{S}$.

We now move to the theme of defining a dynamics for the system. Assuming time to be discrete, the full dynamics will ultimately consist of both a linear and a nonlinear part for each time step. The linear part is the action of a bistochastic map on states. A *stochastic* map in this context is a linear map $P : \mathcal{A} \to \mathcal{A}$ which is positivity preserving (that is, it maps the positive cone of $\mathcal{A}$ to itself) and such that $P1 = 1$. Its dual action on the states is given by a linear map $P^* : \Sigma(\Omega) \to \Sigma(\Omega)$, that is, the restriction to $\Sigma(\Omega)$ of a linear map P^* on $\mathcal{A}$ which preserves both positivity and normalisation. When a basis is specified for $\mathcal{A}$, the map P is given by a positive matrix whose rows add to 1 and P^* by its transpose, that is, a positive matrix whose columns add to 1. A *bistochastic* map is a stochastic map P which is itself dual to a stochastic map. In other words, a stochastic map that also preserves normalisation of probabilities. Given a basis, a bistochastic map is represented by a positive matrix with both rows and columns adding to 1.

Each random variable on Ω divides it into disjoint *shells*. For instance, the energy function $\mathcal{E} : \Omega \to \mathbb{R}$ allows one to write Ω as the disjoint union of the energy shells

$$\Omega_E = \{\omega \in \Omega : \mathcal{E}(\omega) = E\}. \tag{5.11}$$

The dual of a general stochastic map does not necessarily increase the entropy of states, but if P is bistochastic, its dual map P^* does [56, theorem 3.18]. If the restriction of P^* to each energy shell leaves Ω_E invariant and is irreducible on it, the dynamics obtained by iterated application of P^* to any initial state converges to a mixture of microcanonical states, besides conserving the mean energy and increasing the entropy. More precisely [56, theorem 3.26], for any $p_0 \in \Sigma(\Omega)$, we have that $(P^*)^n p_0 \to p_\infty$ as $n \to \infty$, where

$$p_\infty = \sum_E p_0(E) \chi_E \tag{5.12}$$

with

$$p_0(E) = \sum_{\omega \in \Omega_E} p_0(w). \tag{5.13}$$

We see that this dynamics conserves the probability on each energy shell. The nonlinear part of the dynamics is brought in precisely to obtain equilibria states other than the above. We want to obtain Gibbs states, which are a particular mixture of microcanonical states but with fully mixed probabilities among the different shells, regardless of what they were for the initial state p_0. For our chosen set of slow variables $Y_1, \dots, Y_m$, we assume that at any given time the system is in a state $p \in \mathcal{S}_m$. The linear dynamics generally takes it to a point P^*p not in $\mathcal{S}_m$. We supplement it with a projection $Q : \mathcal{S} \to \mathcal{S}_m$, so that QP^*p has the same means as the state P^*p for all the slow variables and maximal von Neumann entropy subject to these constraints. For the case $|\Omega| < \infty$, Streater has rigorously proved [61] that this projection follows a -1-geodesic in $\mathcal{S}$ which cuts $\mathcal{S}_m$ perpendicularly with respect to the Fisher metric

$$g_{ij} := \sum_\Omega p(\omega) \frac{\partial \log p(\omega)}{\partial \theta^i} \frac{\partial \log p(\omega)}{\partial \theta^j}. \tag{5.14}$$

This in turn is equivalent to a well defined procedure in Information Geometry, involving minimisation of a statistical divergence.

The concept of a canonical statistical divergence for a general statistical manifold, parametric or not, classical or quantum, has not been properly introduced in this thesis yet, and now is an opportunity as good as any other to do so. The original parametric definition can be found in [2, p 61]. It is given for a Riemannian manifold $(\mathcal{M}, g)$ eqquiped with a pair of dual flat connections ∇ and ∇^*. The definition is given in terms of potential functions and dual coordinate systems, the existence of which being guaranteed by two of Amari's theorems [1, theorems 3.4 and 3.5]. However, neither potential functions nor dual coordinate systems seem to be fit for generalisations to infinite dimensions. The former are convex functions related by Legendre transforms, a concept difficult to apply for non-reflexive Banach spaces. The latter leads us to the problem that the putative dual coordinates to the exponential coordinate system used here are the so called expectation parameters, which do not constitute a proper chart in infinite dimensions [49].

Alternatively, we propose to generalise an equivalent, coordinate-free, definition of canonical statistical divergence [38, p 174]. Let $\langle\cdot,\cdot\rangle$ be a continuous scalar product on $\mathcal{M}$ and let ∇ be a globally flat connection on $T\mathcal{M}$ such that its dual connection with respect to $\langle\cdot,\cdot\rangle$ is also globally flat.

Definition 5.1.1 The *canonical statistical divergence* associated with $(\langle\cdot,\cdot\rangle,\nabla)$ is a function $D:\mathcal{M}\times\mathcal{M}\to\mathbb{R}$ such that, if $D_p:\mathcal{M}\to\mathbb{R}$ is the function defined by $D_p(q)=D(p,q)$, then

i. $\nabla dD_p=\langle\cdot,\cdot\rangle$,

ii. $dD_p(v)=0$, for all $v\in T_p\mathcal{M}$,

iii. $D_p(p)=0$.

We then have the following theorem [1, theorems 3.8 and 3.9]

Theorem 5.1.2 *If $\mathcal{M}$ is a ∇-flat finite dimensional manifold and $\mathcal{M}'$ is a ∇^*-convex submanifold of it, then given any point $p\in\mathcal{M}$, the unique point p' in $\mathcal{M}'$ which minimises the canonical statistical divergence $D(p|p')$ associated with $(\langle\cdot,\cdot\rangle,\nabla)$ is connected to p by a ∇-geodesic which is orthogonal to $\mathcal{M}'$ with respect to $\langle\cdot,\cdot\rangle$.*

It turns out that, for finite dimensional classical statistical manifolds, the canonical divergence associated with the Fisher metric and $\nabla^{(-1)}$ is the relative entropy (5.9), that is [1, pp 87-88],

$$D_{-1}(p, q) = S(p|q). \tag{5.15}$$

The map Q of Statistical Dynamics is then a special instance of theorem 5.1.2, since our $\mathcal{S}$ is $\nabla^{(-1)}$-flat and its submanifold $\mathcal{S}_m$ is $\nabla^{(1)}$-flat, hence $\nabla^{(1)}$-convex. The full dynamics traces an orbit in $\mathcal{S}_m$, and when P^* is irreducible on the generalised shells determined by the slow variables, it leads to convergence to a great grand canonical equilibrium state in $\mathcal{S}_m$.

5.2 Hydrodynamical Systems

The application we are going to present here consists of using the methods above to obtain the time evolution equations for conserved physical quantities when the system under consideration is a single fluid. We move the theory one step further by taking the formal continuum limit of the discrete equations, thus leading to partial differential equations to be compared with commonly used hydrodynamical equations, such as those of Euler and Navier-Stokes. The equations obtained from the different models we are about to quote have attracted the interest of at least one group of mathematicians [7, 6, 39], who are actively studying the challenging analytical problems they pose.

This line of study can be traced back to [59], where the equations for the density and temperature fields for a gas of Brownian particles moving under a one-dimensional external potential were obtained. The strategy there was to consider particles exchanging energy with the medium, assuming that the correlation between a particle and the temperature field is a fast variable. The slow variables were the density (number of particles at a site) and the fully thermalised energy present at the bonds between sites, so that the kinetic energy of the particles was not explicitly considered (the particle was assumed to be moving in its terminal velocity). The resulting coupled nonlinear equations (called nonlinear heat equations by Streater) were shown to satisfy both laws of thermodynamics and to reduce to the Smoluchowski equation (Brownian motion with drift) when the temperature field is constant [60]. A local existence theorem for its solutions was obtained for special initial data and regularity conditions on the derivatives of the potential [58].

In [62], the more ambitious programme of taking both the number of particles and the particle energy as slow varibles was initiated. The time evolution equations for the density and temperature fields of a one-dimensional fluid under an external potential were obtained and shown to satisfy the laws of thermodynamics. The model had, however, serious oversimplifications, especially the facts that the density of states was not realistic for three dimensions and the diffusion constant was indenpendent of the temperature. A modified version of the model, addressing both problems, was presented in [65]. It still considered only the number of particles and the energy as slow variables, but among its successes is the fact that the hydrodynamical equations for density and temperature exhibited the Soret and Dufour effects, without the need to introduce any inter-particle interaction other than a hard-core repulsion.

The limitations of dealing with only the number of particles and the energy as slow variables were lifted for the first time in [66], where a stochastic model of a fluid with five macroscopic conserved fields, the mass, energy and the three components of momentum, was described. The nonlinear coupled parabolic system obtained for the field equations showed corrections to the Navier-Stokes equations. In particular, the Euler continuity equation acquired a diffusion term, peculiar to the stochastic nature of the dynamics. Streater then argued that these equations are more stable and physically more accurate than the usual Navier-Stokes equations.

In this chapter, we take the model of [66] and put it in an external field Φ. We derive the full set of equations

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\mathbf{u}\rho) = \lambda \operatorname{div}\left(\rho^{-1}\nabla(\Theta^{1/2}\rho) + \frac{\nabla\Phi}{k_B\Theta^{1/2}}\right) \tag{5.16}$$

$$\begin{aligned}\frac{\partial (\rho e)}{\partial t} &+ \operatorname{div}\left[\mathbf{u}(\rho e + P)\right] = \lambda \operatorname{div}\ \Big[2\rho^{-1}\nabla\left(\Theta^{1/2}P\right) + \rho^{-1}\nabla\left(\Theta^{1/2}\rho\right)\phi \\ &+ \frac{\nabla\Phi}{k_B\Theta^{1/2}}\phi + 2\frac{\nabla\Phi}{k_B\Theta^{1/2}}P\Big] \end{aligned} \tag{5.17}$$

$$\begin{aligned}\frac{\partial \rho\mathbf{u}}{\partial t} &+ \operatorname{div}(\rho\mathbf{u}\otimes\mathbf{u}) = \rho\mathbf{f} - \nabla P + \lambda \operatorname{div}\left(\frac{\nabla\Phi}{k_B\Theta^{1/2}}\otimes\mathbf{u}\right) \\ &+ \frac{2\lambda}{5}\partial_i\rho^{-1}\left[3\partial^i\left(\rho(x)\Theta^{1/2}\mathbf{u}\right) + \nabla\left(\rho(x)\Theta^{1/2}u^i\right)\right]. \end{aligned} \tag{5.18}$$

They reduce to those of [66] when $\Phi = 0$, while if $\mathbf{u} = 0$ they reduce to [65] with suitable changes due to the different multiplicity of states in the two models.

5.3 Review of the model

5.3.1 The sample space, conserved quantities and information manifold

Let Λ be a finite subset of the cubic lattice $(a\mathbb{Z})^3$ with spacing $a \approx 10^{-8}$ cm, representing the size of the hard core of the fluid molecules. For a structureless monatomic fluid (e.g. argon) we choose

$$\Omega_x = \left\{\emptyset, (\epsilon\mathbb{Z})^3\right\}, \tag{5.19}$$

where $\epsilon \approx 6.6 \times 10^{-19}$ c.g.s. is the quantum of momentum of a particle confined to a region of size a^3 (see the discussion following equation (6) in [66]). As usual, take $\Omega = \prod_{x\in\Lambda}\Omega_x$. A point $\omega \in \Omega$ is thus the collection $\{\omega_x\}_{x\in\Lambda}$. If ω is such that $\omega_x = \emptyset$ for a certain $x \in \Lambda$ then the configuration ω has no particle at x; we also say that there is a hole, or vacancy, at x. If $\omega_x = \mathbf{k} \in (\epsilon\mathbb{Z})^3$, then in the configuration ω there is a particle at x and its momentum is $\mathbf{k}$.

The slow variables are the $5|\Lambda|$ random variables

$$\mathcal{N}_x(\omega) = \begin{cases} 0 & \text{if } \omega_x = \emptyset \\ 1 & \text{if } \omega_x = \mathbf{k} \in (\epsilon\mathbb{Z})^3 \end{cases} \tag{5.20}$$

$$\mathcal{E}_x(\omega) = \begin{cases} 0 & \text{if } \omega_x = \emptyset \\ (2m)^{-1}\mathbf{k}.\mathbf{k} + \Phi(x) & \text{if } \omega_x = \mathbf{k} \end{cases} \tag{5.21}$$

$$\boldsymbol{\mathcal{P}}_x(\omega) = \begin{cases} 0 & \text{if } \omega_x = \emptyset \\ \mathbf{k} & \text{if } \omega_x = \mathbf{k}, \end{cases} \tag{5.22}$$

where m is the mass of the molecule and Φ is a given real-valued potential, which could be time-dependent. The states in $\mathcal{S}_{5|\Lambda|} \subset \Sigma(\Omega)$ are those of the form $p = \prod_x p_x$, where

$$p_x = \Xi_x^{-1} \exp\left\{-\xi_x \mathcal{N}_x - \beta_x \mathcal{E}_x - \boldsymbol{\zeta}_x \cdot \boldsymbol{\mathcal{P}}_x\right\}. \tag{5.23}$$

Here, ξ_x, β_x and $\boldsymbol{\zeta}_x$ are the fields of $+1$-affine coordinates for $p \in \mathcal{S}_{5|\Lambda|}$, called its 'canonical' coordinates. The great grand partition function Ξ_x at each site is the normalising factor

$$\begin{aligned} \Xi_x &= \sum_{\omega_x \in \Omega_x} \exp\left\{-\xi_x \mathcal{N}_x - \beta_x \mathcal{E}_x - \boldsymbol{\zeta}_x \cdot \boldsymbol{\mathcal{P}}_x\right\} && (5.24) \\ &= 1 + e^{-\xi_x - \beta_x \Phi(x)} Z_1 Z_2 Z_3, && (5.25) \end{aligned}$$

where

$$Z_i = \sum_{k \in \epsilon\mathbb{Z}} \exp\left(-\frac{\beta_x k^2}{2m} - \zeta_x^i k\right). \tag{5.26}$$

This can be calculated explicitly if, due to the small parameter ϵ, we approximate sums by integrals:

$$\begin{aligned} Z_i &\approx \epsilon^{-1} \int_{-\infty}^{\infty} e^{-\frac{\beta_x k^2}{2m} - \zeta_x^i k} dk && (5.27) \\ &= \epsilon^{-1} \left(\frac{2m\pi}{\beta}\right)^{1/2} e^{m(\zeta^i)^2/(2\beta)}. && (5.28) \end{aligned}$$

The means $(N_x, E_x, \boldsymbol{\varpi}_x)$ of the slow variables in a state $p \in \mathcal{S}_{5|\Lambda|}$ are -1-affine coordinates for p, called its 'mixture' coordinates. From theorem 3.2.1 they are related to the canonical coordinates by a Legendre transform:

$$N_x = -\frac{\partial \log \Xi_x}{\partial \xi_x}, \qquad x \in \Lambda \tag{5.29}$$

$$E_x = -\frac{\partial \log \Xi_x}{\partial \beta_x}, \qquad x \in \Lambda \tag{5.30}$$

$$\varpi_x^i = -\frac{\partial \log \Xi_x}{\partial \zeta_x^i}, \qquad i = 1, 2, 3, \quad x \in \Lambda. \tag{5.31}$$

Using the explicit expression for the partition function, these can be used to deduce several equations relating the macroscopic variables in the theory. In particular, if we introduce the mean velocity field

$$\mathbf{u}_x = \frac{\boldsymbol{\varpi}_x}{mN_x}, \tag{5.32}$$

and the temperature field $\Theta_x = 1/(k_B \beta_x)$, it is straightforward to show that

$$\zeta^i = -\frac{\beta_x \varpi_x^i}{mN_x} = -\beta_x u_x^i \tag{5.33}$$

and

$$E_x = N_x \left(\Phi(x) + \frac{3}{2} k_B \Theta_x + \frac{1}{2} m \mathbf{u}_x \cdot \mathbf{u}_x \right). \tag{5.34}$$

The discrete nature of the model allows us to define the thermodynamical entropy as an appropriate multiple of the von Neumann entropy

$$S(p) := -k_B \sum_{\omega} p(\omega) \log p(\omega), \tag{5.35}$$

which for $p \in \mathcal{S}_{5|\Lambda|}$ gives

$$S = k_B \sum_{\mathbf{x} \in \Lambda} \left(\xi_{\mathbf{x}} N_{\mathbf{x}} + \beta_{\mathbf{x}} E_{\mathbf{x}} + \boldsymbol{\zeta}_{\boldsymbol{x}} \cdot \boldsymbol{\varpi}_{\boldsymbol{x}} + \log \Xi_{\mathbf{x}} \right) \tag{5.36}$$

An argument from equilibrium theory [66] then leads to the definition of the thermodynamical pressure as

$$P(\mathbf{x}) = a^{-3} k_B \Theta \log \Xi_{\mathbf{x}}. \tag{5.37}$$

If the ratio V_0/V between the smallest volume that the $N = \sum_x N_x$ particles can occupy, that is $V_0 = a^3 N$, and the total volume V is small, then the formula above for the pressure reduces to the perfect gas approximation

$$P_x = \frac{N_x k_B \Theta_x}{a^3} \tag{5.38}$$

The other macroscopic variables in terms of which the hydrodynamical equations are written are the mass density $\rho(x) = mN_x/a^3$ and the energy-density per unit of mass $e(x) = E_x/(mN_x)$. If we ignore the small term involving $\mathbf{u}_x \cdot \mathbf{u}_x$ in (5.34), we see that

$$e(x) = \frac{\Phi(x)}{m} + \frac{3k_B\Theta_x}{2m} := \phi(x) + \frac{3k_B\Theta_x}{2m}. \tag{5.39}$$

5.3.2 The hopping dynamics and the continuum limit

The linear part of the dynamics is specified by giving hopping rules. We require that P^* should couple only neighbouring points in Λ, where we consider two points to be neighbours if their distance along one of the lattice unit vectors is one mean free path, denoted by ℓ. We assume that ℓ is an integer multiple of the lattice spacing a, but allow it to depend on the local density by taking ℓ/a to be the nearest integer to

$$\frac{\rho_{max}}{\rho(x)} = \frac{m}{a^3\rho(x)}.$$

Suppose that $\omega \in \Omega$ is such that $x \in \Lambda$ is occupied. Consider in turn the possibility of jumping from x along the direction of the unit vectors $\pm e_i$ of the cubic lattice to an empty site $x' := x \pm \ell e_i$, $i = 1, 2, 3$. In the absence of an external potential [66], the jump will take a time $\ell/|v_x^i|$, where

$$\mathbf{v}_x := \mathbf{k}_x/m. \tag{5.40}$$

We then define the (random) hopping *rate* from x to $x + \ell e_i$ to be the inverse of this relaxation time, namely v_x^i/ℓ if $v_x^i \geq 0$ and zero if v_x^i is negative, in which case there is a rate $-v_x^i/\ell$ of hopping to $x - \ell e_i$. The situation in the presence of an external potential is a little more involved, because the potential causes a change in the velocity along the jump. From now on, assume for definiteness that $\Phi(x+\ell e_i) > \Phi(x)$. If the particle at x hops to $x+\ell e_i$, its potential energy increases to $\Phi(x+\ell e_i)$ and so its kinetic energy must decrease by the same amount. Its change in momentum is taken to be entirely in the direction of e_i. So its velocity in the i-direction, $v_x := k_x/m$ (we omit the index i for v and k in the following formulae since it is clear from the rest of the notation what is the component involved) is reduced to $v'_x := k'_x/m$, where

$$k'^2_x/(2m) = k^2_x/(2m) - \Phi(x+\ell e_i) + \Phi(x) = k^2_x/(2m) - \ell\partial_i\Phi(x) + O(\ell^2), \quad (5.41)$$

that is

$$k'_x = \left(k^2_x - 2m\ell\partial_i\Phi(x)\right)^{1/2}. \quad (5.42)$$

In order for the move to be energetically possible, we must have

$$k_x \geq \kappa^i_x := (2\ell m\partial_i\Phi(x))^{1/2}. \quad (5.43)$$

Similarly, if the particle at x with velocity $v_x < 0$ in the i-direction hops to $x - \ell e_i$, its potential energy decreases to $\Phi(x-\ell e_i)$, with a corresponding rise in its kinetic energy. Therefore (again taking the change in momentum to be entirely in the i-direction) its (negative) velocity in the i-direction becomes $v''_x = k''_x/m$, where

$$k''^2_x/(2m) = k^2_x/(2m) + \Phi(x) - \Phi(x-\ell e_i) = k^2_x/(2m) + \ell\partial_i\Phi(x-\ell e_i) + O(\ell^2), \quad (5.44)$$

that is

$$k''_x = -\left(k^2_x + 2m\ell\partial_i\Phi(x-\ell e_i)\right)^{1/2}. \quad (5.45)$$

We take the hopping rate from x to $x' = x \pm \ell e_i$ to be the average of the initial and final rates:

$$r(k_x) = \begin{cases} r_-(k_x) := -\frac{v_x+v''_x}{2m\ell} = \frac{-k_x+\left[k^2_x+(\kappa^i_{x-\ell e_i})^2\right]^{1/2}}{2m\ell}, & \text{if } k_x \leq 0; \\ r_+(k_x) := \frac{v_x+v'_x}{2m\ell} = \frac{k_x+\left[k^2_x-(\kappa^i_x)^2\right]^{1/2}}{2m\ell}, & \text{if } k_x \geq \kappa^i_x. \end{cases} \quad (5.46)$$

These hopping rates increase with k_x and there are infinitely many possible momentum states. To be a Markov chain, the sum of all rates out of a configuration must be less than one. For any k_x this can be achieved by choosing dt small enough. To do this for all k_x with a fixed dt we must put in a cut-off; there are no hops if $|k_x| > K_x$, say. Finally, $r(k_x)dt$ gives the probability of a transition in an interval dt provided that the site x is occupied and the site x' is empty, so the actual entries of the Markov matrix are conditional probabilities and the transition rate above should appear multiplied by factors of the form $N_x(1 - N_{x'})$. As argued in [66], we neglect N_x compared to 1, therefore leaving out the second term in the factors above.

The continuum limit we are going to take in order to obtain the hydrodynamical equations corresponds to $\ell \to 0$, $m \to 0$ such that the product ℓc remains finite and non-zero, where

$$c := \left(\frac{k_B \Theta_0}{m}\right)^{1/2} \tag{5.47}$$

is the approximate velocity of sound at the reference temperature Θ_0. The diffusion constant that appears when we take the limit is then predicted to be

$$\lambda := \frac{\ell c \rho}{(2\pi\Theta_0)^{1/2}} = \frac{a \rho_{max} c}{(2\pi\Theta_0)^{1/2}}. \tag{5.48}$$

5.4 Hydrodynamics in an external field

When a transition from x to $x + \ell e_i$ occurs in a potential Φ, the loss of mass and energy from the site x is equal to the gain at the site $x + \ell e_i$. This is not true of momentum; the loss at x differs from the gain at $x + \ell e_i$ by $\kappa_i := (2\ell m \partial_i \Phi)^{1/2}$. So we deal will $\mathcal{N}$ and $\mathcal{E}$ first.

Before we start the calculations, let us recall that integrals of the form

$$M_n(\zeta) = \int_0^\infty k^n \exp\{-\beta k^2/(2m) - \zeta k\} dk, \qquad n = 0, 1, 2, 3$$

were evaluated up to second order in ζ in Appendix 1 of [66]. For later use, we reproduce the results here up to zeroth order in ζ for $n = 0, 1$,

$$M_0(\zeta) = \left(\frac{\pi m}{2\beta}\right)^{1/2}, \tag{5.49}$$

$$M_1(\zeta) = \frac{m}{\beta}, \tag{5.50}$$

and to first order in ζ for $n = 2$,

$$M_2(\zeta) = \left(\frac{\pi}{2}\right)^{1/2}\left(\frac{m}{\beta}\right)^{3/2} - 2\left(\frac{m}{\beta}\right)^2 \zeta. \tag{5.51}$$

5.4.1 Dynamics of the mass-density in an external field

Since the field $\Phi(x)$ is external, it does not depend on the configuration ω_x of the random fields at x. It therefore cancels in the exponential states. The potential enters only in its supression or enhancement of the transition rate, in that the rate is the average of the initial and final rates. This shows up mainly in the appearance of a non-zero lower limit to the (positive) momentum for any right-going hop to be possible.

Let J^i_x/ℓ be the change in the value of N_x due to the hoppings occuring between x and $x - \ell e_i$ in such a way that the change due to exchanges with both $x \pm \ell e_i$ in an interval δt is

$$\delta_i N_x = -\frac{J^i_{x+\ell e_i} - J^i_x}{\ell}\delta t. \tag{5.52}$$

So the total change in N_x in an interval δt due to hoppings in all directions is

$$\delta N_x = (\delta_1 N_x + \delta_2 N_x + \delta_3 N_x)\delta t. \tag{5.53}$$

Using the hopping rates defined in the previous section, the loss/gain contribution to the particle current involving the exchange between x and $x - \ell e_i$ is

$$J^i_x = -\sum_{k^i \le 0} \ell r_-(k_x) p_x(\mathbf{k})\mathcal{N}_x(\mathbf{k}) + \sum_{k^i \ge \kappa^i} \ell r_+(k_{x-\ell e_i}) p_{x-\ell e_i}(k)\mathcal{N}_{x-\ell e_i}(k). \tag{5.54}$$

As in [66], the analysis of this expression is best handled by introducing a conditional probability $\bar{p}_x(\omega) = p(\omega|\mathcal{N}_x = 1)$ on the particle space $\Omega - \emptyset$, that is,

$$\bar{p}_x(\mathbf{k}) = (Z_1 Z_2 Z_3)^{-1} \exp\{-\beta_x |\mathbf{k}|^2/(2m) - \boldsymbol{\zeta}_x \cdot \mathbf{k}\}. \tag{5.55}$$

We now use the fact that $p_x(k) = N_x \bar{p}_x(k)$ and $\mathcal{N}_x(k) = 1$ on the particle space $\Omega - \emptyset$, replace the sums by integrals in (5.54), and add and subtract the term

$$F_x^i = N_x (Z_i \epsilon)^{-1} \int_{k \geq \kappa} \ell r_+(k) \exp\left\{ -\frac{\beta k^2}{2m} - \zeta k \right\} dk, \tag{5.56}$$

to obtain

$$\begin{aligned} J_x^i &= F_x^i - \frac{N_x}{Z_i \epsilon} \int_{k \leq 0} \ell r_-(k_x) \exp\left(-\frac{\beta k^2}{2m} - \zeta^i k \right) dk \\ & \quad -\ell \left(F_x^i - F_{x-\ell e_i}^i \right) / \ell. \end{aligned} \tag{5.57}$$

We start by calculating F_x^i. In the term

$$\int_{k \geq \kappa} \frac{k + (k^2 - \kappa^2)^{1/2}}{2m} \exp\left\{ -\frac{\beta k^2}{2m} - \zeta^i k \right\} dk,$$

we make the change of variable

$$k'^2 = k^2 - \kappa^2.$$

Then $k\, dk = k'\, dk'$, and the integral becomes

$$\int_{k' \geq 0} \frac{k' \left[(k'^2 + \kappa^2)^{1/2} + k' \right]}{2m (k'^2 + \kappa^2)^{1/2}} \exp\left\{ -\frac{\beta (k'^2 + \kappa^2)}{2m} - \zeta^i (k'^2 + \kappa^2)^{1/2} \right\} dk',$$

which can be written as

$$\begin{aligned} & \int_{k' \geq 0} \frac{k' \left[(k'^2 + \kappa^2)^{1/2} + k' \right]}{2m (k'^2 + \kappa^2)^{1/2}} \exp\left\{ -\frac{\beta k'^2}{2m} - \zeta^i k' \right\} \\ & \quad \times \exp\left\{ -\frac{\beta \kappa^2}{2m} - \left((k'^2 + \kappa^2)^{1/2} - k' \right) \zeta^i \right\} dk'. \end{aligned}$$

The arguments of the exponentials are small, and we expand them to first order:

$$\exp\left\{ -\frac{\beta \kappa^2}{2m} - \left[\left(k'^2 + \kappa^2 \right)^{1/2} - k' \right] \zeta^i \right\} = 1 - \frac{\beta \kappa^2}{2m} - \left[k' - \left(k'^2 + \kappa^2 \right)^{1/2} \right] \zeta^i.$$

This gives us the three terms

$$\int_{k' \geq 0} \frac{k' \left[(k'^2 + \kappa^2)^{1/2} + k' \right]}{2m (k'^2 + \kappa^2)^{1/2}} \exp\left\{ -\frac{\beta k'^2}{2m} - \zeta^i k' \right\} dk' \tag{5.58}$$

$$-\frac{\beta \kappa^2}{2m} \int_{k' \geq 0} \frac{k' \left[(k'^2 + \kappa^2)^{1/2} + k' \right]}{2m (k'^2 + \kappa^2)^{1/2}} \exp\left\{ -\frac{\beta k'^2}{2m} - \zeta^i k' \right\} dk' \tag{5.59}$$

$$-\frac{\kappa^2 \zeta^i}{2m} \int_{k' \geq 0} \frac{k'}{(k'^2 + \kappa^2)^{1/2}} \exp\left\{ -\frac{\beta k'^2}{2m} - \zeta^i k' \right\} dk'. \tag{5.60}$$

The dominant term is (5.58), in which we may replace the factor

$$\frac{k'\left[(k'^2+\kappa^2)^{1/2}+k'\right]}{2m(k'^2+\kappa^2)^{1/2}}$$

by the velocity when $\Phi = 0$, namely k'/m, with an error of $O(\ell\log\ell)$ [19, appendix 1]. So the contribution of this term to the mass current can be approximated in the limit by

$$\frac{N_x}{Z_i\epsilon}\int_0^\infty \frac{k'}{m}\exp\left\{-\frac{\beta k'^2}{2m}-\zeta^i k'\right\}dk. \tag{5.61}$$

Making the same replacement in (5.59), with the same error, we obtain

$$\begin{aligned} & -\frac{\beta\kappa^2}{2m}\int_0^\infty \frac{k'}{m}\exp\left\{-\frac{\beta k'^2}{2m}-\zeta^i k'\right\}dk \\ = & -\frac{\beta\kappa^2}{2m^2}M_1(\zeta^i). \end{aligned}$$

Therefore, the contribution coming from (5.59) to the mass current is

$$-\frac{N_x\beta\kappa^2}{2m^2 Z_i\epsilon}M_1(\zeta^i) = -\frac{N_x\beta\kappa^2}{2m^2}\left(\frac{\beta}{2\pi m}\right)^{1/2}e^{-m(\zeta^i)^2/2\beta}M_1(\zeta^i),$$

which, to zeroth order in ζ^i, gives

$$-N_x\left(\frac{\beta}{2\pi m}\right)^{1/2}\ell\partial_i\Phi(x) = -\frac{N_x}{k_B\Theta^{1/2}}\frac{\ell c}{2\pi\Theta_0}\partial_i\Phi(x) = -\frac{\lambda N_x}{\rho k_B\Theta^{1/2}}\partial_i\Phi(x). \tag{5.62}$$

When multiplied by m/a^3 this is what we call the Smoluchowski, or drift, current:

$$J_S^i = -\lambda\frac{\partial_i\Phi(x)}{k_B\Theta^{1/2}}. \tag{5.63}$$

The integral in (5.60) is bounded by

$$-\frac{\kappa^2\zeta^i}{2m}M_0(\zeta^i).$$

Expanding M_0 to zeroth order in ζ^i, this gives

$$-\frac{\zeta^i\kappa^2}{2m}\frac{1}{2}\left(\frac{2\pi m}{\beta}\right)^{1/2} = -\frac{\zeta^i\ell\partial_i\Phi(x)}{2}\left(\frac{2\pi m}{\beta}\right)^{1/2},$$

so that it can be ignored in the limit.

We now turn our attention to the second term in the current (5.57). We can again replace the factor ℓr_- by $|k^i|/m$ [19, appendix 2], so that the contribution to the mass current from this term is

$$\frac{N_x}{Z_i\epsilon}\int_{-\infty}^{0}\frac{k}{m}\exp\left\{-\frac{\beta k^2}{2m}-\zeta^i k\right\}dk. \tag{5.64}$$

When we combine this with (5.61) and multiply it all by m/a^3 what we find is simply ρu^i.

Finally, we need to deal with the last term in (5.57), which in the limit becomes just $-\ell\partial_i F_x^i$. As we have just shown, the only non-negligible terms in F_x^i itself are (5.61) and (5.62). But (5.62) is already of order ℓc and therefore can be ignored when multiplied by the additional ℓ above. The only term that survives is

$$\begin{aligned} & -\ell\partial_i\left[\frac{N_x}{Z_i\epsilon}\int_0^{\infty}\frac{k}{m}\exp\left\{-\frac{\beta k^2}{2m}-\zeta^i k\right\}dk\right] \\ = & \quad -\ell\partial_i\left[\frac{N_x}{m}\left(\frac{\beta}{2\pi m}\right)^{1/2}e^{-m(\zeta^i)^2/2\beta}M_1(\zeta^i)\right], \end{aligned}$$

which, to zeroth order in ζ^i, is

$$-\ell\partial_i\left(\frac{N_x}{(2\pi m\beta)^{1/2}}\right)=-\frac{\ell c}{(2\pi\Theta_0)^{1/2}}\partial_i\left(\frac{N_x}{k_B^{1/2}\beta^{1/2}}\right)=-\frac{\lambda}{\rho}\partial_i(\Theta^{1/2}N_x). \tag{5.65}$$

When multiplied by m/a^3 this is what we call the diffusion current

$$J_d^i=-\frac{\lambda}{\rho}\partial_i(\Theta^{1/2}\rho), \tag{5.66}$$

which is made up of the Fick current

$$-\lambda\Theta^{1/2}\partial_i(\log\rho) \tag{5.67}$$

and the Soret current

$$-\lambda(2\Theta^{1/2})^{-1}\partial_i\Theta. \tag{5.68}$$

Therefore, we obtain the total mass current by collecting together (5.61), (5.64), (5.62) and (5.65), that is

$$J_x^i=N_x u_x^i-\frac{\lambda}{\rho}\frac{N_x}{k_B\Theta^{1/2}}\partial_i\Phi(x)-\frac{\lambda}{\rho}\partial_i(\Theta^{1/2}N_x). \tag{5.69}$$

We now go back to (5.52) and expand the finite difference in there as

$$\frac{J^i_{x+\ell e_i} - J^i_x}{\ell} = \frac{\partial J^i_x}{\partial x^i} + \frac{\ell}{2}\frac{\partial^2 J^i_x}{\partial x^{i2}} + O(\ell^2).$$

Since the expression we just found for J^i_x does not contain any term with a large factor c, we see that, in the limit $\ell \to 0$ subject to keeping ℓc finite, equation (5.53) becomes

$$\frac{\partial N_x}{\partial t} + \mathrm{div} J = 0. \tag{5.70}$$

Multiplying both sides of (5.70) by m/a^3 gives us the equation for the time evolution of the particle's density

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\mathbf{u}\rho) = \lambda \ \mathrm{div}\left(\rho^{-1}\nabla(\Theta^{1/2}\rho) + \frac{\nabla\Phi}{k_B \Theta^{1/2}}\right), \tag{5.71}$$

or

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(J_\rho) = 0 \tag{5.72}$$

where the conserved density current is found to be

$$J_\rho = \mathbf{u}\rho + J_d + J_S. \tag{5.73}$$

5.4.2 Dynamics of the energy in an external potential

Let J^i_x/ℓ be now the change in the value of E_x due to the hoppings occuring between x and $x - \ell e_i$. As before, the change due to exchanges with both $x \pm \ell e_i$ in an interval δt is

$$\delta_i E_x = -\frac{J^i_{x+\ell e_i} - J^i_x}{\ell}\delta t \tag{5.74}$$

and the total change due in E_x in an interval δt due to hoppings in all directions is

$$\delta E_x = (\delta_1 E_x + \delta_2 E_x + \delta_3 E_x)\delta t. \tag{5.75}$$

We have that

$$J^i_x = -\sum_{k^i \le 0} \ell r_-(k_x) p_x(\mathbf{k})\mathcal{E}_x(\mathbf{k}) + \sum_{k^i \ge \kappa^i} \ell r_+(k_{x-\ell e_i}) p_{x-\ell e_i}(\mathbf{k})\mathcal{E}_{x-\ell e_i}(\mathbf{k}), \tag{5.76}$$

where $\mathcal{E} = \mathbf{k}\cdot\mathbf{k}/2m + \Phi(x)$.

The analogue of the quantity F_x of the previous section is now

$$G_x^i = N_x(Z\epsilon^3)^{-1}\int_{k^i\geq\kappa^i} \ell r_+(k^i)\left(\frac{\mathbf{k}\cdot\mathbf{k}}{2m} + \Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k}\right\}d^3\mathbf{k}. \tag{5.77}$$

Adding and subtracting this to (5.76), replacing sums by integrals and again using that $p_x(k) = N_x\bar{p}_x(k)$, we obtain

$$\begin{aligned} J_x^i &= G_x^i - \frac{N_x}{Z\epsilon^3}\int_{k^i\leq 0}\ell r_-(k_x)\left(\frac{\mathbf{k}\cdot\mathbf{k}}{2m} + \Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k}\right\}d^3\mathbf{k} \\ &\quad -\ell\left(G_x^i - G_{x-\ell e_i}^i\right)/\ell. \end{aligned} \tag{5.78}$$

We calculate G_x first (for $i = 1$). In the integral

$$\int_{k_1\geq\kappa}\frac{k_1 + (k_1^2-\kappa^2)^{1/2}}{2m}\left(\frac{\mathbf{k}\cdot\mathbf{k}}{2m} + \Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k}\right\}d^3\mathbf{k},$$

we make the change of variables $k_1'^2 = k_1^2 - \kappa^2$ while keeping $k_2' = k_2$ and $k_3' = k_3$. Note that

$$k_1^2/2m + \Phi(x) = k_1'^2/2m + \Phi(x + \ell e_1)$$

and $k_1 dk_1 = k_1' dk_1'$, so defining

$$A(\mathbf{k}') = \left(\frac{\mathbf{k}'\cdot\mathbf{k}'}{2m} + \Phi(x+\ell e_1)\right)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k}'\right\},$$

the integral becomes

$$\int_{k_1'\geq 0}\frac{k_1'\left[(k_1'^2+\kappa^2)^{1/2} + k_1'\right]}{2m(k_1'^2+\kappa^2)^{1/2}}A(\mathbf{k}')\exp\left\{-\frac{\beta\kappa^2}{2m} - \left((k_1'^2+\kappa^2)^{1/2} - k_1'\right)\zeta^1\right\}d^3\mathbf{k}' \tag{5.79}$$

We now expand the exponential to first order

$$\exp\left\{-\frac{\beta\kappa^2}{2m} - \left[\left(k_1'^2+\kappa^2\right)^{1/2} - k_1'\right]\zeta^1\right\} = 1 - \frac{\beta\kappa^2}{2m} - \left[k_1' - \left(k_1'^2+\kappa^2\right)^{1/2}\right]\zeta^1.$$

Thus the integral splits in the following three terms

$$
\begin{aligned}
&\frac{N_x}{Z\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'\left[(k_1'^2+\kappa^2)^{1/2}+k_1'\right]}{2m(k_1'^2+\kappa^2)^{1/2}}A(\mathbf{k}')d^3\mathbf{k}' && (5.80)\\
-\quad&\frac{N_x\beta\kappa^2}{2mZ\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'\left[(k_1'^2+\kappa^2)^{1/2}+k_1'\right]}{2m(k_1'^2+\kappa^2)^{1/2}}A(\mathbf{k}')d^3\mathbf{k}' && (5.81)\\
&-\frac{N_x\kappa^2}{Z\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'}{2m(k_1'^2+\kappa^2)^{1/2}}A(\mathbf{k}')d^3\mathbf{k}'. && (5.82)
\end{aligned}
$$

If, we approximate the hopping rates appearing above simply by k_1'/m [19, appendix 3] and use that $\Phi(x+\ell e_1)=\Phi(x)+\ell\partial_1\Phi(x)$, the integral in (5.80) becomes

$$
\begin{aligned}
&\frac{N_x}{Z\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'}{m}\left(\frac{\mathbf{k}'\cdot\mathbf{k}'}{2m}+\Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}'\\
+\quad&\frac{N_x}{Z\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'}{m}\ell\partial_1\Phi(x)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}' \qquad (5.83)
\end{aligned}
$$

The first part of (5.83) is later going to be combined with the integral over the negative values of k_1 appearing in the second term of (5.78). As for the second part of (5.83), we have

$$
\begin{aligned}
&\frac{N_x}{Z\epsilon^3}\int_{k_1'\geq 0}\frac{k_1'}{m}\ell\partial_1\Phi(x)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}'\\
=\quad&\frac{N_x\ell\partial_1\Phi(x)}{Z_1\epsilon m}\int_{k_1'\geq 0}k_1'\exp\left\{-\frac{\beta k_1'^2}{2m}-\zeta^1 k_1\right\}dk_1'\\
=\quad&\frac{N_x\ell\partial_1\Phi(x)}{Z_1\epsilon m}M_1(\zeta^1)=\frac{N_x\ell\partial_1\Phi(x)}{m}\left(\frac{\beta}{2\pi m}\right)^{1/2}M_1(\zeta^1)e^{-\frac{m(\zeta^1)^2}{2\beta}}
\end{aligned}
$$

Expanding it to zeroth order in ζ^1, we get

$$
\frac{N_x\ell\partial_1\Phi(x)}{\beta}\left(\frac{\beta}{2\pi m}\right)^{1/2}=\frac{\lambda}{\rho}N_x\Theta^{1/2}\partial_1\Phi(x). \qquad (5.84)
$$

For the integral (5.81), we again approximate the hopping rate by k_1'/m and use that $\Phi(x+\ell e_1)=\Phi(x)+\ell\partial_1\Phi(x)$, to obtain

$$
\begin{aligned}
&-\frac{N_x\beta\kappa^2}{2m^2Z\epsilon^3}\int_{k_1'\geq 0}k_1'\left(\frac{\mathbf{k}'\cdot\mathbf{k}'}{2m}\right)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}'\\
&-\frac{N_x\beta\kappa^2}{2m^2Z\epsilon^3}\int_{k_1'\geq 0}k_1'(\Phi(x)+\ell\partial_1\Phi(x))\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}' \quad (5.85)
\end{aligned}
$$

The first term above divides into three integrals. The first one is

$$
\begin{aligned}
& -\frac{\beta N_x \kappa^2}{2m^2 Z \epsilon^3} \int_{k_1' \geq 0} \frac{k_1'^3}{2m} \exp\left\{-\frac{\beta \mathbf{k}' \cdot \mathbf{k}'}{2m} - \boldsymbol{\zeta} \cdot \mathbf{k}'\right\} d^3\mathbf{k}' \\
= & -\frac{\beta N_x \ell \partial_1 \Phi}{m Z_1 \epsilon} \frac{M_3(\zeta^1)}{2m} \\
= & -N_x \left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_1 \Phi(x) k_B \Theta;
\end{aligned} \tag{5.86}
$$

the second one is

$$
\begin{aligned}
& -\frac{\beta N_x \kappa^2}{2m^2 Z_1 Z_2 \epsilon^2} \int_{k_1' \geq 0} \frac{k_1'^3}{2m} e^{-\frac{\beta k_1^2}{2m} - \zeta^1 k_1} dk_1 \int_{k_2} k_2^2 e^{-\frac{\beta k_2^2}{2m} - \zeta^2 k_2} dk_2 \\
= & -\frac{\beta N_x \ell \partial_1 \Phi(x)}{m Z_1 \epsilon} \frac{M_1(\zeta^1)}{2m} \left[\frac{m^2 (\zeta^2)^2}{\beta^2} + \frac{m}{\beta}\right] \\
= & -N_x \left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_1 \Phi(x) \left[\frac{m(u^2)^2}{2} + \frac{1}{2} k_B \Theta\right];
\end{aligned} \tag{5.87}
$$

while the third one is

$$
\begin{aligned}
& -\frac{\beta N_x \kappa^2}{2m^2 Z_1 Z_3 \epsilon^2} \int_{k_1' \geq 0} \frac{k_1'^3}{2m} e^{-\frac{\beta k_1^2}{2m} - \zeta^1 k_1} dk_1 \int_{k_3} k_3^2 e^{-\frac{\beta k_3^2}{2m} - \zeta^3 k_3} dk_3 \\
= & -\frac{\beta N_x \ell \partial_1 \Phi(x)}{m Z_1 \epsilon} \frac{M_1(\zeta^1)}{2m} \left[\frac{m^2 (\zeta^3)^2}{\beta^2} + \frac{m}{\beta}\right] \\
= & -N_x \left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_1 \Phi(x) \left[\frac{m(u^3)^2}{2} + \frac{1}{2} k_B \Theta\right].
\end{aligned} \tag{5.88}
$$

So the total contribution from the first term of (5.85) is

$$
-N_x \left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_1 \Phi(x) \left(\frac{m(u^2)^2}{2} + \frac{m(u^3)^2}{2} + 2k_B \Theta\right). \tag{5.89}
$$

The terms involving the velocities disappear in the limit. The remaining term is

$$
-2\frac{\lambda}{\rho} N_x \Theta^{1/2} \partial_1 \Phi(x). \tag{5.90}
$$

As for the second part of (5.85) we have

$$
-\frac{N_x \beta \kappa^2}{2m^2 Z \epsilon^3} \left[\Phi(x) + \ell \partial_1 \Phi(x)\right] M_1(\zeta^1) = -N_x \left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_1 \Phi(x) \left[\Phi(x) + \ell \partial_1 \Phi(x)\right], \tag{5.91}
$$

of which only the first term survives in the limit, leaving us with

$$-\lambda\frac{\partial_1\Phi(x)}{k_B\Theta^{1/2}}\frac{N_x}{\rho}\Phi(x) \tag{5.92}$$

The integral in (5.82) is itself of smaller order and can be ignored [19, appendix 4]. This completes the contribution of G_x to the energy current.

We move on to deal with the second term in (5.78). The factor ℓr_- can be replaced by $-k_1/m$ [19, appendix 5], leading to a contribution of the form

$$\frac{N_x}{Z\epsilon^3}\int_{k_1\le 0}\frac{k_1}{m}\left(\frac{\mathbf{k}\cdot\mathbf{k}}{2m}+\Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}\cdot\mathbf{k}}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}\right\}d^3\mathbf{k}. \tag{5.93}$$

As promised, this joins the first part of (5.83) to give

$$\begin{aligned} N_x\mathbf{E}_{\bar{p}}\left[\frac{\mathcal{P}_1}{m}\mathcal{E}_x\right] &= N_x\mathbf{E}_{\bar{p}}\left[\frac{\mathcal{P}_1}{m}\right]\mathbf{E}_{\bar{p}}[\mathcal{E}]+N_x\overline{cor}\left(\frac{\mathcal{P}^1}{m},\mathcal{E}_x\right)\\ &= u_x^1E_x+\frac{N_x}{m}\frac{\partial^2\log Z}{\partial\zeta_1\partial\beta}\\ &= u_x^1E_x+N_xk_B\Theta_xu_x^1 \end{aligned} \tag{5.94}$$

We finally look at the last part of (5.78), which in the limit becomes $-\ell\partial_1G_x$. As we have just seen, all the contribution coming from G_x are already of order ℓc (and can therefore be discarded when multiplied by the additional ℓ above) with the exception of the first part of (5.83), that is,

$$\frac{N_x}{Z\epsilon^3}\int_{k_1'\ge 0}\frac{k_1'}{m}\left(\frac{\mathbf{k}'\cdot\mathbf{k}'}{2m}+\Phi(x)\right)\exp\left\{-\frac{\beta\mathbf{k}'\cdot\mathbf{k}'}{2m}-\boldsymbol{\zeta}\cdot\mathbf{k}'\right\}d^3\mathbf{k}'. \tag{5.95}$$

We recognise the term not involving Φ as being equal to the first term in (5.85) times a factor $-2m/\beta\kappa^2 = -(\beta\ell\partial_1\Phi(x))^{-1}$. Therefore, its contribution to the energy current is

$$-\ell\partial_1\left[2\frac{\lambda}{\ell\rho}N_xk_B\Theta^{3/2}\right] = -2\frac{\lambda}{\rho}\partial_1(N_xk_B\Theta^{3/2}). \tag{5.96}$$

As for the term involving Φ, we again recognise it as being equal to the term involving Φ in the second part of (5.85) times the factor $-2m/\beta\kappa^2 = -1/\beta\ell\partial_1\Phi$. Therefore, its contribution to the energy current is

$$-\ell\partial_1\left[\frac{\lambda}{\rho}\frac{N_x\Theta^{1/2}}{\ell}\Phi(x)\right] = -\frac{\lambda}{\rho}\partial_1(N_x\Theta^{1/2})\Phi(x)-\frac{\lambda}{\rho}N_x\Theta^{1/2}\partial_1\Phi(x) \tag{5.97}$$

We can now take a breath and collect all the terms we have obtained for the energy current to put back in (5.78). They are (5.84), (5.90), (5.92), (5.94), (5.96) and (5.97) and the end result is

$$\begin{aligned} J_x^i &= u_x^i(E_x + N_x k_B \Theta_x) - 2\frac{\lambda}{\rho} N_x \Theta^{1/2} \partial_i \Phi(x) \\ &- \lambda \frac{\partial_i \Phi(x)}{k_B \Theta^{1/2}} \frac{N_x}{\rho} \Phi(x) - 2\frac{\lambda}{\rho} \partial_i (N_x k_B \Theta^{3/2}) - \frac{\lambda}{\rho} \partial_i (N_x \Theta^{1/2}) \Phi(x). \end{aligned} \tag{5.98}$$

Again we see that none of the terms in J_x^i contains a large factor c, so that the discussion preceding (5.70) applies here as well and in the limit we obtain

$$\frac{\partial E_x}{\partial t} + \mathrm{div} J = 0, \tag{5.99}$$

that is

$$\begin{aligned} \frac{\partial E_x}{\partial t} + \ \mathrm{div}\ [\mathbf{u}_x(E_x + N_x k_B \Theta_x)] = \ \mathrm{div}\ \Big[2\frac{\lambda}{\rho} N_x \Theta^{1/2} \nabla \Phi(x) \\ + \ \lambda \frac{\nabla \Phi(x)}{k_B \Theta^{1/2}} \frac{N_x}{\rho} \Phi(x) + 2\frac{\lambda}{\rho} \nabla (N_x k_B \Theta^{3/2}) + \frac{\lambda}{\rho} \nabla (N_x \Theta^{1/2}) \Phi(x) \Big]. \end{aligned} \tag{5.100}$$

We now use that $e(x) = E_x/(mN_x)$, $\phi(x) = \Phi(x)/m$, $\rho(x) = mN_x/a^3$ and $P_x = N_x K_B \Theta_x / a^3$, divide both sides of the previous equation by a^3 and obtain

$$\begin{aligned} \frac{\partial(\rho e)}{\partial t} + \ \mathrm{div}\ [\mathbf{u}(\rho e + P)] &= \lambda\ \mathrm{div}\ \Big[2\rho^{-1} \nabla \left(\Theta^{1/2} P \right) + \rho^{-1} \nabla \left(\Theta^{1/2} \rho \right) \phi \\ &\quad + \frac{\nabla \Phi}{k_B \Theta^{1/2}} \phi + 2 \frac{\nabla \Phi}{k_B \Theta^{1/2}} P \Big], \end{aligned} \tag{5.101}$$

which can be written as (recall (5.63) and (5.66))

$$\frac{\partial(\rho e)}{\partial t} + \ \mathrm{div}\ [\mathbf{u}(\rho e + P) + (J_d + J_S)\phi + 2J_S P] = 2\lambda\ \mathrm{div}\ \left[\rho^{-1} \nabla \left(P \Theta^{1/2} \right) \right] \tag{5.102}$$

5.4.3 Dynamics of the momentum in an external field

Since momentum is not conserved (as there are body-forces due to the external field), the rate of change of momentum density will not be the divergence of something; we expect the extra term to be ρf where $f := -\nabla \Phi / m$ is the force per unit mass. To see this, let us define the current J_x^i as in the previous sections, namely

$$J_x^i = -\sum_{k^i \le 0} \ell r_-(k_x^i) p_x(k) \mathcal{P}_x^j(k) + \sum_{k^i \ge \kappa^i} \ell r_+(k_{x-\ell e_i}^i) p_{x-\ell e_i}(k) \mathcal{P}_{x-\ell e_i}^j(k) \tag{5.103}$$

Then the change in ϖ_x^j due to exchanges with both $x \pm \ell e_i$ in an interval δt will only be given by the usual

$$\delta_i \varpi_x^j = -\frac{J_{x+\ell e_i}^i - J_x^i}{\ell} \delta t \tag{5.104}$$

for $i \neq j$, because it is implicit in this formula that the particles hopping from $x \pm \ell e_i$ to x have their j-component of the momentum unchanged during the jump. We do this case first. The analogue of F_x^i and G_x^i from the previous sections is now

$$H_x^i = N_x (Z\epsilon^3)^{-1} \int_{k^i \ge \kappa^i} \ell r_+(k^i) k^j \exp\left\{ -\frac{\beta \mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k} \right\} d^3\mathbf{k}. \tag{5.105}$$

Adding and subtracting this to (5.103), replacing sums by integral and using that $p_x(k) = N_x \bar{p}_x(k)$, we obtain the familiar form

$$\begin{aligned} J_x^i &= H_x^i - \frac{N_x}{Z\epsilon^3} \int_{k^i \le 0} \ell r_-(k^i) k^j \exp\left\{ -\frac{\beta \mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k} \right\} d^3\mathbf{k} \\ &- \ell \left(H_x^i - H_{x-\ell e_i}^i \right) / \ell. \end{aligned} \tag{5.106}$$

Notice that since $\varpi_x^j = m N_x u_x^j$, in this section we shall keep all terms of order $m u_x^j$. If we now perform in (5.105) the integrations over k_r ($r \neq i$, $r \neq j$) and k_j we find

$$\begin{aligned} H_x^i &= m u_x^j \frac{N_x}{Z_i \epsilon} \int_{k \ge \kappa} \ell r_+(k) \exp\left\{ -\frac{\beta k^2}{2m} - \zeta k \right\} dk \\ &= m u_x^j F_x^i. \end{aligned}$$

But now we can use the calculation we have already done for F_x^i, that is, in the limit we have

$$\begin{aligned} H_x^i &= m u_x^j \frac{N_x}{Z_i \epsilon} \int_0^\infty \frac{k'}{m} \exp\left\{ -\frac{\beta k'^2}{2m} - \zeta^i k' \right\} dk \\ &- \left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}} \partial_i \Phi(x) \right] m u_x^j \end{aligned} \tag{5.107}$$

The second term above, when multiplied by m/a^3, simply gives $mu_x^j J_S^i$.

As for the second term in (5.106), we can also perform the integrations over k_r ($r \neq i, r \neq j$) and k_j, as well as to replace the factor $\ell r_-(k^i)$ by $-k^i/m$ to find

$$mu_x^j \frac{N_x}{Z_i \epsilon} \int_{-\infty}^{0} \frac{k}{m} \exp\left\{-\frac{\beta k^2}{2m} - \zeta^i k\right\} dk. \tag{5.108}$$

When we add this to what we have found in the first term in (5.107) and multiply them m/a^3, the resulting term is $m\rho u_x^j u_x^i$.

Finally, we look at the last term in (5.106), which in the limit becomes $-\ell \partial_i H_x^i$. Since the second term in (5.107) is already of order ℓc, we see that the only surviving contribution here is

$$-\ell \partial_i \left[mu_x^j \frac{N_x}{Z_i \epsilon} \int_0^{\infty} \frac{k'}{m} \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^i k'\right\} dk \right] = -\ell \partial_i \left[mu_x^j \frac{N_x}{Z_i \epsilon m} M_1(\zeta^i) \right], \tag{5.109}$$

which, to zeroth order in ζ^i, gives

$$\begin{aligned} -\ell \partial_i \left[mu_x^j \frac{N_x}{m} \left(\frac{\beta}{2\pi m}\right)^{1/2} \frac{m}{\beta} \right] &= -\frac{\ell c}{(2\pi \Theta_0)^{1/2}} \partial_i (N_x \Theta^{1/2} m u^j) \\ &= -\frac{\lambda}{\rho} \partial_i (N_x \Theta^{1/2} m u^j). \end{aligned} \tag{5.110}$$

Therefore, the momentum current, for $i \neq j$, is given by

$$J_x^i = mN_x u_x^j u_x^i - \left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}} \partial_i \Phi(x) \right] mu_x^j - \frac{\lambda}{\rho} \partial_i (N_x \Theta^{1/2} m u^j). \tag{5.111}$$

Once more, none of the above terms contain a large factor c, so in the limit $\ell \to 0$ subject to ℓc finite, we can approximate the term $(J_{x+\ell e_i}^i - J_x^i)/\ell$ in (5.104) simply by $\partial_i J_x^i$.

For $i = j$, equation (5.104) is *not* correct. In this case, if $k_{x-\ell e_j} > \kappa_{x-\ell e_j}^j$ is the j-component of the momentum for a particle at $x - \ell e_j$, then

$$k'_{x-\ell e_j} = [k_{x-\ell e_j}^2 - 2m\ell \partial_j \Phi(x - \ell e_j)]^{1/2}$$

is the j-component of its momentum when it arrives at x. Similarly, if $k_{x+\ell e_j} < 0$ is the j-component of the momentum for a particle at $x + \ell e_j$, then

$$k''_{x+\ell e_j} = -[k_{x+\ell e_i}^2 + 2m\ell \partial_j \Phi(x)]^{1/2}$$

is the j-component of its momentum when it gets to x. Therefore, the total rate of change in ϖ_x^j due to exchanges with $x \pm \ell e_j$ consists of the following four terms

$$\frac{\delta_j \varpi_x^j}{\delta t} = \sum_{k_{x+\ell e_j} \leq 0} r_-(k_{x+\ell e_j}) k''_{x+\ell e_j} p_{x+\ell e_j}(\mathbf{k}_{x+\ell e_j}) \tag{5.112}$$

$$- \sum_{k_x \geq \kappa_x^j} r_+(k_x) k_x p_x(\mathbf{k}_x) - \sum_{k_x \leq 0} r_-(k_x) k_x p_x(\mathbf{k}_x) \tag{5.113}$$

$$+ \sum_{k_{x-\ell e_j} \geq \kappa^j_{x-\ell e_j}} r_+(k_{x-\ell e_j}) k'_{x-\ell e_j} p_{x-\ell e_j}(\mathbf{k}_{x-\ell e_j}) \tag{5.114}$$

If we now recall from (5.46) what the hopping rates look like, and use that

$$\begin{aligned}
\left(k_{x+\ell e_j} + k''_{x+\ell e_j}\right) k''_{x+\ell e_j} &= \left(k_{x+\ell e_j} + k''_{x+\ell e_j}\right) k_{x+\ell e_j} + 2m\ell \partial_j \Phi(x) \\
\left(k_{x-\ell e_j} + k'_{x-\ell e_j}\right) k'_{x-\ell e_j} &= \left(k_{x-\ell e_j} + k'_{x-\ell e_j}\right) k_{x-\ell e_j} - 2\ell m \partial_j \Phi(x - \ell e_j),
\end{aligned}$$

then we can rewrite the above as

$$\begin{aligned}
\frac{\delta_j \varpi_x^j}{\delta t} &= \sum_{k_{x+\ell e_j} \leq 0} r_-(k_{x+\ell e_j}) k_{x+\ell e_j} p_{x+\ell e_j}(\mathbf{k}_{x+\ell e_j}) \\
&- \sum_{k_x \geq \kappa_x^j} r_+(k_x) k_x p_x(\mathbf{k}_x) - \sum_{k_x \leq 0} r_-(k_x) k_x p_x(\mathbf{k}_x) \\
&+ \sum_{k_{x-\ell e_j} \geq \kappa^j_{x-\ell e_j}} r_+(k_{x-\ell e_j}) k_{x-\ell e_j} p_{x-\ell e_j}(\mathbf{k}_{x-\ell e_j}) \\
&- \partial_j \Phi(x) \sum_{k_{x+\ell e_j} \leq 0} p_{x+\ell e_j}(\mathbf{k}_{x+\ell e_j}) \\
&- \partial_j \Phi(x - \ell e_j) \sum_{k_{x-\ell e_j} \geq \kappa^j_{x-\ell e_j}} p_{x-\ell e_j}(\mathbf{k}_{x-\ell e_j})
\end{aligned}$$

which we then recognise as

$$\begin{aligned}
\frac{\delta_j \varpi_x^j}{\delta t} &= -\frac{J^j_{x+\ell e_j} - J^j_x}{\ell} \\
&- \partial_j \Phi(x) \sum_{k_{x+\ell e_j} \leq 0} p_{x+\ell e_j}(\mathbf{k}_{x+\ell e_j}) \\
&- \partial_j \Phi(x - \ell e_j) \sum_{k_{x-\ell e_j} \geq \kappa^j_{x-\ell e_j}} p_{x-\ell e_j}(\mathbf{k}_{x-\ell e_j}).
\end{aligned} \tag{5.115}$$

In the limit, the last two terms above add up to $N_x \partial_j \Phi(x)$ [19, appendix 6].

Let us now calculate J_x^j, that is

$$\begin{aligned} J_x^j &= H_x^j - \frac{N_x}{Z\epsilon^3}\int_{k^j\le 0} \ell r_-(k^j)k^j \exp\left\{-\frac{\beta\mathbf{k}\cdot\mathbf{k}}{2m} - \boldsymbol{\zeta}\cdot\mathbf{k}\right\} d^3\mathbf{k} \\ &- \ell\left(H_x^j - H_{x-\ell e_j}^j\right)/\ell. \end{aligned} \tag{5.116}$$

We find that H_x^j reduces to

$$H_x^j = \frac{N_x}{Z_j\epsilon}\int_{k\ge\kappa} \frac{k+(k^2-\kappa^2)^{1/2}}{2m} k \exp\left\{-\frac{\beta k^2}{2m} - \zeta^j k\right\} dk,$$

where we can make the familiar change of variables $k'^2 = k^2 - \kappa^2$ to obtain

$$\begin{aligned} H_x^j = \frac{N_x}{Z_j\epsilon}&\int_{k'\ge 0}\left[\frac{(k'^2+\kappa^2)^{1/2}+k'}{2m}\right] k' \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} \\ \times\ & \exp\left\{-\frac{\beta\kappa^2}{2m} - \left((k'^2+\kappa^2)^{1/2} - k'\right)\zeta^j\right\} dk'. \end{aligned}$$

Expanding the exponential to first order gives the usual three terms

$$\begin{aligned} H_x^j &= \frac{N_x}{Z_j\epsilon}\left(\int_{k'\ge 0}\left[\frac{(k'^2+\kappa^2)^{1/2}+k'}{2m}\right] k' \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk'\right. && (5.117)\\ &-\frac{\beta\kappa^2}{2m}\int_{k'\ge 0}\left[\frac{(k'^2+\kappa^2)^{1/2}+k'}{2m}\right] k' \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk' && (5.118)\\ &\left.-\frac{\kappa^2\zeta^j}{2m}\int_{k'\ge 0} k' \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk'.\right) && (5.119) \end{aligned}$$

In (5.117), we can replace the factor

$$\frac{(k'^2+\kappa^2)^{1/2}+k'}{2m}$$

by k'/m [19, appendix 7], so that it amounts to

$$\frac{N_x}{Z_j\epsilon m}\int_{k'\ge 0}(k')^2 \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk'. \tag{5.120}$$

In (5.118), the same replacement gives

$$-\frac{N_x\beta\kappa^2}{2Z_j\epsilon m^2}\int_{k'\ge 0}(k')^2 \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk' = -\frac{N_x\beta\kappa^2}{2Z_j\epsilon m^2} M_2(\zeta^j), \tag{5.121}$$

which, to first order in ζ_j (recall that we are keeping terms proportional to mu^j in this section, since $\varpi_x^j = N_x m u_x^j$), gives

$$\begin{aligned} &- \frac{2N_x\beta m\ell\partial_j\Phi(x)}{2m^2}\left(\frac{\beta}{2\pi m}\right)^{1/2}\left[\left(\frac{\pi}{2}\right)^{1/2}\left(\frac{m}{\beta}\right)^{3/2} - 2\left(\frac{m}{\beta}\right)^2\zeta^j\right] = \\ &- \frac{\ell N_x\partial_j\Phi(x)}{2} - 2N_x\left(\frac{\beta}{2\pi m}\right)^{1/2}\ell\partial_j\Phi(x) m u^j \end{aligned} \tag{5.122}$$

of which only the second term survives in the limit, resulting in

$$-2\left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}}\partial_j \Phi(x)\right] m u_x^j. \tag{5.123}$$

Similarly in (5.119), we obtain

$$\begin{aligned}
-\frac{N_x \kappa^2 \zeta^j}{2Z_j \epsilon m} M_1(\zeta^j) &= \frac{2N_x m \ell \partial_j \Phi \beta u_x^j}{2m}\left(\frac{\beta}{2\pi m}\right)^{1/2}\frac{m}{\beta} \\
&= N_x\left(\frac{\beta}{2\pi m}\right)^{1/2} \ell \partial_j \Phi(x) m u_x^j \\
&= \left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}}\partial_j \Phi(x)\right] m u_x^j.
\end{aligned} \tag{5.124}$$

Moving to the second term in (5.116), we replace the factor $\ell r_-(k^j)$ by $-k^j/m$ [19, appendix 8], so it contributes with

$$\frac{N_x}{Z_j \epsilon m}\int_{k\le 0} k^2 \exp\left\{-\frac{\beta k^2}{2m} - \zeta^j k\right\} dk. \tag{5.125}$$

Combining (5.120) and (5.125), what we obtain is

$$\begin{aligned}
\frac{N_x}{m Z_j \epsilon}\int_{-\infty}^{\infty}(k^j)^2 \exp\left\{-\frac{\beta k^2}{2m} - \zeta^j k\right\} dk^j &= \frac{N_x}{m} E_{\bar{p}}\left[(k^j)^2\right] \\
&= \frac{N_x}{m} E_{\bar{p}}\left[k^j\right]^2 + \frac{N_x}{m}\frac{\partial^2 \log Z_j}{\partial (\zeta^j)^2} \\
&= m N_x (u_x^j)^2 + N_x k_B \Theta_x.
\end{aligned} \tag{5.126}$$

Finally, for the last term in (5.116), which in the limit becomes $-\ell \partial_j H_x^j$, only (5.120) contributes, since the other terms in H_x^j are already of order ℓc. We get

$$\begin{aligned}
-\ell \partial_j H_x^j &= -\ell \partial_j \left[\frac{N_x}{Z_j \epsilon m}\int_{k'\ge 0}(k')^2 \exp\left\{-\frac{\beta k'^2}{2m} - \zeta^j k'\right\} dk'\right] \\
&= -\ell \partial_j \left[\frac{N_x}{Z_j \epsilon m} M_2(\zeta^j)\right],
\end{aligned}$$

which, to first order in ζ^j, gives

$$\begin{aligned}
&- \quad \ell \partial_j \left\{\frac{N_x}{m}\left(\frac{\beta}{2\pi m}\right)^{1/2}\left[\left(\frac{\pi}{2}\right)^{1/2}\left(\frac{m}{\beta}\right)^{3/2} - 2\left(\frac{m}{\beta}\right)^2 \zeta^j\right]\right\} = \\
&-\frac{\ell}{2}\partial_j \left(N_x k_B \Theta_x\right) - 2\ell \partial_j \left[m u^j N_x \left(\frac{\beta}{2\pi m}\right)^{1/2}\frac{1}{\beta}\right] = \\
&-\frac{\ell}{2}\partial_j \left(N_x k_B \Theta_x\right) - 2\frac{\lambda}{\rho}\partial_j \left(N_x \Theta^{1/2} m u_x^j\right)
\end{aligned} \tag{5.127}$$

Therefore, the momentum current, for $i = j$ is given by the sum of (5.123), (5.124), (5.126) and (5.127), that is

$$\begin{aligned} J_x^j = mN_x(u_x^j)^2 &+ N_x k_B \Theta_x - \left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}} \partial_j \Phi(x)\right] m u_x^j \\ &- \frac{\ell}{2} \partial_j \left(N_x k_B \Theta_x\right) - 2\frac{\lambda}{\rho} \partial_j \left(N_x \Theta^{1/2} m u_x^j\right). \end{aligned}$$

We see that the term $N_x k_B \Theta_x$ above is of a larger order than the others and is not negligible in the expansion of the finite difference $\left(J_{x+\ell e_j} - J_x\right)/\ell$. What we obtain is

$$\begin{aligned} \frac{J_{x+\ell e_j} - J_x}{\ell} &= \frac{\partial J_x^j}{\partial x^j} + \frac{\ell}{2} \frac{\partial^2 J_x^j}{\partial x^{j^2}} + O(\ell^2) \\ &= \frac{\partial J_x^j}{\partial x^j} + \frac{\ell}{2} \frac{\partial^2 (N_x k_B \Theta_x)}{\partial x^{j^2}} + O(\ell^2) \end{aligned}$$

Thus, in this case, the finite difference $\left(J_{x+\ell e_j} - J_x\right)/\ell$ can be approximated by

$$\partial_j \left\{ mN_x(u_x^j)^2 + N_x k_B \Theta_x - \left[\frac{\lambda N_x}{\rho k_B \Theta^{1/2}} \partial_j \Phi(x)\right] m u_x^j - 2\frac{\lambda}{\rho} \partial_j \left(N_x \Theta^{1/2} m u_x^j\right) \right\}. \tag{5.128}$$

Therefore, collecting together the contributions from $i = 1, 2, 3$, we see from (5.104), (5.115) and the equation above, that the change in ϖ^j is governed by the equation

$$\begin{aligned} \frac{\partial \varpi_x^j}{\partial t} = \frac{\partial}{\partial x^j} \left[\frac{\lambda}{\rho} \frac{\partial}{\partial x^j} \left(N_x \Theta^{1/2} m u_x^j\right)\right] - N_x \partial_j \Phi(x) - \sum_{i=1}^{3} \frac{\partial}{\partial x^i} \Big[m N_x u_x^i u_x^j \\ + N_x k_B \Theta_x \delta_{ij} - \left(\frac{\lambda N_x}{\rho k_B \Theta^{1/2}} \partial_i \Phi\right) m u_x^j - \frac{\lambda}{\rho} \frac{\partial}{\partial x^i} \left(N_x \Theta^{1/2} m u_x^j\right) \Big] \end{aligned} \tag{5.129}$$

The first term above is not covariant and we need to average it over $SO(3)$. The averaging procedure is explained in [66], and its result is

$$\sum_{i=1}^{3} \frac{\partial}{\partial x^i} \frac{\lambda}{5\rho} \left[\frac{\partial}{\partial x^i} \left(N_x \Theta^{1/2} m u_x^j\right) + 2 \frac{\partial}{\partial x^j} \left(N_x \Theta^{1/2} m u_x^i\right)\right]. \tag{5.130}$$

We can now put it back into (5.129), divide both sides of it by a^3 and use that

$\varpi_x^j = mN_x u_x^j$, $\rho(x) = mN_x/a^3$, $f_x^j = -\partial_j \Phi(x)/m$ and $P_x = N_x k_B \Theta_x / a^3$ to obtain

$$\begin{aligned} \frac{\partial \rho(x) u_x^j}{\partial t} &= \rho(x) f_x^j - \sum_{i=1}^{3} \frac{\partial}{\partial x^i} \left\{ \rho(x) u_x^i u_x^j + P_x \delta_{ij} - \frac{\lambda \partial_i \Phi}{k_B \Theta^{1/2}} u_x^j \right. \\ &- \left. \frac{2\lambda}{5\rho} \left[3 \frac{\partial}{\partial x^i} \left(\rho(x) \Theta^{1/2} u_x^j \right) + \frac{\partial}{\partial x^j} \left(\rho(x) \Theta^{1/2} u_x^i \right) \right] \right\}. \end{aligned} \tag{5.131}$$

In vector notation, this reads

$$\begin{aligned} \frac{\partial \rho \mathbf{u}}{\partial t} &+ \operatorname{div}(\rho \mathbf{u} \otimes \mathbf{u}) = \rho \mathbf{f} - \nabla P + \lambda \operatorname{div} \left(\frac{\nabla \Phi}{k_B \Theta^{1/2}} \otimes \mathbf{u} \right) \\ &+ \frac{2\lambda}{5} \partial_i \rho^{-1} \left[3 \partial^i \left(\rho(x) \Theta^{1/2} \mathbf{u} \right) + \nabla \left(\rho(x) \Theta^{1/2} u^i \right) \right], \end{aligned} \tag{5.132}$$

which can be written as

$$\begin{aligned} \frac{\partial \rho \mathbf{u}}{\partial t} &+ \operatorname{div} \left(\rho \mathbf{u} \otimes \mathbf{u} + J_S \otimes \mathbf{u} \right) = \rho \mathbf{f} - \nabla P \\ &+ \frac{2\lambda}{5} \partial_i \rho^{-1} \left[3 \partial^i \left(\rho(x) \Theta^{1/2} \mathbf{u} \right) + \nabla \left(\rho(x) \Theta^{1/2} u^i \right) \right]. \end{aligned} \tag{5.133}$$

Chapter 6

A Personal Look at the Future

Conclusions are the place to look back and recollect the achievements of one's work. Apart from doing this here, we also want to use this chapter to bring together the things which were not proved in this thesis, but for which we consider that further research is worth pursuing.

For nonparametric classical information geometry, the construction presented here, using M^{Φ_1} as the model Banach space, allows us to have well defined exponential and mixture connections which are dual with respect to the Fisher metric. The next step in generalising the parametric results is to prove that the Kullback-Leibler relative entropy is the canonical statistical divergence associated with $\left(\langle\cdot,\cdot\rangle, \nabla^{(1)}, \nabla^{(-1)}\right)$. This involves taking directional derivatives of functionals on the manifold, in the same sense as those we took in chapter 4 for the free energy functional, so all the necessary tools seem to be available for the job. As in the parametric case, the α-connections given here are flat in the extended manifold $\widetilde{\mathcal{M}}$ of weights obtained from $\mathcal{M}$. The same investigation is then valid for the α-divergences with respect to the dualistic flat triple $\left(\langle\cdot,\cdot\rangle, \nabla^{(\alpha)}, \nabla^{(-\alpha)}\right)$.

A more ambitious result is then to obtain the minimisation theorem 5.1.2 for infinite dimensional manifolds. One could start with the case where $p \in \mathcal{M}$ is an arbitrary point but we want to minimise the relative entropy $S(p|q)$ for q in a *finite* dimensional submanifold $\mathcal{S}$. This would be enough for applications in Statistical

Dynamics, where the target for the projection Q is always a set of great grand canonical states parametrised by a finite number of slow variables.

Still on ambitious results in the classical domain, we notice that Chentsov's uniqueness theorem for the Fisher metric is yet to be proved for the nonparametric case.

For finite dimensional quantum systems, if ones chooses the α-embeddings as the prefered way to define the α-connections (as opposed to a convex mixture of the ± 1-connections) then it is known that, for each value of α, the connections $\nabla^{(\alpha)}, \nabla^{(-\alpha)}$ are dual with respect to the *WYD* (Wigner-Yanase-Dyson) metric g^α [22]. The *BKM* metric is a limiting case for the *WYD* metrics for $\alpha \to \pm 1$, so a natural starting point for further research is to extend our characterisation of the *BKM* metric to them. The theorem to be proved is that for each fixed $\alpha \in (-1, 1)$, scalar multiples of the *WYD* metric g^α are the only monotone metric with respect to which the connections $\nabla^{(\alpha)}, \nabla^{(-\alpha)}$ are mutually dual. This would have a corollary that the α-connections defined in this way are *not* the convex mixture of the ± 1-connections, because if they were they would be dual with respect to the *BKM* metric for all $\alpha \in (-1, 1)$, due to the same proof as in corollary 2.5.3.

In the nonparametric quantum case, we have succeeded in constructing a Banach manifold with an infitine dimensional quantum analogue of the exponential connection. Each component $\mathcal{M}(H_0)$ does not cover the whole set $\mathcal{M}$ at once. It could not possibly do so, since our small ε-bounded perturbations do not change the domain of the original Hamiltonian H_0, and $\mathcal{M}$ certainly contains states defined by Hamiltonians with plenty of different domains. Also, although we can reach far removed points with a finite number of small perturbations, we cannot move in arbitrary directions. As an example [63], we can never reach a point where $X = -H_0$ as the result of a chain of perturbations, since the identity is not an operator of trace class. The whole manifold thus obtained consists of several disconnected parts, generally point towards positive directions with respect to the initial Hamiltonians.

As we have seen, the hardest part in the construction is to make sure that the perturbations give rise to well defined new Hamiltonians for the perturbed state. The peculiarities of the quantum case show up at every step. For instance, in the classical case it is a matter of convenience to take the centred unit ball in M^{Φ_1} as coordinates. Any other radius would do just as well. In the quantum case, the very new Hamiltonian $H_X = H_0 + X$ ceases to make sense if X is not a small perturbation of H_0. This is the major technical difficulty in obtaining a quantum analogue of the Luxemburg norm (2.11). As a result, although each connected component $\mathcal{M}(H_0)$ is $+1$-convex, we have not been able to prove that it is -1-convex. For instance, given two states $\rho_1, \rho_2 \in \mathcal{M}_0$, the -1-convex mixture

$$\rho = \lambda\rho_1 + (1-\lambda)\rho_2, \qquad \lambda \in (0,1), \tag{6.1}$$

is obviously another state, but we want to write it as

$$\rho = Z_V^{-1} e^{-(H_0+V)}, \tag{6.2}$$

and prove that V can be made up of finitely many ε-bounded perturbations $X_1, \ldots, X_n$. Only when -1-convexity is established can one try to define the infinite dimensional quantum analogue of the mixture connection, and then prove duality with respect to the infinite dimensional generalised BKM scalar product given in (4.3.11).

On top of these hard problems, the infinite dimensional quantum α-connections share all the difficulties of both the infinite dimensional classical case and the finite dimensional quantum case. Therefore, the same criticism we raised to [15] applies to [13], where exactly the same method is used. Nevertheless, in the same way that we have used many technical tools from [15], we also consider invaluable the efforts that Gibilisco and Isola make [13] in order to have noncommutative L^p-spaces as target spaces for the α-embeddings. An approach similar to what we have done in chapter 2 could then be employed to obtain the α-connections acting all on the same tangent bundle also in the infinite dimensional quantum case. We notice that [13] takes for granted the existence of a Banach manifold of density operators with respect to a von Neumann algebra M and a normal semifinite faithful trace

τ, whereas the only explicit constructions of such a manifold we are aware of are the Hilbert space cases described in chapter 4 of this thesis.

As a final word, we see Information Geometry as a subject that has reached a state of completion in the foundations of the parametric classical case and relative maturity in the foundations of nonparametric classical case. It can claim several successes in finite dimensional quantum case, but still presents fresh and short term problems there. It is also enjoying the excitment of early successes in the infinite dimensional quantum version, where most of the hard problems are still to be posed and solved. As for applications, a quick look at the table of contents of [54] is enough to suggest its broad range: neural networks, thermodynamics, spin systems, financial mathematics, to quote a few. To summarise, a promising field of research.

Bibliography

[1] S.-i. Amari. *Differential-geometrical methods in statistics.* Springer-Verlag, New York, 1985.

[2] S.-i. Amari and H. Nagaoka. *Methods of information geometry.* American Mathematical Society, Providence, RI, 2000. Translated from the 1993 Japanese original by Daishi Harada.

[3] T. M. Apostol. *Mathematical analysis: a modern approach to advanced calculus.* Addison-Wesley Publishing Company, Inc., Reading, Mass., second edition, 1974.

[4] H. Araki. Relative Hamiltonian for faithful normal states of a von Neumann algebra. *Publ. Res. Inst. Math. Sci.*, 9:165–209, 1973/74.

[5] R. Balian, Y. Alhassid, and H. Reinhardt. Dissipation in many-body systems: a geometric approach based on information theory. *Phys. Rep.*, 131(1-2):1–146, 1986.

[6] P. Biler. Steady states and long time asymptotics of solutions of Streater's models. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 237–242. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[7] P. Biler, G. Karch, J. Dolbeault, and M. J. Esteban. Stationary solutions, intermediate asymptotics and large-time behaviour of type II Streater's models. *Adv. Differential Equations*, 6(4):461–480, 2001.

[8] N. N. Čencov. *Statistical decision rules and optimal inference.* American Mathematical Society, Providence, R.I., 1982. Translation from the Russian edited by Lev J. Leifman.

[9] E. B. Davies. *One-parameter semigroups.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1980.

[10] A. P. Dawid. On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. Roy. Statist. Soc. Ser. B*, 37:248–258, 1975.

[11] B. Efron. Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.*, 3(6):1189–1242, 1975. With a discussion by C. R. Rao, Don A. Pierce, D. R. Cox, D. V. Lindley, Lucien LeCam, J. K. Ghosh, J. Pfanzagl, Neils Keiding, A. P. Dawid, Jim Reeds and with a reply by the author.

[12] R. H. Fowler. *Statistical mechanics.* Cambridge University Press, Cambridge, 1980. The theory of the properties of matter in equilibrium, Reprint of the second edition of 1936.

[13] P. Gibilisco and T. Isola. Connections on statistical manifolds of density operators by geometry of noncommutative L^p-spaces. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 2(1):169–178, 1999.

[14] P. Gibilisco and T. Isola. Monotone metrics on statistical manifolds of density matrices by geometry of noncommutative L^2-spaces. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 129–139. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[15] P. Gibilisco and G. Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 1(2):325–347, 1998.

[16] M. R. Grasselli. Dual connection in nonparametric classical information geometry. Submitted to Communications in Mathematical Physics, 2001.

[17] M. R. Grasselli. Infinite dimensional quantum information geometry. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 141–146. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[18] M. R. Grasselli and R. F. Streater. The quantum information manifold for ϵ-bounded forms. *Rep. Math. Phys.*, 46(3):325–335, 2000.

[19] M. R. Grasselli and R. F. Streater. Hydrodynamics in an external field. To appear in Reports on Mathematical Physics, 2001.

[20] M. R. Grasselli and R. F. Streater. On the uniqueness of the Chentsov metric in quantum information geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 4(2):173–182, 2001.

[21] M. R. Grasselli and R. F. Streater. The uniqueness of the Chentsov metric. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 165–170. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[22] H. Hasegawa. Noncommutative extension of the information geometry. In *Quantum communications and measurement (Nottingham, 1994)*, pages 327–337. Plenum, New York, 1995.

[23] H. Hasegawa. Exponential and mixture families in quantum statistics: dual structure and unbiased parameter estimation. *Rep. Math. Phys.*, 39(1):49–68, 1997.

[24] H. Hasegawa and D. Petz. Noncommutative extension of the information geometry ii. In *Quantum communications and measurement*, pages 109–118. Plenum, New York, 1997.

[25] R. S. Ingarden, H. Janyszek, A. Kossakowski, and T. Kawaguchi. Information geometry of quantum statistical systems. *Tensor (N.S.)*, 37(1):105–111, 1982.

[26] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev. (2)*, 106:620–630, 1957.

[27] E. T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev. (2)*, 108:171–190, 1957.

[28] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London. Ser. A.*, 186:453–461, 1946.

[29] A. Jenčová. Geometry of quantum states: dual connections and divergence functions. *Rep. Math. Phys.*, 47(1):121–138, 2001.

[30] R. E. Kass and P. W. Vos. *Geometrical foundations of asymptotic inference.* John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[31] T. Kato. *Perturbation theory for linear operators.* Springer-Verlag New York, Inc., New York, 1966.

[32] M. A. Krasnosel′skiĭ and J. B. Rutickiĭ. *Convex functions and Orlicz spaces.* P. Noordhoff Ltd., Groningen, 1961.

[33] R. Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Japan*, 12:570–586, 1957.

[34] S. Lang. *Differential and Riemannian manifolds.* Springer-Verlag, New York, third edition, 1995.

[35] A. Lesniewski and M. B. Ruskai. Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. *J. Math. Phys.*, 40(11):5702–5724, 1999.

[36] H. Mori. Transport, collective motion and brownian motion. *Prog. Theor. Phys.*, 33:423–455, 1965.

[37] E. A. Morozova and N. N. Chentsov. Markov invariant geometry on state manifolds. In *Current problems in mathematics. Newest results, Vol. 36 (Russian)*, pages 69–102, 187. Akad. Nauk SSSR Vsesoyuz. Inst. Nauchn. i Tekhn.

Inform., Moscow, 1989. Translated in J. Soviet Math. **56** (1991), no. 5, 2648–2669.

[38] M. K. Murray and J. W. Rice. *Differential geometry and statistics.* Chapman & Hall, London, 1993.

[39] T. Nadzieja. A note on nonlocal equations in mathematical physics. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 255–259. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[40] H. Nagaoka. Differential geometrical aspects of quantum state estimation and relative entropy. In *Quantum communications and measurement (Nottingham, 1994)*, pages 449–452. Plenum, New York, 1995.

[41] D. Petz. Quasi-entropies for finite quantum systems. *Rep. Math. Phys.*, 23(1):57–65, 1986.

[42] D. Petz. Geometry of canonical correlation on the state space of a quantum system. *J. Math. Phys.*, 35(2):780–795, 1994.

[43] D. Petz. Monotone metrics on matrix spaces. *Linear Algebra Appl.*, 244:81–96, 1996.

[44] D. Petz. Information-geometry of quantum states. In *Quantum probability communications*, pages 135–157. World Sci. Publishing, River Edge, NJ, 1998.

[45] D. Petz and C. Sudár. Geometries of quantum states. *J. Math. Phys.*, 37(6):2662–2673, 1996.

[46] D. Petz and G. Tóth. The Bogoliubov inner product in quantum statistics. *Lett. Math. Phys.*, 27(3):205–216, 1993.

[47] A. Pietsch. *Nuclear locally convex spaces.* Springer-Verlag, New York, 1972. Translated from the second German edition by William H. Ruckle, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 66.

[48] G. Pistone. New ideas in nonparametric estimation. In P. Sollich et al., editors, *Disordered and Complex Systems*, pages 159–164. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[49] G. Pistone and M. P. Rogantin. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, 5(4):721–760, 1999.

[50] G. Pistone and C. Sempi. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*, 23(5):1543–1561, 1995.

[51] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcuta Mathematical Society*, 37:81–91, 1945.

[52] M. M. Rao and Z. D. Ren. *Theory of Orlicz spaces*. Marcel Dekker Inc., New York, 1991.

[53] M. Reed and B. Simon. *Methods of modern mathematical physics. II*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1975. Functional analysis.

[54] P. Sollich et al., editors. *Disordered and Complex Systems*. American Institute of Physics, 2001. AIP Conference Proceedings 553.

[55] R. F. Streater. Statistical dynamics. In *Proceedings of the XXV Symposium on Mathematical Physics (Toruń, 1992)*, volume 33, pages 203–219, 1993.

[56] R. F. Streater. *Statistical dynamics*. Imperial College Press, London, 1995. A stochastic approach to nonequilibrium thermodynamics.

[57] R. F. Streater. Information geometry and reduced quantum description. In *Proceedings of the XXVIII Symposium on Mathematical Physics (Toruń, 1995)*, volume 38, pages 419–436, 1996.

[58] R. F. Streater. Dynamics of Brownian particles in a potential. *J. Math. Phys.*, 38(9):4570–4575, 1997.

[59] R. F. Streater. A gas of Brownian particles in statistical dynamics. *J. Statist. Phys.*, 88(1-2):447–469, 1997.

[60] R. F. Streater. Nonlinear heat equations. In *Proceedings of the XXIX Symposium on Mathematical Physics (Toruń, 1996)*, volume 40, pages 557–564, 1997.

[61] R. F. Streater. Statistical dynamics and information geometry. In *Geometry and nature (Madeira, 1995)*, pages 117–131. Amer. Math. Soc., Providence, RI, 1997.

[62] R. F. Streater. A model of dense fluids. In *Quantum probability (Gdańsk, 1997)*, pages 381–393. Polish Acad. Sci., Warsaw, 1998.

[63] R. F. Streater. The analytic quantum information manifold. In *Stochastic processes, physics and geometry: new interplays, II (Leipzig, 1999)*, pages 603–611. Amer. Math. Soc., Providence, RI, 2000.

[64] R. F. Streater. The information manifold for relatively bounded potentials. *Tr. Mat. Inst. Steklova*, 228(Probl. Sovrem. Mat. Fiz.):217–235, 2000. On the ninetieth anniversary of Academician Nikolaĭ Nikolaevich Bogolyubov.

[65] R. F. Streater. The Soret and Dufour effects in statistical dynamics. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 456(1993):205–221, 2000.

[66] R. F. Streater. Correction to fluid dynamics. Submitted to Jornal of Physics A, 2001.

[67] A. Uhlmann. Density operators as an arena for differential geometry. In *Proceedings of the XXV Symposium on Mathematical Physics (Toruń, 1992)*, volume 33, pages 253–263, 1993.

[68] M. M. Vainberg. *Variational methods for the study of nonlinear operators.* Holden-Day Inc., San Francisco, Calif., 1964.

[69] J. von Neumann. *Mathematical foundations of quantum mechanics.* Princeton University Press, Princeton, NJ, 1996. Translated from the German and with a preface by Robert T. Beyer, Twelfth printing, Princeton Paperbacks.

[70] E. Zeidler. *Nonlinear functional analysis and its applications. I.* Springer-Verlag, New York, 1986. Fixed-point theorems, Translated from the German by Peter R. Wadsack.