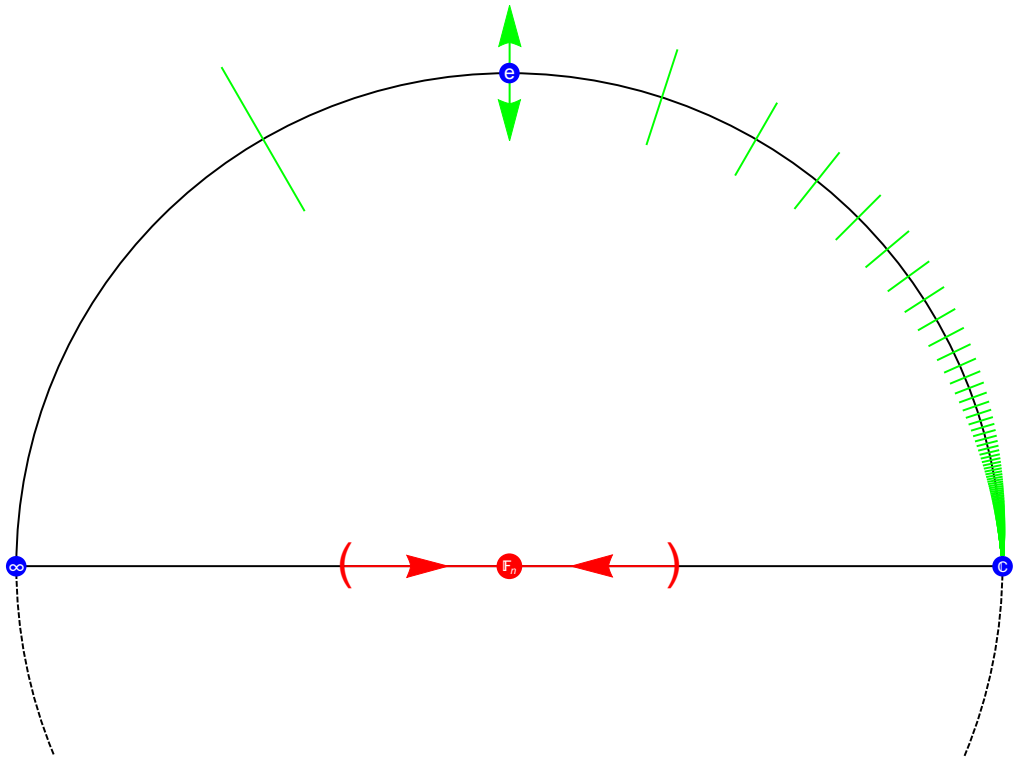
Utrecht University

Deformed CCR/CAR and Free Monotone Transport: Quons and Fock Parafermions



Author:

F.E. Elzinga
3926850

Supervisors:

Prof. Dr. E.P. van den Ban (MS)
Dr. M.P.T. Caspers (MS, Ext.)
Dr. D. Schuricht (TP)

June 7, 2019

Abstract

This thesis consists of two parts, both centred around the theme of deformations of the canonical (anti-)commutation relations.

In the first part, we provide an accessible and self-contained, yet complete, account of the recent partial resolution of the isomorphism problem concerning the q-Gaussian von Neumann algebras of Bożejko and Speicher by the free monotone transport method of Guionnet and Shlyakhtenko. We cover the necessary background in free probability, describe the construction of q-Gaussians and prove some of their elementary properties due to Ricard and Voiculescu, and provide a detailed proof of the existence of free monotone transport and its application to q-Gaussians. Some special attention is given to the commutative case and its links to random matrices, large deviations, and optimal transportation.

The second part starts with a brief review of the physics of quons, which are the particles in the Fock space realisation of the q-Gaussians. Afterwards, we turn to parafermions and their importance to edge modes and topological phases. The Fock parafermion operators due to Cobanera and Ortiz are introduced, which allow for a Fock representation of parafermions. Then, we study Fock parafermions as fundamental degrees of freedom, considering both Potts-like Hamiltonians in the spirit of Calzona et alia, and simple tight-binding Hamiltonians in the spirit of Rossini et alia. This includes mappings of Fock parafermions to electrons and mixed Fermion-Boson systems, the exact ground states for the Potts-like models, analytical evidence that the tight-binding models generally have conformal charge $c = 1$, and first steps towards constructing their phase diagram.

Contents

Introduction

On some level, the world appears to be noncommutative. A key feature of quantum mechanics, one of our fundamental descriptions of nature, is the failure of position X and momentum P of a particle to commute. Despite seeming rather innocent, the canonical commutation relation (CCR)

$$XP - PX = i\hbar,$$

which captures this failure, has all sorts of counter-intuitive consequences, like the Heisenberg uncertainty principle and the quantisation of certain observables.

When one wants to describe systems consisting of many quantum mechanical particles, it becomes convenient to reformulate the CCR in the language of second quantisation. Here, one works with operators a_j and $a_j^\dagger$ acting on a (separable) Hilbert space $\mathcal{F}(\mathcal{H})$, called a Fock space, whose actions are to respectively annihilate and create a particle, with j labelling some quantum number(s). In this thesis, the quantum number will usually be interpreted as the site number on a chain, and sometimes a spin degree of freedom will also be present. The operators a_j and $a_j^\dagger$ are each other's adjoints[1], and they satisfy certain relations derived from the CCR.

When carrying out the second quantisation procedure, the resulting reformulation depends strongly on the statistics of the particles one considers, i.e. how their combined wave function behaves under the interchange of a pair. In our own $(3+1)$-dimensional spacetime (and in higher dimensions), the problem separates neatly into two cases, depending on the spin of the particles under consideration. This is known as the spin-statistics theorem, and it states that particles with half-integer spin pick up a minus sign under exchange, while integer spin particles do not. These are called Fermionic and Bosonic statistics respectively. The former are described by the canonical anti-commutation relations (CAR)

$$a_j a_k^\dagger + a_k^\dagger a_j = \delta_{jk},$$

while the latter obey

$$a_j a_k^\dagger - a_k^\dagger a_j = \delta_{jk},$$

which are also called the CCR[2]. Models with a spacetime dimension less than $3+1$ are also of physical and mathematical interest, but there the spin-statistics theorem does not hold. This gives rise to many variations on the CCR and CAR, which one usually refers to as anyonic statistics.

In this thesis, we will study two families of commutation relations that can be viewed as deformations of the the CAR and CCR. The first of these two, called the *q-mutation relations*, we will study primarily as mathematicians, while the second, called the *Fock parafermion algebras* will be studied exclusively as physicists. Correspondingly, this thesis is split into two parts, one containing mathematics, the other physics.

Let us discuss the context of part I. During his quest to find a rigorous mathematical description of quantum mechanics, John von Neumann introduced the following objects that now bear his name. Let $\mathcal{H}$ be a separable Hilbert space and denote by $B(\mathcal{H})$ the bounded operators on $\mathcal{H}$. If $A \subset B(\mathcal{H})$, define its *commutant* A' to be the set of elements in $B(\mathcal{H})$ that commute with A. Then a *von Neumann algebra* M is a $*$-subalgebra of $B(\mathcal{H})$, containing the identity operator $\mathrm{id}_\mathcal{H}$, such that $M = M''$. From the point of view of physics, one may think of von Neumann algebras as a way to capture the behaviour of the observables of a quantum system. They are a special class of C^*-algebras that arise naturally in mathematical physics, ergodic theory, geometric group theory, and other areas of mathematics.

[1] There is the cultural difference between physicists and mathematicians that one group uses the symbol † for the adjoint, and the other group uses instead the symbol $*$. For the introduction, we will stick to the physicists conventions.

[2] From now on, we will always take CCR to refer to these relations, so there is no risk of confusion.

The q-mutation relations are now

$$a_j a_k^\dagger - q a_k^\dagger a_j = \delta_{jk},$$

with $q \in (-1, 1)$. Formally, if one were to set $q = \pm 1$, one would recover the CCR/CAR, but this is a singular limit. Nevertheless, these relations were also considered by physicists to model particles with small violations of Bosonic/Fermionic statistics (we will discuss this in Chapter 8 of Part II). Mathematically, these relations are interesting for several reasons, one being that they describe a 'noncommutative Brownian motion' with a symmetry related to the famous quantum group $\mathrm{S}_\nu\mathrm{U}(n)$ of Woronowicz [13]. One considers, mimicking the free Gaußian functor of Voiculescu (Section 2.6 in [86]), the von Neumann algebra generated by the self-adjoint operators $a_1 + a_1^\dagger, \ldots, a_n + a_n^\dagger$, which is called the *$q$-Gaußian von Neumann algebra* (with n generators). Our primary motivation for considering these relations is the case $q = 0$, as this corresponds to the so-called *free group factors* $\mathcal{L}(\mathbb{F}_n)$. These are von Neumann algebras that arise from the representation theory of the free groups on n letters.

Despite there being such a concrete description of these von Neumann algebras, it is unknown whether or not $\mathcal{L}(\mathbb{F}_n)$ and $\mathcal{L}(\mathbb{F}_m)$ are isomorphic for $n \neq m$. This is the motivating open problem behind the field of *free probability*, which was initiated by Voicelescu (and others). Since one can view the q-Gaußian von Neumann algebras as a deformation of these free groups factors, it is a natural question to ask how their isomorphism class changes with q. Starting in 1991, when Bożejko and Speicher gave the first construction of the q-Gaußian von Neumann algebras [13], the question went unanswered for almost 25 years. Many properties of these operators and algebras were investigated (as a very small sample, see [12, 14, 26, 30, 71, 78, 89]), but the isomorphism problem remained elusive. In 2014, Guionnet and Shlyakhtenko published a ground-breaking paper in *Inventiones Mathematicae* where they provided a partial resolution to the problem by using ideas from the theory of optimal transportation [43]. They found, for fixed numbers of generators, that the algebras are isomorphic for q close enough to zero. The techniques Guionnet and Shlyakhtenko introduced, have led to several partial resolutions to similar problems and are still being actively expanded, see for instance [45, 61, 62, 63, 64] and the preprint [27] by the original authors (together with Dabrowski).

In this thesis, our goal is to give an accessible, self-contained, and complete exposition of the important original result by Guionnet and Shlyakhtenko and its application to the q-Gaußian von Neumann algebras, assuming only general knowledge of the theory of operator algebras. This will require significant preparation, as we will need to venture deep into the field of free probability and understand several of its key ideas and techniques. Our start is a very brief overview of the necessary ingredients from 'classical' probability theory and optimal transportation theory in Chapter 1. Chapters 2, 4, and 5 are dedicated to developing the necessary background in free probability, culminating in the technical proof of Guionnet and Shlyakhtenko in Chapter 6. Finally, Chapter 7 covers the application of these tools to the q-Gaußian von Neumann algebras, whose construction and basic properties are treated in Chapter 3.

In part II of the thesis, we begin by reviewing some of the work from the physics community on the q-mutation relations, in Chapter 8. In particular, we discuss both the practical problems and more fundamental non-locality issues of the resulting QFTs, and the experimental evidence against their relevance. The rest of the physics section will then focus instead on a different deformation of the CCR/CAR, which has more physically relevant uses. This is the so-called Fock parafermion algebra

$$a_j a_k^\dagger - \omega^{\mathrm{sgn}(j-k)} a_k^\dagger a_j = 0, \ (k \neq j)$$

which was introduced by Cobanera and Ortiz in 2014 [22]. Here, $\omega = \exp(2\pi i/n)$ is an n-th root of unity, and the sgn is computed with respect to some fixed ordering of the labels (e.g. left to right on a chain). We will focus especially on the cases $n = 2, 3, 4$. The operators satisfying the Fock

parafermion algebra are closely related to parafermions [34], and were designed to allow for a Fock space representation of them [22]. One way to think of parafermions is to view them as a natural generalisation of the Majorana Fermion operators appearing in the solution of the Kitaev chain [47, 34]. This also explains their physical relevance, they are frequently associated with topological phases and edge zero modes [2, 21, 34]. The presence of edge zero modes implies a degenerate groundstate manifold, which if gapped, provides a dream platform to implement topological quantum computing by exploiting the anyonic properties of parafermions [48, 60]. Chapter 9 explains the above statements about parafermions in detail (aside from their precise relation to quantum computing), and Chapter 10 establishes the Fock parafermions.

Having amply motivated their relevance and utility, we will proceed to investigate the physics of the Fock parafermion algebra from a fundamental point of view. In Chapter 11, we consider Fock parafermion Hamiltonians that are 'close' to the Potts model, and introduce a powerful mapping from $n = 4$ Fock parafermions to electrons [18]. We find the groundstates of these models, using techniques from [46], determine their electronic representations, and compute spectral functions at two special points following [18]. The last chapter of the thesis, Chapter 12, deals with elementary tight-binding models for Fock parafermions as initiated by [74]. We attempt to apply the techniques developed in the previous chapter to give analytical support to the known numerical results from [54, 74], and take the first steps to explore the phase diagram of these models; this is joint work with Iman Mahyaeh (Stockholm), Jurriaan Wouters, and Dirk Schuricht.

Part I: Mathematics

1 Background and Prerequisites

1.1 Measure Theoretic Probability

As we will be working in the setting of 'non-commutative' probability theory (whatever that may be), a brief recapitulation of the basics of 'ordinary' probability theory may be useful (for a more complete treatment one could consult for instance [73]). The treatment here will be far from complete, it will only cover the concepts that we will generalise or refer to again later. Moreover we assume that the reader is already familiar with measure theory.

Definition 1.1. Let Ω be a set, $\mathcal{F}$ a σ-algebra on Ω, and $\mathbb{P}$ a (positive) finite measure on the measurable space $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. Then we refer to the triple $(\Omega, \mathcal{F}, \mathbb{P})$ as a *probability space*. In particular, we say that Ω is the set of *outcomes*, $\mathcal{F}$ is the σ-algebra of *events*, $\mathbb{P}$ is a *probability measure*, and we say *almost surely* instead of $\mathbb{P}$-almost everywhere.

Up to normalisation, any finite measure space is a probability space, hence the above definition is quite general. However, this also means that the notion of a probability space is quite abstract. To probe the structure of a given probability space, we can look at functions defined on it, which have the real or complex numbers (the symbol $\mathbb{K}$ may refer to either) as target. Whenever we talk about the real or complex numbers as a measurable space in this section, we will always assume that they come equipped with their respective Borel σ-algebras, and we denote the Lebesgue measure by λ. This motivates the following definition.

Definition 1.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, then a measurable function $X : \Omega \to \mathbb{K}$ is called a *random variable*. A measurable function $(X_1, \ldots, X_n)$ $(n \in \mathbb{N})$ from Ω to $\mathbb{K}^n$ is called a *random n-vector*. We write $\sigma(X) \subset \mathcal{F}$ for the smallest σ-algebra on Ω that makes X measurable. In the event that $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, we define the *expectation* of X as its integral over Ω, with the notation

$$\mathbb{E}[X] := \int_\Omega X \, \mathrm{d}\mathbb{P}.$$

An important observation is the following.

Remark 1.3. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X on it, we have a probability measure μ_X on $\mathbb{K}$ defined by the formula

$$\mu_X[B] := \mathbb{P}[X^{-1}(B)],$$

where B is a Borel set of $\mathbb{K}$. Notice that this is simply the pushforward of $\mathbb{P}$ by X, which we write in the sequel as $\mu_X = X_*\mathbb{P}$.

Definition 1.4. The measure obtained in the preceding remark is called the *distribution* or *law* of X. If $Y = (Y_1, \ldots, Y_n)$ is a random n-vector, its law on $\mathbb{K}^n$ is called the *joint law* of $Y_1, \ldots, Y_n$. Conversely, for a probability measure μ on $\mathbb{K}^n$, the probability measures μ_i $(1 \leq i \leq n)$ on $\mathbb{K}$ obtained by setting

$$\mu_i[B] = \mu[\mathbb{K}^{i-1} \times B \times \mathbb{K}^{n-i}],$$

are called the *marginal laws* of μ.

The next result shows that if we can determine the law of a random variable, we also obtain information on related random variables.

Proposition 1.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and X a random variable. If g is a measurable function from $\mathbb{K}$ to $\mathbb{K}$, then $g \circ X$ is again a random variable, which has finite expectation if and only if g is $X_*\mathbb{P}$ integrable. In this case, we have the formula*

$$\mathbb{E}[g \circ X] = \int_{\mathbb{K}} g \ \mathrm{d}(X_*\mathbb{P}) .$$

In the case that $\mathbb{K} = \mathbb{R}$, there is an important inequality that compares applying g to X and then $\mathbb{E}$ with the reversed order.

Lemma 1.6 (Jensen's Inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and X a real-valued random variable. Suppose that $g, h : \mathbb{R} \to \mathbb{R}$ are both measurable functions, then*

(i) if g is convex, $g(\mathbb{E}[X]) \leq \mathbb{E}[g \circ X]$,

(ii) if h is concave, $\mathbb{E}[h \circ X] \leq h(\mathbb{E}[X])$.

See Definition 1.23 for the definition of a convex function (the definition of a concave function has the inequality reversed). Proposition 1.5 also allows us (at least formally) to associate an important family of numbers to a random variable.

Definition 1.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and X a random variable. The *moments* of X are the numbers $\mathbb{E}[X^n]$ $(n \in \mathbb{N})$, provided they exist. The exponent n is referred to as the *order* of a moment. The first moment is also called the *mean* (as well as expectation), and the combination $\mathbb{E}[X^2] - \mathbb{E}[X]^2$ is called the *variance* (notice that it is positive by Jensen's inequality).

A convenient way to combine all the moments together is the *moment generating function*, which if it exists is defined by

$$M(t) := \mathbb{E}[e^{tX}].$$

Formally, this means that

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}[X^n], \ \mathbb{E}[X^m] = \left(\frac{\mathrm{d}}{\mathrm{d}t} \right)^m \Bigg|_{t=0} M(t).$$

It turns out to be sometimes useful to consider instead the logarithm of M, called K, the *cumulant generating function*,

$$K(t) := \log(M(t)) =: \sum_{n=1}^{\infty} \frac{t^n}{n!} \kappa_n(X),$$

where the κ_n are called *cumulants*.

We now consider the crucial concept of 'independence', which imposes (additional) algebraic structure on the theory.

Definition 1.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, then a sequence of random variables, $\{X_i\}_{i \in \mathbb{N}}$, is called *independent*, iff for all natural numbers n, and for any choice of $E_j \in \sigma(X_{i_j})$ $(j = 1, \ldots, n)$, we have the identity $\mathbb{P}[E_1 \cap \cdots \cap E_n] = \prod_{i=1}^{n} \mathbb{P}[E_i]$.

Remark 1.9. As an example of the algebraic structure imposed by independence, consider the well-known result that if X and Y are independent random variables with finite expectations, then their product has finite expectation and we have the relation $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ between them.

We now briefly discuss densities for random variables.

Definition 1.10. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X a random variable. If X allows for the existence of a non-negative measurable function $f_X : \mathbb{K} \to \mathbb{K}$ with the property that

$$(X_*\mathbb{P})[B] = \int_{\mathbb{K}} \chi_B(x) f_X(x) \, \mathrm{d}\lambda(x),$$

for all Borel sets B, we call this function f_X a *density* for X.

The following result identifies when a random variable admits a density.

Theorem 1.11 (Radon-Nikodym)**.** *Let ν be a probability measure on $\mathbb{K}$, then there exists a unique decomposition $\nu = \nu_{ac} + \nu_s$ with the following properties. There is a positive Lebesgue-integrable function n on $\mathbb{K}$ such that*

$$\nu_{ac}[B] = \int_{\mathbb{K}} \chi_B(x) n(x) \, \mathrm{d}\lambda(x),$$

for all Borel sets B. In particular ν_{ac} is Lebesgue-absolutely continuous. The function n is unique in the sense that for any other function n' satisfying the above relation, it holds that $n = n'$ Lebesgue-almost everywhere. The remaining term in the decomposition, ν_s, is mutually singular with respect to λ. It follows that the law of a random variable admits a density precisely when it is Lebesgue-absolutely continuous.

Definition 1.12. Assuming the setting of the above theorem, the function n is called the *Radon-Nikodym derivative* of ν, and is denoted by $n = \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}$.

Example 1.13. Notice that in the right hand side of the formula given in Proposition 1.5, the random variable (and hence the underlying abstract probability space) only enter implicitly in the integration measure. This suggests that one could also proceed 'axiomatically' by studying probability measures on $\mathbb{K}$ (or $\mathbb{K}^n$) without worrying about the underlying structure. In this example we will pursue this route and calculate the 'moments' of the probability measure defined by the density

$$f(x) = \chi_{\{|x| \leq \rho\}}(x) \frac{2}{\pi \rho^2} \sqrt{\rho^2 - x^2},$$

where $\rho > 0$. This function is called the *Wigner semicircle density*. The corresponding measure will be denoted by σ. Will we still use the formal notation $\mathbb{E}[X^n]$ for the integrals

$$\mathbb{E}[X^n] := \frac{2}{\pi \rho^2} \int_{-\rho}^{\rho} x^n \sqrt{\rho^2 - x^2} \, \mathrm{d}x,$$

and note that by definition $\mathbb{E}[X^0] = 1$. By symmetry, all the odd order moments are identically 0, so we may assume that $n = 2k$. Using the substitution $x = \rho \sin \vartheta$, a trigonometric identity, and a cleverly chosen partial integration, one can derive a recurrence relation for the moments, namely

$$\mathbb{E}[X^{2k}] = \rho^2 \frac{2k-1}{2k+2} \mathbb{E}[X^{2k-2}].$$

One easily shows (or checks) that this is solved by

$$\mathbb{E}[X^{2k}] = \left(\frac{\rho}{2}\right)^{2k} C_k,$$

where the *Catalan numbers* C_k are defined as

$$C_k = \frac{1}{k+1} \binom{2k}{k}.$$

Example 1.14. In this example, we consider the *Gaußian density* given by the formula

$$\gamma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

where $\sigma > 0$ is a number, not to be confused with the Wigner semicircle measure. The case $\sigma = 1$ is usually referred to as the *standard Gaußian density*. Its moments are simple to calculate using partial integration and symmetry arguments, they are

$$\mathbb{E}[X^n] = \begin{cases} 0 & \text{if } n = 2k+1, \\ \sigma^{2k}(2k-1)!! & \text{if } n = 2k, \end{cases}$$

where we have used the double factorial. There is also a multivariate version of this density for the case of a random N-vector. Let Σ be a symmetric positive definite matrix, called the *covariance matrix*, and write $x = (x_1, \ldots, x_N)$. Then the multivariate Gaußian density is

$$\gamma_N(x_1, \ldots, x_N) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right). \tag{1.1.1}$$

Note that Σ^{-1} exists by positive definiteness of Σ. Again, it is not hard to show (by using the fact that Σ can be diagonalised) that

$$\mathbb{E}[x_i] = 0, \ \mathbb{E}[x_i x_j] = \Sigma_{ij}, \ \forall \ 1 \leq i, j \leq N.$$

This is why Σ is called the covariance matrix, as the numbers $\mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$ are called *covariances*. We can also compute the cumulants κ_n of γ, by completing the square in the integral

$$\mathbb{E}[e^{tX}] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(tx - \frac{x^2}{2\sigma^2}\right) \, \mathrm{d}x = \exp\left(\frac{t^2\sigma^2}{2}\right),$$

so that $K(t) = \sigma^2 \, (t^2/2)$, and hence $\kappa_2 = \sigma^2$ and all other cumulants vanish. While it may not appear so now, the semi-circle and Gaußian densities are analogous in many ways.

1.2 Entropy and Fisher Information

In this section, we continue our summary of some important ideas from (measure theoretic) probability. We introduce the concepts of entropy, Fisher information, and score functions (mostly following [25]). Any probability measure considered in this section will be assumed to have a real density (usually denoted p or q), hence the brackets around 'measure theoretic'.

Entropy S has a long history and has found manifold application in physics and mathematics (and elsewhere), but for our purposes it is simply some quantity given by the formula

$$S(p) := -\int p \log(p) \, \mathrm{d}\lambda.$$

Left implicit in our definition, the integral is over the support of p. Note that the entropy need not exist for every density p. In general we will pursue a less formal style of exposition in this section. If we now take two densities p and q, we can define their *relative entropy*

$$S(p||q) := \int p \log\left(\frac{p}{q}\right) \, \mathrm{d}\lambda,$$

provided that we interpret $0\log(0) = 0 = 0\log(0/0)$ and the support of p is contained in that of q. If this is the case, we can use Jensen's inequality to derive the bound

$$S(p||q) = \mathbb{E}_p\left[-\log\left(\frac{q}{p}\right)\right] \geq -\log\left(\mathbb{E}_p\left[\frac{q}{p}\right]\right) = \log(\mathbb{E}_q[1]) = \log(1) = 0.$$

We used the notation that a subscript on the expectation indicates with respect to which density the integration is to be done, and we used that $-\log(x)$ is convex. This bound has the following important application.

Proposition 1.15. *Let $N \geq 1$ be a natural number, and consider probability densities on $\mathbb{R}^N$. Among all densities p whose mean vanishes and whose covariance matrix is Σ (and which have a well-defined entropy), the Gaußian density γ_N is the one with maximal entropy. This implies, for all such p, the bound*

$$S(p) \leq \frac{1}{2}\log\left[(2\pi e)^N \det(\Sigma)\right].$$

Proof. As the support of γ_N is all of $\mathbb{R}^N$, the relative entropy of any p with respect to γ_N exists. Thus we have

$$0 \leq S(p||\gamma_N) = \mathbb{E}_p[\log(p)] - \mathbb{E}_p[\log(\gamma_N)].$$

The crucial observation is now that $\log(\gamma_N)$ is a quadratic form according to (1.1.1), and since this is a linear combination of products of coordinates $x_i x_j$, its expectation is completely determined by the covariance matrix of the density, which is Σ for both p and γ_N. Hence, we may replace the p average of $\log(\gamma_N)$ with the γ_N average. This yields

$$0 \leq -S(p) - \mathbb{E}_{\gamma_N}[\log(\gamma_N)] = S(\gamma_N) - S(p),$$

and we have proven that the Gaußian density maximises the entropy.

For the explicit bound given above, we must calculate the value of $S(\gamma_N)$. This can be done by writing out

$$\begin{aligned} S(\gamma_N) &= -\int_{\mathbb{R}^N} \gamma_N(x)\left(-\frac{1}{2}x^T\Sigma^{-1}x - \log\left[\sqrt{(2\pi)^N\det(\Sigma)}\right]\right)\,\mathrm{d}x, \\ &= \frac{1}{2}\log\left[(2\pi)^N\det(\Sigma)\right] + \mathbb{E}_{\gamma_N}\left[\frac{1}{2}\sum_{i,j=1}^N x_i\left(\Sigma^{-1}\right)_{ij}x_j\right], \end{aligned}$$

but this last term is simply

$$\frac{1}{2}\sum_{i,j=1}^N \left(\Sigma^{-1}\right)_{ij}\mathbb{E}_{\gamma_N}[x_j x_i] = \frac{1}{2}\operatorname{Tr}\left[\Sigma^{-1}\Sigma\right] = \frac{N}{2} = \frac{1}{2}\log\left(e^N\right).$$

Thus

$$S(\gamma_N) = \frac{1}{2}\log\left[(2\pi)^N\det(\Sigma)\right] + \frac{1}{2}\log\left(e^N\right) = \frac{1}{2}\log\left[(2\pi e)^N\det(\Sigma)\right],$$

as claimed. □

We now turn to the Fisher information. Given a probability density p on $\mathbb{R}$, one might wonder how 'stable' it is with respect to 'relaxation'. One way to quantify this, would be to consider a solution p_t of the diffusion equation

$$\dot{p}_t = p_t'',$$

with initial condition $p_0 = p$, where a dot indicates a time derivative and a prime a spatial derivative. Then along this solution we have the entropies $S(p_t)$, and we may look at

$$\frac{\mathrm{d}}{\mathrm{d}t} S(p_t) = -\int \left[\dot{p}_t \log(p_t) + \frac{p_t}{p_t} \dot{p}_t \right] \mathrm{d}\lambda = -\int \dot{p}_t \log(p_t)\ \mathrm{d}\lambda - \frac{\mathrm{d}}{\mathrm{d}t} \int p_t\ \mathrm{d}\lambda.$$

Now, the last term on the right vanishes because the diffusion equation preserves normalisation. In the other term on the right we can use the diffusion equation and then integrate by parts (assuming the boundary terms vanish) to find

$$\frac{\mathrm{d}}{\mathrm{d}t} S(p_t) = \int \frac{(p_t')^2}{p_t} \mathrm{d}\lambda.$$

Taking this at $t = 0$ we find the *Fisher information*

$$\Phi(p) := \frac{\mathrm{d}}{\mathrm{d}t} S(p_t)\bigg|_{t=0} = \int \frac{(p')^2}{p} \mathrm{d}\lambda = \mathbb{E}_p \left[\left(-\frac{p'}{p} \right)^2 \right] =: \mathbb{E}_p \left[\xi^2 \right],$$

where we have also defined the *score function* $\xi = -p'/p$. While we will not do very much with the Fisher information, the score functions we will encounter again later. The reason for this is that they satisfy a certain relation, namely let f be a sufficiently regular and integrable function, then

$$\mathbb{E}_p[\xi f] = -\int p' f\ \mathrm{d}\lambda = \int p f'\ \mathrm{d}\lambda = \mathbb{E}_p[f'].$$

This can be phrased another way if the support of p is all of $\mathbb{R}$, in this case we have an inner product given by

$$\langle \xi, f \rangle = \mathbb{E}_p[\xi f] = \mathbb{E}_p[f'] = \langle 1, f' \rangle = \left\langle \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^* 1, f \right\rangle,$$

so that in some sense $\xi = (\mathrm{d}_x)^* 1$. In the multivariate case, the above generalises to give score functions ξ_i, as many as the dimension N, which satisfy

$$\xi_i = -\frac{\partial_i p}{p},\ \mathbb{E}_p[\xi_i g] = \mathbb{E}_p[\partial_i g],\ \xi_i = \partial_i^* 1.$$

Example 1.16. To conclude this section, we will derive the score functions for the Wigner semi-circle and Gaußian densities discussed in examples 1.13 and 1.14 respectively. For the former,

$$p'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{2}{\pi\rho^2} \sqrt{\rho^2 - x^2} \right] = -\frac{2}{\pi\rho^2} \frac{x}{\sqrt{\rho^2 - x^2}},$$

so that

$$\xi_{\mathrm{Wss}} = -\frac{-4\pi\rho^2 x}{4\pi\rho^2 \left(\sqrt{\rho^2 - x^2} \right)^2} = \frac{x}{\rho^2 - x^2}.$$

While for the latter,

$$\gamma'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{x^2}{2\sigma^2}\right)\right] = -\left(\frac{2x}{2\sigma^2}\right)\gamma(x),$$

so that

$$\xi_G = \frac{x}{\sigma^2}.$$

In particular, for the standard Gaußian density we have the interesting result (with some slight abuse of notation) that $\mathrm{d}_x^* 1 = x$ for the inner product

$$\langle f, g\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)g(x)e^{-x^2/2}\,\mathrm{d}x$$

on $L^2(\mathbb{R})$ (with respect to the Gaußian density). It is also worth remarking here, that in this setting applying the Gram-Schmidt algorithm to the set of monomials (which is contained in $L^2(\mathbb{R})$ with this inner product) results in a collection of polynomials that are closely related to the Hermite polynomials (up to normalisation and scaling).

1.3 Optimal Transportation Theory

Optimal transportation theory came to be when mathematicians formalised the notion of moving mass from one configuration in space to another in the cheapest way possible (for example moving ore from mines to refineries). Our primary interest in the theory, is that it produces transport maps with good properties between measures, see Brenier's theorem below. Our presentation will follow the one in [85]. Throughout this section, μ and ν denote probability measures on $\mathbb{R}^n$.

Definition 1.17. A *transport map* between μ and ν is a measurable function $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T_*\mu = \nu$. A non-negative continuous function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ will be called a *cost function*.

The idea is now to search for a transport map T that minimises the total cost, which we achieve by demanding that T is a minimiser of the functional

$$I[T] := \int_{\mathbb{R}^n} c(x, T(x))\,\mathrm{d}\mu(x).$$

However, this is a very harsh condition. One sees that we are not allowed to divide any of the mass at a point x. Rather, it all has to be sent to the unique destination $T(x)$. Moreover, the existence of any transport map at all is not a priori clear. Suppose that the measures μ and ν are regular enough to admit densities m and n respectively, and assume that there exists a transport map T that is a C^1 diffeomorphism. By the change of variables formula, the condition that $T_*\mu = \nu$ can be reformulated as

$$m(x) = n(T(x))|\det DT(x)|,$$

which is a non-linear partial differential equation.

This suggests that we should search for a 'relaxation' of the problem, which is hopefully easier to solve. As a first step, one can observe that we can rewrite

$$I[T] = \int_{\mathbb{R}^n\times\mathbb{R}^n} c(x, y)\,\mathrm{d}\mu(x)\delta[y = T(x)].$$

Now, we are integrating with respect to a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$, and it is easy to check that its marginals are μ and ν. It is now natural to set up the problem in the following way.

Definition 1.18. Let π be a probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ and ν (see Definition 1.4), then we call π a *transference plan* and denote the set of all transference plans by $\Pi(\mu, \nu)$. We say that π is an *optimal* transference plan, if it is a minimiser of the functional

$$I[\pi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) \, \mathrm{d}\pi(x, y).$$

One way to see that this is a considerably easier problem, is that this functional is linear in π. However, one should note that the set $\Pi(\mu, \nu)$ is not a linear space, although it is convex and non-empty (it always contains the product measure $\mu \otimes \nu$).

A great advantage of this relaxed formulation is the following result.

Theorem 1.19 (Kantorovich Duality)**.** *We have the equality*

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi) \in F} J(\varphi, \psi),$$

where the functional J is defined on $L^1(\mu) \times L^1(\nu)$ by the formula

$$J(\varphi, \psi) := \int_{\mathbb{R}^n} \varphi(x) \, \mathrm{d}\mu(x) + \int_{\mathbb{R}^n} \psi(y) \, \mathrm{d}\nu(y),$$

and

$$F := \left\{ (\varphi, \psi) \in L^1(\mu) \times L^1(\nu) \middle| \, \varphi(x) + \psi(y) \leq c(x, y) \text{ for } \mu\text{−almost all } x, \ \nu\text{−almost all } y \right\}.$$

We will now specify to the case $c(x, y) = \|x - y\|^2/2$ and assume that μ and ν have finite second moments. Then we can immediately prove that existence of a minimiser. Note that $\Pi(\mu, \nu)$ is non-empty, because it contains the measure $\mu \otimes \nu$. Consider the weak topology induced on $\Pi(\mu, \nu)$ by $C_b(\mathbb{R}^n \times \mathbb{R}^n)$, then one can show that $\Pi(\mu, \nu)$ is weakly compact. Thus a minimising sequence $(\pi_n)_{n \in \mathbb{N}}$ admits a limit point π. Notice that there exists a non-decreasing sequence $(c_m)_{m \in \mathbb{N}}$ of bounded continuous functions that approximates our cost function point-wise. Then by an application of the monotone convergence theorem one deduces that the limit point π is in fact a minimiser.

However, one can prove much more. To state the technical conditions, we need a bit of terminology.

Definition 1.20. Let U be a subset of $\mathbb{R}^n$, then one can define its *Hausdorff dimension*

$$\dim_{\mathrm{H}}(U) := \inf \left\{ d \geq 0 \middle| \inf \left\{ \sum_{i \in I} r_i^d \middle| (B(x_i; r_i))_{i \in I} \text{ is a cover of } U \right\} = 0 \right\}.$$

A subset V of $\mathbb{R}^n$ is *small* iff $\dim_{\mathrm{H}}(V) \leq n - 1$. We call a measure ρ on $\mathbb{R}^n$ *lipophilic* iff every small set is (contained in) a ρ-null set. Notice that Lebesgue absolutely continuous measures are lipophilic.

Theorem 1.21 (Brenier)**.** *If μ is lipophilic (in addition to our other standing assumptions), then there is a unique optimal transference plan $\overline{\pi}$, given by the formula $\mathrm{d}\overline{\pi}(x, y) = \mathrm{d}\mu(x)\delta[y = \nabla\overline{\varphi}(x)]$. Here, $\nabla\overline{\varphi}$ is the μ-almost every unique gradient of a convex function $\overline{\varphi}$, satisfying $\nabla\overline{\varphi}_* \mu = \nu$ and* $\operatorname{supp} \nu = \overline{\nabla\overline{\varphi}(\operatorname{supp} \mu)}$ *Moreover, $\nabla\overline{\varphi}$ is the unique solution to our original problem, that is*

$$\inf_{T_*\mu = \nu} \int_{\mathbb{R}^n} \frac{1}{2} \|x - T(x)\|^2 \, \mathrm{d}\mu(x) = \int_{\mathbb{R}^n} \frac{1}{2} \|x - \nabla\overline{\varphi}(x)\|^2 \, \mathrm{d}\mu(x).$$

Remark 1.22. $\nabla\overline{\varphi}$ is often called the *Brenier map*.

In the remainder of this section, we will sketch the proof of this theorem and discuss its main ingredients. The first step is to exploit the algebraic properties of our cost function to reformulate the problem. To this end, one notices that

$$\varphi(x) + \psi(y) \leq \frac{1}{2}\|x-y\|^2 \;\Leftrightarrow\; x \cdot y \leq \left[\frac{\|x\|^2}{2} - \varphi(x)\right]\left[\frac{\|y\|^2}{2} - \psi(y)\right].$$

This will turn out to be a more convenient condition to work with, so with abuse of notation we again write φ and ψ for the above expressions in square brackets. Additionally, recall our notation for moments of a probability measure, in particular we will write $\mathbb{E}[X^n]$ for the moments of μ and $\mathbb{E}[Y^n]$ for those of ν. Then we can write

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \frac{1}{2}\mathbb{E}\left[X^2\right] + \frac{1}{2}\mathbb{E}\left[Y^2\right] - \sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} (x\cdot y)\,\mathrm{d}\pi(x,y),$$

while

$$\sup_{(\varphi,\psi)\in F} J(\varphi,\psi) = \frac{1}{2}\mathbb{E}\left[X^2\right] + \frac{1}{2}\mathbb{E}\left[Y^2\right] - \inf_{(\varphi,\psi)\in G} J(\varphi,\psi),$$

where

$$G := \left\{ (\varphi,\psi) \in L^1(\mu)\times L^1(\nu) \middle|\, x\cdot y \leq \varphi(x)+\psi(y) \text{ for } \mu\text{-almost all } x,\ \nu\text{-almost all } y \right\}.$$

By Kantorovich duality, these expressions are equal, and since μ and ν have finite second moments, we can subtract those terms from both sides to obtain

$$\sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} (x\cdot y)\,\mathrm{d}\pi(x,y) = \inf_{(\varphi,\psi)\in G} \left(\int_{\mathbb{R}^n} \varphi(x)\ \mathrm{d}\mu(x) + \int_{\mathbb{R}^n} \psi(y)\ \mathrm{d}\nu(y) \right).$$

Let us take a closer look at the condition for a pair to belong to G. It can also be written

$$\psi(y) \geq x\cdot y - \varphi(x).$$

Intuitively, a pair for which the inequality is as sharp as possible should give a small value of J, so it makes sense to take the supremum[3] over x,

$$\psi(y) \geq \sup_x \left(x\cdot y - \varphi(x)\right) =: \varphi^*(y),$$

where φ^* is the *convex conjugate* (or *Legendre transform*) of φ. By linearity of J, this means that for every pair (φ,ψ) we can minimise further by replacing ψ by φ^* (assuming integrability). This is a crucial observation, as we can now restrict attention to convex conjugate pairs of functions, and bring into play the technology of convex analysis.

Let us therefore take a moment to introduce some of the machinery that will aid us later on.

Definition 1.23. Let $\varphi : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ be a function, not identically ∞, satisfying for all $x, y \in \mathbb{R}^n$ and all $t \in [0,1]$ the inequality

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

Then we call φ a *proper convex function*, and iff equality in this condition implies that $x = y$ or $t = 0, 1$, we say that φ is *strictly convex*. Write $\mathrm{Dom}(\varphi)$ for the set of those points where φ is not equal to ∞.

[3]There are some measure theoretical subtleties we are omitting here.

Some immediate consequences of this definition are that $\mathrm{Dom}(\varphi)$ is convex, and hence that its boundary is a small set. Moreover, we have that φ is both continuous and even locally Lipschitz in the interior of $\mathrm{Dom}(\varphi)$. By Rademacher's theorem, this is enough to conclude that φ is differentiable almost everywhere in the interior of its domain. Suppose that φ is differentiable at x, with gradient $\nabla\varphi$, then an investigation of the difference quotient yields the estimate

$$\varphi(z) \geq \varphi(x) + \nabla\varphi(x) \cdot (z - x),$$

for all $z \in \mathbb{R}^n$.

Definition 1.24. A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is called *monotone*, iff $(f(x) - f(y)) \cdot (x - y) \geq 0$ for all $x, y \in \mathbb{R}^n$. If φ is a convex function, introduce its *subdifferential* at x as the set

$$\partial\varphi(x) := \{\, y \in \mathbb{R}^n |\, \varphi(z) \geq \varphi(x) + y \cdot (z - x),\ \forall z \in \mathbb{R}^n \}\,.$$

Notice that the gradient of a convex function is monotone by the estimate given above. It can be shown that the subdifferential is always non-empty, and that if φ is differentiable at x with gradient $\nabla\varphi(x)$, then $\partial\varphi(x) = \{\nabla\varphi(x)\}$. A simple calculation gives the following equivalent characterisation of a vector y belonging to the subdifferential at x,

$$y \in \partial\varphi(x) \ \Leftrightarrow\ x \cdot y = \varphi(x) + \varphi^*(y).$$

Compare this with the condition for the pair (φ, φ^*) to belong to G, which suggests part of the statement of Brenier's theorem: an optimal transport map should send the mass at a point x to a 'point' in the subdifferential of φ at x. One now proceeds to use what we have discussed so far to prove that there indeed exists a minimising pair (φ, φ^*) (which includes proving integrability of the limit functions, etc.). We will not discuss the details here, as there is a technical issue that has to be overcome, related to the symmetry of J that

$$J(\varphi, \psi) = J(\varphi + c, \psi - c),$$

for any $c \in \mathbb{R}$.

To recapitulate, so far we have argued that there exists an optimal transference plan $\overline{\pi}$, and a minimising pair $(\overline{\varphi}, \overline{\varphi}^*)$ of convex conjugate functions. It is now time to explore the connection between the two solutions. Recall that Kantorovich duality gave us the equality

$$\sup_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y)\, \mathrm{d}\pi(x, y) = \inf_{(\varphi,\psi) \in G} \left(\int_{\mathbb{R}^n} \varphi(x)\ \mathrm{d}\mu(x) + \int_{\mathbb{R}^n} \psi(y)\ \mathrm{d}\nu(y) \right),$$

which can be recast, using the properties of $\overline{\pi}$ and the above, into

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y)\, \mathrm{d}\overline{\pi}(x, y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \big[\overline{\varphi}(x) + \overline{\varphi}^*(y)\big]\, \mathrm{d}\overline{\pi}(x, y).$$

It follows that

$$0 = \int_{\mathbb{R}^n \times \mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(y) - x \cdot y]\, \mathrm{d}\overline{\pi}(x, y),$$

and as the integrand is non-negative by construction, it must hold that

$$\overline{\varphi}(x) + \overline{\varphi}^*(y) - x \cdot y = 0,$$

$\overline{\pi}$-almost everywhere. As we saw above, this is equivalent to saying that $y \in \partial\overline{\varphi}(x)$, $\overline{\pi}$-almost everywhere.

We now strengthen this result. Since $\overline{\varphi}$ is μ-integrable, it must hold that $\mu[\{\overline{\varphi} = \infty\}] = 0$, so that $\mu[\mathrm{Dom}(\overline{\varphi})] = 1$. As the boundary of $\mathrm{Dom}(\overline{\varphi})$ is a small set and μ is lipophilic, it even holds that $\mu[\mathrm{Int}(\mathrm{Dom}(\overline{\varphi}))] = 1$. By the properties of convex functions discussed earlier, $\overline{\varphi}$ is μ-almost everywhere differentiable and at these points $\partial\overline{\varphi} = \{\nabla\overline{\varphi}\}$. It follows that $y = \nabla\overline{\varphi}(x)$ $\overline{\pi}$-almost everywhere, and therefore that $\mathrm{d}\overline{\pi}(x, y) = \mathrm{d}\mu(x)\delta[y = \nabla\overline{\varphi}(x)]$ and $(\nabla\overline{\varphi})_*\mu = \nu$. Showing μ-almost everywhere uniqueness of $\nabla\overline{\varphi}$ is easy, and the proof also implies that $\nabla\overline{\varphi}$ solves the original problem. We make no comment on the result that $\mathrm{supp}\ \nu = \overline{\nabla\overline{\varphi}(\mathrm{supp}\ \mu)}$.

1.4 The Monge-Ampère Equation

Let μ and ν be two probability measures on $\mathbb{R}^n$, both Lebesgue absolutely continuous and with finite second moments. Then by the Radon-Nikodym theorem we have densities m and n for μ and ν respectively, which we will assume are both strictly positive. By Brenier's theorem, there exists a μ-almost everywhere unique gradient of a convex function, $\nabla\varphi$, such that $\nabla\varphi_*\mu = \nu$. In particular, if we let $\psi \in C_b(\mathbb{R}^n)$ be any test function, it holds that

$$\int_{\mathbb{R}^n} \psi(y)n(y)\ \mathrm{d}y = \int_{\mathbb{R}^n} \psi(\nabla\varphi(x))m(x)\ \mathrm{d}x.$$

Making the regularity assumption that $\nabla\varphi$ is C^1, and the non-degeneracy assumption that $\nabla\varphi$ is injective, we can change coordinates to $\nabla\varphi(x)$ in the left integral (recall that $\mathrm{supp}\ \nu = \overline{\nabla\varphi(\mathrm{supp}\ \mu)}$) to find

$$\int_{\mathbb{R}^n} \psi(y)n(y)\ \mathrm{d}y = \int_{\mathbb{R}^n} \psi(\nabla\varphi(x))n(\nabla\varphi(x))\det\big(D^2\varphi(x)\big)\ \mathrm{d}x.$$

Combining the above two equalities, and recalling that the test function ψ was arbitrary, yields

$$\det\big(D^2\varphi(x)\big) = \frac{m(x)}{n(\nabla\varphi(x))}.$$

Definition 1.25. A *Monge-Ampère* equation, is a partial differential equation of the form

$$\det\big(D^2 f(x)\big) = F(x, f(x), Df(x)),$$

where the function f is the unknown.

While this is in general a non-linear PDE, there is some hope. Assume that $\nabla\varphi$ is close to the identity, then n must also be close to m. More precisely, suppose that

$$\varphi(x) = \frac{\|x\|^2}{2} + \epsilon\eta(x) + \mathcal{O}(\epsilon^2),\ \ n(x) = \big(1 + \epsilon h(x) + \mathcal{O}(\epsilon^2)\big)\, m(x).$$

Plugging this into our Monge-Ampère equation and expanding in ϵ, one finds

$$\begin{aligned} m(x) &= [m(\nabla\varphi(x)) + \epsilon h(\nabla\varphi(x))m(\nabla\varphi(x))]\det\big(I_{n\times n} + \epsilon D^2\eta(x)\big) + \mathcal{O}(\epsilon^2), \\ &= m(x)\left[1 + \epsilon\,\mathrm{Tr}\big(D^2\eta(x)\big)\right] + \epsilon\left[h(x)m(x) + \nabla m(x)\cdot\nabla\eta(x)\right] + \mathcal{O}(\epsilon^2). \end{aligned}$$

Rearranging terms, one finally arrives at

$$\left[-\Delta + \nabla\left(-\log(m(x))\right)\cdot\nabla\right]\eta(x) = h(x).$$

This is a linear and elliptic PDE, a class whose existence and uniqueness theory is very well understood (see for instance Chapter 6 in [32]). Much later in the thesis, we will tackle a noncommutative analogue of the Monge-Ampére equation, and there we will also assume that the solution is close to the identity, which will allow for the equation to be solved.

2 Free Probability Theory

2.1 Noncommutative Probability Theory

This chapter is based on Part 1 of [65], we assume familiarity with operator algebra theory at the level of [57]. There are many reasons to seek a generalisation of the theory of probability that includes non-commutative objects, see for instance the survey article [80]. Motivation to study 'non-commutative probability' also comes from outside of mathematics, in particular from quantum physics. In that formalism, one associates to a quantum system an algebra of 'observables' acting on a Hilbert space of 'states' (not to be confused with a state on a C*-algebra, although the two are related) of the system. This algebra is not commutative, as is already evident from the 'canonical commutation relations'

$$[X, P] = i\hbar I,$$

where X and P are operators representing position and momentum respectively, and $\hbar$ is Planck's reduced constant. Nevertheless, one interprets, if for instance the quantum system is a single particle in a potential, expressions of the form

$$\langle X\psi, \psi\rangle$$

as the expected value of the position of the particle when it is in the state ψ.

As a hint towards the definition of a 'non-commutative probability space', we recall from our exposition of measure theoretic probability that the primary tool to investigate 'classical' probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ were its random variables, i.e. the measurable functions from Ω to $\mathbb{C}$. The collection of such random variables naturally has a $*$-algebra structure inherited from the one on $\mathbb{C}$. Moreover, two of the most famous results of operator algebra theory assert that commutative unital C*-algebras and Abelian von Neumann algebras are always of the form $C(\Omega)$ and $L^\infty(\Omega, \mu)$ respectively, for Ω a compact Hausdorff space and μ a positive measure. This motivates the following definition

Definition 2.1.

(i) Let $\mathcal{A}$ be a unital $*$-algebra over $\mathbb{C}$, and $\varphi : \mathcal{A} \to \mathbb{C}$ a linear, unital, and positive functional, then we call the pair $(\mathcal{A}, \varphi)$ a *$*$-probability space*.

(ii) If $\mathcal{A}$ has the additional structure of a unital C*-algebra and φ is a faithful state, we say that $(\mathcal{A}, \varphi)$ is a *C*-probability space*.

(iii) If $\mathcal{A}$ is a von Neumann algebra and φ is a faithful normal (σ-weakly continuous) state, $(\mathcal{A}, \varphi)$ is designated as a *W*-probability space*.

(iv) If φ is a trace in any of the cases above, we add the adjective *tracial* in front of the name, and for the W* case we then adopt the notation (M, τ) instead of $(\mathcal{A}, \varphi)$.

Remark 2.2.

(i) If it does not matter which of the three types a pair $(\mathcal{A}, \varphi)$ is, we shall speak of a *non-commutative probability space* (shortened to NCPS in the sequel).

(ii) In the case that $\mathcal{A}$ is merely a $*$-algebra, a functional is positive if $\varphi(a^*a) \geq 0$ for all $a \in \mathcal{A}$, we will also be sloppy in the future and refer to φ as a state.

(iii) Morally, one should think of the state φ as a non-commutative version of the functional given by integration against the probability measure $\mathbb{P}$ (hence the requirement that φ is unital).

(iv) To continue the analogy, we shall call elements of a non-commutative probability space simply *random variables*.

(v) Definition (i) above is necessary as (ii) and (iii) only allow bounded random variables, and many important examples do not satisfy this (think of the standard Gaußian on $\mathbb{R}$).

(vi) In all three cases, φ satisfies $\varphi(a^*) = \overline{\varphi(a)}$ and $|\varphi(b^*a)|^2 \leq \varphi(a^*a)\varphi(b^*b)$ (Cauchy-Schwartz) for all $a, b \in \mathcal{A}$.

(vii) Recall that a positive functional on a C*-algebra is automatically continuous, so there is no continuity assumption required in the definition.

Definition 2.3. For the three types of non-commutative probability space we have defined, a *morphism* is a morphism of the algebra which additionally intertwines the states. If the target algebra is the algebra of bounded operators on a Hilbert space, and the target state is a vector state, we call the morphism a *representation*.

Remark 2.4. In the W^* case, by an isomorphism of von Neumann algebras we mean a σ-weakly continuous $*$-isomorphism. It turns out that this is a pleonasm, as any bijective $*$-homomorphism between van Neumann algebras is automatically σ-weakly continuous. We sketch the proof of this statement. Let π be a $*$-isomorphism between two von Neumann algebras M and N. Denote by M_* and N_* their unique preduals (uniqueness is a theorem of Sakai, see corollary 1.13.3 in [75]). Notice that π gives us a map $\hat{\cdot} : M \to (N_*)^*$ by setting $\hat{x}(\omega) = (\omega \circ \pi)(x)$. This is an isomorphism, as it is inverted by the map $\check{\cdot} : N \to (M_*)^*$ defined by $\check{y}(\eta) = (\eta \circ \pi^{-1})(y)$. Hence $M \cong (N_*)^*$ and $N \cong (M_*)^*$, but by uniqueness it must then be true that $M_* \cong N_*$. As convergence of sequences in the σ-weak topology is controlled by the elements of the predual, it follows that the map π is σ-weakly continuous. Finally, note that the above argument fails if the map π is merely a $*$-homomorphism.

Let now $(\mathcal{A}, \varphi)$ be a NCPS and a an element of $\mathcal{A}$. We would like to talk about the 'moments' of our generalised random variables, but we should pause for a moment to think about the implications of non-commutativity. It is clear that the natural notion of 'moments' in this context should be the numbers obtained by feeding certain expressions containing a to the state φ. However, it will not suffice to consider only powers of a. Instead, we should keep track of all the words that can be made with the alphabet $\{a, a^*\}$, as also the order of the letters is important. Indeed, the numbers $\varphi(a^*a)$ and $\varphi(aa^*)$ need not be the same. In short, we should study the state φ restricted to the unital $*$-algebra generated by a,

$$\langle a \rangle := \operatorname{span} \left\{ a^{\varepsilon(1)} \cdots a^{\varepsilon(k)} \middle| \, k \in \mathbb{Z}_{\geq 0},\ \varepsilon(j) \in \{1, *\} \right\}.$$

Definition 2.5. Let $(\mathcal{A}, \varphi)$ be a NCPS and a an element of $\mathcal{A}$. An expression of the form $\varphi(a^{\varepsilon(1)} \cdots a^{\varepsilon(k)})$ with $k \in \mathbb{Z}_{\geq 0}$ and $\varepsilon(j) \in \{1, *\}$ is called a $*$-*moment* of a. The $*$-*distribution* of a is the functional μ on the unital $*$-algebra of non-commutative polynomials in the two formal indeterminates X and X^*, denoted $\mathbb{C}\langle X, X^* \rangle$, defined by setting $\mu(X^{\varepsilon(1)} \cdots X^{\varepsilon(k)}) = \varphi(a^{\varepsilon(1)} \cdots a^{\varepsilon(k)})$ for all $*$-moments.

There are two special cases that warrant some additional discussion. First, if a is self-adjoint, then our entire discussion before the previous definition is rendered moot. It is easy to see that $\langle a \rangle$ is commutative and simply the span of the powers of a, and the $*$-moments are reduced to the set of numbers of the form $\varphi(a^n)$ for $n \in \mathbb{Z}_{\geq 0}$. In this case, we will simply call these numbers the *moments* of a (notice that they are all real). We can also talk unambiguously about the 'mean' of a as $\varphi(a)$ and its 'variance' as $\varphi(a^2) - \varphi(a)^2$ (note that we do not require that φ is a character).

The second, slightly more general case occurs if a is a normal element of $\mathcal{A}$. The generated $*$-algebra $\langle a \rangle$ is then still commutative, but not quite as small. It is easy to see that it is spanned by the elements $a^k(a^*)^\ell$, for $k, \ell \in \mathbb{Z}_{\geq 0}$. If we are in the situation where our NCPS $(\mathcal{A}, \varphi)$ admits

a faithful representation, we can construct a C*-probability space and a morphism from $(\mathcal{A}, \varphi)$ to it (since closed $*$-subalgebras of the bounded operators on a Hilbert space are C*-algebras). This then gives us access to spectral theory, which provides us with a notion of $*$-distribution that is more reminiscent of the classical case. If there exists a compactly supported probability measure μ on $\mathbb{C}$ such that the $*$-moments $\varphi(a^k(a^*)^\ell)$ can be obtained by integrating the polynomials $z^k\overline{z}^\ell$ with respect to it, we call μ the *analytic $*$-distribution* of a. Notice that by the Stone-Weierstraß theorem and the fact that μ is a functional on $C(\text{supp }\mu)$, μ is uniquely determined by fixing the values of the integrals of the polynomials $z^k\overline{z}^\ell$. It is an easy exercise in spectral theory to show that such a measure exists and that its support is precisely the spectrum of a. In particular, if a is self-adjoint its analytic $*$-distribution is supported in the real number line, and we only need to consider the polynomials t^n for $n \in \mathbb{Z}_{\geq 0}$. Finally, notice that giving the analytic $*$-distribution of a normal element also fixes its $*$-distribution.

In classical probability, one often refers to a random variable only by its law, we will introduce some examples of this in the non-commutative context. Notice that if $a \in \mathcal{A}$ is unitary, the generated $*$-algebra is spanned by the integer powers a^k, $k \in \mathbb{Z}$, of a.

Definition 2.6. Let $(\mathcal{A}, \varphi)$ be a NCPS. An element u of $\mathcal{A}$ is called a *Haar unitary* iff it is a unitary and all of its non-trivial moments vanish. For $p \in \mathbb{Z}_{\geq 1}$, $v \in \mathcal{A}$ is a *p-Haar unitary* iff it is a unitary, $v^p = I$, and $\varphi(v^k) = 0$ for all integers k not divisible by p.

Example 2.7. Let u be a Haar unitary in a C*-probability space. Then its spectrum is contained in the unit circle $\mathbb{T}$ in the complex plane. Consider the normalised Lebesgue measure on $\mathbb{T}$, it is well known that this is its Haar measure. One straightforwardly computes

$$\int_{\mathbb{T}} z^k\overline{z}^\ell \frac{\mathrm{d}z}{2\pi} = \delta_{k-\ell,0}.$$

Therefore, the analytic $*$-distribution of u is the Haar measure of $\mathbb{T}$, which explains the terminology. Suppose now that v is a p-Haar unitary, then by the spectral mapping theorem its spectrum is contained in the set of p-th roots of unity ρ_p. The natural measure on this discrete space is the following normalised superposition of Dirac measures $(\delta(1) + \cdots + \delta(e^{2\pi i(p-1)/p}))/p$. Indeed, one sees that

$$\frac{1}{p}\int_{\rho_p} z^k\overline{z}^\ell \,\mathrm{d}\big(\delta(1) + \cdots + \delta(e^{2\pi i(p-1)/p})\big) = \begin{cases} 1, & p|(k-\ell) \\ \dfrac{1}{p}\dfrac{1-e^{2\pi i(k-\ell)}}{1-e^{2\pi i(k-\ell)/p}} = 0, & \text{else} \end{cases}.$$

So far, we have only considered distributions associated to a single random variable. It is now time to generalise the notion of a joint law from classical probability.

Definition 2.8. Let $(\mathcal{A}, \varphi)$ be a NCPS. Denote by $\mathbb{C}\langle X_1, X_1^*, \ldots, X_k, X_k^*\rangle$ the unital $*$-algebra of non-commutative polynomials in the $2k$ indeterminates $X_1, X_1^*, \ldots, X_k, X_k^*$. For $a_1, \ldots, a_k \in \mathcal{A}$, their *joint $*$-moments* are the numbers $\varphi(a_{j_1}^{\varepsilon(1)} \cdots a_{j_\ell}^{\varepsilon(\ell)})$, where $\ell \in \mathbb{Z}_{\geq 0}$ and $\varepsilon(j) \in \{1, *\}$. The *joint $*$-distribution* of the random variables $a_1, \ldots, a_k$ is the functional $\mu : \mathbb{C}\langle X_1, X_1^*, \ldots, X_k, X_k^*\rangle \to \mathbb{C}$, defined by $\mu(X_{j_1}^{\varepsilon(1)} \cdots X_{j_\ell}^{\varepsilon(\ell)}) = \varphi(a_{j_1}^{\varepsilon(1)} \cdots a_{j_\ell}^{\varepsilon(\ell)})$ for all joint $*$-moments.

Joint $*$-distributions are of incredible importance not just to non-commutative probability, but also to the theory of operator algebras in general. The reason for this is that under certain assumptions, they provide a complete isomorphism invariant for the algebra underlying a NCPS. These results are particularly valuable to the theory of von Neumann algebras, where this is one of the simplest isomorphism theorems available and K theory is less useful than in the C*-algebraic setting. We expand slightly upon this last statement. At least for factors, it follows from Examples 3.3.2, 3.3.3,

and Exercise 3.12 in [72] that the K_0 group is trivial for types I_∞, II_∞, and III, while it is $\mathbb{Z}$ for type I_f and $\mathbb{R}$ for type II_1. The K_1 group for any von Neumann algebra is trivial by Example 8.1.2 (a) in [10]. The isomorphism result considered here comes in three flavours, one for each type of NCPS we have introduced. We conclude this section with the statements and proofs of these theorems.

Theorem 2.9 ($*$-algebras)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be $*$-algebras, and let φ and ψ be faithful functionals such that $(\mathcal{A},\varphi)$ and $(\mathcal{B},\psi)$ are $*$-probability spaces. Assume that $\mathcal{A}=\langle a_1,\ldots,a_k\rangle$ and $\mathcal{B}=\langle b_1,\ldots,b_k\rangle$. If $a_1,\ldots,a_k$ and $b_1,\ldots,b_k$ have identical joint $*$-distributions (relative to φ and ψ respectively), then the map Φ defined by $a_j\mapsto b_j$ is a uniquely determined $*$-isomorphism. Moreover, Φ is also an isomorphism of the $*$-probability spaces $(\mathcal{A},\varphi)$ and $(\mathcal{B},\psi)$.*

Proof. Write a and b for the generators of $\mathcal{A}$ and $\mathcal{B}$ respectively. Since these are generators, we know that we can write any element as some non-commutative polynomial in them. In other words,

$$\mathcal{A}=\{P(a)\,|P\in\mathbb{C}\langle X_1,X_1^*,\ldots,X_k,X_k^*\rangle\},$$

and similarly for $\mathcal{B}$. This will allow us to compare elements of $\mathcal{A}$ and $\mathcal{B}$, as the joint $*$-distributions of their generators are the same functional μ on $\mathbb{C}\langle X_1,X_1^*,\ldots,X_k,X_k^*\rangle$. Clearly, the map Φ defined in the statement of the theorem extends to a $*$-homomorphism. Therefore, we have that $\Phi(P(a))=P(\Phi(a))=P(b)$. Now, we will prove that all polynomial relations between a and between b must be the same. Let P and Q be non-commutative polynomials such that $P(a)=Q(a)$. Then as φ is faithful, $\varphi((P(a)-Q(a))^*(P(a)-Q(a)))=0$. By definition of μ, this can be written as $\mu((P-Q)^*(P-Q))=0$. Since this is the same joint $*$-distribution as that of b, and ψ is also faithful, this is equivalent to $P(b)=Q(b)$. In particular, Φ is bijective, so it is a $*$-isomorphism. The fact that it intertwines φ and ψ follows immediately from the reasoning above. Uniqueness is also clear. □

Theorem 2.10 (C^*-algebras)**.** *Let A and B be C^*-algebras, and let φ and ψ be faithful states such that (A,φ) and (B,ψ) are C^*-probability spaces. Assume that $A=C^*\langle a_1,\ldots,a_k\rangle$ and $B=C^*\langle b_1,\ldots,b_k\rangle$. If $a_1,\ldots,a_k$ and $b_1,\ldots,b_k$ have identical joint $*$-distributions (relative to φ and ψ respectively), then the map Φ defined by $a_j\mapsto b_j$ extends to a uniquely determined isometric $*$-isomorphism. Moreover, Φ is also an isomorphism of the C^*-probability spaces (A,φ) and (B,ψ).*

Proof. Write a and b for the generators of A and B respectively. We can again consider the $*$-algebra

$$\mathcal{A}=\{P(a)\,|P\in\mathbb{C}\langle X_1,X_1^*,\ldots,X_k,X_k^*\rangle\}\subset A,$$

and similarly for $\mathcal{B}\subset B$. As a and b generate A and B as C^*-algebras, $\mathcal{A}$ is norm-dense in A and $\mathcal{B}$ is norm-dense in B. By the preceding flavour, the map Φ as defined in the statement of the theorem is a $*$-isomorphism of $\mathcal{A}$ and $\mathcal{B}$. The idea of the rest of the proof is to show that this map is in fact isometric, and then a standard extension argument suffices to finish the job. To achieve this, we need to connect the norm of an element in A to something involving φ, which can then be connected to ψ via μ (the shared joint $*$-distribution of the generators), and so to the norm on B. As φ is a state, we know a priori that $\varphi(x)\leq\|x\|$ for any $x\in A$. We know even more, by positivity of φ and the C^* identity, it holds that $\varphi(x^*x)^{1/2}\leq\|x\|$. By the spectral mapping theorem, the elements $(x^*x)^n$ are all positive. Iterating, we obtain $\varphi((x^*x)^n)^{1/2n}\leq\|x\|$, so that $\varphi((x^*x)^n))^{1/2n}$ is a bounded sequence of non-negative numbers. For convenience, we look at the square of the sequence, and we want to show that its limit is then $\|x^*x\|$. Since x^*x is positive, we know that its spectrum is contained in the closed interval $[0,\|x^*x\|]$. Moreover, as φ is faithful, the analytic $*$-distribution ν of x^*x has as support the entire spectrum of x^*x. In particular, the ν-measure of any interval of the form $[\|x^*x\|-2\varepsilon,\|x^*x\|]$ is positive, where $\varepsilon>0$ is arbitrary. Now, using the definition of the analytic $*$-distribution, we can compute

$$\varphi((x^*x)^n)=\int_{\sigma(x^*x)}t^n\,\mathrm{d}\nu(t)\geq\int_{\sigma(x^*x)\cap[\|x^*x\|-2\varepsilon,\|x^*x\|]}t^n\,\mathrm{d}\nu(t)\geq(\|x^*x\|-2\varepsilon)^nC_\varepsilon,$$

where $C_\varepsilon > 0$ is the ν-measure of $[\|x^*x\| - 2\varepsilon, \|x^*x\|]$. It is easy to see that C_ε^n tends to 1 as n grows. From these estimates we conclude that $\varphi((x^*x)^n)^{1/n}$ for n large enough is greater than $\|x^*x\| - \varepsilon$, and as ε was arbitrary, we conclude that

$$\lim_{n\to\infty} \varphi((x^*x)^n)^{1/n} = \|x^*x\|.$$

Using this, we have that

$$\begin{aligned}
\|P(a)\|_A^2 &= \lim_{n\to\infty} \varphi((P(a)^*P(a))^n)^{1/n}, \\
&= \lim_{n\to\infty} \mu((P^*P)^n)^{1/n}, \\
&= \lim_{n\to\infty} \psi((P(b)^*P(b))^n)^{1/n}, \\
&= \|P(b)\|_B^2, \\
&= \|\Phi(P(a))\|_B^2.
\end{aligned}$$

The rest of the proof is now an easy extension argument, which we leave to the diligent reader. □

Theorem 2.11 (von Neumann algebras)**.** *Let M and N be von Neumann algebras, and let φ and ψ be faithful normal states such that (M, φ) and (N, ψ) are W^*-probability spaces. Assume that $M = \mathrm{vN}\langle a_1, \ldots, a_k\rangle$ and $N = \mathrm{vN}\langle b_1, \ldots, b_k\rangle$. If $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ have identical joint $*$-distributions (relative to φ and ψ respectively), then the map Φ defined by $a_j \mapsto b_j$ extends to a $*$-isomorphism. Moreover, Φ is also an isomorphism of the W^*-probability spaces (M, φ) and (N, ψ).*

Proof. Write a and b for the generators of M and N. Denote by $L^2(M, \varphi)$ and $L^2(N, \psi)$ the GNS representations of M and N with respect to φ and ψ. As the states are faithful and normal, we may identify M and N with their GNS representations. Let Ω_M and Ω_N be cyclic vectors in $L^2(M, \varphi)$ and $L^2(N, \psi)$ (for instance the images of the units). Again the $*$-algebra

$$\mathcal{M} = \{P(a) \,|P \in \mathbb{C}\langle X_1, X_1^*, \ldots, X_k, X_k^*\rangle\} \subset B(L^2(M, \varphi)),$$

is dense, and similarly for $\mathcal{N} \subset B(L^2(N, \psi))$. Now consider $U : L^2(M, \varphi) \to L^2(N, \psi)$, where $P(a)\Omega_M \mapsto U(P(a)\Omega_M) = P(b)\Omega_N$. We show that U is unitary.

$$\begin{aligned}
\langle U(P(a)\Omega_M), U(Q(a)\Omega_M)\rangle_N &= \langle P(b)\Omega_N, Q(b)\Omega_N\rangle_N, \\
&= \psi\left(Q(b)^*P(b)\right), \\
&= \mu(Q^*P), \\
&= \varphi(Q^*(a)P(a)), \\
&= \langle P(a)\Omega_M, Q(a)\Omega_M\rangle_M.
\end{aligned}$$

We can then construct $\Phi : M \to N$ as a spatial isomorphism that sends $x \mapsto UxU^*$. This is clearly injective and a $*$-homomorphism. Moreover,

$$UP(a)U^*(Q(b)\Omega_N) = UP(a)Q(a)\Omega_M = P(b)(Q(b)\Omega_N),$$

so that Φ is surjective as its image is dense. Showing that Φ intertwines φ and ψ proceeds as before. □

2.2 Freeness

The problem of computing the joint $*$-distribution of some random variables appears as a rather formidable task. In classical probability, one often focuses on random variables that are independent. This comes with a collection of rules to calculate joint moments, the simplest of which we have already seen in our exposition of classical probability, remark 1.9. However, it is not immediately obvious how to lift this notion to the non-commutative setting. In fact, there are several different ways to implement this idea, and we will present only one.

Definition 2.12. Throughout, $(\mathcal{A}, \varphi)$ is a NCPS and I is a fixed index set.

(i) Let $\{\mathcal{A}_i\}_{i\in I}$ be unital $*$-subalgebras of $\mathcal{A}$, then a product $a_1 \cdots a_k$ ($k \in \mathbb{Z}_{\geq 1}$) with $a_j \in \mathcal{A}_{i_j}$ is called *alternating* iff $i_1 \neq i_2$, $i_2 \neq i_3$,..., $i_{k-1} \neq i_k$.

(ii) An element $a \in \mathcal{A}$ is called *centred* iff $\varphi(a) = 0$. A product $a_1 \dots a_k$ is called *centred* iff all the a_j are centred.

(iii) We say that the unital $*$-subalgebras $\{\mathcal{A}_i\}_{i\in I}$ are *freely independent* (or often just *free*) iff for all centred alternating products $a_1 \cdots a_k$ we have that $\varphi(a_1 \cdots a_k) = 0$.

(iv) If $\{a_i\}_{i\in I}$ are elements of $\mathcal{A}$, we say that they are *freely independent* (or just *free*) iff the unital $*$-subalgebras they generate are free.

Remark 2.13. Notice first that in any NCPS, the scalar multiples of the identity are free with respect to any other subalgebra. Suppose now that a and b are some self-adjoint free elements of a NCPS, and that they commute. From the free independence condition one can derive that the product of the variance of a and b vanishes. This implies that at least one of them is a constant, showing that free independence is a truly non-commutative notion.

One might wonder why this is a good notion of independence. We will spent the remainder of this section arguing that this is the case, by showing that this definition enjoys several good properties. The first one is that, as we insinuated earlier, free independence comes with some rules for calculating joint $*$-moments. We explain that the notion of freeness is analogous (in some sense to be made precise) to the free product of groups. Then we will show that the sentence 'we assume that the random variables $a_1, \dots, a_k$ are freely independent' is both meaningful and without loss of generality. Finally, we show that freeness at the level of $*$-algebras propagates to freeness at the level of C*-and van Neumann algebras.

Proposition 2.14. *Suppose that $\{\mathcal{A}_i\}_{i\in I}$ are free in some NCPS $(\mathcal{A}, \varphi)$. Let $a_1 \cdots a_k$ and $b_1, \dots, b_\ell$ be centred alternating products for $\mathcal{A}_{i_1}, \cdots \mathcal{A}_{i_k}$ and $\mathcal{A}_{j_1}, \cdots, \mathcal{A}_{j_\ell}$ respectively. Then*

$$\varphi(a_1 \cdots a_k b_\ell \cdots a_1) = \begin{cases} \varphi(a_1 b_1) \cdots \varphi(a_k b_k) & \text{if } k = \ell \text{ and } (i_1, \dots, i_k) = (j_1, \dots, j_k), \\ 0 & \text{else} \end{cases}.$$

Proof. If $i_k \neq j_\ell$, then the result is 0 by free independence. So we assume that $i_k = j_\ell$, and for an element $c \in \mathcal{A}$ we introduce its *centering* $\dot{c} = c - \varphi(c)$. Then

$$\begin{aligned} \varphi(a_1 \cdots a_k b_\ell \cdots b_1) &= \varphi(a_1 \cdots_{k-1} (\dot{a_k b_\ell} + \varphi(a_k b_\ell)) b_{\ell-1} \cdots b_1), \\ &= 0 + \varphi(a_k b_\ell)\varphi(a_1 \cdots a_{k-1} b_{\ell-1} \cdots b_1). \end{aligned}$$

Iterating this procedure yields the desired result. □

A straightforward application of the results and techniques of this proposition is the following.

Proposition 2.15. *Let $(\mathcal{A}, \varphi)$ be a NCPS and $\{\mathcal{A}_i\}_{i\in I}$ be free subalgebras for some fixed index set I. Denote by $\mathcal{B}$ the subalgebra of $\mathcal{A}$ generated by the $\mathcal{A}_i$'s. Then $\varphi|_{\mathcal{B}}$ is completely determined by $\{\varphi|_{\mathcal{A}_i}\}_{i\in I}$ and the free independence condition. Moreover, if all the restrictions of φ to $\mathcal{A}_i$'s are tracial, so is $\varphi|_{\mathcal{B}}$.*

We now turn to freeness in the context of groups.

Definition 2.16.

(i) Let G be a group, and $\{G_i\}_{i\in I}$ be subgroups for some fixed index set I. We say that the G_i's are *free* iff when an alternating product $g_1\cdots g_k$ (for all $k\in\mathbb{Z}_{\geq 1}$) is such that $g_j\neq e$, we have $g_1\cdots g_k\neq e$.

(ii) Let H be a group, we define its *group algebra* $\mathbb{C}H$ as the complex vector space given by all formal finite linear combinations of group elements, with the obvious multiplication.

(iii) Let $\{G_i\}_{i\in I}$ be groups with identity elements e_i for some fixed index set I, then their *free product* $G:=*_{i\in I}G_i$ is the unique group generated by all elements in the G_i's subject only to the relations within them and the demand that $e_i=e$.

By checking the definitions, one can prove the following.

Proposition 2.17. *$\{G_i\}_{i\in I}\in G$ are free iff $\{\mathbb{C}G_i\}_{i\in I}$ are free in the NCPS $(\mathbb{C}G,\tau_G)$, where*

$$\tau_G(\sum_{g\in G}\alpha_g g)=\alpha_e.$$

It turns out that one can also define a (reduced) free product of NCPS.

Theorem 2.18 ($*$-probability spaces)**.** *Let $\{(\mathcal{A}_i,\varphi_i)\}_{i\in I}$ be $*$-probability spaces, then there exists a unique $*$-probability space $(\mathcal{A},\varphi)$ such that the $\mathcal{A}_i$ are free subalgebras of $\mathcal{A}$, and the restriction of φ to a $\mathcal{A}_i$ agrees with φ_i. This* free product *$(\mathcal{A},\varphi)$ also has the property that any family of morphisms $\{\Phi_i\}_{i\in I}$, from $(\mathcal{A}_i,\varphi_i)$ into a fixed NCPS $(\mathcal{B},\psi)$, such that the images are free, admits a unique extension to all of $\mathcal{A}$. Moreover, if all the φ_i are traces or faithful, then φ is a trace or faithful respectively.*

Theorem 2.19 (C*-probability spaces)**.** *Let $\{(A_i,\varphi_i)\}_{i\in I}$ be C*-probability spaces, then there exists a unique C*-probability space (A,φ) such that the A_i are free subalgebras of A, and the restriction of φ to a A_i agrees with φ_i. This* free product *(A,φ) is such that A is generated as a C*-algebra by the union of the A_i.*

Remark 2.20.

(i) One can show that with these definitions, $(\mathbb{C}*_{i\in I}G_i,\tau_{*_{i\in I}G_i})$ and $*_{i\in I}(\mathbb{C}G_i,\tau_{G_i})$ are isomorphic NCPS.

(ii) There is also a notion of free product for W*-probability spaces, but this is more involved (see [19]).

Proposition 2.21.

(i) Let (A,φ) be a C-probability space, and $B_1,\ldots,B_k$ free unital $*$-subalgebras. Put $A_i=\overline{B_i}^{\|\cdot\|}$, then the A_i are free.*

(ii) Let (M,ψ) be a W-probability space, and $N_1,\ldots,N_k$ free unital $*$-subalgebras. Put $M_i=N_i''$, then the M_i are free.*

Proof. For (i), suppose that $a_1\cdots a_\ell$ is a centred alternating product. As B_i is dense in A_i, there is a sequence $(b_{i,n})$ in B_i approximating a_i in norm. By continuity of φ, we may assume that the $b_{i,n}$ are centred. Then, using that φ and multiplication are norm continuous,

$$\varphi(a_1\cdots a_\ell)=\lim_{n\to\infty}\varphi(b_{1,n}\cdots b_{\ell,n})=0.$$

For (ii), one repeats the argument above, except that now that $b_{i,n}$ approximate a_i in the strong operator topology. By Kaplanski's theorem, we may assume that the sequences are each bounded in norm by the norm of the element they approximate. As multiplication is SOT-continuous when restricted to bounded sets, the result follows. □

2.3 Semicircular Elements

In this section we will work out an example that will be of importance later. Let $(\mathcal{A}, \varphi)$ be a NCPS such that the algebra $\mathcal{A}$ is generated by some element a^*. The reason why we write the generator as a^* will become clear when we discuss q-Gaußians later. We assume that a^* is a *non-unitary isometry*, i.e. $aa^* = I \neq a^*a$. Since any occurrence of aa^* can be removed from a word in a and a^*, we have that

$$\mathcal{A} = \operatorname{span}\{(a^*)^m a^n \,|m, n \in \mathbb{Z}_{\geq 0}\}.$$

In particular, we assume that the $(a^*)^m(a^n)$ are linearly independent elements of $\mathcal{A}$, and that

$$\varphi((a^*)^m a^n) = \delta_{m,0}\delta_{n,0}.$$

The element a^* is not normal, so we consider instead its self-adjoint part $a + a^*$, and try to compute its $*$-distribution. For this we will need to introduce some notions from combinatorics.

Definition 2.22.

(i) An *n-Dyck path* is a path on the lattice $\mathbb{Z}^2$, starting in the origin an ending at $(2n, 0)$, never going below the x-axis (touching it is allowed) and consisting of steps of the form $(1, 1)$ and $(1, -1)$.

(ii) An n-Dyck path of *exceedance* k, with $n \geq k \geq 0$, has the same properties as a Dyck path, except that there are precisely k steps that start below the x-axis.

(iii) The *n-th Catalan number* C_n, for $n \in \mathbb{Z}_{\geq 0}$, is the number defined by the formula $C_n = \binom{2n}{n}/(n+1)$.

Remark 2.23. Note that an n-Dyck path can also be described as a list of $2n$ numbers $\{\lambda_1, \ldots \lambda_{2n}\} \in \{-1, 1\}^{2n}$ with the properties that

$$\begin{cases} \lambda_1 + \cdots + \lambda_j \geq 0, & \forall\ 1 \leq j \leq 2n \\ \lambda_1 + \cdots + \lambda_{2n} = 0 \end{cases}.$$

The relation between the different parts of this definition is given by the following celebrated result in combinatorics.

Theorem 2.24 (Chung-Feller)**.** *For every $n \in \mathbb{Z}_{\geq 0}$ and $k = 0, 1, \ldots, n-1$, the number of n-Dyck paths with exceedance k is equal to the n-th Catalan number C_n. In particular, the number of n-Dyck paths is equal to C_n.*

Proof. We follow the rather elegant proof presented in [20]. The case $n = 1$ is trivial, so assume that $n \geq 2$. It is clear that there are $\binom{2n}{n}$ paths with steps $(1, 1)$ and $(1, -1)$ starting at the origin and ending at $(2n, 0)$. We will shows that there exists a bijection between the n-Dyck paths with exceedance k and those with exceedance $k + 1$, for every $k = 0, \ldots, n-1$. This divides the total number of paths into $n + 1$ families of equal size, and the result follows.

Let D be an n-Dyck path of exceedance k, then we decompose it in the following way. The path starts in the origin, and it may first contain a part that is all below the x-axis, say with exceedance $k \geq k' \geq 0$, call this part of the path A (note that it may be empty). Denote now by u the first step that ends above the x-axis (this exists because $k < n$). This is followed by a part of the path that is completely above the x-axis so it has exceedance 0. Call this part B (note again that it may be empty). Denote by d the first step down that ends on the x-axis after u (this exists because the

path must end on the x-axis). The remainder of the path, which has exceedance $k-k'$, will be denoted by C (note again that this may be empty, but note also that A, B, and C can not be empty simultaneously because $n \geq 2$). We have that $D = AuBdC$, and we now define the new path by the rearrangement $D' = BdAuC$. Notice that B has exceedance 0, dAu has exceedance $k'+1$, and C has exceedance $k-k'$, so that D' is an n-Dyck path with exceedance $k+1$.

Going the other way, let now E be an n-Dyck path with exceedance $k+1$ and decompose it as follows. First E has a (possibly empty) part A with exceedance 0 (i.e. all above the x-axis). Then a first step ending below the x-axis called d, a part B all below the x-axis, and finally the first step ending again on the x-axis u. The remainder of the path is denoted by C. Say that dBu has exceedance $k'+1$, for some $k \geq k' \geq 0$, and notice that C has exceedance $k-k'$. Now rearrange $E = AdBuC$ into a new path given by $E' = BuAdC$, which has exceedance k. Clearly applying both parts of the above procedure to the same path returns the original, and we are done. □

Proposition 2.25. *Consider $a^{\varepsilon(1)} \cdots a^{\varepsilon(\ell)} \in \mathcal{A}$, where $\ell \in \mathbb{Z}_{\geq 0}$ and $\varepsilon(j) \in \{1, *\}$. Define λ_j to be 1 if $\varepsilon(j) = *$, and -1 if $\varepsilon(j) = 1$. Then*

$$\varphi(a^{\varepsilon(1)} \cdots a^{\varepsilon(\ell)}) = \begin{cases} 1 & \text{if } \{\lambda_\ell, \ldots, \lambda_1\} \text{ is an } \ell\text{-Dyck path,} \\ 0 & \text{else} \end{cases}.$$

Proof. Consider the Hilbert space of square summable sequences $\ell^2(\mathbb{Z}_{\geq 0})$, which has orthonormal basis $\{\delta_n\}_{n \in \mathbb{Z}_{\geq 0}}$. We can define the right shift operator R by the equations $R\delta_n = \delta_{n+1}$. Its adjoint is the left shift operator L, which acts as $L\delta_n = \delta_{n-1}$, with the convention that $\delta_{-1} = 0$. Clearly, R is a non-unitary isometry, and it generates a $*$-algebra $\mathcal{R}$. On $B(\ell^2(\mathbb{Z}_{\geq 0}))$ one has the vector state $\Omega(x) = \langle x\delta_0, \delta_0 \rangle$, and a simple computation shows that $\Omega(R^m L^n) = \delta_{m,0}\delta_{n,0}$. Define a unital $*$-homomorphism $\Phi : \mathcal{A} \to \mathcal{R}$ by $\Phi((a^*)^n a^m) = R^n L^m$. As the elements $R^n L^m$ are linearly independent, this is in fact a $*$-isomorphism. Moreover, as the $*$-moments of a and R agree, Φ is a morphism of $*$-probability spaces. We can now compute

$$\varphi(a^{\varepsilon(1)} \cdots a^{\varepsilon(\ell)}) = \Omega(R^{\varepsilon(1)} \cdots R^{\varepsilon(\ell)}) = \langle R^{\varepsilon(1)} \cdots R^{\varepsilon(\ell)} \delta_0, \delta_0 \rangle.$$

This vanishes unless $R^{\varepsilon(1)} \cdots R^{\varepsilon(\ell)} \delta_0 = \delta_0$. Keeping in mind the action of L and R on basis elements, one arrives at the advertised relation. □

We are now finally in a position to compute the $*$-distribution of $a + a^*$.

Proposition 2.26.

$$\varphi((a+a^*)^\ell) = \begin{cases} C_{\ell/2} \text{ if } \ell \text{ is even,} \\ 0 \text{ if } \ell \text{ is odd} \end{cases}.$$

Proof. A quick computation shows that

$$\varphi((a+a^*)^\ell) = \sum_{\varepsilon(1),\ldots,\varepsilon(\ell) \in \{1,*\}} \varphi(a^{\varepsilon(1)} \cdots a^{\varepsilon(\ell)}) = \sum_{\{\lambda_\ell, \ldots \lambda_1\} \text{ is an } \ell\text{-Dyck path}} 1 = \begin{cases} C_{\ell/2} \text{ if } \ell \text{ is even,} \\ 0 \text{ if } \ell \text{ is odd} \end{cases},$$

where we used the previous proposition in the second equality. □

From the proof of Proposition 2.25, we can deduce through spectral theory that $a + a^*$ has an analytic $*$-distribution. Even better, we already know what this measure is. From example 1.13 we know that the moments of the Wigner semi-circle density with $\rho = 2$ are precisely the moments of $a + a^*$.

Definition 2.27. Let $(\mathcal{A}, \varphi)$ be a NCPS. If x a self-adjoint element such that it has an analytic $*$-distribution, and this distribution is the Wigner semi-circle density, we call x a *semicircular* element. If its distribution is the Wigner semi-circle law with $\rho = 2$, we say that x is *standard semicircular*, as it has mean 0 and variance 1.

2.4 Free Cumulants

In this section we introduce an alternative bookkeeping device to moments, and discuss how it interacts with freeness, in particular with free families of semicircular elements. We being with some notation

Definition 2.28. Let $n \geq 1$ be a natural number, then we write $[n] := \{1, \ldots, n\}$. Define the set of *partitions* of $[n]$ as

$$\mathcal{P} := \{\pi = (V_1, \ldots, V_k) | \, V_j \subset [n] \text{ and } V_j \neq \varnothing \; \forall \; j, \; V_i \cap V_j = \varnothing \; \forall \; i \neq j\},$$

where the V_j are called *blocks*. We say that two blocks V and W of a partition π *cross* if there exist $i, j \in V$ and $k, \ell \in W$ such that $i < k < j < \ell$. If no such blocks exist, we say that π is *non-crossing*, and denote the set of all non-crossing partitions of $[n]$ by $\mathrm{NC}(n)$. In particular, if all blocks of π contain exactly two elements, we call π a *pairing* of $[n]$, and the set of non-crossing pairings of $[n]$ is denoted by $\mathrm{NC}_2(n)$

Remark 2.29. Clearly, $\mathrm{NC}_2(n) = \varnothing$ iff n is odd. We have also already computed the cardinality of $\mathrm{NC}_2(2n)$, namely it is equal to the Catalan number C_n. To see this, one can convince oneself that there exists a bijection between non-crossing pairings of $[2n]$ and n-Dyck paths, with one possible rule being the following. Take a non-crossing pairing, and connect the members of each pair by a line, for example

$\leftrightarrow \; (\{1,4\},\{2,3\},\{5,6\})$,

$\leftrightarrow \; (\{1,6\},\{2,5\},\{3,4\})$.

Since our pairings are non-crossing, these lines can be drawn without crossings. Now imagine traversing $[2n]$ from left to right, and build an n-Dyck path by taking a step $(1,1)$ whenever a line starts, and a step $(-1,1)$ whenever a line ends. Since one cannot see the end of a line before its start, and upon reaching $2n$ all lines have ended, this is indeed an n-Dyck path.

Recall from the text below the definition of 'classical' moments, Definition 1.7, that one could instead also talk about ('classical') cumulants. In probability, it turns out that cumulants interact well with independence in the sense that the cumulants of a sum of independent random variables are simply the sums of their cumulants. We now introduce cumulants into free probability (albeit with a definition that looks very different from the classical one) and discuss how they interact with free independence.

Definition 2.30. Let $(\mathcal{A}, \varphi)$ be a NCPS. We define, for natural numbers $n \geq 1$, the multi-linear maps $\kappa_n : \mathcal{A}^n \to \mathbb{C}$, called *cumulants*, recursively by the *moment-cumulant formula*

$$\varphi(a_1 \cdots a_n) = \sum_{\pi \in \mathrm{NC}(n)} \kappa_\pi(a_1, \ldots, a_n),$$

where $a_1, \ldots, a_n \in \mathcal{A}$ and if $\pi = (V_1, \ldots, V_k)$,

$$\kappa_\pi(a_1, \ldots, a_n) = \prod_{\substack{V \in \pi \\ V = \{i_1, \ldots, i_\ell\}}} \kappa_\ell(a_{i_1}, \ldots, a_{i_\ell}).$$

Remark 2.31. The first two cumulants are in fact not new objects, when restricted to $a_i = a$. It is obvious that $\kappa_1(a) = \varphi(a)$, so that the first cumulant is simply the mean. For the second cumulant, we have

$$\varphi(a^2) = \kappa_{\{}\{1,2\}\}(a,a) + \kappa_{\{\{1\},\{2\}\}}(a,a) = \kappa_2(a,a) + \kappa_1(a)^2 = \kappa_2(a,a) + \varphi(a)^2,$$

from which we conclude that the second cumulant is the variance.

Cumulants have the following significant advantage over moments when it comes to recognising freeness.

Theorem 2.32. *Let $a_1, \ldots, a_n$ be elements of a NCPS, then they are free if and only if all mixed cumulants vanish. To be more precise, the a_i are free if and only if for all tuples $i_1, \ldots, i_\ell$ such that not all members are equal, we have that $\kappa_\ell(a_{i_1}, \ldots, a_{i_\ell}) = 0$. Note that the a_i do not have to be centred.*

The proof of this theorem can be found in Part 2 of [65], but the main ingredients are two properties of the free cumulants (proofs of which can be found in the same reference). The first is that there exists an explicit formula to compute cumulants where some of the entries are products, in terms of the cumulants of the factors of said product. The second is that any cumulant where an entry is proportional to the identity operator vanishes (this explains why centering is irrelevant for cumulants).

To apply this new technology to semicircular families, we should first compute the cumulants of single semicircular element s. We will write $\kappa_\ell(s) = \kappa_\ell(s, \ldots, s)$ for brevity. Since s is centred and the first cumulant is the mean, $\kappa_1(s) = 0$, and we know that $\kappa_2(s) = (\rho/2)^2 C_1 = (\rho/2)^2$ as κ_2 is the variance. For the third cumulant we can calculate

$$0 = \varphi(s^3) = \kappa_3(s) + 2\kappa_1(s)\kappa_2(s) + \kappa_1(s)^3 = \kappa_3(s),$$

so that it vanishes. We now want to prove using induction that in fact every cumulant of order $\ell \geq 3$ is identically zero. First consider the case that $\ell = 2k + 1$ for $k \geq 1$, then

$$0 = \varphi(s^{2k+1}) = \sum_{\pi \in \mathrm{NC}(2k+1)} \kappa_\pi(s) = \kappa_{2k+1}(s) + \sum_{\pi \in \mathrm{NC}(2k+1)\setminus\{[n]\}} 0 = \kappa_{2k+1}(s),$$

because any π must contain a block of odd length, and by the induction hypothesis all odd cumulants of degree less than 2k+1 vanish. Next is the case that $\ell = 2k$ for $\kappa \geq 2$, then

$$\left(\frac{\rho}{2}\right)^{2k} C_k = \varphi(s^{2k}) = \kappa_{2k}(s) + \sum_{\pi \in \mathrm{NC}_2(2k)} (\kappa_2(s))^k + \text{ rest.}$$

By the induction hypothesis, all the other terms vanish because they contain some moment that is not of degree 2. However, it now follows that

$$\left(\frac{\rho}{2}\right)^{2k} C_k = \kappa_{2k}(s) + \left(\frac{\rho}{2}\right)^{2k} \sum_{\pi \in \mathrm{NC}_2(2k)} 1 = \kappa_{2k}(s) + \left(\frac{\rho}{2}\right)^{2k} C_k,$$

whence it must also hold that $\kappa_{2k}(s) = 0$. Notice that this is the same structure as that which we found for the 'classical' cumulants of the Gaußian density in 1.14. If we now want to describe a free semicircular family in terms of cumulants, it follows by Theorem 2.32 that all mixed cumulants vanish, and that only the second cumulant from the non-mixed ones does not vanish. We capture all this information in the following definition.

Definition 2.33. Let $(\mathcal{A}, \varphi)$ be a NCPS, $a_1, \ldots, a_n \in \mathcal{A}$, and denote the cumulants on this space by κ_ℓ. Then the a_i are a *free semicircular family* iff each a_i is a semicircular element, and we have the cumulant identities

$$\kappa_m(a_{i_1}, \ldots, a_{i_m}) = \delta_{m,2}\delta_{i_1,i_2}\frac{\rho_{i_1}^2}{4},$$

for all $m \geq 1$ and $1 \leq i_1, \ldots, i_m \leq n$. In particular, a *standard semicircular family* satisfies

$$\kappa_m(a_{i_1}, \ldots, a_{i_m}) = \delta_{m,2}\delta_{i_1,i_2}.$$

2.5 Free Group Factors

Let G be a countable group. In principle the ideas of this section are more general, but the countable case is sufficient for our purposes. We already introduced the group algebra $\mathbb{C}G$, and now we will consider some extensions of this idea. There is a natural inner product on the vector space $\mathbb{C}G$, which is given on G by $\langle g, g\rangle = 1$ and $\langle g, h\rangle = 0$ for $g \neq h$, and then extended sesquilinearly. This allows us to define

$$L^2(G) := \overline{\mathbb{C}G}^{\langle \cdot, \cdot \rangle},$$

the Hilbert space of square summable functions on G. Notice that $\mathbb{C}G$ can be interpreted as the subspace of finitely supported functions on G. There is a natural orthonormal basis of $L^2(G)$ given by $\{\delta_g\}_{g\in G}$. We can now define two unitary representations of G on $L^2(G)$, i.e. homomorphisms $\lambda, \rho : G \to B(L^2(G))$. They are given by the formulas $\lambda_g\delta_h = \delta_{gh}$ and $\rho_g\delta_h = \delta_{gh^{-1}}$. One readily checks that these assignments respect the group structure ($\lambda_g\lambda_h = \lambda_{gh}$, etc.) and that the operators λ_g and ρ_g are unitary ($\lambda_g^* = \lambda_{g^{-1}}$). These are called the *left* (λ) and *right* (ρ) *regular representations* of G. Consider now the von Neumann algebras generated by these representations,

$$\mathcal{L}(G) := \{\lambda_g \,|g \in G\}'', \ \mathcal{R}(G) := \{\rho_g \,|g \in G\}'',$$

which we say are the *left* and *right group von Neumann algebras*. They turn out not to be independent, as summarised in the following proposition.

Proposition 2.34. *We have that $\mathcal{L}(G)' = \mathcal{R}(G)$ and $\mathcal{R}(G)' = \mathcal{L}(G)$.*

Proof. It suffices to show that $\mathcal{R}(G)' = \mathcal{L}(G)$. For $u, v \in L^2(G)$ such that at least one is finitely supported, we introduce their *convolution* $u \star v$ as

$$(u \star v)(g) := \sum_{h\in G} u(h)v(h^{-1}g),$$

which lies again in $L^2(G)$ by an easy estimate. If u is finitely supported, the operator $u\star$ is an element of $\mathcal{L}(G)$, because it can be written as

$$u\star = \sum_{g\in G} u(g)\lambda_g.$$

Similarly, if v is finitely supported, one can write

$$\star v = \sum_{g\in G} v(g)\rho_g^*,$$

so that $\star v$ is in $\mathcal{R}$.

As $\lambda_g\rho_h = \rho_h\lambda_g$ for any $g, h \in G$, we have the inclusion $\mathcal{L}(G) \subset \mathcal{R}(G)'$. For the reverse inclusion, suppose that $x \in \mathcal{R}(G)'$ and $v \in L^2(G)$ is finitely supported. Then

$$xv = \sum_{g\in G} v(g)x\delta_g = \sum_{g\in G} v(g)\rho_g^*(x\delta_e) = (x\delta_e) \star v.$$

We conclude that for any $x \in \mathcal{R}'$ there is a function ξ in $L^2(G)$ such that $xv = \xi \star v$ for all v finitely supported. One can show that for any $y \in \mathcal{L}'$ there is also a function η in $L^2(G)$ such that $yu = u \star \eta$ for all u finitely supported.

To finish the proof, we argue that any $x \in \mathcal{R}(G)'$ is also in $\mathcal{L}(G)'' = \mathcal{L}(G)$. To this end, let $y \in \mathcal{L}(G)'$. Then there exist functions ξ and η as above. Recall now that the finitely supported

functions are dense in $L^2(G)$, because they can be identified with $\mathbb{C}G$. Hence there are sequences (ξ_n) and (η_n) of finitely supported functions approximating ξ and η in norm. Moreover, it suffices to prove that x and y commute when acting on finitely supported functions. Let u be such a function, then

$$xyu = x(u \star \eta) = \lim_{n\to\infty} \xi \star (u \star \eta) = \lim_{n\to\infty} \lim_{m\to\infty} \xi_m \star (u \star \eta_n).$$

We leave it to the reader to show that the convolution operation is associative on finitely supported functions, and that the limits can be interchanged (take the inner product with a finitely supported function v). Using this, we finally arrive at

$$xyu = \lim_{m\to\infty} \lim_{n\to\infty} (\xi_m \star u) \star \eta_n = \lim_{m\to\infty} y(\xi_m \star u) = yxu.$$

□

From now on we will focus on $\mathcal{L}(G)$. We collect some basic results concerning the properties of this von Neumann algebra coming from properties of G, following Chapter 6 of [56].

Theorem 2.35. *$\mathcal{L}(G)$ is a factor iff G is an ICC group (all non-identity conjugacy classes are infinite).*

Proof. Let x be in the center of $\mathcal{L}(G)$, then by the previous proposition x commutes with all elements of both $\mathcal{L}(G)$ and $\mathcal{R}(G)$. Consider the function $x\delta_e \in L^2(G)$, it has the property that $\lambda_g\rho_g(x\delta_e) = x(\lambda_g\rho_g\delta_e) = x\delta_e$, i.e. it is constant on conjugacy classes. If G is ICC, then square summability forces $x\delta_e$ to vanish on all non-identity conjugacy classes, and it follows that $x \in \mathbb{C}I$. Conversely, if G is not ICC, then the indicator function of a finite non-identity conjugacy class comes from a central element that is not a multiple of the identity, so that $\mathcal{L}(G)$ cannot be a factor. □

Proposition 2.36. *The functional $\tau(x) = \langle xe, e\rangle$ defines a faithful trace on $\mathcal{L}(G)$.*

Proof. We check the trace property on the generating elements λ_g, λ_h, for $g, h \in G$. It holds that $\tau(\lambda_g\lambda_h)$ vanishes unless $\lambda_g\lambda_h e = e = \lambda_e e$, but if $\lambda_g\lambda_h \neq \lambda_e$ then $\lambda_h\lambda_g \neq \lambda_e$, and if $\lambda_g\lambda_h = \lambda_e$ then $\lambda_h\lambda_g = \lambda_e$, so τ is a trace.

Suppose that $x \in \mathcal{L}(G)$ is such that $0 = \tau(x^*x) = \langle xe, xe\rangle$, then $xe = 0$. Let again $g, h \in G$, then we can write

$$\langle xg, h\rangle = \langle h^{-1}xge, e\rangle = \tau(\lambda_{h^{-1}}x\lambda_g) = \tau(\lambda_g\lambda_{h^{-1}}x) = \langle gh^{-1}xe, e\rangle = 0,$$

as $xe = 0$. □

Theorem 2.37. *If G is ICC, then $\mathcal{L}(G)$ is a factor of type II_1.*

Proof. By Theorem 2.6 in Chapter V of [82], it suffices to construct a faithful normal tracial state, as $\mathcal{L}(G)$ is infinite dimensional. The faithful trace from the previous proposition is a vector state, so it is normal (see Theorem 2.6 in chapter II of [82]). □

As mentioned earlier, freeness is analogous to freeness in groups, so we specialise to the case of free groups on a finite number of generators, $\mathbb{F}_n$ for some $n \in \mathbb{Z}_{\geq 1}$. Here are some elementary properties of these groups that are relevant to us.

Proposition 2.38. *Let m and n be positive integers.*

(i) $\mathbb{F}_n \cong \mathbb{F}_m$ *if and only if* $n = m$.

(ii) $\mathbb{F}_n \cong *_{i=1}^n \mathbb{Z}$, *hence* $\mathbb{F}_{n+m} \cong \mathbb{F}_n * \mathbb{F}_m$.

(iii) $\mathbb{F}_n$ *is an ICC group.*

Proof. For the first statement see Section 6.2 in [56]. The second statement holds essentially by definition. For the final statement see Exercise 2 in Chapter 6 of [56]. □

Point (iii) of this theorem justifies the following definition.

Definition 2.39. The von Neumann algebras $\mathcal{L}(\mathbb{F}_n)$, for n a positive integer, are called the *free group factors*

The free group factors are at the heart of a long standing open problem in von Neumann algebra theory. It is known that their little reduced C* brothers are pairwise non-isomorphic [68]. By contrast, $B(L^2(\mathbb{F}_n))$ and $B(L^2(\mathbb{F}_m))$ are always isomorphic because the Hilbert spaces are all separable. The free group factors live in between these two cases, and it is unknown on which side of the fence they fall. However, the following is known.

Theorem 2.40 (Free group factor dichotomy)**.** *Either the free group factors are all pairwise isomorphic, or they are all pairwise non-isomorphic.*

Proof. Due to Dykema [29] and Rădulescu [69]. □

3 q-Gaußians

3.1 Construction of q-Gaußian von Neumann Algebras

In this section we discuss a family of von Neumann algebras that will turn out to be related to the free group factors. Recall that $\mathcal{L}(\mathbb{F}_n)$ is generated by the elements λ_{a_i}, where the a_i are the n letters generating $\mathbb{F}_n$. These can be thought of as 'shift' operators in some sense, as they prefix the letter a_i to a word in the group algebra. This is reminiscent of the semi-circular element we considered earlier. We started that construction with a non-unitary isometry a^*, which satisfied $aa^* = I \neq a^*a$ (amongst other things), and we saw that this admitted a representation as shift operators. We now consider a deformation of this relation, inspired by the canonical commutation relations of quantum mechanics, as these also describe a kind of 'shift' operators that create and annihilate particles.

Let q be a number in the interval $[-1, 1]$ and define the *q-mutator* of two elements x and y as $[x, y]_q := xy - qyx$. Then we can search for operators a_i and a_i^* acting on a Hilbert space satisfying the *q-mutation* (or *q-deformed* commutation) relations

$$[a_i, a_j^*]_q = a_i a_j^* - q a_j^* a_i = \delta_{ij} I. \tag{3.1.1}$$

Notice that for $q = -1, 1$ we recover the well known canonical commutation and anti-commutation relations. For $q = 0$ we recognize the non-unitary isometries. In the extremal cases where $q = -1, 1$, one often constructs these operators acting on a so-called *Fock space* of quantum states. We will be inspired by this approach, but will face the technical difficulty that the 'obvious' definitions of a_i and a_j^* do not define adjoint operators with respect to the inner product. This will be overcome by also deforming the inner product in such a way that it forces a_i and a_i^* to be true adjoints of each other.

Let now $\mathcal{H}_\mathbb{R}$ be a real, separable Hilbert space, and denote by $\mathcal{H}$ its complexification with inner product $(\cdot, \cdot)$. We will use $\boxtimes$ to denote the ordinary Hilbertian tensor product, i.e. the vector space tensor product endowed with the canonical extension of the inner product $(\cdot, \cdot)$. We will however be somewhat sloppy and still denote elements in it as $e_1 \otimes e_2$. Consider the *Fock space* $\mathcal{F}(\mathcal{H})$ associated to $\mathcal{H}$, which is defined as

$$\mathcal{F}(\mathcal{H}) := \mathbb{C}\Omega \oplus \bigoplus_{n \geq 1} \mathcal{H}^{\boxtimes n},$$

where Ω is a unital vector, called the *vacuum*. By definition $\mathcal{F}(\mathcal{H})$ is spanned by words of the form $e_1 \otimes \cdots \otimes e_n$, with $e_i \in \mathcal{H}$ and the vacuum Ω. The vacuum expectation value functional on $B(\mathcal{F}(\mathcal{H}))$, $(\cdot\Omega, \Omega)$, will be denoted by ω.

Definition 3.1. Let $e \in \mathcal{H}_\mathbb{R}$. The *left* and *right creation operators* $a^*(e)$ and $a_r^*(e)$ associated to e are defined by the formulae

$$\begin{cases} a^*(e)\Omega = e \\ a^*(e)(e_1 \otimes \cdots \otimes e_n) = e \otimes e_1 \otimes \cdots \otimes e_n \end{cases}, \quad \begin{cases} a_r^*(e)\Omega = e \\ a_r^*(e)(e_1 \otimes \cdots \otimes e_n) = e_1 \otimes \cdots \otimes e_n \otimes e \end{cases}.$$

On the other hand, the *left* and *right annihilation operators* $a(e)$ and $a_r(e)$ associated to e are defined by the formulae

$$a(e)\Omega = 0, \ a(e)(e_1 \otimes \cdots \otimes e_n) = \sum_{k=1}^{n} q^{k-1}(e, e_k)(e_1 \otimes \cdots \otimes \widehat{e_k} \otimes \cdots \otimes e_n),$$

and

$$a_r(e)\Omega = 0, \ a_r(e)(e_1 \otimes \cdots \otimes e_n) = \sum_{k=1}^{n} q^{n-k}(e, e_k)(e_1 \otimes \cdots \otimes \widehat{e_k} \otimes \cdots \otimes e_n),$$

where the hat means omission of the vector from the tensor.

Remark 3.2. The definition of the annihilation operator can be motivated by imagining that one q-mutes the annihilation operator to each vector in the tensor, supposing that each vector was put there by a creation operator acting on the vacuum.

It is now an easy computation to check that these operators witness the q-mutation relations. Let f_1 and f_2 be vectors in $\mathcal{H}_{\mathbb{R}}$, then

$$a(f_1)a^*(f_2)\Omega = a(f_1)f_2 = (f_1, f_2)\Omega,$$

and

$$\begin{aligned} a(f_1)a^*(f_2)(e_1 \otimes \cdots \otimes e_n) &= a(f_1)(f_2 \otimes e_1 \otimes \cdots \otimes e_n), \\ &= (f_1, f_2)(e_1 \otimes \cdots \otimes e_n) + \sum_{k=1}^{n} q^k (f_1, e_k)(f_2 \otimes e_1 \otimes \cdots \otimes \widehat{e_k} \otimes \cdots \otimes e_n), \\ &= \left[(f_1, f_2) + q a^*(f_2)a(f_1)\right](e_1 \otimes \cdots \otimes e_n). \end{aligned}$$

One recovers the precise form of (3.1.1) if one chooses the f_i to be orthonormal. The computation for the right operators is similar.

Unfortunately, as the reader may check, the operators defined in this way are not adjoints with respect to our current inner product on $\mathcal{F}(\mathcal{H})$. Fortunately, it is not very hard to write down something that looks like an inner product for which this is clearly the case.

Definition 3.3. Define the sesquilinear map $\langle \cdot, \cdot \rangle_q : \mathcal{F}(\mathcal{H}) \times \mathcal{F}(\mathcal{H}) \to \mathbb{C}$ recursively by the rules that it vanishes on pure tensors of different rank, and when the ranks agree write

$$\begin{aligned} \langle e_1 \otimes \cdots \otimes e_n, f_1 \otimes \cdots \otimes f_n \rangle_q &= \langle e_2 \otimes \cdots \otimes e_n, a(e_1)(f_1 \otimes \cdots \otimes f_n) \rangle_q, \\ &= \sum_{k=1}^{n} q^{k-1} (e_1, f_k) \langle e_2 \otimes \cdots \otimes e_n, f_1 \otimes \cdots \otimes \widehat{f_k} \otimes \cdots \otimes f_n \rangle_q \end{aligned}$$

The thorough reader is invited to check that (at least formally) $a_r(e)$ and $a_r^*(e)$ are true adjoints with respect to $\langle \cdot, \cdot \rangle_q$ as well.

It turns out to be surprisingly difficult to show that $\langle \cdot, \cdot \rangle_q$ is in fact an inner product. Bożejko and Speicher were the first to achieve this in their seminal paper [13]. We outline their approach.

Definition 3.4. Define an operator $P_q : \mathcal{F}(\mathcal{H}) \to \mathcal{F}(\mathcal{H})$ by first decomposing it as

$$P_q = \bigoplus_{n \geq 0} P_q^{(n)}, \ P_q^{(n)} : \mathcal{H}^{\boxtimes n} \to \mathcal{H}^{\boxtimes n}.$$

Consider the natural unitary action U of S_n on $\mathcal{H}^{\boxtimes n}$ given by permuting letters, i.e. $U_\sigma(e_1 \otimes \cdots \otimes e_n) = e_{\sigma(1)} \otimes \cdots \otimes e_{\sigma(n)}$. Set

$$|\sigma| := |\{(i,j) \,|1 \leq i < j \leq n, \ \sigma(i) > \sigma(j)\}|,$$

which is the number of involutions of σ, and finally put $P_q^{(n)} := \sum_{\sigma \in S_n} q^{|\sigma|} U_\sigma$, and $P_q^{(0)} = I$.

Proposition 3.5. *The operator P_μ is positive, and even strictly positive if $|q| < 1$. Moreover, it holds that*

$$\langle \xi, \eta \rangle_q = (\xi, P_q \eta).$$

In particular, $\langle \cdot, \cdot \rangle_q$ is an inner product when $|q| < 1$, and in the cases $q = \pm 1$ it as an inner product after dividing out the kernel.

Proof. See Proposition 1 and Lemma 3 in [13]. □

Remark 3.6. Observe in the definition of P_q that $P_q^{(0)}$ and $P_q^{(1)}$ are both equal to the identity operator, so that the deformation of the inner product takes place away from the vacuum sector and the canonical embedding of $\mathcal{H}$ into $\mathcal{F}(\mathcal{H})$.

Definition 3.7. We write $\mathcal{F}_q(\mathcal{H})$ for the completion of $\mathcal{F}(\mathcal{H})$ with respect to the inner product $\langle \cdot, \cdot \rangle_q$. The tensor product with respect to this inner product will be denoted by the ordinary tensor symbol $\otimes$.

Before moving on, it is interesting to compute the norms of the operators we have introduced. It is well known that the $q = 1$ case, $aa^* - a^*a = I$, cannot be represented by operators that are bounded. This result should be recovered, and it is thus also not a priori clear that the operators we defined are bounded. This does however turn out to be the case.

Lemma 3.8. *For any $e \in \mathcal{H}$, the associated annhiliation operator $a(e)$ is bounded as long as q is not equal to* 1*. Moreover, one has that*

$$\|a(e)\|_q = \begin{cases} \|e\|_{\mathcal{H}}, & -1 \leq q \leq 0 \\ \dfrac{\|e\|_{\mathcal{H}}}{\sqrt{1-q}}, & 0 \leq q < 1 \end{cases}.$$

Proof. The case $-1 \leq q \leq 0$ is very straightforward. Let $\xi \in \mathcal{F}_q(\mathcal{H})$, then by the q-mutation relations,

$$\|a^*(e)\xi\|_q^2 = \|e\|_{\mathcal{H}}^2 \|\xi\|_q^2 + q\|a(e)\xi\|_q^2 \leq \|e\|_{\mathcal{H}}^2 \|\xi\|_q^2,$$

and $\|a(e)\Omega\| = \|e\|_{\mathcal{H}}$. As taking adjoints is isometric, it follows that $\|a(e)\|_q = \|a^*(e)\|_q = \|e\|_{\mathcal{H}}$.

Assume now that $0 \leq q < 1$, and without loss of generality that ξ is in $\mathcal{H}^{\otimes n}$. Then $a^*(e)\xi = e \otimes \xi$ is a rank $n+1$ tensor, and we have by definition

$$\|a^*(e)\xi\|_q^2 = \left(e \otimes \xi, P_q^{(n+1)} e \otimes \xi\right).$$

We could factor the norm of e out of this expression if we had a way to replace $P_q^{(n+1)}$ by $I \otimes P_q^{(n)}$. To do this, we must find a way to factorise the permutations we sum over in the operator P_q. Hence consider the transpositions σ_i in S_{n+1} that interchange i and $i+1$ and leave all other numbers fixed. We can decompose each permutation of $\{1, \ldots, n+1\}$ into a part that swaps 1 and its preimage via transpositions, and a part that does the rest. As we sum over all such permutations, the following identity is not difficult to see

$$P_q^{(n+1)} = \left(I \otimes P_q^{(n)}\right) \circ \left(q^0 I + q^1 U_{\sigma_1} + \cdots + q^n U_{\sigma_1 \cdots \sigma_n}\right).$$

Recall from Proposition 3.5 that the operators $P_q^{(m)}$ are all strictly positive and self-adjoint, moreover all coefficients q^k are positive, so that

$$0 < \left(P_q^{(n+1)}\right)^2 \leq \left[(1 + q + \cdots + q^n)\left(I \otimes P_q^{(n)}\right)\right]^2.$$

Therefore

$$0 < P_q^{(n+1)} \leq (1 + q + \cdots + q^n)\left(I \otimes P_q^{(n)}\right) < \frac{1}{1-q}\left(I \otimes P_q^{(n)}\right).$$

It follows that

$$\|a^*(e)\xi\|_q^2 \leq \frac{\|e\|_{\mathcal{H}}^2}{1-q} \|\xi\|_q^2.$$

To finish the proof, notice that if we set $\xi = e^{\otimes n}$ we have that

$$\left\|a^*(e)e^{\otimes n}\right\|_q = (1+q+\cdots+q^n)^{1/2}\|e\|_{\mathcal{H}}\left\|e^{\otimes n}\right\|_q,$$

which implies that we can produce a sequence of vectors that forces the operator norm to be at least $\|e\|_{\mathcal{H}}/\sqrt{1-q}$. □

With the above construction in hand we are ready to introduce the central objects of this chapter and prove some of their most elementary properties.

Definition 3.9. For $e \in \mathcal{H}_{\mathbb{R}}$, write

$$W(e) = a(e) + a^*(e),\ W_r(e) = a_r(e) + a_r^*(e),$$

which define self-adjoint operators. If $\{e_i\}_{i\in I}$ is an orthonormal basis of $\mathcal{H}$, we often write W_{i_j} for $W(e_{i_j})$. These elements generate a von Neumann algebra that we shall call the *q-Gaußian*,

$$\Gamma_q(\mathcal{H}_{\mathbb{R}}) := \{W(e)\,|e \in \mathcal{H}_{\mathbb{R}}\}''.$$

Similarly there is a *right q-Gaußian* $\Gamma_{q,r}(\mathcal{H}_{\mathbb{R}})$. For this chapter, if we write $\Gamma_{q(,r)}$, we mean $\Gamma_{q(,r)}(\mathcal{H}_{\mathbb{R}})$.

Lemma 3.10. *Recall that $\omega(\cdot) = \langle \cdot\Omega, \Omega\rangle_q$ is the vacuum expectation. The following statements are true for both Γ_q and $\Gamma_{q,r}$*

(i) ω is a trace

(ii) The vacuum vector Ω is separating and cyclic.

(iii) The commutant of one is the other, that is $\Gamma_q' = \Gamma_{q,r}$.

Proof. Throughout this proof, we will only present the arguments for the left q-Gaußian, as those for the right one are completely analogous. Also, $\{e_i\}_{i\in I}$ is an orthonormal basis of $\mathcal{H}_{\mathbb{R}}$, whence the span of the words on these basis elements forms a dense subset of $\mathcal{F}_q(\mathcal{H})$.

For the trace property (part (i)), first notice that by definition

$$\omega(W_{i_1}\cdots W_{i_{2r}}) = \sum_{\varepsilon(1)=1}^{*}\cdots\sum_{\varepsilon(2r)=1}^{*}\omega(a_{i_1}^{\varepsilon(1)}\cdots a_{i_{2r}}^{\varepsilon(2r)}),$$

where we took the product to be of even length because otherwise the result is zero by a rank argument. The value of the summands can be determined as follows. It is necessary that the product contains an equal number of creation and annihilation operators, otherwise the result is again zero by a rank argument. Moreover, this argument applies to each vector e_{i_j} that occurs in the product, as these are orthonormal, so there must be an equal number of creation and annihilation operators for each e_{i_j}. Starting from the right, if the first time a vector e_{i_j} appears is in an annihilation operator, the result is zero as this operator can be q-muted to the vacuum and annihilate it. It follows that any non-zero term must allow pairing of creation and annihilation operators belonging to the same vector, where the annihilation operator occurs to the left of the creation operator in the pair, and all these pairs exhaust the product. An example would be

$$\omega(a_{i_a}^{\varepsilon(a)}a_{i_b}^{\varepsilon(b)}\cdots a_{i_\beta}^{\varepsilon(\beta)}\cdots a_{i_c}^{\varepsilon(c)}\cdots a_{i_d}^{\varepsilon(d)}\cdots a_{1_\alpha}^{\varepsilon(\alpha)}\cdots a_{i_\gamma}^{\varepsilon(\gamma)}\cdots a_{i_\delta}^{\varepsilon(\delta)}),$$

where the operators in a pair are connected with a line. To evaluate this, one can q-mute the elements until the operators in each pair are adjacent, resulting in some power of q as the result, as $a_{i_j}a_{i_j}^*\Omega = \Omega$.

We claim that this power is precisely the number of crossings in the expression above. This is not hard to see, but for a formal proof one could proceed to prove the following formula (see Proposition 3.12 for the notation),

$$\omega(a_{i_1}^{\varepsilon(1)}\cdots a_{i_m}^{\varepsilon(m)}) = \begin{cases} 0, & m \text{ odd}, \\ \sum_{k,\ell=0}^{m} \sum_{(I,J)\in P_2^{k,\ell}(m)} \delta_{\text{pairs}}(I,J) q^{i(I,J)}, & m = 2r \end{cases}, \tag{3.1.2}$$

where the δ_{pairs} enforces the conditions explained above. One could show this for products of the form $a_{i_1}^*\cdots a_{i_n}^* a_{i_{n+1}}\cdots a_{i_m}$ and then observe that both formulas transform correctly under the q-mutation relations.

To see that this formula defines a trace on Γ_q, observe what happens when we cyclically permute our expression to

$$\omega(\overbrace{a_{i_b}^{\varepsilon(b)}\cdots a_{i_\beta}^{\varepsilon(\beta)}}\cdots \overbrace{a_{i_c}^{\varepsilon(c)}\cdots \overbrace{a_{i_d}^{\varepsilon(d)}\cdots \overbrace{a_{1_\alpha}^{\varepsilon(\alpha)}\cdots \overbrace{a_{i_\gamma}^{\varepsilon(\gamma)}\cdots a_{i_\delta}^{\varepsilon(\delta)}}\cdots a_{i_a}^{\varepsilon(a)}}).$$

This has not changed the number of crossings, but it has swapped the creation and annihilation operators in one pair. As we sum over all possible values of the $\varepsilon(j)$, this does not matter, as the term with the correct order is also in the sum.

To show that Ω is cyclic (part (ii)), we prove that $\Gamma_q\Omega$ contains all pure tensors of finite rank. The proof will proceed via induction on the rank of the tensor. The case of rank 1 is clear, as then $W(e_{i_1})\Omega = a^*(e_{i_1})\Omega = e_{i_1}$. Suppose now that all tensors of rank k are in $\Gamma_q\Omega$, and moreover that the operators in Γ_q that create rank k tensors ξ are denoted by $W(\xi)$. These exist by the induction hypothesis, and we will later see that they are in fact unique, but we do not use this here. Moreover the notation is consistent with the original meaning of W. Consider then

$$W_{i_1}W(e_{i_2}\otimes\cdots\otimes e_{i_{k+1}})\Omega - W(a(e_{i_1})(e_{i_2}\otimes\cdots\otimes e_{i_{k+1}}))\Omega = e_{i_1}\otimes\cdots\otimes e_{i_{k+1}}.$$

Correspondingly, there exist operators $W_r(\xi)$ in the right q-Gaußian.

To show that Ω is separating, we show that it is cyclic for the commutant $(\Gamma_q)'$. To this end, define the involution S on $\Gamma_q\Omega$ by $Sx\Omega = x^*\Omega$. By the trace property of ω this is isometric, and hence well-defined. Moreover, $S\Gamma_q S \subset (\Gamma_q)'$ as for any $y\in\Gamma_q$

$$SxSy\Omega = Sxy^*\Omega = yx^*\Omega = ySx\Omega = ySxS\Omega,$$

but then Ω is cyclic for the commutant as it is already cyclic for $S\Gamma_q S$.

For later use, define the isometric involution J on $\mathcal{F}_q(\mathcal{H}_\mathbb{R})$ by

$$J(e_{i_1}\otimes\cdots\otimes e_{i_n}) = e_{i_n}\otimes\cdots\otimes e_{i_1},$$

and demanding that it is anti-linear. One can check that $Ja^*(e_i)J = a_r^*(e_i)$ and $Ja(e_i)J = a_r(e_i)$, so in particular $JW(e_i)J = W_r(e_i)$.

To determine the commutant (part (iii)), it is worth digging a little bit deeper into $S\Gamma_q S$ constructed above. In particular what happens when we apply S to an operator of the form $W(\xi)$, with ξ a pure tensor, is that $SW(\xi) = W(J\xi)$. We prove this claim by induction on the rank of a basis tensor $e_{i_1}\otimes\cdots\otimes e_{i_n}$. The case $n=2$ is a straightforward verification. For the induction step we need the observation that $\Gamma_{q,r}\subset(\Gamma_q)'$ because the generators of the left and right q-Gaußians commute. Then

$$\begin{aligned} W(e_{i_1}\otimes\cdots\otimes e_{i_n})^*\Omega &= W(e_{i_2}\otimes\cdots\otimes e_{i_n})^*W(e_{i_1})\Omega - W(a(e_{i_1})(e_{i_2}\otimes\cdots\otimes e_{i_n}))^*\Omega, \\ &= W_r(e_{i_1})(e_{i_n}\otimes\cdots\otimes e_{i_2}) - J\left(a(e_{i_1})(e_{i_2}\otimes\cdots\otimes e_{i_n})\right), \\ &= a_r^*(e_{i_1})(e_{i_n}\otimes\cdots\otimes e_{i_2}) + \left[a_r(e_{i_1})J - Ja(e_{i_1})\right](e_{i_2}\otimes\cdots\otimes e_{i_n}), \\ &= W(J(e_{i_1}\otimes\cdots\otimes e_{i_n}))\Omega \end{aligned}$$

From this it actually follows that $S\Gamma_q S = \Gamma_{q,r}$, as it turns out that S and J are the same operator, and conjugation by J interchanges the generators of the left and right q-Gaußians. The equality of S and J follows from the computation

$$S(e_{i_1} \otimes \cdots \otimes e_{i_n}) = SW(e_{i_1} \otimes \cdots \otimes e_{i_n})\Omega = W(J(e_{i_1} \otimes \cdots \otimes e_{i_n}))\Omega = J(e_{i_1} \otimes \cdots \otimes e_{i_n}).$$

It remains to prove the inclusion $(\Gamma_q)' \subset S\Gamma_q S$ to arrive at the advertised relations. For this we prove that ω is also a trace on $(\Gamma_q)'$, so that by repeating the calculation from the proof that Ω is separating we get $S(\Gamma_q)'S \subset \Gamma_q'' = \Gamma_q$. Notice that our proof of part (ii) of the lemma also implies that the vacuum is cyclic and separating for $(\Gamma_q)'$. If $x' \in (\Gamma_q)'$, it holds that

$$\langle Sx\Omega, x'\Omega\rangle_q = \overline{\langle x\Omega, (x')^*\Omega\rangle_q}.$$

For any $\xi, \eta \in \mathcal{F}_q(\mathcal{H})$ we have by the polarisation identity and anti-linearity of S that

$$\langle S\xi, S\eta\rangle_q = \overline{\langle \xi, \eta\rangle_q},$$

in particular

$$\langle S\xi, \eta\rangle_q = \overline{\langle \xi, S\eta\rangle_q}.$$

However, this means that

$$\overline{\langle x\Omega, (x')^*\Omega\rangle_q} = \langle Sx\Omega, x'\Omega\rangle_q = \overline{\langle x\Omega, Sx'\Omega\rangle_q},$$

and we must conclude that $Sx' = (x')^*$. As S is isometric, the norms of $x'\Omega$ and $(x')^*\Omega$ must agree, which in turn implies that

$$\langle x'(x')^*\Omega, \Omega\rangle_q = \langle (x')^* x'\Omega, \Omega\rangle_q.$$

By polarisation, this shows that ω is a trace on $(\Gamma_q)'$. As discussed at the start of this paragraph, this implies that $S(\Gamma_q)'S \subset \Gamma_q$. Combining this with the other inclusion (which we derived in the proof of (ii)) we conclude that $\Gamma_q' = S\Gamma_q S = \Gamma_{q,r}$ to finish the proof. □

Remark 3.11. The key tool of the proof above was the introduction of the *Wick operators* $W(\xi)$. We can now a posteriori use part (ii) of the lemma to declare that for any $\xi \in \Gamma_q\Omega$, there exists a unique operator $W(\xi)$ defined by the fact that $W(\xi)\Omega = \xi$. On the other hand, it follows also that any $x \in \Gamma_q$ is entirely determined by $x\Omega$. We thus find a one to one correspondence between the orbit of the vacuum under the q-Gaußians, and their elements themselves. These operators will be very useful later to prove that the q-Gaußians are factors. There is even an explicit formula for the case that ξ is a word $e_1 \otimes \cdots \otimes e_n$ in $\mathcal{F}_q(\mathcal{H})$.

Proposition 3.12. *Denote by $P_2^{k,\ell}(n)$ all two-set partitions (I, J) of $\{1, \ldots, n\}$ such that $|I| = k$ and $|J| = \ell$. Naturally this is empty unless $k + \ell = n$, and we write $I = \{i(1), \ldots, i(k)\}$ and similarly for J. Let $e_1 \otimes \cdots \otimes e_n$ be in $\mathcal{F}_q(\mathcal{H})$, then we have the* Wick formula

$$W(e_1 \otimes \cdots \otimes e_n) = \sum_{k,\ell=0}^{n} \sum_{(I,J)\in P_2^{k,\ell}(n)} a^*(e_{i(1)}) \cdots a^*(e_{i(k)}) a(e_{j(1)}) \cdots a(e_{j(\ell)}) \cdot q^{i(I,J)}, \tag{3.1.3}$$

where

$$i(I, J) := |\{(p, q)\,|1 \le p \le k,\ 1 \le q \le \ell,\ i(p) > j(q)\}|.$$

Proof. See Proposition 2.7 in [12]. □

3.2 0-Gaußians are Free Group Factors

Aside from the physically interesting points $q = \pm 1$, there is also the case $q = 0$, which turns out to correspond to a von Neumann algebra that we have already seen.

Theorem 3.13 (Theorem 2.6.2. in [86])**.** *Let $n \in \mathbb{Z}_{>0}$, and suppose that $dim(\mathcal{H}_{\mathbb{R}}) = n$. Then $\Gamma_0(\mathcal{H}_{\mathbb{R}}) \cong \mathcal{L}(\mathbb{F}_n)$.*

One of the goals of this thesis is to show much more than this. It is also true that the q-Gaußians are type II_1 factors (this will be shown in the next section), and point (iii) of Lemma 3.10 looks rather similar to Proposition 2.34. In fact, we will prove in a later chapter that there exist numbers $q_*(n)$ such that all q-Gaußians with $|q| \leq q_*(n)$ (for $\dim(\mathcal{H}_{\mathbb{R}}) = n$) are isomorphic.

The proof of Theorem 3.13 will be split into a few steps, but most of the hard work has already been done. Naturally the result that we would like to use is Theorem 2.11. This has the corollary that if a von Neumann algebra is generated by say n free elements that have the same $*$-distribution, then this is the unique such von Neumann algebra (due to Proposition 2.15). This suggests the following strategy, we first show the easy result that the free group factors are generated by free Haar unitaries (recall definition 2.6). Then we prove that the 0-Gaußians are generated by free semi-circular elements, which we already encountered earlier. The final piece of the puzzle is then to show that these two types of generators are interchangeable, i.e. they can be deformed into one another.

Proposition 3.14. *$\mathcal{L}(\mathbb{F}_n)$ is generated as a von Neumann algebra by n free Haar unitaries.*

Proof. As $\mathbb{F}_n$ is the n-fold free product of groups of $\mathbb{Z}$, it follows by Proposition 2.17 that the $*$-algebras generated by the generators $a_1, \ldots, a_n$ of $\mathbb{F}_n$ are free. Via the left regular representation, we have $\lambda : \mathbb{C}\mathbb{F}_n \hookrightarrow \mathcal{L}(\mathbb{F}_n)$, and this is even an inclusion of NCPS. By Proposition 2.21, the von Neumann algebras generated by the $\lambda(a_i)$ are also free. It is a trivial computation to verify that they are Haar unitaries. Hence we have that $\mathcal{L}(\mathbb{F}_n)$ is generated as a von Neumann algebra by the n free Haar unitaries $\lambda(a_i)$. □

Proposition 3.15. *$\Gamma_0(\mathcal{H}_{\mathbb{R}})$ is generated as a von Neumann algebra by $dim(\mathcal{H}_{\mathbb{R}})$ free semi-circular elements.*

Proof. Let $e_1, \ldots, e_n$ be an ONB of $\mathcal{H}_{\mathbb{R}}$, then $\Gamma_0(\mathcal{H}_{\mathbb{R}})$ is generated by the elements $W(e_i)$. We thus have to show that these are free semi-circular elements. To show that they are semi-circular, we would like to prove that the $W(e_i)$ (or rather, the $a(e_i)$) satisfy the assumptions at the beginning of section 2.3 in the free probability part. This is however immediate, as by the 0-mutation relations it holds that $a(e_i)a^*(e_i) = I$, and one can check from their definitions that $a^*(e_i)a(e_i) \neq I$ (as for instance $a(e_i)\Omega = 0$). We invite the reader to check the linear independence of the elements $a^*(e_i)^m a(e_i)^n$ and that they have the right moments, namely $\omega((a^*)^m a^n) = \delta_{m,0}\delta_{n,0}$. Then by the results of section 2.3 we have that the $W(e_i)$ are standard semi-circular. Their freeness follows from orthogonality, and is the result of the next lemma. □

Lemma 3.16. *Let $e_1, \ldots, e_m$ be an ONS in $\mathcal{H}_{\mathbb{R}}$, and $\mathcal{M}_i$ the von Neumann algebras generated by the $W(e_i)$. The latter are free in the W*-probability space $(\Gamma_0(\mathcal{H}_{\mathbb{R}}), \omega)$*

Proof. By Proposition 2.21, it suffices to show that the $*$-algebras $\mathcal{A}_i$ generated by the $a(e_i)$ are free. Let x_i be an element of $\mathcal{A}_i$, then it can be written in the form

$$x_i = \alpha_i + \sum_{j=1}^{p} \beta_j (a_i^*)^{n(j)} (a_i)^{m(j)}$$

Here, α_i and β_j are complex numbers, and $(n(k), m(\ell)) \in \mathbb{Z}^2_{\geq 0} \setminus \{(0,0)\}$ Clearly, $\omega(x_i) = \alpha_i$, so that the centred elements of $\mathcal{A}_i$ are precisely those for which there is no identity term.

Let now $x_{i_1} \cdots x_{i_k}$ be an alternating centred product, then we may assume that each x_{i_j} is of the form $(a_{i_j}^*)^{n(i_j)}(a_{i_j})^{m(i_j)}$. We distinguish two cases. First, suppose that there is an $1 \leq \ell \leq k-1$ such that neither of $m(i_\ell)$ and $n(i_{\ell+1})$ vanishes. Then $x_{i_1} \cdots x_{i_k}$ contains a factor $a_{i_\ell} a_{i_{\ell+1}}^* = 0$, by the 0-mutation relations. Suppose there is no such ℓ. Then it holds that $x_{i_1} \cdots x_{i_k}$ is of the form $a_{j_1}^* \cdots a_{j_n}^* a_{j_{n+1}} \cdots a_{j_{n+m}}$, with $n+m > 0$. However, it is immediate that then $\omega(x_{i_1} \cdots x_{i_k}) = 0$. We conclude that the $*$-algebras $\mathcal{A}_i$ are free. □

Lemma 3.17. *A semi-circular element in a W*-probability space can be deformed into a Haar unitary.*

Proof. Write (M, τ) for the W*-probability space. Let s be a semi-circular element in M, then its spectral measure is $\sqrt{4-t^2}/(2\pi)\,\mathrm{d}t$, on the interval $[-2, 2]$. In particular, for any bounded measurable function f on $[-2, 2]$, we have that

$$\tau(f(s)) = \frac{1}{2\pi} \int_{-2}^{2} f(t)\sqrt{4-t^2}\ \mathrm{d}t$$

Define the function

$$H(t) = \frac{t}{4\pi}\sqrt{4-t^2} + \frac{1}{\pi}\arcsin(t/2),$$

then notice that the density of the semi-circular measure is precisely $H'(t)$. Through spectral calculus we define an element u of M, given by $u = \exp(2\pi i H(s))$. Notice that u generates the same von Neumann algebra as s. It follows that this is a Haar unitary,

$$\tau(u^k) = \int_{-2}^{2} \exp(2\pi i H(t)) H'(t)\ \mathrm{d}t = \int_{-1/2}^{1/2} e^{2\pi i k h}\ \mathrm{d}h = \delta_{k,0}.$$

□

We now have everything we need to prove Theorem 3.13.

Proof of Theorem 3.13. By Proposition 3.15 and Theorem 2.11, $\Gamma_0(\mathcal{H}_\mathbb{R})$ is the unique von Neumann algebra generated by n free semi-circular elements. Using Lemma 3.17 we can trade these for n Haar unitaries as generators. Since freeness is a property of the subalgebra generated by a random variable, it follows that these Haar unitaries are also free. We conclude that $\Gamma_0(\mathcal{H}_\mathbb{R})$ is also isomorphic to the unique von Neumann algebra generated by n free Haar unitaries, but this is $\mathcal{L}(\mathbb{F}_n)$ by Proposition 3.14. □

3.3 All q-Gaußians are type $\mathbf{II}_1$ Factors

Throughout this section, q will denote a real number of magnitude strictly less than 1, and we assume that $\dim(\mathcal{H}_\mathbb{R}) \geq 2$. In the previous section, we showed that if $q = 0$, the q-Gaußian algebras become free group factors. It is now natural to wonder which (if any) properties of the free group factors are shared with the q-Gaußians for non-zero q. At this point, we have developed (almost) enough technique to prove the following.

Theorem 3.18. Γ_q *is a type* II_1 *factor.*

We will present the proof of this statement in several parts. The key ingredients are Proposition 3.19 and Theorem 3.24. The former will prove to be an immediate consequence of what we have already done. The proof of the latter statement will take up most of this section, and requires some preparation.

Proposition 3.19. *The vacuum expectation ω is a faithful, tracial, and normal state on $\Gamma_q(\mathcal{H}_\mathbb{R})$.*

Proof. By definition, ω is a vector state, so it is normal (see again Theorem 2.6 in chapter II of [82]). We proved earlier, in Lemma 3.10, that ω is both tracial and faithful, as the vacuum vector is separating. □

We now begin our preparations to prove Theorem 3.24. The first tools we need are the *first and second quantisations* of a contractive map.

Proposition 3.20. *Let $T : \mathcal{H}_\mathbb{R} \to \mathcal{H}_\mathbb{R}$ be a linear contraction. Then there exists a canonical ($\mathbb{C}$-)linear contraction extending T on $\mathcal{F}_q(\mathcal{H})$, written as $\mathcal{F}_q(T)$ and called the* first quantisation *of T. Moreover, there exists a unique, unital, and completely positive map $\Gamma_q(T) : \Gamma_q(\mathcal{H}_\mathbb{R}) \to \Gamma_q(T\mathcal{H}_\mathbb{R})$, called the* second quantisation *of T, which satisfies*

$$\Gamma_q(T)(W(\xi)) = W(\mathcal{F}_q(T)\xi),$$

for all ξ in $\Gamma_q(\mathcal{H}_R)\Omega$.

Proof. For the first quantisation, let $\tilde{T}$ be the complexification of T. Then it is clear that

$$\mathcal{F}_q(T) := I_{\mathbb{C}\Omega} \oplus \left(\bigoplus_{n \in \mathbb{Z}_{\geq 1}} \tilde{T}^{\otimes n} \right)$$

witnesses the statement of the theorem.

We will only prove the proposition for the case that $T = P$ is an orthogonal projection onto a subspace $\mathcal{K}_\mathbb{R}$, as this is all we will need. For the full proof, see Theorem 2.11 in [12]. We have $P : \mathcal{H}_\mathbb{R} \to \mathcal{K}_\mathbb{R}$ and $P^* : \mathcal{K}_R \hookrightarrow \mathcal{H}_\mathbb{R}$ the canonical inclusion. We define for all $x \in \Gamma_q(\mathcal{H}_\mathbb{R})$

$$\Gamma_q(P)x := \mathcal{F}_q(P)x\mathcal{F}_q(P^*).$$

It suffices to prove that this expression works for all normal ordered products of creation and annihilation operators. Let $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ be vectors in $\mathcal{H}_\mathbb{R}$. From the definitions one can check that

$$a(e_i)\mathcal{F}_q(P^*) = \mathcal{F}_q(P^*)a(Pe_i),\ \mathcal{F}_q(P)a^*(f_j) = a^*(Pf_j)\mathcal{F}_q(P).$$

Then the following calculation finishes the proof,

$$\mathcal{F}_q(P)a^*(f_1)\cdots a^*(f_m)a(e_1)\cdots a(e_n)\mathcal{F}_q(P^*) = a^*(Pf_1)\cdots a^*(Pf_m)\mathcal{F}_q(PP^*)a(Pe_1)\cdots a(Pe_n),$$

as $PP^* = I_{\mathcal{K}_\mathbb{R}}$. □

Remark 3.21. The (somewhat awkward) notation for the second quantisation is due to the fact that one would like to think of the assignment $(\mathcal{H}_\mathbb{R}, T) \mapsto ((\Gamma_q(\mathcal{H}_\mathbb{R}), \omega), \Gamma_q(T))$ as a functor, the q-Gaussian functor, from pairs of (real) Hilbert spaces and contractions to von Neumann algebras with a given tracial state and unital, trace-preserving, completely positive maps. We will not need these properties and we will not pursue this line of thought further. For details the reader may consult Section 2 of [12] and the references therein.

One of the ideas at the heart of the proof of Theorem 3.24 is to create a sequence with certain properties, modelled on a certain family of functions. To accomplish this, we will need the following result, and then we need to prove that we are allowed to use it.

Theorem 3.22. *An Abelian, diffuse (all generators have diffuse spectral measures), separably generated von Neumann algebra is isomorphic to $L^\infty([0,1],\lambda)$.*

Proof. See Theorem 1.22 in chapter III of [82]. □

Proposition 3.23. *Let e be a unital vector in $\mathcal{H}_\mathbb{R}$, then the element $W(e)$ has a diffuse spectral measure.*

Proof. For details see [14]. We start by investigating the moments of $W(e)$, which are the numbers

$$\omega(W(e)^k) = \left\langle W(e)^k\Omega, \Omega\right\rangle_q,$$

for $k \in \mathbb{Z}_{\geq 0}$. To calculate these moments, we need to know what happens when we apply $W(e)$ to an element of the form $e^{\otimes n}$. To simplify notation, we take $e^{\otimes 0}$ to mean Ω, and $e^{\otimes(-1)} = 0$. Then

$$W(e)e^{\otimes n} = e^{\otimes(n+1)} + \frac{1-q^n}{1-q}e^{\otimes(n-1)},$$

follows straight from the definitions. If we rescale according to

$$E(n) := e^{\otimes n}\prod_{i=0}^{n}\frac{1}{1-q^i},$$

then we obtain the recursion relation

$$W(e)E(n) = (1-q^{n+1})E(n+1) + \frac{1}{1-q}E(n-1).$$

Now notice that all the data that is required to determine the $*$-distribution of $W(e)$ is contained in this recursion relation, together with the fact that the $E(n)$ are orthogonal. Recall that the $*$-distribution in the present case is a functional on the algebra of all polynomials in a single formal indeterminate. To make the jump to spectral theory, we want to think of the operator $W(e)$ as multiplication with the identity function. This suggests that we look for a family of polynomials $p_n(x)$ on $\mathbb{R}$ (as $W(e)$ is self-adjoint) that satisfy

$$xp_n(x) = (1-q^{n+1})p_{n+1}(x) + \frac{1}{1-q}p_{n-1}(x).$$

If we can then find the orthogonalising measure μ for this family, we obtain (via Theorem 2.11) an isomorphism of $W(e)''$ with $L^\infty(\mathbb{R},\mu)$, that sends $W(e)$ to the identity. It then follows that the spectrum of $W(e)$ is the support of μ, and that μ is the spectral measure of $W(e)$. Fortunately, this calculation has been done long ago, and the result can be found in (for instance) [4] as Equation 3.31. The measure μ is diffuse, as it admits a density $\mathrm{d}\mu/\mathrm{d}\lambda$ given by the formula

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = \chi_{\{|x|<2/\sqrt{1-q}\}}(x)\frac{\sqrt{1-q}}{\pi}\sin(\vartheta(x))\prod_{k=1}^{\infty}\left[(1-q^k)\left|1-e^{-2i\vartheta(x)}q^k\right|^2\right],$$

where $\vartheta(x) = \arccos\left(x\sqrt{1-q}/2\right)$. One can check that this indeed reduces to the Wigner semi-circle density when $q=0$ (using $\sin(\arccos(x/2)) = \sqrt{1-(x/2)^2}$). □

We will now state Theorem 3.24, and use it to prove the result claimed at the start of this section (Theorem 3.18).

Theorem 3.24 (Ricard [71])**.** *Let $e \in \mathcal{H}$ be unital, then $W(e)'' \subset \Gamma_q$ is a MASA.*

Proof of Theorem 3.18. According to Ricard's Theorem 3.24, $W(e)''$ is a MASA of Γ_q if $e \in \mathcal{H}$ is unital. It follows that if $x = W(\xi)$ is in the center of Γ_q, it is also in every $W(e)''$. However, this implies that ξ is in every $\mathcal{F}_q(\mathbb{C}e)$. This can only be satisfied if $\xi \in \mathbb{C}\Omega$, say $\xi = \alpha\Omega$. Therefore $x = \alpha I$, as they agree on Ω. We conclude that $\Gamma_q \cap (\Gamma_q)' = \mathbb{C}I$, and hence that the q-Gaußians are factors.

On the other hand, by Proposition 3.19, the q-Gaußians admit a faithful, tracial, and normal state. By Theorem 2.6 in Chapter V of [82], we have that they are of type II_1, and as such II_1 factors. □

The remainder of this section is entirely technical in nature, covering some notation and estimates needed to prove Theorem 3.24, and of course the proof itself.

Definition 3.25. Let $n \in \mathbb{Z}_{\geq 0}$, then define

(i) the q-bracket $[n]_q := (1-q^n)/(1-q)$ ($[0]_q := 0$),

(ii) the q-factorial $[n]_q! := [n]_q \cdots [1]_q$ ($[0]_q! := 1$),

(iii) the q-Pochhammer symbol $(z;q)_\infty := \prod_{i=0}^{\infty}(1-zq^i)$ for $z \in \mathbb{C}$,

(iv) and finally $C_q = 1/(|q|;|q|)_\infty$.

Lemma 3.26.

(i) $(z;q)_\infty$ is well defined and C_q is finite and at least unity.

(ii) If $e \in \mathcal{H}$ is unital, then $\|e^{\otimes n}\|^2_{\mathcal{H}^{\otimes n}} = [n]_q!$.

(iii) For all $k \in \mathbb{Z}_{k\geq 0}$, $0 < [k]_q \leq C_q$.

(iv) For unital vectors $e_1, \ldots, e_n, e$ in $\mathcal{H}$ we have the inequality

$$\left\|e_1 \otimes \cdots \otimes e_n \otimes e^{\otimes m}\right\|_{\mathcal{H}^{\otimes(n+m)}} \leq C_q^{n/2}\sqrt{[m]_q!}.$$

Proof. (i): We have the elementary estimate

$$|(z;q)_\infty| \leq \prod_{i=0}^{\infty}(1+|z||q|^i) \leq \exp\left(|z|\sum_{i=0}^{\infty}|q|^i\right) < \infty,$$

by which $(z;q)_\infty$ exists. We now need to show that $C_q < \infty$ for all $-1 < q < 1$. For this we bound $(|q|,|q|)_\infty > 0$, via

$$(|q|,|q|)_\infty = \prod_{i=1}^{\infty}(1-|q|^i) > (1-|q|)^N \prod_{i=N+1}^{\infty}(1-|q|^i) > 0.$$

The final claim follows from

$$C_q = \prod_{i=1}^{\infty}\frac{1}{1-|q|^i} = \prod_{i=1}^{\infty}\left(1+|q|^i+|q|^{2i}+\ldots\right) \geq 1$$

(ii): By definition,

$$\left\|e^{\otimes n}\right\|^2_{\mathcal{H}^{\otimes n}} = \langle e^{\otimes n}, e^{\otimes n}\rangle_q = \sum_{k=1}^{n} q^{k-1}(e,e)\langle e^{\otimes(n-1)}, e^{\otimes(n-1)}\rangle_q = [n]_q\left\|e^{\otimes(n-1)}\right\|^2_{\mathcal{H}^{\otimes(n-1)}},$$

and since $\langle \cdot, \cdot \rangle_q$ agrees with the original inner product when we reach rank 1, the result follows.

(iii): A small computation shows

$$0 < [k]_q = \frac{1-q^k}{1-q} = 1 + q + \cdots + q^{k-1} \leq 1 + |q| + \cdots + |q|^{k-1} = \frac{1-|q|^k}{1-|q|} \leq \frac{1}{1-|q|} \leq C_q.$$

(iv): This will require the most work. Suppose that we have an operator $R_{n+m,m}$ that factorises $P_q^{(n+m)}$ into its lower rank components, i.e.

$$P_q^{(n+n)} = R_{n+m,m} \left(P_q^{(n)} \otimes P_q^{(m)} \right),$$

then we would have, writing $\xi_n \otimes \eta_m = e_1 \otimes \cdots \otimes e_n \otimes e^{\otimes m}$, that

$$\|\xi_n \otimes \eta_m\|^2_{\mathcal{H}^{\otimes(n+m)}} = \left(\xi_n \otimes \eta_m, R_{n+m,m} \left(P_q^{(n)} \otimes P_q^{(m)} \right) \xi_n \otimes \eta_m \right) \leq \|R_{n+m,m}\| \|\xi_n \otimes \eta_m\|^2_{\mathcal{H}^{\otimes n} \boxtimes \mathcal{H}^{\otimes m}},$$

where the norm of $R_{n+m,m}$ is taken in $B(\mathcal{H}^{\boxtimes(n+m)})$. By iteration and item (ii) this would give

$$\left\| e_1 \otimes \cdots \otimes e_n \otimes e^{\otimes m} \right\|_{\mathcal{H}^{\otimes(n+m)}} \leq \|R_{n+m,m}\|^{n/2} \sqrt{[m]_q!},$$

so it remains to find the operator $R_{n+m,m}$ and show that its (appropriate) norm is less than C_q.

For this, consider the subgroup $S_n \times S_m$ of S_{n+m}. This is not a normal subgroup, but the right cosets $S_n \times S_m / S_{n+m}$ still partition the group, and it is known that each coset admits a unique representative with a minimal number of inversions. Let σ be such a representative, then for any $\pi \in S_{n+m}$, $\sigma(\sigma^{-1}\pi)$ is a factorisation into a coset part and an element of $S_n \times S_m$. It follows that the operator

$$R_{n+m,m} := \sum_{\sigma \in S_{n+m}/(S_n \times S_m)} q^{|\sigma|} U_\sigma,$$

does the job. To bound its operator norm, we begin with the straightforward estimate

$$\|R_{n+m,m}\| \leq \sum_{\sigma \in S_{n+m}/(S_n \times S_m)} |q|^{|\sigma|}.$$

The next step is to use the same concrete representation of the sum as in equation 3.1.3, which yields

$$\|R_{n+m,m}\| \leq \sum_{(I,J) \in P_2^{n,m}(n+m)} |q|^{i(I,J)}.$$

To bound the coefficients appearing in this sum, first assume that $n+m$ is very large so that we never run out of elements in the steps below. Now consider what partitions could give rise to $|q|^0$, in fact there is only one, the first n elements in I_0 and the rest in J_0. We now imagine building partitions with higher powers of $|q|$ from this $|q|^0$ partition. There is only one way to obtain a partition with $|q|^1$, namely swapping the last element of I_0 with the first of J_0. For $|q|^2$, we can either move the last element of I_0 two places into J_0 ($jji = 2$) or the last two elements on place each into J_0 ($jii = 2$). For $|q|^3$ one has the configurations $jjji$, $jiji$, and $jiii$. For $|q|^4$ one has $jjjji$, $jijji$, $jjiij$, $jiiji$, and $jiiii$. One can translate these into partitions of the exponent of $|q|^i$ as follows. For $|q|^3$ our configurations correspond to $jjji = 3$, $jiji = 1+2$, and $jiii = 1+1+1$. For $|q|^4$ one has $jjjji = 4$, $jijji = 1+3$, $jjiij = 2+2$, $jiiji = 1+1+2$, and $jiiii = 1+1+1+1$. The numbers arise by counting the number of j's to the left of each i and adding them up. A little thought shows that this pattern is general. To summarise, we have argued that

$$\|R_{n+m,m}\| \leq \sum_{i=0}^{\infty} p(i) |q|^i,$$

where $p(i)$ is the number of partitions of the natural number i. Thus, the norm we seek is bounded by the generating function of the partition numbers $p(i)$, evaluated in $|q|$. This generating functions is well known from number theory (see for instance Section 1.2 in [3]), and it is also easy to check that the expression below is correct by writing out the product. We conclude that

$$\|R_{n+m,m}\| \leq \sum_{i=0}^{\infty} p(i)|q|^i = \prod_{i=0}^{\infty} \frac{1}{1-|q|^i} = C_q.$$

□

Proof of Theorem 3.24. We follow Ricard's original proof from [71]. Suppose that $x \in W(e)''$, then there is a $\xi \in \mathcal{F}_q(\mathcal{H})$ such that $x = W(\xi)$. If (x_i) is a sequence in the $*$-algebra generated by $W(e)$ that SOT-converges to x, we get that $x_i\Omega$ approaches ξ in norm. It follows that ξ is in fact in $\mathcal{F}_q(\mathbb{C}e)$. Conversely, let $y = W(\eta)$ with η in $\mathcal{F}_q(\mathbb{C}e)$. By second quantisation, there exists a map $\Gamma_q(P_{\mathbb{R}e}) : \Gamma_q(\mathcal{H}_\mathbb{R}) \to \Gamma_q(\mathbb{R}e) = W(e)''$, which implies

$$\Gamma_q(P_{\mathbb{R}e})(y)\Omega = \mathcal{F}_q(P_{\mathbb{R}e})\eta = P_{\mathcal{F}_q(\mathbb{R}e)}\eta = \eta = y\Omega.$$

It follows that $\Gamma_q(P_{\mathbb{R}e})(y) = y$, as Ω is separating, and so that $y \in W(e)''$. To summarise, we have shown that $z = W(\zeta)$ is in $\Gamma_q(\mathbb{R}e)$ if and only if $\zeta \in \mathcal{F}_q(\mathbb{C}e)$.

Now fix an ONB of $\mathcal{H}_\mathbb{R}$, $\{e_i\}_{i\in\mathbb{Z}_{\geq 0}}$ with the property that $e_0 = e$. By definition, $W(e)''$ is a MASA iff $\Gamma_q(\mathcal{H}_\mathbb{R}) \cap W(e)' \subset W(e)''$. Let x be an element in this intersection, then it can be written as $W(\xi)$, and we will show that it is in $W(e)''$ by proving that ξ is in $\mathcal{F}_q(\mathbb{C}e)$.

To this end, suppose that $y = W(\eta)$ for an $\eta \in \mathcal{F}_q(\mathbb{C}e)$. Then by construction, $xy - yx = 0$, so in particular $(W(\xi)W(\eta) - W(\eta)W(\xi))\Omega = 0$. Notice that for any $\alpha, \beta \in \mathcal{F}_q(\mathcal{H})$ we have that

$$W(\alpha)\beta = W(\alpha)W_r(\beta)\Omega = W_r(\beta)W(\alpha)\Omega = W_r(\beta)\alpha. \tag{3.3.1}$$

Hence, it must hold that $(W_r(\eta) - W(\eta))\xi = 0$, and as this must be true for any y, we get that

$$\xi \in \bigcap_{\eta\in\mathcal{F}_q(\mathbb{C}e)} \ker\left(W_r(\eta) - W(\eta)\right).$$

Dualising and using that the $W(\eta)$s are self-adjoint, we want to show instead that

$$\mathcal{F}_q(\mathbb{C}e)^\perp \subset \overline{\text{span}}\left\{\text{im}\left(W_r(\eta) - W(\eta)\right) \mid \eta \in \mathcal{F}_q(\mathbb{C}e)\right\} =: G.$$

Notice that we can describe $\mathcal{F}_q(\mathbb{C}e)^\perp$ as follows. Let F denote the collection of all words $e_{i_1} \otimes \cdots \otimes e_{i_n}$ on the basis elements of $\mathcal{H}_\mathbb{R}$, such that at least one letter is different from $e = e_0$. A brief inspection reveals that then $\mathcal{F}_q(\mathbb{C}e)^\perp$ is the (norm-)closed linear span of F. On the other hand, a geometric version of Hahn-Banach (see for instance Corollary 2.6 in Chapter III of [24]) implies that the norm and weak closures of G coincide. This means that we can reduce the problem to finding, for every word $z = e_{i_1} \otimes \cdots \otimes e_{i_n}$ in F a sequence (z_i) in G that weakly approximates z.

By Proposition 3.23 above, $W(e)''$ is an Abelian, diffuse, and separably generated von Neumann algebra, so by Theorem 3.22 we can identify it with $L^\infty([0,1],\lambda)$. Consider now the Rademacher functions (r_i), which live in $W(e)''$ by this identification. These satisfy the properties that they are self-adjoint, square to unity, and the sequence (r_i) is an (incomplete) ONS in $L^2([0,1],\lambda)$ (see for instance [76], Lemma 24.21). It follows from this last point that the sequence of Rademacher functions converges weakly to 0 in $W(e)''$. We thus obtain a sequence of vectors (η_i) in $\mathcal{F}_q(\mathbb{C}e)$, such that $W(\eta_i) = r_i$. As $\mathcal{F}_q(\mathcal{H})$ is a Hilbert space, its weak topology is (by the Riesz representation theorem) determined by the family of semi-norms $|\langle\cdot,\zeta\rangle|$, $\zeta \in \mathcal{F}_q(\mathcal{H})$. It holds that

$$|\langle\eta_i,\zeta\rangle| = |\langle r_i\Omega,\zeta\rangle| \overset{i\to\infty}{\longrightarrow} 0,$$

and we conclude that also the sequence (η_i) converges weakly to 0 (in $\mathcal{F}_q(\mathcal{H})$).

This suggests that we take a closer look at the sequence

$$z_i = (W(\eta_i) - W_r(\eta_i))\,(W(\eta_i)(z)) = z - W_r(\eta_i)W(\eta_i)(z),$$

which is contained in G. The remainder of the proof will be dedicated to showing that the difference $y_i := z - z_i = W_r(\eta_i)W(\eta_i)(z)$ weakly converges to 0. As a first step, we have

$$\left\|W_{(r)}(\eta_i)\right\|^2 = \left\|W_{(r)}(\eta_i)W^*_{(r)}(\eta_i)\right\| = \left\|W_{(r)}(\eta_i)^2\right\| = 1,$$

so that

$$\|y_i\| = \|W_r(\eta_i)W(\eta_i)(z)\| \leq \|z\|.$$

This means that the sequence (y_i) is norm-bounded, and hence that it suffices to prove that $\langle y_i, t\rangle$ converges to 0 for all $t = e_{j_1} \otimes \cdots \otimes e_{j_k} \in \mathcal{F}_q(\mathcal{H})$ (2ε argument). By the identity (3.3.1) above, we can write

$$\langle y_i, t\rangle = \langle W(\eta_i)(z), W_r(\eta_i)(t)\rangle = \langle W_r(z)(\eta_i), W(t)(\eta_i)\rangle.$$

As z and t are both words on the basis $\{e_i\}$, we can use the Wick formula (3.1.3) to expand the operators $W_r(z)$ and $W(t)$. As the Wick formula involves only finitely many terms, we only need to control terms of the form

$$w^i_{m,\ell} = \left\langle a^*_{r,i_1} \cdots a^*_{r,i_m} a_{r,i_{m+1}} \cdots a_{r,i_n}(\eta_i), a^*_{j_1} \cdots a^*_{j_\ell} a_{j_{\ell+1}} \cdots a_{j_k}(\eta_i)\right\rangle,$$

where at least one of the i_j is non-zero.

We must have $i_{m+1} = \cdots = i_n = 0$ (and $j_{\ell+1} = \cdots = j_k = 0$), otherwise $a_{r,i_{m+1}} \cdots a_{r,i_n}(\eta_i)$ vanishes (as $\eta_i \in \mathcal{F}_q(\mathbb{C}e)$). Recall from the proof of Lemma 3.26 part (ii) that

$$a(e)e^{\otimes s} = [s]_q e^{\otimes(s-1)},\ a_r(e)e^{\otimes s} = [s]_q e^{\otimes(s-1)}.$$

As the η_i live in $\mathcal{F}_q(\mathbb{C}e)$, we can write

$$\eta_i = \sum_{s\in\mathbb{Z}_{\geq 0}} h^i_s e^{\otimes s},$$

with the coefficients h^i_s all real, as the r_i are self-adjoint. Combining these two pieces of information yields

$$a(e)^{n-m}e^{\otimes s} = \frac{[s]_q!}{[s+m-n]_q!}e^{\otimes(s+m-n)},\ a_r(e)e^{\otimes s} = [s]_q e^{\otimes(s-1)},$$

and

$$w^i_{m,\ell} = \sum_{s\geq n-m}\sum_{t\geq k-\ell} h^i_s h^i_t \frac{[s]_q!}{[s+m-n]_q!}\frac{[t]_q!}{[t+\ell-k]_q!}\left\langle a^*_{r,i_1}\cdots a^*_{r,i_m}e^{\otimes(s+m-n)}, a^*_{j_1}\cdots a^*_{j_\ell}e^{\otimes(t+\ell-k)}\right\rangle.$$

For notational simplicity, we will use the convention that $h^i_s = 0$ if $s < 0$. Now we change our summation indices to $s' = s - n + 2m$ and $t' = t - k + 2\ell$, this results in

$$w^i_{m,\ell} = \sum_{s'\geq m}\sum_{t'\geq \ell} h^i_{s'+n-2m}h^i_{t'+k-2\ell}\frac{[s'+n-2m]_q!}{[s'-m]_q!}\frac{[t'+k-2\ell]_q!}{[t'-\ell]_q!}$$
$$\times\left\langle a^*_{r,i_1}\cdots a^*_{r,i_m}e^{\otimes(s'-m)}, a^*_{j_1}\cdots a^*_{j_\ell}e^{\otimes(t'-\ell)}\right\rangle.$$

Comparing the ranks of the two tensors in the inner product, we realize that only terms with $s' = t'$ are non-zero. Additionally, let $v \leq m$ be the first position such that the index $i_v \neq 0$. With this, and dropping the prime on s again, we arrive at

$$w^i_{m,\ell} = \sum_{s \geq m,\ell} h^i_{s+n-2m} h^i_{s+k-2\ell} \frac{[s+n-2m]_q!}{[s-m]_q!} \frac{[s+k-2\ell]_q!}{[s-\ell]_q!} \\ \times \left\langle a^*_{r,i_{v+1}} \cdots a^*_{r,i_m} e^{\otimes(s-m)}, a_{r,i_v} \cdots a_{r,i_1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right) \right\rangle \tag{3.3.2}$$

We investigate the right argument in the inner product. As $i_1 = \cdots = i_{v-1} = 0$, we should compute

$$a_r(e)^{v-1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right).$$

Using the definition of the annihilation operator, this is

$$\sum_{\alpha_{v-1}=1}^{s-v+2} \cdots \sum_{\alpha_1=1}^{s} \Delta(\alpha_1, \ldots, \alpha_{v-1})\ q^{\alpha_1 - 1} \cdots q^{\alpha_{v-1}-1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right)_{(\alpha_1,\ldots,\alpha_{v-1})},$$

where $(\dots)_{(\alpha_1,\ldots,\alpha_{v-1})}$ indicates that the α_1-st letter from the right has been removed from the word, the α_2-nd letter from the remaining word, and so on, and $\Delta(\dots)$ is 0 unless all the letters removed were e's, in which case it is unity. If we now act on this result with the operator $a_{r,v}$, the result is only non-zero if at least one of the $e_{j_1}, \ldots, e_{j_\ell}$ is different from e. Consequently, this can result in at most ℓ terms appearing. The factors of q in those terms are all majorised by the $(s-\ell-v+1)$-st power of q, as $(\dots)_{(\alpha_1,\ldots,\alpha_{v-1})}$ has length $(s-v+1)$ and ends with at least $(s-\ell-v+1)$ consecutive e's. Notice the we can use part (iv) of Lemma 3.26 above to estimate the norm of these words (together with point (iii) of the same lemma). In particular,

$$\left\| a_{r,i_v} \cdots a_{r,i_1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right) \right\| \leq \sum_{\alpha_{v-1}=1}^{s-v+2} \cdots \sum_{\alpha_1=1}^{s} \ell |q|^{(\sum \alpha_i)-v+1} |q|^{s-\ell-v+1} C_q^{\ell/2} \sqrt{[s-\ell-v+1]_q!}.$$

The sums over the α_i can now be carried out, and we can estimate $|q|^{s-\ell-v+1}$ by $|q|^s$ (as $v \geq 1$). Carrying this out, we end up with the estimate

$$\ell [s]_{|q|} \cdots [s-v+2]_{|q|} C_q^{\ell/2} |q|^s \sqrt{[s-\ell-v+1]_q!} \leq \left(\ell C_q^{\ell/2+v-1} \right) |q|^s \sqrt{[s-\ell-v+1]_q!},$$

where we used part (iii) of Lemma 3.26 again. As s is the only variable index (everything else is fixed per term), we will plainly write C for any s-independent constant occurring in our estimates. With this notation, we finally arrive at

$$\left\| a_{r,i_v} \cdots a_{r,i_1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right) \right\| \leq C |q|^s \sqrt{[s-\ell-v+1]_q!}$$

We return to (3.3.2) to finish the proof. As the η_i tend weakly to 0, all their coefficients h^i_s must also tend to zero. It follows that we only need to control the tail of the sum over s. We introduce some number N, taken to be large and focus on the sum over $s \geq N$. We need an estimate on the coefficients h^i_s for this, but it follows from $\|\eta_i\| \leq 1$ and point (ii) of Lemma 3.26 that $|h^i_s| \leq 1/\sqrt{[s]_q!}$. Using this and Cauchy-Schwartz, we can estimate the tail T_N by

$$T_N \leq \sum_{s \geq N} \frac{\sqrt{[s+n-2m]_q! [s+k-2\ell]_q!}}{[s-m]_q! [s-\ell]_q!} \left\| a^*_{r,i_{v+1}} \cdots a^*_{r,i_m} e^{\otimes(s-m)} \right\| \left\| a_{r,i_v} \cdots a_{r,i_1} \left(e_{j_1} \otimes \cdots \otimes e_{j_\ell} \otimes e^{\otimes(s-\ell)} \right) \right\|.$$

We can use (iv) of Lemma 3.26 again to estimate the first norm, and we derived an estimate for the second norm above. Plugging these in, we find the bound

$$T_N \leq \sum_{s \geq N} \frac{\sqrt{[s+n-2m]_q![s+k-2\ell]_q!}\sqrt{[s-\ell-v+1]_q!}\sqrt{[s-m]_q!}}{[s-m]_q![s-\ell]_q!} C|q|^s$$

The final observation is now that the sequence $[s-\ldots]_q!$ behaves (for large s) always like a geometric sequence with ratio $(1-q)^{-1}$. By powercounting, it follows that the large fraction in our bound converges for s heading to infinity, so in particular it must be bounded uniformly in s. This implies that we in fact have the estimate

$$T_N \leq \sum_{s \geq N} C|q|^s = |q|^N C \sum_{s=0}^{\infty} |q|^s \leq C|q|^N.$$

We conclude that, for any given $\varepsilon > 0$, there exists N_ε such that

$$\limsup_{i \to \infty} \left| w^i_{m,\ell} \right| \leq C|q|^{N_\varepsilon} < \varepsilon,$$

and this finishes the proof. □

4 Free Probability Theory II

4.1 Noncommutative Derivatives

Before we can continue on to describing free monotone transport, we need some further ideas and techniques from free probability that are a little more advanced than what we introduced in Chapter 2. In particular, we have yet to introduce analogues of the concepts discussed in Section 1.2. Those are the ideas of Entropy, score functions, and Fisher information, the former two being the more important ones for our purposes. Immediately, one runs into some obstacles, as for both of these ideas we exploited the classical probability concept of a density. While in the case that we work with random variables that have an analytic distribution, we have a probability measure available, it need not admit a density (for instance if the random variable is a p-Haar unitary). Additionally, we saw that for fixed covariances, the entropy is maximised by a Gaußian distribution, but our framework cannot contain a random variable that has a Gaußian as its analytic distribution, as such an element would have to be unbounded (because the support of a Gaußian is the entire real line). Finally, the definition of the score functions (and thus that of the Fisher information) includes taking derivatives, which we have also not discussed in a noncommutative context. It is the goal of the present chapter (based on Chapters 7 and 8 of [56]) to remedy the situation.

Throughout this chapter, we will work in a tracial W*-probability space (M, τ), thus we assume that the state τ on the von Neumann algebra M is a faithful, normal trace. Recall that the score function satisfied $\mathbb{E}[\xi f] = \mathbb{E}[f']$ for f a nice enough function of the underlying random variable. This suggests that we should find a way to make sense of derivatives of functions of our noncommutative random variables. A natural place to start is the case that these functions are restricted to be noncommutative polynomials, as we have already employed these with some success to define concepts in noncommutative probability (such as the distribution of a random variable).

Definition 4.1. The *partial noncommutative derivatives* are the linear maps

$$\partial_i : \mathbb{C}\langle X_1, \ldots, X_n\rangle \to \mathbb{C}\langle X_1, \ldots, X_n\rangle \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle,$$

where $\mathbb{C}\langle X_1, \ldots, X_n\rangle$ is once again the algebra of noncommutative polynomials in the formal indeterminates $X_1, \ldots, X_n$, and i runs between 1 and n, that are defined by the relations

$$\partial_i 1 = 0,\ \partial_i X_j = \delta_{ij} 1 \otimes 1,\ \partial_i (P_1 P_2) = (\partial_i P_1) \cdot (1 \otimes P_2) + (P_1 \otimes 1) \cdot (\partial_i P_2),$$

with $P_1, P_2 \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$.

Remark 4.2. The above definition indeed defines a unique family of maps, as on monomials they are given by

$$\partial_i \left(X_{j(1)} \cdots X_{j(m)}\right) = \sum_{k=1}^{m} \delta_{i,j(k)} X_{j(1)} \cdots X_{j(k-1)} \otimes X_{j(k+1)} \cdots X_{j(m)}.$$

This also represents a useful strategy for proving many elementary properties of the ∂_i, by checking that it is true for all monomials. Notice that the third property of the ∂_i is precisely the Leibniz rule, so that they are derivations. Another convenient way to represent that action of the δ_i on a monomial q is to write

$$\partial_i q = \sum_{q = A X_i B} A \otimes B.$$

As when we discussed moments in Chapter 2, these formal noncommutative polynomials give elements of the von Neumann algebra when we replace the formal indeterminate X_1 by some $x_1 \in M$ and so on. Thus the partial derivative with respect to x_i of the element $P(x_1, \ldots, x_n)$ is given by subjecting $\partial_i P$ to this replacement procedure, and is an element of $M \otimes M$. In formulas, $\partial_i(P(x_1, \ldots, x_n)) = (\partial_i P)(x_1, \ldots, x_n)$. We will discuss the details of this procedure with more rigour later on in this section. Three questions may immediately come to mind regarding the definition we just gave. First and foremost, what does this reduce to in the commutative case when $n = 1$? Second, to what extend do these partial derivatives 'commute', as in the classical case? Finally, how big is the kernel of these operators, in particular, in what capacity is it still true that only constant polynomials are sent to zero? The first and second questions are easy but instructive to answer, and feature the first instances of the proof techniques mentioned in the remark above.

Proposition 4.3. *Let $n = 1$ and notice that $\mathbb{C}\langle X_1\rangle \otimes \mathbb{C}\langle X_1\rangle \simeq \mathbb{C}[X, Y]$ via the identification $X = X_1 \otimes 1$, $Y = 1 \otimes X_1$, where $\mathbb{C}[X, Y]$ is the algebra of polynomials in two commuting formal indeterminates X and Y. Write ∂ for ∂_1 composed with this identification. Then for any $P \in \mathbb{C}\langle X_1\rangle$, ∂ is the* free difference quotient *given by*

$$(\partial P)(X, Y) = \frac{P(X) - P(Y)}{X - Y}.$$

Proof. It suffices to show that the advertised relation holds on monomials, which are all of the form X_1^m for some natural number m. Thus

$$\begin{aligned} \partial X_1^m &= 1 \otimes X_1^{m-1} + X_1 \otimes X_1^{m-2} + \cdots + X_1^{m-1} \otimes 1, \\ &= Y^{m-1} + XY^{m-2} + \cdots + X^{m-1}, \end{aligned}$$

and the realisation that

$$(X - Y)(Y^{m-1} + XY^{m-2} + \cdots + X^{m-1}) = X^m - Y^m$$

finishes the proof. □

Proposition 4.4. *The ∂_i are 'self-co-associative' in the sense that $(1 \otimes \partial_i)\partial_i = (\partial_i \otimes 1)\partial_i$. If two differing indices are involved, we have instead that $(1 \otimes \partial_i)\partial_j = (\partial_j \otimes 1)\partial_i$.*

Proof. Again it suffices to check these identities on monomials q. This is now an easy check using a formula from the remark above. First,

$$(\partial_i \otimes 1)\partial_i q = (\partial_i \otimes 1) \sum_{q = AX_iB} A \otimes B = \sum_{q = AX_iBX_iC} A \otimes B \otimes C = (1 \otimes \partial_i) \sum_{AX_iB} A \otimes B = (1 \otimes \partial_i)\partial_i q.$$

Second,

$$(1 \otimes \partial_i)\partial_j q = \sum_{AX_jBX_iC} A \otimes B \otimes C = (\partial_j \otimes 1) \sum_{q = AX_iB} A \otimes B = (\partial_j \otimes 1)\partial_i q.$$

□

Our final question will require more effort to answer. It is certainly conceivable that the kernels of the ∂_i are bigger than in the classical case due to there being algebraic relations between the $x_1, \ldots, x_n$ on which they are evaluated. However, in the following special case, things behave entirely 'classically'.

Proposition 4.5. *Let $x_1, \ldots, x_n$ be self-adjoint elements of M and $P = P^* \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$. If all the partial derivatives $(\partial_i P)(x_1, \ldots, x_n)$ vanish, then $P(x_1, \ldots, x_n) =: p$ is equal to the scalar multiple of the identity $\tau(p)I$.*

Proof. Write for any $x \in M$ that $\tau(x^*x)^{1/2} = \|x\|_2$, it is well known that the faithfulness of τ implies that this is indeed a norm (we will have more to say about this later). Then the proposition is an immediate consequence of the following inequality,

$$\|p - \tau(p)I\|_2 \leq \left[\sqrt{2} \max_{1\leq i\leq n} \|x_i\|_2\right] \sum_{j=1}^{n} \|\partial_i p\|_2, \tag{4.1.1}$$

which is in some sense a noncummutative version of the well-known Poincaré inequality. By centring p if necessary, we may assume that $\tau(p) = 0$. As a consequence of the assumptions, p is self-adjoint. Thus

$$\begin{aligned}
\tau(p^2) &= \frac{1}{2}\left(\tau(p^2) + \tau(p^2) - 2\tau(p)^2\right), \\
&= \frac{1}{2}\left(\tau \otimes \tau\right)\left(p^2 \otimes I + I \otimes p^2 - 2p \otimes p\right), \\
&= \frac{1}{2}\left(\tau \otimes \tau\right)\left((p \otimes I - I \otimes p)^2\right),
\end{aligned}$$

which implies that $\sqrt{2}\|p\|_2 = \|p \otimes I - I \otimes p\|_2$. However, as one may show by considering monomials,

$$P \otimes 1 - 1 \otimes P = \sum_{i=1}^{n} \left[(\partial_i P) \cdot (X_i \otimes 1) - (1 \otimes X_j) \cdot (\partial_i P)\right].$$

We can use this to obtain the estimate

$$\|p \otimes I - I \otimes p\|_2 \leq \sum_{j=1}^{n} \|\partial_i p\|_2 \left[\|(x_i \otimes I)\|_2 + \|(I \otimes x_i)\|_2\right].$$

It is an easy check that $\|(x_i \otimes I)\|_2 + \|(I \otimes x_i)\|_2 = 2\|x_i\|_2$, and estimating these norms by their maximum gives (4.1.1). □

We now describe more formally how to view the ∂_i as acting on M. Let $x_1, \ldots, x_n$ again be self-adjoint elements of M, then the unital $*$-subalgebra of M they generate is $\langle x_1, \ldots, x_n\rangle$. The procedure of 'plugging in' the x_i in place of the X_i in some $P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ is captured formally in the concept of an *evaluation map* $\varepsilon : \mathbb{C}\langle X_1, \ldots, X_n\rangle \to \langle x_1, \ldots, x_n\rangle \subset M$, which is the unique ($*$-)algebra morphism defined by $\varepsilon(X_i) = x_i$. If this map turns out to be an algebra isomorphism, we say that the x_i are *algebraically independent*, that is to say that there exists no noncommutative polynomial P such that $\varepsilon(P) = 0$. In this case, we can explicitly define (with some abuse of notation) that ∂_i acts on $\langle x_1, \ldots, x_n\rangle$ directly by $(\varepsilon \otimes \varepsilon) \circ \partial_i \circ (\varepsilon)^{-1}$, which of course then lands in $\langle x_1, \ldots, x_n\rangle \otimes \langle x_1, \ldots, x_n\rangle$. As we saw in the proof of Proposition 4.5, the ∂_i interact well with $\|\cdot\|_2$, and this motivates the following definition.

Definition 4.6. Write $\|\cdot\|_2 := \tau((\cdot)^*\cdot)^{1/2}$, which is indeed a norm on M, and write $L^2(M)$ for the closure of M in this norm. Then this is a Hilbert space with inner product $\langle x, y\rangle = \tau(y^*x)$ (compare with the GNS representation). Let $x_1, \ldots, x_n$ be self-adjoint elements of M and denote

$$L^2(x_1, \ldots, x_n) := \overline{\langle x_1, \ldots, x_n\rangle}^{\|\cdot\|_2} \subset L^2(M).$$

In the case that the x_i are algebraically independent, we can view the ∂_i as unbounded operators

$$\partial_i : L^2(x_1, \ldots, x_n) \supset D(\partial_i) = \langle x_1, \ldots, x_n\rangle \to L^2(x_1, \ldots, x_n) \otimes L^2(x_1, \ldots, x_n).$$

The canonical question to ask about an unbounded operator on a Hilbert space, is whether or not it is closable. Since ∂_i is by definition densely defined, it is enough to show that also ∂_i^* is densely defined. Remarkably, the algebraic structure of the partial derivative operators forces this to be true as soon as the domains of the adjoint operators contain a certain element.

Theorem 4.7. *Suppose that $x_1, \ldots, x_n$ are algebraically independent self-adjoint elements of M. If $(I \otimes I) \in D(\partial_i^*)$, then in fact all of $\langle x_1, \ldots, x_n \rangle \otimes \langle x_1, \ldots, x_n \rangle \subset D(\partial_i^*)$. This implies that ∂_i^* is densely defined, and thus ∂_i is closable. Moreover, for elementary tensors $p \otimes q$ in $D(\partial_i^*)$ we have the formula*

$$\partial_i^*(p \otimes q) = p\left[\partial_i^*(I \otimes I)\right] q - p\left[(\tau \otimes \mathrm{id})\partial_i q\right] - \left[(\mathrm{id} \otimes \tau)\partial_i p\right] q. \tag{4.1.2}$$

Proof. The proof boils down to checking that the given expression for ∂_i^* on elementary tensors witnesses the definition of an adjoint operator. This is sufficient, because the right hand side of (4.1.2) is well-defined by the assumption that $I \otimes I \in D(\partial_i^*)$. For this, one first proves by considering monomials that

$$(\tau \otimes \mathrm{id})\left[(\partial_i p^*)^*\right] = (\mathrm{id} \otimes \tau)\left[\partial_i p\right] \tag{4.1.3}$$

for all $p \in \langle x_1, \ldots, x_n \rangle$. Then to conclude the proof, one shows that for all $r \in \langle x_1, \ldots, x_n \rangle$ the above formula is such that

$$\langle \partial_i^*(p \otimes q), r \rangle = \langle p \otimes q, \partial_i r \rangle.$$

One way to do this is to proceed as follows,

$$\begin{aligned}
\langle p\left[\partial_i^*(I \otimes I)\right] q, r \rangle &= \tau\left(r^* p\left[\partial_i^*(I \otimes I)\right] q\right), \\
&= \tau\left(q r^* p[\partial_i^*(I \otimes I)]\right), \\
&= \langle \partial_i^*(I \otimes I), p^* r q^* \rangle, \\
&= \langle I \otimes I, \partial_i(p^* r q^*) \rangle, \\
&= \langle I \otimes I, (\partial_i p^*) \cdot (I \otimes r q^*) + (p^* \otimes I) \cdot (\partial_i r) \cdot (I \otimes q^*) + (p^* r \otimes I) \cdot (\partial_i q^*) \rangle.
\end{aligned}$$

The middle term is

$$\begin{aligned}
\langle I \otimes I, (p^* \otimes I) \cdot (\partial_i r) \cdot (I \otimes q^*) \rangle &= (\tau \otimes \tau)\left((I \otimes q) \cdot (\partial_i r)^* \cdot (p \otimes I)\right), \\
&= (\tau \otimes \tau)\left((p \otimes q) \cdot (\partial_i r)^*\right), \\
&= \langle p \otimes q, \partial_i r \rangle,
\end{aligned}$$

by the trace property. The first term gives

$$\begin{aligned}
\langle I \otimes I, (\partial_i p^*) \cdot (I \otimes r q^*) &= (\tau \otimes \tau)\big((I \otimes q r^*) \cdot (\partial_i p^*)^*\big), \\
&= \tau\big(q r^* \left[(\tau \otimes \mathrm{id})(\partial_i p^*)^*\right]\big), \\
&= \tau\big(q r^* \left[(\mathrm{id} \otimes \tau)(\partial_i p)\right]\big), \\
&= \tau\big(r^* \left[(\mathrm{id} \otimes \tau)(\partial_i p)\right] q\big), \\
&= \langle \left[(\mathrm{id} \otimes \tau)\partial_i p\right] q, r \rangle,
\end{aligned}$$

where we used (4.1.3) in the third line, and the trace property in the fourth. Similarly, the last term gives

$$\langle I \otimes I, (p^* r \otimes I) \cdot (\partial_i q^*) \rangle = \langle p\left[(\tau \otimes \mathrm{id})\partial_i q\right], r \rangle.$$

Comparing with (4.1.2) finishes the proof. □

Clearly, the objects $\partial_i^*(I \otimes I)$ are of some importance. In the next section we will explore their interpretation in free probability and use this to improve the preceding theorem. Morally, we will show that if one can make sense of the $\partial_i^*(I \otimes I)$, then this immediately forces $x_1, \ldots, x_n$ to be algebraically independent, meaning that we can drop this assumption.

4.2 Conjugate Variables

Recall from Section 1.2 that at least formally, one could characterise the score functions ξ_i by writing $\partial_i^* 1$, for the ordinary partial derivatives. This is very similar to the objects $\partial_i^*(I \otimes I)$ we saw played a central role in proving when the noncommutative partial derivatives ∂_i could be closed. The score functions also satisfied $\mathbb{E}[\xi f] = \mathbb{E}[f']$, and this we can now make noncommutative.

Definition 4.8. Let $x_1, \ldots, x_n$ be self-adjoint elements of M. We extend τ to $L^2(M)$ by writing $\tau(\eta) := \langle \eta, I \rangle$ for any $\eta \in L^2(M)$. Now, we say that $\xi_1, \ldots, \xi_n \in L^2(M)$ satisfy the *conjugate relations* for $x_1, \ldots, x_n$, iff for all noncommutative polynomials P in n indeterminates and all $1 \leq i \leq n$ we have

$$\tau\left(\xi_i P(x_1, \ldots, x_n)\right) = (\tau \otimes \tau)\left((\partial_i P)(x_1, \ldots, x_n)\right). \tag{4.2.1}$$

If the ξ_i satisfy this, and moreover are elements of the smaller space $L^2(x_1, \ldots, x_n)$, we call them a *conjugate system* to the x_i, and the individual ξ_i are called *conjugate variables*.

Remark 4.9. One can also phrase the conjugate relations more concretely by demanding that for any monomial $x_{j(1)} \cdots x_{j(m)}$ it holds that

$$\tau\left(\xi_i x_{j(1)} \cdots x_{j(m)}\right) = \sum_{k=1}^{m} \delta_{i,j(k)} \tau\left(x_{j(1)} \cdots x_{j(k-1)}\right) \tau\left(x_{j(k+1)} \cdots x_{j(m)}\right).$$

Theorem 4.10.

(i) If a conjugate system exists, it is unique.

(ii) If there exist elements in $L^2(M)$ witnessing the conjugate relations, then a conjugate system exists.

(iii) The elements of a conjugate system are self-adjoint with respect to the isometric extension S of the involution to $L^2(M)$.

(iv) If $x_1, \ldots, x_n$ are self-adjoint elements of M admitting a conjugate system, then they are algebraically independent, thus conjugate variables can always be written as $\partial_i^*(I \otimes I)$ if they exist.

Proof. The first three points are trivial. For part (i), recall that $\langle x_1, \ldots, x_n \rangle$ is by definition dense in $L^2(x_1, \ldots, x_n)$, and notice that the conjugate relations fix the inner products of the ξ_i with this subspace. For part (ii), use the orthogonal projection from $L^2(M)$ onto $L^2(x_1, \ldots, x_n)$. For part (iii), look back at the proofs of items (ii) and (iii) of Lemma 3.10. There we showed that for a tracial state, one can isometrically extend the involution operation to $M.I$, and thus to $L^2(M)$. Moreover we saw that $\langle S\xi, \eta \rangle = \overline{\langle \xi, S\eta \rangle}$ for all $\xi, \eta \in L^2(M)$. One can check on monomials that $\langle \xi_i, p \rangle = \overline{\langle \xi_i, p^* \rangle}$ for all $p \in \langle x_1, \ldots, x_n \rangle$, and of course $Sp = p^*$. From this we conclude that $S\xi_i = \xi_i$ for all conjugate variables.

For part (iv), let ε denote the evaluation map induced by the x_i. We first show that if $P \in \mathbb{C}\langle X_1, \ldots, X_n \rangle$, then $\varepsilon P = 0$ implies that $\varepsilon \partial_i P = 0$ (shorthand for $(\varepsilon \otimes \varepsilon)\partial_i P$) for all i. It is enough to show that the inner product of $\varepsilon \partial_i P$ with any elementary tensor $\varepsilon Q_1 \otimes \varepsilon Q_2$ vanishes, for $Q_1, Q_2 \in \mathbb{C}\langle X_1, \ldots, X_n \rangle$. Using the Leibniz rule and the fact that $\varepsilon P = 0$,

$$\varepsilon \partial_i (Q_1 P Q_2) = (\varepsilon Q_1 \otimes I) \cdot (\varepsilon \partial_i P) \cdot (I \otimes \varepsilon Q_2).$$

We can now exploit the existence of a conjugate system to write

$$0 = \tau\left[\xi_i \varepsilon\left(Q_1 P Q_2\right)\right] = (\tau \otimes \tau)\left[(\varepsilon Q_1 \otimes I) \cdot (\varepsilon \partial_i P) \cdot (I \otimes \varepsilon Q_2)\right] = (\tau \otimes \tau)\left[(\varepsilon \partial_i P) \cdot (\varepsilon Q_1 \otimes \varepsilon Q_2)\right],$$

where the last step follows from the trace propperty. We conclude that $\varepsilon\partial_i P = 0$, thus any algebraic relation between the x_i implies an entire family of relations in the tensor product. In particular, we can combine this with the operator $\mathrm{id} \otimes \tau$ to get relations within the algebra itself, where we slightly abuse notation by saying that τ acts on noncommutative polynomials directly (while we should write $\tau \circ \varepsilon$). Thus we have that $\varepsilon \circ (\mathrm{id} \otimes \tau) \circ \partial_i P = 0$. Notice that the operator $(\mathrm{id} \otimes \tau) \circ \partial_i$, when acting on a monomial of degree m, produces only terms that have a strictly lower degree, and only produces a term of precisely one degree lower if the final factor of the monomial happens to be X_i. Hence

$$\varepsilon \left[(\mathrm{id} \otimes \tau) \circ \partial_{i(1)} \circ \cdots \circ (\mathrm{id} \otimes \tau) \circ \partial_{i(m)} \right] \left(\alpha X_{i(1)} \cdots X_{i(m)} \right) = \alpha,$$

and this operator applied to any other monomial of equal or lesser degree yields zero. Now, we can use this technique to isolate all the coefficients of the highest order terms in P, but we also have that applying any operator $(\mathrm{id} \otimes \tau) \circ \partial_i$ to P and then applying ε gives zero. Thus, all the coefficients of the highest order terms in P must vanish, and so P can only be the zero polynomial, and we can conclude that the x_i are algebraically independent. □

In preparation for computing the conjugate variables to a semicircular family, we restate the conjugate relations in terms of cumulants.

Proposition 4.11. *Consider an n-tuple $x_1 \ldots, x_n \in M$ such that each x_i is self-adjoint, and $\xi_1, \ldots, \xi_n$ are elements of $L^2(M)$. Then the following are equivalent:*

(i) The ξ_i satisfy the conjugate relations for the x_j

$$\tau\left(\xi_i x_{j(1)} \cdots x_{j(m)}\right) = \sum_{k=1}^{m} \delta_{i,j(k)} \tau\left(x_{j(1)} \cdots x_{j(k-1)}\right) \tau\left(x_{j(k+1)} \cdots x_{j(m)}\right),$$

for all $1 \leq i \leq n$, $1 \leq j(1), \ldots, j(m) \leq n$ and $m \geq 1$.

(ii) For all $1 \leq i \leq n$, $1 \leq j(1), \ldots, j(\ell) \leq n$ and $\ell \geq 1$, the ξ satisfy the equations

$$\kappa_1(\xi) = 0,\ \kappa_2(\xi_i, x_{j(1)}) = \delta_{i,j(1)},\ \kappa_{\ell+1}(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}) = 0 \text{ (when } \ell \geq 2). \tag{4.2.2}$$

Proof. Before we begin we should remark how we actually define cumulants where precisely one entry comes from $L^2(M)$. Recall that for $\eta \in L^2(M)$, we had set $\tau(\eta) = \langle \eta, 1 \rangle$ and this was well-defined. From the definition of cumulants, it follows that any cumulant is eventually just a product of terms that are τ applied to products of the entries. However, only one of these will contain the one entry that comes from $L^2(M)$, and since there is precisely one of those, everything is well-defined.

We first show that (i) implies (ii), and we do this by strong induction on ℓ. The cases $\ell = 0, 1$ are immediate. Now assume that we have shown that the cumulant relations are satisfied for all numbers less than $\ell + 1$. We know that

$$\tau\left(\xi_i x_{j(1)} \cdots x_{j(\ell)}\right) = \sum_{\pi \in \mathrm{NC}(\ell+1)} \kappa_\pi(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}),$$

and by the induction hypothesis, ξ_i can only occur in some κ_2 or in $\kappa_{\ell+1}$. Thus

$$\tau\left(\xi_i x_{j(1)} \cdots x_{j(\ell)}\right) = \kappa_{\ell+1}(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}) + \sum_{\pi = (1,k) \cup \pi_1 \cup \pi_2} \kappa_\pi(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}),$$

where $\pi = (1,k) \cup \pi_1 \cup \pi_2$ means that 1 (which corresponds to ξ_i) is paired with k (corresponding to $x_{j(k)}$), and as we are considering non-crossing partitions, π_1 must be a non-crossing partition of $\{x_{j(1)}, \ldots, x_{j(k-1)}\}$, and π_2 must be a non-crossing partition of $\{x_{j(k+1)}, \ldots, x_{j(\ell)}\}$. This implies that

$$\begin{aligned}
\sum_{\pi=(1,k)\cup\pi_1\cup\pi_2} \kappa_\pi(\ldots) &= \sum_{\pi=(1,k)\cup\pi_1\cup\pi_2} \delta_{i,j(k)}\ \kappa_{\pi_1}(x_{j(1)}, \ldots, x_{j(k-1)})\kappa_{\pi_2}(x_{j(k+1)}, \ldots, x_{j(\ell)}), \\
&= \sum_k \delta_{i,j(k)} \left(\sum_{\pi_1 \in \mathrm{NC}(k-1)} \kappa_{\pi_1}(x_{j(1)}, \ldots, x_{j(k-1)}) \right) \left(\sum_{\pi_2 \in \mathrm{NC}(\ell-k)} \kappa_{\pi_2}(x_{j(k+1)}, \ldots, x_{j(\ell)}) \right), \\
&= \sum_k \delta_{i,j(k)} \tau(x_{j(1)}, \ldots, x_{j(k-1)})\tau(x_{j(k+1)}, \ldots, x_{j(\ell)}), \\
&= \tau\left(\xi_i x_{j(1)} \cdots x_{j(\ell)}\right),
\end{aligned}$$

where in the last two lines we used the definition of cumulants and the conjugate relations. Plugging this back in, we find that

$$\tau\left(\xi_i x_{j(1)} \cdots x_{j(\ell)}\right) = \kappa_{\ell+1}(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}) + \tau\left(\xi_i x_{j(1)} \cdots x_{j(\ell)}\right),$$

so that indeed

$$\kappa_{\ell+1}(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}) = 0.$$

We now prove the converse, that (ii) implies (i). Consider the moment-cumulant formula for $\tau\left(\xi_i x_{j(1)} \cdots x_{j(m)}\right)$. By assumption, $\kappa_\pi\left(\xi_i x_{j(1)} \cdots x_{j(m)}\right) \neq 0$ only if ξ_i belongs to a block of size 2. Thus we obtain

$$\begin{aligned}
\tau\left(\xi_i x_{j(1)} \cdots x_{j(m)}\right) &= \sum_{\pi=(1,k)\cup\pi_1\cup\pi_2} \kappa_\pi(\xi_i, x_{j(1)}, \ldots, x_{j(\ell)}), \\
&= \sum_k \delta_{i,j(k)} \tau(x_{j(1)}, \ldots, x_{j(k-1)})\tau(x_{j(k+1)}, \ldots, x_{j(\ell)}),
\end{aligned}$$

like above, which is precisely the conjugate relation. □

Remark 4.12. Notice that the conjugate relations in cumulant fom (4.2.2) are extremely similar to the cumulant relations of a free semicircular family given in Definition 2.33. It is in fact immediate that the conjugate system to a free standard semicircular family is itself (interpreted as elements of $L^2(M)$). In general, the conjugate system to a free semicircular family $x_1, \ldots, x_n$ is $\xi_i = (2/\rho_i)^2 x_i$, which is exactly the same as the score functions we found for the Gaußian density in Example 1.14.

To close this section, we discuss a formula for the conjugate variable in the commutative ($n = 1$) case, and use it to give another proof that the conjugate variable to a semicircular one is itself. For this we need the Hilbert transform H, which can be defined as

$$H(f)(s) = \frac{1}{\pi} \int \frac{f(s)}{s-t}\, \mathrm{d}\lambda(t),$$

where the integral is taken in the principal value sense, and $f \in L^p(\mathbb{R})$ for some $p \in (1, \infty)$. It is known that this is a bounded linear operator from any such $L^p(\mathbb{R})$ to itself. We will use it here to describe how the conjugate variables can be defined if the distribution of the random variable admits a suitably integrable density.

Proposition 4.13. *Let x be a self-adjoint element of M, such that its analytic distribution has density $p \in L^3(\mathbb{R})$. Then we can use spectral calculus to define*

$$\xi = 2\pi H(p)(x),$$

and this is the conjugate variable for x.

Proof. We need the following fact about the Hilbert transform. Instead of taking the principal value definition, one can also define it via

$$H(p)(s) = \lim_{\varepsilon \downarrow 0} H_\varepsilon(p)(s) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int p(t) \frac{s-t}{(s-t)^2 + \varepsilon^2} \mathrm{d}\lambda(t),$$

where this limit is taken in $L^3(\mathbb{R})$, as p is a member of this space.

The reason why we need p to be cube-integrable, is that then $p \in L^2(\mathbb{R}, p\lambda)$, which is also the space that we want ξ to live in, as it is isomorphic to $L^2(x)$. Thus we need to show that $H(p)$ and the $H_\varepsilon(p)$ all live in $L^2(\mathbb{R}, p\lambda)$ as well. This is an application of Hölders inequality. Notice that $(3)^{-1} + (3/2)^{-1} = 1$, thus

$$\int |H(p)(s)|^2 p(s)\ \mathrm{d}\lambda(s) \leq \|p\|_3 \left\|H(p)^2\right\|_{3/2} = \|p\|_3 \|H(p)\|_3^2 < \infty,$$

as $H(p) \in L^3(\mathbb{R})$ because $p \in L^3(\mathbb{R})$. The same argument also applies to $H_\varepsilon(p)$. However, we also need convergence in $L^2(\mathbb{R}, p\lambda)$ of $H_\varepsilon(p)$ to $H(p)$, which follows again from Hölders inequality

$$\int |H_\varepsilon(p)(s) - H(p)(s)|^2 p(s)\ \mathrm{d}\lambda(s) \leq \|p\|_3 \|H_\varepsilon(p) - H(p)\|_3^2 \to 0.$$

We take f to be a polynomial, and thus f is bounded on the support of p (which is compact by boundedness of x), and hence is a member of $L^\infty(\mathbb{R}, p\lambda)$. It is immediate that then also $fH_\varepsilon(p)$ converges to $fH(p)$ in $L^2(\mathbb{R}, p\lambda)$. Thus, taking all subsequent limits in $L^2(\mathbb{R}, p\lambda)$,

$$\tau(f(x)\xi) = 2\pi \int f(s)H(p)(s)p(s)\ \mathrm{d}\lambda(s) = \lim_{\varepsilon \downarrow 0} 2\pi \int f(s)H_\varepsilon(p)(s)p(s)\ \mathrm{d}\lambda(s).$$

Now, since for all $\varepsilon > 0$

$$\left|(f(s) - f(t)) \frac{s-t}{(s-t)^2 + \varepsilon^2}\right| \leq \left|\frac{f(s) - f(t)}{s-t}\right| = |\partial f(s,t)|,$$

and this final expression is bounded on the support of p (because $(s-t)$ divides $f(s) - f(t)$ and so the quotient is again a polynomial), we also have that

$$\begin{aligned}(\tau \otimes \tau)(\partial f(x)) &= \iint \frac{f(s) - f(t)}{s-t} p(s)p(t)\ \mathrm{d}\lambda(s)\mathrm{d}\lambda(t), \\ &= \lim_{\varepsilon \downarrow 0} \iint (f(s) - f(t)) \frac{s-t}{(s-t)^2 + \varepsilon^2} p(s)p(t)\ \mathrm{d}\lambda(s)\mathrm{d}\lambda(t).\end{aligned}$$

However, the rightmost expression can also be manipulated, using Fubini and renaming dummy integration variables, to read

$$\begin{aligned}\iint (f(s) - f(t)) \frac{s-t}{(s-t)^2 + \varepsilon^2} p(s)p(t)\ \mathrm{d}\lambda(s)\mathrm{d}\lambda(t) &= 2 \iint f(s) \frac{s-t}{(s-t)^2 + \varepsilon^2} p(s)p(t)\ \mathrm{d}\lambda(t)\mathrm{d}\lambda(s), \\ &= 2\pi \int f(s)H_\varepsilon(p)(s)p(s)\ \mathrm{d}\lambda(s),\end{aligned}$$

but we already saw that this converges to $\tau(f(x)\xi)$. We conclude that $\tau(f(x)\xi) = (\tau \otimes \tau)(\partial f(x))$ for all polynomials f, but these are precisely the conjugate relations, and so ξ is the conjugate variable to x (notice that it is in $L^2(x)$ by functional calculus). □

Example 4.14. We end this section by computing again the conjugate variable to a single standard semicircular element x. According to Remark 4.12 and the proposition we just proved, it should be true that

$$\xi = x = 2\pi H(\sigma)(x),$$

or in other words

$$s = 2\pi H(\sigma)(s)$$

as functions. Thus we should compute the principal value integral

$$\frac{2\pi}{\pi}\int_{\mathrm{supp}(\sigma)} \frac{\frac{1}{2\pi}\sqrt{4-t^2}}{s-t}\mathrm{d}t = \frac{1}{\pi}\int_{-2}^{2}\frac{\sqrt{4-t^2}}{s-t}\mathrm{d}t.$$

This is a rather messy business, but requires nothing more than elementary calculus. One approach is sketched here (the cases $s = -2, 0, 2$ require some special care). As the domain of integration is symmetric, it may be fruitful to split the integrand into an even and an odd part in t as

$$\frac{\sqrt{4-t^2}}{s-t} = t\frac{\sqrt{4-t^2}}{s^2-t^2} + s\frac{\sqrt{4-t^2}}{s^2-t^2}.$$

The first term can be integrated using the successive substitutions $t = 2\sin(u)$, $v = \cos(u)$, yielding the primitive

$$\frac{\sqrt{4-s^2}}{2}\left[\mathrm{Log}\left(1+\sqrt{\frac{4-t^2}{4-s^2}}\right) - \mathrm{Log}\left(1-\sqrt{\frac{4-t^2}{4-s^2}}\right)\right] - \sqrt{4-t^2} + \text{ constant},$$

where Log denotes some fixed branch of the logarithm such that it is defined on all non-zero real numbers. The second and third terms are irrelevant to the definite integral, as they are regular and take on the same values on the boundary. Using well known properties of logarithms, one can take the principal value of the first term to find zero. In particular, there are two singularities, at $t = \pm s$ respectively, neither giving a real contribution, and the imaginary contributions of the two cancel. The second integrand can be done with the somewhat less common combination of substitutions $t = 2\sin(u)$, $v = (\tan(u))^{-1}$, yielding the primitive

$$\frac{\sqrt{4-s^2}}{2}\left[\mathrm{Log}\left(1+\frac{s}{t}\sqrt{\frac{4-t^2}{4-s^2}}\right) - \mathrm{Log}\left(1-\frac{s}{t}\sqrt{\frac{4-t^2}{4-s^2}}\right)\right] - s\arctan\left(\frac{\sqrt{4-t^2}}{t}\right) + \text{ constant}.$$

The constant is again irrelevant. One now proceeds for the first term as before, taking care that there are now three singularities, namely at zero and at the two points where $(s\sqrt{4-t^2})/(t\sqrt{4-s^2})$ takes on the values ± 1 respectively, and again one finds zero. The second, arctangent, term has a lone singularity at zero, where one can use that $\arctan(\pm\infty) = \pm\pi/2$ to find that the principal value gives precisely πs. Combining everything, one recovers our desired result that $2\pi H(\sigma)(s) = s$.

4.3 Free Fisher Information and Non-Microstates Free Entropy

Having constructed a noncommutative analogue to score functions in the past two sections, there is now a natural way to define the free Fisher information, and from there a notion of free entropy. Unfortunately, this will not turn out to be the most practical version for our purposes. Therefore this section will not contain any proofs, and is included only for completeness.

Recall from Section 1.2 that the score functions ξ_i and the Fisher information Φ were related via

$$\Phi = \sum_{i=1}^{n} \mathbb{E}[\xi_i^2].$$

With the technology from the previous two sections, we can define the following.

Definition 4.15. The *free Fisher information* Φ^* of the self-adjoint elements $x_1, \ldots, x_n \in M$ is defined to be ∞ if the x_i do not admit a conjugate system ξ_i, and as

$$\Phi^*(x_1, \ldots, x_n) = \sum_{i=1}^{n} \|\xi_i\|_2^2,$$

otherwise.

Remark 4.16. In Theorem 4.10 we showed that a conjugate system is unique if it exists, so that Φ^* is well-defined as an extended real number.

The free Fisher information has the following nice behaviour with respect to freeness.

Theorem 4.17. *Let $x_1, \ldots, x_n, y_1, \ldots, y_m \in M$ be self-adjoint. If the x_i and y_j are free, then*

$$\Phi^*(x_1, \ldots, x_n, y_1, \ldots, y_m) = \Phi^*(x_1, \ldots, x_n) + \Phi^*(y_1, \ldots, y_m).$$

Under the additional assumption that all three free Fisher informations are finite, the converse is also true.

In the commutative case, one can obtain the following extension of Proposition 4.13.

Proposition 4.18. *Adopt the setting and notation of Proposition 4.13, then the free Fisher information is given by the formula*

$$\Phi^*(x) = \frac{4\pi}{3}\|p\|_3^3 < \infty.$$

Conversely, if x is a self-adjoint element of M such that $\Phi^*(x) < \infty$, then its distribution admits a cube-integrable density, and the same formula is valid.

In Section 1.2, we defined the Fisher information as the derivative of the entropy along a diffusion flow applied to the density. So to invert this relation, we would like to define a notion of free entropy as an integral of Φ^* along a 'diffusion flow'. The question is what object from free probability is related to diffusion. Classically, it is well-known that the diffusion equation has a fundamental solution given by a Gaußian with mean zero and variance proportional to the time elapsed t. As we have established throughout the past few chapters, the role of the Gaußians in classical probability are played by semicircular elements in free probability. Thus it seems meaningful to consider standard semicircular elements s_i, which are then multiplied by a factor $\sqrt{t}$ to obtain a variance proportional to t. Now, fundamental solutions act by convolution, for which we do not have a free probability analogue. However, the convolution of densities associated to two independent random variables in classical probability gives the probability density for their sum. Since we are acting on the level of random variables and not densities, we could work with the elements $x_i + \sqrt{t}s_i$ and demand that the s_j are free from the x_k. Thus we have motivated the following definition, which we will not use further in this thesis.

Definition 4.19. Let $x_1, \ldots, x_n \in M$ be self-adjoint, then their *non-microstates free entropy* χ^* is defined by the formula

$$\chi^*(x_1, \ldots, x_n) = \frac{1}{2}\int_0^\infty \left[\frac{n}{1+t} - \Phi^*(x_1 + \sqrt{t}s_1, \ldots, x_n + \sqrt{t}s_n)\right] \mathrm{d}t + \frac{n}{2}\ln(2\pi e),$$

where $s_1, \ldots, s_n$ is a free standard semicircular family that is free from the x_i.

4.4 Microstates Free Entropy

In this section, we define a different notion of free entropy. Instead of relating it to the Fisher information, we will draw inspiration from the physical interpretation of entropy originally provided by Boltzman. In statistical mechanics, one attempts to describe physical systems that contain a large number of particles (say for example 10^{23}, the approximate value of the Avogadro number). However, for most practical and physical purposes, the precise individual positions and momenta of all the particles are not very interesting, nor particularly important. More generally, it is not necessary to know the precise location, called the *microstate*, of the system in phase space. Despite possessing a large number of internal degrees of freedom, the behaviour of these systems as a whole can usually be satisfactorily captured by a handful of derived quantities, such as temperature, pressure, average density, and volume. These are called *macroscopic* quantities, and together they define a *macrostate* of a physical system. The reason that the precise microstate of a system is frequently of lesser importance, is that many microstates give rise to the same macrostate. The fundamental postulate of statistical mechanics, is that in equilibrium, the system is equally likely to be in any particular microstate. Thus, when measuring the macroscopic properties of a system, one is most likely to observe that macrostate to which the most microstates give rise. To describe this behaviour quantitatively, Boltzman defined the entropy of macrostates, which was to be taken as proportional to the logarithm of the number of microstates leading to the selected macrostate. Now as each microstate is equally likely, and the system should always be in *some* microstate, this can be interpreted as measuring the logarithm of the volume of phase space that belongs to a certain macrostate.

So far, we have been discussing systems with a large, but finite number of degrees of freedom. In some sense, random variables in a type II_1 factor have an infinite number of degrees of freedom, as their spectra can contain infinitely many points (as opposed to elements of a matrix algebra). Physicists usually deal with this by taking the *thermodynamic limit*, that is, letting the number of particles formally diverge during calculations (although often in such a way that quantities like the density remain finite).

To apply this to free probability, we imagine that an n-tuple of self-adjoint elements $x_1, \ldots, x_n \in M$ define a macrostate of some system in the thermodynamic limit (self-adjointness is also physically appropriate as observables should be real). The question is then, what is a microstate? To the x_i is associated a collection of real numbers, namely their joint moments. The key idea is now to consider finite dimensional approximations to the x_i in the sense of moments. Naturally, we cannot expect to approximate all moments of all orders simultaneously using matrices, as a self-adjoint $N \times N$ matrix has only N^2 real degrees of freedom (see the text after the definition below). Thus we should introduce some cut-off for the order of moments, and for technical reasons we should also prevent norms of the matrices from becoming too large (otherwise their combined volume might become infinite). We have now motivated and explained the following definition.

Definition 4.20. Write $(\mathfrak{M}_N^{sa}, \mathrm{tr}_N)$ for the $*$-probability space of self-adjoint $N \times N$ matrices with the normalised trace $\mathrm{tr}_N = N^{-1}\,\mathrm{Tr}$. Let $x_1, \ldots, x_n \in M$ be self-adjoint, then the associated sets of *microstates* are

$$\Gamma_R(x_1, \ldots, x_n; N, m, \varepsilon) := \Bigg\{ (A_1, \ldots, A_n) \in (\mathfrak{M}_N^{sa})^n \,\Bigg|\, \big|\tau(x_{j(1)} \cdots x_{j(k)}) - \mathrm{tr}_N(A_{j(1)} \cdots A_{j(k)})\big| < \varepsilon \\ \forall\ 1 \leq k \leq m, 1 \leq j(1), \ldots, j(k) \leq n, \text{ and } \|A_i\| \leq R \text{ for all } 1 \leq i \leq n \Bigg\},$$

where $m \in \mathbb{Z}_{\geq 1}$ is the order cut-off, $R > 0$ regulates the norms, and $\varepsilon > 0$ controls the accuracy of the approximation.

If we now want to determine how large these sets of microstates are, we should agree on a measure on $\mathfrak{M}_N^{sa}$. The measure on the sets of n-tuples then follows by taking products. As remarked earlier,

a self-adjoint $N \times N$ matrix represents N^2 real degrees of freedom, as there are N real elements on the diagonal, and the upper triangular elements represent the rest of the degrees of freedom. There are $N(N-1)/2$ entries in the upper triangle, and as each is complex, it contributes two real degrees of freedom. Hence we find that the total number of degrees of freedom is $N + 2N(N-1)/2 = N^2$. Thus it would be natural to port over the Lebesgue measure from $\mathbb{R}^{N^2}$, but this requires a choice of isomorphism. Since we are considering $\mathfrak{M}_N^{sa}$ as a $*$-probability space, the obvious norm is the one induced by the inner product $(A, B) \mapsto \mathrm{Tr}(B^*A)$. We can then fix an isomorphism by demanding that it intertwines this inner product with the Euclidean one on $\mathbb{R}^{N^2}$. The pull-back of the Lebesgue measure along this isomorphism will be denoted by Λ. Note that this will not be a map that takes the elements a_{ij} of a matrix A and simply puts them into a vector in some order. The reason for this is the asymmetry between on-and off diagonal elements for the norm

$$\mathrm{Tr}\big(A^2\big) = \sum_{i=1}^{N} \mathfrak{Re}[a_{ii}]^2 + 2 \sum_{1 \leq i < j \leq N} \big(\mathfrak{Re}[a_{ij}]^2 + \mathfrak{Im}[a_{ij}]^2\big)\,.$$

Incidentally, this is the norm we imply in the definition of free microstates entropy.

We should now think about how to take the thermodynamic limit $N \to \infty$. It is conceivable that in the limit, every degree of freedom tends to some fixed, finite, non-zero contribution to the entropy, which would cause a divergence. Hence we will instead look at the entropy per degree of freedom, that is to say we regularise via division by N^2. However, this is not enough to ensure convergence, as the following argument indicates. Notice that for $m \geq 2$, any microstate $(A_1, \ldots, A_n)$ satisfies

$$\mathrm{tr}_N(A_1^2 + \cdots + A_n^2) \leq \tau(x_1^2 + \cdots + x_n^2) + n\varepsilon =: C,$$

and thus (neglecting to write the isomorphism)

$$\Gamma_R(x_1, \ldots, x_n; N, m, \varepsilon) \subset \mathrm{Ball}_{nN^2}(\sqrt{N}C) \subset \mathbb{R}^{nN^2}.$$

We derive the asymptotic behaviour of

$$\frac{1}{N^2} \log \lambda \left(\mathrm{Ball}_{nN^2}(\sqrt{N}C)\right).$$

It is well known that the volume $V_d(r)$ of a d-dimensional ball with radius r is given by the formula

$$V_d(r) = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} r^d,$$

where Γ is the Euler Gamma function, not to be confused with a set of microstates. Using this, we find

$$\frac{1}{N^2} \log \lambda \left(\mathrm{Ball}_{nN^2}(\sqrt{N}C)\right) = \frac{n}{2} \log\big(\pi C^2\big) + \frac{n}{2} \log N - \frac{1}{N^2} \log \Gamma\left(\frac{nN^2}{2} + 1\right).$$

Recall the leading terms in Stirling's approximation

$$\log \Gamma(x) = x \log x - x + \mathcal{O}(\log x).$$

This yields the asymptotic behaviour for $N \to \infty$

$$\frac{1}{N^2} \log \lambda \left(\mathrm{Ball}_{nN^2}(\sqrt{N}C)\right) \sim -\frac{n}{2} \log N,$$

which is a logarithmic divergence. This implies that we should add counter-terms to the expression we use to define a notion of entropy. It turns out that the one we have just discussed is the only one, and we will prove this immediately after finally giving the definition

Definition 4.21. Let $x_1, \ldots, x_n \in M$ be self-adjoint. Define their *free microstates entropy* $\chi(x_1, \ldots, x_n)$ as an extended real number ($-\infty$ is allowed) according to the following procedure. Write

$$\chi_R(x_1, \ldots, x_n; N, m, \varepsilon) := \log \Lambda \left(\Gamma_R(x_1, \ldots, x_n; N, m, \varepsilon) \right),$$

take the thermodynamic limit according to

$$\chi_R(x_1, \ldots, x_n; m, \varepsilon) := \limsup_{N \to \infty} \left[\frac{1}{N^2} \chi_R(x_1, \ldots, x_n; N, m, \varepsilon) + \frac{n}{2} \log N \right],$$

consider perfect approximations to all orders via

$$\chi_R(x_1, \ldots, x_n) := \lim_{\substack{m \to \infty \\ \varepsilon \to 0}} \chi_R(x_1, \ldots, x_n; m, \varepsilon),$$

and finally, remove the norm cut-off to find

$$\chi(x_1, \ldots, x_n) := \sup_{R > 0} \chi_R(x_1, \ldots, x_n).$$

Proposition 4.22. *The free microstates entropy is well-defined.*

Proof. We provide a more detailed version of the proof in [87] (Proposition 2.2). Notice that in passing to perfect approximation to all orders, $\chi_R(x_1, \ldots, x_n; m, \varepsilon)$ is an increasing function of ε, and a decreasing function in m, so that the limit makes sense since we allow $-\infty$ as an answer. We will prove the following estimate,

$$\chi_R(x_1, \ldots, x_n; N, m, \varepsilon) \leq \frac{n}{2} N^2 \left[\log\left(\frac{2\pi e}{n} (T^2 + n\varepsilon) \right) - \log N \right], \tag{4.4.1}$$

where we wrote $T^2 = \tau(x_1^2 + \cdots + x_n^2)$. This is uniform in R, and we see that it also implies

$$\left[\frac{1}{N^2} \chi_R(x_1, \ldots, x_n; N, m, \varepsilon) + \frac{n}{2} \log N \right] \leq \frac{n}{2} \log\left(\frac{2\pi e}{n} (T^2 + n\varepsilon) \right),$$

which is uniform in N. Thus it suffices to prove this estimate to establish that χ is well-defined. In fact, we have already done half the work back in Proposition 1.15. We can exploit that result by viewing $\Gamma_R(x_1, \ldots, x_n; N, m, \varepsilon)$ as a subset of $\mathbb{R}^{nN^2}$ and defining f to be its normalised indicator function

$$f(A_1, \ldots, A_n) = \frac{1}{\Lambda(\Gamma_R(\ldots))} 1_{\Gamma_R(\ldots)}(A_1, \ldots, A_n),$$

so that the integral of f over $\mathbb{R}^{nN^2}$ with respect to the Lebesgue measure is unity. Here it is important that we have a norm cut-off, as otherwise it might be possible for the set of microstates to have infinite measure. If we have a collection of parameters for which the set of microstates has zero volume, than any real number is trivially an upper bound for the logarithm of its measure. Therefore f is a compactly supported probability density, so that it has a well-defined covariance matrix Σ. Now we can interpret

$$\chi_R(x_1, \ldots, x_n; N, m, \varepsilon) = \log \Lambda \left(\Gamma_R(x_1, \ldots, x_n; N, m, \varepsilon) \right) = -\int f \log f \, \mathrm{d}\lambda,$$

which is the classical entropy of f. To this we can apply the estimate of Proposition 1.15 to find

$$\chi_R(x_1, \ldots, x_n; N, m, \varepsilon) \leq \frac{n}{2} N^2 \log(2\pi e) + \frac{1}{2} \log \det(\Sigma).$$

Since Σ is a positive definite symmetric matrix, we can employ the identity $\det(\Sigma) = \exp(\operatorname{Tr}\log\Sigma)$, whose proof is trivial using the properties of Σ. This yields

$$\chi_R(x_1,\ldots,x_n;N,m,\varepsilon) \leq \frac{n}{2}N^2\log(2\pi e) + \frac{nN^2}{2}\operatorname{tr}_{nN^2}\log\Sigma.$$

As tr_{nN^2} is an expectation on $\mathbb{R}^{nN^2}$ (modulo identification with $(\mathfrak{M}_N^{sa})^n$), we can use a small generalisation of Jensen's inequality from Lemma 1.6 to interchange it with the logarithm, so that

$$\chi_R(x_1,\ldots,x_n;N,m,\varepsilon) \leq \frac{n}{2}N^2\log(2\pi e) + \frac{n}{2}N^2\log\operatorname{tr}_{nN^2}\Sigma$$

It is easy to see that the formula for the trace of Σ is

$$\operatorname{tr}_{nN^2}\Sigma = \frac{1}{nN}\int \operatorname{tr}_N(A_1^2+\cdots+A_n^2)f\ \mathrm{d}\lambda,$$

but by definition of the microstate sets, this is at most $(T^2+n\varepsilon)/nN$. Putting all of this together, we have indeed that

$$\chi_R(x_1,\ldots,x_n;N,m,\varepsilon) \leq \frac{n}{2}N^2\left[\log\left(\frac{2\pi e}{n}(T^2+n\varepsilon)\right) - \log N\right],$$

and the proof is complete. □

While the definition of χ is complicated, there are some properties that one can prove rather straightforwardly by exploiting inclusions of certain sets of microstates into others. One example is the obvious inclusion

$$\Gamma_R(x_1,\ldots,x_n;N,m,\varepsilon) \subset \Gamma_R(x_1,\ldots,x_k;N,m,\varepsilon)\times\Gamma_R(x_{k+1},\ldots,x_n;N,m,\varepsilon),$$

for any $1 \leq k < n$, which implies that

$$\chi(x_1,\ldots,x_n) \leq \chi(x_1,\ldots,x_k) + \chi(x_{k+1},\ldots,x_n),$$

and in particular

$$\chi(x_1,\ldots,x_n) \leq \chi(x_1)+\cdots+\chi(x_n).$$

Another nice property of χ (See [87], Proposition 5.2), which is reminiscent of χ^*, is the fact that for x_i a free family one even has

$$\chi(x_1,\ldots,x_n) = \chi(x_1)+\cdots+\chi(x_n).$$

The question of how similar χ and χ^* are, in particular whether or not they are the same, is one of the major open problems in free probability. For $n = 1$ it is known that they agree, but the proof hinges on the fact that there are explicit formulas available in this case (see the next chapter). In general, it was shown by Biane, Capitaine, and Guionnet that $\chi \leq \chi^*$ [9], but more than this is not currently known.

5 Commutative Free Entropy and Monotone Transport

5.1 Commutative Free Entropy

In this chapter we describe how free monotone transport works in the commutative case. This allows us to introduce the strategy of the general proof in a setting where there are far fewer technical complications. One reason for the decreased technical difficulty, is that the free microstates entropy admits an integral formula in terms of the analytic distribution of the self-adjoint element under consideration.

Theorem 5.1. *Let (M,τ) be a tracial W*-probability space, and x a self-adjoint element of M. If μ_x denotes the analytic distribution of x, then its free entropy is given by the formula*

$$\chi(x) = \iint \log|s-t|\ \mathrm{d}\mu_x(s)\mathrm{d}\mu_x(t) + \frac{1}{2}\log(2\pi) + \frac{3}{4}. \tag{5.1.1}$$

We will not prove this theorem (for a proof see Section 4 of [87]), but we will give some heuristic arguments that make this formula plausible. It is known from random matrix theory, that a standard semicircular element arises in some sense as the limit of certain random $N\times N$ matrices A_N as N grows to infinity (see the first two chapters of [56]). These matrices are randomly selected from $\mathfrak{M}_N^{sa}$ according to the probability density

$$P_N(A) \propto \exp\left(-\frac{N}{2}\operatorname{Tr}\left[A^2\right]\right).$$

This is called the *Gaußian Unitary Ensemble* (GUE). These random matrices give rise to *empirical eigenvalue distributions*, defined as

$$\mu_A^{(N)} = \frac{1}{N}\left(\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}\right).$$

What it means in this context for these random matrices to converge to a semicircular element, is that this empirical measure should converge in some sense to the Wigner standard semicircle law. One can now ask how likely it is to observe deviations from this law. Classically, the answer is given by Sanov's Theorem, which states that the probability to observe an empirical law 'close' to a measure ν is asymptotically of the form $\exp(-N^2\mathcal{I}(\nu))$. Here, $\mathcal{I}(\nu)\geq 0$ is called a *rate function*, and it satisfies $\mathcal{I}(\sigma)=0$ (as the probability of observing the semicircle law should tend to unity). Moreover, Sanov's Theorem states that this rate function can be interpreted as the relative entropy of ν and σ (see Section 1.2). Thus there should be some relation between this rate function $\mathcal{I}$ and the free entropy χ.

It is known that the density P_N induces the density

$$\tilde{P}_N(\lambda_1,\ldots,\lambda_N) \propto \left[\prod_{1\leq i<j\leq n}(\lambda_i-\lambda_j)^2\right]\exp\left(-\frac{N}{2}\sum_{i=1}^N\lambda_i^2\right) \tag{5.1.2}$$

on the space of eigenvalues. Therefore, the probability that we observe $N^{-1}(\delta_{\lambda_1}+\cdots+\delta_{\lambda_N})\approx\nu$ can be computed using this density as the integral

$$C_N\int_{\{N^{-1}(\delta_{\lambda_1}+\cdots+\delta_{\lambda_N})\approx\nu\}}\left[\prod_{1\leq i<j\leq n}(\lambda_i-\lambda_j)^2\right]\exp\left(-\frac{N}{2}\sum_{i=1}^N\lambda_i^2\right)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_N,$$

where C_N is an N-dependant constant that is related to the normalisation of $\tilde{P}_N$. Now notice that on the domain of integration,

$$-\frac{N}{2}\sum_{i=1}^N\lambda_i^2 = -N^2\int\frac{t^2}{2}\,\mathrm{d}\left(\frac{\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}}{N}\right)(t) \approx -N^2\int\frac{t^2}{2}\mathrm{d}\nu(t),$$

and

$$\prod_{1\leq i<j\leq n}(\lambda_i-\lambda_j)^2=\exp\left(\sum_{i\neq j}\log|\lambda_i-\lambda_j|\right)\approx\exp\left(N^2\iint\log|s-t|\;\mathrm{d}\nu(s)\mathrm{d}\nu(t)\right).$$

This suggests that the appropriate rate function is

$$\mathcal{I}(\nu)=-\iint\log|s-t|\;\mathrm{d}\nu(s)\mathrm{d}\nu(t)+\int\frac{t^2}{2}\mathrm{d}\nu(t)-\frac{3}{4}, \tag{5.1.3}$$

where the constant arises from the limit of $N^{-2}\log C_N$.

Some simple estimates show that this is a meaningful quantity as an extended real number (in this case $+\infty$ is allowed, meaning the probability of observing that particular measure is zero). Rewrite the integral as

$$\mathcal{I}(\nu)=\frac{1}{4}\iint\left[s^2+t^2-4\log|s-t|\right]\mathrm{d}\nu(s)\mathrm{d}\nu(t)-\frac{3}{4}.$$

For $s\neq t$, this integrand can be cast into the form

$$s^2+t^2-2\log\bigl(s^2+t^2\bigr)+4\log\left(\frac{\sqrt{s^2+t^2}}{|s-t|}\right).$$

Elementary calculus shows that the first three terms are at least $2-2\log(2)$, and for the second term use $2(s^2+t^2)\geq(s-t)^2$ to arrive at

$$\mathcal{I}(\nu)\geq-\frac{1}{4}\left(1+4\log(2)\right).$$

One can also show by direct computation that $\mathcal{I}(\sigma)=0$, as it should be. Comparing (5.1.1) and (5.1.3), one sees that there is an extra constant $\log(2\pi)/2$ in χ, and the second moment term is missing. The extra constant is there, so that when computing $\chi(\sigma)$ with this formula, the bound (4.4.1) is saturated.

Remark 5.2. Notice that the eigenvalue density of the GUE given in (5.1.2), cannot be factorised, so that the eigenvalues are not independent. One can understand why this must be the case via the following argument (see [83]). Let $H(t)$ be a smooth curve in $\mathfrak{M}_N^{sa}$ with $t\in(-\varepsilon,\varepsilon)$ for some $\varepsilon>0$. This is to be interpreted as a perturbation of the self-adjoint matrix $H(0)$. We can now derive variation formulas for the eigenvalues $\lambda_1(t),\ldots,\lambda_N(t)$ and corresponding orthonormal eigenbasis $v_1(t),\ldots,v_N(t)$. In writing this, we assume for the sake of simplicity that all eigenvalues are distinct. For notational convenience, we do not write the explicit time dependence in the sequel. Differentiate the eigenvalue equation and the orthonormal eigenbasis relations to find

$$\begin{aligned}Hv_j=\lambda_jv_j\;&\Rightarrow\;\dot{H}v_j+H\dot{v}_j=\dot{\lambda}_jv_j+\lambda_j\dot{v}_j,\\ v_j^*v_j=1\;&\Rightarrow\;\dot{v}_j^*v_j=0=v_j^*\dot{v}_j,\\ v_i^*v_j=0\;&\Rightarrow\;\dot{v}_i^*v_j=-v_i^*\dot{v}_j,\end{aligned}$$

where $i\neq j$. For the eigenvalue equation, multiply by v_j^* to find

$$v_j^*\dot{H}v_j+v_j^*H\dot{v}_j=\dot{\lambda}_jv_j^*v_j+\lambda_jv_j^*\dot{v}_j\;\Rightarrow\;\dot{\lambda}_j=v_j^*\dot{H}v_j,$$

using the other relations derived above, and the adjoint eigenvalue equations. Thus we have found the first variation of the eigenvalues. Now multiply the eigenvalue equation anew, this time with v_i^*, to find

$$v_i^*\dot{H}v_j+v_i^*H\dot{v}_j=\dot{\lambda}_jv_i^*v_j+\lambda_jv_i^*\dot{v}_j\;\Rightarrow\;v_i^*\dot{H}v_j+(\lambda_i-\lambda_j)v_i^*\dot{v}_j=0.$$

Using the basis properties of the eigenvectors and their adjoints, this yields the first variation of the eigenvectors

$$\dot{v}_j = c_j v_j + \sum_{i \neq j} \frac{v_i^* \dot{H} v_j}{\lambda_j - \lambda_i} v_i, \;\; \dot{v}_j^* = -c_j v_j^* + \sum_{i \neq j} \frac{v_j^* \dot{H} v_i}{\lambda_j - \lambda_i} v_i^*,$$

where the c_j are some scalars (they arise from the fact that the eigenvectors may be scaled). Taking an additional time derivative of the first variation formula for the eigenvalues, we find

$$\ddot{\lambda}_j = \dot{v}_j^* \dot{H} v_j + v_j^* \ddot{H} v_j + v_j^* \dot{H} \dot{v}_j.$$

Plugging in the first variation equations for the eigenvectors and simplifying, we uncover

$$\ddot{\lambda}_j = v_j^* \ddot{H} v_j + 2 \sum_{i \neq j} \frac{\left| v_j^* \dot{H} v_j \right|^2}{\lambda_j - \lambda_i}.$$

This can be interpreted as the forces that 'accelerate' the eigenvalues. There is a 'diagonal' term that originates from the acceleration along the curve $H(t)$, and another term that looks like a repulsion between the eigenvalues. This is a well-known phenomenon in quantum physics, where it is folklore knowledge that perturbations of degenerate Hamiltonians generically lift these degeneracies.

5.2 Minimising Rate Functionals

In this section we make rigorous some of the observations concerning the rate function I from the previous section, following [7] and the references therein. In particular we will prove that I has a unique minimiser, which is of course the standard semicircle measure. For this, we introduce some notation. Write $\mathscr{M}_1^+(B)$ for the convex space of probability measures on B, where $B \subset \mathbb{R}$ is Borel. If A is a bounded Borel set, define its *logarithmic capacity* $\gamma(A)$ as the number

$$\gamma(A) := \exp\left[- \inf_{\nu \in \mathscr{M}_1^+(A)} \iint \log \frac{1}{|x-y|} \, \mathrm{d}\nu(x) \mathrm{d}\nu(y) \right],$$

which by the boundedness of A is finite. Let $f : \mathbb{R}^2 \to [0, \infty]$ be the continuous proper function given by

$$f(x,y) = \begin{cases} \frac{1}{4}(x^2+y^2) - \log|x-y| & x \neq y, \\ +\infty & x = y, \end{cases},$$

and for $\mu \in \mathscr{M}_1^+(\mathbb{R})$, set

$$F(\mu) = \iint f(x,y) \, \mathrm{d}\mu(x) \mathrm{d}\mu(y).$$

Then $I = F - 3/4$, and we can obtain the following properties of I by proving them for F.

Proposition 5.3.

(i) F is well-defined on $\mathscr{M}_1^+(\mathbb{R})$, and takes on values in $[-(1+4\log(2))/4, \infty]$.

(ii) If $F(\mu)$ is to be finite, μ must have finite second moment, be diffuse, and not give mass to sets of vanishing logarithmic capacity (so μ must be log-lipophilic *in the terminology of Section 1.3).*

(iii) F is a strictly convex function on $\mathscr{M}_1^+(\mathbb{R})$.

Proof. We already proved part (i) in the previous section. The first two statements in part (ii) are trivial. For the other one, let B be a Borel set of $\mathbb{R}^2$, then as f is bounded from below, say by $L \in \mathbb{R}$, we have

$$F(\mu) \geq \int_B f \ \mathrm{d}(\mu \otimes \mu) + L\left[1 - (\mu \otimes \mu)(B)\right].$$

Let A be a Borel set of $\mathbb{R}$, take $B = A \times A$, and focus on the integral in this lower bound, which can be estimated by

$$\begin{aligned}\int_{A\times A} f \ \mathrm{d}(\mu \otimes \mu) &\geq \int_A \int_A \log \frac{1}{|x-y|} \ \mathrm{d}\mu(x)\mathrm{d}\mu(y), \\ &\geq -\mu(A)^2 \left[- \inf_{\nu \in \mathscr{M}_1^+(A)} \int_A \int_A \log \frac{1}{|x-y|} \ \mathrm{d}\nu(x)\mathrm{d}\nu(y) \right], \\ &= -\mu(A)^2 \log \gamma(A).\end{aligned}$$

Thus if $\gamma(A) = 0$, but $\mu(A) > 0$, F must be infinite.

For part (iii), it is clearly enough to show that

$$H(\mu) = \iint \log|x-y| \ \mathrm{d}\mu(x)\mathrm{d}\mu(y),$$

is strictly concave in μ (as the other term in F is linear). To achieve this, we use the identity

$$\log|x-y| = \int_0^\infty \frac{1}{2t}\left[\exp\left(-\frac{1}{2t}\right) - \exp\left(-\frac{(x-y)^2}{2t}\right)\right] \mathrm{d}t.$$

This is an example of a Frullani integral, which is a general formula, stating that for a function g that is continuous on $(0,\infty)$ and has finite limits at 0^+ and $+\infty$,

$$\int_0^\infty \frac{g(at) - g(bt)}{t} \ \mathrm{d}t = (g(0^+) - g(+\infty)) \ln\left(\frac{b}{a}\right),$$

for any two $a, b > 0$. We will use it in the following way. Define for all $T \in [1,\infty)$

$$H_T(\mu) = \iint \int_{1/T}^T \frac{1}{2t}\left[\exp\left(-\frac{1}{2t}\right) - \exp\left(-\frac{(x-y)^2}{2t}\right)\right] \mathrm{d}t \ \mathrm{d}\mu(x)\mathrm{d}\mu(y),$$

then as $T \to \infty$ this converges to $H(\mu)$ by the monotone convergence theorem (notice that the integrand is always of fixed sign, but which sign depends on $|x-y|$). Now use Fubini to interchange the order of integration,

$$H_T(\mu) = \int_{1/T}^T \frac{e^{-1/2t}}{2t} \ \mathrm{d}t - \int_{1/T}^T \frac{1}{2t}\left[\iint \exp\left(-\frac{(x-y)^2}{2t}\right) \mathrm{d}\mu(x)\mathrm{d}\mu(y)\right] \ \mathrm{d}t.$$

To decouple x and y, we use the well-known Fourier integral identity

$$\exp\left(-\frac{(x-y)^2}{2t}\right) = \sqrt{\frac{t}{2\pi}} \int_{-\infty}^\infty e^{i\omega x} e^{-i\omega y} e^{-t\omega^2/2} \ \mathrm{d}\omega.$$

Inserting this results in

$$H_T(\mu) = \int_{1/T}^{T} \frac{e^{-1/2t}}{2t}\,\mathrm{d}t - \int_{1/T}^{T} \sqrt{\frac{t}{2\pi}} \int_{-\infty}^{\infty} e^{-t\omega^2/2} \left| \int e^{i\omega x}\,\mathrm{d}\mu(x) \right|^2 \mathrm{d}\omega \mathrm{d}t,$$

which is concave, so that also the limit function H is concave. Strict concavity can also be established, first one checks that

$$H\left(\alpha\mu_1 + (1-\alpha)\mu_2\right) - \left[\alpha H(\mu_1) + (1-\alpha)H(\mu_2)\right] = \alpha(\alpha-1)H(\mu_1 - \mu_2), \tag{5.2.1}$$

for any $\alpha \in [0,1]$ and $\mu_1, \mu_2 \in \mathscr{M}_1^+(\mathbb{R})$ such that $H(\mu_1)$ and $H(\mu_2)$ are finite. Then one shows using H_T that

$$H(\mu_1 - \mu_2) = -\int_0^\infty \frac{1}{2t} \left[\iint \exp\left(-\frac{(x-y)^2}{2t} \right)\,\mathrm{d}(\mu_1 - \mu_2)(x)\,\mathrm{d}(\mu_1 - \mu_2)(y) \right]\,\mathrm{d}t,$$

which is clearly strictly negative for $\mu_1 \neq \mu_2$. Thus the RHS of (5.2.1) must be nonzero for all $\alpha \in (0,1)$, and strict concavity follows. □

Skipping ahead a little bit, the following result gives necessary and sufficient conditions for a measure μ to minimise $\mathcal{I}$. The proof will be given after exploring the consequences of this characterisation (see Theorem 5.9), but one can think about it as follows. Suppose that μ were such that it makes F attain its infimum value (and assume that this value is finite), then we would have

$$F_{\inf} = \frac{1}{2} \iint \left[\frac{1}{2}x^2 + \frac{1}{2}y^2 - 2\log|x-y| \right]\,\mathrm{d}\mu(x)\mathrm{d}\mu(y)$$

where $F_{\inf} = \inf_{\nu \in \mathscr{M}_1^+(\mathbb{R})} F(\nu)$. Rearranging terms, and using that $\int \mathrm{d}\mu(x) = 1$, this can be written as

$$\int \left[2\int \log|x-y|\,\mathrm{d}\mu(y) \right]\,\mathrm{d}\mu(x) = \int \left[\frac{1}{2}x^2 + \int \frac{1}{2}y^2\,\mathrm{d}\mu(y) - 2F_{\inf} \right]\,\mathrm{d}\mu(x).$$

This suggests that we could ignore one of the integrals when looking for a minimiser, and indeed this is true.

Lemma 5.4. *μ is a minimiser of $\mathcal{I}$ if and only if for μ-almost all x*

$$2\int \log|x-y|\,\mathrm{d}\mu(y) = \frac{1}{2}x^2 + \int \frac{1}{2}y^2\,\mathrm{d}\mu(y) - 2F_{\inf}.$$

Before we can continue, we also need the following fact.

Lemma 5.5. *The value of $F_{\inf}$ is $\frac{3}{4}$, so that the minimal value of $\mathcal{I}$ is zero. Moreover, there is a unique measure that attains this value.*

Proof. This follows from [7] (see Remark 2.2), where it is proved that the interpretation of $\mathcal{I}$ as a rate function can be made rigorous, and thus that $\mathcal{I} \geq 0$. The fact that there exists a measure that attains this value follows from the fact that $\mathscr{M}_1^+(\mathbb{R})$ is compact in the weak-$*$ topology induced by $C_b(\mathbb{R})$. Uniqueness follows from the strict convexity of $\mathcal{I}$. □

We can now simply check that the standard semicircle law σ satisfies the condition of Lemma 5.4.

Proposition 5.6. *For all $x \in [-2,2]$, we have that*

$$2\int \log|x-y|\,\mathrm{d}\sigma(y) = \frac{1}{2}x^2 - 1.$$

Proof. We invoke some techniques from the theory of distributions. It is well known that $\log|x|$ defines a tempered distribution, whose derivative is the distribution P.V.$\frac{1}{x}$. Thus we may take the derivative in the sense of distributions to find that

$$\partial_x \; 2\int \log|x-y| \; \mathrm{d}\sigma(y) = 2 \; \text{P.V.} \int \frac{\mathrm{d}\sigma(y)}{x-y}.$$

We have already computed this integral in Example 4.14, and the result was x, which is precisely the derivative of the right hand side of our desired relation. This implies that there exists a real constant C such that

$$2\int \log|x-y| \; \mathrm{d}\sigma(y) = \frac{1}{2}x^2 + C, \tag{5.2.2}$$

and the proof is complete if we manage to show that $C = -1$.

For this, evaluate (5.2.2) in $x = 0, \pm 2$, and change variables to $y = 2\sin\vartheta$ to find the conditions

$$\begin{aligned}
C &= 2\log 2 + \frac{8}{\pi}\int_0^{\pi/2} \cos^2(\vartheta)\log\sin(\vartheta) \; \mathrm{d}\vartheta, \\
C &= 2\log 2 - 2 + \frac{4}{\pi}\int_{-\pi/2}^{\pi/2} \cos^2(\vartheta)\log(1-\sin\vartheta) \; \mathrm{d}\vartheta, \\
C &= 2\log 2 - 2 + \frac{4}{\pi}\int_{-\pi/2}^{\pi/2} \cos^2(\vartheta)\log(1+\sin\vartheta) \; \mathrm{d}\vartheta.
\end{aligned}$$

Taking the average of the last two conditions and shifting the cosines into sines yields

$$C = 2\log 2 - 2 + \frac{8}{\pi}\int_0^{\pi/2} \sin^2(\vartheta)\log\sin(\vartheta) \; \mathrm{d}\vartheta,$$

which can be added to the first condition, so that

$$2C = 4\log 2 - 2 + \frac{8}{\pi}\int_0^{\pi/2} \log\sin(\vartheta) \; \mathrm{d}\vartheta.$$

The integral is in fact the famous Euler log-sine integral, and it can be evaluated using elementary means. Using some symmetry arguments

$$\mathcal{E} := \int_0^{\pi/2} \log\sin(\vartheta) \; \mathrm{d}\vartheta = \int_0^{\pi/2} \log\cos(\vartheta) \; \mathrm{d}\vartheta,$$

so that

$$\mathcal{E} = \frac{1}{2}\int_0^{\pi/2} \log[\sin(\vartheta)\cos(\vartheta)] \; \mathrm{d}\vartheta = \frac{1}{2}\int_0^{\pi/2} \log\left[\frac{1}{2}\sin(2\vartheta)\right] \; \mathrm{d}\vartheta = \frac{1}{2}\mathcal{E} - \frac{\pi}{4}\log 2,$$

whence $\mathcal{E} = -\frac{\pi}{2}\log 2$. Plugging this back in

$$2C = 4\log 2 - 2 - \frac{8}{\pi}\frac{\pi}{2}\log 2 = -2,$$

which is the desired result. □

It remains to prove Lemma 5.4. This will follow from a more general result that we prove below. To state it, we need some additional terminology.

Definition 5.7. By an *admissible weight function* (AWF) w, we mean a continuous function $w : \mathbb{R} \to (0, \infty)$ such that $|x|w(x) \to 0$ as $|x| \to \infty$. Any AWF w defines a *potential* $V(x) := \log[1/w(x)]$, and a functional

$$\mathcal{I}^w(\mu) := \iint \Big[V(x) + V(y) - \log|x - y| \Big] \, \mathrm{d}\mu(x)\mathrm{d}\mu(y)$$

on $\mathscr{M}_1^+(\mathbb{R})$. Write $\mathcal{I}^w_{\inf}$ for the infimum of $\mathcal{I}^w$ over $\mathscr{M}_1^+(\mathbb{R})$.

Remark 5.8. Notice that our rate functional $\mathcal{I}$ fits into this framework by choosing the AWF $x \mapsto \exp(-x^2/4)$.

The remainder of this section will be dedicated to proving the following theorem, which is a special case of Theorem 2.3 in [55]. Notice that combining parts (ii) and (iii) gives Lemma 5.4.

Theorem 5.9. *Let w be an AWF, then the following are true.*

(i) $\mathcal{I}^w_{\inf}$ is a finite number, and there exists a unique measure μ^w in $\mathscr{M}_1^+(\mathbb{R})$, whose support is compact and of positive logarithmic capacity, such that $\mathcal{I}^w(\mu^w) = \mathcal{I}^w_{\inf}$.

(ii) For μ^w-almost all $x \in \mathbb{R}$,

$$\int \log|x - y| \, \mathrm{d}\mu^w(y) \leq V(x) - \mathcal{I}^w_{\inf} + \int V(y) \, \mathrm{d}\mu^w(y). \tag{5.2.3}$$

(iii) For all x in the support of μ^w,

$$\int \log|x - y| \, \mathrm{d}\mu^w(y) \geq V(x) - \mathcal{I}^w_{\inf} + \int V(y) \, \mathrm{d}\mu^w(y). \tag{5.2.4}$$

Proof. For part (i), introduce the function

$$f(x, y) = V(x) + V(y) - \log|x - y| = -\log\left[w(x)w(y)|x - y|\right],$$

which is bounded from below since $|x|w(x)$ goes to zero at infinity. Thus $\mathcal{I}^w$ exists as an extended real number (allowing $+\infty$), and the number $\mathcal{I}^w_{\inf}$ is finite (consider the uniform law on $[0, 1]$ to see that there exists a measure for which $\mathcal{I}^w < \infty$). Next, we show that it suffices to consider measures with compact support to find a minimiser. Naturally, we can restrict our attention to measures μ for which $\mathcal{I}^w_{\inf} + \varepsilon > \mathcal{I}^w(\mu)$ for some $\varepsilon > 0$. We know that f is large when $|x|$ and $|y|$ are both large (again by the fact that $|x|w(x)$ goes to zero at infinity), so that there must exist an integer $N > 0$ with the property that $\mathcal{I}^w_{\inf} + \varepsilon < f(x, y)$ on the complement of $[-N, N]^2$. Assume that μ has unbounded support, so that its intersection with $[-N, N]^c$ is non-empty. Then we can estimate

$$\begin{aligned}
0 &> \mathcal{I}^w(\mu) - (\mathcal{I}^w_{\inf} + \varepsilon), \\
&= \iint (f - (\mathcal{I}^w_{\inf} + \varepsilon)) \, \mathrm{d}(\mu \otimes \mu), \\
&= \iint_{[-N,N]^2} (f - (\mathcal{I}^w_{\inf} + \varepsilon)) \, \mathrm{d}(\mu \otimes \mu) + \iint_{([-N,N]^2)^c} (f - (\mathcal{I}^w_{\inf} + \varepsilon)) \, \mathrm{d}(\mu \otimes \mu), \\
&> \mu([-N, N])^2 \left[\mathcal{I}^w(\mu_N) - (\mathcal{I}^w_{\inf} + \varepsilon)\right], \\
&\geq \left[\mathcal{I}^w(\mu_N) - (\mathcal{I}^w_{\inf} + \varepsilon)\right],
\end{aligned}$$

where μ_N is the probability measure with support $[-N, N]$ given by $\mu|_{[-N,N]}/\mu([-N, N])$. Note that $\mu([-N, N]) = 0$ is impossible, as this would imply that the second integral in the third line above is

negative, contradicting the fact that on the complement of $[-N, N]$ the function $f(x, y) - (\mathcal{I}^w_{\text{inf}} + \varepsilon)$ is strictly positive. However, according to this estimate, $\mathcal{I}^w(\mu_N) < \mathcal{I}^w(\mu)$, thus any minimising measure must be compactly supported (it must even be supported in $[-N, N]$). Existence and uniqueness of a minimising measure follows from similar arguments as before (or see [55]). The fact that the support of the minimising measure μ^w must have positive logarithmic capacity, is due to the realisation that it has compact support (so the integrals of the potentials V exist), has a finite value under $\mathcal{I}^w$ (so the log part must be finite), and a close look at the definition of γ.

For part (ii) (and also (iii)), we will only sketch the main line of the argument, all the details that we leave out can be found in Chapter III of [84] and [55]. Begin by defining

$$U^w(x) := \int \log|x - y| \, \mathrm{d}\mu^w(y) - V(x),$$

then it is easy to show, using pointwise approximation by functions bounded from above, that U^w is an upper semi-continuous proper function (allowing the value $-\infty$). This means that all the sets of the form $\{U^w(x) \geq \alpha\}$ for $\alpha \in \mathbb{R}$ are closed. We now argue by contradiction, so assume that the set on which (5.2.3) is violated has positive logarithmic capacity. Then there must exist a positive integer n_0 with the property that the set

$$E_1 := \left\{ U^w(x) \geq \int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} + \frac{1}{n_0} \right\} \cap [-n_0, n_0],$$

has positive logarithmic capacity and is compact (E_1 is bounded by construction, and closed by upper semi-continuity of U^w). This is a useful set to define, as $\int U^w \, \mathrm{d}\mu^w$ must equal $\int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}}$, so the idea is that the set E_1 cannot be too large, in fact it must be so small that it has vanishing logarithmic capacity, leading to a contradiction. To see this, notice that there must exist another compact set E_2, disjoint from E_1, satisfying $\mu^w(E_2) > 0$ and

$$U^w(x) < \int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} + \frac{1}{2n_0},$$

for all $x \in E_2$. As E_1 is compact with positive logarithmic capacity, there exists a measure ν on it with the property that $\mathcal{I}^w(\nu)$ is finite, and we can scale ν such that $\nu(E_1) = \mu^w(E_2)$. Now construct a signed measure ν' on $\mathbb{R}$ by demanding that $\nu' = \nu$ on E_1, $\nu' = -\mu^w$ on E_2, and $\nu' = 0$ else. Thus the total mass of ν' is zero, and when added to μ^w (multiplied by an $\varepsilon > 0$ sufficiently small) it represents moving mass from E_2 to E_1. Naturally, this leads to a contradiction as now one can check that $\mathcal{I}^w(\mu^w + \varepsilon\nu') < \mathcal{I}^w_{\text{inf}}$. It follows from the arguments used to prove point (ii) of Proposition 5.3 that (5.2.3) must even hold μ^w-almost everywhere.

For part (iii), let x_0 be in the support of μ^w and suppose that (5.2.4) is violated at this point. By upper semi-continuity of U^w, there then has to be an open neighbourhood of x_0 on which (5.2.4) is violated with some $\varepsilon > 0$ of room, denote the intersection of this neighbourhood with the support of μ^w by E. Notice that E has positive μ^w-measure, as it is an open subset of its support. Using (5.2.3), we have

$$\begin{aligned}
\int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} &= \int U^w \, \mathrm{d}\mu^w, \\
&\leq \mu^w(E) \left[\int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} - \varepsilon \right] + (1 - \mu^w(E)) \left[\int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} \right], \\
&= \int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}} - \varepsilon\mu^w(E), \\
&< \int V \, \mathrm{d}\mu^w - \mathcal{I}^w_{\text{inf}},
\end{aligned}$$

which is a contradiction. This concludes the proof. □

Remark 5.10. It is a testament to the merit of the definition of the microstates free entropy that it forces the 'interesting' measures to have compact support, as these are the only ones that can possibly arise within free probability as we have developed it here.

5.3 Commutative Free Monotone Transport

At last we have developed enough techniques to explain the commutative case of free monotone transport (Section 1.6 in [43]). Throughout this section, we will as usual denote the standard Wigner semicircle measure by σ, and X and Y denote random variables. In the case of a single standard semicircular random variable X, we know that the generated von Neumann algebra can be viewed as acting on the Hilbert space $L^2([-2,2],\sigma)$ by identifying X with multiplication by the identity function x. In this case the state τ can be represented by the formula

$$\tau(h) = \int h \, \mathrm{d}\sigma,$$

with h an L^∞ function on $[-2,2]$. $L^\infty([-2,2],\sigma)$ is also isomorphic to the von Neumann algebras $\mathcal{L}(\mathbb{F}_1) \cong \mathcal{L}(\mathbb{Z}) \cong \Gamma_0(\mathbb{R})$, as we have shown in the chapter on q-Gaußians. In this $n = 1$ case, we in fact have that the isomorphism class of the q-Gaußians does not change at all when varying q in $(-1,1)$. This is due to our knowledge that the analytic distribution of such a 'q-semicircular element' is diffuse for these q. In this section, we sketch a different proof, which generalises to $n > 1$, which will be the content of the next chapter.

For now, we need a convenient way to characterise the standard semicircular measure σ. In the preceding sections, we discussed the form of the free entropy in the commutative setting, and saw that it was related to a rate function. We proved that σ was the unique minimiser of this rate function, and we even saw that analogous statements were true for a much larger class of functionals. To exploit this, we define

$$\mathcal{X}(\mu) = \iint \log|y-x| \, \mathrm{d}\mu(x)\mathrm{d}\mu(y) - \int \frac{1}{2}x^2 \, \mathrm{d}\mu(x),$$

which is essentially the free entropy minus the expectation of the potential. This is the negative of the rate function we considered earlier, so in this case our arguments carry over to prove that σ is the unique maximiser of $\mathcal{X}$. We now slightly deform this functional, by choosing an analytic function W on a disk of radius $A > 4$, and setting $V(x) = x^2/2 + W(x)$. It follows from a variation on Theorem 5.9 that the existence of a unique maximiser is unspoilt by such a perturbation, and thus there also exists a unique (compactly supported, etc.) measure, which we shall denote by σ_V, that maximises

$$\mathcal{X}_V(\mu) = \iint \log|y-x| \, \mathrm{d}\mu(y)\mathrm{d}\mu(x) - \int V(x)\mathrm{d}\mu(x).$$

This also gives us a new state

$$\tau_V(h) = \int h \, \mathrm{d}\sigma_V.$$

We would now like to prove a result analogous to Brenier's Theorem 1.21, that allows us to conclude that for W a small enough perturbation (in some appropriate norm) there exists an analytic function F (on some disk with radius $A' > 4$ but possibly smaller than A), such that $Y = F(X)$ has precisely σ_V as its distribution. The map F is then the analogue of the Brenier map arising from the aforementioned theorem, as it satisfies $F_*\sigma = \sigma_V$, which can be written in terms of integrals as the equality

$$\int h \circ F \, \mathrm{d}\sigma = \int h \, \mathrm{d}\sigma_V,$$

for all h. It will also turn out that F is indeed the derivative of a convex function. The remainder of this section will be dedicated to sketching a proof, which will then be generalised and made fully rigorous in a later chapter. All integrals whose integrands have singularities are to be taken as principal value integrals.

To begin, we exploit the fact that σ_V maximises $\mathcal{X}_V$ in the following way. If we let f be some nice function, then we can consider the assignment

$$\delta \mapsto \mathcal{X}_V\left((x+\delta f)_*\sigma_V\right)$$

for δ real numbers in some neighbourhood of zero. Since we have a maximum for $\delta = 0$, its derivative there should vanish (provided it exists of course). This derivative can be computed by finding the terms linear in δ in

$$\begin{aligned}\mathcal{X}_V\left((x+\delta f)_*\sigma_V\right) &= \iint \log|y+\delta f(y)-x-\delta f(x)|\,\mathrm{d}\sigma_V(y)\mathrm{d}\sigma_V(x) - \int V(x+\delta f(x))\,\mathrm{d}\sigma_V(x),\\ &= \mathcal{X}_V(\sigma_V) + \delta\left[\iint \frac{f(x)-f(y)}{|x-y|}\frac{x-y}{|x-y|}\mathrm{d}\sigma_V(y)\mathrm{d}\sigma_V(x) - \int f(x)V'(x)\,\mathrm{d}\sigma_V(x)\right] + \mathcal{O}(\delta^2),\end{aligned}$$

where we employed several power series expansions around $\delta = 0$. This implies that σ_V must satisfy the following for any f, namely

$$\iint \frac{f(x)-f(y)}{x-y}\mathrm{d}\sigma_V(y)\mathrm{d}\sigma_V(x) = \int f(x)V'(x)\mathrm{d}\sigma_V(x),$$

which is called a *Schwinger-Dyson* equation. On the left we recognise the finite difference derivative of f, which we also encountered when discussing non-commutative derivatives in Section 4.1. It thus seems profitable to introduce the notations

$$\mathscr{J}f(x,y) = \frac{f(x)-f(y)}{x-y},\ \mathscr{D}V(x) = V'(x),$$

which play the role of a free Jacobian and a free gradient respectively. Recall that the non-commutative derivatives were to be viewed as derivations from some space of non-commutative polynomials to the tensor product of this space with itself. Hence for the second integral it makes sense to identify $L^\infty([-2,2]\times[-2,2],\sigma_V\otimes\sigma_V)$ with the closure of the tensor product of $L^\infty([-2,2],\sigma_V)$ with itself and rewrite the equation abstractly as

$$(\tau_V\otimes\tau_V)\left(\mathscr{J}f\right) = \tau_V\left(f\mathscr{D}V\right).$$

These are very similar to the conjugate relations (4.2.1). One can prove that for W small enough, this equation fully and uniquely determines τ_V and hence also σ_V. This equation is satisfied precisely when

$$\mathscr{J}^*(1\otimes 1) = \mathscr{D}V,$$

so that we should try to compute the adjoint of $\mathscr{J}^*$ acting on $1\otimes 1$. To this end, expand the definitions

$$\begin{aligned}(\tau_V\otimes\tau_V)\left(\mathscr{J}f\right) &= (\tau_V\otimes\tau_V)\left((1\otimes 1)\mathscr{J}f\right),\\ &= \iint \frac{f(x)-f(y)}{x-y}\mathrm{d}\sigma_V(y)\mathrm{d}\sigma_V(x),\\ &= \int\left(\int \frac{1}{x-y}\mathrm{d}\sigma_V(y)\right)f(x)\,\mathrm{d}\sigma_V(x) + \int\left(\int \frac{1}{y-x}\mathrm{d}\sigma_V(x)\right)f(y)\,\mathrm{d}\sigma_V(y),\\ &= 2\int\left(\int \frac{\mathrm{d}\sigma_V(y)}{x-y}\right)f(x)\,\mathrm{d}\sigma_V(x),\\ &= 2\tau_V\left(f\mathscr{J}^*(1\otimes 1)\right)\end{aligned}$$

where in the third step we changed the order of integration in the second integral, and in the last line we relabelled the dummy integration variables x and y. Thus we conclude that

$$\mathscr{J}^*(1 \otimes 1) = 2 \int \frac{\mathrm{d}\sigma_V(y)}{x - y},$$

which is (up to constants) the Hilbert transform of the probability measure σ_V. Thus we have derived that the measure σ_V maximises the functional $\mathcal{X}_V$ if and only if it is a solution of the equation

$$2 \int \frac{\mathrm{d}\sigma_V(y)}{x - y} = V'(x),$$

for all x in the support of σ_V. It is illuminating to compare this to the result of Proposition 4.13. This casts new light on the role of the potential V, its derivative determines the conjugate variable to the element whose distribution maximises $\mathcal{X}_V$.

We now introduce our transport map. Suppose that we have an analytic function F (on a large enough disk) such that $F_*\sigma = \sigma_V$, then we can rewrite the previous equation as

$$V'(F(x)) = F(x) + W'(F(x)) = 2 \int \frac{\mathrm{d}\sigma(y)}{F(x) - F(y)},$$

for all $x \in [-2, 2]$. One can interpret this as saying that the map F (as a function) deforms the conjugate variable to the standard semicircular element into the conjugate variable of whatever element maximises $\mathcal{X}_V$. We make the Ansatz $F(x) = x + f(x)$, with the reasoning behind this being that the size of f should be controllable through the size of W. This yields

$$\begin{aligned} x + f(x) + W'(x + f(x)) &= 2 \int \frac{\mathrm{d}\sigma(y)}{x + f(x) - y - f(y)}, \\ &= 2 \int \frac{x - y}{(x - y) + f(x) - f(y)} \frac{1}{x - y} \mathrm{d}\sigma(y), \\ &= 2 \int \frac{1}{1 + \mathscr{J} f(x, y)} \frac{\mathrm{d}\sigma(y)}{x - y}. \end{aligned} \tag{5.3.1}$$

We are now in the position to do something rather clever, as we have computed in a previous section that

$$x = 2 \int \frac{\mathrm{d}\sigma(y)}{x - y},$$

for all $x \in [-2, 2]$, we can use this to replace the leftmost x occurring in the equation above. Doing this, and bringing it over to the other side, we now have

$$\begin{aligned} f(x) + W'(x + f(x)) &= 2 \int \left[\frac{1}{1 + \mathscr{J} f(x, y)} - 1 \right] \frac{\mathrm{d}\sigma(y)}{x - y}, \\ &= -2 \int \frac{\mathscr{J} f(x, y)}{1 + \mathscr{J} f(x, y)} \frac{\mathrm{d}\sigma(y)}{x - y}. \end{aligned}$$

Our next attempt at simplification will target both the derivative of W on the left and the finite difference derivatives on the right. The obvious object to connect the two is the derivative of f. We will do this by multiplying both sides (next to the equation number and the left side) of (5.3.1) by $f'(x)$, and adding the left part to the the left part of the previous equation, idem for the right part. This results in

$$\begin{aligned} 2 \int \frac{f'(x) - \mathscr{J} f(x, y)}{1 + \mathscr{J} f(x, y)} \frac{\mathrm{d}\sigma(y)}{x - y} &= f(x) + W'(x + f(x)) + x f'(x) + f(x) f'(x) + W'(x + f(x)) f'(x), \\ &= \frac{\mathrm{d}}{\mathrm{d}x} \left[x f(x) + \frac{1}{2} f(x)^2 + W(x + f(x)) \right]. \end{aligned}$$

For this equation to contain the same information as the last, we need to ensure that $\mathscr{J} f(x,y) = -1$ only occurs on a σ-null set, or equivalently almost nowhere as σ is Lebesgue absolutely continuous. A convenient way to arrange this is to demand that $f'(x) \neq -1$ almost everywhere, as then the mean value theorem prevents the finite difference derivative from taking on this value too often as well. The natural next step is to also extract an x-derivative out of the left side of the previous equation. It turns out that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\log(1+\mathscr{J} f(x,y)) &= \frac{1}{1+\mathscr{J} f(x,y)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)-f(y)}{x-y}\right),\\ &= \frac{1}{1+\mathscr{J} f(x,y)}\left(\frac{f'(x)}{x-y}-\frac{1}{x-y}\frac{f(x)-f(y)}{x-y}\right),\\ &= \frac{f'(x)-\mathscr{J} f(x,y)}{1+\mathscr{J} f(x,y)}\frac{1}{x-y}.\end{aligned}$$

Now we can integrate both sides of our equation to get rid of the derivatives, and we arive at

$$xf(x)+\frac{1}{2}f(x)^2+W(x+f(x))+C = 2\int \log(1+\mathscr{J} f(x,y))\,\mathrm{d}\sigma(y), \tag{5.3.2}$$

where C is a constant of integration.

For the final few steps, we make the additional Ansatz that $f = g'$ and rewrite the equation a little bit,

$$xg'(x) = -W(x+g'(x)) - \frac{1}{2}\left(g'(x)\right)^2 + 2\int \log\left(1+\frac{g'(x)-g'(y)}{x-y}\right)\mathrm{d}\sigma(y) - C.$$

It now makes sense to fix C by demanding that both sides of the equation are identically zero for $x = 0$. We now introduce two new operators, both acting on commutative polynomials. First, define $\mathscr{N}$ by $\mathscr{N} g = xg'$, which has a particularly simple action on the basis of monomials, as it just multiplies a monomial by its degree. Hence the symbol $\mathscr{N}$ for 'number operator'. If we consider the subspace of polynomials without constant term, then $\mathscr{N}$ can be inverted, and we denote this inverse by $\mathscr{S}$. It is not difficult to derive the explicit formula

$$\mathscr{S}(\hat{g})(x) = \int_0^x \frac{\hat{g}(y)}{y}\mathrm{d}y,$$

for $\mathscr{S}$. Now write $\hat{g} = \mathscr{N} g$, so that $g = \mathscr{S}\hat{g}$. This allows us to rephrase our equation in the following way,

$$\hat{g}(x) = -W(x+(\mathscr{S}\hat{g})'(x) - \frac{1}{2}\left((\mathscr{S}\hat{g})(x)\right)^2 + 2\int \log\left(1+\frac{(\mathscr{S}\hat{g})'(x)-(\mathscr{S}\hat{g})'(y)}{x-y}\right) - C.$$

For each fixed x, this has the form of a fixed-point equation

$$\hat{g}(x) = \Psi_x\left(\hat{g}(x)\right),$$

with $\Psi_x(0) = W(x)$. In the sequel, what we will essentially prove is that Ψ is a Lipschitz function in $\hat{g}$ such that the fix-point equation allows for a solution by iteration. Our transport map F will then read $F = x + g'$, or defining $G(x) = x^2/2 + g(x)$, $F = G'$. Notice that G is a convex function for W small enough, as all functions are analytic and convexity admits an open characterisation in terms of the second derivative being non-negative, so that our transport map is indeed monotone.

To further explore the analogy between the procedure above and optimal transportation theory, we show that what we have just derived is essentially a free version of the Monge-Ampère equation.

Recall from Section 1.4 that the Monge-Ampère equation for two measures admitting densities m and n respectively was

$$\det((J\nabla\varphi)(x)) = \frac{m(x)}{n(\nabla\varphi(x))},$$

where J is now the commutative Jacobian. If we suppose that the densities are given by

$$m(x) = \exp\left(-\frac{x^2}{2}\right),\ n(x) = \exp(-V(x)),$$

then the Monge-Ampère equation becomes

$$\det((J\nabla\varphi)(x)) = \exp\left(V(\nabla\varphi(x)) - \frac{x^2}{2}\right). \tag{5.3.3}$$

The reason that these densities are sensible, is that the laws σ_V in some sense arise as limits of $N \to \infty$ of measures on the space of $N \times N$ matrices given by the densities $\exp(-N \operatorname{Tr} V)$. Now if we notice in (5.3.2) that $\mathscr{J} F = 1 + \mathscr{J} f$, then this becomes

$$2\int \log(\mathscr{J} F(x,y))\ \mathrm{d}\sigma(y) = V(F(x)) - \frac{1}{2}x^2,$$

which is remarkably similar to (5.3.3).

6 Free Monotone Transport

6.1 Noncommutative Power Series and More on Noncommutative Derivatives

This chapter serves to explain the technical heart of [43], Chapters 2 and 3. In this first section we develop some machinery to deal with 'noncommutative analytic functions'. Throughout this chapter, we set $\mathscr{A} := \mathbb{C}\langle X_1, \ldots, X_n\rangle$, the unital algebra of polynomials in n noncommutative formal indeterminates, and we write $\mathscr{A}_0 \subset \mathscr{A}$ for the subalgebra of such polynomials without constant term.

Definition 6.1. For all monomials $q \in \mathscr{A}$ we define the functional $\gamma_q : \mathscr{A} \to \mathbb{C}$ to be the one sending $P \in \mathscr{A}$ to the coefficient of q in its decomposition into monomials, in other words $P = \sum_q \gamma_q(P)\, q$. Also define $\deg q$ to be the length of q, so $\deg q = k$ if $q = X_{i(1)} \cdots X_{i(k)}$ with $1 \leq i(1), \ldots, i(k) \leq n$. If $\rho > 1$, define the norm $\|\cdot\|_\rho$ on $\mathscr{A}$ by the formula

$$\|P\|_\rho = \sum_q |\gamma_q(P)| \rho^{\deg q}.$$

It is trivial to see that this is indeed a norm, and has the property that $\|PQ\|_\rho \leq \|P\|_\rho \|Q\|_\rho$. Thus, the completion of $\mathscr{A}$ with respect to $\|\cdot\|_\rho$, denoted $\mathscr{A}^{(\rho)}$, is a unital Banach algebra, called the algebra of *absolutely convergent noncommutative power series with radius of convergence at least* ρ. Likewise, we write $\mathscr{A}_0^{(\rho)}$ for the completion of $\mathscr{A}_0$, which is also a Banach algebra.

Remark 6.2. Under suitable conditions, we can also compose such noncommutative power series. For instance, if $Y_1, \ldots, Y_n \in \mathscr{A}^{(\rho)}$ are such that $\|Y_j\|_\rho \leq \rho$ for all j, then for any $P \in \mathscr{A}^{(\rho)}$ we can form the composition $P \circ Y$ which is the noncommutative power series obtained by replacing each X_j in P by the corresponding Y_j. This is again a member of $\mathscr{A}^{(\rho)}$, as

$$\|P \circ Y\|_\rho = \left\| \sum_q \gamma_q(P) q \circ Y \right\|_\rho \leq \sum_q |\gamma_q(P)| \|q \circ Y\|_\rho \leq \sum_q |\gamma_q(P)| \rho^{\deg q} = \|P\|_\rho.$$

By repeating this argument, we see that if $\|Y_j\|_{\rho'} \leq \rho'$ for some $1 < \rho' < \rho$ instead, the composition $P \circ Y$ is an element of $\mathscr{A}^{(\rho')}$. Notice also that in this case, $\|P\|_{\rho'} \leq \|P\|_\rho$, so that $\mathscr{A}^{(\rho)} \subset \mathscr{A}^{(\rho')}$. In general, if we have another Banach algebra $\mathscr{B}$, and elements $T_1, \ldots, T_n \in \mathscr{B}$ such that $\|T_j\|_{\mathscr{B}} \leq \rho$ for all j, we obtain a contractive mapping from $\mathscr{A}^{(\rho)}$ to $\mathscr{B}$, induced by sending X_j to the corresponding T_j. The relevant conclusion for our purposes, is that we may view the algebra $\mathscr{A}^{(\rho)}$ as containing elements that can be thought of as both 'functions' and 'variables'.

Definition 6.3. The *number operator* $\mathscr{N} : \mathscr{A} \to \mathscr{A}$ is defined on monomials q by $\mathscr{N} q = (\deg q)\, q$. Its kernel is precisely the span of the unit, so that $\mathscr{N}$ is invertible when seen as an operator on $\mathscr{A}_0$. Denote its inverse by $\mathscr{S}$, which explicitly acts on monomials as $\mathscr{S} q = (\deg q)^{-1}\, q$. Write $\mathscr{P}_0$ for the projection of $\mathscr{A}$ onto $\mathscr{A}_0$. Finally, $P \in \mathscr{A}$ is called *cyclically symmetric* iff for all monomials $q = X_{i(1)} \cdots X_{i(k)}$, the number $\gamma_q(P)$ is invariant under cyclic permutations of the $i(1), \ldots, i(k)$. The *cyclic symmetrisation operator* $\mathscr{C}$ on $\mathscr{A}_0$ acts on monomials as

$$\mathscr{C} q = \mathscr{C} X_{i(1)} \cdots X_{i(k)} = \frac{1}{k} \sum_{\ell=1}^{k} X_{i(\ell+1)} \cdots X_{i(k)} X_{i(1)} \cdots X_{i(\ell)}.$$

Clearly, $\mathscr{C} P$ is cyclically symmetric and $\|\mathscr{C} P\|_\rho \leq \|P\|_\rho$ for any $P \in \mathscr{A}_0$.

We now want to introduce some notions of derivatives on the algebras $\mathscr{A}$ and $\mathscr{A}^{(\rho)}$. Recall our discussion of noncommutative partial derivatives in Section 4.1. Here, we work with a slightly different definition, namely that we view the ∂_i as operators from $\mathscr{A}$ to $\mathscr{A} \otimes \mathscr{A}^{op}$, with $\mathscr{A}^{op}$ the opposite algebra of $\mathscr{A}$, equipped with the same norms $\|\cdot\|_\rho$, and $\otimes$ denoting the algebraic tensor product. As we would like a notion of derivative for which the derivative of a noncommutative power series is again a noncommutative power series (although possibly with a smaller radius of convergence), we make the following definitions.

Definition 6.4. The *cyclic derivatives* on $\mathscr{A}$ are the maps $\mathscr{D}_j : \mathscr{A} \to \mathscr{A}$, $1 \leq j \leq n$, such that on monomials q,

$$\mathscr{D}_j q = \sum_{q = AX_jB} BA,$$

where A and B are also monomials, or more explicitly, if $q = X_{i(1)} \cdots X_{i(k)}$,

$$\mathscr{D}_j q = \sum_{\ell=1}^{k} \delta_{ji(\ell)} X_{i(\ell+1)} \cdots X_{i(k)} X_{i(1)} \cdots X_{i(\ell-1)}.$$

The *cyclic gradient* $\mathscr{D}$ is the vector valued operator from $\mathscr{A}$ to $\mathscr{A}^n$, such that $\mathscr{D}P$ has j-th component $\mathscr{D}_j P$ for $P \in \mathscr{A}$ and $1 \leq j \leq n$. If $f \in \mathscr{A}^n$, that is $f = (f_1, \ldots, f_n)$ with all $f_\ell \in \mathscr{A}$, its *noncommutative Jacobian* $\mathscr{J} f \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$ is the matrix with i,j-th entry $\partial_j f_i$.

Remark 6.5. Another way to view these cyclic derivatives, is as the composition of noncommutative partial derivatives with the maps t and ι, where $t(a \otimes b) = b \otimes a \in \mathscr{A} \otimes \mathscr{A}^{op}$ and $\iota(a \otimes b) = ab \in \mathscr{A}$ for $a \otimes b \in \mathscr{A} \otimes \mathscr{A}^{op}$. To be precise, $\mathscr{D}_j = \iota \circ t \circ \partial_j$. It is not difficult to check that $\mathscr{D}\mathscr{C}P = \mathscr{D}P$ for all $P \in \mathscr{A}$. It is also easy to see that $\mathscr{D}_j 1 = 0$, so that $\mathscr{D}_j = \mathscr{D}_j \mathscr{P}_0$.

A good property of these cyclic derivatives, is that every 'analytic function' $F \in \mathscr{A}^{(\rho)}$ is 'infinitely cyclically differentiable' in any of the spaces $\mathscr{A}^{(\rho')}$ with $1 < \rho' < \rho$.

Proposition 6.6. *Let $F \in \mathscr{A}^{(\rho)}$ and $1 < \rho' < \rho$, then $\mathscr{D}_{j(1)} \cdots \mathscr{D}_{j(k)} F \in \mathscr{A}^{(\rho')}$ for all $k \in \mathbb{Z}_{\geq 1}$ and all $1 \leq j(1), \ldots, j(k) \leq n$.*

Proof. Let $q \in \mathscr{A}^{(\rho)}$ be a monomial, then

$$\|\mathscr{D}_j q\|_{\rho'} \leq \sum_{q = AX_jB} \|BA\|_{\rho'} = (\deg q)\ (\rho')^{\deg q - 1},$$

as the sum contains at most $\deg q$ terms, and the ρ'-norm of a degree k monomial is precisely $(\rho')^k$. By iterating the idea of this argument, it is easy to see that

$$\left\|\mathscr{D}_{j(1)} \cdots \mathscr{D}_{j(k)} q\right\|_{\rho'} \leq \left[\frac{(\deg q)!}{(\deg q - k)!}\right] (\rho')^{\deg q - k},$$

where we assume that $k \leq \deg q$, as otherwise the iterated derivative becomes zero. Therefore, by the triangle inequality,

$$\begin{aligned}
\left\|\mathscr{D}_{j(1)} \cdots \mathscr{D}_{j(k)} F\right\|_{\rho'} &\leq \sum_{q \text{ s.t. } \deg q \geq k} |\gamma_q(F)| \left[\frac{(\deg q)!}{(\deg q - k)!}\right] (\rho')^{\deg q - k}, \\
&= (\rho')^{-k} \sum_{q \text{ s.t. } \deg q \geq k} |\gamma_q(F)| \left[\frac{(\deg q)!}{(\deg q - k)!} \left(\frac{\rho'}{\rho}\right)^{\deg q}\right] \rho^{\deg q}, \\
&\leq \max\left\{C(\rho, \rho', k), 1\right\} \|F\|_\rho,
\end{aligned}$$

where

$$C(\rho, \rho', k) = (\rho')^{-k} \sup_{\ell \geq k} \left[\frac{\ell!}{(\ell - k)!} \left(\frac{\rho'}{\rho} \right)^{\ell} \right].$$

This constant is finite, because $\rho'/\rho < 1$, and up to constants, the sequence over which we take the supremum are the coefficients of the k-th formal derivative of the (complex) power series $\sum_k z^k$, which has radius of convergence equal to unity (see Theorem 5.1 in Chapter II, §5 of [49]). Alternatively, one can prove by direct inspection that this sequence tends to zero, and so must be bounded. □

Since we have now also introduced the spaces $\mathscr{A}^n$, $\mathscr{A} \otimes \mathscr{A}^{op}$, and even $M_n(\mathscr{A} \otimes \mathscr{A}^{op})$ into the mix, we should devise some ways of combining elements of these spaces, such as notions of matrix multiplication, inner products, et cetera.

Definition 6.7.

(i) On $\mathscr{A} \otimes \mathscr{A}^{op}$, we of course have the multiplication $(a \otimes b).(c \otimes d) = (ac) \otimes (db)$, denoted by a period.

(ii) Combined with 'ordinary' matrix multiplication, this gives a multiplication on $M_n(\mathscr{A} \otimes \mathscr{A}^{op})$, also denoted by a period. Explicitly, if $A, B \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$, then the i, j-th component of $A.B$ is $\sum_{k=1}^n A_{ik}.B_{kj}$.

(iii) We also introduce $\mathrm{Tr} : M_n(\mathscr{A} \otimes \mathscr{A}^{op}) \to \mathscr{A} \otimes \mathscr{A}^{op}$ by putting $\mathrm{Tr}(A) = \sum_{i=1}^n A_{ii}$.

(iv) If $f, g \in \mathscr{A}^n$, we define their *dot product* as $f \cdot g = \sum_{j=1}^n f_j g_j \in \mathscr{A}$.

(v) We now define an action $\#$ of $\mathscr{A} \otimes \mathscr{A}^{op}$ on $\mathscr{A}$, by demanding that $(a \otimes b)\#c = acb$ for all $a \otimes b \in \mathscr{A} \otimes \mathscr{A}^{op}$ and $c \in \mathscr{A}$, and extending linearly.

(vi) This induces $\# : M_n(\mathscr{A} \otimes \mathscr{A}^{op}) \times \mathscr{A}^n \to \mathscr{A}^n$ with $(A\#f)_j = \sum_{i=1}^n A_{ji}\#f_i$.

Having endowed these spaces with algebraic structures, we turn to norms. The straightforward extension of $\|\cdot\|_\rho$ to $\mathscr{A}^n$ is given by $\|f\|_\rho = \max_{1 \leq j \leq n} \|f_j\|_\rho$. It is less straightforward to come up with a useful norm on $\mathscr{A} \otimes \mathscr{A}^{op}$, where by 'useful' we mean that that $\#$ maps and matrix multiplication become contractive. In the next definition, we mimic the construction of the maximal C* tensor product norm.

Definition 6.8. Consider $\mathscr{A} \otimes \mathscr{A}^{op}$ and the function $\|\cdot\|_{\rho \otimes_\pi \rho}$ defined by

$$\left\| \sum_i a_i \otimes b_i \right\|_{\rho \otimes_\pi \rho} = \sup_\pi \left\| \pi \left(\sum_i a_i \otimes b_i \right) \right\|_{\mathscr{B}},$$

where the supremum is taken over all algebra morphisms $\pi : \mathscr{A} \otimes \mathscr{A}^{op} \to \mathscr{B}$, with $\mathscr{B}$ a Banach algebra with norm $\|\cdot\|_{\mathscr{B}}$, such that $\|\pi(a \otimes 1)\|_{\mathscr{B}} \leq \|a\|_\rho$ and $\|\pi(1 \otimes b)\|_{\mathscr{B}} \leq \|b\|_\rho$ for all $a, b \in \mathscr{A}$. Call this function the *projective tensor product norm*.

Proposition 6.9. *$\|\cdot\|_{\rho \otimes_\pi \rho}$ is a norm and induces a Banach algebra structure on $\mathscr{A} \otimes \mathscr{A}^{op}$.*

Proof. The triangle inequality is obvious, and thus

$$\left\| \sum_i a_i \otimes b_i \right\|_{\rho \otimes_\pi \rho} \leq \sum_i \|a_i \otimes b_i\|_{\rho \otimes_\pi \rho} = \sum_i \sup_\pi \|\pi(a_i \otimes 1)\pi(1 \otimes b_i)\|_{\mathscr{B}} \leq \sum_i \|a_i\|_\rho \|b_i\|_\rho < \infty,$$

where we used that $\|\cdot\|_{\mathscr{B}}$ is a Banach algebra norm and the boundedness assumption on π. The same argument also shows that the multiplication on $\mathscr{A} \otimes \mathscr{A}^{op}$ is continuous with respect to $\|\cdot\|_{\rho\otimes_\pi\rho}$. It follows that $\|\cdot\|_{\rho\otimes_\pi\rho}$ is a semi-norm, and so it remains to show that it is non-degenerate. For this, recall the existence of the minimal Banach algebra tensor product norm $\|\cdot\|_\lambda$ (see Section 2 of Chapter IV in [82]), which is given by

$$\left\|\sum_i a_i \otimes b_i\right\|_\lambda = \sup\left\{\left\|\sum_i a_i \otimes b_i\right\| \alpha \in (\mathscr{A})^*,\ \|\alpha\|_{\rho*} \leq 1;\ \beta \in (\mathscr{A}^{op})^*,\ \|\beta\|_{\rho*} \leq 1\right\}.$$

Denote the resulting Banach algebra by $\mathscr{A} \otimes_\lambda \mathscr{A}^{op}$ and consider π the inclusion map from $\mathscr{A} \otimes \mathscr{A}^{op}$ into this minimal tensor product. It is easy to see that this gives a valid π for the supremum in the definition of $\|\cdot\|_{\rho\otimes_\pi\rho}$. Moreover, this map is injective and $\|\cdot\|_\lambda$ is a norm, so that also $\|\cdot\|_{\rho\otimes_\pi\rho}$ is a norm. □

Remark 6.10. An argument similar to the proof of Proposition 6.6 shows that for any $f \in \mathscr{A}^{(\rho)}$, its j-th noncommutative partial derivative has finite $\rho' \otimes_\pi \rho'$-norm for all $1 < \rho' < \rho$.

Now extend this norm to $F \in (\mathscr{A} \otimes \mathscr{A}^{op})^n$ by $\|F\|_{\rho\otimes_\pi\rho} = \max_{1\leq j\leq n} \|F_j\|_{\rho\otimes_\pi\rho}$, and to $A \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$ by $\|A\|_{\rho\otimes_\pi\rho} = \max_{1\leq j\leq n} \sum_{i=1}^n \|A_{ji}\|_{\rho\otimes_\pi\rho}$, then we get the continuity results we are after.

Lemma 6.11.

(i) For any $F \in \mathscr{A} \otimes \mathscr{A}^{op}$ and $c \in \mathscr{A}$, $\|F\#c\|_\rho \leq \|F\|_{\rho\otimes_\pi\rho}\|c\|_\rho$.

(ii) For any $A, B \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$, $\|A.B\|_{\rho\otimes_\pi\rho} \leq \|A\|_{\rho\otimes_\pi\rho}\|B\|_{\rho\otimes_\pi\rho}$.

(iii) For any $f \in \mathscr{A}^n$ and $A \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$, $\|A\#f\|_\rho \leq \|A\|_{\rho\otimes_\pi\rho}\|f\|_\rho$.

Proof. For part (i), notice that $F\# \cdot$ defines an operator on $\mathscr{A}^{(\rho)}$, and that the assignment $F \mapsto F\# \cdot$ is a valid map π for the projective tensor norm. Thus

$$\|F\#c\|_\rho \leq \|F\# \cdot\|_{B(\mathscr{A}^{(\rho)})}\|c\|_\rho \leq \|F\|_{\rho\otimes_\pi\rho}\|c\|_\rho.$$

For part (ii),

$$\begin{aligned}
\|A.B\|_{\rho\otimes_\pi\rho} &= \max_{1\leq j\leq n} \sum_{i=1}^n \left\|\sum_{k=1}^n A_{jk}.B_{ki}\right\|_{\rho\otimes_\pi\rho}, \\
&\leq \max_{1\leq j\leq n} \sum_{i=1}^n \sum_{k=1}^n \|A_{jk}\|_{\rho\otimes_\pi\rho}\|B_{ki}\|_{\rho\otimes_\pi\rho}, \\
&\leq \left(\max_{1\leq j\leq n} \sum_{k=1}^n \|A_{jk}\|_{\rho\otimes_\pi\rho}\right)\left(\max_{1\leq \ell\leq n} \sum_{i=1}^n \|B_{\ell i}\|_{\rho\otimes_\pi\rho}\right), \\
&= \|A\|_{\rho\otimes_\pi\rho}\|B\|_{\rho\otimes_\pi\rho}.
\end{aligned}$$

For part (iii),

$$\|A\#f\|_\rho \leq \max_{1\leq i\leq n} \sum_{j=1}^n \|A_{ij}\#f_j\|_\rho \leq \max_{1\leq i\leq n} \sum_{j=1}^n \|A_{ij}\|_{\rho\otimes_\pi\rho}\|f_j\|_\rho$$

by part (i), now estimate

$$\|A\#f\|_\rho \leq \left(\max_{1\leq i\leq n} \sum_{j=1}^n \|A_{ij}\|_{\rho\otimes_\pi\rho}\right)\left(\max_{1\leq k\leq n} \|f_k\|_\rho\right) = \|A\|_{\rho\otimes_\pi\rho}\|f\|_\rho.$$

□

We close this section with a lemma about some interactions between the various operations we have introduced so far, which will be useful later.

Lemma 6.12.

(i) For all $P \in \mathscr{A}$, $\mathscr{C}P = \mathscr{S}[(\mathscr{D}P) \cdot X]$.

(ii) If $g \in \mathscr{A}^{(\rho)}$, then $\mathscr{D}g$ satisfies $(\mathscr{J}\mathscr{D}g)_{ij} = t[(\mathscr{J}\mathscr{D}g)_{ji}]$ (where still $t(a \otimes b) = b \otimes a$).

(iii) Let $g \in \mathscr{A}^{(\rho)}$ and define $\curvearrowright (a \otimes b \otimes c) = b \otimes c \otimes a$ and $\curvearrowleft (a \otimes b \otimes c) = c \otimes a \otimes b$, then

$$(A) = \curvearrowright [(\partial_j \otimes 1)\partial_k \mathscr{D}_i g] = (\partial_k \otimes 1)\partial_i \mathscr{D}_j g = (1 \otimes \partial_i)\partial_k \mathscr{D}_j g,$$
$$(B) = \curvearrowleft [(1 \otimes \partial_j)\partial_k \mathscr{D}_i g] = (1 \otimes \partial_k)\partial_i \mathscr{D}_j g = (\partial_i \otimes 1)\partial_k \mathscr{D}_j g.$$

(iv) Let $F, G \in \mathscr{A}^{(\rho)} \otimes (\mathscr{A}^{(\rho)})^{op}$, then $t(F.G) = t(G).t(F)$.

(v) Define the contractions $(a \otimes b \otimes c)\#_1 d = (adb) \otimes c$ and $(a \otimes b \otimes c)\#_2 d = a \otimes (bdc)$, $a, b, c, d \in \mathscr{A}$, then

$$t\left[(a \otimes b \otimes c)\#_1 d\right] = \left[\curvearrowleft (a \otimes b \otimes c)\right]\#_2 d,$$
$$t\left[(a \otimes b \otimes c)\#_2 d\right] = \left[\curvearrowright (a \otimes b \otimes c)\right]\#_1 d.$$

(vi) For F and g like above, it holds that

$$\partial_j\left[F\#g\right] = F.(\partial_j g) + \left[(1 \otimes \partial_j)F\right]\#_1 g + \left[(\partial_j \otimes 1)F\right]\#_2 g.$$

Proof. Part (i) is trivial.

For part (ii), we may assume that g is a monomial, in particular $g = X_{i(1)} \cdots X_{i(k)}$. Then

$$\partial_i \mathscr{D}_j g = \partial_i \sum_{g = AX_jB} BA = \sum_{g = AX_jBX_iC} B \otimes (CA) + \sum_{g = AX_iBX_jC} (CA) \otimes B,$$

while

$$\partial_j \mathscr{D}_i g = \sum_{g = AX_iBX_jC} B \otimes (CA) + \sum_{g = AX_jBX_iC} (CA) \otimes B,$$

and we see that applying t to $\partial_j \mathscr{D}_i g$ gives $\partial_i \mathscr{D}_j g$.

For part (iii), we may again assume that g is a monomial q. Then, suppressing the q's in the sums for brevity,

$$\begin{aligned}
(A) =& \curvearrowright \left[(\partial_j \otimes 1) \sum_{AX_kBX_iC} (CA) \otimes B + (\partial_j \otimes 1) \sum_{AX_iBX_kC} B \otimes (AC)\right], \\
=& \curvearrowright \left[\sum_{AX_jBX_kCX_iD} (DA) \otimes B \otimes C + \sum_{AX_kBX_iCX_jD} C \otimes (DA) \otimes B + \sum_{AX_iBX_jCX_kD} B \otimes C \otimes (DA)\right], \\
=& \sum_{AX_jBX_kCX_iD} B \otimes C \otimes (DA) + \sum_{AX_kBX_iCX_jD} (DA) \otimes B \otimes C + \sum_{AX_iBX_jCX_kD} C \otimes (DA) \otimes B, \\
=& \sum_{AX_jBX_iC} (\partial_k \otimes 1)\left[B \otimes (CA)\right] + \sum_{AX_iBX_jC} (\partial_k \otimes 1)\left[(CA) \otimes B\right], \\
=& (\partial_k \otimes 1)\partial_i \mathscr{D}_j q.
\end{aligned}$$

The calculation for (B) is entirely analogous, but we include it for completeness.

$$\begin{aligned}(B) &= \curvearrowleft \left[\sum_{AX_kBX_jCX_iD} (DA) \otimes B \otimes C + \sum_{AX_jBX_iCX_kD} C \otimes (DA) \otimes B + \sum_{AX_iBX_kCX_jD} B \otimes C \otimes (DA) \right], \\ &= \sum_{AX_kBX_jCX_iD} C \otimes (DA) \otimes B + \sum_{AX_jBX_iCX_kD} B \otimes C \otimes (DA) + \sum_{AX_iBX_kCX_jD} (DA) \otimes B \otimes C, \\ &= (1 \otimes \partial_k) \left[\sum_{A_iBX_jC} (CA) \otimes B + \sum_{AX_jBX_iC} B \otimes (CA) \right], \\ &= (1 \otimes \partial_k)\partial_i \mathscr{D}_j q.\end{aligned}$$

The other forms of the identities follow from Proposition 4.4.

For part (iv), we may suppose that $F = \sum_i a_i \otimes b_i$ and $G = \sum_j c_j \otimes d_j$ (with the sums finite), thus

$$\begin{aligned}t(F.G) &= \sum_{i,j} t\left[(a_i \otimes b_i).(c_j \otimes d_j)\right] = \sum_{i,j} t\left[(a_ic_j) \otimes (d_jb_i)\right] = \sum_{i,j} (d_jb_i) \otimes (a_ic_j) = \sum_{i,j} (d_j \otimes c_j).(b_i \otimes a_i), \\ &= t(G).t(F).\end{aligned}$$

For part (v), we can simply compute

$$\begin{aligned}t\left[(a \otimes b \otimes c)\#_1 d\right] &= t\left[(adb) \otimes c\right] = c \otimes (adb) = (c \otimes a \otimes b)\#_2 d = \left[\curvearrowleft (a \otimes b \otimes c)\right]\#_2 d, \\ t\left[(a \otimes b \otimes c)\#_2 d\right] &= t\left[a \otimes (bdc)\right] = (bdc) \otimes a = (b \otimes c \otimes a)\#_1 d = \left[\curvearrowright (a \otimes b \otimes c)\right]\#_1 d.\end{aligned}$$

For part (vi), we may suppose that F is as in the proof of part (iv). The stated result is then little more than the Leibniz rule,

$$\begin{aligned}\partial_j\left[F\#g\right] &= \sum_i \partial_j\left[a_igb_i\right] = \sum_i \left[(\partial_j a_i)(1 \otimes (gb_i)) + (a_i \otimes 1)(\partial_j g)(1 \otimes b_i) + ((a_ig) \otimes 1)(\partial_j b_i)\right], \\ &= \left[(\partial_j \otimes 1)F\right]\#_2 g + F.(\partial_j g) + \left[(1 \otimes \partial_j)F\right]\#_1 g.\end{aligned}$$

□

6.2 The Schwinger-Dyson Equation

Let now (M, τ) be a tracial W*-probability space, thus M is a von Neumann algebra, and τ is a faithful normal trace. Assume that $x_1, \ldots, x_n \in M$ are self-adjoint and algebraically free elements that generate M. Then with the identification $x_i \leftrightarrow X_i$, $\mathscr{A}$ is a strongly dense subset of M, and M may be viewed as a completion of $\mathscr{A}$. Notice that $\mathscr{A}$ now receives a $*$-algebra structure, which complex conjugates coefficients and reverses the order of monomials. The trace τ restricts to a trace, which we will denote τ_X, on $\mathscr{A}$. This is precisely the joint distribution of the random variables X_i, and we write that $M \cong W^*(\tau_X)$. We can write this because we know that a (finite) set of generators together with their joint law contain all the information of the generated von Neumann algebra. Suppose that $N \cong W^*(\tau_Y)$, that is to say that there are self-adjoint generators $Y_1, \ldots, Y_n \in N$, and their joint law is τ_Y. Finally, we are in a position to define what is meant by free monotone transport.

Definition 6.13. By *transport* between the laws τ_X and τ_N, we mean elements $\hat{Y}_1, \ldots, \hat{Y}_n \in W^*(\tau_X)$, which are self-adjoint, and whose joint distribution is precisely τ_Y. Recall that $\|\cdot\|_2$ denotes the norm on $L^2(M, \tau)$, then if for all $1 \leq j \leq n$ we have

$$\hat{Y}_j \in \overline{\left\{\mathscr{D}g \,|g \in \mathscr{A} \text{ and } \mathscr{J}\mathscr{D}g \geq 0\right\}}^{\|\cdot\|_2},$$

we say that the transport is *monotone*. Notice that now, $\mathscr{J}\mathscr{D}g \in M_n(M \otimes M^{op})$, which indeed carries a notion of positivity.

Remark 6.14. We should compare this definition to the classical one given in Section 1.3, Definition 1.17, Theorem 1.21, and Definition 1.24. If $n = 1$, the von Neumann algebra M is Abelian, and is simply an algebra of bounded functions on some measure space. Thus our definition above reduces to asking for a measurable function h such that $\hat{Y} = h(X)$ and hence $h_*\tau_X = \tau_Y$, which is the same as in Definition 1.17. Our notion of monotone transport is chosen to be reminiscent of the Brenier map of Theorem 1.21, where the transport map was the (ordinary) gradient of a convex function, while here it is the cyclic gradient. Moreover, in the $n = 1$ case, the demand $\mathscr{J}\mathscr{D}g \geq 0$ becomes the condition that the function $f = \mathscr{D}g$ satisfies $(f(x) - f(y))/(x - y) \geq 0$, which is compatible with Definition 1.24.

It is now the question in which situations we can prove that monotone transport exists. As the aim is to construct elements in M, we should choose the generators X_j to be random variables that we understand rather well. Given the material we have covered so far, it is a natural choice to pick a free family of n (standard) semicircular elements. Next, we need a mechanism that connects the semicircular law to others. Recall that the conjugate variables to a free semicircular family were rather simple, namely $\partial_i^*(1 \otimes 1) = X_i$, the elements themselves. In general, conjugate variables are elements of $L^2(X_1, \ldots, X_n)$, but what happens if we demand instead that they belong to some $\mathscr{A}^{(\rho)}$ for a big enough ρ (recall that these spaces shrink with increasing ρ)? From Example 1.13 in the first section of this thesis, one obtains the bound $C_k \leq 4^k$ (with C_k the Catalan numbers), so that it is not difficult to see that we need $\rho > 4$ for $\mathscr{A}^{(\rho)} \subset L^2(X_1, \ldots, X_n)$. Assuming this to be the case, we can write down the equation (dropping the hats for convenience)

$$\partial_{Y_j}^*(1 \otimes 1) = Y_j + \mathscr{D}_{Y_j}\left[W(Y_1, \ldots, Y_n)\right], \tag{6.2.1}$$

which is called a *Schwinger-Dyson equation*, and is to be solved for the joint law τ_V of the Y_j. Here, the idea is to have W be a 'small' (with respect to $\|\cdot\|_\rho$ of course) perturbation of the conjugate variable relations of the semicircular elements, and $V = (1/2)\sum_j Y_j^2 + W$ is the *potential* of the equation. In this section, we will show that this is a meaningful equation, in the sense that it has unique solutions under reasonable assumptions. For this, we present and prove a slightly stronger version of Theorem 2.1 in [42]. While in the next two sections, we will show that we can construct transport between the free semicirculars X_n and the Y_j arising from the solution of the Schwinger-Dyson equation.

Theorem 6.15. *Fix $R > 1$ and assume that $\rho > 2R$, moreover write $T_1, \ldots, T_n$ for formal indeterminates generating $\mathscr{A}$[4]. Then there exists a $\delta > 0$, such that for all $W \in \mathscr{A}^{(\rho)}$ with $\|W\|_\rho < \delta$, there is at most one functional τ_V solving the Schwinger-Dyson equation* (6.2.1) *with potential $V = (1/2)\sum_j T_j^2 + W$ and additionally satisfying the bound $|\tau_V(q)| \leq R^{\deg q}$ for all monomials q.*

Proof. We begin by rewriting (6.2.1) into a form where the functional appears more explicitly (as it stands it enters implicitly through the adjoint). It is not hard to see that this results in the system of equations

$$(\tau \otimes \tau)(\partial_j P) = \tau\left((T_j + \mathscr{D}_j W)P\right), \tag{6.2.2}$$

for all $1 \leq j \leq n$ and $P \in \mathscr{A}$, together with the constraint $\tau(1) = 1$. Assume now that we have two solutions τ_V and τ_V', then it suffices to show that $\tau_V(q) = \tau_V'(q)$ for all monomials q to conclude that they are in fact equal. It is therefore natural to consider the numbers

$$\Delta_\ell := \max_{\deg q = \ell} |\tau_V(q) - \tau_V'(q)|,$$

[4]We do this to avoid confusion with the use of X and Y above. Note that also all differentiation is with respect to these T's.

where $\ell \geq 1$ (as trivially $\Delta_0 = 0$). By the bound stated in the assumptions, it holds that $\Delta_\ell \leq 2R^\ell$, so that it can grow at most geometrically in ℓ. This suggests that we can meaningfully combine these numbers into

$$d(t) := \sum_{\ell \geq 0} \Delta_\ell t^\ell,$$

for any $0 < t < 1/R$. As each term in $d(t)$ is nonnegative, we can conclude that all the Δ_ℓ's are zero if we manage to show that $d(t) = 0$ for some conveniently chosen t.

Now we have an opportunity to exploit that τ_V and τ'_V both solve (6.2.2). Let q be a monomial of degree ℓ, then it can be written as $q = T_j p$ for some monomial p of degree $\ell - 1$. The combination $T_j p$ occurs in (6.2.2), and so we can isolate it,

$$\tau(T_j p) = (\tau \otimes \tau)(\partial_j p) - \tau((\mathscr{D}_j W)p),$$

and take the difference of this relation for τ_V and τ'_V, which after some rearrangement yields

$$(\tau_V - \tau'_V)(T_j p) = \left((\tau_V - \tau'_V) \otimes \tau_V\right)(\partial_j p) + \left(\tau'_V \otimes (\tau_V - \tau'_V)\right)(\partial_j p) - (\tau_V - \tau'_V)((\mathscr{D}_j W)p).$$

Consider the first term on the right hand side,

$$\begin{aligned} \left|\left((\tau_V - \tau'_V) \otimes \tau_V\right)(\partial_j p)\right| &\leq \sum_{p = AT_jB} |\tau_V(B)||(\tau_V - \tau'_V)(A)|, \\ &\leq \sum_{p = AT_jB} R^{\deg B} \Delta_{\deg A}, \\ &\leq \sum_{k=0}^{\ell-2} \Delta_k R^{\ell-2-k}, \end{aligned}$$

where we used that all the A's and B's are monomials as p is a monomial. The second term on the right hand side van be estimated in exactly the same way. Thus we have so far that

$$\Delta_\ell = \max_{1 \leq j \leq n} \max_{\deg p = \ell - 1} |(\tau_V - \tau'_V)(T_j p)| \leq 2 \sum_{k=0}^{\ell-2} \Delta_k R^{\ell-2-k} + \max_{1 \leq j \leq n} \max_{\deg p = \ell - 1} |(\tau_V - \tau'_V)((\mathscr{D}_j W)p)|.$$

To estimate the other term, write $W = \sum_q q\, \gamma_q(W)$, so that

$$\begin{aligned} |(\tau_V - \tau'_V)((\mathscr{D}_j W)p)| &\leq \sum_q \sum_{q = AT_jB} |\gamma_q(W)||(\tau_V - \tau'_V)(BAp)|, \\ &\leq \sum_q (\deg q)\, |\gamma_q(W)| \Delta_{\ell - 1 + \deg q - 1}. \end{aligned}$$

We combine the estimates derived so far on Δ_ℓ to bound $d(t)$,

$$d(t) \leq \sum_{\ell \geq 1} t^\ell \left[2 \sum_{k=0}^{\ell-2} \Delta_k R^{\ell-2-k} + \sum_q (\deg q)\, |\gamma_q(W)| \Delta_{\ell - 1 + \deg q - 1} \right].$$

For the second term, we have

$$\sum_q (\deg q) |\gamma_q(W)| t^{2 - \deg q} \sum_{\ell \geq 1} t^{\ell - 1 + \deg q - 1} \Delta_{\ell - 1 + \deg q - 1} \leq d(t) t^2 \sum_q (\deg q) |\gamma_q(W)| \left(\frac{1}{t}\right)^{\deg q}.$$

We now choose $t = 1/2R$, so that this becomes

$$\frac{d(1/2R)}{(2R)^2}\sum_q(\deg q)|\gamma_q(W)|\,(2R)^{\deg q} \leq \frac{d(1/2R)}{(2R)^2}\sum_q|\gamma_q(W)|\left[(\deg q)\left(\frac{2R}{\rho}\right)^{\deg q}\right]\rho^{\deg q},$$
$$\leq C\frac{d(1/2R)}{(2R)^2}\|W\|_\rho,$$

where $C > 0$ arises as a bound on the term in square brackets (note that we assumed $\rho > 2R$). Plugging this back into our bound on $d(t)$ and recalling that we set $t = 1/2R$ yields

$$d(1/2R) \leq 2\sum_{\ell\geq 1}\sum_{k=1}^{\ell-2}\Delta_k\left(\frac{1}{2R}\right)^\ell R^{\ell-2-k} + C\frac{d(1/2R)}{(2R)^2}\|W\|_\rho.$$

Focus on the first term, which can also be written as

$$2\sum_{\ell\geq 1}\sum_{k=1}^{\ell-2}\Delta_k\left(\frac{1}{2R}\right)^\ell R^{\ell-2-k} = \frac{1}{2R^2}\sum_{\ell\geq 1}\sum_{k=1}^{\ell-2}\Delta_k\left(\frac{1}{2R}\right)^k\left(\frac{1}{2}\right)^{\ell-2-k},$$
$$= \frac{1}{2R^2}\sum_{k\geq 0}\Delta_k\left(\frac{1}{2R}\right)^k\sum_{\ell\geq k+2}\left(\frac{1}{2}\right)^{\ell-2-k},$$
$$= \frac{1}{R^2}d(1/2R).$$

Finally then, we have that

$$d(1/2R) \leq d(1/2R)\left[\frac{1}{R^2} + \frac{C}{(2R)^2}\|W\|_\rho\right].$$

It follows that if $\delta > 0$ is such that

$$\frac{1}{R^2} + \frac{C\delta}{4R^2} < 1,$$

which is possible because $1/R^2 < 1$, $d(1/2R) = 0$ holds true in case $\|W\|_\rho < \delta$. As argued above, this then forces $\tau_V = \tau_V'$, and hence there is at most one solution. □

6.3 Manipulations and Estimates

In Section 5.3, we sketched how to proceed for the commutative ($n = 1$) case. Here, we will begin filling in the details of this sketch, and tackle the technical obstacles that come from noncommutativity. Throughout this section, we make the convention that all differential operators that are written without or with numerical indices are with respect to the X_j, which are taken to be an n-tuple of free standard semicircular elements in a tracial W*-probability space (M, τ). If we want to take a derivative with respect to an Y_j, we will explicitly write a subscript Y_j. Also, we will write X and Y for the n-tuples $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ respectively. To recapitulate the first few steps taken in Section 5.3, we want to make the Ansatz $Y = X + f(X)$ for some 'nice' $f \in (\mathscr{A}^{(\rho)})^n$. Using this to simplify (6.2.1), we can arrive at an equation that only involves f and the semicircle law, which can be further simplified using some of the particular properties of the semicircle law. Making the additional Ansatz that $f = \mathscr{D}g$ for some 'nice' $g \in \mathscr{A}^{(\rho)}$, and imposing some invertibility condition on $\mathscr{J}f$, we can (with considerable effort) extract a cyclic derivative from both sides of the equation, yielding an equation for g. To carry out these first few steps, we need one more gadget.

Definition 6.16. As we have that $1 \otimes 1$ belongs to the domain of ∂_j^* for all j (as the X_j are semicircular), both ∂_j and ∂_j^* are densely defined unbounded operators, and all of $\mathscr{A} \otimes \mathscr{A}^{op}$ lies in the domains of the ∂_j^* (see Section 4.1). Let $A \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$, and define the adjoint to $\mathscr{J}$ as

$$\mathscr{J}^* A = \left(\sum_{i=1}^n \partial_i^* A_{ji} \right)_{j=1}^n \in \left(L^2(X_1, \ldots, X_n) \right)^n .$$

Lemma 6.17. *Let $A \in M_n(\mathscr{A} \otimes \mathscr{A}^{op})$ and $h \in \mathscr{A}^{(\rho)}$, then*

$$\tau \left[(\mathscr{J}^* A) \cdot h \right] = \sum_{i,j=1}^n (\tau \otimes \tau) \left[A_{ij} . t(\mathscr{J} h)_{ij} \right] . \tag{6.3.1}$$

Proof. Using the trace property of τ, we have

$$\tau \left[(\mathscr{J}^* A) \cdot h \right] = \sum_{i,j=1}^n \tau \left[(\partial_j^* A_{ij}) h_i \right] = \sum_{i,j=1}^n \tau \left[h_i (\partial_j^* A_{ij}) \right] = \sum_{i,j=1}^n \langle \partial_j^* A_{ij}, h_i^* \rangle = \sum_{i,j=1}^n \langle A_{ij}, \partial_j h_i^* \rangle .$$

We claim that $\partial_j h_i^* = t(\partial_j h_i)^*$, where $(a \otimes b)^* = a^* \otimes b^*$. It suffices to check this on monomials q. Thus

$$\partial_j q^* = \sum_{q^* = A X_j B} A \otimes B = \sum_{q = B^* X_j A^*} A \otimes B = t \left(\sum_{q = B^* X_j A^*} B^* \otimes A^* \right)^* = t(\partial_j h_i)^* ,$$

as claimed. Now,

$$\tau \left[(\mathscr{J}^* A) \cdot h \right] = \sum_{i,j=1}^n \langle A_{ij}, t(\partial_j h_i)^* \rangle = \sum_{i,j=1}^n (\tau \otimes \tau) \left[t(\partial_j h_i) A_{ij} \right] = \sum_{i,j=1}^n (\tau \otimes \tau) \left[A_{ij} . t(\mathscr{J} h)_{ij} \right] ,$$

where in the last step we used again the trace property of τ (and checking on monomials). □

Our first step towards carrying out the programme described above is to prove the following proposition, which implements the Ansatz $Y = X + f(X)$, and reformulates (6.2.1) entirely in terms of f and X.

Proposition 6.18. *Assume that $g = g^* \in \mathscr{A}^{(\rho)}$ and that $1 + \mathscr{J} f$ is invertible, where $f_j = \mathscr{D}_j g$. Putting $G = (1/2) \sum_{j=1}^n X_j^2 + W(X)$ and $Y = \mathscr{D} G = X + f(X)$, we have that (6.2.1) is true if and only if*

$$\mathscr{J}^* \left(\frac{1}{1 + \mathscr{J} f} \right) = X + f + (\mathscr{D} W)(X + f), \tag{6.3.2}$$

where $(\mathscr{D} W)(X + f)$ means the cyclic gradient of W, viewed as a function of X, with X replaced by $X + f(X)$ in the result.

The proof requires the following lemma, which features a kind of noncommutative chain rule, and collects some properties of the Jacobian of the transformation from X to Y.

Lemma 6.19. *Assume that the $\mathscr{J} Y \in M_n(M \otimes M^{op})$ is bounded and invertible, then*

(i) the operator

$$\hat{\partial}_j (\cdot) := \sum_{i=1}^n \partial_{X_i} (\cdot) . \left((\mathscr{J} Y)^{-1} \right)_{ij} ,$$

satisfies $\hat{\partial}_j = \partial_{Y_j}$ and $\partial_{Y_j} X_i = \left((\mathscr{J} Y)^{-1} \right)_{ij}$.

(ii) the conjugate variables to the Y_j can be written $\partial^*_{Y_j}(1\otimes 1) = \sum_k \partial^*_{X_k}\left((\mathscr{J}Y)^{-1}\right)^*_{kj}$.

(iii) if additionally $G = G^* \in \mathscr{A}^{(\rho)}$ is such that $\mathscr{D}_j G = Y_j$, we have $t(\mathscr{J}Y)^*_{ij} = (\mathscr{J}Y)_{ij}$ and $(\mathscr{J}Y)^{-1}$ is a self-adjoint.

Proof. For part (i), it is easy to see that the $\hat{\partial}_j$ are derivations, so that we only need to show that $\hat{\partial}_j Y_i = \delta_{ij} 1\otimes 1$. This follows from the computation

$$\hat{\partial}_j Y_i = \sum_{k=1}^n \partial_{X_k}(Y_i).\left((\mathscr{J}Y)^{-1}\right)_{kj} = \left[\mathscr{J}Y.(\mathscr{J}Y)^{-1}\right]_{ij} = \delta_{ij} 1\otimes 1.$$

Thus also

$$\partial_{Y_j} X_i = \hat{\partial}_j X_i = \sum_{k=1}^n \partial_{X_k}(X_i).\left((\mathscr{J}Y)^{-1}\right)_{kj} = \sum_{k=1}^n \delta_{ik}\left((\mathscr{J}Y)^{-1}\right)_{kj} = \left((\mathscr{J}Y)^{-1}\right)_{ij}.$$

We turn to (ii), where by definition

$$\langle \partial^*_{Y_j}(1\otimes 1), q\rangle = \langle 1\otimes 1, \partial_{Y_j} q\rangle = \langle 1\otimes 1, \hat{\partial}_j q\rangle = \sum_{k=1}^n \langle 1\otimes 1, (\partial_{X_k} q).\left[(\mathscr{J}Y)^{-1}\right]_{kj}\rangle,$$

with q a monomial. Using the trace property of τ, we can continue by writing

$$\langle \partial^*_{Y_j}(1\otimes 1), q\rangle = \sum_{k=1}^n \langle \left[(\mathscr{J}Y)^{-1}\right]^*_{kj}, (\partial_{X_k} q)\rangle = \sum_{k=1}^n \langle \partial^*_{X_k}\left[(\mathscr{J}Y)^{-1}\right]^*_{kj}, q\rangle.$$

As this is true for any monomial q, the desired result follows.

Now for (iii), we have already shown in part (ii) of Lemma 6.12 that $(\mathscr{J}\mathscr{D}g)_{ij} = t[(\mathscr{J}\mathscr{D}g)_{ji}]$. To complete the proof, it therefore suffices to argue that also $(\mathscr{J}\mathscr{D}g)^*_{ij} = t[(\mathscr{J}\mathscr{D}g)_{ij}]$. It then follows immediately that $\mathscr{J}Y$ and its inverse are self-adjoint. Hence assume without loss of generality that $G = q + q^*$, whence

$$(\mathscr{J}Y)_{ij} = \sum_{q=AX_jB}\ \sum_{BA=CX_iD} (C\otimes D + D^*\otimes C^*),$$

from which the desired conclusion is readily drawn. □

Proof of Proposition 6.18. By points (ii) and (iii) of the previous lemma, the left hand side of (6.2.1) is equal to

$$\partial^*_{Y_j}(1\otimes 1) = \sum_k \partial^*_{X_k}\left((\mathscr{J}Y)^{-1}\right)^*_{kj} = \sum_k \partial^*_{X_k}\left((\mathscr{J}Y)^{-1}\right)_{jk} = \left[\mathscr{J}^*(\mathscr{J}Y)^{-1}\right]_j.$$

As $Y = X + f(X)$, $\mathscr{J}Y = 1 + \mathscr{J}f$, and so $(\mathscr{J}Y)^{-1} = 1/(1+\mathscr{J}f)$, which exists by assumption. For the right hand side of (6.2.1), it is not difficult to see that this equals

$$Y_j + \mathscr{D}_{Y_j}[W(Y_1,\ldots,Y_n)] = X_j + f_j(X) + (\mathscr{D}_j W)(X + f(X)).$$

Combining all the above yields

$$\mathscr{J}^*\left(\frac{1}{1+\mathscr{J}f}\right) = X + f + (\mathscr{D}W)(X+f).$$

□

Our next step is to try and extract a cyclic gradient from both sides of (6.3.2).

Proposition 6.20. *Assume that $f = \mathscr{D}g$ for a self-adjoint, cyclically symmetric $g \in \mathscr{A}^{(\rho)}$, such that $\|\mathscr{J}f\|_{\rho\otimes_\pi\rho} < 1$. Write*

$$Q(g) := (1\otimes\tau + \tau\otimes 1)\operatorname{Tr}\left[\mathscr{J}\mathscr{D}g - \log(1+\mathscr{J}\mathscr{D}g)\right],$$

then (6.3.2) *is equivalent to*

$$\mathscr{D}\big[(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D}g) - \mathscr{N}g\big] = \mathscr{D}\big[W(X+\mathscr{D}g) + Q(g) + \frac{1}{2}(\mathscr{D}g)\cdot(\mathscr{D}g)\big] \tag{6.3.3}$$

The proof of this proposition will require some significant preparation, the first part of which is rewriting (6.3.2) in a more convenient form.

Lemma 6.21. *Assume that $f = \mathscr{D}g$ with $g = g^* \in \mathscr{A}^{(\rho)}$ such that $h \mapsto (1+\mathscr{J}f)\#h$ defines an invertible mapping on $(\mathscr{A}^{(\rho)})^n$. Write $K = -\mathscr{J}^*\mathscr{J} - \mathrm{id}$, then* (6.3.2) *can be expressed also as*

$$K(f) = \mathscr{D}\left[W(X+f)\right] + (\mathscr{J}f)\#f + (\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right]. \tag{6.3.4}$$

Proof. Let us begin by rewriting (6.3.2) as

$$f + (\mathscr{D}W)(X+f) = \mathscr{J}^*\left(\frac{1}{1+\mathscr{J}f}\right) - X.$$

Now we use that τ is the semicircle law, which implies $X = \mathscr{J}^*(1\otimes 1)$, and

$$f + (\mathscr{D}W)(X+f) = \mathscr{J}^*\left(\frac{1}{1+\mathscr{J}f} - 1\right) = -\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right).$$

Thus, we have

$$\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) + f + (\mathscr{D}W)(X+f) = 0.$$

By assumption, we may # both sides by $(1+\mathscr{J}f)$ to obtain the equivalent equation

$$\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) + f + (\mathscr{D}W)(X+f) + (\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) + (\mathscr{J}f)\#f + (\mathscr{J}f)\#(\mathscr{D}W)(X+f) = 0.$$

The first term in this equation can split into $\mathscr{J}^*\mathscr{J}f - \mathscr{J}^*[(\mathscr{J}f)^2/(1+\mathscr{J}f)]$, because $x/(1+x) = x - x^2/(1+x)$. Recognising K and rearranging terms gives

$$K(f) = (\mathscr{D}W)(X+f) + (\mathscr{J}f)\#\left[(\mathscr{D}W)(X+f)\right] + (\mathscr{J}f)\#f + (\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right].$$

We now claim that the first two terms on the right hand side combine to give

$$\mathscr{D}\left[W(X+f)\right] = (\mathscr{D}W)(X+f) + (\mathscr{J}f)\#\left[(\mathscr{D}W)(X+f)\right],$$

which we prove by showing that it holds true if W is a monomial, $W(X) = X_{i(1)} \cdots X_{i(k)}$. One sees that

$$\mathscr{D}\left[W(X+f)\right] = (\iota \circ t)\left\{\partial_j\left[\left(X_{i(1)} + f_{i(1)}(X)\right)\cdots\left(X_{i(k)} + f_{i(k)}(X)\right)\right]\right\},$$

where ι is as in Remark 6.5. We investigate the expression in curly brackets, slightly abusing notation for brevity,

$$\begin{aligned}\{\dots\} = &\sum_{\ell=1}^{k} \delta_{ji(\ell)}\left[(X+f)_{i(1)}\cdots(X+f)_{i(\ell-1)}\right]\otimes\left[(X+f)_{i(\ell+1)}\cdots(X+f)_{i(k)}\right]\\ &+\sum_{\ell=1}^{k}\left[(X+f)_{i(1)}\cdots(X+f)_{i(\ell-1)}\otimes 1\right](\partial_j f_{i(\ell)})\left[1\otimes(X+f)_{i(\ell+1)}\cdots(X+f)_{i(k)}\right].\end{aligned}$$

It is clear that the upper term on the right hand side will become $(\mathscr{D}W)(X+f)$. Notice that

$$(\iota\circ t)\left[(a\otimes 1)(b\otimes c)(1\otimes d)\right] = (\iota\circ t)\left[(ab)\otimes(cd)\right] = cdab = (c\otimes b)\#(da) = t(b\otimes c)\#(da),$$

which implies

$$\begin{aligned}\mathscr{D}\left[W(X+f)\right] = &(\mathscr{D}W)(X+f)\\ &+\sum_{\ell=1}^{k} t(\partial_j f_{i(\ell)})\#\left[(X+f)_{i(\ell+1)}\cdots(X+f)_{i(k)}\,(X+f)_{i(1)}\cdots(X+f)_{i(\ell-1)}\right].\end{aligned}$$

By our assumptions on f, we may invoke part (iii) of Lemma 6.19 to say that $t(\partial_j f_{i(\ell)}) = \partial_{i(\ell)} f_j$. Introducing an additional sum over m (from 1 to n) together with a $\delta_{mi(\ell)}$, we can swap the order of the sums over ℓ and m to find

$$\begin{aligned}\mathscr{D}\left[W(X+f)\right] &= (\mathscr{D}W)(X+f) + \sum_{m=1}^{n}(\partial_m f_j)\#\left[(\mathscr{D}_m W)(X+f)\right],\\ &= (\mathscr{D}W)(X+f) + (\mathscr{J}f)\#\left[(\mathscr{D}W)(X+f)\right],\end{aligned}$$

as claimed. With this identity, we finally arrive at

$$K(f) = \mathscr{D}\left[W(X+f)\right] + (\mathscr{J}f)\#f + (\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right].$$

□

We now try to extract cyclic gradients from both sides of (6.3.4) to the best of our abilities, the result of which is given by the following two lemmata.

Lemma 6.22. *Assume that $f = \mathscr{D}g$ with $g = g^* \in \mathscr{A}^{(\rho)}$ and cyclically symmetric, and let K be as in Lemma 6.21, then*

$$K(f) = \mathscr{D}\big[(1\otimes\tau + \tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D}g) - \mathscr{N}g\big]$$

Lemma 6.23. *Under the same assumptions as Proposition 6.20 and using the same notation,*

$$\mathscr{D}Q(g) = (\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right].$$

The main ingredient in their proofs is the following identity.

Lemma 6.24. *Under the same assumptions as Proposition 6.20, we have for any $m \in \mathbb{Z}_{\geq -1}$ that*

$$\frac{1}{m+2}\mathscr{D}\left\{(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left[(\mathscr{J}f)^{m+2}\right]\right\} = -\mathscr{J}^*\left[(\mathscr{J}f)^{m+2}\right] + (\mathscr{J}f)\#\mathscr{J}^*\left[(\mathscr{J}f)^{m+1}\right]. \quad (6.3.5)$$

We now proceed as follows. First, we use (6.3.5) to give short proofs of Lemma 6.22 and Lemma 6.23 respectively. Then, we can use these to give another short proof, this time of Proposition 6.20. It then remains to prove Lemma 6.24, whose proof will be by far the longest.

Proof of Lemma 6.22. Consider the $m=-1$ case of (6.3.5), which reads

$$\mathscr{D}\left[(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}f)\right] = -\mathscr{J}^*(\mathscr{J}f) + (\mathscr{J}f)\#\mathscr{J}^*(1) = -\mathscr{J}^*(\mathscr{J}f) + (\mathscr{J}f)\#X,$$

where we used again that $\partial_j^*(1\otimes 1) = X_j$ for semicircular variables. By checking on monomials, it is easy to see that $(\mathscr{J}f)\#X = \mathscr{N}f$ (where $\mathscr{N}$ is applied componentwise). Using our assumptions on f, $\mathscr{N}f = \mathscr{N}\mathscr{D}g = \mathscr{D}(\mathscr{N}-1)g$, as taking cyclic derivatives decreases degree by one. Combining what we have so far yields

$$\mathscr{D}\left[(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}f)\right] = -\mathscr{J}^*(\mathscr{J}f) - f + \mathscr{D}\mathscr{N}g = K(f) + \mathscr{D}\mathscr{N}g,$$

which can be rearranged into

$$K(f) = \mathscr{D}\big[(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left(\mathscr{J}\mathscr{D}g\right) - \mathscr{N}g\big],$$

which was to be shown. □

Proof of Lemma 6.23. By assumption, $\|\mathscr{J}f\|_{\rho\otimes_\pi\rho} < 1$, so that we may perform series expansions on

$$(\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) = (\mathscr{J}f)\#\mathscr{J}^*\left(\sum_{m=0}^{\infty}(-1)^m(\mathscr{J}f)^{m+1}\right),$$
$$\mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right] = \mathscr{J}^*\left(\sum_{m=0}^{\infty}(-1)^m(\mathscr{J}f)^{m+2}\right).$$

Thus,

$$\begin{aligned}
(\mathscr{J}f)\#\mathscr{J}^*\left(\frac{\mathscr{J}f}{1+\mathscr{J}f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J}f)^2}{1+\mathscr{J}f}\right] &= \sum_{m=0}^{\infty}(-1)^m\left\{-\mathscr{J}^*\left[(\mathscr{J}f)^{m+2}\right] + (\mathscr{J}f)\#\mathscr{J}^*\left[(\mathscr{J}f)^{m+1}\right]\right\},\\
&= \sum_{m=0}^{\infty}\frac{(-1)^m}{m+2}\mathscr{D}\left\{(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left[(\mathscr{J}f)^{m+2}\right]\right\},\\
&= \mathscr{D}\left\{(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left[\sum_{m=0}^{\infty}(-1)^m(\mathscr{J}f)^{m+2}\right]\right\},\\
&= \mathscr{D}\left\{(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left[\mathscr{J}f - \log\left(1+\mathscr{J}f\right)\right]\right\},\\
&= \mathscr{D}Q(g).
\end{aligned}$$

Here, we used (6.3.5) in the second step, and the Taylor series of $x - \log(1+x)$ in the last step. □

Proof of Proposition 6.20. Since we assume that $\|\mathscr{J}f\|_{\rho\otimes_\pi\rho} < 1$, the conditions of Lemma 6.21 are satisfied. Indeed, by part (iii) of Lemma 6.11, $\|(\mathscr{J}f)\#h\|_\rho \leq \|\mathscr{J}f\|_{\rho\otimes_\pi\rho}\|h\|_\rho < \|h\|_\rho$, so the map

$h \mapsto (\mathscr{J} f)\#h$ is contractive, and thus $h \mapsto h + (\mathscr{J} f)\#h$ is invertible. Thus we have that (6.3.2) is equivalent to (6.3.4), which was

$$K(f) = \mathscr{D}\left[W(X+f)\right] + (\mathscr{J} f)\#f + (\mathscr{J} f)\#\mathscr{J}^*\left(\frac{\mathscr{J} f}{1+\mathscr{J} f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J} f)^2}{1+\mathscr{J} f}\right].$$

Now, Lemma 6.22 allows us to replace the left hand side of this by

$$K(f) = \mathscr{D}\big[(1\otimes\tau + \tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D} g) - \mathscr{N} g\big],$$

while we may replace the last two term on the right hand side by

$$(\mathscr{J} f)\#\mathscr{J}^*\left(\frac{\mathscr{J} f}{1+\mathscr{J} f}\right) - \mathscr{J}^*\left[\frac{(\mathscr{J} f)^2}{1+\mathscr{J} f}\right] = \mathscr{D} Q(g)$$

on account of Lemma 6.23. Recalling finally that $f = \mathscr{D} g$, we see that (6.3.2) is equivalent to

$$\mathscr{D}\big[(1\otimes\tau + \tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D} g) - \mathscr{N} g\big] = \mathscr{D}\big[W(X+\mathscr{D} g) + Q(g)\big] + (\mathscr{J}\mathscr{D} g)\#\mathscr{D} g,$$

and we see that it only remains to show that

$$\mathscr{D}\left[\frac{1}{2}(\mathscr{D} g)\cdot(\mathscr{D} g)\right] = (\mathscr{J}\mathscr{D} g)\#(\mathscr{D} g).$$

For this, suppose that $q = (q_1, \ldots, q_n)$ with q_j monomials, then

$$\left[\mathscr{D}(q\cdot q)\right]_j = \sum_{i=1}^n \mathscr{D}_j(q_i q_i) = 2\sum_{i=1}^n \sum_{q_i = AX_jB} Bq_iA = 2\sum_{i=1}^n t\left(\sum_{q=AX_jB} A\otimes B\right)\#q_i = 2\sum_{i=1}^n t(\mathscr{J} q)_{ij}\#q_i.$$

Using part (ii) of Lemma 6.12, we conclude that

$$\left[\mathscr{D}(\mathscr{D} g\cdot\mathscr{D} g)\right]_j = 2\sum_{i=1}^n t(\mathscr{J}\mathscr{D} g)_{ij}\#(\mathscr{D} g)_i = \left[2(\mathscr{J}\mathscr{D} g)\#(\mathscr{D} g)\right]_j.$$

□

Proof of Lemma 6.24. Since (6.3.5) is an equality of vectors in $(L^2(X_1,\ldots,X_n))^n$, we can show that it is true by taking the inner product with an arbitrary test function $h \in (\mathscr{A}^{(\rho)})^n$, and arguing that the resulting equations must hold (recall that $\mathscr{A}^{(\rho)} \subset L^2(\ldots)$ for $\rho > 4$ and that it is dense with respect to $\|\cdot\|_2$, as it contains $\mathscr{A}$). We shall begin with the right hand side of (6.3.5). Define

$$\Lambda(h) := \tau\left\{\left[-\mathscr{J}^*\left((\mathscr{J} f)^{m+2}\right) + (\mathscr{J} f)\#\mathscr{J}^*\left((\mathscr{J} f)^{m+1}\right)\right]\cdot h\right\}.$$

Expanding definitions,

$$\Lambda(h) = -\tau\left[\mathscr{J}^*\left((\mathscr{J} f)^{m+2}\right)\cdot h\right] + \tau\left\{\sum_{i,j=1}^n \left([\mathscr{J} f]_{ij}\#\left[\mathscr{J}^*\left((\mathscr{J} f)^{m+1}\right)\right]_j\right) h_i\right\}.$$

To get rid of $\mathscr{J}^*$, we would like to use (6.3.1), but this requires us to bring it to the left side in the second term above. A quick algebraic calculation, using the trace property of τ, gives that

$$\tau\left\{\left[\left(\sum_i a_i\otimes b_i\right)\#c\right] d\right\} = \tau(a_i c b_i d) = \tau\left(\sum_i c b_i d a_i\right) = \tau\left\{c\left[t\left(\sum_i a_i\otimes b_i\right)\#d\right]\right\}.$$

Applying this to Λ, we find

$$\begin{aligned}\Lambda(h) &= -\tau\left[\mathscr{J}^*\left((\mathscr{J}f)^{m+2}\right)\cdot h\right] + \tau\left\{\sum_{i,j=1}^n \left[\mathscr{J}^*\left((\mathscr{J}f)^{m+1}\right)\right]_j \left[t(\mathscr{J}f)_{ij}\#h_i\right]\right\},\\ &= -\tau\left[\mathscr{J}^*\left((\mathscr{J}f)^{m+2}\right)\cdot h\right] + \tau\left\{\sum_{i,j=1}^n \left[\mathscr{J}^*\left((\mathscr{J}f)^{m+1}\right)\right]_j \left[(\mathscr{J}f)_{ji}\#h_i\right]\right\},\\ &= -\tau\left[\mathscr{J}^*\left((\mathscr{J}f)^{m+2}\right)\cdot h\right] + \tau\left\{\mathscr{J}^*\left((\mathscr{J}f)^{m+1}\right)\cdot[(\mathscr{J}f)\#h]\right\},\end{aligned}$$

where we also used part (ii) of Lemma 6.12. Thus we have brought ourselves into a position to use (6.3.1), which yields

$$\Lambda(h) = -\sum_{i,j=1}^n (\tau\otimes\tau)\left\{\left[(\mathscr{J}f)^{m+2}\right]_{ij}.\, t(\mathscr{J}h)_{ij}\right\} + \sum_{i,j=1}^n (\tau\otimes\tau)\left\{\left[(\mathscr{J}f)^{m+1}\right]_{ij}.\, t\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij}\right\}.$$

We continue to simplify this, focusing first on $t\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij}$. For this, we can start our offensive by the version of the Leibniz rule recorded in part (vi) of Lemma 6.12. Hence,

$$\begin{aligned}\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij} &= \sum_{k=1}^n \partial_j\left[(\partial_k f_i)\#h_k\right],\\ &= \sum_{k=1}^n (\partial_k f_i).(\partial_j h_k) + \sum_{k=1}^n \left[(1\otimes\partial_j)(\partial_k f_i)\right]\#_1 h_k + \sum_{k=1}^n \left[(\partial_j\otimes 1)(\partial_k f_i)\right]\#_2 h_k.\end{aligned}$$

We can take the next step on the strength of parts (iv) and (v) of Lemma 6.12, which is

$$\begin{aligned}t\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij} &= \sum_{k=1}^n t\left[(\partial_k f_i).(\partial_j h_k)\right] + \sum_{k=1}^n t\left\{\left[(1\otimes\partial_j)(\partial_k f_i)\right]\#_1 h_k\right\} + \sum_{k=1}^n t\left\{\left[(\partial_j\otimes 1)(\partial_k f_i)\right]\#_2 h_k\right\},\\ &= \sum_{k=1}^n t(\mathscr{J}h)_{kj}.t(\mathscr{J}f)_{ik} + \sum_{k=1}^n \curvearrowleft\left[(1\otimes\partial_j)(\partial_k f_i)\right]\#_2 h_k + \sum_{k=1}^n \curvearrowright\left[(\partial_j\otimes 1)(\partial_k f_i)\right]\#_1 h_k,\\ &= \sum_{k=1}^n t(\mathscr{J}h)_{kj}.(\mathscr{J}f)_{ki} + \sum_{k=1}^n \curvearrowleft\left[(1\otimes\partial_j)(\partial_k f_i)\right]\#_2 h_k + \sum_{k=1}^n \curvearrowright\left[(\partial_j\otimes 1)(\partial_k f_i)\right]\#_1 h_k,\end{aligned}$$

where we again used part (ii) of Lemma 6.12 in the last step. Our next weapon is part (iii) of Lemma 6.12, combined with the fact that $f = \mathscr{D}g$, which gives

$$\begin{aligned}t\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij} &= \sum_{k=1}^n t(\mathscr{J}h)_{kj}.(\mathscr{J}f)_{ki} + \sum_{k=1}^n \curvearrowleft\left[(1\otimes\partial_j)(\partial_k\mathscr{D}_i g)\right]\#_2 h_k + \sum_{k=1}^n \curvearrowright\left[(\partial_j\otimes 1)(\partial_k\mathscr{D}_i g)\right]\#_1 h_k,\\ &= \sum_{k=1}^n t(\mathscr{J}h)_{kj}.(\mathscr{J}f)_{ki} + \sum_{k=1}^n \left[(\partial_i\otimes 1)\partial_k\mathscr{D}_j g\right]\#_2 h_k + \sum_{k=1}^n \left[(1\otimes\partial_i)\partial_k\mathscr{D}_j g\right]\#_1 h_k,\\ &= \sum_{k=1}^n t(\mathscr{J}h)_{kj}.(\mathscr{J}f)_{ki} + \sum_{k=1}^n \left[(\partial_i\otimes 1)\partial_k f_j\right]\#_2 h_k + \sum_{k=1}^n \left[(1\otimes\partial_i)\partial_k f_j\right]\#_1 h_k.\end{aligned}$$

Introducing the notations

$$\begin{aligned}R^h_{ji} &= \sum_{k=1}^n \left\{\left[(\partial_i\otimes 1)\partial_k f_j\right]\#_2 h_k + \left[(1\otimes\partial_i)\partial_k f_j\right]\#_1 h_k\right\},\\ H_{jk} &= t(\mathscr{J}h)_{kj},\end{aligned}$$

this result can be written compactly as

$$t\left[\mathscr{J}\left((\mathscr{J}f)\#h\right)\right]_{ij} = \left[H.(\mathscr{J}f)\right]_{ji} + R^h_{ji}.$$

Returning our attention to Λ, we see that we have

$$\begin{aligned}
\Lambda(h) &= -\sum_{i,j=1}^{n}(\tau\otimes\tau)\left\{\left[(\mathscr{J}f)^{m+2}\right]_{ij}.\,H_{ji}\right\} + \sum_{i,j=1}^{n}(\tau\otimes\tau)\left\{\left[(\mathscr{J}f)^{m+1}\right]_{ij}.\left[H.(\mathscr{J}f)\right]_{ji}\right\} \\
&\quad + \sum_{i,j=1}^{n}(\tau\otimes\tau)\left\{\left[(\mathscr{J}f)^{m+1}\right]_{ij}.\,R^h_{ji}\right\}, \\
&= (\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{m+1}.H.(\mathscr{J}f)\right] - (\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{m+2}.H\right] + (\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{m+1}.R^h\right], \\
&= (\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{m+1}.R^h\right], \\
&= \frac{1}{m+2}\sum_{\ell=0}^{m+1}(\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{\ell}.R^h.(\mathscr{J}f)^{m+1-\ell}\right],
\end{aligned}$$

where the last two steps follow readily from the trace property of τ. Thus, we have reduced showing (6.3.5) to proving that

$$\sum_{\ell=0}^{m+1}(\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{\ell}.R^h.(\mathscr{J}f)^{m+1-\ell}\right] = \tau\left(\mathscr{D}\left\{(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}\left[(\mathscr{J}f)^{m+2}\right]\right\}\cdot h\right)$$

holds for all test functions h. We will accomplish this by introducing a certain function of $s\in(-\varepsilon,\varepsilon)$ for some $\varepsilon>0$, and showing that both sides of the above equation can be viewed as the derivative of this function in $s=0$. Writing $X^s_j = X_j + sh_j$ for all $1\le j\le n$, the derivative we want to consider is given by

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}\left[(\mathscr{J}f)(X^s)\right]_{ij} = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\partial_j f_i)(X^s).$$

First, we claim that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\partial_j f_i)(X^s) = \sum_{k=1}^{n}\left\{\left[(1\otimes\partial_k)(\partial_j f_i)\right]\#_2 h_k + \left[(\partial_k\otimes 1)(\partial_j f_i)\right]\#_1 h_k\right\},$$

which we prove by considering f_i to be a monomial $q = X_{i(1)}\cdots X_{i(\ell)}$. With this,

$$(\partial_j f_i)(X^s) = \sum_{m=1}^{\ell}\delta_{ji(m)}\left(X^s_{i(1)}\cdots X^s_{i(m-1)}\right)\otimes\left(X^s_{i(m+1)}\cdots X^s_{i(\ell)}\right),$$

and

$$\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\partial_j f_i)(X^s) &= \sum_{m=1}^{\ell}\delta_{ji(m)}\left[\left(\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}X^s_{i(1)}\cdots X^s_{i(m-1)}\right)\otimes\left(X^0_{i(m+1)}\cdots X^0_{i(\ell)}\right)\right.\\
&\qquad\left.+\left(X^0_{i(1)}\cdots X^0_{i(m-1)}\right)\otimes\left(\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}X^s_{i(m+1)}\cdots X^s_{i(\ell)}\right)\right],\\
&= \sum_{m=1}^{\ell}\delta_{ji(m)}\left[\sum_{p=1}^{m-1}\left(X^0_{i(1)}\cdots X^0_{i(p-1)}h_{i(p)}X^0_{i(p+1)}\cdots X^0_{i(m-1)}\right)\otimes\left(X_{i(m+1)}\cdots X_{i(\ell)}\right)\right.\\
&\qquad\left.+\sum_{p=m+1}^{\ell}\left(X_{i(1)}\cdots X_{i(m-1)}\right)\otimes\left(X^0_{i(m+1)}\cdots X^0_{i(p-1)}h_{i(p)}X^0_{i(p+1)}\cdots X^0_{i(\ell)}\right)\right],\\
&= \sum_{m=1}^{\ell}\delta_{ji(m)}\left\{\sum_{p=1}^{m-1}\left[\left(X_{i(1)}\cdots X_{i(p-1)}\otimes X_{i(p+1)}\cdots X_{i(m-1)}\otimes X_{i(m+1)}\cdots X_{i(\ell)}\right)\right]\#_1 h_{i(p)}\right.\\
&\qquad\left.+\sum_{p=m+1}^{\ell}\left[\left(X_{i(1)}\cdots X_{i(m-1)}\otimes X_{i(m+1)}\cdots X_{i(p-1)}\otimes X_{i(p+1)}\cdots X_{i(\ell)}\right)\right]\#_2 h_{i(p)}\right\},\\
&= \sum_{m=1}^{\ell}\delta_{ji(m)}\left\{\sum_{k=1}^{n}\left[(\partial_k\otimes 1)\left(X_{i(1)}\cdots X_{i(m-1)}\otimes X_{i(m+1)}\cdots X_{i(\ell)}\right)\right]\#_1 h_k\right.\\
&\qquad\left.+\sum_{k=1}^{n}\left[(1\otimes\partial_k)\left(X_{i(1)}\cdots X_{i(m-1)}\otimes X_{i(m+1)}\cdots X_{i(\ell)}\right)\right]\#_2 h_k\right\},\\
&= \sum_{k=1}^{n}\left\{[(1\otimes\partial_k)(\partial_j f_i)]\#_2 h_k+[(\partial_k\otimes 1)(\partial_j f_i)]\#_1 h_k\right\},
\end{aligned}$$

as claimed. Using the rules from Proposition 4.4, this can be recast into

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\partial_j f_i)(X^s)=\sum_{k=1}^{n}\left\{[(\partial_j\otimes 1)(\partial_k f_i)]\#_2 h_k+[(1\otimes\partial_j)(\partial_k f_i)]\#_1 h_k\right\}=R^h_{ij}.$$

It follows that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\tau\otimes\tau)\operatorname{Tr}\left\{[(\mathscr{J}f)(X^s)]^{m+2}\right\}=\sum_{\ell=0}^{m+1}(\tau\otimes\tau)\operatorname{Tr}\left[(\mathscr{J}f)^{\ell}.R^h.(\mathscr{J}f)^{m+1-\ell}\right].$$

On the other had, we claim that for any matrix $A\in M_n(\mathscr{A}\otimes\mathscr{A}^{op})$, it is also true that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\tau\otimes\tau)\operatorname{Tr}\left[A(X^s)\right]=\tau\left\{\mathscr{D}\left[(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(A)\right]\cdot h\right\}.$$

Notice that applying this to $A=(\mathscr{J}f)^{m+2}$ concludes the proof. To prove the claim, we can without loss of generality suppose that $A=(b\otimes c)e_{ij}$, with e_{ij} the matrix with a unit in the i,j-th position, and only zeros in all other entries. Then

$$\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(\tau\otimes\tau)\operatorname{Tr}\left[A(X^s)\right] &= \delta_{ij}\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}\tau[b(X^s)]\tau[c(X^s)],\\
&= \delta_{ij}\left\{\tau\left[\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}b(X^s)\right]\tau(c)+\tau(b)\tau\left[\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}c(X^s)\right]\right\}.
\end{aligned}$$

Using similar arguments as in the proof of the previous claim and the trace property of τ, one can show that

$$\tau\left[\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} b(X^s)\right] = \tau\left[(\mathscr{D}b)\cdot h\right],$$

and analogously for c. Thus,

$$\begin{aligned}\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} (\tau\otimes\tau)\,\mathrm{Tr}\left[A(X^s)\right] &= \delta_{ij}\left\{\tau\left[(\mathscr{D}b)\cdot h\right]\tau(c) + \tau(b)\tau\left[(\mathscr{D}c)\cdot h\right]\right\},\\ &= \delta_{ij}\tau\left\{(1\otimes\tau)\left[(\mathscr{D}b\cdot h)\otimes c\right] + (\tau\otimes 1)\left[b\otimes(\mathscr{D}c\cdot h)\right]\right\},\\ &= \delta_{ij}\tau\left\{\mathscr{D}\left[(1\otimes\tau)(b\otimes c)\right]\cdot h + \mathscr{D}\left[(\tau\otimes 1)(b\otimes c)\right]\cdot h\right\},\\ &= \tau\left\{\mathscr{D}\left[(1\otimes\tau+\tau\otimes 1)\,\mathrm{Tr}\,(A)\right]\cdot h\right\},\end{aligned}$$

and we are done. □

For the remainder of this section, we develop estimates to control the function $g \mapsto Q(g)$ that appears in (6.3.3). We will need these in the next section to prove existence of some g satisfying this equation. The first estimate is left as an exercise to the reader.

Lemma 6.25. *For any monomial $q \in \mathscr{A}$, we have $|\tau(q)| \leq 2^{\deg q}$.*

Introduce for $m \in \mathbb{Z}_{\geq 1}$ the multilinear maps $Q_m : \mathscr{A}^m \to \mathscr{A}$ defined by

$$Q_m(a_1,\ldots,a_m) = (1\otimes\tau+\tau\otimes 1)\,\mathrm{Tr}\left[(\mathscr{J}\mathscr{D}a_1)\cdots(\mathscr{J}\mathscr{D}a_m)\right].$$

We write $Q_m(a)$ for $Q_m(a,\ldots,a)$, and if $a \in \mathscr{A}_0^{(\rho)}$ with $\|a\|_\rho < \rho^2/2$ for some $\rho > 4$, we set

$$Q(\mathscr{S}a) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m+2} Q_{m+2}(\mathscr{S}a).$$

Our main goal is to prove the following proposition.

Proposition 6.26. *Let $\rho > 4$ be given, and suppose that $g \in \mathscr{A}_0^{(\rho)}$ is such that $\|g\|_\rho < \rho^2/2$. Then the series defining $Q(\mathscr{S}g)$ converges in $\|\cdot\|_\rho$, and*

$$Q(\mathscr{S}g) = (1\otimes\tau+\tau\otimes 1)\,\mathrm{Tr}\left[\mathscr{J}\mathscr{D}\mathscr{S}g - \log(1+\mathscr{J}\mathscr{D}\mathscr{S}g)\right]$$

in the sense of analytic functional calculus on the $n\times n$ matrices over the von Neumann algebra generated by $\mathscr{A}\otimes\mathscr{A}^{op}$ and the obvious trace. On its domain $\{\|g\|_\rho < \rho^2/2\}$, the function Q is locally Lipschitz. In particular,

$$\|Q(\mathscr{S}f) - Q(\mathscr{S}g)\|_\rho \leq \|f-g\|_\rho \frac{2}{\rho^2}\left[\frac{1}{(1-2\|f\|_\rho/\rho^2)(1-2\|g\|_\rho/\rho^2)} - 1\right],$$

and so we also have the bound

$$\|Q(\mathscr{S}g)\|_\rho \leq \frac{(2\|g\|_\rho/\rho^2)^2}{1-2\|g\|_\rho/\rho^2}.$$

For the proof of this proposition, we require two technical lemmas.

Lemma 6.27. *Let $g_1, \dots, g_m \in \mathscr{A}_0$, then*

$$\|Q_m(\mathscr{S}g_1, \dots, \mathscr{S}g_m)\|_\rho \leq 2(2/\rho^2)^m \prod_{k=1}^m \|g_k\|_\rho.$$

Proof. Let $q_1, \dots, q_m$ be monomials, then

$$\begin{aligned}
[(\mathscr{J}\mathscr{D}\mathscr{S}q_1)\cdots(\mathscr{J}\mathscr{D}\mathscr{S}q_m)]_{j(0)j(m)} \prod_{i=1}^m \deg q_i &= \sum_{j(1),\dots,j(m-1)} (\partial_{j(1)}\mathscr{D}_{j(0)}q_1)\cdots(\partial_{j(m)}\mathscr{D}_{j(m-1)}q_m), \\
&= \sum_{j(1),\dots,j(m-1)} \sum_{\substack{q_k = A_k X_{j(k-1)} B_k \\ B_k A_k = C_k X_{j(k)} D_k}} (C_1 \otimes D_1)\cdots(C_m \otimes D_m)\,, \\
&= \sum_{j(1),\dots,j(m-1)} \sum_{\substack{q_k = A_k X_{j(k-1)} B_k \\ B_k A_k = C_k X_{j(k)} D_k}} (C_1 \cdots C_m) \otimes (D_m \cdots D_1)\,.
\end{aligned}$$

Therefore we can estimate

$$\left\|(1\otimes\tau)\,[(\mathscr{J}\mathscr{D}\mathscr{S}q_1)\cdots(\mathscr{J}\mathscr{D}\mathscr{S}q_m)]_{j(0)j(m)}\right\|_\rho \leq \frac{1}{\prod_{i=1}^m \deg q_i} \sum_{\substack{j(1),\dots,j(m-1) \\ q_k = A_k X_{j(k-1)} B_k \\ B_k A_k = C_k X_{j(k)} D_k}} \rho^{\deg(C_1\cdots C_m)} 2^{\deg(D_m\cdots D_1)}.$$

We know that $\deg(C_k) + \deg(D_k) = \deg(q_k) - 2$, so if we write $\ell_k = \deg(C_k)$, we have $0 \leq \ell_k \leq \deg(q_k) - 2$, and $\deg(D_k) = \deg(q_k) - 2 - \ell_k$. We can use the m-tuples $(\ell_1, \dots, \ell_m)$ to parametrise the terms coming from the Jacobians $\mathscr{J}$, and since every sum $\sum_j \mathscr{D}_j q_k$ has at most $\deg(q_k)$ terms, we can simplify this estimate to

$$\begin{aligned}
\left\|(1\otimes\tau)\,[(\mathscr{J}\mathscr{D}\mathscr{S}q_1)\cdots(\mathscr{J}\mathscr{D}\mathscr{S}q_m)]_{j(0)j(m)}\right\|_\rho &\leq \sum_{\ell_1,\dots,\ell_m} \rho^{\ell_1+\cdots+\ell_m}\ 2^{\deg(q_1)-2-\ell_1+\cdots+\deg(q_m)-2-\ell_m}, \\
&= \rho^{\deg(q_1\cdots q_m)}\rho^{-2m} \sum_{\ell_1,\dots,\ell_m} \left(\frac{2}{\rho}\right)^{\deg(q_1)-2-\ell_1+\cdots+\deg(q_m)-2-\ell_m}, \\
&= \prod_{k=1}^m \left[\rho^{-2}\rho^{\deg(q_k)} \sum_{\ell_k=0}^{\deg(q_k)-2} \left(\frac{2}{\rho}\right)^{\deg(q_k)-2-\ell_k}\right], \\
&\leq \prod_{k=1}^m \left[\rho^{-2}\rho^{\deg(q_k)} \frac{1}{1-2/\rho}\right], \\
&\leq \prod_{k=1}^m \left[2\rho^{-2}\|q_k\|_\rho\right].
\end{aligned}$$

In the last line we used that $4 < \rho$, and that $\|q_k\|_\rho = \rho^{\deg(q_k)}$. By the same arguments, one can show that also

$$\left\|(\tau\otimes 1)\,[(\mathscr{J}\mathscr{D}\mathscr{S}q_1)\cdots(\mathscr{J}\mathscr{D}\mathscr{S}q_m)]_{j(0)j(m)}\right\|_\rho \leq \prod_{k=1}^m \left[2\rho^{-2}\|q_k\|_\rho\right].$$

Thus, we obtain

$$\|Q_m(\mathscr{S}g_1, \dots, \mathscr{S}g_m)\|_\rho \leq 2(2/\rho^2)^m \prod_{k=1}^m \|g_k\|_\rho.$$

Notice that this estimate is independent of n, which is somewhat surprising. Indeed, a naive application of the triangle inequality would yield an additional factor of n coming from the trace in the definition of Q_m. The reason that this factor is absent in the final estimate given above, is that our estimates earlier were rather crude, namely the sums we estimated as having at most $\prod_k \deg(q_k)$ terms. This estimate is still true when one adds sums over $j(0)$ and $j(m)$, so that this in fact also estimates the trace. □

Lemma 6.28. *Let $f, g \in \mathscr{A}_0$, then*

$$\|Q_m(\mathscr{S}f) - Q_m(\mathscr{S}g)\|_\rho \leq 2(2/\rho^2)^m \|f-g\|_\rho \sum_{k=0}^{m-1} \|g\|_\rho^k \|f\|_\rho^{m-1-k}.$$

Thus also

$$\|Q_m(\mathscr{S}g)\|_\rho \leq 2(2/\rho^2)^m \|g\|_\rho^m,$$

and hence the maps $g \mapsto Q_m(g)$ extend to all of $\mathscr{A}_0^{(\rho)}$.

Proof. Write $(\mathscr{S}g)^{(k)}$ for the k-tuple $(\mathscr{S}g, \ldots, \mathscr{S}g)$, and similarly for f. Now exploit multilinearity to write a telescoping sum,

$$\begin{aligned}
\|Q_m(\mathscr{S}g) - Q_m(\mathscr{S}f)\|_\rho &\leq \sum_{k=0}^{m-1} \left\| Q_m\left((\mathscr{S}g)^{(k)}, (\mathscr{S}f)^{(m-k)}\right) - Q_m\left((\mathscr{S}g)^{(k+1)}, (\mathscr{S}f)^{(m-k-1)}\right)\right\|_\rho, \\
&= \sum_{k=0}^{m-1} \left\| Q_m\left((\mathscr{S}g)^{(k)}, \mathscr{S}g - \mathscr{S}f, (\mathscr{S}f)^{(m-1-k)}\right)\right\|_\rho, \\
&\leq 2(2/\rho^2)^m \sum_{k=0}^{m-1} \|g\|_\rho^k \|f-g\|_\rho \|f\|_\rho^{m-1-k}.
\end{aligned}$$

The other estimate is obtained by setting $f = 0$, as then

$$\|Q_m(\mathscr{S}g)\|_\rho \leq 2(2/\rho^2)^m \left[\|g\|_\rho^m + \sum_{k=0}^{m-2} \|g\|_\rho^{k+1} 0^{m-k-1}\right] = 2(2/\rho^2)^m \|g\|_\rho^m.$$

□

Proof of Proposition 6.26. By the previous lemma and the assumptions of this proposition,

$$\|Q(\mathscr{S}g)\|_\rho \leq \sum_{m=0}^{\infty} \left(\frac{\|g\|_\rho}{\rho^2/2}\right)^{m+2} \leq \frac{1}{1 - 2\|g\|_\rho/\rho^2} < \infty,$$

so the series converges. For the functional calculus identity, simply notice that

$$x - \log(1+x) = \sum_{m \geq 0} \frac{(-x)^{m+2}}{m+2}.$$

To see that Q is locally Lipschitz, use again the previous lemma to find

$$\|Q(\mathscr{S}f) - Q(\mathscr{S}g)\|_\rho \leq \|f-g\|_\rho \frac{2}{\rho^2} \sum_{m=0}^{\infty} (2/\rho^2)^{m+1} \sum_{k=0}^{m+1} \|g\|_\rho^k \|f\|_\rho^{m+1-k}.$$

Replace the summation index m by ℓ via $m+1=k+\ell$, and since we want ℓ to range from 0 to infinity, we need to add and subtract the term in the sum with $k=\ell=0$ (this is not present above as $k+l=m+1\geq 1$). This yields, where we also let k run unbounded,

$$\begin{aligned}\|Q(\mathscr{S}f)-Q(\mathscr{S}g)\|_\rho &\leq \|f-g\|_\rho \frac{2}{\rho^2}\left[\sum_{\ell,k=0}^{\infty}(2/\rho^2)^{\ell+k}\|g\|_\rho^k\|f\|_\rho^\ell - 1\right],\\ &= \|f-g\|_\rho \frac{2}{\rho^2}\left[\frac{1}{(1-2\|f\|_\rho/\rho^2)(1-2\|g\|_\rho/\rho^2)} - 1\right].\end{aligned}$$

The other bound can be obtained by setting $f=0$. □

6.4 The Main Theorem and its Proof

At last we are able to state the main result that we have been working toward.

Theorem 6.29 (Existence of Free Monotone Transport)**.** *Let (M,τ) be the tracial von Neumann algebra generated by free standard semicircular elements $X_1,\ldots,X_n$. For every pair $\rho>\rho'>4$, there is a constant $C>0$, depending only on ρ and ρ', satisfying that if $\|W\|_{\rho+1}<C$ and $W=W^*$, then there exists a noncommutative power series $G=G^*\in\mathscr{A}^{(\rho')}$ such that*

(i) the law τ_V of $Y=\mathscr{D}G$ solves the Schwinger-Dyson equation (6.2.1) for $V=\frac{1}{2}\sum_j X_j^2+W$,

(ii) $\|Y-X\|_{\rho'}$ goes to zero if $\|W\|_{\rho+1}$ tends to zero,

(iii) the Hessian of G, $\mathscr{J}\mathscr{D}G\in M_n(M\otimes M^{op})$, is strictly positive,

(iv) there exists $H\in(\mathscr{A}^{(\rho')})^n$ with $X=H(Y)$.

Remark 6.30. Notice that part (i) of this theorem gives transport between τ and τ_V, in the sense of Definition 6.13. Part (iii) expresses that this transport is in fact monotone. One should compare the form of the transport map, namely the (cyclic) gradient of a 'strictly convex' function (having strictly positive Hessian), to the Brenier map of Theorem 1.21.

We will now prove Theorem 6.29 over the course of two lemmas, two propositions, and a theorem. As a starting point, motivated by equation (6.3.3), we are led to consider the equation for g given by

$$(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D}g)-\mathscr{N}g = W(X+\mathscr{D}g)+Q(g)+\frac{1}{2}(\mathscr{D}g)\cdot(\mathscr{D}g).$$

Assuming that $g\in\mathscr{A}_0^{(\rho)}$ for some ρ large enough, this can be rearranged into

$$\begin{aligned}\mathscr{N}g = -W(X)+\Big\{&(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D}\mathscr{S}\mathscr{N}g)-[W(X+\mathscr{D}\mathscr{S}\mathscr{N}g)-W(X)]\\ &-Q(\mathscr{S}\mathscr{N}g)-\frac{1}{2}(\mathscr{D}\mathscr{S}\mathscr{N}g)\cdot(\mathscr{D}\mathscr{S}\mathscr{N}g)\Big\},\end{aligned}$$

which is a fixed point equation for $\mathscr{N}g$. In the next lemma, we will 'change variables' to $\mathscr{N}g$ instead of g, and derive some useful properties of the right hand side of this fixed point equation.

Lemma 6.31. *Let $g\in\mathscr{A}_0^{(\rho)}$ for some $\rho>4$, assume that $\|g\|_\rho<\rho^2/2$, and let $W\in\mathscr{A}^{(R)}$ with $R\geq\rho+\|g\|_\rho$. Set $F(g)$ equal to the expression*

$$-W(X)+(1\otimes\tau+\tau\otimes 1)\operatorname{Tr}(\mathscr{J}\mathscr{D}\mathscr{S}g)-[W(X+\mathscr{D}\mathscr{S}g)-W(X)]-Q(\mathscr{S}g)-\frac{1}{2}(\mathscr{D}\mathscr{S}g)\cdot(\mathscr{D}\mathscr{S}g),$$

then the assignment $g \mapsto F(g)$ gives a well-defined function from $\{\|g\|_\rho < \rho^2/2\}$ to $\mathscr{A}^{(\rho)}$. This function is locally Lipschitz, with

$$\|F(g) - F(f)\|_\rho \leq \|g - f\|_\rho \Bigg\{ \frac{2}{\rho^2} \left[\frac{1}{(1 - 2\|g\|_\rho/\rho^2)(1 - 2\|f\|_\rho/\rho^2)} + 1 \right] + \sum_{j=1}^{n} \|\partial_j W\|_{R \otimes_\pi R} + \frac{1}{2} \Big(\|f\|_\rho + \|g\|_\rho \Big) \Bigg\},$$

and can be bounded by

$$\|F(g)\|_\rho \leq \|W\|_\rho + \|g\|_\rho \left\{ \frac{2}{\rho^2} \left[\frac{1}{(1 - 2\|g\|_\rho/\rho^2)} + 1 \right] + \sum_{j=1}^{n} \|\partial_j W\|_{R \otimes_\pi R} + \frac{1}{2}\|g\|_\rho \right\}.$$

Proof. We begin by noting some useful estimates. Looking back at the proof of Proposition 6.6, it is easy to see that $\|\mathscr{D}\mathscr{S} g\|_\rho \leq \|g\|_\rho$, and hence that $\|X + \mathscr{D}\mathscr{S} g\|_\rho \leq R$, which means that the composition $W(X + \mathscr{D}\mathscr{S} g)$ is meaningful (recall Remark 6.2). In a similar vein, one readily shows that $\sum_j \|\mathscr{D}_j \mathscr{S} g\|_\rho \leq \|g\|_\rho$. Checking on monomials using a telescoping sum (similar to the proof of Lemma 6.28), it is also not difficult to realise that

$$\|W(X + \mathscr{D}\mathscr{S} g) - W(X + \mathscr{D}\mathscr{S} f)\|_\rho \leq \sum_{j=1}^{n} \|\partial_j W\|_{R \otimes_\pi R} \|\mathscr{D}_j \mathscr{S} g - \mathscr{D}_j \mathscr{S} f\|_\rho \leq \|f - g\|_\rho \sum_{j=1}^{n} \|\partial_j W\|_{R \otimes_\pi R}.$$

Finally, consider

$$\begin{aligned} \frac{1}{2}\|(\mathscr{D}\mathscr{S} g) \cdot (\mathscr{D}\mathscr{S} g) - (\mathscr{D}\mathscr{S} f) \cdot (\mathscr{D}\mathscr{S} f)\|_\rho &\leq \sum_{i=1}^{n} \Big(\|\mathscr{D}_i \mathscr{S} g\|_\rho + \|\mathscr{D}_i \mathscr{S} f\|_\rho \Big) \|\mathscr{D}_i \mathscr{S} g - \mathscr{D}_i \mathscr{S} f\|_\rho, \\ &\leq \frac{1}{2}\|\mathscr{D}\mathscr{S} g - \mathscr{D}\mathscr{S} f\|_\rho \Big(\|\mathscr{D}\mathscr{S} g\|_\rho + \|\mathscr{D}\mathscr{S} f\|_\rho \Big), \\ &\leq \|f - g\|_\rho \frac{1}{2} \Big(\|g\|_\rho + \|f\|_\rho \Big). \end{aligned}$$

To obtain the Lipschitz condition for F, first notice that $(1 \otimes \tau + \tau \otimes 1) \operatorname{Tr}(\mathscr{J}\mathscr{D}\mathscr{S} g) = Q_1(\mathscr{S} g)$. Therefore,

$$\begin{aligned} \|F(g) - F(f)\|_\rho \leq &\|Q_1(\mathscr{S} g) - Q_1(\mathscr{S} f)\|_\rho + \|W(X + \mathscr{D}\mathscr{S} g) - W(X + \mathscr{D}\mathscr{S} f)\|_\rho \\ &+ \|Q(\mathscr{S} g) - Q(\mathscr{S} f)\|_\rho + \frac{1}{2}\|(\mathscr{D}\mathscr{S} g) \cdot (\mathscr{D}\mathscr{S} g) - (\mathscr{D}\mathscr{S} f) \cdot (\mathscr{D}\mathscr{S} f)\|_\rho. \end{aligned}$$

The first term can be estimated using Lemma 6.28, which gives a bound of $2(2/\rho^2)\|f - g\|_\rho$. The third term can instead be estimated using Proposition 6.26. The remaining two terms we have shown how to estimate above. Combining all of this yields

$$\begin{aligned} \|F(g) - F(f)\|_\rho \leq \|f - g\|_\rho 2 \left(\frac{2}{\rho^2} \right) + \|f - g\|_\rho \sum_{j=1}^{n} \|\partial_j W\|_{R \otimes_\pi R} \\ + \|f - g\|_\rho \frac{2}{\rho^2} \left[\frac{1}{(1 - 2\|f\|_\rho/\rho^2)(1 - 2\|g\|_\rho/\rho^2)} - 1 \right] + \|f - g\|_\rho \frac{1}{2} \Big(\|g\|_\rho + \|f\|_\rho \Big), \end{aligned}$$

which gives the desired estimate. The other bound follows immediately since $F(0) = -W(X)$, as

$$\|F(g)\|_\rho \leq \|W\|_\rho + \|F(g) - F(0)\|_\rho.$$

□

Next, we show that under the right additional conditions, this local Lipschitz constant can actually be uniformly bounded on some ball by a number strictly less than unity.

Lemma 6.32. *In the situation of the previous lemma, if in addition*

$$\|W\|_\rho < \frac{t}{12},\ \sum_{j=1}^{n} \|\partial_j W\|_{(\rho+t)\otimes_\pi(\rho+t)} < \frac{1}{8},\ 0 < t \leq 1, \tag{6.4.1}$$

then $\{\|g\|_\rho < t/4\}$ *is mapped into itself by* F*, which is uniformly contractive with constant at most* $7/8$*.*

Proof. Assume that $\|f\|_\rho, \|g\|_\rho < t/4$, then as $t \leq 1$, we have $\|f\|_\rho, \|g\|_\rho \leq 1/4 \leq 1$. Also, set $R = \rho + t \geq \rho + t/4 > \rho + \|g\|_\rho$. Thus,

$$\frac{1}{1 - 2\|f\|_\rho/\rho^2} \leq \frac{1}{1 - 2/4^2} = \frac{8}{7},$$

and the same for g. It follows that

$$\frac{2}{\rho^2}\left[\frac{1}{(1 - 2\|g\|_\rho/\rho^2)(1 - 2\|f\|_\rho/\rho^2)} + 1\right] \leq \frac{1}{8}\left[\left(\frac{8}{7}\right)^2 + 1\right] = \frac{113}{392} \leq \frac{1}{2}.$$

Now,

$$\begin{aligned}
\|F(g) - F(f)\|_\rho &\leq \|g - f\|_\rho \left[\frac{1}{2} + \sum_{j=1}^{n} \|\partial_j W\|_{R\otimes_\pi R} + \frac{1}{2}\left(\|f\|_\rho + \|g\|_\rho\right)\right], \\
&\leq \|g - f\|_\rho \left(\frac{1}{2} + \frac{1}{8} + \frac{1}{2}\frac{2}{4}\right), \\
&\leq \frac{7}{8}\|g - f\|_\rho.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|F(g)\|_\rho &\leq \|W\|_\rho + \|g\|_\rho \left\{\frac{2}{\rho^2}\left[\frac{1}{(1 - 2\|g\|_\rho/\rho^2)} + 1\right] + \sum_{j=1}^{n} \|\partial_j W\|_{R\otimes_\pi R} + \frac{1}{2}\|g\|_\rho\right\}, \\
&\leq \frac{t}{12} + \frac{t}{4}\left[\frac{1}{8}\left(\frac{8}{7} + 1\right) + \frac{1}{8} + \frac{1}{2}\frac{1}{4}\right], \\
&= \frac{t}{4}\left(\frac{143}{168}\right), \\
&< \frac{t}{4}.
\end{aligned}$$

So that indeed $\{\|g\|_\rho < t/4\}$ is mapped into itself. □

With these properties of F in hand, it is a relatively straightforward matter to employ a fixed point argument to obtain solutions, and to transform this solution back to the original 'variable' g.

Proposition 6.33. *Let* $W = W^* \in \mathscr{A}^{(\rho)}$ *(with* $\rho > 4$*), cyclically symmetric, be given, and assume that the conditions in* (6.4.1) *are satisfied. Then there exists a* $\hat{g}$ *with the properties*

(i) $\hat{g} \in \mathscr{A}_0^{(\rho)}$*, and* $g = \mathscr{S}\hat{g}$ *obeys* $\|g\|_\rho \leq t/4$ *and* $\|g\|_\rho \leq 3\|W\|_\rho$*,*

(ii) additionally, $g = g^*$, and g belongs to the closure of the cyclically symmetric polynomials,

(iii) $\hat{g}$ witnesses the equation $\hat{g} = \mathscr{C}\mathscr{P}_0 F(\hat{g})$,

(iv) g instead satisfies $\mathscr{C}\mathscr{P}_0 \mathscr{N} g = \mathscr{C}\mathscr{P}_0 F(\mathscr{N} g)$.

Proof. As noted in Definition 6.3, $\mathscr{C}\mathscr{P}_0$ is contractive, so that the combination $\mathscr{C}\mathscr{P}_0 F$ is also uniformly contractive on $\{\|g\|_\rho < t/4\}$ with constant at most $7/8$, and maps this ball into itself (see the previous Lemma). Notice that W is in this ball, as by assumption $\|W\|_\rho < t/12 < t/4$. Pick $\hat{g}_0 = -W(X)$ as a starting point, and then define the iterates $\hat{g}_k = \mathscr{C}\mathscr{P}_0 F(\hat{g}_{k-1})$ for all $k \in \mathbb{Z}_{\geq 1}$, which all lie in the ball. It follows immediately that the sequence of iterates is Cauchy, and hence converges to a fixed point $\hat{g}$ in the closure of the ball. Clearly, $\hat{g} \in \mathscr{A}_0^{(\rho)}$ and hence so does $g = \mathscr{S}\hat{g}$, which by construction satisfies $\|g\|_\rho \leq \|\hat{g}\|_\rho \leq t/4$. It is also trivial that $\hat{g}$ and thus g are cyclically symmetric.

To see that g is self-adjoint, it suffices to show that $\hat{g}$ is self-adjoint. For this, we show using induction that all $\hat{g}_k$ are self-adjoint. Clearly, $(\hat{g}_0)^* = -W(X)^* = -W(X) = \hat{g}_0$, and we claim that $(\hat{g}_k)^* = \mathscr{C}\mathscr{P}_0 F((\hat{g}_{k-1})^*) = \mathscr{C}\mathscr{P}_0 F(\hat{g}_{k-1}) = \hat{g}_k$ with the induction hypothesis. The proof of part (iii) of Lemma 6.19 tells us that $(\mathscr{D}_i \hat{g}_{k-1})^* = \mathscr{D}_i \hat{g}_{k-1}$, $W(X + \mathscr{D}\mathscr{S}\hat{g}_{k-1})^* = W(X + \mathscr{D}\mathscr{S}\hat{g}_{k-1})$, and also that $\mathscr{J}\mathscr{D}\mathscr{S}\hat{g}_{k-1}$ is a self-adjoint matrix. By spectral mapping, the logarithm of the identity plus this matrix is also self-adjoint. It is straightforward to check that the other operations involved in F preserve this self-adjointness, and hence that the claim is true.

Since the inequality $\|W\|_\rho < t/12$ is strict, there exists some $0 < t' < t$ such that $\|W\|_\rho < t'/12$ is also true. As the norm $\|\cdot\|_{(\rho+t)\otimes_\pi(\rho+t)}$ is increasing in t, W also satisfies the conditions (6.4.1) with t' instead of t'. It follows that $\|g\|_\rho \leq t'/4$ for all $t' > 12\|W\|_\rho$, and hence it must hold that $\|g\|_\rho \leq 12\|W\|_\rho/4 = 3\|W\|_\rho$. This proves parts (i), (ii), and (iii). Part (iv) is immediate from (iii) when one notices that $\mathscr{N} g = \hat{g}$, so we are done. □

We now argue that the solution obtained in the previous proposition indeed gives a solution to the Schwinger-Dyson equation.

Proposition 6.34. *Let $\rho > \rho' > 4$, then there is a constant $C > 0$, depending only on ρ and ρ', satisfying that if $\|W\|_{\rho+1} < C$ and $W = W^*$, there exists $f \in (\mathscr{A}^{(\rho')})^n$ solving* (6.3.2)*. In particular, f is of the form $\mathscr{D}g$ for some $g \in \mathscr{A}^{(\rho')}$. Moreover, if $\|W\|_\rho \to 0$, the ρ-norms of the corresponding solutions f vanish as well.*

Proof. Let $\rho > \tilde{\rho} > \rho'$. By Proposition 6.6, there is a constant $c_1(\rho, \tilde{\rho}) > 0$ such that $\|\mathscr{D}g\|_{\tilde{\rho}} \leq c_1(\rho, \tilde{\rho})\|g\|_\rho$. Employing similar techniques and some from the proof of Proposition 6.9, there is a constant $c_2(\rho, \tilde{\rho}) > 0$ such that

$$\|\mathscr{J}\mathscr{D}g\|_{\tilde{\rho}\otimes_\pi\tilde{\rho}} = \max_{1\leq i\leq n} \sum_{j=1}^n \|\partial_j \mathscr{D}_i g\|_{\tilde{\rho}\otimes_\pi\tilde{\rho}} \leq \sum_q |\gamma_q(g)|(\deg g)^2(\tilde{\rho})^{\deg g - 2} \leq c_2(\rho, \tilde{\rho})\|g\|_\rho.$$

Additionally, one can prove using the same ideas that

$$\sum_{j=1}^n \|\partial_j W\|_{(\tilde{\rho}+1)\otimes_\pi(\tilde{\rho}+1)} \leq c_1(\rho+1, \tilde{\rho}+1)\|W\|_{\rho+1}.$$

Setting $t = 1$ in the conditions (6.4.1), they read

$$\|W\|_{\tilde{\rho}} < \frac{1}{12},\ \sum_{j=1}^n \|\partial_j W\|_{(\tilde{\rho}+1)\otimes_\pi(\tilde{\rho}+1)} < \frac{1}{8}.$$

Obviously, $\|W\|_{\tilde{\rho}} \leq \|W\|_{\rho+1}$, and we showed above that the other condition can also be controlled by $\|W\|_{\rho+1}$. Therefore, there exists a constant $C > 0$, depending only on ρ and ρ', such that the conditions (6.4.1) with $t = 1$ are satisfied for all W with $\|W\|_{\rho+1} < C$. Since we considered the combination $\mathscr{C}\mathscr{P}_0 F$ in the previous proposition, we may assume without loss of generality that W is also cyclically symmetric, and hence all the conditions of this previous proposition are met. It follows that there exists $g \in \mathscr{A}_0^{(\tilde{\rho})} \subset \mathscr{A}_0^{(\rho')}$, solving $\mathscr{C}\mathscr{P}_0\mathscr{N} g = \mathscr{C}\mathscr{P}_0 F(\mathscr{N} g)$, and with $\mathscr{D} g \in \mathscr{A}_0^{(\rho')}$ by the estimate above. Set $f = \mathscr{D} g$, then by the above (perhaps choosing C smaller) we may assume that $\|\mathscr{J} f\|_{\rho'\otimes_\pi \rho'} < 1$, as

$$\|\mathscr{J} f\|_{\rho'\otimes_\pi\rho'} \leq c_2(\tilde{\rho},\rho')\|g\|_{\tilde{\rho}} \leq 3c_2(\tilde{\rho},\rho')\|W\|_{\tilde{\rho}} \leq 3c_2(\tilde{\rho},\rho')\|W\|_{\rho+1}.$$

Now, we want to show that $f = \mathscr{D} g$ witnessing $\mathscr{C}\mathscr{P}_0\mathscr{N} g = \mathscr{C}\mathscr{P}_0 F(\mathscr{N} g)$ implies that it also satisfies (6.3.2). By Proposition 6.20, it suffices to show that (6.3.3) holds. However, this is immediate when applying $\mathscr{D}$ to $\mathscr{C}\mathscr{P}_0\mathscr{N} g = \mathscr{C}\mathscr{P}_0 F(\mathscr{N} g)$ and rearranging terms, and recalling that $\mathscr{D} = \mathscr{D}\mathscr{C}\mathscr{P}_0$. □

Since our solution has the form of a perturbation of the identity, it should also be possible to invert it. The next theorem, which is a noncommutative inverse function theorem, affirms that this is the case.

Theorem 6.35 (Inverse Function Theorem)**.** *Consider a triple $0 < \rho' < \rho < R$, and $f \in (\mathscr{A}^{(R)})^n$, $Y \in (\mathscr{A}^{(\rho')})^n$ with $\|Y\|_{\rho'} \leq \rho$. Then there exists a constant $C > 0$, only dependent on ρ', ρ, and R, so that whenever $\|f\|_R < C$, there exists a $H \in (\mathscr{A}^{(\rho)})^n$ such that $X = H(Y)$ solves the equation $Y = X + f(X)$.*

Proof. Rewrite the equation as $X = Y - f(X)$, which can be iterated. Define $H_0 = X$, and $H_k = X - f \circ H_{k-1}$, for all $k \in \mathbb{Z}_{\geq 1}$. So $H_1 = X - f(X)$, $H_2 = X - f(X - f(X))$, and so on. First, we claim that there exists a constant $\tilde{C} = C(R, R', \|f\|_R) > 0$, such that for all k with $\max_{1\leq \ell\leq k}\|H_\ell\|_\rho \leq R' < R$ (note that $\|H_0\|_\rho = \rho < R$), it holds that

$$\|H_{k+1} - H_k\|_\rho \leq \tilde{C}\|H_k - H_{k-1}\|_\rho.$$

We prove this as follows,

$$\begin{aligned}
\Big\|(H_{k+1} - H_k)_j\Big\|_\rho &= \|f_j \circ H_k - f_j \circ H_{k-1}\|_\rho, \\
&\leq \sum_q |\gamma_q(f_j)| \|q \circ H_k - q \circ H_{k-1}\|_\rho, \\
&= \sum_{\ell=1}^{\infty} \sum_{i(1),\ldots,i(\ell)} \left|\gamma^j_{i(1),\ldots,i(\ell)}\right| \Big\|(H_k)_{i(1)} \cdots (H_k)_{i(\ell)} - (H_{k-1})_{i(1)} \cdots (H_{k-1})_{i(\ell)}\Big\|_\rho, \\
&\leq \sum_{\ell=1}^{\infty} \sum_{i(1),\ldots,i(\ell)} \sum_{m=1}^{\ell} \left|\gamma^j_{i(1),\ldots,i(\ell)}\right| \Big\|(H_k)_{i(1)} \cdots \big[(H_k)_{i(m)} - (H_{k-1})_{i(m)}\big] \cdots (H_{k-1})_{i(\ell)}\Big\|_\rho, \\
&\leq \|H_k - H_{k-1}\|_\rho \sum_q |\gamma_q(f_j)| (\deg g)(R')^{\deg q - 1}, \\
&\leq \left[\|f\|_R \sup_{\ell\geq 0} \ell (R')^{-1} \left(\frac{R'}{R}\right)^\ell\right] \|H_k - H_{k-1}\|_\rho,
\end{aligned}$$

and the constant in square brackets is finite by the reasoning of Proposition 6.6. It is easy to see that $\|H_1 - H_0\|_\rho = \|f\|_\rho$, and we can iterate the estimate above to find

$$\|H_{k+1} - H_0\|_\rho \leq \tilde{C}^k \|f\|_\rho.$$

Thus if we make C so small that $\tilde{C} < 1$, it holds that

$$\|H_{k+1} - H_0\|_\rho \leq \frac{1}{1-\tilde{C}}\|f\|_\rho \leq \frac{1}{1-\tilde{C}}\|f\|_R < \frac{C}{1-\tilde{C}}.$$

Since $\|H_0\|_\rho = \rho$, if we make C so small that $\rho + C/(1-\tilde{C}) \leq R'$, it holds that $\|H_\ell\|_\rho \leq R'$ for all $\ell \geq 0$.

Now fix C such that all of the above is satisfied, then the sequence of H_ks is Cauchy and it converges to a limit $H \in (\mathscr{A}^{(\rho)})^n$. This H has the properties that $\|H\|_\rho \leq R'$ and $H(X) = X - f(H(X))$. Since both $\|Y\|_{\rho'}, \|X\|_{\rho'} \leq \rho$, we can write that

$$X + f(X) = Y = H(Y) + f(H(Y)).$$

Suppose that $X \neq H(Y)$, then

$$\|X - H(Y)\|_{\rho'} = \|f(X) - f(H(Y))\|_{\rho'} \leq \tilde{X}\|X - H(Y)\|_{\rho'} < \|X - H(Y)\|_{\rho'},$$

which is a contradiction, so we conclude that $X = H(Y)$. □

Finally, we are ready to combine all of the above to prove the existence of free monotone transport.

Proof of Theorem 6.29. By Proposition 6.34, if we set $G = \sum_j X_j^2 + g$, with g as in that proposition, then $\mathscr{D}G = X + f$ solves the Schwinger-Dyson equation with potential $V = \frac{1}{2}\sum_j V_j^2 + W$. We now want to use Theorem 6.15 to conclude that this solution is unique, so that that the law of Y is indeed τ_V. Notice that as $\|q\| \leq 2^{\deg q}$ and $\|q\|_\rho = \rho^{\deg q}$ for any monomial q, it holds that $\|\cdot\| \leq \|\cdot\|_\rho$ whenever $\rho \geq 2$, which is the case here. Therefore, $\|Y\| = \|X + f(X)\| \leq 2 + \|f(X)\| \leq 2 + \|f\|_\rho$, and in particular

$$\left|\tau(Y_{i(1)} \cdots Y_{i(k)})\right| \leq \|Y\|^k \leq (2 + \|f\|_\rho)^k.$$

Thus the joint law of Y, which can be identified with the restriction of τ to the $*$-algebra that Y generates, satisfies the bound in the hypothesis of Theorem 6.15 for $R = 2 + \|f\|_\rho > 1$. Since we control the $(\rho+1)$-norm of W, we need that $\rho + 1 > 2R = 4 + 2\|f\|_\rho$ for us to benefit from the conclusion of Theorem 6.15, thus we want $2R < 5$. As $\|f\|_\rho$ goes to zero as $\|W\|_{\rho+1}$ goes to zero, this can be arranged if the constant C is made small enough. This proves (i). Part (ii) follows directly from the same results. Part (iii) holds because $\mathscr{J}\mathscr{D}G = 1 + \mathscr{J}f$ and $\|\mathscr{J}f\|_{\rho' \otimes_\pi \rho'} < 1$. Part (iv) follows from the inverse function theorem. □

7 Free Monotone Transport Applied to q-Gaußians

7.1 On the C*-Algebra Generated by the q-mutation Relations

At last, we return to the q-mutation relations

$$a_i a_j^* - q a_j^* a_i = \delta_{ij} I,$$

of Chapter 3. In this context, q is a real number strictly between 1 and -1 (not a monomial in $\mathscr{A}$). However, before we tackle the isomorphism question for the q-Gaußian von Neumann algebras Γ_q, which are generated by the combinations $a_i + a_i^*$, we first give some attention to the C*-algebras G_q generated by the a_i themselves. Note that we do not consider the universal C*-algebra generated by the abstract q-mutation relations, but rather the 'reduced' version generated by the annihilation and creation operators acting on Fock spaces as constructed in Chapter 3. For definiteness, we take the 'left' versions of all these operators, as before. In this section, we will sketch[5] the results of the 1993 paper [30], where they partially settled the isomorphism problem for the G_q. We include this section not only for thematic reasons, but also because along the way we will obtain some useful relations and estimates.

Since $q = 0$ will act as the 'base case' for the results in this section, we introduce some special notation for it. For $q = 0$, write $v_1, \ldots, v_n$ instead of $a_1, \ldots, a_n$ for the n-tuple of operators witnessing the q-mutation relations, assuming throughout this chapter that $n \geq 2$. Additionally define $p = \sum_{i=1}^n v_i^* v_i$, which is the projection onto the orthogonal complement of the vacuum Ω. We set $o = \sum_{j=1}^n a_i^* a_i$, which is a positive operator and the deformed version of p. The results of Dykema & Nica (see the Theorem in Section 1 of [30]) can now be summarised as follows.

Theorem 7.1.

(i) The equation

$$r^2 = p + q \sum_{i,j=1}^n (v_i r v_j)^* (v_j r v_i),$$

together with the demand that the subspaces spanned by the tensors of a given length are invariant, defines a unique positive operator $r \in B(\mathcal{F}_0(\mathcal{H}))$.

(ii) We can construct a unitary map $\mathcal{U} : \mathcal{F}_q(\mathcal{H}) \to \mathcal{F}_0(\mathcal{H})$, satisfying $r = \mathcal{U} o^{1/2} \mathcal{U}^*$ and $v_i r = \mathcal{U} a_i \mathcal{U}^*$ (for all $1 \leq i \leq n$).

(iii) We have the inclusion $G_0 \subset \mathcal{U} G_q \mathcal{U}^*$ of C-algebras.*

(iv) If q is such that[6]

$$q^2 < 1 - 2|q| + 2|q|^4 - 2|q|^9 + \cdots + 2(-1)^k |q|^{k^2} + \ldots,$$

then $G_0 \cong G_q$.

Remark 7.2. It follows from part (ii) of the theorem that $v_1 r, \ldots, v_n r$ in $B(\mathcal{F}_0(\mathcal{H}))$ witness the q-mutation relations. The relation between this fact and the equation for r^2 can be seen as follows. Notice first that r annihilates the vacuum, so that

$$r^2 = p r^2 p = \left(\sum_{i=1}^n v_i^* v_i \right) r^2 \left(\sum_{j=1}^n v_j^* v_j \right) = \sum_{i,j=1}^n v_j^* \left[(v_j r)(v_i r)^* \right] v_i.$$

[5] Accordingly, for all details and proofs of claims not included here, we refer to their paper.

[6] One can check numerically that this bound is satisfied, for instance, if $|q| < 0.44$.

Using the q-mutation relations for the $v_k r$, this becomes

$$r^2 = \sum_{i,j=1}^{n} v_j^* \left[\delta_{ij} I + q(v_i r)^*(v_j r)\right] v_i = p + q \sum_{i,j=1}^{n} (v_i r v_j)^*(v_j r v_i).$$

Introduce the notation $\mathcal{E}_q^{(m)}$ for the subspace of $\mathcal{F}_q(\mathcal{H})$ spanned by the elementary tensors of length m, and note that $\dim(\mathcal{E}_q^{(m)}) = n^m$. While the elementary tensors of length m form by definition a basis of $\mathcal{E}_q^{(m)}$, they are in general not an orthogonal basis. Nevertheless, the $n^m \times n^m$ matrix $\mathcal{G}_m$, whose elements are their the inner products, is strictly positive due to the fact that $\langle \cdot, \cdot \rangle_q$ is an inner product (and the elementary tensors are linearly independent). It is not hard to see that the operator o defined above maps any $\mathcal{E}_q^{(m)}$ into itself, and that we can therefore define $o^{(m)} \in B(\mathcal{E}_q^{(m)})$ as its restriction, and represent this operator by an $n^m \times n^m$ matrix O_m with respect to the basis of elementary tensors. The operators $o^{(m)}$ also admit a nice description in terms of the action U of the permutation group S_m discussed in Definition 3.4. Let $\sigma(k, \ell)$, for $1 \leq k \leq \ell \leq m$, be the cycle that send k to $k+1$, $k+1$ to $k+2$, and so on, until $\ell - 1$ is sent to ℓ, and ℓ is sent to k. One can then show that

$$o^{(m)} = \sum_{k=1}^{m} q^{k-1} U_{\sigma(1,k)}.$$

This information can be combined to give a recursion relation for the matrices $\mathcal{G}_m$.

Lemma 7.3. *For all $m \geq 1$, we have the recursion relation*

$$\mathcal{G}_m = (I_n \otimes \mathcal{G}_{m-1}) O_m.$$

Proof. As discussed in Chapter 3 (Lemma 3.8 and Proposition 3.12), any permutation σ in S_m can be decomposed into a cycle $\sigma(1, k)$ and a permutation $\sigma' \in S_{m-1}$, in such a way that the number of inversions add. Thus we have the identity

$$\sum_{\sigma \in S_m} q^{|\sigma|} U_\sigma = \left(\sum_{\sigma' \in S_{m-1}} q^{|\sigma'|} U_{\sigma'} \right) \left(\sum_{k=1}^{m} q^{k-1} U_{\sigma(1,k)} \right).$$

The left hand side of this identity is the operator $P_q^{(m)}$ discussed in Definition 3.4, which determines the inner product on the subspace $\mathcal{E}_q^{(m)}$. Therefore, the matrix of $P_q^{(m)}$ is precisely $\mathcal{G}_m$, and recognising also $o^{(m)}$ on the right hand side yields

$$\mathcal{G}_m = (I_n \otimes \mathcal{G}_{m-1}) O_m.$$

□

Using this technical result, one can construct the unitary $\mathcal{U}$ recursively on the subspaces $\mathcal{E}_q^{(m)}$. In the vacuum sector, define $\mathcal{U}^{(0)} : \mathcal{E}_q^{(0)} \to \mathcal{E}_0^{(0)}$ by $\mathcal{U}^{(0)}\Omega = \Omega$. The rest of the operator follows from imposing the recursion relation $\mathcal{U}^{(m)} = (I_n \otimes \mathcal{U}^{(m-1)}) \left[o^{(m)}\right]^{1/2}$. It is then merely a computation to verify the properties in part (ii) of the theorem. In fact, we use part (ii) to define the solution r of part (i). We already indicated how the equation for r^2 follows in the remark after the theorem, and uniqueness is straightforward. This concludes our discussion of items (i) and (ii).

To prove the inclusion of part (iii), all we need is to show that the v_i are in $\mathcal{U} G_q \mathcal{U}^*$. For this, we would like to relate them to the $v_i r$ and r itself, of which we know that they are in $\mathcal{U} G_q \mathcal{U}^*$. Since the v_i annihilate the vacuum, we have $v_i = v_i p$. As r is related to p through the equation for r^2, we might hope to construct p out of r through functional calculus. We know that r also annihilates the

vacuum, but if we assume that 0 is an isolated point in its spectrum, we can do this. Take φ to be the continuous function on $\sigma(r)$ defined by $\varphi(0) = 0$ and $\varphi(z) = 1/z$ else, then $r\varphi(r) = p$. We could then conclude the argument by noting that $v_i = (v_i r)\varphi(r) \in \mathcal{U}G_q\mathcal{U}^*$. Thus it remains to show that $\ker(r) = \mathbb{C}\Omega$ and that 0 is isolated in the spectrum. This is the content of the next lemma, concluding our discussion of part (iii).

Lemma 7.4. *As usual, write $r^{(m)}$ for the restriction of r to $\mathcal{E}_q^{(m)}$. Then for every $m \geq 1$ we have the bounds*

$$\left(\frac{1}{1-|q|}\prod_{k=1}^{\infty}\frac{1-|q|^k}{1+|q|^k}\right) I \leq \left[r^{(m)}\right]^2 \leq \frac{1}{1-|q|} I,$$

which are uniform in m, and the coefficient of the identity operator in the lower bound is strictly positive.

Proof. We may show instead that these bounds hold for the operators $o^{(m)}$, as they are related to the $[r^{(m)}]^2$ by conjugation with a unitary. The upper bound is immediate from the expression for $o^{(m)}$ in terms of permutations. For the lower bound, we first prove that $o^{(m)}$ is invertible, and then give an upper bound on this inverse. Consider, for $2 \leq \ell \leq m$, the identity

$$\left(\sum_{k=1}^{\ell} q^{k-1}\sigma(1,k)\right)\left(I - q^{\ell-1}\sigma(1,\ell)\right) = \left(I - q^{\ell}\sigma(2,\ell)\right)\left(\sum_{k=1}^{\ell-1} q^{k-1}\sigma(1,k)\right),$$

which arises from $\sigma(1,i)\sigma(1,j) = \sigma(2,j)\sigma(1,i-1)$ (where $2 \leq i \leq j \leq m$). Notice that the operator $I - q^{\ell-1}\sigma(1,\ell)$ is invertible, because it is close to the identity, so that we may multiply our identity from the right by this inverse. This gives

$$\sum_{k=1}^{\ell} q^{k-1}\sigma(1,k) = \left(I - q^{\ell}\sigma(2,\ell)\right)\left(\sum_{k=1}^{\ell-1} q^{k-1}\sigma(1,k)\right)\left(I - q^{\ell-1}\sigma(1,\ell)\right)^{-1},$$

and this can be iterated to yield

$$\sum_{k=1}^{m} q^{k-1}\sigma(1,k) = \left[\prod_{i=0}^{m-2}\left(I - q^{m-i}\sigma(2,n-i)\right)\right]\left[\prod_{j=0}^{m}\left(I - q^{j}\sigma(1,j+1)\right)^{-1}\right],$$

which can be inverted, as it is a product of operators that are close to the identity. Thus,

$$\begin{aligned}
\left\|\left[o^{(m)}\right]^{-1}\right\| &\leq \left[\prod_{i=0}^{m-2}\left\|\left(I - q^{m-i}\sigma(2,n-i)\right)^{-1}\right\|\right]\left[\prod_{j=0}^{m}\left\|I - q^{j}\sigma(1,j+1)\right\|\right], \\
&\leq \left[\prod_{i=0}^{m-2}\frac{1}{1-|q|^{m-i}}\right]\left[\prod_{j=0}^{m} 1+|q|^{j}\right], \\
&\leq (1-|q|)\prod_{k=1}^{\infty}\frac{1+|q|^k}{1-|q|^k},
\end{aligned}$$

so that we obtain the desired lower bound. □

We will not have much to say about part (iv), as it plays no role in what follows. Note that the reverse inclusion (with respect to (iii)) is true if $r \in G_0$. It turns out that one can construct a sequence in G_0 that approximates r, but one can only ensure that this sequence is Cauchy if

$\liminf_{m\to\infty}\lambda_m(q) > q^2/(1-|q|)$, where $\lambda_m(q)$ is the smallest eigenvalue of $[r^{(m)}]^2$ (with $m \geq 1$). By the lower bound in the previous lemma, this is the case when

$$q^2 < \prod_{k=1}^{\infty}\frac{1-|q|^k}{1+|q|^k} = \sum_{k\in\mathbb{Z}}(-1)^k|q|^{k^2} = 1-2|q|+2|q|^4-2|q|^9+\cdots+2(-1)^k|q|^{k^2}+\ldots,$$

with the first equality coming from Corollary 2.10 in [3].

7.2 Existence of Analytic Conjugate Variables for Small q

The strategy of this section is based on [78] (Section 4) and [26] (Sections 4.2 and 4.3).

Definition 7.5. Define the self-adjoint operator $\Xi_q = \sum_{m\geq 0} q^m P^{(m)}$, with $P^{(m)}$ the projector onto $\mathcal{E}_q^{(m)}$. Notice that it is trace class when $qN < 1$, and that it is Hilbert-Schmidt when $q^2N < 1$. If the latter condition is satisfied, let $\partial_j^{(q)} : \mathscr{A} \to \text{HS}$, with $1 \leq j \leq n$, be the derivation defined by $\partial_j^{(q)}(X_i) = \delta_{ij}\Xi_q$.

Remark 7.6. The motivation of this definition is as follows. A natural source of derivations on $\mathscr{A}$ is to take the bracket $[\cdot, L]$ with some element L in $B(\mathcal{F}_q(\mathcal{H}))$. Out of all the possible choices, a meaningful one could be to take the right creation operators $L = a_{r,j}^*$, and define the derivations $\partial_j^{(q)}(\cdot) = [\cdot, a_{r,j}^*]$. Indeed, as the notation suggests, we have

$$[X_i, a_{r,j}^*] = [a_i + a_i^*, a_{r,j}^*] = [a_i, a_{r,j}^*],$$

and using their definitions we see

$$\begin{aligned} a_i a_{r,j}^*(e_{i(1)}\otimes\cdots\otimes e_{i(k)}) &= a_i(e_{i(1)}\otimes\cdots\otimes e_{i(k)}\otimes e_j), \\ &= \sum_{\ell=1}^{k} q^{m-1}\delta_{i(\ell)i}(e_{i(1)}\otimes\cdots\otimes\hat{e}_{i(\ell)}\otimes\cdots\otimes e_{i(k)}\otimes e_j) + q^k\delta_{ij}(e_{i(1)}\otimes\cdots\otimes e_{i(k)}), \\ &= (a_{r,j}^* a_i + \delta_{ij}q^k)(e_{i(1)}\otimes\cdots\otimes e_{i(k)}), \end{aligned}$$

and thus $[X_i, a_{r,j}^*] = \delta_{ij}\Xi_q$.

Remark 7.7. The main advantage of the derivation $\partial_j^{(q)}$ is due to the following identification. The Hilbert-Schmidt operators form a self-adjoint ideal in $B(\mathcal{F}_q(\mathcal{H}))$, and thus we can consider it as a Γ_q-bimodule, with real structure given by the adjoint. Recall that $L^2(\Gamma_q) \cong \mathcal{F}_q(\mathcal{H})$ via Wick operators, and consider the map from $L^2(\Gamma_q)\otimes L^2(\Gamma_q)$ to HS given by assigning to $a\otimes b$ the finite rank operator $a\omega(b\cdot)$ (where ω denotes the vacuum expectation $\langle\cdot\Omega,\Omega\rangle$). This gives an isomorphism as Γ_q-bimodules of HS $\cong L^2(\Gamma_q)\otimes L^2(\Gamma_q)$ which intertwines the real structures, if we define $c(a\otimes b)d = (ca)\otimes(bd)$ and the real structure $\mathcal{J}(a\otimes b) = b^*\otimes a^*$. It follows that we can regard $\partial_j^{(q)}$ as a densely defined unbounded operator from $L^2(\Gamma_q)$ to $L^2(\Gamma_q)\otimes L^2(\Gamma_q)$, just like the noncommutative partial derivatives ∂_j. It is worthwhile to notice that under this identification, the elementary tensor $1\otimes 1$ is identified with the vacuum projection operator $P^{(0)}$, hence $\partial_j^{(0)} = \partial_j$.

The following proposition can be found in [78] as Lemma 10.

Proposition 7.8. *The $\partial_j^{(q)}$ are real derivations, closable, and $\partial_j^{(q)*}(1\otimes 1) = X_j$.*

When we dealt with the closability of the noncommutative partial derivatives in Chapter 4, we managed to prove a result analogous to the above proposition using a formula for the action of ∂_j^* on elementary tensors (see Theorem 4.7). We would like a similar formula for $\partial_j^{(q)*}$, and indeed there exists one.

Lemma 7.9. *On $\Gamma_q \otimes \Gamma_q^{op}$ we have the formula*

$$\partial_j^{(q)*}(a \otimes b) = aX_j b - (1 \otimes \omega)(\partial_j^{(q)} a)\ b - a\ (\omega \otimes 1)(\partial_j^{(q)} b).$$

Proof. We show that the formula holds (following [88]) for $a, b \in \mathscr{A}$, which implies that $\mathscr{A} \otimes \mathscr{A}^{op}$ is contained in the domain of $\partial_j^{(q)*}$. For the extension to all of $\Gamma_q \otimes \Gamma_q^{op}$, see the proof of the claim in step 1 in the proof of Theorem 34 of [26].

Let $c \in \mathscr{A}$, $\eta \in D(\partial_j^{(q)*})$, and consider

$$\langle \eta(1 \otimes b), \partial_j^{(q)} c\rangle = \langle \eta, \partial_j^{(q)}(c)(1 \otimes b^*), \rangle$$

using the trace property of ω. Now, exploit the Leibniz rule to write

$$\begin{aligned}
\langle \eta(1 \otimes b), \partial_j^{(q)} c\rangle &= \langle \eta, \partial_j^{(q)}(cb^*) - (c \otimes 1)\partial_j^{(q)}(b^*)\rangle, \\
&= \langle \partial_j^{(q)*}(\eta), cb^*\rangle - \langle \eta\ \partial_j^{(q)}(b^*)^*, c \otimes 1\rangle, \\
&= \langle \partial_j^{(q)*}(\eta) b - (1 \otimes \omega)\left[\eta\ \partial_j^{(q)}(b^*)^*\right], c\rangle.
\end{aligned}$$

Recall now that $\partial_j^{(q)}$ is a real derivation, with respect to the real structure $\mathcal{J}(a \otimes b) = b^* \otimes a^* = t(a \otimes b)^*$ on $L^2(\Gamma_q) \otimes L^2(\Gamma_q)$, thus by density we conclude that

$$\partial_j^{(q)*}\left[\eta(1 \otimes b)\right] = \partial_j^{(q)*}(\eta) b - (1 \otimes \omega)\left[\eta\ t(\partial_j^{(q)} b)\right].$$

Using similar arguments, one can show also that

$$\partial_j^{(q)*}\left[(a \otimes 1)\eta\right] = a\ \partial_j^{(q)*}(\eta) - (\omega \otimes 1)\left[t(\partial_j^{(q)} a)\eta\right].$$

Taking $\eta = 1 \otimes 1$ (which is in the domain by the Proposition 7.8) shows that $a \otimes 1$ and $1 \otimes b$ themselves are indeed in the domain of $\partial_j^{(q)*}$. We now combine these two identities as follows,

$$\begin{aligned}
\partial_j^{(q)*}(a \otimes b) &= \partial_j^{(q)*}\left[(a \otimes 1)(1 \otimes b)\right], \\
&= a\ \partial_j^{(q)*}(1 \otimes b) - (\omega \otimes 1)\left[t(\partial_j^{(q)} a)(1 \otimes b)\right], \\
&= a\ \partial_j^{(q)*}(1 \otimes 1) b - \left[(\omega \otimes 1) \circ t(\partial_j(q) a)\right] b - a\left[(1 \otimes \omega) \circ t(\partial_j^{(q)} b)\right], \\
&= aX_j b - (1 \otimes \omega)(\partial_j^{(q)} a)\ b - a\ (\omega \otimes 1)(\partial_j^{(q)} b),
\end{aligned}$$

where we used that $\partial_j^{(q)*}(1 \otimes 1) = X_j$. □

Remark 7.10. In the sequel, we will need to slightly expand our notion of a noncommutative (formal) power series to include expressions of the form

$$\sum_{k,\ell=0}^{\infty} \sum_{\substack{|I|=k \\ |J|=\ell}} \gamma_{IJ}\left(T_{i(1)} \cdots T_{i(k)}\right) \otimes \left(T_{j(1)} \cdots T_{j(\ell)}\right).$$

It is not difficult to convince oneself that all our (necessary) results about noncommutative power series so far have natural extensions to include these objects.

We now sketch a proof of the key technical Lemma of this section, where we establish that for a certain range of (small) q, the operator Ξ_q can be viewed as a noncommutative power series. This will lead to the required regularity results for the conjugate variables.

Lemma 7.11. *Fix some $R+1>5$. There exists a function $\mu_{R+1}(q,n)$, vanishing as $|q|$ approaches zero, such that if $\mu_{R+1}(q,n)<1$, there exists a noncommutative power series $\Xi_q(T)$, with radius of convergence $R+1$, satisfying $\Xi_q(X)=\Xi_q$. Moreover, there also exists a function $\nu_{R+1}(q,n)$, vanishing as $|q|$ approaches zero, such that*

$$\|1\otimes 1-\Xi_q(T)\|_{R+1}\leq \nu_{R+1}(q,n)$$

Proof. We follow [78], see Corollary 29 in [26] for an improved construction. Recall that $\Xi_q=\sum_m q^m P^{(m)}$, so essentially our goal is to construct the projectors $P^{(m)}$, via the identification of HS with $L^2(\Gamma_q)\otimes L^2(\Gamma_q)$. This amounts to finding a way of constructing ONBs for the subspaces $\mathcal{E}_q^{(m)}$. As we can identify $L^2(\Gamma_q)$ with $\mathcal{F}_q(\mathcal{H})$ via Wick operators, we have two options. Either we construct ONBs by applying Gramm-Schmidt (with respect to $\langle\cdot,\cdot\rangle_q$) to the basis of elementary tensors, or we try to do something similar on the level of the Wick operators induced by these elementary tensors. We will opt for the latter.

Let $\{e_i\}_{i=1}^n$ be an ONB for $\mathcal{H}$ (recall $n=\dim\mathcal{H}$), and write $W_{i(1),\ldots,i(m)}$ for the Wick operator $W(e_{i(1)}\otimes\ldots\otimes e_{i(m)})$, uniquely defined by $W_{i(1),\ldots,i(m)}\Omega=e_{i(1)}\otimes\ldots\otimes e_{i(m)}$. It is straightforward to verify that these operators satisfy the recursive relation

$$W_{i(1),\ldots,i(m)}=X_{i(1)}W_{i(2),\ldots,i(m)}-\sum_{j=2}^m q^{j-2}\delta_{i(1)i(j)}W_{i(2),\ldots,\hat{i(j)},\ldots,e(m)}.$$

From here we see that the W's are polynomials in the q-semicircular variables X. Write $c_W(m,k)$ for the maximal absolute value of a coefficient of a monomial of degree k in a Wick operator of length m. It is simple to see that the degree of $W_{i(1),\ldots,i(m)}$ is m, and so we restrict attention to $k\leq m$. By induction, one can prove that (see Lemma 11 in [78])

$$c_W(m,k)\leq 2^{m-k}\left(\frac{1}{1-|q|}\right)^{m-k},\ \left\|W_{i(1),\ldots,i(m)}\right\|_2^2\leq 2^m(1-|q|)^{-m}.$$

The claim is now that there exist elements $V_{i(1),\ldots,i(m)}\in\mathscr{A}$ of degree m, such that the vectors $V_{i(1),\ldots,i(m)}\Omega$ form an ONB of $\mathcal{E}_q^{(m)}$, and

$$c_V(m,k)\leq\left(\frac{2n}{\sqrt{1-2|q|}}\right)^m\left(\frac{1-|q|}{2}\right)^k.$$

To prove the claim, consider again the matrices $\mathcal{G}_m$, and recall that we proved the recursion

$$\mathcal{G}_m=(I_n\otimes\mathcal{G}_{m-1})O_m,$$

in Lemma 7.3. We also showed that O_m is positive and provided a lower bound on its spectrum. Combining these two pieces of data, we have that the least eigenvalue of $\mathcal{G}_m$, denoted g_m satisfies

$$g_m\geq\left(\frac{1}{1-|q|}\prod_{k=1}^\infty\frac{1-|q|^k}{1+|q|^k}\right)^m=\left(\frac{1}{1-|q|}\sum_{k\in\mathbb{Z}}|q|^{k^2}\right)^m=\left(\frac{1}{1-|q|}\left[1-\sum_{k=1}^\infty|q|^{k^2}\right]\right)^m.$$

This can be estimated from below as

$$g_m\geq\left(\frac{1}{1-|q|}\left[1-\sum_{k=1}^\infty|q|^k\right]\right)^m=\left(\frac{1}{1-|q|}\left[1-\frac{|q|}{1-|q|}\right]\right)^m=\left(\frac{1-2|q|}{(1-|q|)^2}\right)^m.$$

Hence if we define $\mathcal{B} = \mathcal{G}_m^{-1/2}$, then $|\mathcal{B}_{IJ}| \leq g_m^{-1/2}$ for all multi-indices I and J on $\{1, \ldots, n\}$ of length m, and a simple computation shows that $V_I = \sum_J \mathcal{B}_{JI} W_J$ are the polynomials we are looking for (Lemma 12 in [78]). The estimates on their coefficients follow from

$$c_V(m,k) \leq n^m g_m^{-1/2} c_W(m,k) \leq n^m \left(\frac{(1-|q|)^2}{1-2|q|}\right)^{m/2} 2^{m-k} \left(\frac{1}{1-|q|}\right)^{m-k} = \left(\frac{2n}{\sqrt{1-2|q|}}\right)^m \left(\frac{1-|q|}{2}\right)^k.$$

In Lemma 3.8, we computed the norms of the annihilation and creation operators, and this tells us that $\|X\| \leq 2/\sqrt{1-|q|} \leq 2/(1-|q|) =: R_0$. Suppose that $\alpha > 1$, and consider $0 < \rho < \alpha R_0$, then

$$\left\|V_{i(1),\ldots,i(m)}\right\|_\rho \leq \sum_{k=0}^m n^k c_V(m,k) \rho^k \leq \left(\frac{2n\alpha}{\sqrt{1-2|q|}}\right)^m \sum_{k=0}^m \left(\frac{n\rho}{\alpha R_0}\right)^k \leq \frac{1}{1-n\rho/(\alpha R_0)} \left(\frac{2n\alpha}{\sqrt{1-2|q|}}\right)^m.$$

Therefore

$$\left\|V_{i(1),\ldots,i(m)} \otimes V_{i(1),\ldots,i(m)}\right\|_\rho \leq \left\|V_{i(1),\ldots,i(m)}\right\|_\rho^2 \leq \frac{1}{(1-n\rho/(\alpha R_0))^2} \left(\frac{4n^2\alpha^2}{1-2|q|}\right)^m,$$

and the power series

$$\Xi_q(T) := \sum_{m=0}^\infty q^m \sum_{i(1),\ldots,i(m)} V_{i(1),\ldots,i(m)}(T) \otimes V_{i(1),\ldots,i(m)}(T),$$

has finite ρ-norm if $\rho < \alpha R_0$ and

$$\frac{4n^3\alpha^2|q|}{1-2|q|} < 1.$$

It is not difficult to see that under the identification of HS with $L^2(\Gamma_q) \otimes L^2(\Gamma_q)$, $\Xi_q(X) = \Xi_q$ (see the text at the start of this proof). Choosing α such that $\alpha R_0 > R+1$ and then setting $\rho = R+1$ gives $\mu_{R+1}(q,n)$.

To find the function $\nu_{R+1}(q,n)$ it suffices to note that $1 \otimes 1$ is precisely the zero-th order term in $\Xi_q(T)$, so that

$$\|1 \otimes 1 - \Xi_q(T)\|_{R+1} \leq \frac{1}{(1-n\rho/(\alpha R_0))^2} \sum_{m=1}^\infty \left(\frac{4n^2\alpha^2}{1-2|q|}\right)^m = \frac{1}{(1-n\rho/(\alpha R_0))^2} \frac{4n^3\alpha^2|q|}{1-(4n^3\alpha^2+2)|q|}.$$

□

Remark 7.12. Under the assumptions of the preceding Lemma, we have a nice formula for $\partial_j^{(q)}$ in terms of the ordinary noncommutative partials ∂_j, namely $\partial_j^{(q)}(\cdot) = \partial_j(\cdot).\Xi_q$, or equivalently $\partial_j(\cdot) = \partial_j^{(q)}(\cdot).\Xi_q^{-1}$ if $\nu_{R+1}(q,n) < 1$ (where Ξ_q^{-1} means $\Xi_q^{-1}(X)$). This follows immediately upon checking that the expressions agree on the X_i.

Proposition 7.13. *For q small enough, all of the above can be combined to give power series $\xi_j(T)$ with radii of convergence $R+1$, such that $\xi_i(X)$ is the conjugate variable to X_i.*

Proof. Take q so small that both $\mu_{R+1}(q,n), \nu_{R+1}(q,n) < 1$. From the previous Lemma, we obtain power series $\Xi_q(T)$ and $\Xi_q^{-1}(T)$ with radii of convergence $R+1$, such that $\Xi_q(T).\Xi_q^{-1}(T) = \Xi_q^{-1}(T).\Xi_q(T) = 1 \otimes 1$, and $\Xi_q(X) = \Xi_q$. By the previous remark,

$$\xi_j = \partial_j^*(1 \otimes 1) = \left[\partial_j^{(q)}(\cdot).\Xi_q^{-1}\right]^* (1 \otimes 1) = \partial_j^{(q)*}(\Xi_q^{-1}).$$

We now define

$$\xi_j(T) = \Xi_q^{-1}(T)\#T_j - m\left\{\left[(1\otimes\omega)\circ\partial_j^{(q)}\right]\otimes 1\ \left(\Xi_q^{-1}(T)\right) + 1\otimes\left[(\omega\otimes 1)\circ\partial_j^{(q)}\right]\left(\Xi_q^{-1}(T)\right)\right\},$$

where $m(a\otimes b) = ab$. As under these conditions we can write $\partial_j^{(q)}(\cdot) = \partial_j(\cdot).\Xi_q$, this is again a noncommutative power series. The radius of convergence of this series can still be taken as $R+1$, as convergence loss due to derivatives is infinitesimal (Proposition 6.6), so that we could take $|q|$ so small that both $\mu_{R+1+\varepsilon}(q,n), \nu_{R+1+\varepsilon}(q,n) < 1$. If we evaluate this in $T = X$, one finds $\xi_j(X) = \xi_j$ by the formula in Lemma 7.9 above. We are allowed to use this formula and do this evaluation, because $R+1 > \|X\|$ and $\left\|\Xi_q^{-1}(X)\right\| \leq \left\|\Xi_q^{-1}(T)\right\|_{R+1} < \infty$, so that $\Xi_q^{-1}(X) \in \Gamma_q \otimes \Gamma_q^{op} \subset D(\partial_j^{(q)*})$. □

7.3 Constructing the Schwinger-Dyson Potential

Now that we have proven that the conjugate variables ξ_j can be viewed as noncommutative power series in X, it remains to show that we can antidifferentiate them. By this we mean a noncommutative power series $V(T)$, with radius of convergence at least $R+1$, such that $\xi_j(T) = \mathscr{D}_j V(T)$ in $\mathscr{A}^{(R+1)}$. To feed this into the free monotone transport machinery later, we also require V to be self-adjoint. It turns out that all of these demands can be met.

Proposition 7.14. *Assume q is small enough and let*

$$V(T) = \mathscr{S}\left\{\frac{1}{2}\sum_{j=1}^{n}\left[\xi_j(T)T_j + T_j\xi_j(T)\right]\right\},$$

then this is a self-adjoint element of $\mathscr{A}^{(R+1)}$ and one has that $\mathscr{D}_{T_i}V(T) = \xi_i(T)$.

Proof. It is immediate that $V \in \mathscr{A}^{(R+1)}$, as the $\xi_j \in \mathscr{A}^{(R+1)}$ and $\mathscr{S}$ is contractive. Self-adjointness is straightforward to check using that the ξ_j are self-adjoint. The desired conclusion is then equivalent to the relation

$$\mathscr{D}_{T_i}(\mathscr{N}V(T)) = (1+\mathscr{N})\mathscr{D}_{T_i}V(T) = (1+\mathscr{N})\xi_i(T) = \xi_i(T) + \sum_{j=1}^{n}\left[\partial_{T_j}\xi_i(T)\right]\#T_j.$$

Since the X_i are algebraically free, one can show using a slight strengthening (see Lemma 37 in [26]) of an earlier result (part (iv) of Theorem 4.10) that it is sufficient to check the above relation evaluated at $T = X$. Thus we want to show that

$$\mathscr{D}_i\left[\mathscr{N}V(X)\right] = \xi_i(X) + \sum_{j=1}^{n}\left[\partial_j\xi_i(X)\right]\#X_j.$$

To begin, we derive a certain identity for $\mathscr{N}V(X)$ by rewriting

$$\mathscr{N}V(X) = \frac{1}{2}\sum_{j=1}^{n}\left[\xi_j(X)X_j + X_j\xi_j(T)\right] = \sum_{j=1}^{n}\xi_j(X)X_j - \frac{1}{2}\sum_{j=1}^{n}\left[\xi_j(X), X_j\right].$$

As $\xi_j(X) = \xi_j$ really are the conjugate variables, we can take P to be an arbitrary element of $\mathscr{A}$ and

consider

$$\begin{aligned}
\sum_{j=1}^{n} \omega\left(\left[\xi_j, X_j\right] P\right) &= \sum_{j=1}^{n} \omega\left(\xi_j \left[X_j, P\right]\right), \\
&= \sum_{j=1}^{n} (\omega \otimes \omega)\left(\partial_j \left[X_j, P\right]\right), \\
&= \sum_{j=1}^{n} (\omega \otimes \omega)\left(1 \otimes P + (X_j \otimes 1)\partial_j P - \partial_j P(1 \otimes X_j) - P \otimes 1\right), \\
&= n\left[\omega(1)\omega(P) - \omega(P)\omega(1)\right] + (\omega \otimes \omega)\left\{\sum_{j=1}^{n}\left[\partial_j P(X_j \otimes 1) - (1 \otimes X_j)\partial_j P\right]\right\}, \\
&= (\omega \otimes \omega)\left(P \otimes 1 - 1 \otimes P\right), \\
&= 0.
\end{aligned}$$

Thus we have that

$$\mathscr{N}V(X) = \sum_{j=1}^{n} \xi_j(X) X_j.$$

Applying the Leibniz rule for the cyclic derivatives, we find

$$\mathscr{D}_i(\mathscr{N}V(X)) = \sum_{j=1}^{n} \left[t(\partial_i \xi_j(X)) \# X_j + t(\partial_i X_j) \# \xi_j(X)\right] = \xi_j(X) + \sum_{j=1}^{n} t(\partial_i \xi_j(X)) \# X_j.$$

We claim that $t(\partial_i \xi_j) = \partial_j \xi_i$, and using this indeed yields

$$\mathscr{D}_i(\mathscr{N}V(X)) = \xi_i(X) + \sum_{j=1}^{n} \left[\partial_j \xi_i(X)\right] \# X_j,$$

as advertised.

It remains to prove our claim. For this, recall that $\mathscr{A} \subset D(\partial_j)$ and $\mathscr{A} \otimes \mathscr{A} \subset D(\partial_j^*)$ for all $1 \leq j \leq n$, and that these subsets are necessarily dense in the respective domains (as they are dense in the ambient L^2 spaces). Let $a, b \in \mathscr{A}$ be arbitrary, then we will show that

$$\langle \xi_i, \partial_j^*(a \otimes b)\rangle = \langle \xi_j, \partial_i^*(b \otimes a)\rangle.$$

Recall (Theorem 4.7) that we had the formula

$$\partial_j^*(a \otimes b) = a\partial_j^*(1 \otimes 1)b - a(\omega \otimes 1)(\partial_j b) - (1 \otimes \omega)(\partial_j a)b.$$

Plugging this in and using the definition of the inner products (also that conjugate variables are self-adjoint in the sense of J), we obtain

$$\begin{aligned}
\langle \xi_i, \partial_j^*(a \otimes b)\rangle - \langle \xi_j, \partial_i^*(b \otimes a)\rangle &= \omega(\xi_i a \xi_j b) - \omega(\xi_j b \xi_i a) - \omega\left[\xi_i(1 \otimes \omega)(\partial_j a)b\right] - \omega\left[\xi_i a(\omega \otimes 1)(\partial_j b)\right] \\
&\quad + \omega\left[\xi_j(1 \otimes \omega)(\partial_i b)a\right] + \omega\left[\xi_j b(\omega \otimes 1)(\partial_i a)\right], \\
&= (\omega \otimes \omega)\Big\{ - \left[(1 \otimes \omega)\partial_j a\right] \otimes 1(\partial_i b) + \left[(1 \otimes \omega)\partial_i b\right] \otimes 1(\partial_j a) \\
&\quad - (\partial_i a)1 \otimes \left[(\omega \otimes 1)\partial_j b\right] + (\partial_j b)1 \otimes \left[(\omega \otimes 1)\partial_i a\right] \\
&\quad - (a \otimes 1)\left[(\omega \otimes \partial_i)\partial_j b\right] + \left[(\partial_j \otimes \omega)\partial_i b\right](1 \otimes a) \\
&\quad - \left[(\partial_i \otimes \omega)\partial_j a\right](1 \otimes b) + (b \otimes 1)\left[(\omega \otimes \partial_j)\partial_i a\right]\Big\}.
\end{aligned}$$

By checking on elementary tensors, one readily shows that the first two pairs cancel when their trace is take. For the remaining pairs, we explicitly show how the first pair cancels, and the other one is similar. Notice that we can write

$$\begin{aligned}(\omega\otimes\omega)\big\{[(\partial_j\otimes\omega)\partial_i b]\,(1\otimes a)\big\} &= (\omega\otimes\omega)\big\{[(1\otimes 1\otimes\omega)\circ(\partial_j\otimes 1)\partial_i b]\,(1\otimes a)\big\},\\ &= (\omega\otimes\omega)\big[(1\otimes a\otimes\omega)\circ(\partial_j\otimes 1)\circ\partial_i(b)\big].\end{aligned}$$

Now we can use coassociativity of the partials to obtain

$$\begin{aligned}(\omega\otimes\omega)\big\{[(\partial_j\otimes\omega)\partial_i b]\,(1\otimes a)\big\} &= (\omega\otimes\omega)\big[(1\otimes a\otimes\omega)\circ(1\otimes\partial_i)\circ\partial_j(b)\big],\\ &= \omega\big[(a\otimes\omega)\circ(\omega\otimes\partial_i)\circ\partial)j(b)\big],\\ &= (\omega\otimes\omega)\big\{(a\otimes 1)\,[(\omega\otimes\partial_i)\partial_j b]\big\},\end{aligned}$$

and this concludes the proof. □

7.4 The q-Gaußians are Isomorphic for Small q

After about a hundred pages of hard work, we come to the pièce de résistance of this thesis: the partial resolution of the isomorphism problem for q-Gaußian von Neumann algebras. This result was first achieved by Alice Guionnet and Dimitry Shlyakhtenko in [43], after developing the fundamental machinery of free monotone transport in the same paper. More than twenty years passed between the first construction of q-Gaußians by Marek Bożejko and Roland Speicher in [13] and this result, despite these algebras being very well studied (see for instance the references in the text below Theorem 1.3 in [43]). Without further ado:

Theorem 7.15. *There exists a function $q_*:\mathbb{Z}_{\geq 1}\to(0,1]$, with $q_*(1)=1$, such that for any $n\in\mathbb{Z}_{\geq 1}$, $\Gamma_0(n)\cong\Gamma_q(n)$ whenever $|q|<q_*(n)$.*

Proof. For $n=1$, recall that the law of a single q-semicircular element is diffuse according to Proposition 3.23, and so the von Neumann algebra it generates is simply isomorphic to $L^\infty([0,1],\lambda)$ (Theorem 3.22). As this is true for all $|q|<1$, we have that $q_*(1)=1$.

Now suppose that $n\geq 2$. We proved in Theorem 3.13 that then $\Gamma_0(n)\cong\mathcal{L}(\mathbb{F}_n)$. Hence, if we can prove that the conditions of Theorem 6.29 are met, we are done (note that part (iv) of that Theorem gives invertibility). That is to say, we want to prove that there exists a self-adjoint noncommutative power series W in n variables, such that the unique solution to the Schwinger-Dyson equation with potential $V=\frac{1}{2}\sum_j X_j^2+W$ is the joint law of the n q-semicircular elements. Moreover, we need that W has radius of convergence at least $\rho+1$ for some $\rho>4$, and that the $\rho+1$-norm of W can be estimated by a function w of q, that vanishes as q goes to zero (which then implicitly defines $q_*(n)$). The way to obtain W, is to subtract $\frac{1}{2}\sum_j X_j^2$ from the V constructed in Proposition 7.14. We claim that the function w can be taken to be proportional to the function $\nu_{\rho+1}(q,n)$ constructed in Lemma 7.11. To see this, notice that

$$V(T)-\frac{1}{2}\sum_j T_j^2=\frac{1}{2}\mathscr{S}\sum_j\left[(\xi_j(T)-T_j)\,T_j+T_j\,(\xi_j(T)-T_j)\right].$$

Thus it suffices to estimate

$$\begin{aligned}\left\|\mathscr{S}\left(\xi_j(T)-T_j\right)\right\|_{\rho+1}\leq{}&\left\|\mathscr{S}\left(\Xi_q^{-1}(T)-1\otimes 1\right)\#T_j\right\|_{\rho+1}\\ &+\left\|\mathscr{S}m\left\{\left[(1\otimes\omega)\circ\partial_j^{(q)}\right]\otimes 1\ \left(\Xi_q^{-1}(T)-1\otimes 1\right)\right\}\right\|_{\rho+1}\\ &+\left\|\mathscr{S}m\left\{1\otimes\left[(\omega\otimes 1)\circ\partial_j^{(q)}\right]\left(\Xi_q^{-1}(T)-1\otimes 1\right)\right\}\right\|_{\rho+1},\end{aligned}$$

where we are allowed to introduce the $1 \otimes 1$ in the last two terms because they are in the kernel of $\partial_j^{(q)}$. Notice that

$$\left\| \mathscr{S} \left(\Xi_q^{-1}(T) - 1 \otimes 1 \right) \# T_j \right\|_{\rho+1} \leq (\rho+1) \left\| \Xi_q^{-1}(T) \right\|_{\rho+1} \| 1 \otimes 1 - \Xi_q(T) \|_{\rho+1} \leq C \nu_{\rho+1}(q,n),$$

because $\mathscr{S}$ is contractive and $\|T_j\|_{\rho+1} = \rho + 1$. One can apply similar tricks to the other two terms, as due to the presence of $\mathscr{S}$, the $\partial_j^{(q)}$ term does not lead to a smaller radius of convergence. This concludes the proof. □

Part II: Physics

8 The Quantum Field Theory of Quons

8.1 Second Quantised Quons

The q-mutation relations (3.1.1) discussed in part I,

$$a_i a_j^\dagger - q a_j^\dagger a_i = \delta_{ij}, \tag{8.1.1}$$

with q a real number between 1 and -1, were also studied by physicists. This happened mostly in the early nineties, alongside the mathematical developments. In this chapter, we will take a look at the physics of these relations, which we shall call the *quon algebra* in this context, following the papers [39, 40, 41] by Greenberg.

The most basic observation is that setting $q = 1$ gives the Bosonic commutation relations, and setting $q = -1$ gives the Fermionic anti-commutation relations. In that sense, the quon algebra (8.1.1) represents a linear interpolation between those two algebras. In these familiar cases, one imposes additional relations on the creation and annihilation operators, namely that creation/annihilation operators with different labels (anti-)commute. Analogously, one might naively expect the quon annihilation operators to satisfy $a_i a_j = q a_j a_i$ when $i \neq j$, but this cannot be true unless $q^2 = 1$, as can be seen by plugging the relation into itself. That is to say these extra relations are only consistent in the Bosonic and Fermionic settings. From Chapter 3 we know that the quon algebra admits a Fock space representation with a unique vacuum $|0\rangle$ (there called Ω). Despite the lack of additional relations, one can still normal order any expression with respect to this vacuum using only (8.1.1) to move all creation operators to the left of the annihilation operators. Correspondingly, there is a Wick theorem for these particles, namely (3.1.2), which we derived in the proof of Lemma 3.10. This is entirely analogous to, say, the Wick theorem for Fermions, except that one picks up factors of q instead of -1. The Wick theorem also allows us to recover the extra Bosonic and Fermionic relations in the limits $q \to \pm 1$ from the quon algebra. For this, consider the vacuum expectation value

$$\langle 0 | \, a_i a_j a_k^\dagger a_\ell^\dagger \, | 0 \rangle = \delta_{i\ell}\delta_{jk} + q\delta_{ik}\delta_{j\ell},$$

where we used (3.1.2). Now, one can split this into symmetric and anti-symmetric parts as

$$\langle 0 | \, a_i a_j a_k^\dagger a_\ell^\dagger \, | 0 \rangle = \frac{1+q}{2}\left(\delta_{i\ell}\delta_{jk} + \delta_{ik}\delta_{j\ell}\right) + \frac{1-q}{2}\left(\delta_{i\ell}\delta_{jk} - \delta_{ik}\delta_{j\ell}\right). \tag{8.1.2}$$

From here, we see that if $q = \pm 1$, only the part with the correct symmetry (in the pairs i, j and k, ℓ) survives.

In the Bosonic and Fermionic cases, we know that not every state is non-zero in the Fock representation. Bosons must always occur in symmetric combinations, and Fermions must always occur in antisymmetric combinations. There are no such restrictions on quons, every state is allowed. Indeed, this is trivial for $q = 0$, where the quon algebra becomes

$$a_i a_j^\dagger = \delta_{ij},$$

and thus the states $a_{i(1)}^\dagger \cdots a_{i(n)}^\dagger \, |0\rangle$[7] are orthonormal and span the Fock space. We know from Proposition 3.5 and Lemma 7.4 that the matrices $\mathcal{G}_m$ of inner products between these states (m being the number of particles) remain strictly positive so long as q stays away from ± 1, hence all these states have positive norms and are allowed. Therefore, quons are sometimes said to have *infinite statistics* (although this sometimes reserved to refer only to the case $q = 0$). Alternatively, one can consider

[7] The notation $i(j)$ is to be interpreted as i_j, not as a permutation.

the $m! \times m!$ matrix $\hat{\mathcal{G}}_m$ of m-particle states with no repeated quantum numbers. The determinant of this matrix was computed by Zagier in [89] (see Theorem 2) to be

$$\det \hat{\mathcal{G}}_m = \prod_{k=1}^{m-1} \left(1 - q^{k^2+k}\right)^{\frac{m!(m-k)}{k^2+k}}.$$

This is unity for $q = 0$, and strictly positive as long as $|q| < 1$, so that at least all states without repeating quantum numbers are allowed. As there is no way to permute creation operators (and there are no relations saying that some power of them vanishes), in fact all states must be allowed, and we arrive at the same conclusion. The reason why we restrict q to be a real number between -1 and 1 now also becomes apparent: this determinant can become negative otherwise. Concretely, one can use (8.1.2) to show that

$$\left\| [a_k^\dagger, a_\ell^\dagger] \, |0\rangle \right\|^2 = 2(1-q), \;\; \left\| \{a_k^\dagger, a_\ell^\dagger\} \, |0\rangle \right\|^2 = 2(1+q),$$

one of which will become negative if $|q| > 1$. Taking q to be a complex number (for instance a root of unity), as we will do in later chapters, comes with the additional problem that the quon algebra loses conjugation invariance (as q will be sent to $\overline{q}$), so we do not consider this for now.

Since the case $q = 0$ is the simplest, it will be our primary focus. For convenience, we will call the particles *zerons*. Before we turn to quantum fields, let us flesh out the Fock space formalism a little more. In particular, we would like to have a formula for the number operators n_j, as we could then for instance define a many-body energy operator via the formula $\sum_j \epsilon_j n_j$, with ϵ_j the single-particle energies associated with the orbitals/quantum numbers j. This operator will not be a simple bilinear in creation and annihilation operators, as for instance the state $a_k^\dagger a_j^\dagger \, |0\rangle$ certainly contains a zeron with quantum number j, but this cannot be 'measured' by something like $a_j^\dagger a_j$ because there is an $a_k^\dagger$ in the way. The solution in this case is to add an a_k in front of $a_j^\dagger a_j$ to 'remove' the obstructing $a_k^\dagger$, and to add a $a_k^\dagger$ to the end to put this zeron back. In general then, we need to consider every possible combination of zerons that could occur to the left of any j-zeron that may be present in the state. This is captured by the formula

$$n_j = a_j^\dagger a_j + \sum_{k(1)} a_{k(1)}^\dagger a_j^\dagger a_j a_{k(1)} + \ldots + \sum_{k(1),\ldots,k(m)} a_{k(1)}^\dagger \cdots a_{k(m)}^\dagger a_j^\dagger a_j a_{k(m)} \cdots a_{k(1)} + \ldots. \tag{8.1.3}$$

To further convince ourselves that this is the correct expression, we would like it to be true that this operator is the U(1) transformation generator, i.e.

$$[n_i, a_j] = -\delta_{ij} a_j.$$

This computation may seem impossible at first, because we know nothing about the commutators $[a_j, a_k]$, but it turns out that this is not required. Consider the lowest order term in the commutator,

$$[a_i^\dagger a_i, a_j] = a_i^\dagger a_i a_j - \delta_{ij} a_j,$$

as by the zeron algebra $a_j a_i^\dagger = \delta_{ij}$. Consider now a general term

$$\begin{aligned}\left[a_{k(1)}^\dagger \cdots a_{k(m)}^\dagger a_i^\dagger a_i a_{k(m)} \cdots a_{k(1)}, a_j\right] &= a_{k(1)}^\dagger \cdots a_{k(m)}^\dagger a_i^\dagger a_i a_{k(m)} \cdots a_{k(1)} a_j \\ &\quad - \delta_{k(1)j} \; a_{k(2)}^\dagger \cdots a_{k(m)}^\dagger a_i^\dagger a_i a_{k(m)} \cdots a_{k(2)} a_j,\end{aligned}$$

whence in particular

$$\left[a_{k(1)}^\dagger a_i^\dagger a_i a_{k(1)}, a_j\right] = a_{k(1)}^\dagger a_i^\dagger a_i a_{k(1)} a_j - \delta_{k(1)j} \; a_i^\dagger a_i a_j.$$

We see that the term with a minus sign cancels the trilinear term in the lowest order commutator, and it is not too difficult to see that this pattern is general, resulting in a telescoping sum with only the linear term in the lowest order contribution surviving, but this is precisely $-\delta_{ij}a_j$, as desired. The arguments we just used to construct n_j can be extended to give many more operators, such as n_{ij}, which is the transition operator from orbital j to orbital i. If we write $n_j =; a_j^\dagger a_j;$, using the semi-colons to indicate the sums in (8.1.3), it holds that $n_{ij} =; a_i^\dagger a_j;$.

An amusing consequence of the fact that every state is allowed, is that the partition function for a gas of free zerons takes on a well known form. Suppose that the gas consists of N zerons, and we wish to compute the canonical N-particle partition function Z_N. We interpret the (continuous) quantum number p as the momentum of the zeron. We need to count how many quantum many-body states there are such that $\sum_p n_p = N$. Since every state is allowed, this number is $N!/(\prod_p n_p!)$, given $\{n_p\}$, which is the same as for classical indistinguishable particles, except there is no $1/N!$ factor to cancel the over-counting that arises by labelling in that case. As there is no potential or interaction, these momentum integrals can simply be performed as usual to give Λ^{-DN}, with D the dimension and Λ the thermal wavelength, and the space integrals give a factor V^N. In total then, we find the partition function

$$Z_N = \left(\frac{V}{\Lambda^D}\right)^N,$$

which is the same as the partition function for an ideal gas of N classical, distinguishable particles.

8.2 Non-Relativistic Quons

Now, let us move on to field theory, starting with non-relativistic quons. We will consider one-dimensional field theories. Assume that we have a quon field ψ that satisfies the continuous version

$$\psi(x)\psi^\dagger(y) - q\ \psi^\dagger(y)\psi(x) = \delta(x-y),$$

of (8.1.1). To write down interaction terms in a non-relativistic Hamiltonian, we need continuous analogues of the transition operators $n_{ij} =; a_i^\dagger a_j;$ for one and two particles. These operators, call them $\rho_1(x;x')$ and $\rho_2(x,y;y',x')$, should satisfy the commutation relations

$$\begin{aligned}\big[\rho_1(x;x'),\psi^\dagger(y)\big] &= \delta(x'-y)\psi^\dagger(x),\\ \big[\rho_2(x,y;y',x'),\psi^\dagger(z)\big] &= \delta(x'-z)\psi^\dagger(x)\rho_1(y;y') + \delta(y'-z)\psi^\dagger(y)\rho_1(x;x').\end{aligned}$$

The second relation one can motivate for instance by considering Bosonic operators b_i, for which the two-body operator is $b_i^\dagger b_j^\dagger b_k b_\ell$, so that

$$[b_i^\dagger b_j^\dagger b_k b_\ell, b_m^\dagger] = b_i^\dagger b_j^\dagger [b_k b_\ell, b_m^\dagger] = b_i^\dagger b_j^\dagger\,(b_k\delta_{\ell m} + b_\ell\delta_{km}) = b_i^\dagger(b_j^\dagger b_k)\delta_{\ell m} + b_j^\dagger(b_i^\dagger b_\ell)\delta_{km}.$$

The operator $\rho_1(x;x')$ for zerons is straightforward to construct using the same ideas as before,

$$\begin{aligned}\rho_1(x;x') &= \big(;\psi^\dagger(x)\psi(x');\big)\,,\\ &= \psi^\dagger(x)\psi(x') + \sum_{n\geq 1}\int \mathrm{d}y_1\cdots\mathrm{d}y_n\ \psi^\dagger(y_n)\cdots\psi^\dagger(y_1)\psi^\dagger(x)\psi(x')\psi(y_1)\cdots\psi(y_n).\end{aligned}\tag{8.2.1}$$

Still for zerons, it requires a little more though to realise that the correct 'core' operator for $\rho_2(x,y;y',x')$ is

$$\rho_2(x,y;y',x') = \big(;\psi^\dagger(x)\rho_1(y;y')\psi(x');+;\psi^\dagger(y)\rho_1(x;x')\psi(y');\big)\,.$$

These are precisely the tools we need to write down a Hamiltonian, analogous to for instance equation (6.29) in [81] (but without the 'spin' indices). Thus, any non-relativistic Hamiltonian for quons, containing at most two-body interactions, is of the form

$$H = \frac{1}{2m}\int \mathrm{d}x\ (\nabla_x \cdot \nabla_x)\ \rho_1(x;x) + \frac{1}{2}\int \mathrm{d}x\ \mathrm{d}x'\ V(x-x')\ \rho_2(x,x';x',x).$$

It is now an interesting proposition to derive the equation of motion for the many-body wave function, which can be written as

$$i\hbar\ \partial_t \Psi(z_1,\ldots,z_n) = \left(i\hbar\ \partial_t\ \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)\right)|0\rangle = [H, \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)]\,|0\rangle\,.$$

This commutator can be computed using the commutator relations for ρ_1 and ρ_2 as follows,

$$\begin{aligned}
[H, \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)] &= \frac{1}{2m}\int \mathrm{d}x\ (\nabla_x \cdot \nabla_x)\ [\rho_1(x;x), \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)] \\
&\quad + \frac{1}{2}\int \mathrm{d}x\ \mathrm{d}x'\ V(x-x')\ [\rho_2(x,x';x',x), \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)], \\
&= \left\{-\frac{1}{2m}\sum_{i=1}^{n}(\nabla_{z_i}\cdot\nabla_{z_i}) + \sum_{i<j} V(z_i - z_j)\right\}\psi^\dagger(z_1)\cdots\psi^\dagger(z_n) \\
&\quad + \sum_{i=1}^{n}\int \mathrm{d}x\ V(x-z_i)\ \psi^\dagger(z_1)\cdots\psi^\dagger(z_n)\ \rho_1(x;x).
\end{aligned}$$

Letting this act on the vacuum $|0\rangle$, the second term vanishes, and we find

$$i\hbar\ \partial_t \Psi(z_1,\ldots,z_n) = \left\{-\frac{1}{2m}\sum_{i=1}^{n}(\nabla_{z_i}\cdot\nabla_{z_i}) + \sum_{i<j} V(z_i - z_j)\right\}\ \Psi(z_1,\ldots,z_n),$$

which is the 'ordinary' many-body Schrödinger equation. It is a rather pleasant surprise that the Schrödinger equation not only describes the cases $q = \pm 1$, but in fact the entire line $q \in [-1,1]$. However, there is a catch to all this, as we only have explicit expressions for ρ_1 and ρ_2 in the case that the particles are zerons. The first few terms are known, see equations (19) and (20) in [41], but the general formulas remain elusive.

8.3 Relativistic Quons

Having established some fundamentals of non-relativistic quon field theories, we investigate what happens if we follow the quantisation procedure of the Klein-Gordon field (as can be found in Chapter 2 of [67]). Since we are in one spatial dimension (plus one time), there is no link between spin and statistics, thus we can emulate the aforementioned chapter to hopefully produce a 'scalar' quon field. As the exact expressions for many necessary operators for general quons are unknown, let us stick to zerons for simplicity. We could consider a 'real scalar' zeron field defined by the Fourier transform expression

$$\varphi(x) = \frac{1}{(2\pi)^{3/2}}\int \frac{\mathrm{d}k}{\sqrt{2\omega(k)}}\left[a_k e^{-i(\omega(k)t - xk)} + a_{-k}^\dagger e^{i(\omega(k)t-xk)}\right],$$

compare to (2.25) in [67], but it is more fruitful to consider instead a 'complex scalar' zeron field, as this has a charge and thus a conserved current. We know from ordinary QFT that such a field should consist of both particles and anti-particles, with opposing charges. Write $\alpha_k^\dagger$ for the operator that

creates a zeron, and $\beta_k^\dagger$ for the operator that creates an anti-zeron, then (compare with Exercise 2.2 in [67]) the field can be written as

$$\phi(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{\mathrm{d}k}{\sqrt{2\omega(k)}} \left[\alpha_k e^{-i(\omega(k)t - xk)} + \beta_{-k}^\dagger e^{i(\omega(k)t - xk)} \right].$$

Both α_k and β_k satisfy the zeron algebra, and the additional relation $\alpha_k \beta_\ell^\dagger = 0 = \beta_k \alpha_\ell^\dagger$ to link them. The semi-colon operation (acting on bilinears) from before now needs to be generalised the contain a sum over all combinations of zerons and anti-zerons. Thus, for instance the charge operator (taking the zerons to have positive charge) becomes

$$\begin{aligned} Q =& ; \sum_k \left(\alpha_k^\dagger \alpha_k - \beta_k^\dagger \beta_k \right) ; , \\ =& \sum_k \left(\alpha_k^\dagger \alpha_k - \beta_k^\dagger \beta_k \right) + \sum_{n \geq 1} \sum_{\ell_1, \ldots, \ell_n} \sum_{\gamma_{\ell_i} = \alpha_{\ell_i}, \beta_{\ell_i}} \sum_k \gamma_{\ell_1}^\dagger \cdots \gamma_{\ell_n}^\dagger \left(\alpha_k^\dagger \alpha_k - \beta_k^\dagger \beta_k \right) \gamma_{\ell_n} \cdots \gamma_{\ell_1}. \end{aligned}$$

This is the point where we run into trouble, as

$$\left[; \alpha_k^\dagger \beta_\ell^\dagger ; , \beta_m^\dagger \right] = \alpha_k^\dagger \beta_\ell^\dagger \beta_m^\dagger,$$

which is trilinear. Let $j^\mu(x)$ denote the particle current for the field $\phi(x)$ (see (2.16) in [67], with the continuous version of the semi-colon operation applied to it), then one can show (see [39]) that it does not commute locally with the field, and does not commute with itself at space-like separation, due to these trilinear terms. A mathematical proof of this failure of locality can be found in [38], see the text below equation (3.4). Another philosophical problem is that the continuous version of the semi-colon operation contains integrals over all space, so that the current is not even a local function of the fields. In (8.2.1) this is not so much of an issue, because this is a non-relativistic field theory, and the commutator with the field is still local (due to the delta function).

The non-locality of the zeron quantum field is not the only problem the theory suffers from. One expects these locality problems to also hold for finite values of q (as these are still far away from Bosons and Fermions), but there might be hope that theories with q very close to either ± 1 are somehow viable. They could describe theories with small deviations from pure Bosonic or Fermionic statistics. Perhaps electrons are only Fermions to good approximation, so that in principle states with higher occupation may (briefly) exist in nature. Unfortunately, there is an experimental upper bound on the distance of q from -1 based on X-ray detection experiments on Copper in [70] (see also Section 5 of [41]), which gives

$$|-1 - q| \leq 3.4 \times 10^{-26}.$$

Thus it seems there is not much use for quons, and for the rest of this thesis, we will instead focus on a different deformation of the canonical (anti-)commutation relations.

9 The Road to Parafermions

9.1 The Ising Model and the Jordan-Wigner Transformation

In this chapter, we introduce a deformation of the canonical (anti-)commutation relations that is different from the quon algebra of the previous chapter. We will follow [34].

We begin our journey with the quantum version of the famous Ising model with a transverse field. Consider a chain with L sites, open boundary conditions, and at each site a spin 1/2 degree of freedom. This means that at a single site we have the two-dimensional Hilbert space $\mathbb{C}^2$, and the Hilbert space of the entire chain is $\otimes_{j=1}^{L}\mathbb{C}^2$, which is 2^L-dimensional. We will write

$$\sigma_j^\alpha := I \otimes \cdots \otimes \sigma^\alpha \otimes \cdots \otimes I,$$

for the spin operators at site j. Here, σ^α are the Pauli matrices and I is the identity. The Pauli matrices form an $\mathfrak{su}_2$ representation as they satisfy the commutation relations

$$\left[\sigma^\alpha, \sigma^\beta\right] = 2i\epsilon^{\alpha\beta\gamma}\sigma^\gamma,$$

and by construction the spin operators satisfy

$$\left[\sigma_j^\alpha, \sigma_k^\beta\right] = 2i\delta_{jk}\epsilon^{\alpha\beta\gamma}\sigma_j^\gamma. \tag{9.1.1}$$

Additionally, their on-site anti-commutator is easily seen to be

$$\left\{\sigma_j^\alpha, \sigma_j^\beta\right\} = 2\delta^{\alpha\beta}I, \tag{9.1.2}$$

which is the famous *Clifford algebra* (with respect to the upper index). Let f and J be two positive numbers, meant to characterise the strength of the transverse field and the nearest-neighbour interactions. Then the Hamiltonian is

$$H_{\mathrm{qI}} = -f\sum_{j=1}^{L}\sigma_j^x - J\sum_{j=1}^{L-1}\sigma_j^z\sigma_{j+1}^z. \tag{9.1.3}$$

We take the z axis to be our axis of quantisation, i.e. we take σ^z to be diagonal, so that the first term is a spin flip operator (favouring alignment with the field in the positive x-direction), and the second term is an interaction energy between adjacent spins (favouring ferromagnetic alignment).

There are some natural symmetries of the Hamiltonian (9.1.3). First, notice that we can send $\sigma_j^x \to -\sigma_j^x$ and $\sigma_j^z \to (-1)^{j+1}\sigma_j^z$ while preserving the algebra (9.1.1). Under this mapping, our Hamiltonian H_{qI} is transformed into its negative $-H_{\mathrm{qI}}$, implying that the spectrum of H_{qI} is symmetrically distributed around 0. Second, it is intuitive that H_{qI} should be invariant under the transformation that flips all spins. The natural operator to implement this is

$$\sigma_f = \prod_{j=1}^{L}\sigma_j^x.$$

Indeed, σ_f trivially commutes with the first term in (9.1.3), and a simple calculation [using (9.1.2)] shows that

$$\sigma_f\sigma_j^z\sigma_{j+1}^z = (\sigma_j^x\sigma_j^z)(\sigma_{j+1}^x\sigma_{j+1}^z)(\sigma_j^x\sigma_{j+1}^x\sigma_f) = (-1)^2\sigma_j^z\sigma_{j+1}^z\sigma_f.$$

As the operator σ_f squares to the identity, it implements a $\mathbb{Z}_2$ symmetry.

A slightly less obvious symmetry comes from the *self-duality* of our Hamiltonian (see Sections IIA and IIB in [21]). Define the following two operators, that live on the dual chain (whose sites are labelled by $j+1/2$)

$$\nu_{j+1/2} = \sigma_j^z \sigma_{j+1}^z, \ \mu_{j+1/2} = \prod_{k<j+1/2} \sigma_k^x.$$

It is straightforward to check that these operators also are also Hermitian and satisfy the algebra (9.1.2). In terms of these operators (9.1.3) becomes

$$H_{\text{qI}} = -f \sum_{j=1}^{L} \mu_{j-1/2}\mu_{j+1/2} - J \sum_{j=1}^{L-1} \nu_{j+1/2},$$

with the convention $\mu_{1/2} = I$. This has exactly the same form as our original expression, except that the roles of f and J have been switched. This symmetry is related to the phases of the model as follows. Suppose that we are in a phase where $\mathbb{Z}_2$ symmetry is spontaneously broken, then σ_j^z acquires a finite ground state expectation value (think $J \gg f$). The dual operator $\mu_{j+1/2}$ flips all spins to the left of $j+1$, creating a domain wall defect. If it is then the case that this operator also has a non-zero ground state expectation, these defects destroy the σ_j^z order and restore the $\mathbb{Z}_2$ symmetry. This indicates that we should think of σ_j^z as on *order operator* and of $\mu_{j+1/2}$ as the dual *disorder* operator. By explicitly constructing the ground states for the cases $f = 0$ and $J = 0$, one see that for small f there is an ordered phase with either all spins up or all spins down (so the $\mathbb{Z}_2$ symmetry is spontaneously broken), and for small J there is a unique ground state (namely the product of the eigenvalue 1 states of σ_j^x). As duality interchanges these two phases, there must be a quantum critical point at the self-dual point $J = f$.

The order and disorder operators have another important use. As each lattice site has a two-dimensional Hilbert space, one might wonder if there exists some mapping from the spins σ_j^z to spinless Fermions c_j. These Fermionic creation and annihilation operators naturally satisfy the canonical anti-commutation relations

$$\left\{c_j, c_k^\dagger\right\} = \delta_{jk}, \ \left\{c_j^{(\dagger)}, c_k^{(\dagger)}\right\} = 0.$$

As is well known, these relations imply that a site with a Fermionic degree of freedom can be either occupied or empty, so that, like a spin 1/2, it has two possible states. For a single site, it is indeed simple to construct such a mapping, for instance by using the *number operator*

$$n_j = c_j^\dagger c_j,$$

which takes on the value 0 when the site j is empty, and the value 1 when it is occupied. Since n_j commutes with all creation and annihilation operators associated to different sites, it is a natural building block for a spin operator. If we identify an empty site with the $+1$ eigenstate of σ^x and conversely an occupied site with the -1 eigenstate, we find the Ansatz

$$\sigma_j^x = (+1)(1-n_j) + (-1)n_j = 1 - 2n_j.$$

Recalling that σ^z flips the eigenstates of σ^x, it makes sense to model it with the combination $c_j + c_j^\dagger$, as this operator swaps the occupied and empty states. The last obstacle is now to relate the commutation relations. This turns out to be possible through the so called *Jordan-Wigner transformation*, given by

$$\sigma_j^x = \left(1 - 2c_j^\dagger c_j\right), \ \sigma_j^z = \left[\prod_{k<j} \left(1 - 2c_k^\dagger c_k\right)\right] \left(c_j + c_j^\dagger\right). \tag{9.1.4}$$

The inverse expressions are

$$c_j^\dagger + c_j = \left(\prod_{k<j} \sigma_k^x\right) \sigma_j^z, \quad c_j^\dagger - c_j = \left(\prod_{k<j} \sigma_k^x\right) \sigma_j^z \sigma_j^x,$$

which in terms of the order and disorder operators become

$$c_j^\dagger + c_j = \mu_{j-1/2}\sigma_j^z, \; c_j^\dagger - c_j = -\mu_{j+1/2}\sigma_j^z.$$

We see that the product of order and disorder operators gives rise to Fermionic (anti-commuting) operators. In fact, this is precisely the purpose of the non-local *strings* $\prod_{k<j} \sigma_k^x$, to convert the commuting spins into anti-commuting Fermions. One way in which we could have guessed the form of the strings is to notice that commuting c_j or $c_j^\dagger$ through a site k yields a minus sign if the site is occupied, and no sign change else, but this is precisely how we matched the empty and occupied states at site k to the ± 1 the eigenstates of σ_k^x. Now stringing all these together and using the Ansatz we already had for σ_k^x yields (9.1.4).

It is a straightforward exercise to convert (9.1.3) into a Fermionic Hamiltonian using the Jordan-Wigner transformation (9.1.4). The result is

$$H_{\text{qI}} = 2f \sum_{j=1}^{L} \left(c_j^\dagger c_j - \frac{1}{2}\right) - J \sum_{j=1}^{L-1} \left(c_j^\dagger - c_j\right)\left(c_{j+1}^\dagger + c_{j+1}\right), \tag{9.1.5}$$

$$= \text{constant} - J \sum_{j=1}^{L-1} \left(c_{j+1}^\dagger c_j + \text{H.c.}\right) - J \sum_{j=1}^{L-1} \left(c_j^\dagger c_{j+1}^\dagger + \text{H.c.}\right) + 2f \sum_{j=1}^{L} c_j^\dagger c_j. \tag{9.1.6}$$

This has the form of a tight-binding model with hopping terms, 'Cooper pairing' terms, and a chemical potential term. Actually, that we ended up with this form is a minor miracle. The *non-local* transformation (9.1.4) somehow results in a *local* Fermionic Hamiltonian. However, we know that this is not due to divine intervention, but rather due to the models nice behaviour with respect to duality transformations. The 'Cooper pairing' terms do not conserve the Fermion number, rather they explicitly break the U(1) symmetry down to the $\mathbb{Z}_2$ symmetry of the spin degrees of freedom. This sheds new light on our spin flip operator σ_f, as

$$\sigma_f = \prod_{j=1}^{L} \sigma_j^x = \prod_{j=1}^{L} \left(1 - 2c_j^\dagger c_j\right) = \prod_{j=1}^{L} (-1)^{c_j^\dagger c_j} = (-1)^F,$$

with $F = \sum_j c_j^\dagger c_j$ the Fermion number operator. Thus we see that the spin flip operator becomes the Fermion parity operator under the Jordan-Wigner transformation. As our Fermionic form of H_{qI} is entirely quadratic, it commutes with the Fermion parity $(-1)^F$, and so the spectrum can be split into an even and odd sector.

9.2 Majorana Fermions and Edge Zero Modes

In principle, one could now unleash the technology of Fourier and Bogoliubov transformations upon the Fermionic Hamiltonian (9.1.6) (taking care of the different sectors etc.) and diagonalise it. We will take a different (and more elegant) approach. Look again at the form (9.1.5) of H_{qI}, and notice that the second term is the product of the Hermitian and anti-Hermitian parts of the Fermion operators. This suggests that it might be fruitful to split the 'complex' Fermions into two 'real' particles via

$$a_j = c_j^\dagger + c_j, \; b_j = i\left(c_j^\dagger - c_j\right). \tag{9.2.1}$$

One can check that with these definitions the operators a_j and b_j are Hermitian, square to the identity, and satisfy the algebra

$$\{a_j, a_k\} = \{b_j, b_k\} = 2\delta_{jk},\ \{a_j, b_k\} = 0. \tag{9.2.2}$$

Thus we have two mutually anticommuting copies of the Clifford algebra, but this time with respect to the site index instead of the space index as in (9.1.2). In terms of the spins these operators are

$$a_j = \left(\prod_{k<j} \sigma_k^x\right) \sigma_j^z,\ b_j = i \left(\prod_{k<j} \sigma_k^x\right) \sigma_j^z \sigma_j^x, \tag{9.2.3}$$

and inversely

$$\sigma_j^x = (-i) a_j b_j,\ \sigma_j^z \sigma_{j+1}^z = (-i) b_j a_{j+1},$$

which is remarkably simple. The operators a_j and b_j are called *Majorana* operators (as $a_j^\dagger = a_j$ and $b_j^\dagger = b_j$, so they are their own 'antiparticles'), and later on we will see that these are the simplest example of parafermions.

In terms of the Majoranas, H_{qI} becomes

$$H_{\mathrm{qI}} = if \sum_{j=1}^{L} a_j b_j + iJ \sum_{j=1}^{L-1} b_j a_{j+1}. \tag{9.2.4}$$

In the case that $f = 0$, one can actually diagonalise the model by introducing the operators

$$\mathcal{L}_j = \left(\frac{1}{2} b_j + J(1-i) a_{j+1}\right),$$

upon which the model becomes quadratic, namely

$$H_{\mathrm{qI}} = \sum_{j=1}^{L-1} \mathcal{L}_j^\dagger \mathcal{L}_j + \text{constant},$$

so that the ground state manifold is precisely the intersection of the kernels of the $\mathcal{L}_j$ (see the next chapter for a more detailed explanation of this technique). At the moment, however, we are more interested in the observation that two of the Majoranas at the edges of the system, namely a_1 and b_L, do not show up in the Hamiltonian at all if $f = 0$. By the Majorana algebra (9.2.2), these single operators commute with any quadratic term, so that $[a_1, H_{\mathrm{qI}}] = [b_L, H_{\mathrm{qI}}] = 0$. Moreover, one can express the Fermionic parity operator $(-1)^F$ in terms of the Majoranas as follows,

$$(-1)^F = \prod_{j=1}^{L} \sigma_j^x = \prod_{j=1}^{L} (-i a_j b_j),$$

from which it is easy to see that it anti-commutes with a_1 and b_L. In other words, these edge Majoranas cycle through the even and odd sectors of the ground state manifold of H_{qI}. These properties are worth collecting in a definition.

Definition 9.1. A *Fermionic zero mode* is an operator Ψ that

(i) commutes with the Hamiltonian: $[H, \Psi] = 0$,

(ii) anti-commutes with the Fermionic partity (so that it cycles through the groundstates): $\{(-1)^F, \Psi\} = 0$,

(iii) and survives the thermodynamic limit, i.e. has finite normalisation $\Psi^\dagger \Psi = 1$ even when L grows to infinity.

If the operator Ψ is also localised at the edge of the system, meaning that matrix elements of Ψ involving bulk sites are exponentially small, we say it is an *edge zero mode*.

Edge zero modes are connected to topological order, by which we mean 'hidden' order that cannot be detected locally in the bulk (so a phase with topological order is one that cannot be described by a local order parameter). From the above discussion, it is clear that a_1 and b_L are edge zero modes of H_{qI} when $f = 0$, and they can be combined into a single non-local fermion $(b_L + ia_1)/2$ which is a Fermionic edge zero mode. A natural questions is now whether or not these edge zero modes survive at finite values of the transverse field f. One can compute

$$[H_{\mathrm{qI}}, a_1] = [if a_1 b_1, a_1] = -2if b_1 = -if\,[b_1 a_2, a_2]\,,$$

so that

$$\left[if a_1 b_1 + iJ b_1 a_2, a_1 + \frac{f}{J} a_2\right] = 0.$$

This suggests that there are numbers α_k such that $\sum_k \alpha_k a_k$ (almost) commutes with the entire Hamiltonian. We construct them by induction, suppose that

$$\left[if \sum_{j=1}^{k} a_j b_j + iJ \sum_{j=1}^{k} b_j a_{j+1}, \sum_{\ell=1}^{k+1} \alpha_\ell a_\ell\right] = 0,$$

then to complete the induction step we want to solve

$$[if a_{k+1} b_{k+1} + iJ b_{k+1} a_{k+2}, \alpha_{k+1} a_{k+1} + \alpha_{k+2} a_{k+2}] = 0$$

for α_{k+2}. This condition becomes

$$\begin{aligned} 0 &= if\alpha_{k+1}\,[a_{k+1} b_{k+1}, a_{k+1}] + iJ\alpha_{k+2}\,[b_{k+1} a_{k+2}, a_{k+2}]\,, \\ &= 2i\,(J\alpha_{k+2} - f\alpha_{k+1})\, b_{k+1}, \end{aligned}$$

which has the solution $\alpha_{k+2} = (f/J)\alpha_{k+1}$. If we set

$$\Psi_\ell = a_1 + \left(\frac{f}{J}\right) a_2 + \cdots + \left(\frac{f}{J}\right)^{L-1} a_L,$$

then

$$[H_{\mathrm{qI}}, \Psi_\ell] = -2if \left(\frac{f}{J}\right)^{L-1} b_L,$$

which vanishes exponentially in the thermodynamic limit. Moreover,

$$\Psi_\ell^\dagger \Psi_\ell = \sum_{j=0}^{L-1} \left(\frac{f}{J}\right)^{2j},$$

which converges in the thermodynamic limit if and only if $f < J$. It is clear that Ψ_ℓ anti-commutes with $(-1)^F$, so that it is indeed an edge zero mode. Analogously, there is an operator Ψ_r involving b_j's.

The convergence condition $f < J$ shows that these edge zero modes are the Fermionic manifestation of the order in the corresponding phase of the Ising model. Thus the *local* order of the Ising model is converted to *topological* order in the Fermionic model (9.2.4), which makes sense considering the non-local character of the Jordan-Wigner transformation. A slight modification of this argument shows that these edge zero modes generically also exist if there is some disorder in the couplings (i.e. having site-dependent J_j and f_j).

9.3 Clock Models and Parafermions

There are several ways in which one could generalise the quantum Ising model (9.1.3). One idea is to swap out the spins for a more general degree of freedom that has a large number of states. It is then the question how one extends the $\mathbb{Z}_2$ symmetry of the Ising model. There is an accidental isomorphism between $\mathbb{Z}_2$ and S_2, the permutation group on 2 objects. So in principle one could consider either S_n or $\mathbb{Z}_n$ symmetry for the generalised model. We will focus on the $\mathbb{Z}_n$ case, where the degrees of freedom are called *clocks*. The S_n generalisations are called Potts models, and are quite famous. As $\mathbb{Z}_n$ is a subgroup of S_n (in a natural way as the cyclic permutations), we expect to be able to recover these models if we impose extra conditions on the coefficients of the $\mathbb{Z}_n$ case.

There is a nice representation of $\mathbb{Z}_n$ as the n-th roots of unity lying on the complex unit circle $\mathbb{T}$. The generator 1 of $\mathbb{Z}_n$ (the residue class 1 mod n to be precise), is mapped to the complex number $\omega = e^{2\pi i/n}$. Thus we label the values of a clock variable by the numbers $\omega^0 = 1, \omega, \ldots, \omega^{n-1}$, and we will write s_j for a (classical) clock degree of freedom at the lattice site j. We now want to construct the most general $\mathbb{Z}_n$ invariant coupling between two clocks located on sites j and k respectively. We will take the system to be homogeneous in the sense that the coupling coefficients are site-independent. As multiplication by ω generates the $\mathbb{Z}_n$ symmetry, and $\omega^n = 1$, we see that in $s_j^{\ell_1} s_k^{\ell_2}$ the powers have to add up to a multiple of n for it to be invariant under the $\mathbb{Z}_n$ symmetry. As $s_j^* = 1/s_j = (s_j)^{n-1}$ (because this is true for all integer powers of ω), the powers $(s_j^* s_k)^\ell$ for $1 \leq \ell \leq n-1$ exhaust all the possibilities. This results in interaction terms

$$-J \sum_{\ell=1}^{n-1} \gamma_\ell (s_j^* s_k)^\ell.$$

The coefficients γ_ℓ cannot be chosen completely arbitrarily, as we need to ensure that the interactions are Hermitian. This gives the condition

$$-J \sum_{\ell=1}^{n-1} \gamma_\ell (s_j^* s_k)^\ell = \left(-J \sum_{\ell=1}^{n-1} \gamma_\ell (s_j^* s_k)^\ell \right)^\dagger = -J \sum_{\ell=1}^{n-1} \gamma_\ell^* (s_j^* s_k)^{n-\ell},$$

which is satisfied if $\gamma_\ell^* = \gamma_{n-\ell}$.

To get an idea of the richness of the physics described by these interactions, we will spend some time on the $n = 3$ case. There are now only two coefficients to choose, and we parametrise them by $\gamma_1 = \gamma_2^* = e^{i\varphi}$ and absorb the modulus degree of freedom into J. The interaction is then

$$-J \left(e^{i\varphi} s_j^* s_k + e^{-i\varphi} s_j s_k^* \right) = -2J \mathfrak{Re} \left[e^{i\varphi} s_j^* s_k \right],$$

and the product $s_j^* s_k$ can only take on the values $1, \omega, \omega^2 = \omega^*$. Thus the interaction energy is proportional to the negative of the real part of $e^{i\varphi} s_j^* s_k$, which gives the three possible energy values $-2J\cos(\varphi)$, $-2J\cos(\varphi \pm 2\pi/3)$. If $\varphi = 0$, all coefficients are real, and we end up with a 'ferromagnet'

as there is a unique lowest energy when s_j and s_k are equal. Oppositely, if $\varphi = \pi/3$, the energy is minimal when s_j and s_k are different, yielding an 'anti-ferromagnet'. Varying φ yields hybrids of these interactions (values of φ outside of $[0, \pi/3]$ give no new configurations by symmetry). These hybrid interactions, however, are all *chiral* (meaning that left and right are not treated in the same way, or more formally, a lack of reflection symmetry), as swapping the values of s_j and s_k would give a different energy. Thus for φ not an integer multiple of $\pi/3$, parity symmetry is explicitly broken by the clock interactions. The 'average' of the ferro and anti-ferromagnetic cases is also interesting. This corresponds to $\varphi = \pi/6$, where the set of energy values is invariant under multiplication by -1, reminiscent of the first symmetry of (9.1.3) that we discussed. Finally, one can notice that we recover S_3 symmetry by choosing both coefficients to be the same (in particular then $\gamma_1 = \gamma_1^*$ and the coefficients must be real), and in general by setting $\gamma_1 = \cdots = \gamma_{\ell-1} \in \mathbb{R}$.

To quantise these clock variables, one could proceed as follows. It is clear that the Hilbert space for a single quantum clock should be $\mathbb{C}^n$, and so that the Hilbert space of a chain consisting of L them is $(\mathbb{C}^n)^{\otimes L}$. As we are interested in generalisations of the quantum Ising chain (9.1.3), we only need to construct analogues to σ^x and σ^z. Thus we introduce two new operators σ and τ. As these should form a representation of $\mathbb{Z}_n$, we immediately have the demand that $\sigma^n = \tau^n = 1$. To keep the representation consistent with the one we have chosen for our classical clocks as n-th roots of unity, we should also impose $\sigma^\dagger = \sigma^{n-1}$ and $\tau^\dagger = \tau^{n-1}$. Finally, σ^x cycles through the eigenvalues of σ^z (or vice versa), so that it is natural (also keeping in mind the anti-commutation relations (9.1.2)) to consider the commutation relation $\sigma\tau = \omega\tau\sigma$. Then as before we construct σ_j and τ_k, which again commute off-site.

There is in fact another way to derive these relations (see also Section II in [22]). So far, we have been discussing lattice models, mostly with spin-like degrees of freedom. Usually in quantum mechanics, one postulates the position and momentum operators, X_j and P_j respectively (where j labels the spatial dimensions), and their *canonical commutation relations* $[X_J, P_k] = i\hbar\delta_{jk}$. Now, starting from the realisation that the canonical commutation relations do not admit a finite dimensional representation, we can arrive at the idea of quantum clocks in a very different way. This follows immediately by taking the trace on both sides. However more is true, there does not even exist a representation for which the position and momentum operators are simultaneously bounded. To see this, notice that the canonical commutation relations imply (by a simple induction argument) that

$$\left[X_j^m, P_j\right] = i\hbar m X_j^{m-1},$$

so that by taking norms we find the estimate

$$m\hbar\left\|X_j^{m-1}\right\| \leq 2\left\|X_j^m\right\|\|P_j\|.$$

As X_j is Hermitian, it is normal and it follows from the spectral mapping theorem that taking powers and the norm commute, so that we can refine our estimate to

$$m\hbar \leq 2\|X_j\|\|P_j\|,$$

for any natural number m. This is clearly irreconcilable with X_j and P_j being simultaneously bounded, as m can be an arbitrarily large natural number. Since dealing with infinite dimensional spaces and unbounded operators is a hairy matter, one would still like to have some sort of finite dimensional model for quantum mechanics. One answer is to look at the *Weyl algebra* $\mathcal{W}$ which is generated by the unitary translations

$$V_j(x) = \exp(ixP_j/\hbar),\ U_k(p) = \exp(ipX_k/\hbar),$$

where $x, p \in \mathbb{R}$. These commute for $j \neq k$, and by the Baker-Campbell-Hausdorff formula they obey the commutation relation

$$V_j(x)U_k(p) = \exp(ixp/\hbar)U_k(p)V_j(x).$$

If we now discretise space according to $x = r\delta$, then $p = 2\pi\hbar s/(n\delta)$, where $r, s \in \mathbb{Z}$ and $n \in \mathbb{Z}_{\geq 1}$ is fixed as before. This is appropriate as we are considering chains and lattices. We can now define a finite set of unitary generators

$$\sigma_j = e^{i\delta P_j/\hbar},\ \tau_k = e^{i2\pi X_k/(n\delta)},$$

which commute for $j \neq k$ and otherwise satisfy our desired relation $\sigma\tau = \omega\tau\sigma$. This 'discretised' Weyl algebra $\mathcal{W}_n$ admits a finite dimensional unitary representation, namely precisely the quantum clock variables we constructed above. For concreteness, one could keep in mind the explicit matrices

$$\sigma = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix},\ \tau = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \omega & 0 & \dots & 0 \\ 0 & 0 & \omega^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \omega^{n-1} \end{pmatrix}. \tag{9.3.1}$$

Having motivated our quantum clock variables, we return to cooking up a natural generalisation to (9.1.3). We know how to handle the interaction term, but there remain some words to be said about the flip term. Since there are more states to flip through, we should consider powers again, each with its own coefficient. This gives terms of the form

$$-f\sum_{\ell=1}^{n-1}\beta_\ell s^\ell,$$

and again we need to impose $\beta_\ell^* = \beta_{n-\ell}$ to ensure Hermicity. In total then, we find the family of models

$$H_n = -f\sum_{j=1}^{L}\sum_{\ell=1}^{n-1}\beta_\ell\tau_j^\ell - J\sum_{j=1}^{L-1}\sum_{\ell=1}^{n-1}\gamma_\ell\left(\sigma_j^\dagger\sigma_{j+1}\right)^\ell. \tag{9.3.2}$$

The $\mathbb{Z}_n$ symmetry is now generated by

$$\omega^P = \prod_{j=1}^{L}\tau_j^\dagger,$$

whose notation is chosen to be similar to $(-1)^F$, and will turn out to have an analogous interpretation.

In the first section of this chapter, we saw a solution of the quantum Ising model using Majorana operators. We would therefore now like to generalise (9.2.3) to obtain *parafermion* operators. One correct formula turns out to be

$$\chi_j = \left(\prod_{\ell<j}\tau_\ell\right)\sigma_j,\ \psi_k = \omega^{(n-1)/2}\left(\prod_{\ell<k}\tau_\ell\right)\sigma_k\tau_k. \tag{9.3.3}$$

We check which properties these operators satisfy. It is immediate that $\chi_j^n = 1$, and

$$\psi_k^n = \omega^{n(n-1)/2}(\sigma_k\tau_k)^n = \omega^{n(n-1)/2}\omega^{(n+\cdots+1)}\tau_k^n\sigma_k^n = \omega^{n\cdot n} = 1.$$

We also have that

$$\chi_j^\dagger = \sigma_j^{n-1}\tau_{j-1}^{n-1}\cdots\tau_1^{n-1} = (\tau_1\cdots\tau_{j-1}\sigma_j)^{n-1} = \chi_j^{n-1},$$

and

$$\psi_k^\dagger = \omega^{-(n-1)/2}\tau_k^{n-1}\chi_k^{n-1} = \omega^{-(n-1)/2}\left(\chi_k\tau_k\right)^{n-1}\omega^{-n(n-1)/2} = \omega^{-(n+1)(n-1)/2}\left(\chi_k\tau_k\right)^{n-1} = \psi_k^{n-1},$$

where in the last step we used that $(n-1)^2/2 \equiv -(n^2-1)/2 \bmod n$. On-site we retain the commutation relation $\chi_j\psi_j = \omega\psi_j\chi_j$, but due to the strings in our generalised 'Jordan-Wigner' transformation, we find the offsite ($j \neq k$) commutation relations

$$\chi_j\chi_k = \omega^{\mathrm{sgn}(k-j)}\chi_k\chi_j,\ \psi_j\psi_k = \omega^{\mathrm{sgn}(k-j)}\psi_k\psi_j,\ \chi_j\psi_k = \omega^{\mathrm{sgn}(k-j)}\psi_k\chi_j.$$

These relations are very unfortunate from the point of view of Fourier transformations, as they couple different momenta. In some sense these relations 'break' spatial parity symmetry, as swapping spatial indices has a non-trivial action on them (compare with (9.2.2), where there are no restrictions on j and k).

It is not hard to (partially) invert (9.3.3) to find

$$\tau_j = \omega^{-(n-1)/2}\chi_j^\dagger\psi_j,\ \sigma_j^\dagger\sigma_{j+1} = \omega^{-(n-1)/2}\psi_j^\dagger\chi_{j+1}. \tag{9.3.4}$$

Plugging this into (9.3.2) and cleaning things up a bit, we find the parafermionic version

$$H_n = -f\sum_{j=1}^{L}\sum_{\ell=1}^{n-1}\beta_\ell\omega^{\ell(\ell-n)/2}\chi_j^{n-\ell}\psi_j^\ell - J\sum_{j=1}^{L-1}\sum_{\ell=1}^{n-1}\gamma_\ell\omega^{\ell(\ell-n)/2}\psi_j^{n-\ell}\chi_{j+1}^\ell. \tag{9.3.5}$$

This looks very similar to (9.2.4), and indeed for $n = 2$ they are the same. The $n = 3$ case we discussed earlier now looks like

$$H_3 = -f\sum_{j=1}^{L}\left(e^{i\vartheta}\omega^*\chi_j^\dagger\psi_j + e^{-i\vartheta}\omega\chi_j\psi_j^\dagger\right) - J\sum_{j=1}^{L-1}\left(e^{i\varphi}\omega^*\psi_j^\dagger\chi_{j+1} + e^{-i\varphi}\omega\psi_j\chi_{j+1}^\dagger\right), \tag{9.3.6}$$

with the symmetric case

$$H_3(\varphi = \vartheta = -\pi/6) = if\sum_{j=1}^{L}\left(\chi_j^\dagger\psi_j - \chi_j\psi_j^\dagger\right) + iJ\sum_{j=1}^{L-1}\left(\psi_j^\dagger\chi_{j+1} - \chi_{j+1}^\dagger\psi_j\right).$$

9.4 Parafermion Edge Modes

While we have now managed to formulate the general $\mathbb{Z}_n$ invariant family of models (9.3.2) in terms of parafermions (see (9.3.5)) analogously to the formulation of the quantum Ising model in terms of Majoranas, we have not yet used this technology to uncover any physics. In the present section we will remedy this, by showing that the family of parafermionic models (9.3.5) also generically admit edge modes, and hence have degenerate groundstate manifolds. Moreover, many of the special properties of the edge modes of the quantum Ising model are also present here, such as the existence of non-local operators that cycle through the groundstates. Thus the Hamiltonians H_n all have n-fold degenerate groundstate manifolds, protected by a gap (for small f at least), which can be cycled through with non-local operators. This is the dream platform on which to implement *topological quantum computing*, i.e. unitary rotations in a low-dimensional Hilbert space that are topologically protected from decoherence [48, 60].

We start by generalizing the definition of a Fermionic (edge) zero mode in the appropriate way. Let Ψ be an operator such that $[H_n, \Psi] = 0$, $\omega^P\Psi = \omega\Psi\omega^P$, and $\Psi^\dagger\Psi = 1$ even for $L \to \infty$. Then we say that Ψ is a *parafermionic zero mode*. If additionally the matrix elements of Ψ are concentrated on the edges of the system, with bulk elements being exponentially small, we say that it is a *parafermionic*

edge zero mode. The most obvious example would be the operators χ_1 and ψ_L in the zero-field ($f = 0$) case of H_n, which is the model

$$H_n(f=0) = -J\sum_{j=1}^{L-1}\sum_{\ell=1}^{n-1}\gamma_\ell\omega^{\ell(\ell-n)/2}\psi_j^{n-\ell}\chi_{j+1}^{\ell}.$$

Analogous to the quantum Ising model written in Majorana operators, we see that χ_1 and ψ_L do not appear at all in the Hamiltonian, and it is a simple calculation that see that they also commute with it, since

$$\begin{aligned}\chi_1\psi_j^{n-\ell}\chi_{j+1}^{\ell} &= \omega^{n-\ell}\psi_j^{n-\ell}\chi_1\chi_{j+1}^{\ell} = \omega^n\psi_j^{n-\ell}\chi_{j+1}^{\ell}\chi_1 = \psi_j^{n-\ell}\chi_{j+1}^{\ell}\chi_1,\\ \psi_L\psi_j^{n-\ell}\chi_{j+1}^{\ell} &= (\omega^*)^n\psi_j^{n-\ell}\chi_{j+1}^{\ell}\psi_L = \psi_j^{n-\ell}\chi_{j+1}^{\ell}\psi_L.\end{aligned}$$

The $\mathbb{Z}_n$ generator ω^P can also be expressed in terms of parafermions as

$$\omega^P = \prod_{j=1}^{L}\left(\omega^{(n-1)/2}\psi_j^\dagger\chi_j\right),$$

from which one can show that ψ_L and χ_1 satisfy the required commutation relations with it. The normalisation condition is trivial to check.

Thus we have exact edge modes as long as $f = 0$, but is it again true that they exist for finite, albeit small enough, values of f? The answer turns out to be positive, but getting to that point will require some technical finesse. In the Majorana case, we were able to exploit the fact that applying the operator $[H_{\mathrm{qI}}, \cdot]$ applied to a_j and b_j yielded results that were again linear in Majorana operators. Unfortunately, for parafermions this is no longer the case, as trying to commute say ψ_j past a field term in H_n gives

$$[H_n, \psi_j] \propto [\chi_j^\dagger\psi_j, \psi_j] \propto \chi_j^\dagger\psi_j^2,$$

which is non-linear in both χ_j and ψ_j. This is the main technical obstacle that we will have to overcome. For notational convenience, split the Hamiltonian H_n into its fields terms $\mathcal{F}$ (those with f) and its interaction terms $\mathcal{J}$ (those with J). Then our goal is to study the two operators

$$\mathrm{ad}_{\mathcal{F}} = [\mathcal{F}, \cdot],\ \mathrm{ad}_{\mathcal{J}} = [\mathcal{J}, \cdot],$$

and prove that an iterative procedure like the one for the Majorana case exists. Naturally, $\mathrm{ad}_{\mathcal{J}}(\chi_1) = 0$, as it is an exact zero mode for $f = 0$, but

$$[H_n, \chi_1] = \mathrm{ad}_{\mathcal{F}}\ \chi_1 = f\sum_{\ell=1}^{n-1}\beta_\ell\omega^{\ell(\ell-n)/2}(1-\omega^{-\ell})\chi_1^{n-\ell+1}\psi_1^{\ell}.$$

If this can be written as a commutator with $\mathcal{J}$, we can bring it over to the other side and complete another step in the iterative procedure. In other words, we want to prove that

$$\left[(\mathrm{ad}_{\mathcal{J}})^{-1}(\mathrm{ad}_{\mathcal{F}})\right]^N\chi_1 \neq \varnothing,$$

for all $N \leq L-1$. Additionally, we want to make sure that the elements we pick admit a well-defined powercounting in f, so that we obtain a series that is convergent for f small enough.

To carry out this procedure, we first construct a convenient space to work in. Let $\mathcal{P}$ be the Hilbert space spanned by linear combinations of all possible products of parafermions. More formally, we set

$$\mathcal{P} := \mathrm{span}\left\{\chi_1^{p_1}\psi_1^{p_2}\cdots\chi_L^{p_{2L-1}}\psi_L^{p_{2L}}\,\middle|\,1 \leq p_j \leq n\right\}, \tag{9.4.1}$$

introduce the notation

$$|p_1, \ldots, p_{2L}\rangle := \chi_1^{p_1}\psi_1^{p_2}\cdots\chi_L^{p_{2L-1}}\psi_L^{p_{2L}},$$

and the inner product

$$\langle p_1, \ldots, p_{2L}|q_1, \ldots, q_{2L}\rangle = \delta_{p_1 q_1}\cdots\delta_{p_{2L}q_{2L}}.$$

We will use the symbols μ and ν to denote valid (obeying the restraint in (9.4.1)) multi-indices of length $2L$. As each parafermion only admits n non-trivial powers, there are two per site, and there are L sites, the dimension of $\mathcal{P}$ as a Hilbert space is n^{2L}. We may view

$$\mathrm{ad}_{H_n} = -\mathrm{ad}_{\mathcal{F}} - \mathrm{ad}_{\mathcal{J}},$$

as linear operators on this Hilbert space, thus they may be represented by $n^{2L}\times n^{2L}$ matrices. As a preparation to write down the matrices of the operators above, one may work out that

$$\begin{aligned}\left[\chi_j^{n-\ell}\psi_j^{\ell}, \chi_j^{p}\psi_j^{q}\right] &= \left(\omega^{-\ell p} - \omega^{\ell q}\right)\chi_j^{n-\ell+p}\psi_j^{\ell+q},\\ \left[\psi_j^{n-\ell}\chi_{j+1}^{\ell}, \psi_j^{p}\chi_{j+1}^{q}\right] &= \left(\omega^{-\ell p} - \omega^{\ell q}\right)\psi_j^{n-\ell+p}\chi_{j+1}^{\ell+q}.\end{aligned}$$

These relations imply two properties for the ad., not only do they conserve $\mathbb{Z}_n$ charge, they also can be split in the following way

$$\mathrm{ad}_{\mathcal{F}} = f\sum_{j=1}^{L}\mathcal{H}_{2j-1},\ \mathrm{ad}_{\mathcal{J}} = J\sum_{j=1}^{L-1}\mathcal{H}_{2j},$$

where $\mathcal{H}_k$ acts only on the k-th index and the $(k+1)$-th index. Let μ be a multi-index whose k-th index is p, and whose $(k+1)$-th index is q, and let ν be a multi-index whose k-th index is $p-\ell$, and whose $(k+1)$-th index is $q+\ell$. Then by the properties above, $\mathcal{H}_k$ only has non-vanishing matrix elements between multi-indices likes μ and ν (in particular it has only zeroes along the diagonal). A brief computation, using the commutators above, gives the matrix elements

$$\begin{aligned}\left(\mathcal{H}_{2j}\right)_{\mu\nu} &= \omega^{\ell(\ell-n)/2}\gamma_\ell\ \omega^{-\ell p}\left(1-\omega^{\ell q}\right),\\ \left(\mathcal{H}_{2j-1}\right)_{\mu\nu} &= \omega^{\ell(\ell-n)/2}\beta_\ell\ \omega^{-\ell p}\left(1-\omega^{\ell q}\right).\end{aligned} \tag{9.4.2}$$

With these explicit expressions in hand, it is not difficult to check that all matrices $\mathcal{H}_k$ are Hermitian, and hence so is the full matrix $\mathcal{H}$ associated to ad_{H_n}.

As we want to be able to invert the action of $\mathrm{ad}_{\mathcal{J}}$ for our procedure, we should check first whether or not the individual pieces $\mathcal{H}_k$ are invertible. The reason for this is that the 0-th order term in the left edge mode operator will be χ_1, and applying $\mathrm{ad}_{\mathcal{F}}$ gives terms containing both χ_1 and ψ_1, that means we end up with a vector whose first two indices are non-zero and the rest zero. Moreover, notice that at this first step, only the $\mathcal{H}_1$ part of $\mathrm{ad}_{\mathcal{F}}$ has a non-trivial action on χ_1, and that only the $\mathcal{H}_2$ part of $\mathrm{ad}_{\mathcal{J}}$ has a non-trivial action on the result. This is a general pattern, that applying either $\mathrm{ad}_{\mathcal{F}}$ or trying to invert $\mathrm{ad}_{\mathcal{J}}$ moves the boundary between non-zero and zero indices one step farther to the right. Thus we see that we will generically be able to restrict the action and inverting of these operators to smaller subspaces where only certain indices are non-zero. This motivates investigating the individual pieces $\mathcal{H}_{2j}$ of the matrix representing $\mathrm{ad}_{\mathcal{J}}$. Now write $q = Q - p'$ and look again at the matrix elements in (9.4.2). One way to show that this matrix is invertible is to prove that none of its eigenvalues vanish. It turns out that the eigenvectors are labelled by $0 \le r \le n-1$ and have coordinates

$$u_p^r = \omega^{pr}\omega^{p(n-p)/2}.$$

This then gives eigenvalues

$$\lambda^{Q,r} v_p^r = \sum_{p'} \omega^{\ell(\ell-n)/2} \gamma_\ell \; \omega^{-\ell p} \left(1 - \omega^{\ell q}\right) v_{p'}^r,$$

where $\ell \equiv p - p' \bmod n$. To compute this, notice that we can replace ℓ by $p - p'$ without trouble in the latter two powers of ω. However, the first occurrence features a product of ℓ and $\ell - n$, and it is not immediately clear that this value is invariant mod n by this replacement. Yet this turns out to be the case, as supposing that $\ell = p - p' + zn$ gives

$$\frac{1}{2}\ell(\ell - n) = \frac{1}{2}\left(p - p' + zn\right)\left(p - p' - zn - n\right) = \frac{1}{2}\left(p - p'\right)\left(p - p' - n\right) + n\left[z(p - p') + n\frac{z(z-1)}{2}\right].$$

This allows us to rewrite the eigenvalue equation as

$$\lambda^{Q,r} v_p^r = \omega^{pr} \sum_{p'} \gamma_{p-p'} \; \omega^{-(p-p')r} \left(1 - \omega^{(p-p')Q}\right) \omega^{(p-p')(p-p'-n)/2} \omega^{-(p-p')p} \omega^{p'(n-p')/2}.$$

One can check that combining the last three powers of ω yields

$$\frac{1}{2}(p - p')(p - p' - n) - (p - p')p + \frac{1}{2}p'(n - p') = -\frac{1}{2}p(n + p) \equiv \frac{1}{2}p(n - p) \bmod n,$$

which is the last remaining factor of the eigenvector. Finally, we find that the eigenvalues are (reverting to summing over ℓ)

$$\lambda^{Q,r} = \sum_{\ell}^{n-1} \gamma_\ell \; \omega^{-\ell r} \left(1 - \omega^{\ell Q}\right).$$

If $Q = 0$, all the eigenvalues vanish, but this is of no concern as such terms will not appear in our edge mode, since we start with an operator that has $\mathbb{Z}_n$ charge 1 and this is preserved by taking commutators. However, there are enough degrees of freedom present in the γ_ℓ to arrange for an eigenvalue to be zero, but generically this will not be the case. One special case is worth considering, that of choosing all γ_ℓ to be real. Now take some r, and set $Q = 2r$, then (with the convention that $\gamma_{n/2} = 0$ if n is odd)

$$\begin{aligned}
\lambda^{2r,r} &= \gamma_{n/2} \; \omega^{nr/2}(1 - \omega^{nQ/2}) + \sum_{\ell<n/2} \gamma_\ell \left(\omega^{-\ell r}(1 - \omega^{\ell Q})\omega^{\ell r}(1 - \omega^{-\ell Q})\right), \\
&= \gamma_{n/2} \; \omega^{nr/2}(1 - \omega^{nr}) + \sum_{\ell<n/2} \gamma_\ell \left(\omega^{-\ell r} - \omega^{\ell r} + \omega^{\ell r} - \omega^{-\ell r}\right), \\
&= 0.
\end{aligned}$$

Thus if we take all the coefficients in the interaction term of H_n to be real, we cannot construct edge modes using the procedure. This phenomenon has a physical basis, namely one may take all the coefficients to be real in the case of purely ferro-or antiferromagnetic couplings, and in this case parity and time-reversal symmetry are unbroken. We conclude that it is necessary (at finite transverse field) for the interactions in the $\mathbb{Z}_n$ chain to be chiral for edge modes to exist.

For the remainder of this section we will assume that all the eigenvalues above are finite, and thus that the $\mathcal{H}_{2j}$ are all invertible in the subspace where $p_{2j} + p_{2j+1} \not\equiv 0 \bmod n$ (this corresponds to taking $Q \neq 0$ in the above discussion). This is allowed because as remarked earlier, we start on the left side of the chain with χ_1, and each step in the algorithm pushes the position of the right-most non-zero index one site further, meaning that we will always try to invert $\mathcal{H}_{2j}$ when acting on vectors that

satisfy $p_{2j} + p_{2j+1} \not\equiv 0 \bmod n$, as $p_{2j} \not\equiv 0 \bmod n$ but $p_{2j+1} \equiv 0 \bmod n$. In short, we assume that there exist operators $\mathcal{G}_{2j}$ satisfying

$$\mathcal{G}_{2j}\mathcal{H}_{2j} |\ldots, p, q, \ldots\rangle = \mathcal{H}_{2j}\mathcal{G}_{2j} |\ldots, p, q, \ldots\rangle = (1 - \delta_{0,p+q \bmod n}) |\ldots, p, q, \ldots\rangle .$$

Thus the first order correction for Ψ becomes

$$\Psi = \chi_1 - \frac{f}{J}\mathcal{G}_2\mathcal{H}_1\chi_1 + \ldots .$$

Indeed,

$$\begin{aligned} \mathrm{ad}_{H_n}\Psi &= f\mathcal{H}_1\chi_1 - J\mathcal{H}_2\frac{f}{J}\mathcal{G}_2\mathcal{H}_1\chi_1 + \mathcal{O}\left((f/J)^2\right), \\ &= f\left(\mathcal{H}_1 - \mathcal{H}_2\mathcal{G}_2\mathcal{H}_1\right)\chi_1 + \mathcal{O}\left((f/J)^2\right), \\ &= 0 + \mathcal{O}\left((f/J)^2\right). \end{aligned}$$

However, we now run into the next obstacle, namely if we iterate again

$$\mathrm{ad}_{\mathcal{F}}\Psi = f\mathcal{H}_1\chi_1 - f\left(\mathcal{H}_1 + \mathcal{H}_3\right)\frac{f}{J}\mathcal{G}_2\mathcal{H}_1\chi_1 + \ldots,$$

as the first order correction may contain parafermions up to χ_2, which are not in the kernel of $\mathcal{H}_3$. We see that we now need to know whether the 'partial sums' of $\mathrm{ad}_{\mathcal{J}}$ are invertible to continue the procedure. A priori, we cannot see if this is true or not, especially since $\mathrm{ad}_{\mathcal{J}}$ has many null-vectors, for example any vector that satisfies $p_{2j} + p_{2j+1} \equiv 0 \bmod n$ for all j.

At this point it is convenient to introduce some additional notation, we will use $|l\rangle$ to denote the terms in Ψ that are of order l in f/J. Let $|0_i\rangle$ be an orthonormal basis of the kernel of $\mathrm{ad}_{\mathcal{J}}$, such that $\mathrm{ad}_{\mathcal{F}}$ is diagonal (this is possible as both operators are Hermitian). Write φ_i for the diagonal elements of $\mathrm{ad}_{\mathcal{F}}$ in this basis. If there is an index i such that

$$\langle 0_i| \mathrm{ad}_{\mathcal{F}} |l\rangle \neq 0,$$

we cannot directly apply the next step where we invert $\mathrm{ad}_{\mathcal{J}}$. Denote by P the projection operator onto the kernel of $\mathrm{ad}_{\mathcal{J}}$. Then there exist numbers v_i^l such that

$$P\left(\mathrm{ad}_{\mathcal{F}} |l\rangle\right) = \sum_i v_i^l |0_i\rangle = \sum_{i,\varphi_i\neq 0} \varphi_i \left(\frac{v_i^l}{\varphi_i} |0_i\rangle\right) = P\ \mathrm{ad}_{\mathcal{F}} \sum_{i,\varphi_i\neq 0} \left(\frac{v_i^l}{\varphi_i} |0_i\rangle\right).$$

Hence there exists a vector $|l_0\rangle$ such that $\mathrm{ad}_{\mathcal{F}}\left(|l\rangle - |l_0\rangle\right)$ is orthogonal to the kernel of $\mathrm{ad}_{\mathcal{J}}$, and this can then be inverted to continue the procedure. The powercounting is preserved by this procedure as $|l_0\rangle \sim (f/J)^l$ as it is a part of $|v_l\rangle$. We should however argue that this correction does not move Ψ away from the edge, in the sense that the parafermion χ_l should always appear with at least the power $(f/J)^{l-1}$. To show this, we need to consider the leading terms at each step of the procedure. It is not difficult to convince oneself that the leading terms are the following

$$\Psi_{\text{leading}} = \chi_1 - \frac{f}{J}\mathcal{G}_2\mathcal{H}_1\chi_1 + \left(\frac{f}{J}\right)^2 \mathcal{G}_4\mathcal{H}_3\mathcal{G}_2\mathcal{H}_1\chi_1 - \left(\frac{f}{J}\right)^3 \mathcal{G}_6\mathcal{H}_5\mathcal{G}_4\mathcal{H}_3\mathcal{G}_2\mathcal{H}_1\chi_1 + \ldots . \tag{9.4.3}$$

In particular, there are no corrections in the sense we just discussed to these leading terms, and we see that each parafermion occurs with the correct power.

The expression (9.4.3) might look innocent enough, but is is not very practical to use it to calculate the edge mode operators. As an example of this, we suppose that $n = 4$ and compute the first order correction $\propto \mathcal{G}_2\mathcal{H}_1\chi_1$. From (9.4.2) we obtain

$$\mathcal{H}_1 \left|1, 0, \ldots\right\rangle = \sum_{\ell=1}^{3} \beta_\ell \omega^{\ell(\ell-4)/2}(\omega^{-\ell} - 1) \left|1-\ell, \ell, 0, \ldots\right\rangle .$$

To find the action of $\mathcal{G}_2$ on this, recall that $\mathcal{H}_2$ only acts on the second and third index, and that it conserves the sum of these indices mod 4. Thus to determine $\mathcal{G}_2 \left|1-\ell, \ell, 0, \ldots\right\rangle$, we can restrict to the subspace spanned by $\left|1-\ell, \ell, 0, \ldots\right\rangle, \left|1-\ell, \ell+3, 1, 0\ldots\right\rangle, \left|1-\ell, \ell+2, 2, 0\ldots\right\rangle, \left|1-\ell, \ell+1, 3, 0\ldots\right\rangle$, which we shall denote by $\left|0\right\rangle_\ell, \ldots, \left|3\right\rangle_\ell$. Now for $\ell = 1, 2, 3$ one can write out the 4×4 matrix of $\mathcal{H}_2$ in the relevant subspace (denoted $\mathcal{M}_\ell$), and invert it using a computer algebra programme. It turns out that

$$\begin{aligned}
\det(\mathcal{M}_1) &= 4\left(\gamma_1^4 + 4\gamma_2^4 - 8\gamma_1\gamma_2^2\gamma_3 + 2\gamma_1^2\gamma_3^2 + \gamma_3^4\right), \\
\det(\mathcal{M}_2) &= -16\left(\gamma_1^2 - \gamma_3^2\right)^2, \\
\det(\mathcal{M}_3) &= \det(\mathcal{M}_1),
\end{aligned}$$

and

$$\begin{aligned}
(\mathcal{M}_1)^{-1}\left|0\right\rangle_1 = \frac{1}{\det(\mathcal{M}_1)}\big[&-4i\gamma_2(\gamma_1^2 - \gamma_3^2)\left|0\right\rangle_1 - 2\sqrt{2}i(\gamma_1^3 - 2\gamma_2^2\gamma_3 + \gamma_1\gamma_3^2)\left|1\right\rangle_1 + 8(\gamma_2^3 - \gamma_1\gamma_2\gamma_3)\left|2\right\rangle_1 \\
&+ 2\sqrt{2}(\gamma_1^2\gamma_3 - 2\gamma_1\gamma_2^2 + \gamma_3^3)\left|3\right\rangle_1\big], \\
(\mathcal{M}_2)^{-1}\left|0\right\rangle_2 = \frac{8e^{i\pi 3/4}}{\det(\mathcal{M}_2)}&\left[\gamma_1(\gamma_1^2 - \gamma_3^2)\left|1\right\rangle_2 + \gamma_3(\gamma_3^2 - \gamma_1^2)\left|3\right\rangle_2\right], \\
(\mathcal{M}_3)^{-1}\left|0\right\rangle_3 = \frac{1}{\det(\mathcal{M}_3)}\big[&4i\gamma_2(\gamma_1^2 - \gamma_3^2)\left|0\right\rangle_3 + 2\sqrt{2}(\gamma_1^3 - 2\gamma_2^2\gamma_3 + \gamma_1\gamma_3^2)\left|1\right\rangle_3 + 8(\gamma_2^3 - \gamma_1\gamma_2\gamma_3)\left|2\right\rangle_3 \\
&- 2\sqrt{2}i(\gamma_1^2\gamma_3 - 2\gamma_1\gamma_2^2 + \gamma_3^3)\left|3\right\rangle_3\big].
\end{aligned}$$

For the specific choice of coefficients $\gamma_1 = (1/4)(\frac{1+i}{\sqrt{2}})$ and $\gamma_2 = 1/4$ (and of course $\gamma_1 = \gamma_3^*$), this finally results in the leading terms

$$\begin{aligned}
\chi_1 - \frac{f}{J}\Big\{&\beta_1 e^{i\pi/4}(1+i)\left[-2\psi_1 - (1+i)\chi_2 + (1+i)\psi_1^2\chi_2^3\right] + 2\beta_2\left[-i\chi_1^3\psi_1\chi_2 + \chi_1^3\psi_1^3\chi_2^3\right] \\
&+ \beta_3 e^{i\pi/4}(1-i)\left[2\chi_1^2\psi_1^3 + (1-i)\chi_1^2\psi_1^2\chi_2 + (1-i)\chi_1^2\chi_2^3\right]\Big\} + \ldots .
\end{aligned}$$

Finally we turn to the convergence of Ψ. We will not be able to say very much in general, other than that the series does converge for f/J small enough. The argument is as follows. Certainly we need $f/J < 1$ for convergence, but this is not enough, as there are now multiple terms at each order of f/J. However, the amount of such terms that are present is certainly at most exponential, say that they can be bounded by Ce^{cl}, where $C, c > 0$ are constants and l is the order. Since we are in finite dimensions, all other operations are bounded and we only need to worry about the number of terms present versus their power of f/J. Now, as $(f/J)^l = \exp(\ln(f/J)l)$ and $\ln(f/J) < 0$, we can indeed make f/J small enough to make $c - \ln(f/J)$ negative and hence ensure convergence. Determining the precise radii of convergence for various choices of coefficients should be a hard problem in general, as it is related to constructing the phase diagram of the model H_n, which as we have seen may exhibit a wide variety of different physics.

10 Parafermions as Particles: Fock Parafermions

10.1 Constructing the Fock Space

In the $n = 2$ case of the story told in the previous chapter, there was a strong relation between the Majorana (para)Fermions and spinless electrons, codified in the relation (9.2.1). This has several major benefits. First, it increases the chances of finding an experimentally realisable system in which these excitations can be found (in case of the quantum Ising chain that would be the model (9.1.6)). Recall that the spin-statistics theorem is not valid in low (less than 3+1) dimensions, so that particles like spinless fermions are not only allowed, but in fact experimentally realisable in quantum chains with strong spin-orbit coupling to a superconducting substrate and certain magnetic fields (see [53, 66]). Second, it gives access to the powerful technology of Fock spaces. In this section, we will concern ourselves with the second point. We will construct a Fock space that hosts '*Fock parafermion* creation and annihilation operators' and explore their relation to the parafermions introduced in (9.3.3). We will go about this in a somewhat different way to the 'q-deformed' Fock space in Chapter 3, or to the standard construction for Bosons and Fermions (see Chapter 1 of [15]). Instead, we follow [22], where Cobanera and Ortiz gave the first construction, and first derived the important relation (10.2.1).

The construction does start in the standard way. Suppose that $\mathcal{H}$ is the Hilbert space of single particle states, which we will assume to have dimension p. For example, this could represent the Hilbert space of Warnier states related to a p-site chain (if there is one orbital per site). Choose an orthonormal basis $e_1, \ldots, e_p$ of $\mathcal{H}$ and fix an ordering of this basis (we will use the index). These are the single-particle wavefunctions available in the system. Then a many-particle state of m independent, indistinguishable copies of our particle should be completely described by the numbers $o_1, \ldots, o_p$, where o_j is the occupation number of the state j. To each list of such numbers we will associate a unique (up to an inconsequential phase) vector

$$|o_1, \ldots, o_p\rangle \in \operatorname{span}\{e_{j_1} \otimes \cdots \otimes e_{j_m} \mid \#j_k \text{ with } j_k = \ell \text{ is } o_\ell\}.$$

The precise vector we take will be determined by an iterative procedure later (for instance for Bosons one would take the symmetric combination of uniform weight). Our Fock space will then be

$$\mathcal{F}_n(\mathcal{H}) = \overline{\bigoplus_{m\geq 0} \operatorname{span}\{|o_1, \ldots, o_p\rangle \, | o_1 + \cdots + o_p = m\}},$$

where the closure is taken with respect to the inner product

$$\langle o_1, \ldots, o_p | o'_1, \ldots, o'_p\rangle = \prod_{j=1}^{p} \delta_{o_j o'_j}, \tag{10.1.1}$$

and with $m = 0$ we mean the span of the vacuum vector $\mathbb{C}\,|0\rangle$.

We now begin to introduce the peculiar ingredients of parafermions into the soup. These are the exchange statistics of parafermions. To implement these, we should introduce an algebra structure on $\mathcal{F}_n(\mathcal{H})$ that models the adding of particles to states. For notational convenience, we will write

$$|o_j\rangle = |0, \ldots, 0, o_j, 0, \ldots, 0\rangle, \ |o_j, o_k\rangle = |0, \ldots, 0, o_j, 0, \ldots, 0, o_k, 0, \ldots, 0\rangle, \text{ etc.}$$

Rule number one is then $|o_j = 1\rangle^\ell = |o_j = \ell\rangle$, and rule number two is

$$|o_j\rangle\,|o_k\rangle = \omega^{o_j o_k}\,|o_k\rangle\,|o_j\rangle = |o_j, o_k\rangle \text{ for } j < k.$$

This procedure completely determines $\mathcal{F}_n(\mathcal{H})$ as follows, $|o_j = 1\rangle = e_j$ and

$$|o'_1, \ldots, o'_p\rangle = |o_1 = 1\rangle^{o'_i} \cdots |o_p = 1\rangle^{o'_p}.$$

A consequence of these rules is that the algebra is unital with the vacuum vector $|0\rangle$ as unit. We assume that this is the unique unit of the algebra, and moreover that it spans the center of the algebra. In other words, any many-particle state that commutes with all others is a scalar multiple of the vacuum. There are actually many such states, namely any state generated by the vectors $|o_j = n\rangle$. Hence any such state should be a multiple of the vacuum, but as states with differing numbers of particles are orthogonal, this coefficient must be zero. It follows that there is a maximal value of $n-1$ for all the occupation numbers o_j. We conclude that we have constructed a Fock space of particles that satisfy ω-exchange and n-exclusion statistics.

Our next project is to come up with a definition for creation and annihilation operators d and $d^\dagger$ on this Fock space such that they satisfy the right statistics. Let us start with the creation operators $d_j^\dagger$. It is clear that these should send a state with occupation number o_j to one with $o_j + 1$. However we still have the freedom to multiply by a phase, which should be judiciously chosen to yield the correct commutation relations. Thankfully, the algebra structure we defined earlier will do all the work for us if we set

$$d_j^\dagger |o_1, \ldots, o_p\rangle = |o_j = 1\rangle |o_1, \ldots, o_p\rangle = \omega^{-\sum_{k<j} o_k} |o_1, \ldots, o_j + 1, \ldots, o_p\rangle . \tag{10.1.2}$$

The condition that d_j is the adjoint to $d_j^\dagger$ with respect to the inner product (10.1.1) then fixes

$$d_j |o_1, \ldots, o_p\rangle = \omega^{\sum_{k<j} o_k} |o_1, \ldots, o_j - 1, \ldots, o_p\rangle . \tag{10.1.3}$$

Simple computations then show that these operators satisfy the relations for $j \neq k$

$$d_j^{(\dagger)n} = 0,\ d_j^{(\dagger)} d_k^{(\dagger)} = \omega^{\mathrm{sgn}(k-j)} d_k^{(\dagger)} d_j^{(\dagger)},\ d_j^\dagger d_k = (\omega^*)^{\mathrm{sgn}(k-j)} d_k d_j^\dagger,\ d_j d_k^\dagger = (\omega^*)^{\mathrm{sgn}(k-j)} d_k^\dagger d_j. \tag{10.1.4}$$

Unfortunately, now things become more complicated. For ordinary fermions we would have $c_j^\dagger c_j + c_j c_j^\dagger = I$, which can be understood by seeing that the first term is the projection onto the subspace with $o_j = 1$, and the second term is the projection onto $o_j = 0$. If one considers the action of the operators $d_j^{\dagger m} d_j^m$ and $d_j^{n-m} d_j^{\dagger(n-m)}$, one will notice that the first one projects onto the subspace with occupation numbers $o_j \geq m$, and the second to the subspace with $o_j < m$. Therefore we have that

$$d_j^{\dagger m} d_j^m + d_j^{n-m} d_j^{\dagger(n-m)} = I, \tag{10.1.5}$$

for all $1 \leq m < n$. This is substantially more complicated than the $n = 2$ Majorana algebra (9.2.2), as these constraints are non-linear in d_j and $d_j^\dagger$.

Of course we can also define number (or occupation) operators

$$O_j |o_1, \ldots, o_n\rangle = o_j |o_1, \ldots, o_n\rangle .$$

In view of (10.1.5) we cannot expect a relation with the creation and annihilation operators as simple as those in the Fermionic or Bosonic cases. Recalling that $d_j^{\dagger m} d_j^m$ projects onto states with $o_j \geq m$, it is not difficult to see that

$$O_j = \sum_{m=1}^{n-1} d_j^{\dagger m} d_j^m, \tag{10.1.6}$$

which is indeed more complicated. Fortunately, despite this more complicated formula, O_j retains the properties that the Ferminonic number operators have. Take for instance the well known fact that the Fermionic number operators generate the U(1) translations of the Fermionic creation and annihilation operators. We can compute

$$\begin{aligned}
\left[d_j^{\dagger m} d_j^m, d_j\right] &= d_j^{\dagger m} d_j^{m+1} - d_j d_j^{\dagger m} d_j^m, \\
&= d_j^{\dagger m} d_j^{m+1} - \left(I - d_j^{\dagger(n-1)} d_j^{(n-1)}\right) d_j^{\dagger(m-1)} d_j^m, \\
&= d_j^{\dagger m} d_j^{m+1} - d_j^{\dagger(m-1)} d_j^m,
\end{aligned}$$

as $m-1 < n-1$. It follows that

$$\begin{aligned}[O_j, d_j] &= \sum_{m=1}^{n-1} \left(d_j^{\dagger m} d_j^{m+1} - d_j^{\dagger(m-1)} d_j^{m-1} \right) d_j, \\ &= -d_j + d_j^{\dagger(n-1)} d_j^n, \\ &= -d_j.\end{aligned}$$

This also gives us

$$\left[O_j, d_j^\dagger\right] = -\left([O_j, d_j]\right)^\dagger = d_j^\dagger.$$

Combining these identities then yields that the O_j are still U(1) generators.

10.2 Building the Bridge to Parafermions

We now want to generalise (9.2.1) and find the relations connecting the Fock parafermions we have constructed to the 'ordinary' parafermions we saw before. There is a priori no clear way to do this, but it turns out that we can rewrite (9.2.1) a little bit to read

$$a_j = c_j + c_j^\dagger \, , b_k = (-1)^{(2-1)/2} \left(c_k (-1)^{c_k^\dagger c_k} + c_k^\dagger \right),$$

which has a natural generalisation[8] to

$$\chi_j = d_j + d_j^{\dagger(n-1)}, \;\; \psi_k = \omega^{(n-1)/2} \left(d_k \omega^{O_k} + d_k^{\dagger(n-1)} \right). \tag{10.2.1}$$

We check that these indeed define operators with the correct relations. The commutation relations are straightforward to verify. Taking the n-th power of χ_j yields

$$\chi_j^n = \sum_{\ell=0}^{n-1} d_j^{n-1-\ell} d_j^{\dagger(n-1)} d_j^\ell = \sum_{\ell=0}^{n-1} P_j(\ell) = 1,$$

Where $P_j(\ell)$ is the projector onto the subspace with $o_j = \ell$. In the first equality, all other terms vanish because they either remove n particles, or add more than n. The second equality follows from the realisation that only states with $o_j = 0$ survive application of $d_j^{\dagger(n-1)}$, so that we must have that d_j^ℓ brings the input state into that sector. By powercounting we see that the number of particles in the state j is not changed, so all phase factors cancel and we find the projectors. We now check

$$\chi_j^\dagger \chi_j = d_j^\dagger d_j + d_j^n + d_j^{\dagger n} + d_j^{n-1} d_j^{\dagger(n-1)} = I.$$

The calculations for ψ_k are similar, we start again by taking the n-th power to find

$$\begin{aligned}\psi_k^n &= \omega^{n(n-1)/2} \sum_{\ell=0}^{n-1} \left(d_k \omega^{O_k} \right)^{n-1-\ell} d_k^{\dagger(n-1)} \left(d_k \omega^{O_k} \right)^\ell, \\ &= \omega^{n(n-1)/2} \sum_{\ell=0}^{n-1} d_k \omega^{\ell+1} \cdots d_k \omega^{n-1} d_k^{\dagger(n-1)} d_k \omega \cdots d_k \omega^\ell, \\ &= \omega^{n(n-1)} \sum_{\ell=0}^{n-1} P(\ell), \\ &= I,\end{aligned}$$

[8] This is different from (123) in [22] due to the difference in conventions with [34] (which we follow) in the definition of parafermions.

using similar arguments as before. Additionally

$$\psi_k^\dagger \psi_k = \omega^{-O_k} d_k^\dagger d_k \omega^{O_k} + \omega^{-O_k} d_k^{\dagger n} + d_k^n \omega^{O_k} + d_k^{n-1} d_k^{\dagger(n-1)} = d_k^\dagger d_k + d_k^{n-1} d_k^{\dagger(n-1)} = I.$$

Of course we can now also build quantum clocks out of Fock parafermions. For this we improve (9.3.4) to

$$\tau_j = \omega^{-(n-1)/2} \chi_j^\dagger \psi_j,\ \sigma_j = \omega^{-(n-1)/2} \chi_j \prod_{k<j} \omega^{(n-1)/2} \psi_k^\dagger \chi_k. \tag{10.2.2}$$

So that

$$\tau_j = d_j^{n-1} d_j^{\dagger(n-1)} + d_j^\dagger d_j \omega^{O_j} = \omega^{O_j},$$

which is precisely the matrix given in (9.3.1). It follows immediately that

$$\sigma_j = \left(d_j + d_j^{\dagger(n-1)} \right) \omega^{-\sum_{k<j} O_k}. \tag{10.2.3}$$

This can also be cast into a slightly different form using a generalisation of the $n = 2$ identity $(-1)^{c_j^\dagger c_j} = 1 - 2c_j^\dagger c_j$. We construct this as follows, recall that $d_j^{\dagger m} d_j^m$ projects onto the subspace with $o_j \geq m$. This suggests that we try to write ω^ℓ as a sum, which can be achieved by

$$\omega^\ell = 1 - (1-\omega)\left(1 + \omega + \cdots + \omega^{\ell-1}\right).$$

By staring at this for long enough, one sees that

$$\omega^{O_j} = I - (1-\omega) \sum_{m=1}^{n-1} \omega^{m-1} d_j^{\dagger m} d_j^m. \tag{10.2.4}$$

In principle this allows us to express parafermions and quantum clock variables entirely in terms of the d's. It is also worth pointing out that Fock parafermions can be used to model fractionalisation of electrons [23].

We close this section by inverting the relations (10.2.1). For this we start by decomposing σ_j into a 'creation' and an 'annihilation' operator,

$$\sigma_j = \left(d_j + d_j^{\dagger(n-1)} \right) \prod_{k<j} \omega^{-O_k} = d_j \prod_{k<j} \omega^{-O_k} + \left(d_j^\dagger \prod_{k<j} \omega^{O_k} \right)^{n-1},$$

leading to the definitions

$$B_j = d_j \prod_{k<j} \omega^{-O_k},\ B_j^\dagger = d_j^\dagger \prod_{k<j} \omega^{O_k}, \tag{10.2.5}$$

which is usually called the Fradkin-Kadanoff transformation [37]. In terms of our original clock variables σ_j, this reads $\sigma_j = B_j + B_j^{\dagger(n-1)}$. These transformations are 'canonical' in the sense that they preserve the occupation operators

$$O_j = \sum_{m=1}^{n-1} d_j^{\dagger m} d_j^m = \sum_{m=1}^{n-1} B_j^{\dagger m} B_j^m.$$

To compute the commutation relations of the B operators, first notice that for $j \neq k$ we have

$$\left[\omega^{O_j}, d_k\right] = 0,\ \omega^{O_j} d_j = \omega^* d_j \omega^{O_j}.$$

Then for $j \neq k$ (wlog $j < k$)

$$[B_j, B_k] \propto \left[d_j, d_k \omega^{-O_j}\right] = (d_j d_k - \omega d_k d_j) = 0,$$

so that this is also true for the annihilation operators. The mixed commutator evaluates to

$$\left[B_j, B_k^\dagger\right] \propto \left[d_j, d_k^\dagger \omega^{O_j}\right] = \left(d_j d_k^\dagger - \omega^* d_k^\dagger d_j\right) \omega^{O_j} = 0.$$

These are Bosonic commutation relations, but on-site the story is slightly altered, namely

$$B_j^n = B_k^{\dagger n} = 0, \;\; B_j^{\dagger m} B_j^m + B_j^{n-m} B_j^{\dagger(n-m)} = I.$$

We will call these operators *Weyl hard-core Bosons*, as they describe particles that are Bosonic under exchange, but only allow $n-1$ particles per state. Note that the on-site relations cannot be Bosonic, due to the fact that there exist no matrices that witness that canonical commutation relations, and since the on-site Hilbert space is finite dimensional, the operators B_j are (locally) matrices.

Our strategy now is to relate the hard-core Bosons also directly to the quantum clock variables, and upon combining with their definitions we find the inverted form of (10.2.1) that we are after. One can compute the matrix elements

$$\langle o_j = q| \, \sigma_j^\dagger B_j \, |o_j = p\rangle = \delta_{pq} \left(1 - \delta_{p0}\right).$$

Thus the matrix $\sigma_j^\dagger B_j$ is diagonal, and because the n matrices $\tau_j^{\dagger m}$ $(0 \leq m \leq n-1)$ are also diagonal and linearly independent, we can use them as a basis. Actually, the $\tau_j^{\dagger m}$ are even an orthonormal basis of the diagonal matrices with respect to the inner product $\mathrm{tr}\big(N^\dagger M\big)/n$. It follows (by using a basis expansion) that we have

$$\sigma_j^\dagger B_j = \sum_{m=0}^{n-1} \frac{1}{n} \mathrm{tr}\Big(\tau_j^m \sigma_j^\dagger B_j\Big) \tau_j^{\dagger m},$$

which yields that

$$B_j = \frac{n-1}{n} \sigma_j - \frac{1}{n} \sum_{m=1}^{n-1} \sigma_j \tau_j^{\dagger m}.$$

Now we have that

$$d_j = B_j \prod_{k<j} \omega^{O_k} = \left(\frac{n-1}{n} \sigma_j - \frac{1}{n} \sum_{m=1}^{n-1} \sigma_j \tau_j^{\dagger m}\right) \prod_{k<j} \tau_k.$$

Plugging (10.2.2) into this we find

$$\begin{aligned}
d_j &= \left(\frac{n-1}{n} \omega^{-(n-1)/2} \chi_j \prod_{\ell<j} \tau_\ell^\dagger - \frac{1}{n} \sum_{m=1}^{n-1} \omega^{-(n-1)/2} \chi_j \left(\prod_{\ell<j} \tau_\ell^\dagger\right) \left(\omega^{-(n-1)/2} \chi_j^\dagger \psi_j\right)^{\dagger m}\right) \prod_{k<j} \tau_k, \\
&= \omega^{-(n-1)/2} \left[\frac{n-1}{n} \chi_j - \frac{1}{n} \sum_{m=1}^{n-1} \chi_j \left(\omega^{(n-1)/2} \psi_j^\dagger \chi_j\right)^m\right], \\
&= \omega^{-(n-1)/2} \left[\frac{n-1}{n} \chi_j - \frac{1}{n} \sum_{m=1}^{n-1} \omega^{m(m-n)/2} \chi_j^{m+1} \psi_j^m\right].
\end{aligned}$$

10.3 Example: Transforming the 3-and 4-Potts Models

Having established the relation (10.2.1) between parafermions and Fock parafermions, it is interesting to transform the family (9.3.5) into this language. In this section, we will carry out the calculation for the cases $n = 3$ and $n = 4$. Equation (9.3.6) gives the general $\mathbb{Z}_3$-invariant Hamiltonian in terms of parafermions

$$H_3 = -f\sum_{j=1}^{L}\left(\overline{\omega}\chi_j^\dagger\psi_j + \text{H.c.}\right) - J\sum_{j=1}^{L-1}\left(\overline{\omega}\psi_j^\dagger\chi_{j+1} + \text{H.c.}\right),$$

where $\omega = e^{2\pi i/3}$ and $\alpha_1, \hat{\alpha}_1 \in \mathbb{C}$. By polar decomposition, we can absorb the modulus degree of freedom of these coefficients into f and J respectively, so that this model can also be written as

$$H_3 = -f\sum_{j=1}^{L}\left(e^{i\varphi}\overline{\omega}\chi_j^\dagger\psi_j + \text{H.c.}\right) - J\sum_{j=1}^{L-1}\left(e^{i\vartheta}\overline{\omega}\psi_j^\dagger\chi_{j+1} + \text{H.c.}\right). \tag{10.3.1}$$

To get rid of the phase factor in ψ_k in Equation (10.2.1), we can use Equation (10.2.4), which in this case reads

$$\omega^{O_k} = 1 + (\omega - 1)\left[d_k^\dagger d_k + \omega d_k^{\dagger 2} d_k^2\right]. \tag{10.3.2}$$

Thus[9],

$$\begin{aligned}\psi_k &= \omega\left[d_k + (\omega-1)\left(d_k d_k^\dagger d_k + \omega d_k d_k^{\dagger 2} d_k^2\right) + d_k^{\dagger 2}\right],\\ &= \overline{\omega}d_k + (1-\overline{\omega})d_k^\dagger d_k^2 + \omega d_k^{\dagger 2}.\end{aligned}$$

Now, we transform the field terms in (10.3.1), which gives

$$\begin{aligned}\overline{\omega}\chi_j^\dagger\psi_j &= \overline{\omega}\left(d_j^\dagger + d_j^2\right)\left[\overline{\omega}d_j + (1-\overline{\omega})d_j^\dagger d_j^2 + \omega d_j^{\dagger 2}\right],\\ &= \omega d_j^\dagger d_j + (\overline{\omega}-\omega)d_j^{\dagger 2}d_j^2 + d_j^2 d_j^{\dagger 2},\\ &= 1 + (\omega-1)d_j^\dagger d_j + (\omega-1)\omega d_j^{\dagger 2}d_j^2,\\ &= \omega^{O_j},\end{aligned}$$

where we used (10.1.5) in the second-to-last step, and (10.3.2) in the last step. This is precisely the expected result, keeping in mind (9.3.4). Moreover, the field terms now become

$$-f\sum_{j=1}^{L}\left(e^{i\varphi}\overline{\omega}\chi_j^\dagger\psi_j + \text{H.c.}\right) = -f\sum_{j=1}^{L}\left(e^{i\varphi}\omega^{O_j} + \text{H.c.}\right) = -2f\sum_{j=1}^{L-1}\cos\left(\varphi + \frac{2\pi}{3}O_j\right).$$

For the interaction terms,

$$\overline{\omega}\psi_j^\dagger\chi_{j+1} = \left[d_j^\dagger + \omega d_j^2 + (\overline{\omega}-1)d_j^{\dagger 2}d_j\right]\left(d_{j+1} + d_{j+1}^{\dagger 2}\right).$$

In total, (10.3.1) becomes

$$H_3 = -2f\sum_{j=1}^{L-1}\cos\left(\varphi + \frac{2\pi}{3}O_j\right) - J\sum_{j=1}^{L-1}\left\{e^{i\vartheta}\left[d_j^\dagger + \omega d_j^2 + (\overline{\omega}-1)d_j^{\dagger 2}d_j\right]\left(d_{j+1} + d_{j+1}^{\dagger 2}\right) + \text{H.c.}\right\}.$$

[9]Notice that this is different from (129) in [22]. Indeed, it seems the factor ω^{m-1} in the sum of (126) was forgotten (one can also convince oneself that (129) is false by writing out the matrices, ignoring the strings).

To compare this later with the Fock parafermion tight-binding model of [74], Equation (10), we split off the 'hopping' term. Then

$$\begin{aligned}H_3 = &-J\sum_{j=1}^{L-1}\left(e^{i\vartheta}d_j^\dagger d_{j+1} + e^{-i\vartheta}d_{j+1}^\dagger d_j\right)\\ &- J\sum_{j=1}^{L-1}\left\{e^{i\vartheta}\left[d_j^\dagger d_{j+1}^{\dagger 2} + \omega d_j^2 d_{j+1} + \omega d_j^2 d_{j+2}^{\dagger 2} + (\overline{\omega}-1)d_j^{\dagger 2}d_j d_{j+1} + (\overline{\omega}-1)d_j^{\dagger 2}d_j d_{j+1}^{\dagger 2}\right] + \text{H.c.}\right\}\\ &- 2f\sum_{j=1}^{L-1}\cos\left(\varphi + \frac{2\pi}{3}O_j\right).\end{aligned}$$

To recover the Potts Model, one sets all coefficients equal, that is $\vartheta = \varphi = 0$.

Let us move on to the case $n = 4$, but this time we start in the Potts scenario with all coefficients equal. In this case this also means that they are all real, as the second coefficient is always such. We absorb these coefficients into f and J without loss of generality. Then from (9.3.5) with $n = 4$, we have

$$\begin{aligned}H_{4P} &= -f\sum_{j=1}^{l}\left[i^{-3/2}\chi_j^\dagger\psi_j + i^{-2}\chi_j^2\psi_j^2 + i^{-3/2}\chi_j\psi_j^\dagger\right] - J\sum_{j=1}^{L-1}\left[i^{-3/2}\psi_j^\dagger\chi_{j+1} + i^{-2}\psi_j^2\chi_{j+1}^2 + i^{-3/2}\psi_j\chi_{j+1}^\dagger\right],\\ &= f\sum_{j=1}^{l}\left\{\left[e^{i\pi/4}\chi_j^\dagger\psi_j + \text{H.c.}\right] + \psi_j^2\chi_j^2\right\} + J\sum_{j=1}^{L-1}\left\{\left[e^{i\pi/4}\psi_j^\dagger\chi_{j+1} + \text{H.c.}\right] + \psi_j^2\chi_{j+1}^2\right\}.\end{aligned}$$

Now, the relations between parafermions and Fock parafermions become

$$\chi_j = d_j + d_j^{\dagger 3},\ \psi_j = \left(-e^{i\pi/4}\right)\left[id_j + i(i-1)d_j^\dagger d_j^2 - (i-1)d_j^{\dagger 2}d_j^3 + d_j^{\dagger 3}\right].$$

With these relations one can compute that

$$\chi_j^2 = d_j^2 + d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j,\ \psi_j^2 = -d_j^2 + 2d_j^\dagger d_j^3 - d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j.$$

We first transform the field terms.

$$\begin{aligned}\psi_j^2\chi_j^2 &= \left(-d_j^2 + 2d_j^\dagger d_j^3 - d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j\right)\left(d_j^2 + d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j\right),\\ &= -d_j^3 d_j^{\dagger 3} - d_j^2 d_j^{\dagger 3}d_j + 2d_j^\dagger d_j^3 d_j^{\dagger 3}d_j - d_j d_j^{\dagger 3}d_j^2 + d_j^{\dagger 3}d_j d_j^2,\\ &= -(1 - d_j^\dagger d_j) - (d_j^2 d_j^{\dagger 2})(d_j^\dagger d_j) + 2d_j^\dagger(1 - d_j^\dagger d_j)d_j - (d_j d_j^\dagger)(d_j^{\dagger 2}d_j^2) + d_j^{\dagger 3}d_j^3,\\ &= -1 + 3d_j^\dagger d_j - 2d_j^{\dagger 2}d_j^2 + d_j^{\dagger 3}d_j^3 - (1 - d_j^{\dagger 2}d_j^2)(d_j^\dagger d_j) - (1 - d_j^{\dagger 3}d_j^3)(d_j^{\dagger 2}d_j^2),\\ &= -1 + 2d_j^\dagger d_j - 3d_j^{\dagger 2}d_j^2 + d_j^{\dagger 3}d_j^3 + d_j^{\dagger 2}d_j^2 d_j^\dagger d_j + d_j^{\dagger 3}d_j^3 d_j^{\dagger 2}d_j^2,\\ &= -1 + 2d_j^\dagger d_j - 3d_j^{\dagger 2}d_j^2 + d_j^{\dagger 3}d_j^3 + d_j^{\dagger 2}d_j^2 + d_j^{\dagger 3}d_j^3,\\ &= -\left[1 - 2\left(d_j^\dagger d_j - d_j^{\dagger 2}d_j^2 + d_j^{\dagger 3}d_j^3\right)\right],\\ &= (-1)^{O_j}.\end{aligned}$$

The last step follows from (10.2.4) for $n = 4$, either by using the same proof, or taking its square.

The other term is

$$\begin{aligned}
-\left(e^{i\pi/4}\right)\chi_j^\dagger\psi_j &= \left(d_j^3+d_j^\dagger\right)\left[id_j+i(i-1)d_j^\dagger d_j^2-(i-1)d_j^{\dagger 2}d_j^3+d_j^{\dagger 3}\right],\\
&= d_j^3d_j^{\dagger 3}+id_j^\dagger d_j+i(i-1)d_j^{\dagger 2}d_j^2-(i-1)d_j^{\dagger 3}d_j^3,\\
&= 1-d_j^\dagger d_j+id_j^\dagger d_j+i(i-1)d_j^{\dagger 2}d_j^2-(i-1)d_j^{\dagger 3}d_j^3,\\
&= 1+(i-1)\left[d_j^\dagger d_j+id_j^{\dagger 2}d_j^2-d_j^{\dagger 3}d_j^3\right],\\
&= i^{O_j}.
\end{aligned}$$

Thus the field terms are

$$f\sum_{j=1}^{l}\left\{\left[e^{i\pi/4}\chi_j^\dagger\psi_j+\text{H.c.}\right]+\psi_j^2\chi_j^2\right\}=-f\sum_{j=1}^{l}\left\{i^{O_j}+(-1)^{O_j}+(-i)^{O_j}\right\},$$

which can also be written without phase factors as

$$f\sum_{j=1}^{l}\left\{\left[e^{i\pi/4}\chi_j^\dagger\psi_j+\text{H.c.}\right]+\psi_j^2\chi_j^2\right\}=-f\sum_{j=1}^{l}\left\{3-4d_j^\dagger d_j\right\},$$

In the same way, the interaction terms can be seen to be equal to

$$\begin{aligned}
&J\sum_{j=1}^{L}\left(-d_j^2+2d_j^\dagger d_j^3-d_jd_j^{\dagger 3}+d_j^{\dagger 3}d_j\right)\left(d_{j+1}^2+d_{j+1}d_{j+1}^{\dagger 3}d_{j+1}^{\dagger 3}d_{j+1}\right)\\
&-J\sum_{j=1}^{L-1}\left[\left(id_j+i(i-1)d_j^\dagger d_j^2-(i-1)d_j^{\dagger 2}d_j^3+d_j^{\dagger 3}\right)\left(d_{j+1}^\dagger+d_{j+1}^3\right)+\text{H.c.}\right]
\end{aligned}$$

Combining everything and isolating the tight-binding terms yields

$$\begin{aligned}
H_{4P} = &-J\sum_{j=1}^{L-1}\left[d_{j+1}^\dagger d_j+d_j^\dagger d_{j+1}\right]\\
&-J\sum_{j=1}^{L-1}\Big[id_jd_{j+1}^3+i(i-1)d_j^\dagger d_j^2d_{j+1}^\dagger d_{j+1}^3+i(i-1)d_j^\dagger d_j^2d_{j+1}^3-(i-1)d_j^{\dagger 2}d_j^3d_{j+1}^\dagger\\
&\qquad -(i-1)d_j^{\dagger 2}d_j^3d_{j+1}^3+d_j^{\dagger 3}d_{j+1}^\dagger+d_j^{\dagger 3}d_{j+1}^3+\text{H.c.}\Big]\\
&-f\sum_{j=1}^{l}\left\{i^{O_j}+(-1)^{O_j}+(-i)^{O_j}\right\}\ .
\end{aligned}$$

11 Potts-Like Fock Parafermion Models

11.1 The Model and our Goals

In this chapter we will probe the physics of a special case of the general family of Hamiltonians (9.3.5), namely $H_4(f = 0)$ (and $J > 0$). The choice of $n = 4$ is motivated by the fact that the local Hilbert space then has the same dimension as that of spinful Fermions, that is electrons. Our philosophy will be that since these have been far more thoroughly studied than parafermions, as they are fundamental particles, we should try to find a relation between the two algebras and use this to map the parafermionic Hamiltonian to an electronic one. For this, we will follow the procedure outlined in [18] by Calzona, Meng, Sassetti, and Schmidt, and reproduce and expand upon their results. If there is no field ($f = 0$), the general $\mathbb{Z}_4$ invariant Hamiltonian in (9.3.5) reads

$$H_4 = -J \sum_{j=1}^{L-1} \left(\hat{\gamma}_1 \omega^{-3/2} \psi_j^\dagger \chi_{j+1} + \hat{\gamma}_2 \omega^{-2} \psi_j^2 \chi_{j+1}^2 + \hat{\gamma}_1 \omega^{-3/2} \psi_j \chi_{j+1}^\dagger \right),$$

with $\omega = e^{i\pi/2} = i$ and $\hat{\gamma}_1, \hat{\gamma}_2$ both real for convenience. As

$$\omega^{-3/2} = -e^{i\pi/4},\ \omega^{-2} = -1,\ \psi_j^\dagger \chi_{j+1} = e^{i3\pi/2} \chi_{j+1} \psi_j^\dagger,$$

this becomes

$$\begin{aligned} H_4 &= -J \sum_{j=1}^{L-1} \left(-\hat{\gamma}_1 e^{-i\pi/4} \chi_{j+1} \psi_j^\dagger - \hat{\gamma}_2 \chi_{j+1}^2 \psi_j^2 - \hat{\gamma}_1 e^{i\pi/4} \psi_j \chi_{j+1}^\dagger \right), \\ &= \hat{\gamma}_1 \left[J e^{i\pi/4} \sum_{j=1}^{L-1} \psi_j \chi_{j+1}^\dagger + \text{ H.c.} \right] + \hat{\gamma}_2 \left[J \sum_{j=1}^{L-1} \psi_j^2 \chi_{j+1}^2 \right], \\ &= \hat{\gamma}_1 H_A + \hat{\gamma}_2 H_B \end{aligned}$$

We may set $\hat{\gamma}_1 = 1$ without loss of generality. Hence we arrive at

$$H_4 = H_A + \hat{\gamma}_2 H_B. \tag{11.1.1}$$

To relate this to Equation (3) in [18], one sets $\hat{\gamma}_2 = 0$ and identifies $a_{j+1} = \chi_{j+1}$, $b_j = -\psi_j$, the minus sign arising from a difference in phase convention (compare Equation (1)[10] in [18] to (10.2.1)). Correspondingly, we will often consider the case $\hat{\gamma}_2 = 0$.

In this chapter, we will first determine explicit expressions for the groundstates of (11.1.1). To transform it into an electronic Hamiltonian, we will then proceed along the following steps. First, we use the Fock parafermion technology introduced in the previous chapter to relate the local (i.e. single-site) Hilbert space to an electronic one. This will simply amount to a (cleverly chosen) matching of bases. Then, we introduce the appropriate Jordan-Wigner strings to convert the commutation relations. At this point we will confront the main danger of this procedure, namely that the electronic Hamiltonian resulting from the mapping may be non-local, and show that our cleverness in the previous step prevents this from happening. Finally, we apply the mapping and investigate some properties of the resulting electronic model. We end the chapter by computing the fermionic spectral functions of H_A and a deformation of it, to see how the parafermionic edge modes show up in the electronic models.

[10] There is a typo in this equation, it should instead read $b_j = e^{-i\pi/4}(d_j i^{N_j} + d_j^{\dagger 3})$

11.2 Groundstates and Lindblad Operators

To determine the groundstates of (11.1.1), we will use the following approach. Let O be an operator, then for any vector $|\psi\rangle$ it holds that

$$\langle\psi| O^\dagger O |\psi\rangle = \|O\psi\|^2 \geq 0.$$

In particular, if $|\varphi\rangle$ is an eigenvector of $O^\dagger O$ with eigenvalue λ, then

$$\lambda = \langle\varphi| \lambda |\varphi\rangle = \langle\varphi| O^\dagger O |\varphi\rangle \geq 0.$$

It follows that the spectrum of $O^\dagger O$ is entirely contained in $[0, \infty)$, and so people call it a *positive* operator. One can prove (using spectral theory) that every positive operator is of this form. For closed quantum systems, any physically reasonable Hamiltonian has a spectrum that is real and bounded from below. Thus, shifting by a constant if necessary, we can assume that the Hamiltonian is a positive operator. Hence there must exist an operator h and a constant h_0 such that $H = h^\dagger h - h_0$. If the kernel of $h^\dagger h$ is non-empty, it is precisely the groundstate manifold of H. Suppose that $|\vartheta\rangle$ is an element of this kernel, then

$$0 = \langle\vartheta| h^\dagger h |\vartheta\rangle = \|h\vartheta\|^2,$$

and so it must be true that $h|\vartheta\rangle = 0$. This means that to find the groundstates we may 'factor' H, and instead determine the kernel of the operator h. The hope then is that h can be determined explicitly and that the equation $h|\vartheta\rangle = 0$ is simple to solve. We now apply this technique to our Hamiltonian H_4.

We set $\hat{\gamma}_2 = 0$ for now. Introduce the *Lindblad operators*

$$\mathcal{L}_j = e^{i\pi/4}\psi_j + i\chi_{j+1}, \tag{11.2.1}$$

then

$$\begin{aligned}
\mathcal{L}_j^\dagger \mathcal{L}_j &= \left(e^{-i\pi/4}\psi_j^\dagger - i\chi_{j+1}^\dagger\right)\left(e^{i\pi/4}\psi_j + i\chi_{j+1}\right), \\
&= 2 + \left(ie^{-i\pi/4}\psi_j^\dagger\chi_{j+1} - ie^{i\pi/4}\chi_{j+1}^\dagger\psi_j\right), \\
&= 2 + \left(ie^{-i\pi/4}(-i)\chi_{j+1}\psi_j^\dagger - ie^{i\pi/4}(i)\psi_j\chi_{j+1}^\dagger\right), \\
&= 2 + \left(e^{i\pi/4}\psi_j\chi_{j+1}^\dagger + \text{ H.c.}\right),
\end{aligned}$$

so that

$$H_A = J\sum_{j=1}^{L-1}\left(\mathcal{L}_j^\dagger\mathcal{L}_j - 2\right) = -2J(L-1) + J\sum_{j=1}^{L-1}\mathcal{L}_j^\dagger\mathcal{L}_j.$$

We can now find the groundstates of H_A by finding the elements in the intersection of the kernels of the $\mathcal{L}_j$. That is to say we want to find a state $|\psi\rangle$ such that $\mathcal{L}_j|\psi\rangle = 0$ for all j. To find these, we draw inspiration from the supplemental material to [46], Equations (S5) to (S8). We introduce Fock parafermions into (11.2.1) according to (10.2.1) and neglect overall constants (as we are only interested in the kernel), then

$$\mathcal{L}_j \propto d_j i^{O_j} + d_j^{\dagger 3} - id_{j+1} - id_{j+1}^{\dagger 3}. \tag{11.2.2}$$

This acts only on the sites j and $j+1$, so that we can write down the matrix of this operator in the $|o_j, o_{j+1}\rangle$ subspace. If we look back at the action of the Fock parafermion operators on such basis

elements given by equations (10.1.2) and (10.1.3), we see that we will end up with some phase factors involving the sites to the left of j. These are of no importance due to the following argument. Denote the phase factor of d_j by

$$\Phi_j := \exp\left(i\frac{\pi}{2}\sum_{k<j} o_k\right),$$

then the two annihilation operators in (11.2.2) contribute a common phase factor of Φ_j, and the two creation operators carry Φ_j^{-3}. Now notice that Φ_j is some rational power of i, so that $\Phi_j = \Phi_j^{-3}$, as $i = (-i)^3 = i^{-3}$. It follows that we can completely factor the string out of the matrix for $\mathcal{L}_j$, and hence we neglect it in the sequel.

It is now a straightforward computation to arrive at the following matrix, where we have recorded the image of each basis element (modulo common factors as explained above),

$$\left(\begin{array}{llll}
|0,0\rangle & : & |3,0\rangle - i\,|0,3\rangle\,, & |1,0\rangle \;:\; i\,|0,0\rangle + |1,3\rangle\,, \\
|0,1\rangle & : & |3,1\rangle - i\,|0,0\rangle\,, & |1,1\rangle \;:\; i\,|0,1\rangle + |1,0\rangle\,, \\
|0,2\rangle & : & |3,2\rangle - i\,|0,1\rangle\,, & |1,2\rangle \;:\; i\,|0,2\rangle + |1,1\rangle\,, \\
|0,3\rangle & : & |3,3\rangle - i\,|0,2\rangle\,, & |1,3\rangle \;:\; i\,|0,3\rangle + |1,2\rangle\,, \\
 & & & \\
|2,0\rangle & : & -|1,0\rangle + i\,|2,3\rangle\,, & |3,0\rangle \;:\; -i\,|2,0\rangle - |3,3\rangle\,, \\
|2,1\rangle & : & -|1,1\rangle + i\,|2,0\rangle\,, & |3,1\rangle \;:\; -i\,|2,1\rangle - |3,0\rangle\,, \\
|2,2\rangle & : & -|1,2\rangle + i\,|2,1\rangle\,, & |3,2\rangle \;:\; -i\,|2,2\rangle - |3,1\rangle\,, \\
|2,3\rangle & : & -|1,3\rangle + i\,|2,2\rangle\,, & |3,3\rangle \;:\; -i\,|2,3\rangle - |3,2\rangle\,,
\end{array}\right).$$

From this one can read off the kernel as follows. Notice that each basis element is mapped to a linear combination of two other basis elements. If one now focuses on the outputs, one sees that each basis element occurs exactly twice and with opposite signs. For example, $|3,0\rangle$ occurs with a plus sign in the image of $|0,0\rangle$, and with a minus sign in the output of $|3,1\rangle$. Additionally we remark that all 'blocks' of the above matrix are (indirectly) connected. As an example, in the image of $|0,0\rangle$ we find $|3,0\rangle$, in whose image is $|2,0\rangle$, which connects to $|1,0\rangle$, and this returns us to $|0,0\rangle$. It now follows that we have the four orthonormal null-vectors

$$\begin{aligned}
|z=0\rangle &= \frac{1}{2}\left(|0,0\rangle + |3,1\rangle + |1,3\rangle + |2,2\rangle\right), \\
|z=1\rangle &= \frac{1}{2}\left(|0,1\rangle + |3,2\rangle + |1,0\rangle + |2,3\rangle\right), \\
|z=2\rangle &= \frac{1}{2}\left(|0,2\rangle + |3,3\rangle + |1,1\rangle + |2,0\rangle\right), \\
|z=3\rangle &= \frac{1}{2}\left(|0,3\rangle + |3,0\rangle + |1,2\rangle + |2,1\rangle\right).
\end{aligned}$$

The notation $|z=m\rangle$ comes from the fact that these states have different charges under $\mathbb{Z}_4$. This organisation into $\mathbb{Z}_4$ sectors was to be expected, since H_A has $\mathbb{Z}_4$ symmetry by construction. Notice that in all four cases, all coefficients are the same. Moreover, one can trivially check numerically that the operator $\mathcal{L}_j^\dagger\mathcal{L}_j$ indeed has precisely a four-dimensional kernel, and that the gap to the first excited states is $2J$.

To extend this to the intersection of all the kernels of the $\mathcal{L}_j$, one can form the following four states,

$$|\mathrm{GS}_m\rangle = 2^{1-L}\sum_{\substack{\{o_j\}\text{ such that}\\ \sum_j o_j = m \pmod 4}}\bigotimes_{j=1}^{L}|o_j\rangle\,. \tag{11.2.3}$$

This respects both the separation into $\mathbb{Z}_4$ sectors and the $\mathcal{L}_j$ kernels. To see the latter, suppose that $\sum_j o_j = m \pmod 4$ and that $|o_k, o_{k+1}\rangle = |0,0\rangle$ (for concreteness). Then as $0+0$, $3+1$, $2+2$, and $1+3$ are all equal modulo 4, we can replace $|0,0\rangle$ by any one of $|3,1\rangle$, $|2,2\rangle$, or $|1,3\rangle$. Hence those sequences of o_j's are also included in the sum defining $|\mathrm{GS}_m\rangle$, all appearing with equal coefficients, and so we have that $\mathcal{L}_k |\mathrm{GS}_m\rangle = 0$ upon comparison with the local null-vectors $|z = m\rangle_k$.

Can we still write the full Hamiltonian

$$H_4 = H_A + \hat{\gamma}_2 H_B$$

in terms of a single family of Lindblad operators? A simple counting argument already casts doubt on this. If we include the terms χ_{j+1}, ψ_j, and their squares in an operator, its norm-squared operator contains 16 terms, 8 of which are order 0 mod 4. However, there are then 8 terms of orders 1 and 3, which all need to cancel (as the Hamiltonian conserves the $\mathbb{Z}_4$ charge). As we only have 8 (real) coefficients to choose, it seems unlikely that we can enforce the required cancellations and also arrange for the order 0 terms to have the correct coefficients. Indeed, the Anstatz

$$\xi_j = \mathcal{L}_j + \alpha\chi_{j+1}^2 + \beta\psi_j^2$$

leads to

$$\begin{aligned}\xi_j^\dagger \xi_j = 2 + |\alpha|^2 + |\beta|^2 + \left(e^{i\pi/4}\psi_j\chi_{j+1}^\dagger + \text{ H.c.}\right) + (\alpha^*\beta + \beta^*\alpha)\chi_{j+1}^2\psi_j^2 \\ + \left(i\alpha\chi_{j+1} - e^{-i\pi/4}\beta\psi_j + \text{ H.c.}\right) + \left(i\beta\chi_{j+1}^\dagger\psi_j^2 - e^{-i\pi/4}\alpha\psi_j^\dagger\chi_{j+1}^2 + \text{ H.c.}\right),\end{aligned}$$

where it does not seem possible to choose α and β appropriately. With two sets of Lindblad operators the problem is trivial, namely take

$$\mathcal{K}_j = \frac{1}{2}\left(\chi_{j+1}^2 + \psi_j^2\right).$$

This allows us to write

$$H_4 = -2J(L-1)(1+\hat{\gamma}_2/2) + J\sum_{j=1}^{L-1}\left(\mathcal{L}_j^\dagger\mathcal{L}_j + \hat{\gamma}_2\mathcal{K}_j^\dagger\mathcal{K}_j\right).$$

Sadly these operators don't seem to have any nice relations, as

$$[\mathcal{K}_j, \mathcal{L}_j] = -e^{i\pi/4}\chi_{j+1}^2\psi_j - i\psi_j^2\chi_{j+1}, \ \{\mathcal{K}_j, \mathcal{L}_j\} = -e^{i\pi/4}\psi_j^3 - i\chi_{j+1}^3.$$

However, all is not lost. We can still use the operators $\mathcal{K}_j$ to determine the ground states in the case that $\hat{\gamma}_2 > 0$. For this introduce again Fock parafermions according to (10.2.1) to find

$$\begin{aligned}\chi_{j+1}^2 &= \left(d_{j+1} + d_{j+1}^{\dagger 3}\right)^2 = d_{j+1} + d_{j+1}d_{j+1}^{\dagger 3} + d_{j+1}^{\dagger 3}d_j, \\ \psi_j^2 &= -i\left(d_j i^{O_j} + d_j^{\dagger 3}\right)^2 = -d_j(-1)^{O_j+1}d_j - d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j,\end{aligned}$$

so that

$$\mathcal{K}_j \propto -d_j(-1)^{O_j+1}d_j - d_j d_j^{\dagger 3} + d_j^{\dagger 3}d_j + d_{j+1} + d_{j+1}d_{j+1}^{\dagger 3} + d_{j+1}^{\dagger 3}d_j.$$

By the same arguments as before, we can restrict our attention to the $|o_j, o_{j+1}\rangle$ subspace, where we find the matrix

$$\left(\begin{array}{llll}
|0,0\rangle & : & -|2,0\rangle + |0,2\rangle\,, & |1,0\rangle \;:\; |3,0\rangle - |1,2\rangle\,, \\
|0,1\rangle & : & -|2,1\rangle + |0,3\rangle\,, & |1,1\rangle \;:\; |3,1\rangle - |1,3\rangle\,, \\
|0,2\rangle & : & -|2,2\rangle + |0,0\rangle\,, & |1,2\rangle \;:\; |3,2\rangle - |1,0\rangle\,, \\
|0,3\rangle & : & -|2,3\rangle + |0,1\rangle\,, & |1,3\rangle \;:\; |3,3\rangle - |1,1\rangle\,, \\
 & & & \\
|2,0\rangle & : & -|0,0\rangle + |2,2\rangle\,, & |3,0\rangle \;:\; |1,0\rangle - |3,2\rangle\,, \\
|2,1\rangle & : & -|0,1\rangle + |2,3\rangle\,, & |3,1\rangle \;:\; |1,1\rangle - |3,3\rangle\,, \\
|2,2\rangle & : & -|0,2\rangle + |2,0\rangle\,, & |3,2\rangle \;:\; |1,2\rangle - |3,0\rangle\,, \\
|2,3\rangle & : & -|0,3\rangle + |2,1\rangle\,, & |3,3\rangle \;:\; |1,3\rangle - |3,1\rangle\,,
\end{array}\right).$$

Like the $\hat{\gamma}_2 = 0$ case, we neglected overall phase factors (as $i^{-2} = -1 = i^2$). For this matrix, the 'blocks' are not all connected. Namely the left and right 'columns' do not talk to each other, but have the same action (modulo an overall minus sign) on their invariant subspaces. Accordingly, $\mathcal{K}_j$ has a higher-dimensional kernel than $\mathcal{L}_j$. One can check by inspection that the kernel of $\mathcal{K}_j$ is spanned by the eight elements

$$\begin{aligned}
&|0,0\rangle + |2,2\rangle\,,\ |0,1\rangle + |2,3\rangle\,,\ |0,2\rangle + |2,0\rangle\,,\ |0,3\rangle + |2,1\rangle\,, \\
&|1,3\rangle + |3,1\rangle\,,\ |1,0\rangle + |3,2\rangle\,,\ |1,1\rangle + |3,3\rangle\,,\ |1,2\rangle + |3,0\rangle\,.
\end{aligned}$$

Luckily, the states $|z = m\rangle$ discussed above can be formed using these basis elements, by adding the above vectors vertically in pairs. It follows that the kernel of $\mathcal{L}_j$ is contained in that of $\mathcal{K}_j$, and so that H_4 has the same groundstates for all $\hat{\gamma}_2 \geq 0$.

We would of course like to know if this model is gapped away from the Potts-point. In the mathematical physics literature ([5, 33, 50, 59]), there are results and criteria to show the existence of a gap in TDL, for systems that exhibit similar behaviour. Despite a significant time investment, we could not find a straightforward way to adept their techniques.

11.3 The Electronic Representation and Locality

We now describe the parafermion to Fermion mapping constructed in [18]. Consider a single-site system, then the Hilbert space is spanned by $\{|0\rangle\,, |1\rangle\,, |2\rangle\,, |3\rangle\}$ in the Fock parafermion picture. For a single-site electronic system, we also have a four-dimensional Hilbert space, now spanned by the states $\{|\varnothing\rangle\,, |\uparrow\rangle\,, |\downarrow\rangle\,, |\uparrow\downarrow\rangle\}$. The action of the Fock parafermion annihilation operator d is $d\,|n\rangle = |n-1\rangle$ for $3 \geq n \geq 1$ and $d\,|0\rangle = 0$. Since one can write all four electronic states as some number (zero included) of creation and annihilation operators acting on the electronic vacuum $|\varnothing\rangle$, it is possible to construct an operator on the electronic Fock space with the same action as d. The only required input is the designation of four orthonormal states to take the roles of the $|n\rangle$. As there are many choices for this, it pays off to come up with some extra conditions to impose.

The parity of Fock parafermions is described by $\mathbb{Z}_4$, whereas Fermion parity is a $\mathbb{Z}_2$ representation. One can view $\mathbb{Z}_2$ as a subgroup of $\mathbb{Z}_4$ in a natural way as $\{0, 2\} \subset \mathbb{Z}_4$. As d has parafermion parity one, it then makes sense to construct an electronic version in such a way that it has Fermion parity one. In this way, the resulting transformed Hamiltonians will preserve Fermion parity. This will also be favourable for the Jordan-Wigner strings, as the minus signs generated by Fermion commutation can then be cancelled by phase factors of the form $i^{2(\cdots)} = (-1)^{\cdots}$, which is useful as we will need to include powers of i in any case to obtain the Fock parafermion algebra. We are also free to multiply each electronic basis states by a phase, which we will use to ensure that each power of i occurs exactly once in the definition of the electronic version of d.

An identification that satisfies all of the above requirements is

$$\begin{aligned}
|0\rangle &\mapsto |f_0\rangle := |\varnothing\rangle\,, \\
|1\rangle &\mapsto |f_1\rangle := c_\uparrow^\dagger\,|\varnothing\rangle\,, \\
|2\rangle &\mapsto |f_2\rangle := ic_\uparrow^\dagger c_\downarrow^\dagger\,|\varnothing\rangle\,, \\
|3\rangle &\mapsto |f_3\rangle := -ic_\downarrow^\dagger\,|\varnothing\rangle\,.
\end{aligned}$$

Denote by $P(f_n)$ the projection operator onto the subspace spanned by $|f_n\rangle$. These have the usual representations using number operators, namely

$$\begin{aligned}
P(f_0) &= (1-n_\uparrow)(1-n_\downarrow), \\
P(f_1) &= n_\uparrow(1-n_\downarrow), \\
P(f_2) &= n_\uparrow n_\downarrow, \\
P(f_3) &= (1-n_\uparrow)n_\downarrow.
\end{aligned}$$

Using these projectors, we can now construct

$$\begin{aligned}
d &= i\cdot 0\cdot P(f_0) + 1\cdot c_\uparrow\cdot P(f_1) + (-i)\cdot c_\downarrow\cdot P(f_2) + (-1)\cdot c_\uparrow^\dagger\cdot P(f_3), \\
&= c_\uparrow(1-n_\downarrow)n_\uparrow - ic_\downarrow n_\uparrow n_\downarrow - c_\uparrow^\dagger(1-n_\uparrow)n_\downarrow, \\
&= c_\downarrow^\dagger c_\uparrow c_\downarrow - c_\uparrow^\dagger c_\downarrow^\dagger c_\downarrow + ic_\uparrow^\dagger c_\uparrow c_\downarrow + c_\uparrow, \\
&= c_\uparrow - c_\uparrow n_\downarrow - c_\uparrow^\dagger + ic_\downarrow n_\uparrow.
\end{aligned}$$

The next step is to consider an L-site system, with the sites indexed by j. The Jordan-Wigner strings to induce Fermionic commutation behaviour are well known, and are given by

$$\begin{aligned}
c_{j\sigma} &= (-1)^{\sum_{\sigma,k<j} n_k}\; I\otimes\cdots\otimes I\otimes c_\sigma\otimes I\otimes\cdots\otimes I, \\
&= i^{2\sum_{\sigma,k<j} n_k}\; I\otimes\cdots\otimes I\otimes c_\sigma\otimes I\otimes\cdots\otimes I.
\end{aligned}$$

For notational convenience we introduce the shorthand notation

$$D_j = c_{j\uparrow} - c_{j\uparrow}n_{j\downarrow} - c_{j\uparrow}^\dagger + ic_{j\downarrow}n_{j\uparrow}. \tag{11.3.1}$$

The Fock parafermionic Jordan-Wigner strings can be read off from equation (10.1.3) and give us that

$$d_j = i^{\sum_{k<j} O_k} D_j.$$

One can now merge the previous two relations to obtain

$$d_j = i^{\sum_{k<j}\left(O_k + 2n_{k\uparrow} + 2n_{k\downarrow}\right)} D_j.$$

This still involves the Fock parafermion number operators O_k. Again we can find the representation

$$O_k = n_{k\uparrow} - 2n_{k\uparrow}n_{k\downarrow} + 3n_{k\downarrow}. \tag{11.3.2}$$

This is the final ingredient that allows us to relate Fock parafermions and spinful Fermions, with the result being

$$\begin{aligned}
d_j &= i^{\sum_{k<j}\left(n_{k\downarrow} - 2n_{k\uparrow}n_{k\downarrow} - n_{k\uparrow}\right)} D_j, \\
&= i^{\sum_{k<j}\left(n_{k\downarrow} - 2n_{k\uparrow}n_{k\downarrow} - n_{k\uparrow}\right)} \left(c_{j\uparrow} - c_{j\uparrow}n_{j\downarrow} - c_{j\uparrow}^\dagger + ic_{j\downarrow}n_{j\uparrow}\right).
\end{aligned} \tag{11.3.3}$$

To relate the spinful Fermions to the parafermions in which (11.1.1) is written, we again use the relation (10.2.1), yielding

$$H_A = -J\sum_{j=1}^{L-1}\left(d_j i^{O_j} d_{j+1}^\dagger + d_j i^{O_j} d_{j+1}^3 + d_j^{\dagger 3} d_{j+1}^\dagger + d_j^{\dagger 3} d_{j+1}^3\right) + \text{ H.c..} \tag{11.3.4}$$

From (11.3.3) one can immediately find an expression for $d_{j+1}^\dagger$ in terms of spinful Fermions, but we also need more complicated operators, such as d_{j+1}^3. For these, it is simpler to construct them using the same reasoning that went into (11.3.3), rather than taking the third power directly. For instance,

$$\begin{aligned} d_{j+1}^3 &= i^{3\sum_{k<j+1}(O_k+2n_{k\uparrow}+2n_{k\downarrow})} D_{j+1}^3, \\ &= i^{3\sum_{k<j}(O_k+2n_{k\uparrow}+2n_{k\downarrow})}\ i^{3O_j+2n_{j\uparrow}+2n_{j\downarrow}} D_{j+1}^3. \end{aligned}$$

The operator D_{j+1}^3 is characterised by the fact that $D_{j+1}^3 \left|n\right\rangle_{j+1} = 0$ for $0 \le n \le 2$, and $D_{j+1}^3 \left|3\right\rangle_{j+1} = \left|0\right\rangle_{j+1}$. Looking back out our mapping from $\left|n\right\rangle$ to $\left|f_n\right\rangle$, we conclude that

$$\begin{aligned} d_{j+1}^3 &= i^{3\sum_{k<j}(O_k+2n_{k\uparrow}+2n_{k\downarrow})}\ i^{3O_j+2n_{j\uparrow}+2n_{j\downarrow}} D_{j+1}^3, \\ &= i^{3\sum_{k<j}(O_k+2n_{k\uparrow}+2n_{k\downarrow})}\ i^{3O_j+2n_{j\uparrow}+2n_{j\downarrow}} (1-n_{j+1\uparrow}) i c_{j+1\downarrow}, \end{aligned}$$

as

$$\begin{aligned} (1-n_{j+1\uparrow}) i c_{j+1\downarrow} \left|\varnothing\right\rangle_{j+1} &= 0, \\ (1-n_{j+1\uparrow}) i c_{j+1\downarrow} c_{j+1\uparrow}^\dagger \left|\varnothing\right\rangle_{j+1} &= 0, \\ (1-n_{j+1\uparrow}) i c_{j+1\downarrow} i c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger \left|\varnothing\right\rangle_{j+1} = (1-n_{j+1\uparrow}) c_{j+1\uparrow}^\dagger \left|\varnothing\right\rangle_{j+1} &= 0, \\ (1-n_{j+1\uparrow}) i c_{j+1\downarrow} (-i) c_{j+1\downarrow}^\dagger \left|\varnothing\right\rangle = \left|\varnothing\right\rangle_{j+1} &= \left|f_0\right\rangle_{j+1}. \end{aligned}$$

Similarly,

$$d_j^{\dagger 3} = i^{-3\sum_{k<j}(O_k+2n_{k\uparrow}+2n_{k\downarrow})} (1-n_{j\uparrow})(-i c_{j\downarrow}^\dagger).$$

Before we start to use all this fancy new technology, there is a potential problem that we should address. Since we are using Jordan-Wigner strings in the transformation (11.3.3), it is non-local. Thus there is a danger that when we transform (11.3.4), we will end up with an electronic model that contains non-local strings. We would very much like to avoid that situation, especially with an eye on experimental realisability of the result. Thankfully, the condition we imposed earlier on the possible identifications of bases will come to our rescue. Let D denote some local (nearest-neighbour) Fock parafermionic operator that respects $\mathbb{Z}_4$, in other words it preserves the number of Fock parafermions modulo four (as all terms in our Hamiltonian do). Generically, this can be written in the form

$$D = d_j^{a_1} d_j^{\dagger a_2} d_{j+1}^{b_1} d_{j+1}^{\dagger b_2},$$

with $0 \le a_i, b_j \le 3$, satisfying $a_1 - a_2 + b_1 - b_2 = 0 \pmod 4$. When we apply the mapping (11.3.3) to this, the result is a string factor depending on all sites to the left of j and some Fermionic operator C that is local (depends only on operators associated to sites j and $j+1$). To be more precise, we find

$$D = i^{(a_1-a_2+b_1-b_2)\sum_{k<j}(\dots)} C = 1^{\sum_{k<j}(\dots)} C = C,$$

since the first factor in the phase is congruent to zero modulo four, and i is a fourth root of unity. Thus we have the remarkable property that our mapping results in a local Hamiltonian if $\mathbb{Z}_4$ is respected. It is not difficult to see how to extend this argument to operators D that involve a range of sites (say

from j to $j+k$). Hence we may (mostly) ignore the non-local part of the Jordan-Wigner strings in the transformation. However, note that we will need to deal with the j part of strings coming from operators at site $j+1$.

We now have all the necessary tools at hand to transform (11.3.4) into an electronic Hamiltonian. The procedure is now simply to transform each term in (11.3.4) and collect the results, where we will work from right to left, this is carried out in Section 11.5. The result is

$$H_A = H_A^{(2)} + H_A^{(4)} + H_A^{(6)},$$

where we use the notation $H_A^{(m)}$, with $m = 2, 4, 6$, to indicate the $m/2$-fermion terms appearing in the transformed H_A, with

$$H_A^{(2)} = -J \sum_{j,\sigma} \left[c_{j,\sigma}^\dagger c_{j+1,\sigma} - i c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger \right] + \text{ H.c.}, \tag{11.3.5}$$

$$\begin{aligned} H_A^{(4)} = -J \sum_{j,\sigma} \Big[& c_{j,\sigma}^\dagger c_{j+1,\sigma} \left(-n_{j,-\sigma} - n_{j+1,-\sigma} \right) + c_{j,\sigma}^\dagger c_{j+1,-\sigma} i \left(n_{j,-\sigma} + n_{j+1,\sigma} \right) \\ & + c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger i \left(n_{j,-\sigma} + n_{j+1,\sigma} \right) + c_{j,\sigma}^\dagger c_{j+1,\sigma}^\dagger \left(n_{j,-\sigma} - n_{j+1,-\sigma} \right) \Big] + \text{ H.c.}, \end{aligned} \tag{11.3.6}$$

$$H_A^{(6)} = -J \sum_{j,\sigma} \left[-2i c_{j,\sigma}^\dagger c_{j+1,-\sigma} \left(n_{j,-\sigma} n_{j+1,\sigma} \right) - 2i c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger \left(n_{j,-\sigma} n_{j+1,\sigma} \right) \right] + \text{ H.c.}. \tag{11.3.7}$$

We collect some remarks on the resulting electronic Hamiltonian. The first thing to notice is that the coefficients of the terms where the creation and annihilation operators appear with flipped spins are always purely imaginary, whereas when they appear with the same spin they are real. The only processes that appear are hopping (with and without spin flip) and Cooper pairing (with and without spin flip), but they may be 'weighted/correlated' by the densities of other states. These are all well-studied and 'common' processes in electronic Hamiltonians. This entails in particular that the Hamiltonian respects Fermion parity, which we anticipated. In all the terms of degree 4, there is an imbalance in the appearance of the lattice sites, that is any term always contains thrice one site and only once the other. The only cancellation that occurs in the entire calculation are the same-spin terms in the order 6 Hamiltonian, every other possible term is present.

With our electronic representation of H_A in hand, we ask ourselves what happens to H_B under this mapping. The hope is of course that there is some nice choice for the value of $\hat{\gamma}_2$ (ideally non-zero) such that the full Hamiltonian $H_A + \hat{\gamma}_2 H_B$ becomes 'simpler' then H_A (perhaps at $\hat{\gamma}_2 = 1$, as this is the Potts model without a field). We therefore apply the mapping (11.3.3) to the $\hat{\gamma}_2 H_B$ term in our original Hamiltonian (11.1.1). Our result is

$$H_B = H_B^{(4)} = J \sum_{j=1}^{L-1} \prod_{k=0}^{1} \prod_{\sigma=\uparrow}^{\downarrow} \left(c_{j+k,\sigma}^\dagger + c_{j+k,\sigma} \right), \tag{11.3.8}$$

see Section 11.5 for the details. Some remarks on these terms are as follows. First, our transformed H_B contains exclusively 4-Fermion terms. Moreover, these are precisely the possible order 4 terms that did not appear in (11.3.6), and they all appear with exactly the same coefficient. Unfortunately, there does not appear to be an obvious simplification when these terms are added to H_A. In electronic language, H_B contains some 'uncommon' processes, namely 'simultaneous' Cooper pairing on adjacent sites (the term with four daggers), Cooper pairing on one site accompanied by a spin flip on an adjacent site (terms with three daggers), 'anti-simultaneous' Cooper pairing on adjacent sites (term with two adjacent daggers), and finally 'simultaneous' spin-flips on adjacent sites (terms with two non-adjacent daggers). For this reason, we will set $\hat{\gamma}_2 = 0$ for the remainder of this chapter.

Recall from the construction of H_n in (9.3.5) that if we make all coefficients $\hat{\gamma}_i$ equal, we find the n-state Potts model (see the first paragraph of Section 9.3). Thus if we also set $\hat{\gamma}_2 = 1$ (recall the we took $\hat{\gamma}_1 = 1$ at the start of this chapter), we find the following electronic representation of the zero-field 4-Potts model,

$$\begin{aligned} H_{4\mathrm{P}} = -J \sum_{j,\sigma} \Big[& c^\dagger_{j,\sigma} c_{j+1,\sigma} - i c^\dagger_{j,\sigma} c^\dagger_{j+1,-\sigma} + c^\dagger_{j,\sigma} c_{j+1,\sigma} \left(-n_{j,-\sigma} - n_{j+1,-\sigma}\right) + c^\dagger_{j,\sigma} c_{j+1,-\sigma} i \left(n_{j,-\sigma} + n_{j+1,\sigma}\right) \\ & + c^\dagger_{j,\sigma} c^\dagger_{j+1,-\sigma} i \left(n_{j,-\sigma} + n_{j+1,\sigma}\right) + c^\dagger_{j,\sigma} c^\dagger_{j+1,\sigma} \left(n_{j,-\sigma} - n_{j+1,-\sigma}\right) - 2 i c^\dagger_{j,\sigma} c_{j+1,-\sigma} \left(n_{j,-\sigma} n_{j+1,\sigma}\right) \\ & - 2 i c^\dagger_{j,\sigma} c^\dagger_{j+1,-\sigma} \left(n_{j,-\sigma} n_{j+1,\sigma}\right) - (c^\dagger_{j,\sigma} + c_{j,\sigma})(c^\dagger_{j,-\sigma} + c_{j,-\sigma})(c^\dagger_{j+1,\sigma} + c_{j+1,\sigma})(c^\dagger_{j+1,-\sigma} + c_{j+1,-\sigma})/4 \Big] \\ & + \text{ H.c.}, \end{aligned}$$

with which we close this section.

11.4 Fermionic Spectral Functions and a Topological Sweet Spot

As we saw in the last section, the Hamiltonian (11.1.1) becomes rather complicated when transformed into an electronic model. The term H_B even gave rise to some very 'exotic' electronic interactions, which lead us to set $\hat{\gamma}_2 = 0$. However, the electronic model consisting solely of H_A is still not quite as simple as one would like. In particular, the six-Fermion terms (11.3.7) and the correlated hopping with spin flip terms from (11.3.6) are problematic. One might thus be tempted to introduce two new parameters U and V, ranging from 0 to 2, and the more general Hamiltonian

$$\tilde{H}_A(U,V) = H_A^{(2)} + U \left[V \left(H^{(4)} + H^{(6)} \right) + (1-V) \tilde{H}_A^{(4)} \right], \tag{11.4.1}$$

where $\tilde{H}_A^{(4)}$ consists only of the terms of $H_A^{(4)}$ without spin flips (or equivalently only of the terms with real coefficients). Our original H_A corresponds to $\tilde{H}_A(1,1)$, which is an exactly solvable point of (11.4.1) (with solution given by reading the previous two sections in reverse). The natural questions to ask is now how much one can deform the model away from $U = V = 1$ without closing a gap. This was investigated numerically (using DMRG [77]) in [18], see Fig 11.1. They found evidence that there are two phases in the (U,V)-plane, separated by a gapless line. It turns out that the point $(1,0)$ is in the same phase as $(1,1)$, which we shall dub the strongly interacting phase, with Hamiltonian

$$\tilde{H}_A(1,0) = H_A^{(2)} + \tilde{H}_A^{(4)}.$$

This suggests that we can neglect the problematic terms discussed above, without changing the essential physics (modulo exponential groundstate splitting and the like, see Fig 11.2). The other phase, which we shall say is the weakly interacting one, features simpler, even quadratic, Hamiltonians, but as it can only be reached from H_A by closing a gap, we cannot say a priori what the physics of it will be. We will come back to this later in this section.

For now, our goal is to use the solvability of H_A in terms of parafermions to analytically compute the spectral function for the electronic model, as this is a quantity that can be probed experimentally. At zero temperature, and considering only the groundstate manifold, the Fermionic spectral function is given by the formula

$$\mathcal{A}_j^n(\omega) = 2\pi\delta(\omega) \sum_{m,\sigma} \left(\left| \langle \mathrm{GS}_m | \, c^\dagger_{j\sigma} \, | \mathrm{GS}_n \rangle \right|^2 + \left| \langle \mathrm{GS}_m | \, c_{j\sigma} \, | \mathrm{GS}_n \rangle \right|^2 \right) \tag{11.4.2}$$

Recall that we found the groundstates $|\mathrm{GS}_m\rangle$ of H_A in a previous section, and that they are listed in (11.2.3). They can also be expressed conveniently in terms of the Fermionic states $\left| f_{n_j} \right\rangle$ (see the

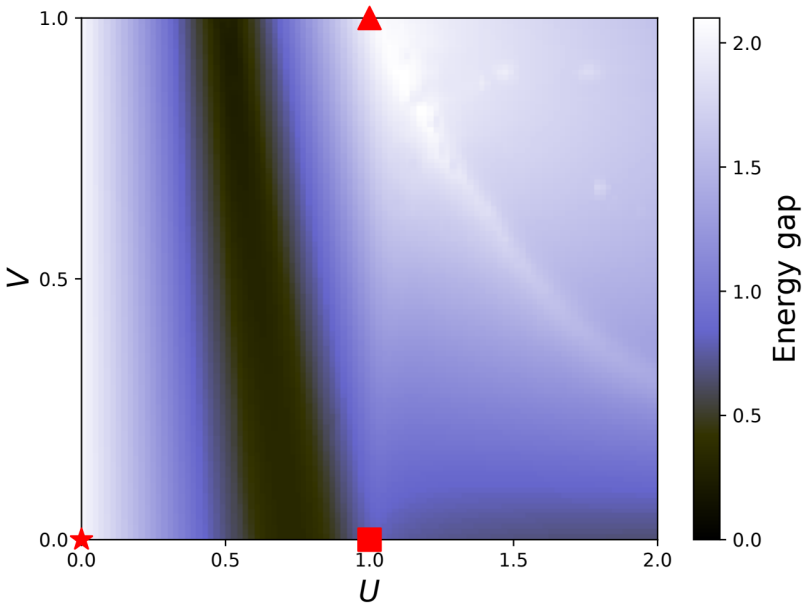


Figure 11.1: 16 Site DMRG simulation of $\tilde{H}_A(U,V)$ for $0 \leq U \leq 2$ and $0 \leq V \leq 1$, resulting in a contour plot of the gap in units of J. The triangle represents our original model H_A, the square is its deformation $\tilde{H}_A(1,0)$ discussed in the text, and finally the star represents the quadratic Hamiltonian $H_A^{(2)}$. Figure taken from [18].

previous section) if we introduce

$$\left|F_n^{L-1}\right\rangle = 2^{1-L} \sum_{\substack{\{m_j\} \text{ such that} \\ \sum_j m_j = n \pmod 4}} \bigotimes_{j=1}^{L-1} \left|f_{m_j}\right\rangle .$$

This allows for a number of useful decompositions. The first two are related to the behaviour of the system on the edges, as we can write

$$|\mathrm{GS}_n\rangle = \frac{1}{2}\Big(\left|F_0^{L-1}\right\rangle \otimes |f_n\rangle + \left|F_1^{L-1}\right\rangle \otimes |f_{n-1}\rangle + \left|F_2^{L-1}\right\rangle \otimes |f_{n-2}\rangle + \left|F_3^{L-1}\right\rangle \otimes |f_{n-3}\rangle\Big), \tag{11.4.3}$$

$$= \frac{1}{2}\Big(|f_n\rangle \otimes \left|F_0^{L-1}\right\rangle + |f_{n-1}\rangle \otimes \left|F_1^{L-1}\right\rangle + |f_{n-2}\rangle \otimes \left|F_2^{L-1}\right\rangle + |f_{n-3}\rangle \otimes \left|F_3^{L-1}\right\rangle\Big). \tag{11.4.4}$$

Now if we wanted to single out a bulk site k, we can instead write

$$|\mathrm{GS}_n\rangle = \frac{1}{4}\sum_{\ell=0}^{3}\Big(\left|F_\ell^{k-1}\right\rangle \otimes |f_n\rangle \otimes \left|F_{-\ell}^{L-k}\right\rangle + \left|F_\ell^{k-1}\right\rangle \otimes |f_{n-1}\rangle \otimes \left|F_{1-\ell}^{L-k}\right\rangle$$
$$+ \left|F_\ell^{k-1}\right\rangle \otimes |f_{n-2}\rangle \otimes \left|F_{2-\ell}^{L-k}\right\rangle + \left|F_\ell^{k-1}\right\rangle \otimes |f_{n-3}\rangle \otimes \left|F_{3-\ell}^{L-k}\right\rangle\Big). \tag{11.4.5}$$

The use of these representations, is that we know what the actions of the $c_{j\sigma}$ and $c_{j\sigma}^\dagger$ are on the states $|f_n\rangle$. Let us first focus on the annihilation operators $c_{j\sigma}$. For a single site, it follows by inspection that only the following matrix elements are non-zero,

$$\langle f_0|\, c_\uparrow\, |f_1\rangle = 1,\ \langle f_3|\, c_\uparrow\, |f_2\rangle = -1,\ \langle f_1|\, c_\downarrow\, |f_2\rangle = -i,\ \langle f_0|\, c_\downarrow\, |f_3\rangle = -i. \tag{11.4.6}$$

Notice that each of them is a complex number of unit norm, and that applying c_σ changes the parafermion parity by one (as discussed in the section where we build our mapping). In particular,

c_σ always annihilates two of the groundstates, and on the others it increases the parafermion parity of one, and lowers that of the other.

We now focus on the sites on the edges of the system, that is $j = 1, L$. An immediate consequence of the above is that the matrix elements

$$\langle \mathrm{GS}_n | \, c_{1\sigma} \, | \mathrm{GS}_n \rangle = \langle \mathrm{GS}_n | \, c_{L\sigma} \, | \mathrm{GS}_n \rangle = \langle \mathrm{GS}_{n+2} | \, c_{1\sigma} \, | \mathrm{GS}_n \rangle = \langle \mathrm{GS}_{n+2} | \, c_{L\sigma} \, | \mathrm{GS}_n \rangle = 0.$$

For the remaining matrix elements, we can use say (11.4.3) to compute

$$\begin{aligned} c_{L\sigma} \, | \mathrm{GS}_n \rangle &= \frac{1}{2} c_{L\sigma} \Big(\left| F_0^{L-1} \right\rangle \otimes | f_n \rangle + \left| F_1^{L-1} \right\rangle \otimes | f_{n-1} \rangle + \left| F_2^{L-1} \right\rangle \otimes | f_{n-2} \rangle + \left| F_3^{L-1} \right\rangle \otimes | f_{n-3} \rangle \Big), \\ &= \frac{1}{2} \left(A^{n,a}_{L,\sigma} \left| F_a^{L-1} \right\rangle \otimes | f_{n-a+1} \rangle + A^{n,b}_{L,\sigma} \left| F_b^{L-1} \right\rangle \otimes | f_{n-b-1} \rangle \right), \end{aligned}$$

where the coefficients[11] $A^{m,c}_{j\tau}$ arise from the single site matrix elements in (11.4.6) and anticommutation with the creation operators contained in the $\left| F_c^{L-1} \right\rangle$. In particular they are complex numbers with modulus unity. Notice that of the two resulting terms, precisely one has overlap only with the $n+1$ groundstate, and the other has overlap only with the $n-1$ groundstate. Therefore,

$$| \langle \mathrm{GS}_{n\pm1} | \, c_{L\sigma} \, | \mathrm{GS}_n \rangle |^2 = \left| \frac{1}{4} A^{n,c}_{L,\sigma} \right|^2 = \frac{1}{16}.$$

The same arguments, this time using (11.4.4), yield

$$| \langle \mathrm{GS}_{n\pm1} | \, c_{1\sigma} \, | \mathrm{GS}_n \rangle |^2 = \frac{1}{16}.$$

This we can combine into

$$\sum_m | \langle \mathrm{GS}_m | \, c_{1\sigma} \, | \mathrm{GS}_n \rangle |^2 = \sum_m | \langle \mathrm{GS}_m | \, c_{L\sigma} \, | \mathrm{GS}_n \rangle |^2 = \frac{1}{8}.$$

Finally, there is a clever argument to show that all the bulk operators $c_{j\sigma}$ have vanishing groundstate matrix elements using (11.4.5). Again by (11.4.6), $c_{j\sigma}$ applied to the n groundstate only has overlap with the $n \pm 1$ groundstates. If we now look closely at (11.4.5), we see that

$$\langle \mathrm{GS}_{n\pm1} | \, c_{j\sigma} \, | \mathrm{GS}_n \rangle \propto \sum_{\ell=0}^{3} (-1)^\ell = 0,$$

where the factor $(-1)^\ell$ arises from anticommuting $c_{j\sigma}$ through $\left| F_\ell^{L-1} \right\rangle$. We can now conclude that

$$\sum_m | \langle \mathrm{GS}_m | \, c_{j\sigma} \, | \mathrm{GS}_n \rangle |^2 = (\delta_{j1} + \delta_{JL}) \, \frac{1}{8},$$

and by analogous arguments we have the same result for the creation operator. Therefore,

$$\mathcal{A}^n_j(\omega) = \pi \delta(\omega) \, (\delta_{j1} + \delta_{jL}) \, . \tag{11.4.7}$$

This describes localised peaks with weight π at the edges of the system, which is what one would expect.

Let us now return to the weakly interacting phase. Here, the simplest possible case would be to reduce to the line $(0, V)$ given by the quadratic Hamiltonian (11.3.5), which we shall refer to as the

[11] Note there is no implied summation over repeated indices.

topological sweet spot. This is reminiscent of the Kitaev chain (9.1.6), except that that there are two spin species, and the Cooper pairs have opposite spins. However, this situation can be remedied as follows. Introduce the two sets of Majorana operators $a_{j\sigma}$, $b_{j\sigma}$ (compare with (9.2.1)) according to

$$a_{j,\sigma} = c^\dagger_{j,\sigma} + c_{j,\sigma},\ b_{j,\sigma} = i\left(c^\dagger_{j,\sigma} - c_{j,\sigma}\right).$$

Since there are spin flips in (11.3.5), these operators will not bring it into a form similar to that of (9.2.4). Hence we perform a 'rotation' in the Majorana operator space to the new (Majorana) operators

$$\gamma_{j,\sigma} = \frac{1}{\sqrt{2}}\left(a_{j,-\sigma} + b_{j,\sigma}\right),\ \eta_{j,\sigma} = \frac{1}{\sqrt{2}}\left(a_{j,\sigma} - b_{j,-\sigma}\right).$$

We expect this to decouple (11.3.5) into two non-interacting Kiteav chains,

$$H^{(2)}_A = Ji\sum_{j,\sigma}\gamma_{j\sigma}\eta_{j+1\sigma}, \tag{11.4.8}$$

and indeed one can show that this is the case by plugging in the above relations. We now proceed as one would for the Kitaev chain, by first noticing that the four Majorana operators $\gamma_{L,\sigma}$ and $\eta_{1,\sigma}$ do not appear in (11.4.8) and commute with it. Hence one defines two non-local fermions

$$f_\sigma = \frac{1}{2}\left(\gamma_{L\sigma} + i\eta_{1\sigma}\right),$$

out of the dangling Majorana modes. This gives a Fermionic edge zero mode, and thus the f_σ cycle through the four ground states as

$$|0\rangle,\ |\downarrow\rangle = f^\dagger_\downarrow|0\rangle,\ |\uparrow\downarrow\rangle = f^\dagger_\uparrow f^\dagger_\downarrow|0\rangle,\ |\uparrow\rangle = f^\dagger_\uparrow|0\rangle, \tag{11.4.9}$$

where $|0\rangle$ is the ground state annihilated by both f operators. We will use the dummy index τ to label these groundstates. This already suggests that the physics of the weakly interacting phase is different than in the strongly interacting phase. In the latter, there exist $\mathbb{Z}_4$ parafermionic edge modes, whereas in this phase there are instead two pairs of Majorana edge modes.

One would like to be able to (experimentally) distinguish between the strongly and weakly interacting phases, as only one of them features the parafermionic edge modes we are after. As we already have the fermionic spectral function for the full Hamiltonian H_A in the strongly interacting phase, we will also compute it for the topological sweet spot[12]. We begin by showing that the dangling edge Majoranas also cycle through the ground states. As

$$\gamma_{L,\sigma} = f_\sigma + f^\dagger_\sigma,\ \eta_{1,\sigma} = i\left(f^\dagger_\sigma - f_\sigma\right),$$

it holds that applying one of the dangling Majoranas to one of the ground states does the following. If the spin σ associated with the dangling Majorana is present, it is removed, and if it is not present, it is added. Hence each dangling Majorana cycles through a pair of groundstates. Moreover, the actions of $\gamma_{L\sigma}$ on any particular groundstate are different, same as for $\eta_{1\sigma}$. We next show that all bulk Majaronas (by which we mean all of them except $\gamma_{L\sigma}$ and $\eta_{1\sigma}$) have vanishing matrix elements in the groundstate manifold. To this end we diagonalise $H^{(2)}_A$ in terms of new Fermions. Specifically, define

$$g_{j\sigma} = \frac{1}{2}\left(\gamma_{j\sigma} - i\eta_{j+1\sigma}\right),$$

[12]The author would like to thank Alessio Calzona for sketching the proof in a private communication.

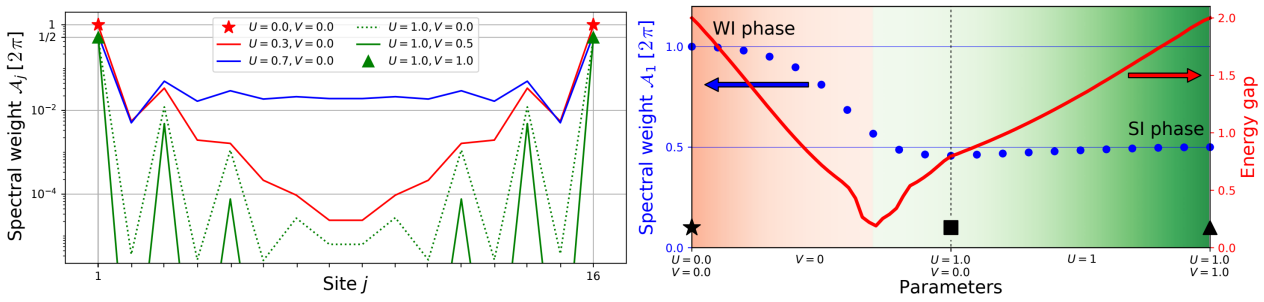


Figure 11.2: **Left:** The Fermionic spectral function (11.4.2) for a 16 site system plotted for several values of U and V. **Right:** The spectral weight concentrated in the left edge ($j = 1$) in blue, and the gap in red. The left half of the plot is on the line $V = 0$, and the right half is on the line $U = 1$. The star, square, and triangle represent the same points as in Fig 11.1. Both taken from [18].

for $j = 1, \ldots, L-1$. These are indeed fermions

$$g_{j\sigma}^2 = 0 = \left(g_{j\sigma}^\dagger\right)^2, \; g_{j\sigma}^\dagger g_{k\tau} = -g_{k\tau} g_{j\sigma}^\dagger.$$

Inverting this yields

$$\gamma_{j\sigma} = g_{j\sigma} + g_{j\sigma}^\dagger, \; \eta_{j+1\sigma} = i\left(g_{j\sigma} - g_{j\sigma}^\dagger\right).$$

Thus

$$\begin{aligned} H_A^{(2)} &= Ji \sum_{j,\sigma} \left(g_{j\sigma} + g_{j\sigma}^\dagger\right) i \left(g_{j\sigma} - g_{j\sigma}^\dagger\right), \\ &= -2J \sum_{j,\sigma} g_{j\sigma}^\dagger g_{j\sigma}, \end{aligned}$$

and the ground states are completely filled systems. However, the bulk Majoranas have g-Fermion parity one, so that they always leave the ground state manifold. This immediately implies that all the bulk spinful Fermion operators have vanishing matrix elements in the ground state manifold as well, since they can be expressed as linear combinations of the bulk Majoranas.

From the above we can already conclude that for any τ' and σ

$$\sum_\tau |\langle \tau | \, c_{j\sigma} \, | \tau' \rangle|^2 = \left(A_{1,\sigma}^{\tau'} \delta_{j1} + A_{L,\sigma}^{\tau'} \delta_{jL}\right),$$

where we use τ and τ' as dummy variables to label the groundstates (11.4.9). In other words, we now know that the spectral function is localised at the edges of the chain, as was the case in the strongly interacting phase. It is now useful to find the edge spinful Fermion operators $c_{1\sigma}^\dagger$ and $c_{L\sigma}^\dagger$ in terms of the edge Majoranas. The result is

$$\begin{aligned} c_{L\sigma}^\dagger &= \frac{1}{\sqrt{8}} \left[\eta_{L,\sigma} - \gamma_{L,-\sigma} + i\left(\eta_{L,-\sigma} - \gamma_{L,\sigma}\right)\right] \sim -\frac{1}{\sqrt{8}} \left[i\gamma_{L,\sigma} + \gamma_{L,-\sigma}\right], \\ c_{1\sigma}^\dagger &= \frac{1}{\sqrt{8}} \left[\eta_{1,\sigma} - \gamma_{1,-\sigma} + i\left(\eta_{1,-\sigma} - \gamma_{1,\sigma}\right)\right] \sim \frac{1}{\sqrt{8}} \left[\eta_{1,\sigma} + i\eta_{1,-\sigma}\right], \end{aligned}$$

where the tilde indicates restriction to the groundstate manifold (recall that $\gamma_{1\sigma}$ and $\eta_{L\sigma}$ count as bulk Majoranas). Now for instance

$$c_{L\sigma}^\dagger \, |\tau\rangle \sim B_1 \, |\tau'\rangle + B_2 \, |\tau''\rangle,$$

where one coefficient B_i is real and the other imaginary, and $\tau' \neq \tau''$. Therefore

$$\sum_\tau \left| \langle \tau | \, c^\dagger_{L\sigma} \, | \tau' \rangle \right|^2 = \frac{1}{8}(1+1) = \frac{1}{4} = A^{\tau'}_{L,\sigma},$$

for all τ'. By analogous arguments this is true for $j = 1$ and the annihilation operators as well. We conclude that

$$\sum_\tau \left| \langle \tau | \, c^\dagger_{j\sigma} \, | \tau' \rangle \right|^2 = \sum_\tau |\langle \tau | \, c_{j\sigma} \, | \tau' \rangle|^2 = (\delta_{j1} + \delta_{jL}) \, \frac{1}{4},$$

and so from (11.4.2) we have $\mathcal{A}_{1/L}(\epsilon) = 2\pi\delta(\epsilon)$. This peak has exactly double the weight of the one in the strongly interacting phase, and thus provides an experimentally accessible object to discern the phase of the model.

The results are plotted in Fig 11.2, with numerical results for parameter values intermediate to our two exactly solvable points. In the left figure, notice the exponential tail of the edge modes in the bulk. This is expected from the discussion of the Marjorana edge modes of the Kitaev chain in the previous chapter. This in particular means that for finite system sizes, there is an exponential splitting of the groundstate energies, and the four-fold degeneracy is only exact in the thermodynamic limit $L \to \infty$.

11.5 Appendix: The Sanguine Details of Certain Calculations

In this section we derive equations (11.3.5), (11.3.6), (11.3.7), and (11.3.6). We transform each term in (11.3.4) using (11.3.3) and the expressions for powers of the d's below it, and collect the results, working from right to left. First up is

$$\begin{aligned} d^{\dagger 3}_j d^3_{j+1} &= (1-n_{j\uparrow})(-ic^\dagger_{j\downarrow}) \; i^{3O_j + 2n_{j\uparrow} + 2n_{j\downarrow}} i(1-n_{j+1\uparrow})c_{j+1\downarrow}, \\ &= c^\dagger_{j\downarrow} i^{3n_{j\uparrow} + 5n_{j\downarrow} - 2n_{j\uparrow}n_{j\downarrow}} c_{j+1\downarrow}(1-n_{j\uparrow})(1-n_{j+1\uparrow}), \\ &= c^\dagger_{j\downarrow}(1-n_{j\uparrow}) i^{n_{j\downarrow} - n_{j\uparrow} - 2n_{j\uparrow}n_{j\downarrow}} c_{j+1\downarrow}(1-n_{j+1\uparrow}). \end{aligned}$$

Now notice that only $|\varnothing\rangle_j$ survives the combination $c^\dagger_{j\downarrow}(1-n_{j\uparrow})$, so that the power of i in the last line vanishes. Alternatively one can expand it using projectors and carry on from there. Thus

$$\begin{aligned} d^{\dagger 3}_j d^3_{j+1} &= c^\dagger_{j\downarrow}(1-n_{j\uparrow})c_{j+1\downarrow}(1-n_{j+1\uparrow}), \\ &= \left(c^\dagger_{j\downarrow} c_{j+1\downarrow} \right) + \left\{ -c^\dagger_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} - c^\dagger_{j\downarrow} c_{j+1\downarrow} n_{j+1\uparrow} \right\} + \left[c^\dagger_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} n_{j+1\uparrow} \right], \end{aligned} \tag{11.5.1}$$

where the terms in round brackets are 2-fermion terms, curly brackets are 4-fermion terms, and square brackets are 6-fermion terms (we will use this convention to report the final result of a transformation throughout this section). Next up is

$$\begin{aligned} d^{\dagger 3}_j d^\dagger_{j+1} &= -i(1-n_{j\uparrow})c^\dagger_{j\downarrow} i^{-\left(O_j + 2n_{j\uparrow} + 2n_{j\downarrow}\right)} \left(c_{j+1\uparrow} - c_{j+1\uparrow} n_{j+1\downarrow} - c^\dagger_{j+1\uparrow} n_{j+1\downarrow} + i c_{j+1\downarrow} n_{j+1\uparrow} \right)^\dagger, \\ &= -i(1-n_{j\uparrow})c^\dagger_{j\downarrow}(-i)^{n_{j\downarrow} - n_{j\uparrow} - 2n_{j\uparrow}n_{j\downarrow}} \left(c^\dagger_{j+1\uparrow} - c^\dagger_{j+1\uparrow} n_{j+1\downarrow} - c_{j+1\uparrow} n_{j+1\downarrow} - i c^\dagger_{j+1\downarrow} n_{j+1\uparrow} \right). \end{aligned}$$

The power of i is unity by the same argument as previously, so that

$$\begin{aligned} d^{\dagger 3}_j d^\dagger_{j+1} &= (1-n_{j\uparrow})c^\dagger_{j\downarrow} \left(-ic^\dagger_{j+1\uparrow} + ic^\dagger_{j+1\uparrow} n_{j+1\downarrow} + ic_{j+1\uparrow} n_{j+1\downarrow} + c^\dagger_{j+1\downarrow} n_{j+1\uparrow} \right), \\ &= \left(-ic^\dagger_{j\downarrow} c^\dagger_{j+1\uparrow} \right) + \left\{ +ic^\dagger_{j\downarrow} c^\dagger_{j+1\uparrow} n_{j+1\downarrow} + ic^\dagger_{j\downarrow} c_{j+1\uparrow} n_{j+1\downarrow} - c^\dagger_{j\downarrow} c^\dagger_{j+1\downarrow} n_{j+1\uparrow} + ic^\dagger_{j\downarrow} c^\dagger_{j+1\uparrow} n_{j\uparrow} \right\} \\ &\quad + \left[-ic^\dagger_{j\downarrow} c^\dagger_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} - ic^\dagger_{j\downarrow} c_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} + c^\dagger_{j\downarrow} c^\dagger_{j+1\downarrow} n_{j\uparrow} n_{j+1\uparrow} \right]. \end{aligned} \tag{11.5.2}$$

Our next victim is

$$\begin{aligned} d_j i^{O_j} d_{j+1}^3 &= \left(c_{j\uparrow} - c_{j\uparrow} n_{j\downarrow} - c_{j\uparrow}^\dagger n_{j\downarrow} + i c_{j\downarrow} n_{j\uparrow}\right) i^{O_j} i^{3O_j + 2n_{j\uparrow} + 2n_{j\downarrow}} \left(i(1-n_{j+1\uparrow}) c_{j+1\downarrow}\right), \\ &= i\left(c_{j\uparrow}(1-n_{j\downarrow}) - c_{j\uparrow}^\dagger n_{j\downarrow} + i c_{j\downarrow} n_{j\uparrow}\right) (-1)^{n_{j\uparrow}+n_{j\downarrow}} (1-n_{j+1\uparrow}) c_{j+1\downarrow}, \\ &= \left(-i c_{j\uparrow}(1-n_{j\downarrow}) + i c_{j\uparrow}^\dagger n_{j\downarrow} - c_{j\downarrow} n_{j\uparrow}\right)(1-n_{j+1\uparrow}) c_{j+1\downarrow}, \end{aligned}$$

where the powers of (-1) can be determined either by inspection, or by expanding it in projectors. It is now a matter of working out the product to find

$$\begin{aligned} d_j i^{O_j} d_{j+1}^3 = (-i c_{j\uparrow} c_{j+1\downarrow}) &+ \left\{ i c_{j\uparrow} c_{j+1\downarrow} n_{j\downarrow} + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow} - c_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} + i c_{j\uparrow} c_{j+1\downarrow} n_{j+1\uparrow} \right\} \\ &+ \left[-i c_{j\uparrow} c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} - i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} + c_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} n_{j+1\uparrow} \right]. \end{aligned} \tag{11.5.3}$$

Finally, the nastiest term (dealing with the powers like in the last term) is

$$\begin{aligned} d_j i^{O_j} d_{j+1}^\dagger &= \left(-c_{j\uparrow} + c_{j\uparrow} n_{j\downarrow} + c_{j\uparrow}^\dagger n_{j\downarrow} + i c_{j\downarrow} n_{j\uparrow}\right)\left(c_{j+1\uparrow}^\dagger - c_{j+1\uparrow}^\dagger n_{j+1\downarrow} - c_{j+1\uparrow} n_{j+1\downarrow} - i c_{j+1\downarrow}^\dagger n_{j+1\uparrow}\right), \\ &= \left(-c_{j\uparrow} c_{j+1\uparrow}^\dagger\right) + \left\{ c_{j\uparrow} c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + c_{j\uparrow} c_{j+1\uparrow} n_{j+1\downarrow} + i c_{j\uparrow} c_{j+1\downarrow}^\dagger n_{j+1\uparrow} + c_{j\uparrow} c_{j+1\uparrow}^\dagger n_{j\downarrow} \right. \\ &\quad \left. + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\downarrow} + i c_{j\downarrow} c_{j+1\uparrow}^\dagger n_{j\uparrow} \right\} + \left[-c_{j\uparrow} c_{j+1\uparrow}^\dagger n_{j\downarrow} n_{j+1\downarrow} - c_{j\uparrow} c_{j+1\uparrow} n_{j\downarrow} n_{j+1\downarrow} \right. \\ &\quad - i c_{j\uparrow} c_{j+1\downarrow}^\dagger n_{j\downarrow} n_{j+1\uparrow} - c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\downarrow} n_{j+1\downarrow} - c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j\downarrow} n_{j+1\downarrow} - i c_{j\uparrow}^\dagger c_{j+1\downarrow}^\dagger n_{j\downarrow} n_{j+1\uparrow} \\ &\quad \left. - i c_{j\downarrow} c_{j+1\uparrow}^\dagger n_{j\uparrow} n_{j+1\downarrow} - i c_{j\downarrow} c_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} + c_{j\downarrow} c_{j+1\downarrow}^\dagger n_{j\uparrow} n_{j+1\uparrow} \right]. \end{aligned} \tag{11.5.4}$$

We now collect all the terms, by order, from equations (11.5.1), (11.5.2), (11.5.3), and (11.5.4), yielding

$$(\dots) = -c_{j\uparrow} c_{j+1\uparrow}^\dagger - i c_{j\uparrow} c_{j+1\downarrow} - i c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger + c_{j\downarrow}^\dagger c_{j+1\downarrow}, \tag{11.5.5}$$

$$\begin{aligned} \{\dots\} = {} & c_{j\uparrow} c_{j+1\uparrow}^\dagger (n_{j\downarrow} + n_{j+1\downarrow}) - c_{j\downarrow}^\dagger c_{j+1\downarrow} (n_{j\uparrow} + n_{j+1\uparrow}) + i c_{j\uparrow} c_{j+1\downarrow} (n_{j\downarrow} + n_{j+1\uparrow}) \\ & + i c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\uparrow} + n_{j+1\downarrow}) + \left(c_{j\uparrow} c_{j+1\uparrow} n_{j+1\downarrow} + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\downarrow}\right) + \left(i c_{j\uparrow} c_{j+1\downarrow}^\dagger n_{j+1\uparrow} + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow}\right) \\ & + \left(i c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j+1\downarrow} + i c_{j\downarrow} c_{j+1\uparrow}^\dagger n_{j\uparrow}\right) - \left(c_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} + c_{j\downarrow}^\dagger c_{j+1\downarrow}^\dagger n_{j+1\uparrow}\right), \end{aligned} \tag{11.5.6}$$

$$\begin{aligned} [\dots] = {} & -\left(c_{j\uparrow} c_{j+1\uparrow}^\dagger + c_{j\uparrow}^\dagger c_{j+1\uparrow} + c_{j\uparrow} c_{j+1\uparrow} + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger\right) n_{j\downarrow} n_{j+1\downarrow} + (\uparrow \text{ to } \downarrow)\, n_{j\uparrow} n_{j+1\uparrow} \\ & - i c_{j\uparrow} c_{j+1\downarrow}^\dagger n_{j\downarrow} n_{j+1\uparrow} - i c_{j\downarrow} c_{j+1\uparrow}^\dagger n_{j\uparrow} n_{j+1\downarrow} - i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} - i c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} \\ & - i c_{j\uparrow}^\dagger c_{j+1\downarrow}^\dagger n_{j\downarrow} n_{j+1\uparrow} - i c_{j\downarrow} c_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} - i c_{j\uparrow} c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} - i c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\uparrow} n_{j+1\downarrow}. \end{aligned} \tag{11.5.7}$$

Recall that in (11.3.4) also the Hermitian conjugates of all the above terms appear, and that we must sum over the site index j. We will exploit this fact and some other tricks to bring the result into a more transparent form. In (11.5.5) (+ its Hermitian conjugate and summed over j), replace the first two terms by their Hermitian conjugates and extract a sum over spins to obtain

$$H_A^{(2)} = -J \sum_{j,\sigma} \left[c_{j,\sigma}^\dagger c_{j+1,\sigma} - i c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger \right] + \text{ H.c.}$$

In (11.5.7), notice that the first two terms (those with round brackets) are anti-Hermitian, so that they do not show up in the Hamiltonian. Replace terms with the dagger on the right by their Hermitian conjugate, and do the same to terms with no daggers. The factors of i ensure that the extra minus

sign coming from commuting back into j to $j+1$ order is cancelled. Doing this one notices that there are only two types of terms (those with one dagger, and those with two), which can be combined. A sum over spins is readily extracted and we find

$$H_A^{(6)} = -J \sum_{j,\sigma} \left[-2ic_{j,\sigma}^\dagger c_{j+1,-\sigma} \left(n_{j,-\sigma} n_{j+1,\sigma} \right) - 2ic_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger \left(n_{j,-\sigma} n_{j+1,\sigma} \right) \right] + \text{ H.c.}$$

In (11.5.6), carry out the same replacement by Hermitian conjugates and group the terms together correctly to find

$$\begin{aligned} H_A^{(4)} = -J \sum_{j,\sigma} \Big[& c_{j,\sigma}^\dagger c_{j+1,\sigma} \left(-n_{j,-\sigma} - n_{j+1,-\sigma} \right) + c_{j,\sigma}^\dagger c_{j+1,-\sigma} i \left(n_{j,-\sigma} + n_{j+1,\sigma} \right) \\ & + c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger i \left(n_{j,-\sigma} + n_{j+1,\sigma} \right) + c_{j,\sigma}^\dagger c_{j+1,\sigma}^\dagger \left(n_{j,-\sigma} - n_{j+1,-\sigma} \right) \Big] + \text{ H.c.} \end{aligned}$$

To derive (11.3.6), we have as before by (10.2.1) that H_B, as defined in (11.1.1), becomes

$$\begin{aligned} \chi_{j+1}^2 &= d_{j+1}^2 + d_{j+1} d_{j+1}^{\dagger 3} + d_{j+1}^{\dagger 3} d_{j+1}, \\ \psi_j^2 &= -i \left(d_j i^{O_j} d_j i^{O_j} + d_j i^{O_j} d_j^{\dagger 3} + d_j^{\dagger 3} d_j i^{O_j} \right), \\ &= d_j (-1)^{O_j} d_j - d_j d_j^{\dagger 3} + d_j^{\dagger 3} d_j. \end{aligned}$$

In anticipation of some of the calculations we will have to do, we compute some expressions that will be useful (using again the notation (11.3.1) and the reasoning behind (11.3.3)) Namely,

$$\begin{aligned} D_j (-1)^{O_j} D_j &= i \left(c_{j\downarrow} c_{j\uparrow} + c_{j\uparrow}^\dagger c_{j\downarrow} \right), \\ D_j D_j^{\dagger 3} &= i c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger, \\ D_j^{\dagger 3} D_j &= -i c_{j\downarrow}^\dagger c_{j\uparrow}, \\ D_j^2 &= (-i) \left(c_{j\downarrow} c_{j\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} \right). \end{aligned}$$

We now want to transform

$$\begin{aligned} \psi_j^2 \chi_{j+1}^2 = & \, d_j (-1)^{O_j} d_j d_{j+1}^2 + d_j (-1)^{O_j} d_j d_{j+1} d_{j+1}^{\dagger 3} + d_j (-1)^{O_j} d_j d_{j+1}^{\dagger 3} d_{j+1} \\ & - d_j d_j^{\dagger 3} d_{j+1}^2 - d_j d_j^{\dagger 3} d_{j+1} d_{j+1}^{\dagger 3} - d_j d_j^{\dagger 3} d_{j+1}^{\dagger 3} d_{j+1} \\ & + d_j^{\dagger 3} d_j d_{j+1}^2 + d_j^{\dagger 3} d_j d_{j+1} d_{j+1}^{\dagger 3} + d_j^{\dagger 3} d_j d_{j+1}^{\dagger 3} d_{j+1}. \end{aligned} \tag{11.5.8}$$

We do this term by term (from left to right this time). Again all $k < j$ parts of the strings cancel, so we will not write them. In general, we will give few details in the following calculations, as the arguments are entirely analogous to the computation for H_A. First

$$\begin{aligned} d_j (-1)^{O_j} d_j d_{j+1}^2 &= D_j (-1)^{O_j} D_j i^{2(O_j + 2j_\uparrow + 2n_{j\downarrow})} D_{j+1}^2, \\ &= \left(c_{j\downarrow} c_{j\uparrow} + c_{j\uparrow}^\dagger c_{j\downarrow} \right) (-1)^{n_{j\uparrow} + n_{j\downarrow}} \left(c_{j+1\downarrow} c_{j+1\uparrow} - c_{j+1\uparrow}^\dagger c_{j+1\downarrow} \right), \\ &= \left(c_{j\downarrow} c_{j\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} \right) \left(c_{j+1\downarrow} c_{j+1\uparrow} - c_{j+1\uparrow}^\dagger c_{j+1\downarrow} \right). \end{aligned}$$

Then we have

$$\begin{aligned} d_j (-1)^{O_j} d_j d_{j+1} d_{j+1}^{\dagger 3} &= D_j (-1)^{O_j} D_j (-1)^{O_j} D_{j+1} D_{j+1}^{\dagger 3} = - \left(c_{j\downarrow} c_{j\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} \right) \left(c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger \right), \\ d_j (-1)^{O_j} d_j d_{j+1}^{\dagger 3} d_{j+1} &= \left(c_{j\downarrow} c_{j\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} \right) \left(c_{j+1\downarrow}^\dagger c_{j+1\uparrow} \right). \end{aligned}$$

The next row in (11.5.8) becomes

$$
\begin{aligned}
d_j d_j^{\dagger 3} d_{j+1}^2 &= D_j D_j^{\dagger 3} (-1)^{O_j} D_{j+1}^2 = \left(c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger\right)\left(c_{j+1\downarrow} c_{j+1\uparrow} - c_{j+1\uparrow}^\dagger c_{j+1\downarrow}\right), \\
d_j d_j^{\dagger 3} d_{j+1} d_{j+1}^{\dagger 3} &= -c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger, \\
d_j d_j^{\dagger 3} d_{j+1}^{\dagger 3} d_{j+1} &= c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\downarrow}^\dagger c_{j+1\uparrow}.
\end{aligned}
$$

Finally the last row is

$$
\begin{aligned}
d_j^{\dagger 3} d_j d_{j+1}^2 &= D_j^{\dagger 3} D_j (-1)^{O_j} D_{j+1}^2 = \left(c_{j\downarrow}^\dagger c_{j\uparrow}\right)\left(c_{j+1\downarrow} c_{j+1\uparrow} - c_{j+1\uparrow}^\dagger c_{j+1\downarrow}\right), \\
d_j^{\dagger 3} d_j d_{j+1} d_{j+1}^{\dagger 3} &= -c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger, \\
d_j^{\dagger 3} d_j d_{j+1}^{\dagger 3} d_{j+1} &= c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\downarrow}^\dagger c_{j+1\uparrow}.
\end{aligned}
$$

We collect the terms

$$
\begin{aligned}
\psi_j^2 \chi_{j+1}^2 = \quad & c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger + c_{j\downarrow} c_{j\uparrow} c_{j+1\downarrow} c_{j+1\downarrow} \\
& + c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow} + c_{j\downarrow} c_{j\uparrow} c_{j+1\downarrow}^\dagger c_{j+1\uparrow} - c_{j\downarrow} c_{j\uparrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow} - c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\downarrow}^\dagger c_{j+1\uparrow} \\
& + c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger + c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\downarrow} c_{j+1\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\downarrow} c_{j+1\uparrow} - c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger \\
& + c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow} + c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\downarrow}^\dagger c_{j+1\uparrow} - c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\downarrow}^\dagger c_{j+1\uparrow} - c_{j\downarrow}^\dagger c_{j\uparrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow} \\
& - c_{j\downarrow} c_{j\uparrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger - c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\downarrow} c_{j+1\uparrow}.
\end{aligned}
$$

Notice that each term is paired with its Hermitian conjugate. Commute operators in each term with a minus sign such that the spins read up-down-up-down, then we arrive at the form

$$
\begin{aligned}
\psi_j^2 \chi_{j+1}^2 = \quad & c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger \\
& + c_{j\uparrow} c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger + c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow}^\dagger + c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow} c_{j+1\downarrow}^\dagger + c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger c_{j+1\downarrow} \\
& + c_{j\uparrow}^\dagger c_{j\downarrow}^\dagger c_{j+1\uparrow} c_{j+1\downarrow} + c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\uparrow}^\dagger c_{j+1\downarrow} + c_{j\uparrow}^\dagger c_{j\downarrow} c_{j+1\uparrow} c_{j+1\downarrow}^\dagger \\
& + \text{ H.c.}, \\
= & \left(c_{j\uparrow}^\dagger + c_{j\uparrow}\right)\left(c_{j\downarrow}^\dagger + c_{j\downarrow}\right)\left(c_{j+1\uparrow}^\dagger + c_{j+1\uparrow}\right)\left(c_{j+1\downarrow}^\dagger + c_{j+1\downarrow}\right).
\end{aligned}
$$

Thus the final result is

$$
H_B = H_B^{(4)} = J \sum_{j=1}^{L-1} \prod_{k=0}^{1} \prod_{\sigma=\uparrow}^{\downarrow} \left(c_{j+k,\sigma}^\dagger + c_{j+k,\sigma}\right).
$$

12 Free Fermion-Like Fock Parafermion Models

12.1 Tight-Binding Fock Parafermions

In the previous chapter, we looked at Fock parafermion models 'close' to the Potts model. While this was natural, in the sense that the Potts model becomes very simple in terms of ordinary parafermions, it resulted in rather complicated expressions, both in their Fock parafermion and electron representations. In this chapter we will instead consider the Fock parafermions to be the fundamental degree of freedom, and try to study Hamiltonians that are simple in this language. If one were to write down a one-dimensional, tight-binding Hamiltonian for Fock parafermions (that is not some combination of number operators) for this purpose, one is likely to end up with

$$H_h = -t \sum_j \left(d_j^\dagger d_{j+1} + \text{H.c.} \right), \tag{12.1.1}$$

where $t > 0$ (one usually sets $t = 1$ in numerics). The study of this model in the literature was initiated by Rossini, Carrega, Strinati, and Mazza in [74]. In the present section, we will summarise their findings, and in the following sections, we will attempt to apply the techniques we have learned so far to learn more about the model, and we start to consider directions in the phase diagram[13].

The first thing to notice about (12.1.1), is that it preserves the Fock parafermion number, in contrast to (11.3.4). Thus $O_{\text{tot}} = \sum_j O_j$ is a good quantum number, with the corresponding symmetry being the usual U(1). Compared to the Fermionic version of (12.1.1), inversion symmetry is missing, since $d_j \mapsto d_{-j}$ does not preserve the relations (10.1.4). While the off-site Fock parafermion algebra (10.1.4) is particle-hole symmetric, as is the on-site relation (10.1.5), if we try to apply this transformation to the Hamiltonian

$$H_h \mapsto -t \sum_j \left(d_j d_{j+1}^\dagger + d_{j+1} d_j^\dagger \right) = -t \sum_j \left(\omega d_j^\dagger d_{j+1} + \text{H.c.} \right),$$

it is not invariant (for Fermions $\omega = -1$, but the single-particle spectrum $-2t\cos(k)$ is invariant under this reflection). However, one can combine particle-hole and inversion transformations to obtain a symmetry (see the text at the end of Section IIB in [74]). As a consequence, one only needs to solve the model for densities below 'half-filling'. Note that the maximal occupation number for $\mathbb{Z}_n$ Fock parafermions is $n-1$, thus full filling means $n-1$ particles per site, and the total number of particles is $(n-1)L$. This corresponds to a maximal particle density of $(n-1)$, and so half-filling means a density of $(n-1)/2$.

If there is only a single particle, the behaviour should be the same as in the Fermionic case, as there is no notion of exchange statistics for a single particle. Consider the state $|\ell\rangle = d_\ell^\dagger |0\rangle$, with $|0\rangle$ the vacuum. Then

$$d_j^\dagger d_{j+1} |\ell\rangle = \delta_{\ell,j+1} d_j^\dagger d_{j+1} d_{j+1}^\dagger |0\rangle = \delta_{\ell,j+1} d_j^\dagger \left(1 - d_{j+1}^{\dagger(n-1)} d_{j+1}^{n-1} \right) = \delta_{\ell,j+1} |j\rangle ,$$

which is the same as for Fermions (we used (10.1.4) in the second step). Therefore, the states

$$|k\rangle = \frac{1}{\sqrt{L}} \sum_\ell e^{ik\ell} |\ell\rangle ,$$

[13] This is joint work with Iman Mahyaeh (Stockholm), Jurriaan Wouters, and Dirk Schuricht.

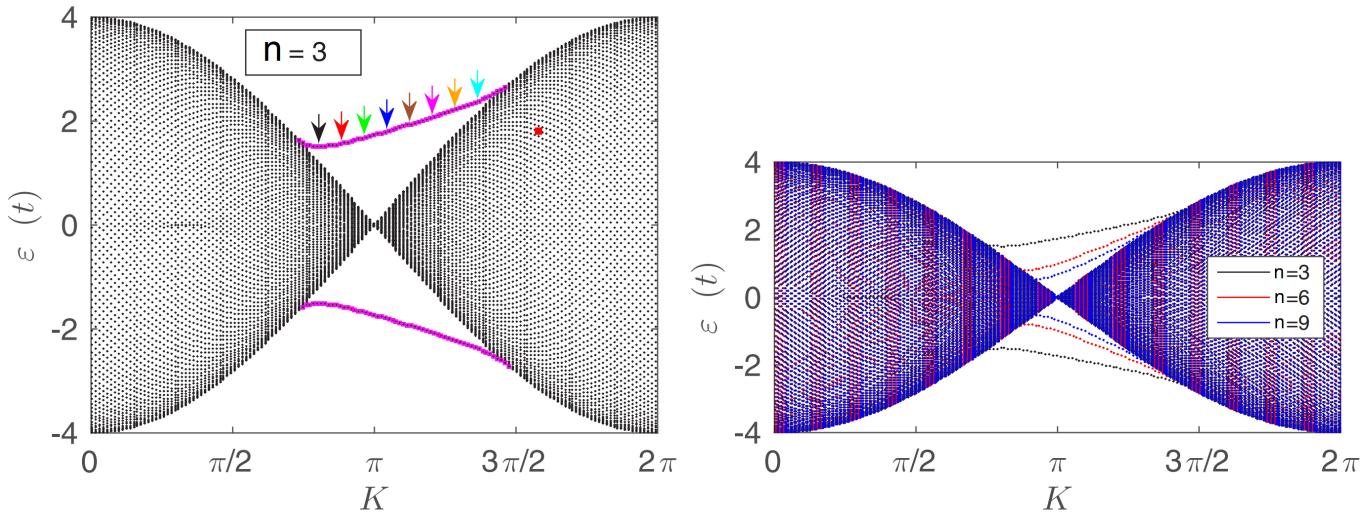


Figure 12.1: Two-particle spectrum of (12.1.1) as a function of center of mass momentum K for several values of n, adapted from [74]. Exact diagonalisation on $L = 151$ sites with periodic boundary conditions. The density associated to the red square in the left image is plotted on the right in Figure 12.2. The different coloured arrows indicate bound states whose densities are plotted (in matching colours) on the left in Figure 12.2. In the right image, notice that the lines of bound states head towards the lobes for increasing n, as expected since this is the hard-core Boson limit.

with $k \in (2\pi/L)\mathbb{Z}_L$ (using periodic boundary conditions), are eigenvectors of (12.1.1),

$$\begin{aligned} H_h |k\rangle &= -\frac{t}{\sqrt{L}} \sum_{j,\ell} \left(d_j^\dagger d_{j+1} + d_{j+1}^\dagger d_j \right) e^{ik\ell} |\ell\rangle \,, \\ &= -\frac{t}{\sqrt{L}} \sum_{j} \left(e^{ik(j+1)} d_j^\dagger + e^{ikj} d_{j+1}^\dagger \right) |0\rangle \,, \\ &= -2t \cos(k) |k\rangle \,, \end{aligned}$$

with the expected dispersion $\varepsilon(k) = -2t\cos(k)$.

Unfortunately, things become more complicated if we add another particle. In the Fermionic case, the eigenstates are simply $|k, k'\rangle$, and the energy is

$$\varepsilon(k, k') = -2t \left[\cos(k) + \cos(k')\right] = -4t \cos(K) \cos\left(\frac{k - k'}{2}\right), \tag{12.1.2}$$

with $K = (k + k')/2$ the center of mass momentum. The two-particle spectrum of (12.1.1) was found in [74] for several values of n, and can be seen in Figure 12.1. The two lobes on the left and right in the plots in Figure 12.1 are the same as in the Fermionic case (the lobes are a 'side view' of (12.1.2)), but there are lines of states between them. The states in the lobes are scattering states, as can be seen on the right in Figure 12.2, as in the Fermionic case. However, as evident from the left plot in Figure 12.2, the lines are made up of 'bound' states. This means that despite (12.1.1) being quadratic, it is not a model of 'free' particles unless $n = 2$ (compare with [35] and the equation above (30) in [34]). Analytically, one could argue this as follows. In the Fermionic ($n = 2$) case, the Hamiltonian can be Fourier transformed to become

$$H_h^{(n=2)} = \sum_k \varepsilon(k) c_k^\dagger c_k,$$

which is made out of commuting pieces,

$$\left[c_k^\dagger c_k, c_{k'}^\dagger c_{k'} \right] = 0.$$

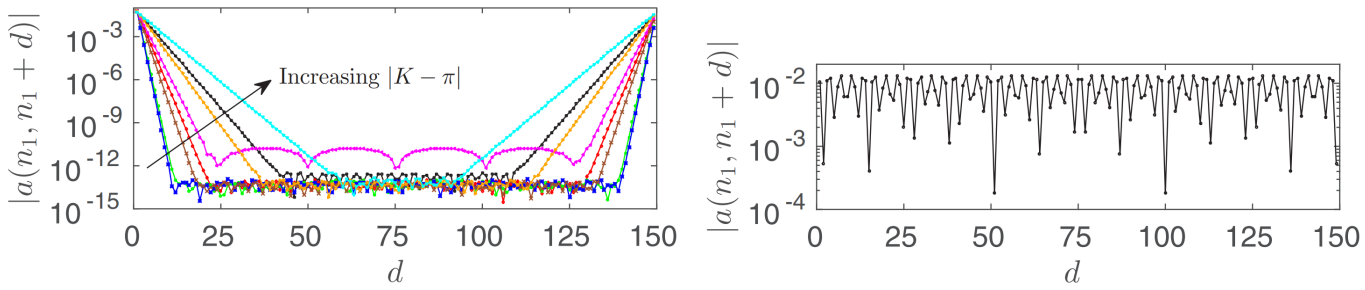


Figure 12.2: Plots of the coefficients $|a(n_1, n_2)|$, defined through $|\psi\rangle = \sum_{n_1,n_2} d^\dagger_{n_1} d^\dagger_{n_2} |0\rangle$, of the states (left) given by the arrows in Figure 12.1 and the state (right) given by the red square in Figure 12.1, with $n_2 - n_1 = d$. Adapted from [74]. Notice on the left that the profiles become wider as the state comes closer to the lobes in Figure 12.1.

If $n \geq 3$, things go wrong. First, if one were to define $d^\dagger_k$ via

$$|k\rangle = d^\dagger_k |0\rangle = \frac{1}{\sqrt{L}} \sum_\ell e^{ik\ell} d^\dagger_\ell |0\rangle$$

as the 'Fourier transform' of $d^\dagger_j$, this would not be a canonical transformation, i.e. the operators $d^\dagger_k$ and d_k do not satisfy the Fock parafermion algebra (10.1.4). Intuitively, this follows from the fact that this algebra is 'chiral', which is incompatible with Fourier transformation. Concretely, commuting a $d^\dagger_k$ and a $d^\dagger_{k'}$ involves commuting $d^\dagger_j$'s and $d^\dagger_{j'}$'s where both the cases $j < j'$ and $j > j'$ occur, but these result in different powers of ω. Moreover, even if it were possible to write

$$H_h = \sum_k \tilde{\varepsilon}(k) d^\dagger_k d_k,$$

then this still does not consist of commuting pieces, as one can check that $d^\dagger_k d_k$ does not commute with $d^\dagger_{k'} d_{k'}$. Recall also that $d^\dagger_\ell d_\ell$ is not the Fock parafermion number operator (10.1.6) on site ℓ.

Further arguments that the case $n \geq 3$ is very different from $n = 2$ can be made using random matrix theory, see [6, 44, 52] for instance. The Fermionic case is integrable (as demonstrated explicitly above), so one expects the spectrum to display *Poisson statistics*. What is meant by this statement, is that the level spacings $s_n = (E_{n+1} - E_n)/\langle s \rangle$, viewed as realisations of a random variable s, are Poisson distributed, that is to say that s has probability density e^{-s}. A priori, this is a strange distribution to see, as we know that levels generically repel each other (see Remark 5.2), and this distribution is largest for s close to zero. Integrable systems have infinitely many conserved quantities (by definition), and that is enough to overcome this level repulsion. In Figure 12.3, the density function of s (induced by H_h) has been plotted for several values of n, and we clearly see that as n moves away from 2, the maximum of the distribution moves away from zero. This is strong numerical evidence that the general Fock parafermion model H_h is not integrable. In general, the type of eigenvalue statistics one observes is dictated by the symmetries of the model. Well known distributions are the Gaußian Unitary Ensemble (GUE) briefly touched upon in Chapter 5, and the Gaußian Orthogonal Ensemble (GOE), which has density $(\pi s/2)e^{-\pi s^2/4}$. This latter density arises when the system has an anti-unitary symmetry, but it can be difficult to distinguish between GUE and GOE for small system sizes, see Appendix B of [74].

We attempt to say something about the low-energy physics of H_h in the many-particle sector. Using DMRG [77], one can compute the gap between the groundstate and the first excited state, see Figure 12.4 and also FIG. 4 in [74]. We observe vanishing gaps in the thermodynamic limit for almost all values of n and densities O_{tot}/L, with the notable exception being the combination $n = 3$ and $O_{\text{tot}}/L = 1$ (so half-filling). Thus one expects the low-energy behaviour to generally be described

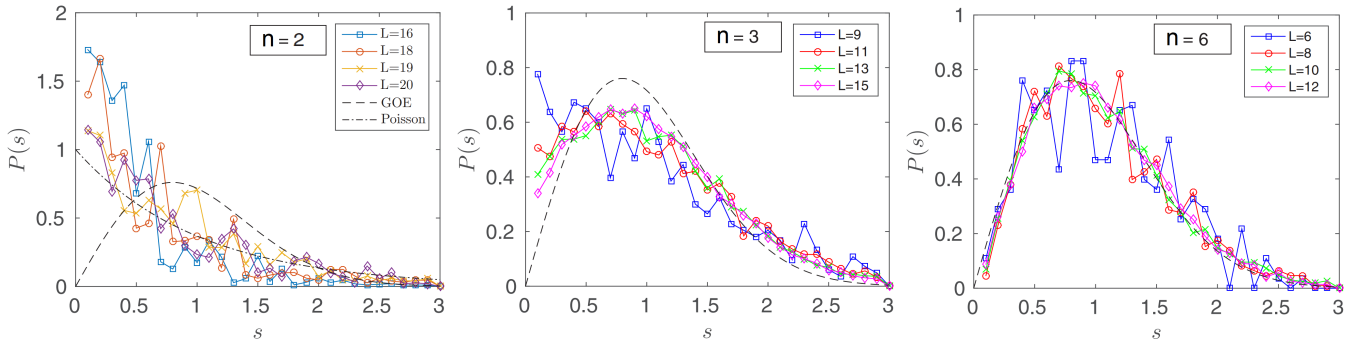


Figure 12.3: Level statistics in the 7-particle sector of (12.1.1) for varying system sizes and several values of n. Exact diagonalisation with open boundary conditions, adapted from [74] (see their FIG. 3 for details). The dashed curve in all three plots is the GOE density. One can clearly see motion away from Poissonian statistics for $n = 2$ towards something resembling GOE for $n = 6$. Recall from Chapter 5 that convergence towards the GUE distribution improves rapidly with growing matrix size, somewhat justifying the increasing quality of the fit with growing n.

by a conformal field theory. For a review of conformal field theory, see Chapter 10 and beyond in [58]. Another interesting quantity that is numerically accessible through DMRG is the entanglement entropy of the system in the groundstate. This contains a lot of information about the correlations between different subsystems, see Chapter 17 of [36] for more on entanglement entropy. Since we are working with a one-dimensional chain, there is a natural way to choose two subsystems. Let ℓ be an integer between 1 and L, the length of the chain, and split the system into the first ℓ and the last $L - \ell$ sites. Write $|GS\rangle$ for the groundstate, and form its density matrix $\rho = |GS\rangle \langle GS|$, and perform a partial trace over the last $L - \ell$ sites,

$$\rho_\ell = \mathrm{Tr}_{L-\ell}(\rho),$$

to obtain the reduced density matrix for the first subsystem. Finally, consider the von Neumann entropy of this reduced density matrix ρ_ℓ,

$$S(\ell) = -\mathrm{Tr}\left[\rho_\ell \log(\rho_\ell)\right],$$

which is the entanglement entropy. Recall that (reduced) density operators are positive, so that the assignment $\rho_\ell \mapsto \rho_\ell \log(\rho_\ell)$ is well-defined if we extend $f(x) = x \log(x)$ to a continuous function on $[0, \infty)$ by setting $f(0) = 0$. In a one-dimensional conformal field theory, $S(\ell)$ has a universal functional form, which for our geometry (and open boundary conditions) was derived by Calabrese and Cardy [16] and reads

$$S(\ell) = C + \frac{c}{6} \log\left[\frac{2\pi}{L} \sin\left(\frac{\pi \ell}{L}\right)\right], \tag{12.1.3}$$

where c is the conformal charge of the theory, and C is some (non-universal) constant. Therefore we have a two-parameter expression to fit to data, which allows us both to check that the low-energy behaviour is (approximately) conformal and to determine the central charge of the corresponding theory. In Figure 12.5, the entanglement entropy is plotted for the $n = 4$ case at half-filling ($O_{\mathrm{tot}}/L = 3/2$), and we see excellent agreement with (12.1.3). The fit suggests that the central charge c should be unity, and this is consistent [54] with the $n = 4$ case in Figure 12.4, since the leading finite size behaviour of the gap depends linearly on the conformal charge [1, 11], providing a consistency check. In [74], additional numerical evidence that $c = 1$ holds (for all n!) is presented in FIG. 5. The upper panel of this figure also shows the deviation from (12.1.3) for the special case $n = 3$ at half-filling,

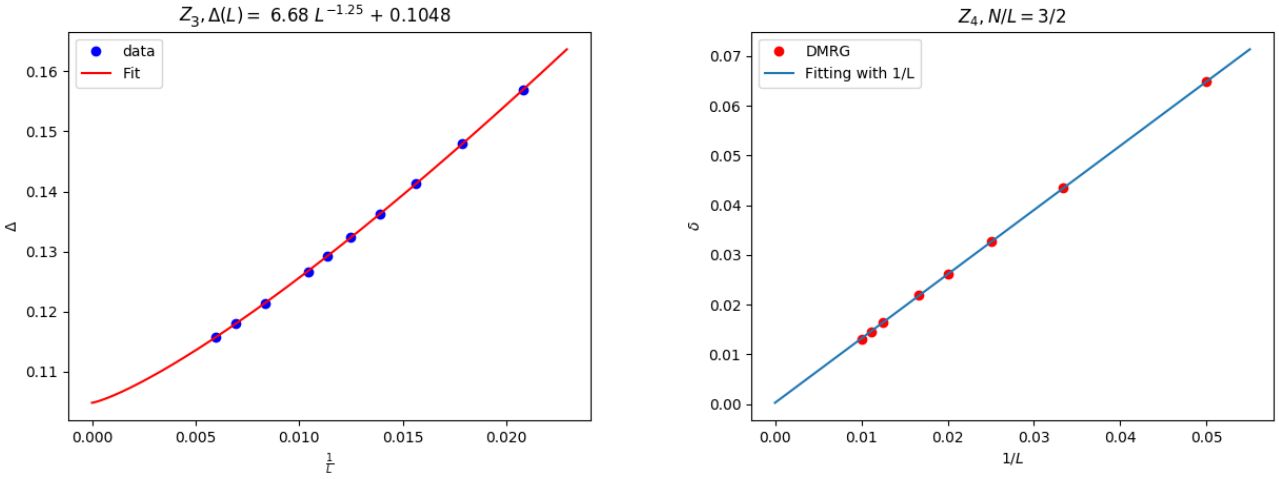


Figure 12.4: DMRG calculations [54] of the gap between the groundstate and the first excited state of H_h as a function of $1/L$, compare with FIG. 4 in [74]. Both plots are at half-filling for the respective values of n. On the right, $n = 4$ and we see a gap that is linear in $1/L$ and vanishes in the thermodynamic limit. On the left, $n = 3$ and we see instead that the gap approaches a finite value (note the ordinate), fitted to the functional form $AL^{-\alpha} + B$, with A, B, α constants.

instead following an area law. This suggests that the critical point of (12.1.1) is described by the same conformal field theory as a free Boson or Fermion, since these also have $c = 1$ (Chapter 12 in [58]). For the n-Potts models (Chapter 14 in [58]), $c = 4/5$ when $n = 3$, and $c = 1$ when $n = 4$, indicating that at least for $n = 3$, H_h is indeed far away from the Potts model, and for $n > 4$, the Potts model has a first order phase transition that cannot even be described by a conformal field theory.

Another observable that is accessible through DMRG is the one-particle many-body correlation function

$$g_1(j, \ell) = \langle GS | \, d_j^\dagger d_\ell \, | GS \rangle \, . \tag{12.1.4}$$

In [17], combining Equation (6) and the text below Equation (7), one finds the leading order prediction

$$g_1(|j - \ell|) \sim |j - \ell|^{-2/n}, \tag{12.1.5}$$

deep in the bulk (recall we use open boundary conditions). This prediction can only be accurate in the gapless cases, because when the system is gapped, the correlation length should be finite and thus the correlations decay exponentially. In the gapless cases, the correlations are predicted to decay only decrease algebraically, with an exponent that is a direct measure of the statistical parameter $2/n$. In Figure 12.6, one can see such a correlation function for $n = 4$, with the exponent resulting from the fit being 0.53, which is in good agreement with $2/4 = .5$. Here however, one should note that there is considerable variation with respect to density in the exponents following from numerics, see FIG. 6 in [74].

Let us finally introduce a hopping model that is slightly different from (12.1.1) (and not considered in [74]), by drawing some inspiration from the Potts model in terms of Fock parafermions (11.3.4). There, we see a 'chiral' hopping term, namely

$$H_{Ph} = -t \sum_j \left(d_j^\dagger \overline{\omega}^{O_j} d_{j+1} + \text{H.c.} \right) .$$

At this point, we have no way of deciding whether H_h or H_{Ph} is the more natural Hamiltonian. Including the phase factor does come with a certain benefit. To perform DMRG calculations, it is

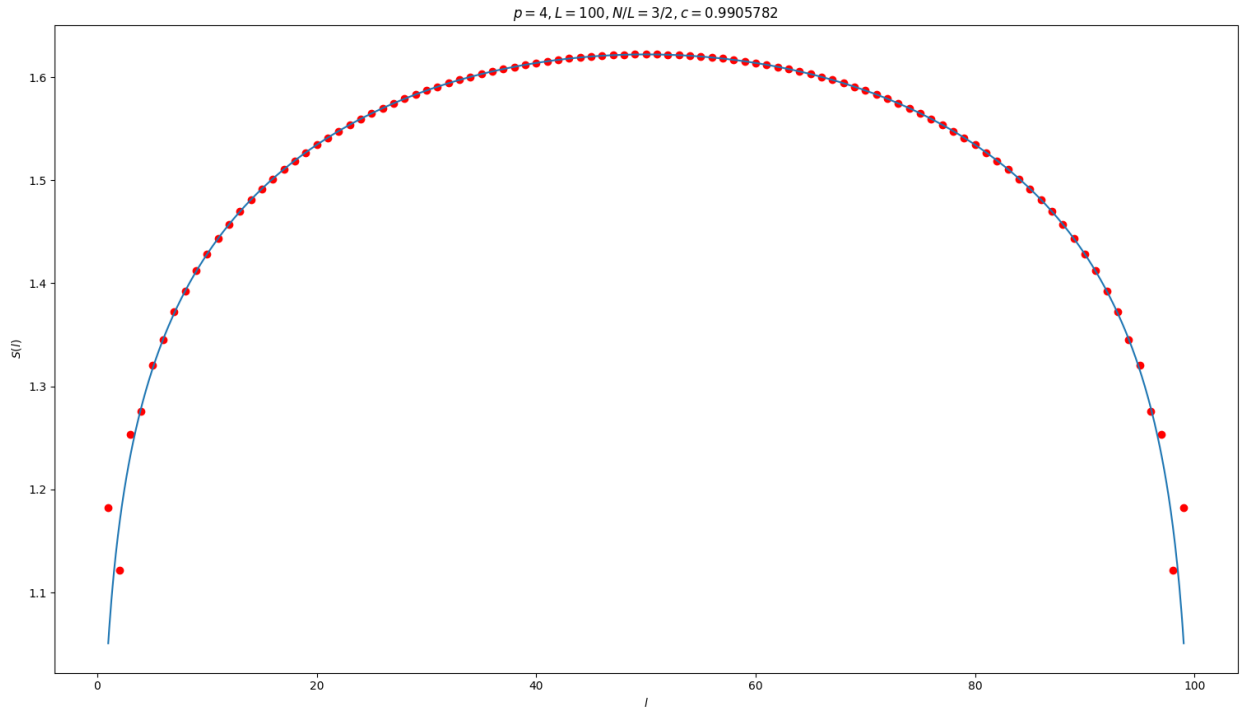


Figure 12.5: DMRG calculation (dots) of the entanglement entropy $S(\ell)$ for the $n = 4$ version of (12.1.1) at half-filling ($O_{\text{tot}}/L = 3/2$), and a plot of the best fit for (12.1.3) (line). We observe excellent agreement, especially away from the boundaries, providing strong evidence that $c = 1$. Adapted from [54].

useful to apply the Fradkin-Kadanoff transformation (10.2.5) to the two hopping models, so that they are written in terms of operators with more conventional commutation relations. It is simple to see that

$$d_j^\dagger d_{j+1} = B_j^\dagger U_{j-1}^\dagger \cdots U_1^\dagger U_1 \cdots U_j B_{j+1} = B_j^\dagger U_j B_{j+1},$$
$$d_j^\dagger \overline{\omega}^{O_j} d_{j+1} = B_j^\dagger U_{j-1}^\dagger \cdots U_1^\dagger U_j^\dagger U_1 \cdots U_j B_{j+1} = B_j^\dagger B_{j+1}.$$

Therefore,

$$H_h = -t \sum_j \left(B_j^\dagger U_j B_{j+1} + B_{j+1}^\dagger U_j^\dagger B_j \right),$$
$$H_{Ph} = -t \sum_j \left(B_j^\dagger B_{j+1} + B_{j+1}^\dagger B_j \right).$$

Now notice that in this representation, H_{Ph} does not have any term involving U, and so it is not only simpler than H_h, but it is a matrix with exclusively real entries. In fact, we see that H_{Ph} is a model of hopping hard-core Bosons. The failure of this fact for H_h follows from its lack of time reversal symmetry (see the text at the end of section IIB in [74]), so it may be the case that adding this phase gives a model with more symmetries.

12.2 Alternative Representations and the Continuum Limit

In the previous section, we provided numerical evidence that the low energy behaviour of (12.1.1) should be described by a conformal field theory with central charge $c = 1$. This suggests that H_h is 'close' to a free Fermion model not only in outward appearance. To provide analytical evidence in

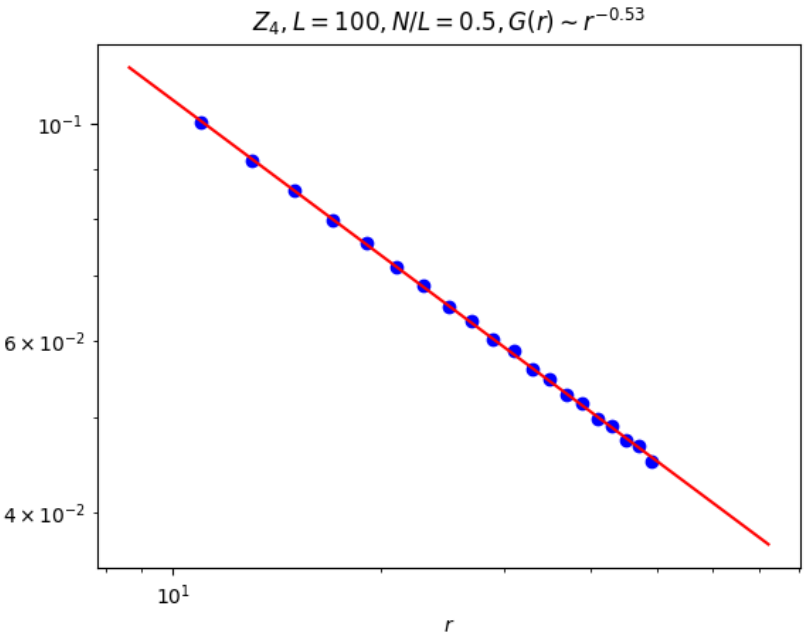


Figure 12.6: DMRG calculation [54] on 100 sites of (12.1.4) for $n = 4$ at filling $O_{\text{tot}}/L = 0.5$ (dots), with the best fit for the functional form r^α (line), which has exponent -0.53, in good agreement with the predicted value of $-1/2$ in (12.1.5).

this direction, we would like to employ the mapping developed in Section 11.3. This does mean that we are (for the moment) restricted to $n = 4$. The result of this transformation, carried out in Section 12.4, for (12.1.1) is

$$\begin{aligned} H_h = -t \sum_j \Big[& c^\dagger_{j\uparrow} c_{j+1\uparrow} \left(1 - n_{j\downarrow} - n_{j+1\downarrow} + 2 n_{j\downarrow} n_{j+1\downarrow}\right) + c^\dagger_{j\uparrow} c^\dagger_{j+1\uparrow} \left(2 n_{j\downarrow} n_{j+1\downarrow} - n_{j\downarrow} - n_{j+1\downarrow}\right) \\ & + i c^\dagger_{j\uparrow} c_{j+1\downarrow} \left(1 - n_{j\downarrow}\right) n_{j+1\uparrow} + c^\dagger_{j\downarrow} c_{j+1\uparrow} n_{j\uparrow} \left(-1 + (1-i) n_{j+1\downarrow}\right) \\ & + (1+i) c^\dagger_{j\downarrow} c^\dagger_{j+1\uparrow} \left(n_{j\uparrow} n_{j+1\downarrow}\right) - i c^\dagger_{j\downarrow} c_{j+1\downarrow} \left(n_{j\uparrow} n_{j+1\uparrow}\right) + \text{H.c.} \Big]. \end{aligned}$$

The only processes in this Hamiltonian are hopping and 'Cooper pairing' (with and without spin flip and coupled to densities), so that electron parity is conserved, as expected. Also, contrary to (11.3.5), (11.3.6), and (11.3.7), the two spins species appear in a strongly asymmetric way. In particular, the only quadratic terms that are present at all, are hopping terms for the up-spins, while the down-spin hopping term comes with a coupling to two densities, so it is a higher order process. While this looks strange from the electronic point of view, it does line up with the prediction that the central charge is unity. This is because the up-spins and down-spins represent a Fermionic degree of freedom each, and so only one of the two should be relevant.

We saw above that the Potts-inspired hopping term

$$H_{Ph} = -t \sum_j \left(d^\dagger_j (-i)^{O_j} d_{j+1} + \text{H.c.} \right)$$

was slightly simpler than H_h in the Fradkin-Kadanoff representation, and one might hope that it is also simpler in the electronic representation. We defer again to Section 12.4 for the details of the

calculation, whose result is

$$H_{Ph} = -t\sum_j \Big[c^\dagger_{j\uparrow}c_{j+1\uparrow}\left(1-n_{j\downarrow}-n_{j+1\downarrow}\right) + c^\dagger_{j\uparrow}c^\dagger_{j+1\uparrow}\left(n_{j\downarrow}-n_{j+1\downarrow}\right) + c^\dagger_{j\downarrow}c_{j+1\downarrow}\left(-n_{j\uparrow}n_{j+1\uparrow}\right) \\ + ic^\dagger_{j\uparrow}c_{j+1\downarrow}n_{j+1\uparrow} + ic^\dagger_{j\downarrow}c_{j+1\uparrow}n_{j\uparrow} - i\sum_\sigma \left(c^\dagger_{j,\sigma}c_{j+1,-\sigma} + c^\dagger_{j,\sigma}c^\dagger_{j+1,-\sigma}\right)\left(n_{j,-\sigma}n_{j+1,\sigma}\right) + \text{H.c.}\Big].$$

This indeed has more structure, in that the spin species appear somewhat more on equal footing, and many of the density couplings are simpler and of lower order. For instance there are no products of densities in the coupling of the up-spin hopping and 'Cooper pairing' terms. The down-spin hopping still only occurs in higher orders, and the model still does not fit on a single line. Let us compare the one-and two-particle (Fock parafermions that is) sectors of H_h and H_{Ph} in their electronic representations. The one Fock parafermion states $d^\dagger_j|0\rangle$ become the states $c^\dagger_{j\uparrow}|\varnothing\rangle$, and it is easy to see that both H_h and H_{Ph} project to the simple hopping Hamiltonian

$$H^{1p}_h = H^{1p}_{Ph} = -t\sum_j \left(c^\dagger_{j\uparrow}c_{j+1\uparrow} + \text{H.c.}\right).$$

This has the same eigenstates and dispersion as derived in the previous section, as expected. The two Fock parafermion sector is slightly more interesting, with the hope being the these electronic representations somehow help us understand the bound states observed in the two-particle spectrum in Figure 12.1. In the electronic language[14], the two Fock parafermion states are

$$\left\{d^\dagger_j d^\dagger_\ell|0\rangle\right\}_{j\le\ell} = \left\{c^\dagger_{j\uparrow}c^\dagger_{\ell\uparrow}|\varnothing\rangle\right\}_{j<\ell} \cup \left\{ic^\dagger_{j\uparrow}c^\dagger_{j\downarrow}|\varnothing\rangle\right\}_j,$$

so either we have two up-spins on separate sites, or we have an up-and down-spin pair on a single site. Projecting the Hamiltonians onto this subspace yields

$$H^{2p}_h = -t\sum_j \left[c^\dagger_{j\uparrow}c_{j+1\uparrow}\left(1-n_{j\uparrow}-n_{j+1\uparrow}\right) + ic^\dagger_{j\uparrow}c_{j+1\downarrow}n_{j+1\uparrow} - c^\dagger_{j\downarrow}c_{j+1\uparrow}n_{j\uparrow} + \text{H.c.}\right],$$
$$H^{2p}_{Ph} = -t\sum_j \left[c^\dagger_{j\uparrow}c_{j+1\uparrow}\left(1-n_{j\uparrow}-n_{j+1\uparrow}\right) + ic^\dagger_{j\uparrow}c_{j+1\downarrow}n_{j+1\uparrow} + ic^\dagger_{j\downarrow}c_{j+1\uparrow}n_{j\uparrow} + \text{H.c.}\right],$$

which are identical up to a single coefficient. These models allow for the following processes. If the two particles are separated, they are freely hopping up-spins, but when they meet, they can only pass each other if one of them flips its spin. It is interesting to note that an up-and down-spin pair on a single site cannot hop as a pair, but the only allowed process from there is for them to separate and both become down-spins. This prevents the bound states from being simple moving up-and down-spin pairs, and unfortunately there does not seem to be a simple way to solve the Hamiltonians in this subspace to find the bound states.

As another attempt to end up with a more favourable form of H_h, one can attempt to modify the mapping by choosing different state identifications. This is explained in Section 12.4, and a new electronic Hamiltonian $\tilde{H}_h$ is derived. Sadly, the full result is rather unwieldy, but the quadratic terms are

$$\tilde{H}^{(2)}_h = -\frac{t}{2}\sum_j\sum_{\sigma,\sigma'}\left(c^\dagger_{j,\sigma}c_{j+1,\sigma'} + \text{H.c.}\right).$$

[14]Here we see the first signs of a problem that we will have to face later, namely that the relations between the N Fock parafermion sectors and their electronic versions are not so straightforward.

This indeed involves all spins, but it is not hard to see that

$$\tilde{H}_h^{(2)} = -\frac{t}{2}\sum_j \left[\left(c_{j\uparrow}^\dagger + c_{j\downarrow}^\dagger\right)\left(c_{j+1\uparrow} + c_{j+1\downarrow}\right) + \text{H.c.}\right],$$

which is only the charge degree of freedom, and so we are back to square one, albeit one sanity check richer.

Hence, it is time for different methods. Transforming H_h and H_{Ph} into electronic models was not for nothing, because we do now have access to many more techniques, such as Fourier transformation (which did not work in Fock parafermion language), but it also represents a more tangible link to field theory. The idea now is to derive such a field theory from (12.1.1), and show using RG (Chapter 8 of [58], Chapter 12 of [67], or Chapter 14 of [81]) that this flows to the free Fermion fixed point, providing analytic evidence for the hypothesis that $c = 1$. Recall that the basic model was

$$H_h = -t\sum_{j=1}^{L-1}\left(d_j^\dagger d_{j+1} + \text{H.c.}\right),$$

which in electronic language (see again Section 12.4) becomes

$$d_j^\dagger d_{j+1} = \left(c_{j\uparrow}^\dagger - c_{j\uparrow}^\dagger n_{j\downarrow} + c_{j\uparrow}n_{j\downarrow} - c_{j\downarrow}^\dagger n_{j\uparrow}\right)\times\left(c_{j+1\uparrow}^\dagger - c_{j+1\uparrow}n_{j+1\downarrow} - c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + ic_{j+1\downarrow}n_{j+1\uparrow}\right). \tag{12.2.1}$$

We would have liked to do Bosonisation [28] to obtain a field theory, but we could not easily find a point around which the dispersion is approximately linear. Instead we take the continuum limit, therefore we first need to fix the scaling dimension of the fields ψ_σ, and then compensate with some power of the lattice constant a to create a dimensionless field, so that we can identify $a^d\psi_\sigma(ja) \sim c_{j\sigma}$. To this end, recall that the action will contain (see (7.10) in [81]) the dynamical term

$$\int \mathrm{d}\tau\mathrm{d}x\ \overline{\psi}_\sigma\partial_\tau\psi_\sigma,$$

which should be dimensionless in natural units ($\hbar = 1$). Thus, using square brackets to denote the scaling dimension of an object, we must have that

$$[\tau]\times[x]\times[\psi_\sigma]^2\times[\partial_\tau] = L\times L\times[\psi_\sigma]^2\times L^{-1} = 1,$$

implying $[\psi_\sigma] = L^{-1/2}$. It follows that the correct identification for the continuum limit is $c_{j\sigma} \sim \sqrt{a}\psi_\sigma(ja)$. We now determine the continuum limit of each individual term arising from the product in $d_j^\dagger d_{j+1}$ (16 in total). Call the four terms on the left A to D, and the terms on the right i to iv. The first one, Ai, is

$$\begin{aligned}-t\sum_j\left(c_{j\uparrow}^\dagger c_{j+1\uparrow} + c_{j+1\uparrow}^\dagger c_{j\uparrow}\right) &= -t\sum_j c_{j\uparrow}^\dagger\left(c_{j+1\uparrow} + c_{j-1\uparrow}\right),\\ &\sim -t\sum_j a\overline{\psi}_\uparrow(ja)\left[\psi_\uparrow(ja+a) + \psi_\uparrow(ja-a)\right].\end{aligned}$$

We now perform a second order Taylor expansion, as a will go to zero, to find

$$Ai \sim -t\sum_j a\overline{\psi}_\uparrow(ja)\left[2\psi_\uparrow(ja) + a^2\partial_{ja}^2\psi_\uparrow(ja) + \mathcal{O}(a^4)\right].$$

We now send a to zero, in such a way that

$$\sum_j a \to \int \mathrm{d}x,\ ja \to x.$$

This yields

$$Ai \sim -2t \int \mathrm{d}x\ \overline{\psi}_\uparrow \psi_\uparrow - (ta^2) \int \mathrm{d}x\ \overline{\psi}_\uparrow \partial_x^2 \psi_\uparrow + \ldots \tag{12.2.2}$$

Under RG, the coefficients in front of these terms are allowed to run, and we determine their scaling dimension using naive powercounting. Recall that the action contains an additional integral over τ (we suppressed the field's dependence on τ for notational brevity), so that the powercounting gives for the first term

$$[\lambda] \times [\tau] \times [x] \times [\psi]^2 = [\lambda] \times L = 1 \ \Rightarrow\ [\lambda] = L^{-1},$$

where λ is a generic symbol for the coefficient. Thus, this term is relevant under RG, and the coefficient scales like an energy, which is consistent with the units of t. For the second term, the powercounting goes as

$$[\lambda] \times L \times [\partial_x^2] = [\lambda] \times L \times L^{-2} = 1 \ \Rightarrow\ [\lambda] = L^1,$$

so that the second term is irrelevant. We see that derivatives make a term less relevant, and as all the terms in the ellipses of (12.2.2) contain more derivatives, they are all irrelevant. Hence we can consistently ignore higher order terms in Taylor expansions, as they will quickly become irrelevant.

We continue with the next term, which is Aii. We will write $n_\sigma(x) = \overline{\psi}_\sigma(x)\psi_\sigma(x)$ for the spin σ density field. If there is no spatial variable written for a field, it is evaluated in x (identified with ja). Here we have

$$\begin{aligned} t\sum_j \left(c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j+1\downarrow} + c_{j+1\uparrow}^\dagger c_{j\uparrow} n_{j+1\downarrow} \right) &= t\sum_j c_{j\uparrow}^\dagger \left(c_{j+1\uparrow} n_{j+1\downarrow} + c_{j-1\uparrow} n_{j+1\downarrow} \right), \\ &\sim t\sum_j a^2 \overline{\psi}_\uparrow \left\{ \left[\psi_\uparrow + a\partial_x \psi_\uparrow \right] \left[n_\downarrow + a\partial_x n_\downarrow \right] + n_\downarrow \left[\psi_\uparrow - a\partial_x \psi_\uparrow \right] \right\}, \\ &\sim (2ta) \int \mathrm{d}x\ n_\uparrow n_\downarrow + (ta^2) \int \mathrm{d}x\ n_\uparrow \partial_x n_\downarrow + \ldots \end{aligned}$$

Notice that $[n_\sigma] = -1$, thus the first term is marginal, and the second is irrelevant (again so are the terms in the ellipses). In short, the presence of densities n_σ decreases the relevancy of a term, thus we can ignore any contribution in (12.2.1) that contains more than one such density. This leaves only Ai to Aiv and Bi, all others will only give irrelevant terms. Thus we only have three more calculations to do, the first of which is $Aiii$. As before,

$$t\sum_j \left(c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + c_{j+1\uparrow} c_{j\uparrow} n_{j+1\downarrow} \right) = t\sum_j \left(c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger - c_{j\uparrow} c_{j+1\uparrow} \right) n_{j+1\downarrow}.$$

We focus only on the term with the daggers, as the second term is similar. Hence

$$\begin{aligned} t\sum_j c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j+1\downarrow} &\sim (ta) \int \mathrm{d}x\ \overline{\psi}_\uparrow \left[\overline{\psi}_\uparrow + a\partial_x \overline{\psi}_\uparrow \right] \left[n_\downarrow + a\partial_x n_\downarrow \right], \\ &\sim (ta) \int \mathrm{d}x\ \left(\overline{\psi}_\uparrow \right)^2 \left[n_\downarrow + a\partial_x n_\downarrow \right] + (ta^2) \int \mathrm{d}x\ \left(\overline{\psi}_\uparrow \partial_x \overline{\psi}_\uparrow \right) n_\downarrow + \ldots \end{aligned}$$

The first term vanishes due to the Fermionic nature of the field, and for the second term the power-counting is

$$[\lambda] \times [\tau] \times [x] \times [\psi_\uparrow] \times [\partial_x] \times [\psi_\uparrow] \times [n_\downarrow] = [\lambda] \times L^{-1} = 1 \;\Rightarrow\; [\lambda] = L,$$

so that it and all higher order terms are irrelevant. Similarly, one can investigate Aiv and it only produces irrelevant contributions. We arrive at Bi, the last one we need to consider,

$$\begin{aligned} t\sum_j \left(c^\dagger_{j\uparrow}c_{j+1\uparrow}n_{j\downarrow} + c^\dagger_{j+1\uparrow}c_{j\uparrow}n_{j\downarrow}\right) &= t\sum_j n_{j\downarrow}\left(c^\dagger_{j\uparrow}c_{j+1\uparrow} + c^\dagger_{j+1\uparrow}c_{j\uparrow}\right), \\ &\sim (ta)\int \mathrm{d}x\; n_\downarrow \left(2\overline{\psi}_\uparrow\psi_\uparrow + a\overline{\psi}_\uparrow\partial_x\psi_\uparrow + a(\partial_x\overline{\psi}_\uparrow)\psi_\uparrow + \dots\right), \end{aligned}$$

and the same powercounting arguments as before give that the first term is marginal, and all other irrelevant. Thus the 'effective' field theory in the long wavelength limit is

$$S = \lambda_1 \int \mathrm{d}\tau \mathrm{d}x\; n_\uparrow\left(1 + \lambda_2 n_\downarrow\right).$$

If we instead desire an 'effective' second quantised Hamiltonian, we could keep only those terms that produce relevant or marginal operators under RG, as outlined above. This gives the model

$$\begin{aligned} H_{\text{eff}} &= -t\sum_j \left(c^\dagger_{j\uparrow}c_{j+1\uparrow} - c^\dagger_{j\uparrow}c_{j+1\uparrow}n_{j+1\downarrow} - c^\dagger_{j\uparrow}c_{j+1\uparrow}n_{j\downarrow} + \text{H.c.}\right), \\ &= -t\sum_j \left[1 - n_{j\downarrow} - n_{j+1\downarrow}\right]\left(c^\dagger_{j\uparrow}c_{j+1\uparrow} + \text{H.c.}\right). \end{aligned} \tag{12.2.3}$$

Up to marginal terms, this is simply a tight-binding model of free Fermions, so it seems safe to conclude (keeping also the numerical evidence in mind) that indeed $c = 1$. We see that the down-spin electrons in this model are completely stationary, they acts as a sort of potential for the up-spin electrons to move in. If an up-spin electron on site j wants to hop to site $j+1$, this is somewhat penalised if a down-spin is present on either site j or $j+1$, and strongly penalised if there are down-spins on both sites. Naively one could think of the down-spins as being localised bumps in a potential that the up-spin electrons feel. Let us take a look at the few-particle sectors, like we did in Section 12.1. If there is a single up-spin and a single down-spin, we essentially face a single-particle scattering problem. Suppose that the down-spin is located at some site ℓ, then there is an obvious bound state $c^\dagger_{\ell\uparrow}c^\dagger_{\ell\downarrow}\left|0\right\rangle$, which is an eigenstate of the effective Hamiltonian with energy zero. With a single down-spin localised at a specific site but multiple up-spins, the model is essentially a chain hosting spinless Fermions and with a 'weak link' in the chain at site ℓ, as here the hopping amplitude drops to zero. This can be solved with elementary techniques, see for instance Chapter 6 in [31]. In general, one can view the problem as a chain of non-interacting spinless fermions, where the hopping between 'ordinary' sites (those without down-spins around them) occurs with amplitude t, and then there are 'weak' links with hopping amplitude zero (between a site with and a site without a down-spin), and 'very weak' links with hopping amplitude $-t$ (between two sites with a down-spin). Since the spinless Fermions are non-interacting, this could in principle be solved with elementary scattering theory (as outlined in [31]). As a final remark on this model, one might be tempted to think of the term in square brackets in (12.2.3) as a spin 1 degree of freedom (as it takes on the same values) living on the bond lattice, but this is in fact not so natural, as the value zero has multiplicity two. Moreover, under this mapping, the Fock parafermion number operator (11.3.2) becomes

$$O_j = n_{j\uparrow} - 2n_{j\uparrow}n_{j\downarrow} + 3n_{j\downarrow},$$

which is rather complicated as it contains a non-linear term. Thus, it is difficult to translate between Fock parafermion density and the up-and down-spin densities. One can also check that H_{Ph} produces the same relevant and marginal terms (only some coefficients are different).

As an attempt to remedy all this, one can also try to construct a mapping from $\mathbb{Z}_4$ Fock parafermions to something other than electrons. Since a pair of spinless Fermions would be essentially the same thing, we instead express d_j in terms of a spinless fermion c_j and a Weyl hard-core Boson b_j with $b_j^2 = 0$. One can then construct the mapping using the same general ideas as before, which is done in Section 12.4, resulting in the Hamiltonian $\overline{H}_h$ in (12.4.2). The number operator then becomes entirely linear,

$$O_j = n_j^f + 2n_j^b,$$

where n^f and n^b are the Fermion and Boson density respectively. The second order terms $\overline{H}_h^{(2)}$ contain only hopping for the fermion, as desired. The third order terms $\overline{H}_h^{(3)}$ give BCS-like terms which allows a pair of Fermions to become a Boson and vice versa. The fourth order terms $\overline{H}_h^{(4)}$ contain simultaneous Fermion and Boson hopping (in opposite directions) and another BCS-like term, but this time coupled to the Boson density. Finally, the fifth order term $\overline{H}_h^{(5)}$ is a BCS-like term that combines the properties of the other two. One can again take the continuum limit and perform the powercounting analysis to find the relevant and marginal terms in this Hamiltonian, to arrive at

$$\overline{H}_{\text{eff}} = -t \sum_j \left[1 - 2n_j^b\right] \left(c_j^\dagger c_{j+1} + \text{H.c.}\right).$$

This is indeed significantly nicer, as the factor in square brackets is now local, and can be considered as a spin $S_j^z = 1 - 2n_j^b$ (this time the values all have the same multiplicity). While one could interpret this model again as a sort of scattering problem, there is a much more elegant solution. The number of Bosons is obviously conserved by $\overline{H}_{\text{eff}}$, so that given a list of values for the n_j^b, we obtain a list of signs S_j^z that do not change. These signs can then just be gauged away by exploiting the U(1) symmetry of the Fermion algebra, by sending

$$c_j \mapsto \left[\prod_{k<j} \left(1 - 2n_k^b\right)\right] c_j.$$

This is inspired by equation (19) in the Supplementary Material to [79]. Under this transformation, $\overline{H}_{\text{eff}}$ becomes the standard tight-binding model with exclusively hopping terms, and the dispersion is $-2t\cos(k)$, while the Boson band is flat with value zero. In Section 12.4 it is also derived that $O_j = n_j^f + 2n_j^b$, and so

$$O_{\text{tot}} = \sum_j O_j = \sum_j \left(n_j^f + 2n_j^b\right).$$

There are now several different density regimes to consider for the groundstate. If the density is very low, that is $O_{\text{tot}}/L < 1/2$ (note that the maximal density is 3), one would expect the system to only contain Fermions, as they have negative-energy states available to them. This lasts until $O_{\text{tot}}/L = 1/2$, because then these states are all occupied. As the other Fermion states have positive energy, one would now expect the Boson states to become occupied, as they have zero energy. This comes with a huge degeneracy, as all Boson states have zero energy (while the negative energies for the Fermion states were different or had only double degeneracy), so that a given number n_{tot}^b of Bosons can be placed into the chain in L choose n_{tot}^b ways. Finally, for densities above 5/2 (recall that a Boson counts as two Fock parafermions), one would expect the remaining Fermion states to

become filled. Unfortunately, this does not line up with the numerics, and so this is too drastic of a simplification. Specifically, the conflict arises when we set for instance $O_{\text{tot}}/L = 1$. According to our naive model, all the negative energy Fermion states should be filled and a quarter of the Boson states should be filled. Thus we have the probabilities p_m of being in state m on some site (in the bulk),

$$\begin{aligned} p_0 &= \frac{1}{2}\frac{3}{4} = \frac{3}{8} = 0.375, \\ p_1 &= \frac{1}{2}\frac{3}{4} = \frac{3}{8}, \\ p_2 &= \frac{1}{2}\frac{1}{4} = \frac{1}{8} = 0.125, \\ p_3 &= \frac{1}{2}\frac{1}{4} = \frac{1}{8}, \end{aligned}$$

due to the particles all being non-interacting, and there being (approximate) translation symmetry These indeed add up to unity, but the numerics (see Figure 12.7) indicate that rather

$$\tilde{p}_0 \approx 0.26,\ \tilde{p}_1 \approx 0.495,\ \tilde{p}_2 \approx 0.222,\ \tilde{p}_3 \approx 0.022.$$

These are not only wildly different values, but it also fails completely to hold that $\tilde{p}_0 = \tilde{p}_1$ and $\tilde{p}_2 = \tilde{p}_3$. One natural way to extend this approach would be to perform perturbation theory in the least irrelevant terms that arise in $\overline{H}_h$, which are those in $\overline{H}_h^{(3)}$, scaling as $L^{-1/2}$. Working with the continuum field theory itself is probably rather difficult, as we would need to impose the constraints of fixed Fock parafermion number and the hard-core Boson relation ($b^2 = 0$) using delta functions in the path integral measure.

Let us return for a second of the mystery of the gapped phase occurring for $n = 3$ at half-filling ($O - \text{tot}/L = 1$) presented in Figure 12.4 on the left. It is well known that H_h at half-filling is gapless for $n = 2$ (the free Fermion case) and it follows from Figure 12.4 (on the right) that it is also gapless for $n = 4$ (for higher $n = 6$ see FIG. 4 in [74], and of course for Bosons the system is also gapless), so what is special about $n = 3$? Perhaps the answer lies instead in considering the three cases $n = 2, 3, 4$ at the 'absolute filling' $O_{\text{tot}}/L = 1$ rather than at the 'relative' half-filling. Under this assumption, the free Fermion case $n = 2$ becomes a trivial band insulator, so it is now 'gapped'. However, the $n = 4$ case remains gapless (as is the free electron hopping model at this density). From Figure 12.7, we know that the occupation probabilities for sites in the groundstate are very similar for $n = 3$ and $n = 4$ (in the later case $\tilde{p}_3$ is very close to zero), suggesting that the states are similar. As the Hamiltonians are formally the same, the difference must lie in the Hilbert space. For $n = 3$, the average density per site is close to unity, and the probability of a doubly occupied site is significant, and this is the maximal occupation, so it presents a 'blockade' to a third particle, possibly gapping the system. It seems likely that the extra room on-site provided by larger n prevents the formation of a gap through this mechanism, and indeed all those systems are gapless.

As a possible direction for further research, one could try to generalise the procedure above to all even n, instead of only $n = 4$. In [21], Section V, a fermionisation procedure for parafermions with even n is developed. It would be interesting to extend to Fock parafermions and analyse the Fermionic field theories that arise from H_h through this mapping.

12.3 Exploring the Phase Diagram

Since we are primarily interested in $n \geq 3$, since $n = 2$ are just Fermions, there are other natural terms[15] one could write down and add to (12.1.1). One such family are 'multiple hoppings' $d_j^{\dagger m} d_{j+1}^m$,

[15] In general, there are many combinations of Fock parafermion creation and annihilation operators that one could write, but we restrict attention to those that preserve the total Fock parafermion number and involve only two neighbouring sites.

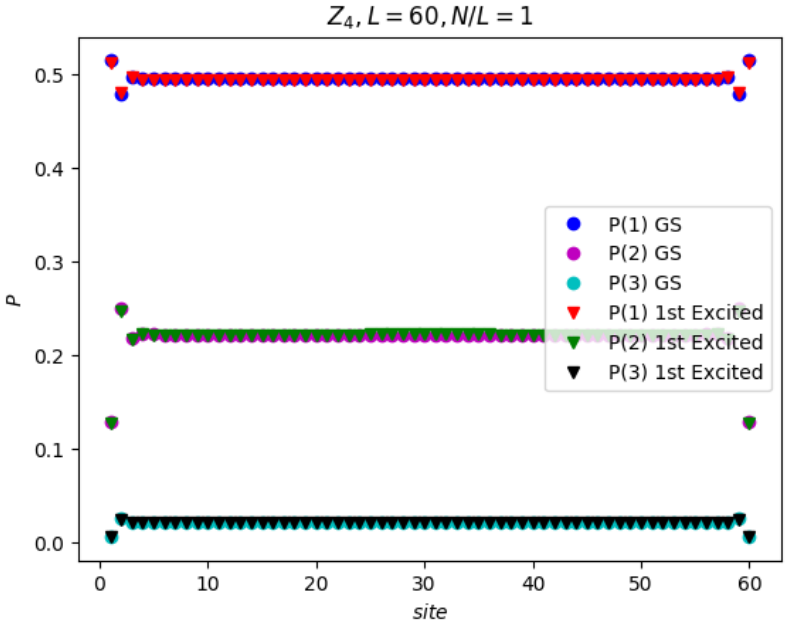


Figure 12.7: A DMRG calculation [54] on 60 sites of the occupation probabilities for the groundstate and first excited states of (12.1.1) for $n = 4$ and $O_{\text{tot}}/L = 1$ (third-filling). Notice that this implies that the Fock parafermions are distributed almost homogeneously, that is $\langle O_j \rangle$ is very close to unity for j in the bulk. Moreover, $\tilde{p}_0 \approx 0.26$ (not shown), and so $\tilde{p}_3 \approx 0.022$ is an order of magnitude smaller than the other probabilities. Analogous calculations (not shown) for $n = 3$ at $O_{\text{tot}}/L = 1$ (half-filling) produce similar values for $\tilde{p}_1$ and $\tilde{p}_2$ (and thus $\tilde{p}_0$).

where m-tuples of Fock parafermions hop in unison. One could consider the exponent m to take any value between 1 and $n-1$, but let us be modest and only consider $m = 2$ for the cases that $n = 3, 4$. Thus we define the new line of Hamiltonians

$$H_h(g) = -(1-g)t \sum_j \left(d_j^\dagger d_{j+1} + \text{H.c.} \right) - gt \sum_j \left(d_j^{\dagger 2} d_{j+1}^2 + \text{H.c.} \right), \tag{12.3.1}$$

where $0 \leq g \leq 1$. For $g = 0$, we recover H_h, and

$$H_{2h} = H_h(1) = -t \sum_j \left(d_j^{\dagger 2} d_{j+1}^2 + \text{H.c.} \right).$$

In this section, we will try to gather some information about the phase diagram in the $\rho = O_{\text{tot}}/L$ and g plane. Recall from the first section that we only need to consider densities up to half-filling due to symmetry.

First, let us think a little bit about the $g = 1$ case of this new Hamiltonian. For $n = 4$, it is not too hard to see from Equations (10.1.4) and (10.1.5) that the operators d_ℓ^2 satisfy the relations

$$d_j^2 d_\ell^2 = d_\ell^2 d_j^2, \ (d_\ell^2)^2 = 0, \ (d_\ell^2)^\dagger d_\ell^2 + d_\ell^2 (d_\ell^2)^\dagger = I,$$

etcetera. Thus, they are hard-core Bosons with maximal occupation of one particle, and H_{2h} for $n = 4$ is a model of free hopping hard-core Bosons. If instead $n = 3$, things are more subtle. Observe that both d_j^2 and $d_j^{\dagger 2}$ annihilate any state containing $|1\rangle_j$. Since this is a hopping model, one expects there to be negative energy states (compare with Figure 12.1), and thus singly-occupied sites should be avoided in the low-energy sector (at least at lower densities). This inspires us to project H_{2h} onto

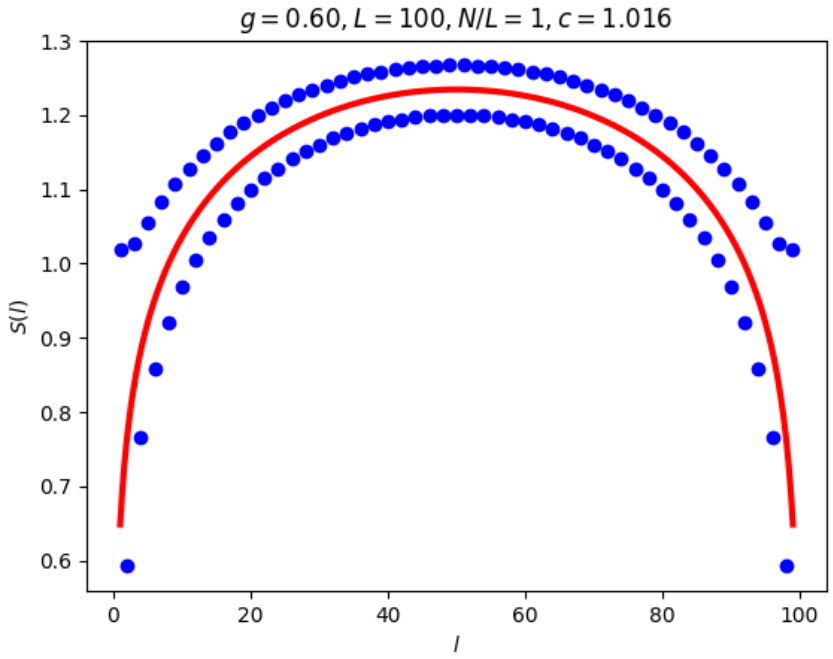


Figure 12.8: Entanglement entropy from DMRG calculations [54] for (12.3.1) at half-filling with $g = 0.6$ (denoted by a cross in Figure 12.9 on the left) and $n = 3$ (dots), and the best fit of (12.1.3) (line). Notice the even-odd effect, which is often a sign of dimerisation. The magnitude of the jumps depend on the value of g (not shown).

the subspace without singly occupied sites. We can do this as follows, apply the Fradkin-Kadanoff transformation (10.2.5) to the operator

$$d_j^{\dagger 2} d_{j+1}^2 = B_j^{\dagger 2} U_j^2 B_{j+1}^2,$$

and use the matrix representations

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \; U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \overline{\omega} \end{pmatrix},$$

where $\omega = \exp(2\pi i/3)$, to find

$$d_j^{\dagger 2} d_{j+1}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

From here, the projection we want is to cut out the middle row and column of the matrices in the tensor product, yielding

$$d_j^{\dagger 2} d_{j+1}^2 \Big|_{\text{eff}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

These are the same matrices as the Pauli (spin-1/2) raising and lowering operators $\sigma^{\pm}$, and so we can write

$$d_j^{\dagger 2} d_{j+1}^2 \Big|_{\text{eff}} = \sigma_j^+ \sigma_{j+1}^-,$$

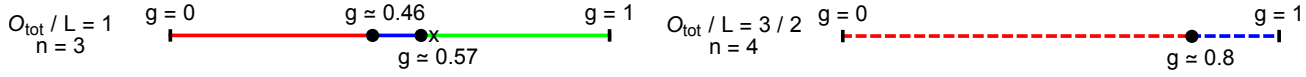


Figure 12.9: Phases of $H_h(g)$ on the half-filling lines in the phase diagrams, determined through DMRG [54]. The lines on the right are dashed to indicate a lack of precise data. The cross in the image on the left refers to Figure 12.8. For an explanation of the phases, see the last paragraph of this section.

if we identify $|0\rangle \leftrightarrow |\downarrow\rangle$ and $|2\rangle \leftrightarrow |\uparrow\rangle$. Thus, the effective Hamiltonian for H_{2h} becomes

$$H_{2h}^{(n=3)}\Big|_{\text{eff}} = -t\sum_j \left(\sigma_j^+\sigma_{j+1}^- + \sigma_j^-\sigma_{j+1}^+\right) = -t\sum_j \left(\sigma_j^x\sigma_{j+1}^x + \sigma_j^y\sigma_{j+1}^y\right),$$

which is the XX-model. It is well-known that this has $c = 1$.

For the phase diagram, it turns out from DMRG simulations [54] that (12.3.1) is gapless everywhere, for both $n = 3, 4$, unless we are at half-filling. In Figure 12.9, the information known at the time of writing about these lines in the respective phase diagrams is drawn. Let us first discuss the case $n = 4$, where we have less information. Most of the line is gapless, with $c = 1$, as one would expect from the previous sections. However, around $g \approx 0.8$ a gap appears to open, which is somewhat unexpected, but further research is needed to say anything more precise. The case $n = 3$ is far more interesting. We know from the previous sections that the model is gapped for $g = 0$, and this phase extends out up to $g \simeq 0.46$. Above, we argued that the effective model for H_{2h} should be the XX-Hamiltonian, which is gapless with $c = 1$, and indeed this agrees with the numerics. This phase extends from $g = 1$ down to the neighbourhood of $g \simeq 0.56$, thus there is a region between the two phases we have discussed so far. This region is also gapless, but has $c = 2$ (the same as free electrons), which is different from everything we have seen so far. It is difficult to speculate on what this phase could be, and further study is needed. Lastly, the entanglement entropy in the gapless $c = 1$ phase ($0.57 < g \leq 1$ and $n = 3$), which is plotted in Figure 12.8, exhibits a rather interesting even-odd effect, for which we do not yet have an explanation.

12.4 Appendix: The Sanguine Details of Certain Other Calculations

We want to express the $n = 4$ case of the model

$$H_h = -t\sum_j \left(d_j^\dagger d_{j+1} + d_{j+1}^\dagger d_j\right)$$

in terms of electrons. For this we discussed the mapping (11.3.3) of [18] in the previous Chapter. Plugging this in yields

$$\begin{aligned} d_j^\dagger d_{j+1} = &\left(c_{j\uparrow}^\dagger - c_{j\uparrow}^\dagger n_{j\downarrow} - c_{j\uparrow}n_{j\downarrow} - ic_{j\downarrow}^\dagger n_{j\uparrow}\right) i^{n_{j\downarrow}-2n_{j\uparrow}n_{j\downarrow}-n_{j\uparrow}} \\ &\times \left(c_{j+1\uparrow} - c_{j+1\uparrow}n_{j+1\downarrow} - c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + ic_{j+1\downarrow}n_{j\uparrow}\right). \end{aligned}$$

We begin by simplifying the first line, getting rid of the phase operator. To this end, group

$$c_{j\uparrow}^\dagger - c_{j\uparrow}^\dagger n_{j\downarrow} - c_{j\uparrow}n_{j\downarrow} - ic_{j\downarrow}^\dagger n_{j\uparrow} = c_{j\uparrow}^\dagger(1-n_{j\downarrow}) - c_{j\uparrow}n_{j\downarrow} - ic_{j\downarrow}^\dagger n_{j\uparrow},$$

then

$$\begin{aligned} c_{j\uparrow}^\dagger(1-n_{j\downarrow})i^{n_{j\downarrow}-2n_{j\uparrow}n_{j\downarrow}-n_{j\uparrow}} &= c_{j\uparrow}^\dagger(1-n_{j\downarrow})i^{0-2\times0\times0-0} = c_{j\uparrow}^\dagger(1-n_{j\downarrow}), \\ c_{j\uparrow}n_{j\downarrow}i^{n_{j\downarrow}-2n_{j\uparrow}n_{j\downarrow}-n_{j\uparrow}} &= c_{j\uparrow}n_{j\downarrow}i^{1-2\times1\times1-1} = -c_{j\uparrow}n_{j\downarrow}, \\ c_{j\downarrow}^\dagger n_{j\uparrow}i^{n_{j\downarrow}-2n_{j\uparrow}n_{j\downarrow}-n_{j\uparrow}} &= c_{j\downarrow}^\dagger n_{j\uparrow}i^{1-2\times1\times0-0} = -ic_{j\downarrow}^\dagger n_{j\uparrow}. \end{aligned}$$

Leading to

$$d_j^\dagger d_{j+1} = \left(c_{j\uparrow}^\dagger - c_{j\uparrow}^\dagger n_{j\downarrow} + c_{j\uparrow} n_{j\downarrow} - c_{j\downarrow}^\dagger n_{j\uparrow}\right)\left(c_{j+1\uparrow} - c_{j+1\uparrow} n_{j+1\downarrow} - c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + i c_{j+1\downarrow} n_{j\uparrow}\right). \quad (12.4.1)$$

Writing out this product and collecting the terms per order yields (using the same notation as in Section 11.5),

$$\begin{aligned}
(\dots) &= c_{j\uparrow}^\dagger c_{j+1\uparrow},\\
\{\dots\} &= -c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j+1\downarrow} - c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow} - c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j\downarrow} + c_{j\uparrow} c_{j+1\uparrow} n_{j\downarrow} - c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow},\\
[\dots] &= c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j\downarrow} n_{j+1\downarrow} + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\downarrow} n_{j+1\downarrow} - i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} - c_{j\uparrow} c_{j+1\uparrow} n_{j\downarrow} n_{j+1\downarrow} - c_{j\uparrow} c_{j+1\uparrow}^\dagger n_{j\downarrow} n_{j+1\downarrow}\\
&\quad + i c_{j\uparrow} c_{j+1\downarrow} n_{j\downarrow} n_{j+1\uparrow} + c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow} n_{j+1\downarrow} + c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger n_{j\uparrow} n_{j+1\downarrow} - i c_{j\downarrow}^\dagger c_{j+1\downarrow} n_{j\uparrow} n_{j+1\uparrow}.
\end{aligned}$$

Thus

$$H_h = H_h^{(2)} + H_h^{(4)} + H_h^{(6)},$$

with, carrying out similar replacements as in Section 11.5,

$$\begin{aligned}
H_h^{(2)} &= -t \sum_j \left(c_{j\uparrow}^\dagger c_{j+1\uparrow} + c_{j+1\uparrow}^\dagger c_{j\uparrow}\right),\\
H_h^{(4)} &= -t \sum_j \left[c_{j\uparrow}^\dagger c_{j+1\uparrow}(-n_{j\downarrow} - n_{j+1\downarrow}) + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger(-n_{j\downarrow} - n_{j+1\downarrow}) + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow} - c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow} + \text{H.c.}\right],\\
H_h^{(6)} &= -t \sum_j \Big[2 c_{j\uparrow}^\dagger c_{j+1\uparrow}(n_{j\downarrow} n_{j+1\downarrow}) + 2 c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\downarrow} n_{j+1\downarrow}) - i c_{j\uparrow}^\dagger c_{j+1\downarrow}(n_{j\downarrow} n_{j+1\uparrow})\\
&\qquad - i c_{j\downarrow}^\dagger c_{j+1\downarrow}\,(n_{j\uparrow} n_{j+1\uparrow}) + (1-i) c_{j\downarrow}^\dagger c_{j+1\uparrow}(n_{j\uparrow} n_{j+1\downarrow}) + (1+i) c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\uparrow} n_{j+1\downarrow}) + \text{H.c.}\Big].
\end{aligned}$$

We also determine the electronic representation of the Potts-like hopping term,

$$\begin{aligned}
d_j^\dagger i^{-O_j} d_{j+1} &= \left(d_j i^{O_j} d_{j+1}^\dagger\right)^\dagger,\\
&= \left(c_{j\uparrow}^\dagger c_{j+1\uparrow}\right) + \Big\{c_{j\uparrow}^\dagger c_{j+1\uparrow}(-n_{j\downarrow} - n_{j+1\downarrow}) + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\downarrow} - n_{j+1\downarrow}) + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow}\\
&\quad + i c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow}\Big\} - \Big[c_{j\downarrow}^\dagger c_{j+1\downarrow}(n_{j\uparrow} n_{j+1\downarrow}) + i c_{j\uparrow}^\dagger c_{j+1\downarrow}(n_{j\downarrow} n_{j+1\uparrow}) + i c_{j\uparrow}^\dagger c_{j+1\downarrow}^\dagger (n_{j\downarrow} n_{j+1\uparrow})\\
&\quad + i c_{j\downarrow}^\dagger c_{j+1\uparrow}(n_{j\uparrow} n_{j+1\downarrow}) + i c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\uparrow} n_{j+1\downarrow})\Big]
\end{aligned}$$

where we used (11.5.4). Thus

$$H_{Ph} = -t \sum_j \left(d_j^\dagger i^{-O_j} d_{j+1} + \text{H.c.}\right) = H_{Ph}^{(2)} + H_{Ph}^{(4)} + H_{Ph}^{(6)},$$

with

$$\begin{aligned}
H_{Ph}^{(2)} &= -t \sum_j \left(c_{j\uparrow}^\dagger c_{j+1\uparrow} + \text{H.c.}\right),\\
H_{Ph}^{(4)} &= -t \sum_j \left[c_{j\uparrow}^\dagger c_{j+1\uparrow}(-n_{j\downarrow} - n_{j+1\downarrow}) + c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\downarrow} - n_{j+1\downarrow}) + i c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow} + i c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow} + \text{H.c.}\right],\\
H_{Ph}^{(6)} &= -t \sum_j \left\{-c_{j\downarrow}^\dagger c_{j+1\downarrow}(n_{j\uparrow} n_{j+1\uparrow}) - i \sum_\sigma \left[c_{j,\sigma}^\dagger c_{j+1,-\sigma}(n_{j,-\sigma} n_{j+1,\sigma}) + c_{j,\sigma}^\dagger c_{j+1,-\sigma}^\dagger (n_{j,-\sigma} n_{j+1,\sigma})\right]\right\}.
\end{aligned}$$

Next, we consider an alternative mapping from $n = 4$ Fock parafermions to electrons. Instead of identifying the bases as in Section 11.3, we propose

$$\begin{aligned}
|0\rangle &\mapsto |\varnothing\rangle\,, \\
|1\rangle &\mapsto 2^{-1/2}(c_\uparrow^\dagger + c_\downarrow^\dagger)\,|\varnothing\rangle\,, \\
|2\rangle &\mapsto c_\uparrow^\dagger c_\downarrow^\dagger\,|\varnothing\rangle\,, \\
|3\rangle &\mapsto 2^{-1/2}(c_\uparrow^\dagger - c_\downarrow^\dagger)\,|\varnothing\rangle\,.
\end{aligned}$$

The projectors onto the even number states are still the same,

$$P(0) = (1-n_\uparrow)(1-n_\downarrow),\ P(2) = n_\uparrow n_\downarrow.$$

However, the odd number state projectors are more complicated,

$$\begin{aligned}
P(1) = |1\rangle\langle 1| &= \frac{1}{2}(c_\uparrow^\dagger + c_\downarrow^\dagger)\,|\varnothing\rangle\langle\varnothing|\,(c_\uparrow + c_\downarrow) = \frac{1}{2}\left(n_\uparrow + n_\downarrow - 2n_\uparrow n_\downarrow + c_\uparrow^\dagger c_\downarrow + c_\downarrow^\dagger c_\uparrow\right), \\
P(3) = |3\rangle\langle 3| &= \frac{1}{2}(c_\uparrow^\dagger - c_\downarrow^\dagger)\,|\varnothing\rangle\langle\varnothing|\,(c_\uparrow - c_\downarrow) = \frac{1}{2}\left(n_\uparrow + n_\downarrow - 2n_\uparrow n_\downarrow - c_\uparrow^\dagger c_\downarrow - c_\downarrow^\dagger c_\uparrow\right).
\end{aligned}$$

Constructing the operator d as in Section 11.3 yields the expression

$$d = 2^{-1/2}\left[c_\uparrow + c_\downarrow(1-2n_\uparrow) - c_\uparrow^\dagger n_\downarrow - c_\downarrow^\dagger n_\uparrow\right].$$

To turn this into d_j, we want to use the same strings as before,

$$d_j = i^{\sum_{k<j}\left(O_k + 2n_{k\uparrow} + 2n_{k\downarrow}\right)} 2^{-1/2}\left[c_{j\uparrow} + c_{j\downarrow}(1-2n_{j\uparrow}) - c_{j\uparrow}^\dagger n_{j\downarrow} - c_{j\downarrow}^\dagger n_{j\uparrow}\right],$$

but we need to find the representation of the Fock parafermion number operators O_k under this mapping. This turns out to be

$$O_k = 2\left(n_{k\uparrow} + n_{k\downarrow} - n_{k\uparrow}n_{k\downarrow}\right) - \left(c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}\right).$$

At first glance this is a disaster, as now O_k and $n_{k\sigma}$ do not commute, which makes it difficult to work with the strings. Thankfully not all is lost, as the combination

$$O_k + 2n_{k\uparrow} + 2n_{k\downarrow} = 4n_{k\uparrow} + 4n_{k\downarrow} - 2n_{k\downarrow}n_{k\uparrow} - \left(c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}\right) \equiv -2n_{k\downarrow}n_{k\uparrow} - \left(c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}\right) \pmod 4,$$

which occurs in the strings, actually consists of two commuting pieces,

$$\left[n_{k\downarrow}n_{k\uparrow},\, c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}\right] = 0.$$

Thus

$$\begin{aligned}
i^{\sum_{k<j}\left(O_k + 2n_{k\uparrow} + 2n_{k\downarrow}\right)} &= \prod_{k<j}\left\{(-1)^{n_{k\downarrow}n_{k\uparrow}}(-i)^{c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}}\right\}, \\
&= \prod_{k<j}\left\{(-1)^{n_{k\downarrow}n_{k\uparrow}}\left[P_k(0) + (-i)P_k(1) + P_k(2) + iP_k(3)\right]\right\}, \\
&= \prod_{k<j}\left\{(-1)^{n_{k\downarrow}n_{k\uparrow}}\left[1 - n_{k\uparrow} - n_{k\downarrow} + 2n_{k\uparrow}n_{k\downarrow} - i\left(c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow}\right)\right]\right\},
\end{aligned}$$

which, while cumbersome, can be worked with. In total,

$$d_j = \frac{1}{\sqrt{2}} \prod_{k<j} \left\{ (-1)^{n_{k\downarrow} n_{k\uparrow}} \left[1 - n_{k\uparrow} - n_{k\downarrow} + 2 n_{k\uparrow} n_{k\downarrow} - i \left(c_{k\uparrow}^\dagger c_{k\downarrow} + c_{k\downarrow}^\dagger c_{k\uparrow} \right) \right] \right\} \\ \times \left[c_{j\uparrow} + c_{j\downarrow}(1 - 2n_{j\uparrow}) - c_{j\uparrow}^\dagger n_{j\downarrow} - c_{j\downarrow}^\dagger n_{j\uparrow} \right].$$

With this expression and the necessary elbow grease, one can derive that

$$\begin{aligned} 2 d_j^\dagger d_{j+1} = & \left(c_{j\uparrow}^\dagger c_{j+1\uparrow} + c_{j\downarrow}^\dagger c_{j+1\uparrow} + c_{j\uparrow}^\dagger c_{j+1\downarrow} + c_{j\downarrow}^\dagger c_{j+1\downarrow} \right) \\ & + \Big\{ -(1+i) c_{j\uparrow}^\dagger c_{j+1\uparrow} n_{j\downarrow} - (1-i) c_{j\downarrow}^\dagger c_{j+1\uparrow} n_{j\uparrow} \\ & \quad + c_{j\uparrow} c_{j+1\uparrow} n_{j\downarrow} + c_{j\downarrow} c_{j+1\uparrow} n_{j\uparrow} - (1+i) c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j\downarrow} - 2 c_{j\uparrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow} - (1-i) c_{j\downarrow}^\dagger c_{j+1\downarrow} n_{j\uparrow} \\ & \quad - 2 c_{j\downarrow}^\dagger c_{j+1\downarrow} n_{j+1\uparrow} + c_{j\uparrow} c_{j+1\downarrow} n_{j\downarrow} + c_{j\downarrow} c_{j+1\downarrow} n_{j\uparrow} - c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger n_{j+1\downarrow} - c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger n_{j+1\downarrow} \\ & \quad - c_{j\uparrow}^\dagger c_{j+1\downarrow}^\dagger n_{j+1\uparrow} - c_{j\downarrow}^\dagger c_{j+1\downarrow}^\dagger n_{j+1\uparrow} \Big\} \\ & + \Big[2(1+i) c_{j\uparrow}^\dagger c_{j+1\downarrow} (n_{j\downarrow} n_{j+1\uparrow}) + 2(1-i) c_{j\downarrow}^\dagger c_{j+1\downarrow} (n_{j\uparrow} n_{j+1\uparrow}) - 2 c_{j\uparrow} c_{j+1\downarrow} (n_{j\downarrow} n_{j+1\uparrow}) \\ & \quad - 2 c_{j\downarrow} c_{j+1\downarrow} (n_{j\uparrow} n_{j+1\uparrow}) + (1+i) c_{j\uparrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\downarrow} n_{j+1\downarrow}) + (1-i) c_{j\downarrow}^\dagger c_{j+1\uparrow}^\dagger (n_{j\uparrow} n_{j+1\downarrow}) \\ & \quad - c_{j\uparrow} c_{j+1\uparrow}^\dagger (n_{j\downarrow} n_{j+1\downarrow}) - c_{j\downarrow} c_{j+1\uparrow}^\dagger (n_{j\uparrow} n_{j+1\downarrow}) + (1+i) c_{j\uparrow}^\dagger c_{j+1\downarrow}^\dagger (n_{j\downarrow} n_{j+1\uparrow}) \\ & \quad + (1-i) c_{j\downarrow}^\dagger c_{j+1\downarrow}^\dagger (n_{j\uparrow} n_{j+1\downarrow}) - c_{j\uparrow} c_{j+1\downarrow}^\dagger (n_{j\downarrow} n_{j+1\uparrow}) - c_{j\downarrow} c_{j+1\downarrow}^\dagger (n_{j\uparrow} n_{j+1\uparrow}) \Big]. \end{aligned}$$

Thus the Hamiltonian consists of the pieces

$$\begin{aligned} \tilde{H}_h^{(2)} &= -\frac{t}{2} \sum_j \sum_{\sigma,\sigma'} \left(c_{j,\sigma}^\dagger c_{j+1,\sigma'} + \text{H.c.} \right), \\ \tilde{H}_h^{(4)} &= \frac{t}{2} \sum_j \Big\{ \left[(1+i) c_{j\uparrow}^\dagger n_{j\downarrow} + (1-i) c_{j\downarrow}^\dagger n_{j\uparrow} \right] (c_{j+1\uparrow} + c_{j+1\downarrow}) + 2 (c_{j\uparrow}^\dagger + c_{j\downarrow}^\dagger) c_{j+1\downarrow} n_{j+1\uparrow} \\ & \qquad + \sum_{\sigma,\sigma'} c_{j,\sigma}^\dagger c_{j+1,\sigma'} (n_{j,-\sigma} + n_{j+1,-\sigma'}) + \text{H.c.} \Big\}, \\ \tilde{H}_h^{(6)} &= -\frac{t}{2} \sum_j \Big\{ 2 \left(c_{j\uparrow}^\dagger n_{j\downarrow} + c_{j\downarrow}^\dagger n_{j\uparrow} \right) c_{j+1\downarrow}^\dagger n_{j+1\uparrow} + 2 \left[(1+i) c_{j\uparrow}^\dagger n_{j\downarrow} + (1-i) c_{j\downarrow}^\dagger n_{j\uparrow} \right] c_{j+1\downarrow} n_{j+1\uparrow} \\ & \qquad + (1+i) c_{j\uparrow}^\dagger n_{j\downarrow} \left(c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + c_{j+1\downarrow}^\dagger n_{j+1\uparrow} \right) + (1-i) c_{j\downarrow}^\dagger n_{j\uparrow} \left(c_{j+1\uparrow}^\dagger n_{j+1\downarrow} + c_{j+1\downarrow}^\dagger n_{j+1\uparrow} \right) \\ & \qquad + \sum_{\sigma,\sigma'} c_{j,\sigma}^\dagger c_{j+1,\sigma'} (n_{j,-\sigma} n_{j+1,-\sigma'}) + \text{H.c.} \Big\}. \end{aligned}$$

Instead of electrons, one can also represent the Fock parafermions using a (spinless) Fermion c and a Weyl hard-core Boson b. These two particle species commute, so $[c, b] = 0$, etcetera. To be able to write down strings, it is important that we construct the Fock parafermion operator d in such a way that it respects Fermion parity (for the Boson it is unimportant because they commute off-site).

Thus we choose the identification

$$\begin{aligned}
|0\rangle &\mapsto |\varnothing\rangle\,, \\
|1\rangle &\mapsto c^\dagger\,|\varnothing\rangle\,, \\
|2\rangle &\mapsto b^\dagger\,|\varnothing\rangle\,, \\
|3\rangle &\mapsto f^\dagger b^\dagger\,|\varnothing\rangle\,.
\end{aligned}$$

Write $n^f = c^\dagger c$ and $n^b = b^\dagger b$ for the Fermion and Boson number operators respectively. It is not difficult to see that the projection operators are

$$P(0) = (1-n^f)(1-n^b),\ P(1) = n^f(1-n^b),\ P(2) = (1-n^f)n^b,\ P(3) = n^f n^b,$$

so that we can construct d as usual,

$$\begin{aligned}
d &= 0P(0) + cP(1) + c^\dagger bP(2) + cP(3), \\
&= c^\dagger b(1-n^f)n^b + cn^f n^b, \\
&= c + c^\dagger b.
\end{aligned}$$

This is a rather nice expression, and in fact the Fock parafermion number operator also has a favourable form,

$$\begin{aligned}
O &= 0P(0) + P(1) + 2P(2) + 3P(3), \\
&= n^f(1-n^b) + 2(1-n^f)n^b + 3n^f n^b, \\
&= n^f + 2n^b.
\end{aligned}$$

To turn d into d_j, we need to write the correct strings. We know how to cancel Fermionic off-site anti-commutation, and we do not need any string for the Boson, thus

$$d_j = i^{\sum_{k<j}\left(O_k + 2n_k^f\right)}\left(c_j + c_j^\dagger b_j\right) = i^{\sum_{k<j}\left(2n_k^b - n_k^f\right)}\left(c_j + c_j^\dagger b_j\right).$$

The string can again be split into two commuting pieces,

$$\begin{aligned}
i^{2n_k^b} &= (-1)^{n_k^b} = P_k(0) + P_k(1) - P_k(2) - P_k(3) = 1 - 2n_k^b, \\
(-i)^{n_k^f} &= P_k(0) + P_k(2) - i\left[P_k(1) + P_k(3)\right] = 1 - n_k^f - in_k^f = 1 - (1+i)n_k^f.
\end{aligned}$$

This yields a considerably simpler expression than the last mapping for

$$\begin{aligned}
d_j^\dagger d_{j+1} &= \left(c_j^\dagger + c_j b_j^\dagger\right)\left[1 - (1+i)n_j^f\right]\left(1 - 2n_j^b\right)\left(c_{j+1} + c_{j+1}^\dagger b_{j+1}\right), \\
&= \left[c_j^\dagger\left(1 - 2n_j^b\right) - ic_j b_j^\dagger\right]\left(c_{j+1} + c_{j+1}^\dagger b_{j+1}\right), \\
&= c_j^\dagger c_{j+1} + c_j^\dagger c_{j+1}^\dagger b_{j+1} - ic_j c_{j+1} b_j^\dagger - ic_j c_{j+1}^\dagger b_j^\dagger b_{j+1} - 2n_j^b c_j^\dagger c_{j+1} - 2n_j^b c_j^\dagger c_{j+1}^\dagger b_{j+1}.
\end{aligned}$$

Therefore

$$\overline{H}_h = \overline{H}_h^{(2)} + \overline{H}_h^{(3)} + \overline{H}_h^{(4)} + \overline{H}_h^{(5)}, \tag{12.4.2}$$

with

$$\begin{aligned}
\overline{H}_h^{(2)} &= -t\sum_j\left(c_j^\dagger c_{j+1} + \text{H.c}\right), \\
\overline{H}_h^{(3)} &= -t\sum_j\left[c_j^\dagger c_{j+1}^\dagger\left(b_{j+1} - ib_j\right) + \text{H.c}\right], \\
\overline{H}_h^{(4)} &= t\sum_j\left[ic_j^\dagger c_{j+1} b_{j+1}^\dagger b_j + 2n_j^b c_j^\dagger c_{j+1} + \text{H.c}\right], \\
\overline{H}_h^{(5)} &= t\sum_j\left(2n_j^b c_j^\dagger c_{j+1}^\dagger b_{j+1} + \text{H.c}\right).
\end{aligned}$$

References

[1] I. Affleck, *"Universal Term in the Free Energy at a Critical Point and the Conformal Anomaly"*, Phys. Rev. Lett. 56, 7 (1986).

[2] J. Alicea, and Paul Fendley, *"Topological Phases with Parafermions: Theory and Blueprints"*, Annu. Rev. Condens. Matter Phys. 7, 119 (2016).

[3] G. E. Andrews, *"The Theory of Partitions"*, Cambridge University Press, Cambridge (1984).

[4] R. Askey and M. Ismail, *"Recurrence Relations, Continued Fractions and Orthogonal Polynomials"* in the series *Memoirs of the American Mathematical Society*, American Mathematical Society, Providence (1984).

[5] S. Bachmann and B. Nachtergaele, *"Product Vacua with Boundary States and the Classification of Gapped Phases"*, Comm. Math. Phys. 329, 509-544 (2014).

[6] C. W. J. Beenakker, *"Random-Matrix Theory of Quantum Transport"*, Rev. Mod. Phys. 69, 3 (1997).

[7] G. Ben Arous and A. Guionnet, *"Large Deviations for Wigner's Law and Voiculescu's Non-Commutative Entropy"*, Probab. Theory Relat. Fields 108, 517-542 (1997).

[8] B. A. Bernevig with T. L. Huges, *"Topological Insulators and Topological Superconductors"*, Princeton University Press, Princeton (2013).

[9] P. Biane, M. Capitaine and A. Guionnet, *"Large Deviation Bounds for Matrix Brownian Motion"*, Invent. Math. 152(2), 433-459 (2003).

[10] B. Blackadar, *"K-Theory for Operator Algebras"* (second edition), Cambridge University Press, Cambridge (1998)

[11] H. W. J. Blöte, J. L. Cardy, and M. P. Nightingale, *"Conformal Invariance, the Central Charge, and Universal Finite-Size Amplitudes at Criticality"*, Phys. Rev. Lett. 56, 7 (1986).

[12] B. Bożejko, B. Kümmerer, and R. Speicher, *"q-Gaußian Processes: Noncommutative and Classical Aspects"*, Comm. Math. Phys. 185, 1, 129-154 (1997).

[13] M. Bożejko and R. Speicher, *"An Example of a Generalized Brownian Motion"*, Comm. Math. Phys. 137, 3, 519-531 (1991).

[14] M. Bożejko and R. Speicher, *"An example of a Generalized Brownian motion II"* in *Quantum Probability and Related Topics VII*, World Scientific, Singapore 67-77 (1992).

[15] H. Bruus and K. Flensberg, *"Many-Body Quantum Theory in Condensed Matter Physics"*, Oxford University Press, Oxford (2004).

[16] P. Calabrese, and J. Cardy, *"Entanglement Entropy and Conformal Field Theory"*, J. Phys. A: Math. Theor. 42, 504005 (2009).

[17] P. Calabrese, and M. Mintchev, *"Correlation Functions of One-Dimensional Anyonic Fluids"*, Phys. Rev. B 75, 233104 (2007).

[18] A. Calzona, T. Meng, M. Sassetti, and T. L. Schmidt, *"$\mathbb{Z}_4$ Parafermions in One-Dimensional Fermionic Lattices"*, Phys. Rev. B 98, 201110(R) (2018).

[19] M. Caspers and P. Fima, *"Graph Products of Operator Algebras"*, J. Noncommut. Geom. 11 (2017), no. 1, 367-411.

[20] Y-M. Chen, *"The Chung-Feller Theorem Revisited"*, Discrete Mathematics Volume 308 Issue 7 (2008), 1328-1329

[21] A. Chew, D. F. Mross, and J. Alicea, *"Fermionized Parafermions and Symmetry-Enriched Majorana Modes"*, Phys. Rev. B 98, 085143 (2018).

[22] E. Cobanera and G. Ortiz, *"Fock Parafermions and Self-Dual Representations of the Braid Group"*, Phys. Rev. A 89, 012328 (2014).

[23] E. Cobanera, *"Modeling Electron Fractionalization with Unconventional Fock Spaces"*, J. Phys.: Condens. Matter 29, 305602 (2017).

[24] J. B. Conway, *"A Course in Functional Analysis"* (second edition), Springer Verlag, New York (2007).

[25] T. M. Cover and J. A. Thomas, *"Elements of Information Theory"*, John Wiley & Sons, Inc. (1991).

[26] Y. Dabrowski, *"A Free Stochastic Partial Differential Equation"*, Ann. Inst. Henri Poincaré B, Probab. Stat. Vol. 50 No. 4 (2014), 1404-1455.

[27] Y. Dabrowski, A. Guionnet, and D. Shlyakhtenko, *"Free Transport for Convex Potentials"*, Preprint, arXiv:1701.00132v1 (2016).

[28] J. von Delft and H. Schoeller, *"Bosonization for Beginners - Refermionization for Experts"*, Ann. Phys. (Leipzig) 7, 4, 225-306 (1998).

[29] K. Dykema, *"Interpolated Free Group Factors"*, Pacific J. Math. Volume 163 Number 1 (1994), 123-135.

[30] K. Dykema and A. Nica, *"On the Fock Representation of the q-Commutation Relations"*, Journal für die reine und angewandte Mathematik 440 (1993), 201-212.

[31] F. E. Elzinga, *"Mesoscopic Electronic Transport through Ring-Shaped Nano-Structures"* (Bachelor Thesis), Utrecht University Repository, https://dspace.library.uu.nl/handle/1874/336066 (2016).

[32] L. C. Evans, *"Partial Differential Equations"* (second edition) in the series *Graduate Studies in Mathematics*, American Mathematical Society, Providence (2010).

[33] M. Fannes, B. Nachtergaele, and R. F. Werner, *"Finitely Correlated States on Quantum Spin Chains"*, Comm. Math. Phys. 144, 443-490 (1992).

[34] P. Fendley, *"Parafermionic Edge Zero Modes in $\mathbb{Z}_n$-Invariant Spin Chains"*, J. Stat. Mech. P11020 (2012).

[35] P. Fendley, *"Free Parafermions"*, J. Phys. A: Math. Theor. 47, 075001 (2014).

[36] E. Fradkin, *"Field Theories of Condensed Matter Physics"* (second edition), Cambridge University Press, Cambridge (2013).

[37] E. Fradkin and L. P. Kadanoff, *"Disorder Variables and Para-Fermions in Two-Dimensional Statistical Mechanics"*, Nucl. Phys. B 170, 1 (1980).

[38] K. Fredenhagen, *"On the Existence of Antiparticles"*, Comm. Math. Phys. 79, 141-151 (1981).

[39] O. W. Greenberg, *"Example of Infinite Statistics"*, Phys. Rev. Letters 64, 7 (1990).

[40] O. W. Greenberg, *"Particles With Small Violations of Fermi or Bose Statistics"*, Phys. Rev. D 43, 12 (1991).

[41] O. W. Greenberg, *"Interactions of Particles Having Small Violations of Statistics"*, Physica A 180, 419-427 (1992).

[42] A. Guionnet and E. Maurel-Segala, *"Combinatorial Aspects of Matrix Models"*, ALEA Lat. Am. J. Probab. Math. Stat. 1 (2006), 241279.

[43] A. Guionnet and D. Shlyakhtenko, *"Free Montone Transport"*, Invent. math. 197, 3, 613-661 (2014).

[44] F. Haake, S. Gnutzmann, and M. Kuś, *"Quantum Signatures of Chaos"* (fourth edition), Springer Nature Switzerland, Cham (2018).

[45] M. Hartglass, and B. Nelson, *"Free Transport for Interpolated Free Group Factors"*, J. Funct. Anal. 274, 1, 222-251 (2018).

[46] F. Iemini, C. Mora, and L. Mazza, *"Topological Phases of Parafermions: a Model with Exactly-Solvable Ground States"*, Phys. Rev. Lett. 118, 170402 (2017).

[47] A. Kitaev, *"Unpaired Majorana Fermions in Quantum Wires"* Sov. Phys.-Usp. 44, 131 (2001).

[48] A. Kitaev, *"Fault-Tolerant Quantum Computation by Anyons"*, Ann. Phys. 303, 2 (2003).

[49] S. Lang, *"Complex Analysis"* (fourth edition), Springer-Verlag, New York, (1999).

[50] M. Lemm and B. Nachtergaele, *"The PVBS Models are Gapped for all Species Numbers and all Dimensions"*, Preprint, arXiv:1902.09678v1 (2019).

[51] N. H. Lindner, E. Berg, G. Rafael, and A. Stern, *"Fractionalizing Majorna Fermions: Non-Abelian Statistics on the Edges of Abelian Quantum Hall States"*, Phys. Rev. X 2, 041002 (2012).

[52] G. Livan, M. Novaes, and P. Vivo, *"Introduction to Random Matrices: Theory and Practice"*, Springer International Publishing (2018).

[53] R. M. Lutchyn, J. D. Sau, and S. Das Sarma, *"Majorana Fermions and a Topological Phase Transition in Semiconductor-Superconductor Heterostructures"*, Phys. Rev. Lett. 105, 077001 (2010).

[54] I. Mahyaeh (Stockholm), private communication.

[55] H. N. Mhaskar and E. B. Saff, *"Where Does the Sup Norm of a Weighted Polynomial Live?"*, Constr. Approx. Volume 1 Issue 1 (1985), 71-91.

[56] J.A. Mingo and R. Speicher, *"Free Probability and Random Matrices"* in the series *Fields Institute Monographs*, Springer New York (2017).

[57] G.J. Murphy, *"C*-Algebras and Operator Theory"*, Academic Press, Inc. San Diego (1990).

[58] G. Mussardo, *"Statistical Field Theory: An Introduction to Exactly Solved Models in Statistical Physics"*, Oxford University Press, Oxford (2010).

[59] B. Nachtergaele, *"Quantum spin systems after DLS 1978"* (English summary), Proc. Sympos. Pure Math. 76, Part 1 (2007).

[60] C. Nayak, S. H. Simon, A. Stern, M. Freedman, and S. Das Sarma, *"Non-Abelian Anyons and Topological Quantum Computation"*, Rev. Mod. Phys. 80, 1083 (2008).

[61] B. Nelson, *"Free Monotone Transport without a Trace"*, Comm. Math. Phys. 334, 3, 1245-1298 (2015).

[62] B. Nelson, *"Free Transport for Finite Depth Subfactor Planar Algebras"*, J. Funct. Anal. 268, 9, 2586-2620 (2015).

[63] B. Nelson, and Q. Zeng, *"An Application of Free Transport to mixed q-Gaussian Algebras"*, Proc. Amer. Math. Soc. 144, 10, 4357-4366 (2016).

[64] B. Nelson, and Q. Zeng, *"Free Monotone Transport for Infinite Variables"*, Int. Math. Res. Not. 17, 5486-5535 (2018).

[65] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, Cambridge (2006).

[66] Y. Oreg, G. Refael, and F. von Oppen, *"Helical Liquids and Majorana Bound States in Quantum Wires"*, Phys. Rev. Lett. 105, 177002 (2010).

[67] M. E. Peskin and D. V. Schroeder, *"An Introduction to Quantum Field Theory"*, Westview Press (1995).

[68] M. Pimsner and D. Voiculescu, *"K-Groups of Reduced Crossed Products by Free Groups"*, J. Operator Theory 8 (1982), No. 1, 131156.

[69] F. Rădulescu, *"Random Matrices, Amalgamated Free Products and Subfactors of the von Neumann Algebra of a Free Group, of Noninteger Index"*, Invent. Math. Volume 115 Number 2 (1994), 347-389.

[70] E. Ramberg and G. A. Snow, *"Experimental Limit on s Small Violation of the Pauli Principle"*, Phys. Lett. B 238, 438 (1990).

[71] E. Ricard, *"Factoriality of q-Gaussian von Neumann Algebras"*, Comm. Math. Phys. 257, 3, 659-665 (2005).

[72] M. Rørdam, F. Larsen, and N. Laustsen, *"An Introduction to K-Theory for C*-Algebras"*, Cambridge University Press, Cambridge (2000).

[73] J. S. Rosenthal, *"A First Look at Rigorous Probability Theory"* (second edition), World Scientific Publishing, Singapore (2006).

[74] D. Rossini, M. Carrega, M. C. Strinati, and L. Mazza, *"Anyonic Tight-Binding Models of Parafermions and of Fractionalised Fermions"*, Phys. Rev. B 99, 085113 (2019).

[75] S. Sakai, *"C*-Algebras and W*-Algebras"* in the Series *Classics in Mathematics* (Volume 60), Springer Verlag, Berlin Heidelberg (1998).

[76] R. L. Schilling, *"Measures, Integrals and Martingales"*, Cambridge University Press, Cambridge (2005).

[77] U. Schollwöck, *"The Density-Matrix Renormalization group"*, Rev. Mod. Phys. 77, 259 (2005).

[78] D. Shlyakhtenko, *"Lower Estimates on Microstates Free Entropy Dimension"*, Analysis and PDE Vol. 2, No. 2 (2009).

[79] A. Smith, J. Knolle, D. L. Kovrizhin, and R. Moessner, *"Disorder-Free Localization"*, Phys. Rev. Lett. 118, 266601 (2017).

[80] R. Speicher, *"Free Probability Theory: and its Avatars in Representation Theory, Random Matrices, and Operator Algebras; also Featuring: Non-Commutative Distributions"* (English summary), Jahresber. Dtsch. Math.-Ver. 119 (2017), No. 1, 330.

[81] H. T. C. Stoof, K. B. Gubbels, and D. B. M. Dickerscheid, *"Ultracold Quantum Fields"*, Springer Verlag, Berlin Heidelberg (2009).

[82] M. Takesaki, *"Theory of Operator Algebras I"* in the series *Encyclopaedia of Mathematical Sciences*, Springer Verlag, Berlin Heidelberg (2001).

[83] T. Tao, *"When Are Eigenvalues Stable?"* (blog post), https://terrytao.wordpress.com/2008/10/28/when-are-eigenvalues-stable/ (2008).

[84] M. Tsuji. *"Potential Theory in Modern Function Theory"* (second edition), Chelsea Publishing Company, New York (1975).

[85] C. Villani, *"Topics in Optimal Transportation"* in the series *Graduate Studies in Mathematics*, American Mathematical Society, Providence (2003).

[86] D. Voiculescu, K. Dykema, and A. Nica, *"Free Random Variables"* in the series *CRM Monograph Series* (volume 1), American Mathematical Society, Providence (1992).

[87] D. Voiculescu, *"The Analogues of Entropy and of Fisher's Information Measure in Free Probability Theory, II"*, Invent. Math. 118, 411-440 (1994).

[88] D. Voiculescu, *"The Analogues of Entropy and of Fisher's Information Measure in Free Probability Theory, V: Non Commutative Hilbert Transforms"*, Invent. Math. 132, 189-227 (1998).

[89] D. B. Zagier, *"Realizability of a Model in Infinite Statistics"*, Comm. Math. Phys. 147, 1, 199-210 (1992).

Acknowledgements

First and foremost, I want to thank my supervisors, Erik van den Ban, Martijn Caspers, and Dirk Schuricht. Thank you Erik, for enthousiastically supervising a project that was both somewhat out of your direct field of research and administratively/logistically challenging. Thank you Martijn, for suggesting the topic of free monotone transport and all your support in finding a PhD position (also thanks to the Analysis group at Delft for their hospitality). Thank you Dirk, for introducing me to 'proper' research and trusting me with such an 'exploratory' project. My gratitude also goes out to Iman Mahyaeh and Jurriaan Wouters, it was a pleasure to work with you.

Thanks to wordclouds.com for providing the nice images at the start of the parts.

I would like to extend a heartfelt thank you to my family for their unwavering support and engagement with my studies. Finally I would like to thank my friends who made this journey with me.