Preprint typeset in JHEP style - HYPER VERSION

A Few Remarks On Topological Field Theory

Gregory W. Moore

Abstract: Physics 695, Fall 2015. September 9 - December 14, 2015. Version April 30, 2016

Contents

1. Introduction

Topological field theory is an excellent pedagogical tool for introducing both some basic ideas of physics along with some beautiful mathematical ideas.

The idea of TFT arose from both the study of two-dimensional conformal field theories and from Witten's work on the relation of Donaldson theory to N=2 supersymmetric field theory and Witten's work on the Jones polynomial and three-dimensional quantum field theories. In conformal field theory, Graeme Segal stated a number of axioms for the definition of a CFT. These were adapted to define a notion of a TFT by Atiyah.

TFT might be viewed as a basic framework for physics. It assigns Hilbert spaces, states, and transition amplitudes to topological spaces in a way that captures the most primitive notions of locality. By stripping away the many complications of "real physics" one is left with a very simple structure. Nevertheless, the resulting structure is elegant, it is related to beautiful algebraic structures which, at least in two dimensions, which have surprisingly useful consequences. This is one case where one can truly "solve the theory."

Of course, we are interested in more complicated theories. But the basic framework here can be adapted to any field theory. What changes is the geometric category under consideration. Thus, it offers one approach to the general question of "What is a quantum field theory?"

2. Basic Ideas

It is possible to speak of physics in 0-dimensional spacetime. From the functional integral viewpoint this is quite natural: Path integrals become ordinary integrals. It is also very fruitful to consider string theories whose target spaces are 0-dimensional spacetimes. Nevertheless, in the vast majority of physical problems we work with systems in d spacetime dimensions with $d > 0$. We will henceforth assume $d > 0$.

What are the most primitive things we want from a physical theory in d spacetime dimensions? In a physical theory one often decomposes spacetime into space and time as in (1). If space is a $(d-1)$-dimensional manifold Y then, in quantum mechanics, we associate to it a vector space of states $\mathcal{H}(Y_{d-1})$.

Of course, in quantum mechanics $\mathcal{H}(Y_{d-1})$ usually has more structure - it is a Hilbert space. But in the spirit of developing just the most primitive aspects we will not incorporate that for the moment. (The notion of a *unitary TFT* captures the Hilbert space, as described below.) Moreover, in a generic physical theory there are natural operators acting on this Hilbert space such as the Hamiltonian. The spectrum of the Hamiltonian and other physical observables depends on a great deal of data. Certainly they depend on the metric on spacetime since a nonzero energy defines a length scale

$$L = \frac{\hbar c}{E}.$$

In topological field theory one ignores most of this structure, and focuses on the dependence of $\mathcal{H}(Y)$ on the topology of Y. For simplicity, we will initially assume Y is compact without boundary.

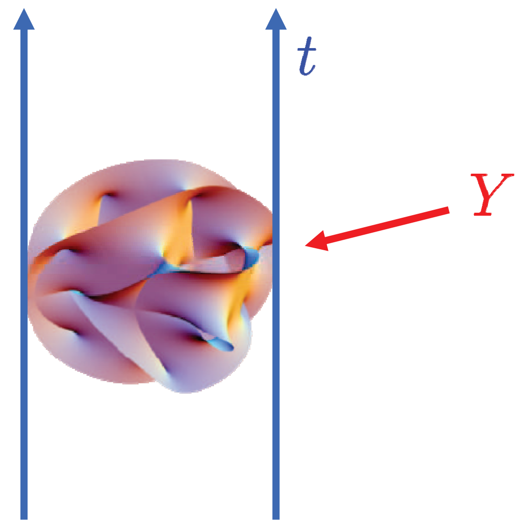


Figure 1: A spacetime $X_d = Y \times \mathbb{R}$. Y is $(d-1)$-dimensional space, possibly with nontrivial topology.

So: In topological field theory we want to have an association:

$(d-1)$-manifolds Y to vector spaces: $Y \to \mathcal{H}(Y)$, such that "$\mathcal{H}(Y)$ is the same for homeomorphic vector spaces." What this means is that if there is a homeomorphism

$$\varphi : Y \to Y' \tag{2.1}$$

then there is a corresponding isomorphism of vector spaces:

$$\varphi_* : \mathcal{H}(Y) \to \mathcal{H}(Y') \tag{2.2}$$

so that composition of homeomorphisms corresponds to composition of vector space isomorphisms. In particular, self-homeomorphisms of Y act as automorphisms of $\mathcal{H}(Y)$: It therefore provides a (possibly trivial) representation of the diffeomorphism group.

Now, we also want to incorporate some form of locality, at the most primitive level. Thus, if we take disjoint unions

$$\mathcal{H}(Y_1 \amalg Y_2) = \mathcal{H}(Y_1) \otimes \mathcal{H}(Y_2) \tag{2.3}$$

Note that (2.3) implies that we should assign to $\mathcal{H}(\emptyset)$ the field of definition of our vector space. For simplicity we will take $\mathcal{H}(\emptyset) = \mathbb{C}$, although one could use other ground fields.

Remark: In algebraic topology it is quite common to assign an abelian group or vector space to a topological space. This is what the cohomology groups do, for example. But here we see a big difference from the standard algebraic topology examples. In those examples the spaces add under disjoint union. In quantum mechanics the spaces multiply. This is the fundamental reason why many topologists refer to the topological invariants arising from topological field theories as "quantum invariants."

Finally, there is an obvious homeomorphism

$$Y \amalg Y' \cong Y' \amalg Y \tag{2.4}$$

and hence there must be an isomorphism

$$\Omega : \mathcal{H}(Y) \otimes \mathcal{H}(Y') \to \mathcal{H}(Y') \otimes \mathcal{H}(Y) \tag{2.5}$$

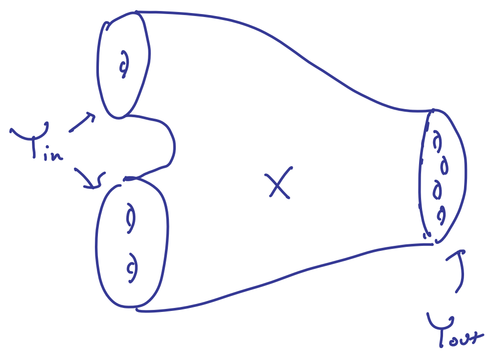


Figure 2: Generalizing the product structure, a d-dimensional bordism X can include topology change between the initial $(d-1)$-dimensional spatial slices Y_{in} and the final spatial slice Y_{out}. The amplitude $F(X)$ determined by a path integral on this bordism is a linear map $\mathcal{H}(Y_{\text{in}}) \to \mathcal{H}(Y_{\text{out}})$.

In addition, in physics we want to speak of transition amplitudes. If there is a spacetime X_d interpolating between two time-slices, then mathematically, we say there is a bordism between Y and Y'. That is, a *bordism* from Y to Y' is a d-manifold with boundary and a disjoint partition of its boundary into two sets the "in-boundary" and the "out-boundary"

$$\partial X_d = (\partial X_d)_{\rm in} \cup (\partial X_d)_{\rm out}$$

so that there is a homeomorphism $(\partial X_d)_{\rm in} \cong Y$ and $(\partial X_d)_{\rm out} \cong Y'$. We will say this a bit more precisely, and discuss some variants, in Section **** below.

If X_d is a bordism from Y to Y' then the Feynman path integral assigns a linear transformation

$$F(X_d) : \mathcal{H}(Y) \to \mathcal{H}(Y').$$

Again, in the general case, the amplitudes depend on much more than just the topology of X_d, but in topological field theory they are supposed only to depend on the topology. More precisely, if $X_d \cong X'_d$ are homeomorphic by a homeomorphism $= 1$ on the boundary of the bordism, then

$$F(X_d) = F(X'_d)$$

One key aspect of the path integral - in quantum mechanics, or functional integral - in quantum field theory, we want to capture - again a consequence of locality - is the idea of summing over a complete set of intermediate states. In the path integral formalism we can formulate the sum over all paths of field configurations from t_0 to t_2 by composing the amplitude for all paths from t_0 to t_1 and then from t_1 to t_2, where $t_0 < t_1 < t_2$, and then summing over all intermediate field configurations at t_1. We refer to this property as the "gluing property." The gluing property is particularly obvious in the functional integral formulation of field theories.

In topological field theory this is formalized as:

If X is a bordism from Y to Y' with

$$(\partial X)_{\rm in} = Y \qquad (\partial X)_{\rm out} = Y'$$

and X' is another oriented bordism from Y' to Y''

$$(\partial X')_{\rm in} = Y' \qquad (\partial X)_{\rm out} = Y''$$

then we can compose $X' \circ X$ as in (**??**) to get a bordism from Y to Y''.

Naturally enough we want the associated linear maps to compose:

$$F(X' \circ X) = F(X') \circ F(X) : \mathcal{H}(Y) \to \mathcal{H}(Y'')$$

What we are describing, in mathematical terms, is a functor between categories. After describing a few variations on the above theme, we will explain that sentence in detail.

2.1 More Structure

We can regard the above picture as a basic framework for building up more interesting theories by enriching the topological and geometric data associated with the spaces X and Y.

For example, we might be able to endow X and Y with

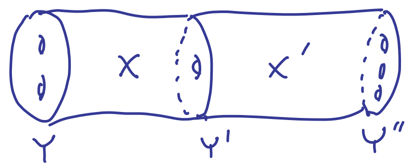


Figure 3: Gluing two bordisms to produce a third bordism.

1. Orientations, spin, pin structures, etc. (for certain X's and Y's).

2. Riemannian metrics.

3. Other fields - Principal G-bundles with connection, sections of associated bundles etc.

One of the motivating examples was two-dimensional conformal field theory. In this case, Segal's axioms were based on two-dimensional bordisms endowed with conformal structure.

Two important complications that will arise when considering nontopological theories are:

1. The notion of scale and renormalization becomes important.
2. The Hilbert space is actually not defined for a $(d-1)$-dimensional manifold but rather for a germ of d-manifolds around a $(d-1)$-dimensional manifold.

3. Some Basic Notions In Category Theory

We will not describe categories in any great detail. See, for example, the book by

S. Maclane, *Categories for the Working Mathematician*, Springer GTM vol.5

This rather abstract mathematical idea has nevertheless found recent application in string theory and conformal field theory. Many physicists object to the high level of abstraction entailed in the category language. However, it seems to be of increasing utility in the further formal development of string theory and supersymmetric gauge theory as well as certain aspects of condensed matter theory and quantum information theory.

3.1 Basic Definitions

Definition A *category* $\mathcal{C}$ consists of

a.) A set $Ob(\mathcal{C})$ of "objects"

b.) A collection $Mor(\mathcal{C})$ of sets $\hom(X,Y)$, defined for any two objects $X,Y \in Ob(\mathcal{C})$. The elements of $\hom(X,Y)$ are called the "morphisms from X to Y." They are often denoted as arrows:

$$X \quad \stackrel{\phi}{\rightarrow} \quad Y \tag{3.1}$$

c.) A composition law:

$$\hom(X,Y) \times \hom(Y,Z) \rightarrow \hom(X,Z) \tag{3.2}$$

$$(\psi_1, \psi_2) \mapsto \psi_2 \circ \psi_1 \tag{3.3}$$

Such that

1. A morphism ϕ uniquely determines its source X and target Y. That is, $\hom(X,Y)$ are disjoint.

2. $\forall X \in Ob(\mathcal{C}) \; \exists \; 1_X : X \rightarrow X$, uniquely determined by:

$$1_X \circ \phi = \phi \qquad \psi \circ 1_X = \psi \tag{3.4}$$

for morphisms ϕ, ψ, when the composition is defined.

3. Composition of morphisms is associative:

$$(\psi_1 \circ \psi_2) \circ \psi_3 = \psi_1 \circ (\psi_2 \circ \psi_3) \tag{3.5}$$

An alternative definition one sometimes finds is that a category is defined by two sets C_0 (the objects) and C_1 (the morphisms) with two maps $p_0 : C_1 \rightarrow C_0$ and $p_1 : C_1 \rightarrow C_0$. The map $p_0(f) = x_1$ is the *range* map and $p_1(f) = x_0$ is the *domain* map. In this alternative definition a category is then defined by a composition law on the set of *composable morphisms*

$$C_2 = \{(f,g) \in C_1 \times C_1 | p_0(f) = p_1(g)\} \tag{3.6}$$

which is sometimes denoted $C_{1 p_1} \times_{p_0} C_1$ and called the *fiber product*. The composition law takes $C_2 \rightarrow C_1$ and may be pictured as the composition of arrows. If $f : x_0 \rightarrow x_1$ and $g : x_1 \rightarrow x_2$ then the composed arrow will be denoted $g \circ f : x_0 \rightarrow x_2$. The composition law satisfies the axioms

1. For all $x \in C_0$ there is an identity morphism in C_1, denoted 1_x, or Id_x, such that $1_x f = f$ and $g 1_x = g$ for all suitably composable morphisms f, g.

2. The composition law is associative. If f, g, h are 3-composable morphisms then $(hg)f = h(gf)$.

Remarks:

1. When defining composition of arrows one needs to make an important notational decision. If $f : x_0 \to x_1$ and $g : x_1 \to x_2$ then the composed arrow is an arrow $x_0 \to x_2$. We will write $g \circ f$ when we want to think of f, g as functions and fg when we think of them as arrows.

2. It is possible to endow the data C_0, C_1 and p_0, p_1 with additional structures, such as topologies, and demand that p_0, p_1 have continuity or other properties.

3. A morphism $\phi \in \hom(C, D)$ is said to be *invertible* if there is a morphism $\psi \in \hom(D, C)$ such that $\psi \circ \phi = 1_C$ and $\phi \circ \psi = 1_D$. If C and D are objects with an invertible morphism between then then they are called *isomorphic objects*. One key reason to use the language of categories is that objects can have nontrivial automorphisms. That is, $\hom(C, C)$ can have more than just 1_C in it. When this is true then it is tricky to speak of "equality" of objects, and the language of categories becomes very helpful. One should be very careful about saying that two mathematical things are "the same."

♣Explain this really important point better. Give an example where literal equality is far too rigid. ♣

One use of categories is that they provide a language for describing precisely notions of "similar structures" in different mathematical contexts. For example:

1. **SET**: The category of sets and maps of sets

2. **TOP**: The category of topological spaces and continuous maps.

3. **TOPH**: The category of topological spaces and homotopy classes of continuous maps.

4. **MAN**: The category of manifolds and maps of manifolds. (One should specify the degree of smoothness here.)

5. **GROUP**: the category of groups and homomorphisms of groups.

6. **AB**: The (sub) category of abelian groups.

7. **VECT**$_\kappa$: The category of finite-dimensional vector spaces over a field κ.

♣Do we want to impose finite dimensionality? Or introduce a category of all vector spaces and a subcategory of finite-dimensional vector spaces. ♣

When discussed in this way it is important to introduce the notion of functors and natural transformations (morphisms between functors) to speak of interesting relationships between categories.

In order to state a relation between categories one needs a "map of categories." This is what is known as a functor:

Definition A *functor* between two categories $\mathcal{C}_1$ and $\mathcal{C}_2$ consists of a pair of maps $F_{\rm obj} : Obj(\mathcal{C}_1) \to Obj(\mathcal{C}_2)$ and $F_{\rm mor} : Mor(\mathcal{C}_1) \to Mor(\mathcal{C}_2)$ so that if

$$x \xrightarrow{f} y \ \in \hom(x,y) \tag{3.7}$$

then

$$F_{\rm obj}(x) \xrightarrow{F_{\rm mor}(f)} F_{\rm obj}(y) \ \in \hom(F_{\rm obj}(x), F_{\rm obj}(y)) \tag{3.8}$$

Moreover we require that if f_1, f_2 are composable morphisms then

$$F_{\rm mor}(f_1 \circ f_2) = F_{\rm mor}(f_1) \circ F_{\rm mor}(f_2) \tag{3.9}$$

and finally we require that for all objects $x \in Obj(\mathcal{C}_1)$ we have

$$F_{\rm mor}(1_x) = 1_{F_{\rm obj}(x)} \tag{3.10}$$

We usually drop the subscript on F since it is clear what is meant from context.

Remarks

1. Above we have described a *covariant functor*. A *contravariant functor* instead satisfies ϕ_2, ϕ_1,

   $$F(\phi_2 \circ \phi_1) = F(\phi_1) \circ F(\phi_2)$$

   for any pair of composable morphisms
2. Some authors use the term *homomorphism of categories*.

Exercise

Using the alternative definition of a category in terms of data $p_{0,1} : X_1 \to X_0$ define the notion of a functor writing out the relevant commutative diagrams.

Example 1: Every category has a canonical functor to itself, called the identity functor $Id_{\mathcal{C}}$.

Example 2: There is an obvious functor, the "forgetful functor" that forgets mathematical structure, so we have, for example, forgetful functors from **TOP**, **MAN** and **GROUP** to **SET**.

Example 3: Since **AB** is a subcategory of **GROUP** there is an obvious functor $\mathcal{F}$: **AB** $\to$ **GROUP**.

Example 4: In an exercise below you are asked to show that the abelianization of a group defines a functor $\mathcal{G} : \mathbf{GROUP} \to \mathbf{AB}$.

Example 5: Homology groups give a nice example of a functor from **TOP** to **AB**. For example, fix a nonnegative integer k, then the functor H_k on objects is $F(X) := H_k(X;\mathbb{Z})$, and for a continuous map of spaces $f : X_1 \to X_2$ we have $F(f) = f_*$. Similarly cohomology groups give an example of a contravariant functor.

When there are functors both ways between two categories we might ask whether they might be, in some sense, "the same." But saying precisely what is meant by "the same" requires some care.

Definition If $\mathcal{C}_1$ and $\mathcal{C}_2$ are categories and $F_1 : \mathcal{C}_1 \to \mathcal{C}_2$ and $F_2 : \mathcal{C}_1 \to \mathcal{C}_2$ are two functors then a *natural transformation* (a.k.a. a morphism of functors) $\tau : F_1 \to F_2$ is a rule which, for every $X \in Obj(\mathcal{C}_1)$ assigns an arrow $\tau_X : F_1(X) \to F_2(X)$ so that, for all $X, Y \in Obj(\mathcal{C}_1)$ and all $f \in \hom(X, Y)$,

$$\tau_Y \circ F_1(f) = F_2(f) \circ \tau_X \tag{3.11}$$

Or, in terms of diagrams.

$$\begin{tikzcd} F_1(X) \arrow[r, "F_1(f)"] \arrow[d, "\tau_X"] & F_1(Y) \arrow[d, "\tau_Y"] \\ F_2(X) \arrow[r, "F_2(f)"] & F_2(Y) \end{tikzcd} \tag{3.12}$$

Note that it makes sense to compose natural transformations: If $\tau : F_1 \to F_2$ and $\tau' : F_2 \to F_3$ are morphisms of functors then $(\tau' \circ \tau)_X$ is the morphism from $F_1(X) \to F_3(X)$ given by composing the morphisms $\tau_X : F_1(X) \to F_2(X)$ and $\tau'_X : F_2(X) \to F_3(X)$. A natural transformation $\tau : F_1 \to F_2$ such that there exists another natural transformation $\tau' : F_2 \to F_1$ such that

$$(\tau' \circ \tau)_X = 1_{F_1(X)} \qquad (\tau \circ \tau')_X = 1_{F_2(X)} \tag{3.13}$$

is called an *isomorphism of functors.*

Example: A good example of various natural transformations are various cohomology operations. For example the cup product gives a natural transformation from H_k to H_{2k}. (This is related to, but not the same as the cohomology operation known as a "Steenrod square.") ♣CHECK!! ♣

Definition Two categories are said to be *equivalent* if there are functors $F : \mathcal{C}_1 \to \mathcal{C}_2$ and $G : \mathcal{C}_2 \to \mathcal{C}_1$ together with isomorphisms (via natural transformations) $FG \cong Id_{\mathcal{C}_2}$ and $GF \cong Id_{\mathcal{C}_1}$. (Note that FG and $Id_{\mathcal{C}_2}$ are both objects in the category of functors $\mathrm{FUNCT}(\mathcal{C}_2, \mathcal{C}_2)$ so it makes sense to say that they are isomorphic.)

♣Explain in the context of the following examples why a definition of equivalence of categories based on $GF = Id_{\mathcal{C}_1}$ etc. is too restrictive. ♣

Many important theorems in mathematics can be given an elegant and concise formulation by saying that two seemingly different categories are in fact equivalent. Here is a (very selective) list: [1]

Example 1: Consider the category with one object for each nonnegative integer n and the morphism space $GL(n, \kappa)$ of invertible $n \times n$ matrices over the field κ. These categories are equivalent. That is one way of saying that the only invariant of a finite-dimensional vector space is its dimension.

Example 2: The basic relation between Lie groups and Lie algebras the statement that the functor which takes a Lie group G to its tangent space at the identity, $T_1 G$ is an equivalence of the category of connected and simply-connected Lie groups with the category of finite-dimensional Lie algebras.

Example 3: Covering space theory is about an equivalence of categories. On the one hand we have the category of coverings of a pointed space (X, x_0) and on the other hand the category of topological spaces with an action of the group $\pi_1(X, x_0)$. Closely related to this, Galois theory can be viewed as an equivalence of categories.

Example 4: As we will see below, the category of unital commutative C^*-algebras is equivalent to the category of compact Hausdorff topological spaces. This is known as Gelfand's theorem.

Example 5: Similar to the previous example, an important point in algebraic geometry is that there is an equivalence of categories of commutative algebras over a field κ (with no nilpotent elements) and the category of affine algebraic varieties.

Example 6: Pontryagin duality is a nontrivial self-equivalence of the category of locally compact abelian groups (and continuous homomorphisms) with itself.

Example 7: A generalization of Pontryagin duality is Tannaka-Krein duality between the category of compact groups and a certain category of linear tensor categories. (The idea is that, given an abstract tensor category satisfying certain conditions one can construct a group, and if that tensor category is the category of representations of a compact group, one recovers that group.)

Example 8: The Riemann-Hilbert correspondence can be viewed as an equivalence of categories of flat connections (a.k.a. linear differential equations, a.k.a. D-modules) with their monodromy representations.

♣This needs a lot more explanation. ♣

In physics, the statement of "dualities" between different physical theories can sometimes be formulated precisely as an equivalence of categories. One important example of

[1] I thank G. Segal for a nice discussion that helped prepare this list.

this is mirror symmetry, which asserts an equivalence of (A_∞)-) categories of the derived category of holomorphic bundles on X and the Fukaya category of Lagrangians on $X^\vee$. But more generally, nontrivial duality symmetries in string theory and field theory have a strong flavor of an equivalence of categories.

Exercise
Give an example to show that equation (3.10) does not follow from (3.10).

Exercise *Playing with natural transformations*

a.) Given two categories $\mathcal{C}_1, \mathcal{C}_2$ show that the natural transformations allow one to define a category FUNCT($\mathcal{C}_1, \mathcal{C}_2$) whose objects are functors from $\mathcal{C}_1$ to $\mathcal{C}_2$ and whose morphisms are natural transformations. For this reason natural transformations are often called "morphisms of functors."

b.) Write out the meaning of a natural transformation of the identity functor $Id_{\mathcal{C}}$ to itself. Show that $End(Id_{\mathcal{C}})$, the set of all natural transformations of the identity functor to itself is a monoid.

Exercise *Freyd's theorem*

A "practical" way to tell if two categories are equivalent is the following:

By definition, a *fully faithful functor* is a functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ where $F_{\rm mor}$ is a bijection on all the hom-sets. That is, for all $X, Y \in Obj(\mathcal{C}_1)$ the map

$$F_{\rm mor} : \hom(X, Y) \to \hom(F_{\rm obj}(X), F_{\rm obj}(Y)) \tag{3.14}$$

is a bijection.

Show that $\mathcal{C}_1$ is equivalent to $\mathcal{C}_2$ iff there is a fully faithful functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ so that any object $\alpha \in Obj(\mathcal{C}_2)$ is isomorphic to an object of the form $F(X)$ for some $X \in Obj(\mathcal{C}_1)$.

c.) Show that the category of finite-dimensional vector spaces over $\mathbb{C}$ is equivalent to the category

---

Exercise

As we noted above, there is a functor $\mathbf{AB} \to \mathbf{GROUP}$ just given by inclusion.

a.) Show that the abelianization map $G \to G/[G, G]$ defines a functor $\mathbf{GROUP} \to \mathbf{AB}$.

b.) Show that the existence of nontrivial perfect groups, such as A_5, implies that this functor cannot be an equivalence of categories.

In addition to the very abstract view of categories we have just sketched, very concrete objects, like groups, manifolds, and orbifolds can profitably be viewed as categories.

One may always picture a category with the objects constituting points and the morphisms directed arrows between the points as shown in Figure 4.

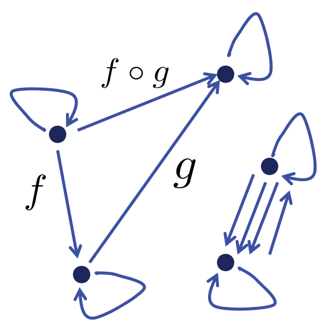


Figure 4: Pictorial illustration of a category. The objects are the black dots. The arrows are shown, and one must give a rule for composing each arrow and identifying with one of the other arrows. For example, given the arrows denoted f and g it follows that there must be an arrow of the type denoted $f \circ g$. Note that every object x has at least one arrow, the identity arrow in $Hom(x, x)$.

As an extreme example of this let us consider a category with only *one object*, but we allow the possibility that there are several morphisms. For such a category let us look carefully at the structure on morphisms $f \in Mor(\mathcal{C})$. We know that there is a binary operation, with an identity 1 which is associative.

But this is just the definition of a monoid!

If we have in addition inverses then we get a group. Hence:

Definition A *group* is a category with one object, all of whose morphisms are invertible.

To see that this is equivalent to our previous notion of a group we associate to each morphism a group element. Composition of morphisms is the group operation. The invertibility of morphisms is the existence of inverses.

We will briefly describe an important and far-reaching generalization of a group afforded by this viewpoint. Then we will show that this viewpoint leads to a nice geometrical construction making the formulae of group cohomology a little bit more intuitive.

3.2 Groupoids

Definition A *groupoid* is a category all of whose morphisms are invertible.

Note that for any object x in a groupoid, $\hom(x, x)$ is a group. It is called the *automorphism group* of the object x.

Example 1. Any equivalence relation on a set X defines a groupoid. The objects are the elements of X. A morphism is an equivalence relation $a \sim b$. Composition of morphisms $a \sim b$ with $b \sim c$ is $a \sim c$. Clearly, every morphism is invertible.

Example 2. Consider time evolution in quantum mechanics with a time-dependent Hamiltonian. There is no sense to time evolution $U(t)$. Rather one must speak of unitary evolution $U(t_1, t_2)$ such that $U(t_1, t_2)U(t_2, t_3) = U(t_1, t_3)$. Given a solution of the Schrodinger equation $\Psi(t)$ we may consider the state vectors $\Psi(t)$ as objects and $U(t_1, t_2)$ as morphisms. In this way a solution of the Schrodinger equation defines a groupoid.

Example 3. Let X be a topological space. The fundamental groupoid $\pi_{\leq 1}(X)$ is the category whose objects are points $x \in X$, and whose morphisms are homotopy classes of paths $f : x \to x'$. These compose in a natural way. Note that the automorphism group of a point $x \in X$, namely, $\hom(x, x)$ is the fundamental group of X based at x, $\pi_1(X, x)$.

Example 4. Gauge theory: Objects = connections on a principal bundle. Morphisms = gauge transformations. This is the right point of view for thinking about some more exotic (abelian) gauge theories of higher degree forms which arise in supergravity and string theories.

Example 5. In the theory of string theory orbifolds and orientifolds spacetime must be considered to be a groupoid.

Exercise

Let X be a set with an action of a group G. Show that there is a natural groupoid (sometimes denoted $X//G$) such that the set of isomorphism classes of objects is naturally identified with the quotient set X/G.

Exercise

For a group G let us define a groupoid denoted $G//G$ whose objects are group elements $Obj(G//G) = G$ and whose morphisms are arrows defined by

$$g_1 \xrightarrow{h} g_2 \tag{3.15}$$

iff $g_2 = h^{-1} g_1 h$. This is the groupoid of principal G-bundles on the circle.

Draw the groupoid corresponding to S_3.

3.3 Tensor Categories

To define a TFT we need the further notion of a tensor category. Note that given a category C, the Cartesian products $C \times C$, $C \times C \times C$, ... are also categories in a natural way.

Definition A *tensor category* (also known as a *monoidal category*) is a category with a functor $\otimes : C \times C \to C$ such that there is an isomorphism $\mathcal{A}$ of the two functors $\otimes \circ \otimes_{12} : C \times C \times C \to C$ and $\otimes \circ \otimes_{23} : C \times C \times C \to C$ satisfying the pentagon identity, and such that there is an identity object 1_C together with natural transformations of functors $C \to C$:

$$\iota_L : 1_C \otimes \cdot \to \mathrm{Id} \tag{3.16}$$

$$\iota_R : \cdot \otimes 1_C \to \mathrm{Id} \tag{3.17}$$

These data are subject to a number of natural compatibility conditions:

♣Do we require the existence of a dual object? ♣

To give an example of the compatibility conditions we consider the the first condition on the natural transformation $\mathcal{A}$: for all objects x, x', x'' in C_0 we have an isomorphism:

$$\mathcal{A}_{x,x',x''} : (x \otimes x') \otimes x'' \to x \otimes (x' \otimes x'') \tag{3.18}$$

which satisfies the pentagon identity:

♣FIX xy matrix ♣

$$\begin{tikzcd}[column sep=tiny]
& ((x_1 \otimes x_2) \otimes x_3) \otimes x_4 \arrow[r] \arrow[dl] & (x_1 \otimes x_2) \otimes (x_3 \otimes x_4) \arrow[dr] & \\
(x_1 \otimes (x_2 \otimes x_3)) \otimes x_4 \arrow[drr] & & & x_1 \otimes (x_2 \otimes (x_3 \otimes x_4)) \\
& & x_1 \otimes ((x_2 \otimes x_3) \otimes x_4)) \arrow[ur] &
\end{tikzcd} \tag{3.19}$$

It is then a theorem (the "coherence theorem") that $x_0 \otimes x_1 \cdots \otimes x_n$ is well-defined up to isomorphism no matter how one brackets the products. The conditions on the natural transformations ι_L and ι_R are fairly obvious.

Example The category $\mathbf{VECT}_\kappa$ is a tensor category. What is the tensor unit $1_{\mathbf{VECT}_\kappa}$?

Let $\sigma : C \times C \to C \times C$ be the exchange functor that switches factors on objects and morphisms.

Definition A *symmetric monoidal category* is a monoidal category with an isomorphism Ω of $\otimes \circ \sigma$ with $\otimes$ which squares to one. Again, there are many rather obvious compatibility conditions with $\mathcal{A}$, ι_L and ι_R.

Again, this means that for all objects x, y we have an isomorphism

$$\Omega_{x,y} : x \otimes y \to y \otimes x \tag{3.20}$$

so that $\Omega_{y,x} \circ \Omega_{x,y} = 1_{x\otimes y}$.

Remark: An important generalization for conformal field theory and for quasiparticle statistics in 2+1 dimensions is the notion of a *braided tensor category* where there is an isomorphism Ω, but it does not square to 1.

Finally, we need the notation of a (symmetric) tensor functor. This is a functor $F : C \to D$ between symmetric tensor categories together with an isomorphism $1_D \to F(1_C)$ and an isomorphism of the two functors $C \times C \to D$ given by $F \circ \otimes$ and $\otimes \circ F \times F$.

3.4 Other Tensor Categories

3.5 $\mathbb{Z}_2$-graded vector spaces

A $\mathbb{Z}_2$ graded vector space is a vector space with a decomposition $V = V_0 \oplus V_1$, where the subscripts are understood as elements of $\mathbb{Z}_2$. In the category of $\mathbb{Z}_2$-graded vector spaces we can introduce two different kinds of tensor categories. For $\mathbb{Z}_2$ graded vector spaces we can and will use the graded tensor product. Then there is an isomorphism

$$\Omega : V \otimes W \to W \otimes V \tag{3.21}$$

but we must be careful to apply the *Koszul sign rule*: If v, w are homogeneous elements then

$$\Omega(v \otimes w) = (-1)^{|v|\cdot|w|} w \otimes v \tag{3.22}$$

This rule has the important consequence that if we have any collection $(V_\alpha)_{\alpha\in I}$ of supervector spaces (where the subscript α denotes different supervector spaces and should not be confused with the $\mathbb{Z}_2$ grading) then there is a single canonical tensor product

$$\otimes_\alpha V_\alpha$$

without the need to specify any ordering.

3.6 Category Of Representations Of A Group

Let G be a group. Then then there is a category whose objects are representations and morphisms are intertwiners of representations (i.e. maps between representations that commute with the G action.

Now let G be a compact group and restrict to the subcategory of finite-dimensional representations. Call this $\mathrm{Rep}(G)$. This is a tensor category. Moreover, there is a set of "simple" objects, the irreducible representations V_λ such that all objects are isomorphic to direct sums of simple objects. The tensor functor is determined by the "fusion rules"

$$V_\lambda \otimes V_\mu \cong D^\rho_{\lambda\mu} \otimes V_\rho \tag{3.23}$$

where $D^\rho_{\lambda\mu}$ is a finite-dimensional real vector space of degeneracies.

4. Bordism

4.1 Unoriented Bordism: Definition And Examples

Here we give the official definition of a bordism:

Definition Let Y_0, Y_1 be two closed $(d-1)$-dimensional manifolds. A *bordism* from Y_0 to Y_1 is

1. A d-manifold X together with a disjoint partition of its boundary:
$$\partial X = (\partial X)_0 \amalg (\partial X)_1 \tag{4.1}$$
2. A pair of embeddings $\theta_0 : [0,1) \times Y_0 \to X$ and $\theta_1 : (-1,0] \times Y_1 \to X$, which are diffeomorphisms onto their images such that the restrictions $\theta_0 : \{0\} \times Y_0 \to (\partial X)_{\rm in}$ and $\theta_1 : \{0\} \times Y_1 \to (\partial X)_{\rm out}$ are homeomorphisms.

The reason for the extra level of complexity in this definition compared to what we said earlier is that this extra data facilitates the gluing of bordisms to produce a new bordism.

It is easy to see that bordism is an equivalence relation and that disjoint union defines an abelian group structure on the space of bordism equivalence classes Ω_n of n-manifolds. The zero element of the abelian group is the equivalence class of the empty set $\emptyset^n$ and any closed n-manifold X is its own inverse since $[0,1] \times X$ can be considered as a bordism of $X \amalg X$ with $\emptyset$. So $2[X] = 0$ in Ω_n.

Examples

1. There is only one nontrivial zero-dimensional manifold, the point, and we have just seen that the disjoint union of two points is null-bordant, hence $\Omega_0 \cong \mathbb{Z}/2\mathbb{Z}$. Note that if we dropped the manifold condition on X then the letter Y would define a bordism of two points (equivalent to zero) with one point, and hence the bordism group would be trivial. Thus, the manifold condition is important.

2. $\Omega_1 = 0$, because the only closed connected one-manifold is the circle, and this clearly bounds a disk.

3. One can show that $\Omega_2 \cong \mathbb{Z}/2\mathbb{Z}$ with generator $[\mathbb{RP}^2]$. Here is the argument (taken from D. Freed's notes "Bordism Old And New," on his homepage). The classification of compact surfaces shows that they are characterized by two invariants: Orientability and the Euler character. Oriented surfaces are clearly bordant to zero. Note well! The Eulcer character is <u>not</u> a bordism invariant! Unorientable surfaces are all obtained by connected sums with $\mathbb{RP}^2$. The connected sum of two copies of $\mathbb{RP}^2$ is a circle bundle over the circle. Take $[0,1] \times S^1$ and quotient by $\{(0,z)\} \sim \{(1,\bar{z})\}$ (where we view S^1 as the unit complex numbers). Note that we can replace the S^1 by the disk D^2 and use the same identification $\{(0,z)\} \sim \{(1,\bar{z})\}$ to produce a bordism of the Klein bottle to zero. Next we claim that $\mathbb{RP}^2$ descends to a nontrivial bordism class. For, if it had a bordism to zero $\partial X = \mathbb{RP}^2$ then triangulation of X gives a triangulation of

the double $X \cup_{\mathbb{RP}^2} X$ with Euler character $2\chi(X) - 1$. On the other hand, the Euler character of a closed 3-fold is zero. Now, the general connected unorientable surface is a connected sum of n copies of $\mathbb{RP}^2$. Separate these in pairs and choose a bordism of the pairs to zero to identify the bordism class with an element $n$mod2 of $\mathbb{Z}/2\mathbb{Z}$.

4. To describe all bordism groups Ω_d it is useful to note that Cartesion product of manifolds is compatible with the bordism equivalence relation and this makes $\Omega_* \cong \amalg_{d \geq 0} \Omega_d$ into a $\mathbb{Z}$-graded ring, with the grading given by the dimension. Thom proved that
$$\Omega_* \cong R[x_2, x_4, x_5, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{16}, x_{17}, ...] \tag{4.2}$$
where $R = \mathbb{Z}/2\mathbb{Z}$ and there is precisely one generator x_k of degree k so long as k is <u>not</u> of the form $2^j - 1$. The even degree generators are the bordism classes of $\mathbb{RP}^k$ and the odd ones are a quotient of $(S^m \times \mathbb{CP}^\ell)/\mathbb{Z}_2$ where the $\mathbb{Z}_2$ acts as (antipodal map, complex conjugation).

5. Moreover, to any manifold there is a series of cohomology classes $w_i(Y) \in H^i(Y; \mathbb{Z}/2\mathbb{Z})$ known as Stiefel-Whitney classes. They are associated with the twisting of the tangent bundle. (For example, $w_1(Y)$ measures whether Y is orientable or not.) The *Stiefel numbers* of a manifold is the sequence of elements of $\mathbb{Z}/2\mathbb{Z}$:
$$\langle w_{i_1}(Y) \cup \cdots w_{i_k}(Y), [Y] \rangle \tag{4.3}$$
and two manifolds are bordant iff all their Stiefel numbers agree. For the last two items see the excellent book by Milnor and Stasheff, *Characteristic Classes.* [2]

4.2 The Bordism Category $\mathrm{Bord}_{\langle d-1,d \rangle}$

Now, we can define a bordism <u>category</u> $\mathrm{Bord}_{\langle d-1,d \rangle}$.

1. Objects: Closed $(d-1)$-manifolds, usually denoted Y.
2. $\hom(Y_0, Y_1)$ is the set of homeomorphism classes of bordisms $X : Y_0 \to Y_1$. A homeomorphism of bordisms X, X' is a homeomorphism of manifolds with boundaries which takes $(\partial X)_{\rm in} \to (\partial X')_{\rm in}$ and commutes with the collars θ_0, θ_1.

The composition of morphisms in the bordism category is by gluing. Since we identify bordisms by homeomorphism the bordism $X = [0,1] \times Y$ from $Y \to Y$ is the identity morphism 1_Y. The category $\mathrm{Bord}_{\langle d-1,d \rangle}$ is a symmetric tensor category: The tensor product is disjoint union, and the empty manifold $\emptyset^{d-1}$ is the tensor unit.

4.3 The Oriented Bordism Category $\mathrm{Bord}^{\rm SO}_{\langle d-1,d \rangle}$

We are often interested in *oriented bordism.*

To define an oriented bordism we modify the definition of bordism slightly. Now, Y_0, Y_1, X are all oriented. The embeddings θ_0 and θ_1 are required to be orientation preserving and we identify bordisms X and X' by oriented diffeomorphisms.

[2] If we cover the chapter on characteristic classes we will prove these two results.

The condition that θ_0 and θ_1 are orientation preserving must be treated with care. Note that if we are given a sum of oriented real vector spaces there is no natural orientation on the direct sum. However, if we are given an exact sequence

$$0 \to V_1 \to V_2 \to V_3 \to 0 \tag{4.4}$$

Then there is a canonical isomorphism $\mathrm{DET}V_3 \cong \mathrm{DET}V_1 \to \mathrm{DET}V_2$ so if two of the three spaces are oriented, we can determine an orientation on the third by requiring this canonical isomorphism to be orientation preserving. In particular, an orientation on a submanifold and the ambient manifold determines an orientation on the normal bundle. When defining θ_0, θ_1 we orient $[0, +1)$ and $(-1, 0]$ with the standard orientation on $\mathbb{R}$, $+\frac{\partial}{\partial x}$ and then we take the product orientation on $[0, +1) \times Y$ and $(-1, 0] \times Y$.

Definition To every oriented bordism $X : Y_0 \to Y_1$ there is a *dual oriented bordism* $X^\vee : Y_1^\vee \to Y_0^\vee$. Let us write it out carefully, since it can cause confusion. $Y^\vee$ denotes Y with the opposite orientation. $X^\vee$ is the manifold with the <u>same</u> orientation. However, we exchange ingoing and outgoing boundaries. Moreover,

$$\theta_0^\vee(t, y_1) = \theta_1(-t, y_1) \qquad \forall t \in [0, +1) \qquad \& \qquad y_1 \in Y_1 \tag{4.5}$$

$$\theta_1^\vee(t, y_0) = \theta_0(-t, y_0) \qquad \forall t \in (-1, 0] \qquad \& \qquad y_0 \in Y_0 \tag{4.6}$$

Note that the relation between $\theta_0^\vee$ and θ_1 involves an orientation-reversing transformation $t \to -t$ and hence we require orientation reversal on Y since $X^\vee$ has the same orientation as X. Forgetting about orientations we also obtain a notion of dual bordism for the unoriented case.

Once again we can define oriented bordism groups Ω_n^{SO}, for $n \geq 0$, the oriented bordism ring Ω_*^{SO} and the oriented bordism category $\mathrm{Bord}_{\langle d-1,d\rangle}^{\mathrm{SO}}$.

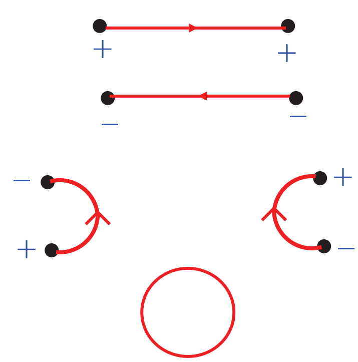


Figure 5: Five connected bordisms in the oriented bordism category. Ingoing boundaries are on the left and outgoing boundaries are on the right.

Example 1 Let us consider the oriented bordism group Ω_0^{SO}. There are two kinds of points pt_+ and pt_-, and five basic connected oriented bordisms, shown in figure 5. Accordingly, $\Omega_0^{\mathrm{SO}} \cong \mathbb{Z}$. The isomorphism takes the difference of the number of $+$ and $-$ points.

Example 2 In dimensions 1 and 2 we again have zero bordism groups.

A summary of the main factors on the oriented bordism ring Ω_*^{SO} is the following. (See Milnor and Stasheff. Several further references are provided in Freed's notes, near Theorem 2.24.) [3]

Theorem

1. All torsion elements in Ω_*^{SO} have order two.

2. Ω_*^{SO}/torsion is a ring with one generator in degrees $4k$, $k \geq 1$.

3. There is an isomorphism
$$\Omega_*^{\mathrm{SO}} \otimes \mathbb{Q} \cong \mathbb{Q}[y_4, y_8, \cdots] \tag{4.7}$$
under which y_{4k} corresponds to the oriented bordism class of $\mathbb{CP}^{2k}$.

4. There are characteristic classes of the tangent bundle of Y, the Stiefel-Whitney classes $w_i(Y) \in H^i(Y; \mathbb{Z}_2)$ and the Pontryagin classes $p_i(Y) \in H^{4i}(Y; \mathbb{Z})$ (the latter depending on the orientation of Y) such that Y_1 and Y_2 and bordant iff all the Stiefel-Whitney and Pontryagin numbers are the same. We defined the Stiefel-Whitney classes above and the Pontryagin numbers are similarly the collection
$$\langle p_{i_1}(Y) \cup \cdots \cup p_{i_k}(Y), [Y] \rangle \in \mathbb{Z} \tag{4.8}$$

4.4 Other Bordism Categories

We can go on an consider other forms of bordism:

1. Framed bordism. (Closely related to the stable homotopy of spheres, by the Pontryagin-Thom construction.)

2. Spin and $\mathrm{Pin}^{\pm}$ bordism.

3. Riemannian bordism.

Accordingly, there are generalizations of the bordism cateogry. In general, if we take into account a structure $\mathcal{S}$ we denote the bordism category by $\mathrm{Bord}^{\mathcal{S}}_{\langle d-1,d\rangle}$, where it is understood that the bordisms are identified by homeomorphisms preserving the structure $\mathcal{S}$. Thus, the oriented bordism category is denoted by $\mathrm{Bord}^{\mathrm{SO}}_{\langle d-1,d\rangle}$ (because the structure group of the tangent bundle is $SO(d-1)$ and $SO(d)$, respectively). Similarly we can define a Riemannian bordism category $\mathrm{Bord}^{\mathrm{Riem}}_{\langle d-1,d\rangle}$, and so on.

5. The Definition Of Topological Field Theory

The definition of a topological field theory can now be given. Let S be a structure on the tangent bundle and C any symmetric monoidal category. Then

Definition A d-dimensional topological field theory of S-manifolds is a symmetric tensor functor from the tensor category $\mathrm{Bord}^{S}_{\langle d-1,d\rangle}$ to some symmetric tensor category C.

[3] If we cover the chapter on characteristic classes we will prove some of these results.

The example we started out with is the case where S is empty and the target category is VECT_κ for some field κ, so a topological field theory is a tensor functor from $\mathrm{Bord}_{\langle d-1,d\rangle}$ to VECT_κ.

For examples of this more general notion:

1. Use the identity functor! This gives what Michael Freedman calls the "lazy TFT" and it leads to a pairing of manifolds with very interesting positivity properties. See [18].

2. We can generalize this as follows: Let K be a closed manifold of dimension k. Then Cartesian product with K defines a symmetric tensor functor t_K
$$t_K : \mathrm{Bord}_{\langle d-1,d\rangle} \to \mathrm{Bord}_{\langle d+k-1,d+k\rangle} \tag{5.1}$$
where $t_K(Y) = Y \times K$, etc. If F is a $(d+k)$-dimensional TFT then we can compose $F \circ t_K$ to obtain a d-dimensional TFT denoted F^{KK}. This is the topological field theory analog of "Kaluza-Klein compactification". For example the state space on $(d-1)$-manifolds is
$$F^{KK}(Y) := F(Y \times K) \tag{5.2}$$

3. If there are sufficiently natural constructions of quantum field theories depending on some geometric category then one can define a TFT whose values are moduli spaces of vacua of the quantum field theory. This is done for the case of a target category of holomorphic symplectic varieties in [37].

6. Some General Properties

Let us deduce some simple general facts following from the above simple remarks.

For the moment take the target category to be SVECT_κ, the category of super-vector spaces over the field κ. (If one prefers, just ignore the signs and work with the category of vector spaces.)

First note that if X is closed then it can be regarded as a bordism from $\emptyset$ to $\emptyset$. Therefore $F(X)$ must be a linear map from κ to κ. But any linear map $T \in \mathrm{Hom}(\kappa,\kappa)$ must be of the form
$$T(z) = tz \tag{6.1}$$
for some scalar $t \in \kappa$. That is, any linear map $\kappa \to \kappa$ is canonically associated to an element of the ground field. For the case of $F(X) : \kappa \to \kappa$ we call that number the *partition function* of X, and denote it $Z(X)$.

There is one bordism which is distinguished, namely $[0,1] \times Y$. This corresponds to a linear map $P : \mathcal{H}(Y) \to \mathcal{H}(Y)$. In Euclidean field theory the amplitude one would associate to a cylindrical spacetime $[0,1] \times Y$ is just
$$\exp[-TH]$$

where H is the Hamiltonian, and T is the Euclidean time interval. Notice that this requires a metric. A change of the length of the cylinder leads to a change in T.

Evidently, by the axioms of topological field theory, $P^2 = P$ and therefore we can decompose

$$\mathcal{H}(Y) = P\mathcal{H}(Y) \oplus (1-P)\mathcal{H}(Y) \tag{6.2}$$

All possible transitions are zero on the second summand since, topologically, we can always insert such a cylinder. It follows that it is natural to assume that

$$F(Y \times [0,1]) = Id_{\mathcal{H}(Y)} \tag{6.3}$$

One can think of this as the statement that the Hamiltonian is zero. Note that this renders the amplitude independent of the length of the cylinder.

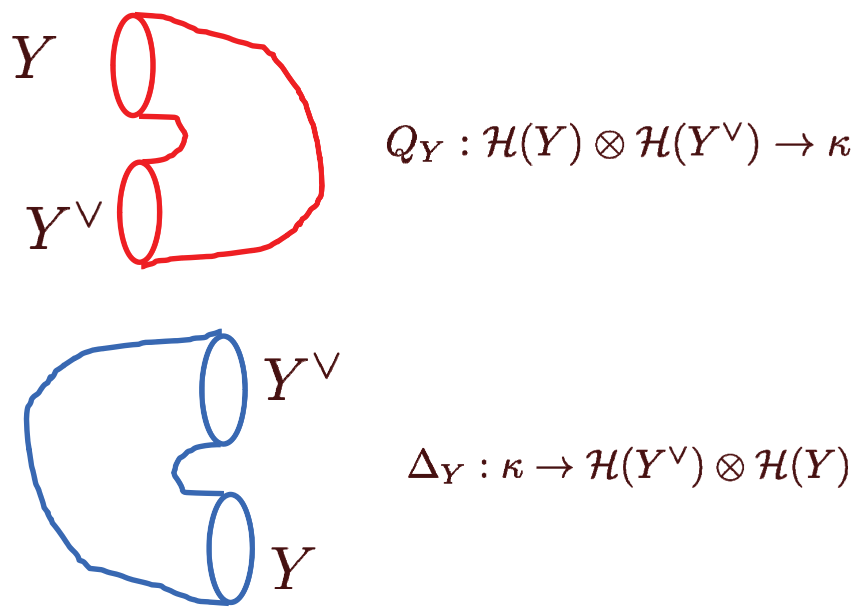


Figure 6: Bending the cylinder to define Δ_Y and Q_Y.

Now, let us consider the oriented bordism category, so Y is oriented. Let $Y^\vee$ denote Y with the opposite orientation. The bordism (6.3) is closely related to the bordism $\emptyset \to Y^\vee \amalg Y$ thus defining a map

$$\Delta_Y : \kappa \to \mathcal{H}(Y^\vee) \otimes \mathcal{H}(Y) \tag{6.4}$$

and also to a bordism $Y \amalg Y^\vee \to \emptyset$ thus defining a quadratic form:

$$Q_Y : \mathcal{H}(Y) \otimes \mathcal{H}(Y^\vee) \to \kappa \tag{6.5}$$

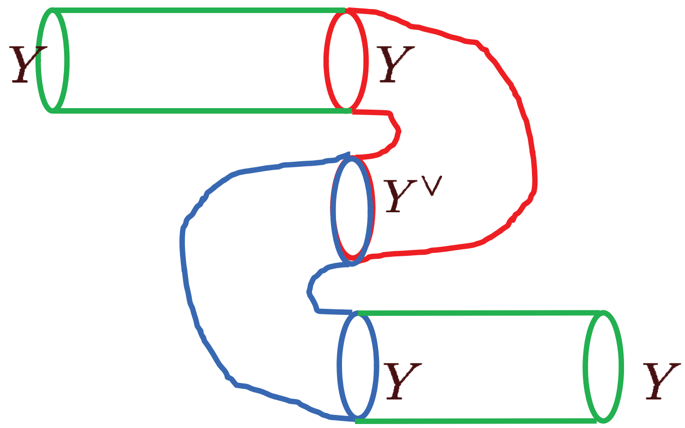


Figure 7: Composing $\Delta \otimes 1$ and $1 \otimes Q$ in a way that gives P.

Let us now compose these bordisms we get the identity map as in 7. It then follows from some linear algebra that Q is a *nondegenerate* pairing, so we have an isomorphism to the linear dual space:

$$\mathcal{H}(Y^\vee) \cong \mathcal{H}(Y)^\vee,$$

under which Q is just the dual pairing. (On the left $Y^\vee$ is Y is the reversal of orientation, and on the right $\mathcal{H}(Y)^\vee$ is the linear dual space.)

To prove this choose a basis $\{\phi_i\}$ for $\mathcal{H}(Y)$ and a basis $\{\psi_a\}$ for $\mathcal{H}(Y^\vee)$. Then we must have

$$\Delta_Y(1) = \sum_{i,a} \Delta^{ai} \psi_a \otimes \phi_i \tag{6.6}$$

The S-diagram shows that

$$\phi \to \sum_{i,a} \Delta^{ai} Q(\phi, \psi_a)\phi_i \tag{6.7}$$

must be the identity map, so, choosing $\phi = \phi_j$ and defining $Q_Y(\phi_j, \psi_a) := Q_{ja}$ we must have

$$\sum_a \Delta^{ai} Q_{ja} = \delta^i{}_j \tag{6.8}$$

In addition to this we can exchange the roles of Y and $Y^\vee$. Including signs for the $\mathbb{Z}_2$-graded case (with a homogeneous basis) we get

$$\sum_i \Delta^{ai} (-1)^{(|a|+|b|)|i|} Q_{ib} = \delta^a{}_b \tag{6.9}$$

It follows that Q is invertible, hence the pairing is nondegenerate. This implies hence there is an isomorphism $\mathcal{H}(Y^\vee) \cong \mathcal{H}(Y)^\vee$ as asserted above. Moreover, choosing an isomorphism so that $Q_{i,a} = \delta_{i,a}$, now labeling the dual basis by an index i and changing notation to $\psi_i \to \psi^i$ in this basis we have simply

$$\Delta_Y(1) = \sum_i \psi^i \otimes \phi_i \tag{6.10}$$

Now, the result (6.10) brings up an important point. It is not obvious that (6.10) will converge if $\mathcal{H}(Y)$ is infinite dimensional. In fact, even if $\mathcal{H}(Y)$ is a normed vector space, or a Hilbert space, so that convergence of infinite sums of vectors does make sense, since ϕ_i and ϕ^i are dual bases the sum will <u>not</u> converge if $\mathcal{H}(Y)$ is infinite dimensional. *Therefore, the space of states $\mathcal{H}(Y)$ must be finite-dimensional!*

There are many examples of interesting "topological field theories" where $\mathcal{H}(Y)$ is decidedly infinite-dimensional. We will comment on this below.

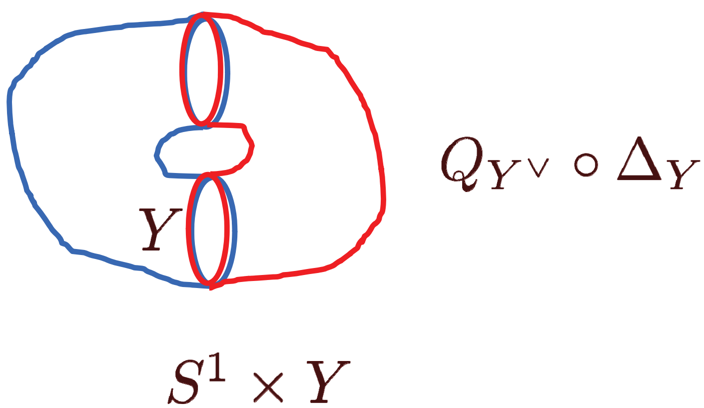


Figure 8: Composing $Q_{Y^\vee}$ with Δ_Y gives the super dimension of $\mathcal{H}(Y)$ in the $\mathbb{Z}_2$-graded case, and $\mathrm{dim}\mathcal{H}(Y) = Z(Y \times S^1)$ in the ungraded case.

Now consider the diagram in 8. On the one hand this is just the partition function $Z(Y \times S^1)$. On the other hand, the linear map $\kappa \to \kappa$ must be the composition $Q_{Y^\vee}\Delta_Y$, or, equivalently, $Q_Y \circ \Omega \circ \Delta_Y : \kappa \to \kappa$. From our formula for $\Delta_Y(1)$ above we see that the value $Z(Y \times S^1)$ is just the dimension $\mathrm{dim}\mathcal{H}(Y)$, or, in the $\mathbb{Z}_2$-graded case, the superdimension

$$\mathrm{sdim}\mathcal{H}(Y) = \mathrm{dim}\mathcal{H}(Y)_0 - \mathrm{dim}\mathcal{H}(Y)_1 \tag{6.11}$$

Remarks

1. Note that if we change the category to the category of manifolds with Riemannian structure and we take the product Riemannian structure on $Y \times S^1$ then

$$Z(Y \times S^1) = \mathrm{Tr} e^{-\beta H} \tag{6.12}$$

where β is the radius of the circle and H is the Hamiltonian.

2. There are important examples of "topological field theories" of interest in the physics literature where this condition is violated. One example is Chern-Simons theory with noncompact gauge group. Another example is two-dimensional Yang-Mills theory with zero area element. These are "partially defined" topological field theories. They are only defined on a subset of objects in the bordism category. ♣Say more. ♣

3. The S-diagram argument above points the way to a definition of a dual object in a symmetric monoidal category. A dual object $x \in \mathrm{Obj}(C)$ is one such that there exists an object $x^\vee \in \mathrm{Obj}(C)$ and morphisms $\delta_x : 1_C \to x \otimes x^\vee$ and $q_x : x^\vee \otimes x \to 1_C$ such that

$$x \xrightarrow{\iota_L(x)^{-1}} 1_C \otimes x \xrightarrow{\delta_x \otimes 1_x} x \otimes x^\vee \otimes x \xrightarrow{1_x \otimes q_x} x \otimes 1_C \xrightarrow{\iota_R(x)} x \tag{6.13}$$

and (omitting the isomorphisms with multiplication by the tensor unit, for simplicity)

$$x^\vee \xrightarrow{1_{x^\vee} \otimes \delta_x} x^\vee \otimes x \otimes x^\vee \xrightarrow{q_x \otimes 1_{x^\vee}} x^\vee \tag{6.14}$$

are the identity morphisms.

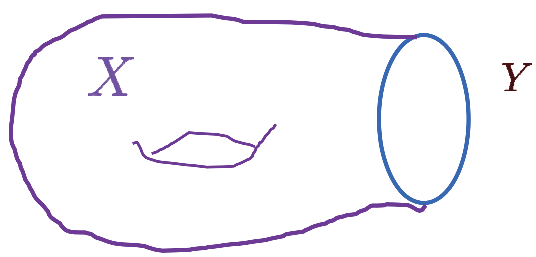


Figure 9: A state created by a bordism of $\emptyset$ to Y.

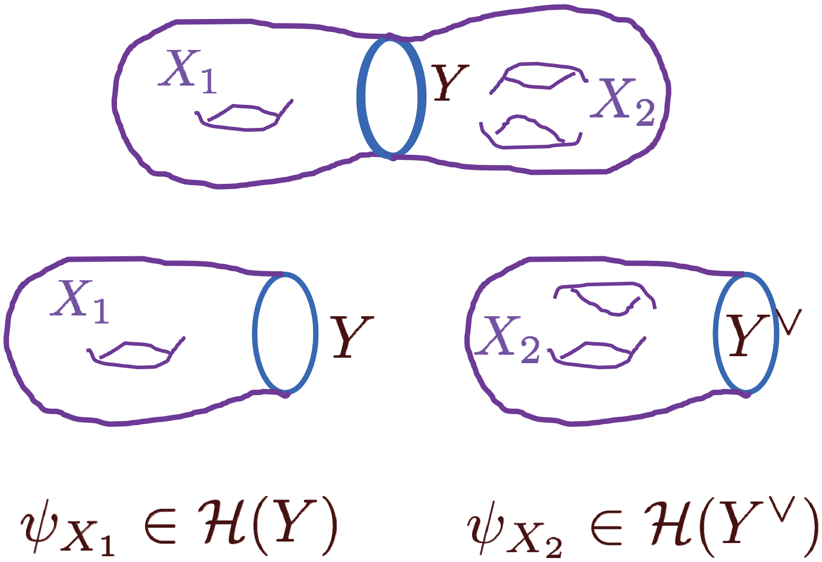


Figure 10: If a closed manifold X is cut along a codimension one submanifold Y that divides X into two pieces X_1 and X_2 then there are two associated states $\psi_{X_1} \in \mathcal{H}(Y)$ and $\psi_{X_2} \in \mathcal{H}(Y^\vee)$, and the value of the partition function $Z(X)$ may be viewed as the natural contraction of these states using the nondegenerate pairing Q_Y.

Exercise *Mapping cylinders and characters of the diffeomorphism group*

Let $f \in \mathrm{Diff}(Y)$ and consider the mapping cylinder $M_f(Y) = ([0,1] \times Y)/\sim$ where we identify $(0,y)$ with $(1, f(y))$. Recall that $\mathcal{H}(Y)$ has a representation $\rho(f)$ of the diffeomorphism group.

Show that

$$Z(M_f(Y)) = \mathrm{Tr}_{\mathcal{H}(Y)} \rho(f) \tag{6.15}$$

is a character of the diffeomorphism group.

In fact, $\rho(f)$ only depends on the image of f in the mapping class group Γ_Y: This is defined as follows: The diffeomorphisms isotopic to the identity form a normal subgroup $\mathrm{Diff}_0(Y)$ of the full diffeomorphism group and $\Gamma_Y := \mathrm{Diff}(Y)/\mathrm{Diff}_0(Y)$.

♣Explain in detail how the independence of $\rho(f)$ under isotopy of f follows from the axioms. ♣

Exercise *Hartle-Hawking sates and partition functions as inner products*

a.) Show that any bordism of $X : \emptyset \to Y$ defines a *state* in the space $\mathcal{H}(Y)$. (See Figure 9.) The functor of the topological field theory defines a map $F(X) : \kappa \to \mathcal{H}(Y)$, and we can define,

$$\psi_X := F(X)(1) \in \mathcal{H}(Y) \tag{6.16}$$

This simple observation is very important in physics.

The state, of course, depends on the (topological) details of the bordism. For example, any Riemann surface with a single hole defines a bordism of the circle to zero and there are many such topologies. This is a primitive version of the notion of the "Hartle-Hawking" state in quantum gravity. It is also related to the state/operator correspondence in conformal field theory.

b.) Show that, in the oriented bordism category, by exchanging in and out boundaries (but not the orientation of X) the same manifold defines a bordism $X^\vee : -Y \to \emptyset$, and hence a linear functional on $\mathcal{H}(Y^\vee)$.

c.) Show that applying this linear functional to $\Delta_Y(1)$ gives back the original vector in $\mathcal{H}(Y)$ associated to X.

d.) Show that if a closed manifold X is cut along an oriented manifold Y to produce X_1 and X_2 then $Z(X)$ can be interpreted as a contraction of a state $\psi_{X_1} \in \mathcal{H}(Y)$ and $\psi_{X_2}^\vee \in \mathcal{H}(Y^\vee)$:

$$Z(X) = \langle \psi_{X_2}^\vee, \psi_{X_1} \rangle \tag{6.17}$$

See Figure 10.

6.1 Unitarity

In unitary theories, and certainly in the axioms of quantum mechanics, one wants the state space to be a complex Hilbert space, and $F(X)$ for a bordism X should be a unitary operator.

Now, in general, a sesquilinear form on a complex vector space V is a linear map $V \to \bar{V}^\vee$. Therefore, in a unitary theory changing orientation of Y complex conjugates the Hilbert space

$$\mathcal{H}(Y^\vee) \cong \bar{\mathcal{H}}(Y) \tag{6.18}$$

Moreover, in physical unitary theories there is a positivity condition on Q_Y. If $X : Y_1 \to Y_2$ is a bordism then, if we change the orientation of X and take the dual we get a bordism

$$\bar{X}^\vee : Y_2 \to Y_1 \tag{6.19}$$

It is natural to add a condition that

$$F(\bar{X}^\vee) = F(X)^\dagger \tag{6.20}$$

In particular, changing orientation of the manifold invariant $Z(X)$ for a closed manifold complex conjugates the invariant.

7. One Dimensional Topological Field Theory

Consider the oriented case. Then the objects in $\mathrm{Bord}^{\mathrm{SO}}_{\langle 0,1\rangle}$ are disjoint unions of points $\mathrm{pt}_\pm$ with $+$ and $-$ orientation.

The topological field theory with symmetric monoidal category C gives two objects $y_\pm = F(\mathrm{pt}_\pm)$ with data δ and q as described above. The general object is a disjoint union of $n_\pm$ points of type $\mathrm{pt}_\pm$. The diffeomorphism group of this manifold is just $S_{n_+} \times S_{n_-}$ and it acts in a natural way on the "state space" $y_+^{\otimes n_+} \otimes y_-^{\otimes n_-}$.

Specializing to VECT_κ, we get a pair of finite-dimensional vector spaces $V_\pm$ together with the data mentioned above: A nondegenerate pairing $Q : V_+ \otimes V_- \to \kappa$ and the "inverse" $\Delta : \kappa \to V_+ \otimes V_-$. As mentioned above, these constitute duality data for $V_- = V_+^\vee$.

A good example of a physical origin of such a topological field theory is to consider quantization of a compact symplectic manifold (M, ω).

A useful concrete example to keep in mind is $M = S^2$ with a symplectic form

$$\omega = \frac{1}{2\hbar} \sin\theta d\theta d\phi \tag{7.1}$$

where here $\hbar$ is some dimensionless normalization of the form.

In the Hamiltonian formulation of the path integral we consider paths in phase space M. We form a path integral of the form

$$\int_{\mathcal{P}} [d\gamma] \exp[\mathrm{i}S] \tag{7.2}$$

where $\mathcal{P}$ is a space of paths in M, $[d\gamma]$ is an induced measure on the space of paths from the symplectic form, and S is an action. There are many issues to settle in making sense of this expressions. We will just touch on a few of them here.

If the symplectic form ω is globally exact then we can write $\omega = d\lambda$ where, in terms of local Darboux coordinates

$$\lambda = \frac{1}{\hbar} p dq \tag{7.3}$$

A good example of this is the case $M = T^*X$ for some manifold X. Note that the Hamiltonian associated with the action principle:

$$S[\gamma] = \frac{1}{\hbar} \int_\gamma p dq \tag{7.4}$$

is zero.

But what if ω is not exact? (As in our above example with $M = S^2$.) Let us suppose first that M is simply connected. Then, if γ is a closed path we can attempt to define the action by choosing a disk $\Sigma \subset M$ such that $\partial\Sigma = \gamma$ and then take

$$S_\Sigma[\gamma] := \int_\Sigma \omega \tag{7.5}$$

If ω is exact this reduces to the previous definition.

Now there is a problem because there can be more than one disk bounding γ. If Σ_1, Σ_2 both bound γ then $\Sigma_{12} := \Sigma_1 \cup_\gamma \Sigma_2^\vee$ is a closed 2-cycle and the ambiguity in the action is

$$S_{\Sigma_1}[\gamma] - S_{\Sigma_2}[\gamma] = \int_{\Sigma_{12}} \omega \tag{7.6}$$

So the action is not well-defined. However, all we need for the quantum path integral is that the weight

$$\exp[\mathrm{i}S] = \exp[\mathrm{i} \int_\Sigma \omega] \tag{7.7}$$

should be well-defined. The ambiguity in the exponentiated action is:

$$\exp[\mathrm{i}S_1]/\exp[\mathrm{i}S_2] = \exp[\mathrm{i}\int_{\Sigma_{12}} \omega] \tag{7.8}$$

The LHS will be one - and there will be an unambiguous weight in the path integral - if the periods of ω are integral multiples of 2π. Notice that this quantizes "$1/\hbar$" to be an integer.

Now, suppose that γ is not closed. Let us consider a path space

$$\mathcal{P} = \{\gamma : [0,1] \to M | \gamma(0) = x_0 \quad \& \quad \gamma(1) = x_1\} \tag{7.9}$$

(We assume x_0, x_1 are in the same path-connected component of M.) Choose a basepoint path γ_0 in $\mathcal{P}$. Then any other path homotopic to γ_0 will be such that $\gamma_0^{-1} * \gamma$ bounds a disk Σ. We then use this data to define an action as

$$S_{\gamma_0,\Sigma}[\gamma] := \int_\Sigma \omega. \tag{7.10}$$

For a fixed basepath the exponentiated action will be independent of the choice of Σ if the periods of ω are in $2\pi\mathbb{Z}$. If we change the basepath γ_0 to another one in the same homotopy class then the action only shifts by a constant, and in fact with ω quantized as above, the choice will again not matter in the exponentiated action.

If M is not simply connected further considerations are needed because there will be paths in $\mathcal{P}$ not in the path-component of γ_0 even when x_0, x_1 are in the same path component of M. One way to deal with this is to work on the universal cover $\widetilde{M}$. It is best to couple the theory to a flat connection on M to keep track of the fundamental group. ♣EXPLAIN MORE! ♣

An important special case of the quantization above is the case of coadjoint orbits of a compact simple Lie group defined by integral weights $\lambda \in \mathfrak{g}^*$. There is a natural integrally-quantized symplectic form - the Kirillov-Kostant symplectic form, and quantization gives a representation with dominant weight vector a suitable Weyl rotation of λ. Pursuing this line of thought leads to a path integral interpretation of the Borel-Weil-Bott theorem. In ♣Do it later? ♣ the topological field theory the space V_+ is the representation with dominant weight λ and the space V_- is the conjugate representation with anti-dominant weight $-\lambda$. The duality data Q is the standard pairing of a representation and its conjugate to form the singlet while Δ is the embedding of the singlet into $R \otimes R^\vee$.

8. Two Dimensional TFT And Commutative Frobenius Algebras

Some beautiful extra structure shows up when we consider the case $d = 2$, due to the relatively simple nature of the topology of 1-manifolds and 2-manifolds.

We restrict attention to the oriented case. The unoriented case presents new and interesting features related to the theory of "orientifolds."

To begin, we restrict attention to closed $(d-1)$-manifolds, that is, we consider a theory of closed strings.

In this case, the spatial $(d-1)$ manifolds are necessarily of the form $S^1 \cup S^1 \cup \cdots \cup S^1$, i.e. disjoint unions of n S^1's.

The circle with have two orientations $S^1_\pm$. We choose one, say ccw on the boundary of the unit disk and set $\mathcal{C} := \mathcal{H}(S^1_+)$. Then, as usual we have duality data for $\mathcal{H}(S^1_-)$. We henceforth focus on $\mathcal{C}$.

$$\mathcal{H}(S^1 \amalg S^1 \amalg \cdots \amalg S^1) = \mathcal{C}^{\otimes n}$$

Thus, thanks to the simple topology of closed 1-manifolds, the entire theory is built from a single vector space $\mathcal{C}$.

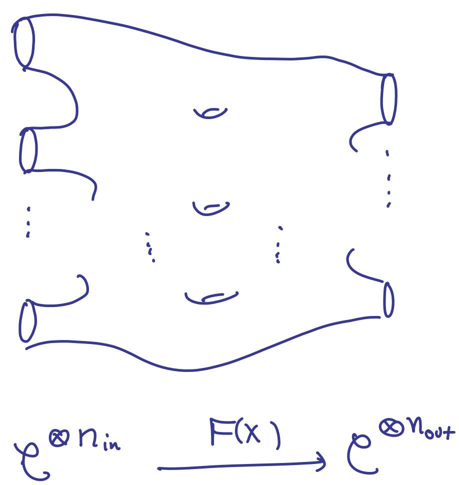


Figure 11: Typical Riemann surface amplitude. If there are n_{in} ingoing circles and n_{out} outgoing circles then the corresponding amplitude is a linear map $F(X) : \mathcal{C}^{n_{\text{in}}} \to \mathcal{C}^{n_{\text{out}}}$.

Transition amplitudes can be pictured as in 11. We therefore get a linear map:

$$F(\Sigma) : \mathcal{C}^{\otimes n} \to \mathcal{C}^{\otimes m}$$

where Σ is a Riemann surface. There are n ingoing circles and m outgoing circles.

Now, topology dictates that the vector space $\mathcal{C}$ in fact must carry some interesting extra structure.

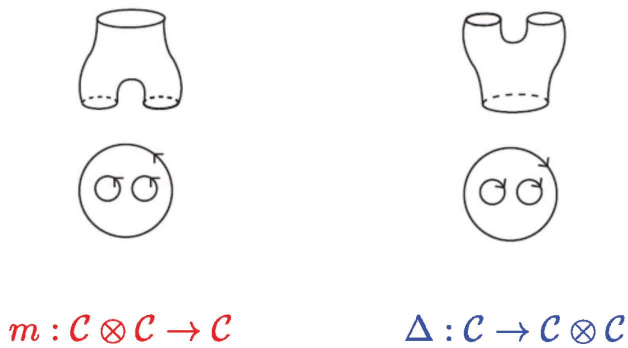


Figure 12: The sphere with 3 holes defines m and Δ

In Figure 12 we see that the sphere with three holes defines a product

$$m : \mathcal{C} \otimes \mathcal{C} \to \mathcal{C}$$

Exchanging in and out boundaries we get a coproduct:

$$\Delta : \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$$

In 13 we see that there is a trace (a.k.a a *counit*):

$$\theta : \mathcal{C} \to \kappa$$

In addition, in Figure 13 we see that there is a map $\kappa \to \mathcal{C}$. This is completely determined by its value on $1 \in \kappa$. The image of $1 \in \kappa$ in $\mathcal{C}$ under this map is denoted $1_{\mathcal{C}}$, and this element indeed functions as a unit for the multiplication m. From the diagram in Figure 14 we see that the image of 1 must be in fact a unit for the multiplication.

Moreover, from 15 we see the multiplication is associative.

We can now consider the compositions $\theta \circ m : \mathcal{C} \otimes \mathcal{C} \to \kappa$ and $\Delta \circ UNIT : \kappa \to \mathcal{C} \otimes \mathcal{C}$. Note well that these operations are <u>different</u> from the duality data Q, Δ_Y discussed before because they involve the <u>same</u> space $\mathcal{C}$, rather than $\mathcal{C}$ and its dual! However, the same S-diagram argument shows that the quadratic form

$$Q(\phi_1, \phi_2) := \theta(\phi_1 \phi_2) \tag{8.1}$$

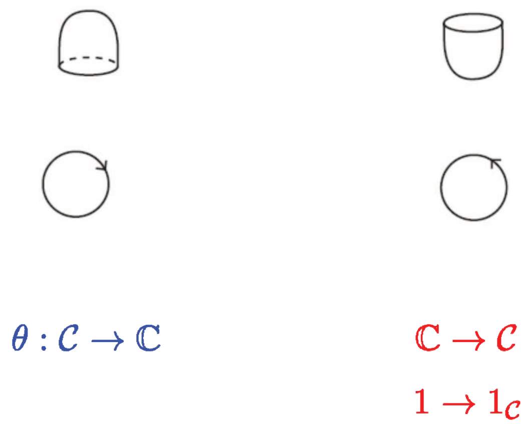


Figure 13: The trace map and the unit.

is nondegenerate.

Finally, we can make a diffeomorphism of the disk with 2 holes, holding the outer circle fixed and rotating the inner two circles. This shows that the product must be (graded) *commutative*.

The algebraic structure we have discovered is known as a *Frobenius algebra*.

Definition. A vector space V is an *associative algebra* if there is a multiplication $v_1 v_2 \in V$ satisfying

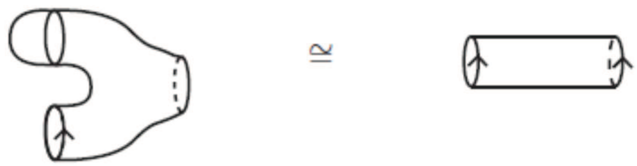
1. $v_1(v_2 v_3) = (v_1 v_2) v_3$
2. $(v_1 + v_2)v_3 = v_1 v_3 + v_2 v_3 \qquad v_3(v_1 + v_2) = v_3 v_1 + v_3 v_2$
3. $\alpha(v_1 v_2) = (\alpha v_1) v_2 = v_1(\alpha v_2)$

for all vectors $v_1, v_2, v_3 \in V$ and all scalars α in the ground field.

Definition A *Frobenius algebra* A is an associative algebra over a field k with a trace $\theta : A \to k$ such that the quadratic form $A \otimes A \to k$ given by $a \otimes b \to \theta(ab)$ defines a nondegenerate bilinear form on A.

What we have shown at this point is that $\mathcal{H}(S^1)$ is a (graded) commutative Frobenius algebra.

The data of the Frobenius algebra is sufficient to compute arbitrary amplitudes: Any oriented surface can be decomposed into the basic building blocks we have used above. However, the same surface can be decomposed in many different ways. When we have



$$1_{\mathcal{C}} \cdot \phi = \phi$$

Figure 14: Proof that $1_{\mathcal{C}}$ really is a unit.

different decompositions we get algebraic relations on the basic data $m, \Delta, \theta_{\mathcal{C}}$. At this point you might well ask: "Can we get more elaborate relations on the algebraic data by cutting up complicated surfaces in different ways?" That is, are we required to consider only special kinds of Frobenius algebras? The beautiful answer is: "No, the above relations are the only independent relations, so, conversely, any (graded) commutative Frobenius algebra defines an oriented d=2 TFT." We call this statement the "sewing theorem."

Exercise

a.) If A is an algebra, then it is a module over itself, via the left-regular representation (LRR). $a \to L(a)$ where

$$L(a) \cdot b := ab$$

Show that if we choose a basis a_i then the structure constants

$$a_i a_j = c_{ij}{}^k a_k$$

define the matrix elements of the LRR:

$$(L(a_i))_j{}^k = c_{ij}^k$$

b.) If A is an algebra, the *opposite algebra* is the algebra with the new product

$$v_1 \circ v_2 := v_2 v_1 \tag{8.2}$$

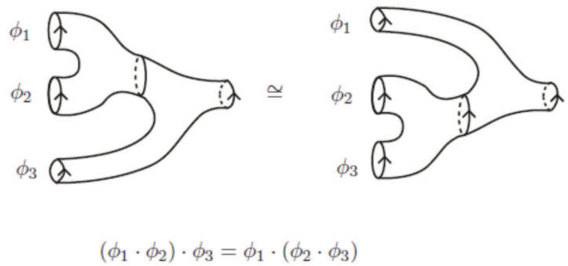


Figure 15: Associativity.

Show that A is a bimodule over $A \otimes A^o$ where A^o is the opposite algebra.

c.) Show that if (A, θ) is a Frobenius algebra then the dual algebra A^* is a left A-module which is isomorphic to A as a left A-module.

Exercise

a.) Show the equality of maps $\mathcal{C} \otimes \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$

$$(Id \otimes m) \circ (\Delta \otimes Id) = \Delta \circ m = (m \otimes Id) \circ (Id \otimes \Delta) \tag{8.3}$$

b.) Show that

$$(Id \otimes \theta) \circ \Delta = Id = (\theta \otimes Id) \circ \Delta \tag{8.4}$$

8.1 The Sewing Theorem

The Sewing Theorem. To give a 2d topological field theory is equivalent to giving a commutative associative finite dimensional Frobenius algebra.

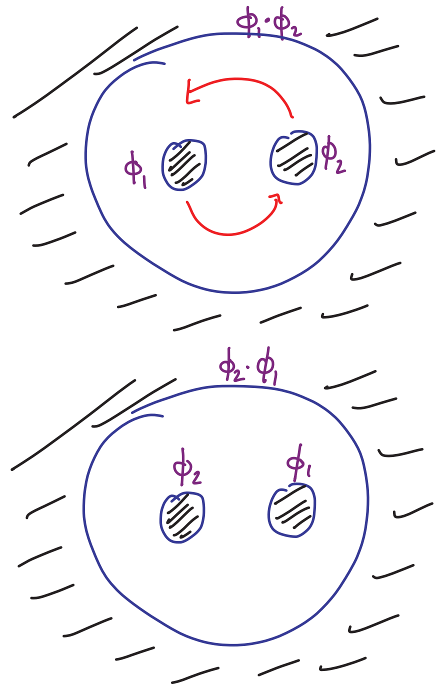


Figure 16: Commutativity. The outer boundary of the disk is held fixed and the two inner boundaries are rotated in a basic braiding action. In order for the diffeomorphism of the coboundary to be one on the ingoing circles we must compose with Ω, so in the supercase we will have graded commutativity: $\phi_1\phi_2 = (-1)^{\deg\phi_1 \deg\phi_2}\phi_2\phi_1$.

Proof: In one direction the theorem is obvious. Given a 2d topological field theory one recovers a commutative Frobenius algebra as described above.

What is not immediately obvious is the converse. Given a commutative Frobenius algebra $(\mathcal{C}, \theta)$ one defines the functor on the special surfaces as above, but in principle further restrictions on the data could arise from consistency of gluing.

Consider an oriented surface with boundary Σ. Different sewings correspond to different choices of "time-slicings." We can make this precise by associating the time-slicings with level sets of a suitable smooth function $f : \Sigma \to \mathbb{R}$, as in 17. We can assume Σ is an orientable surface in $\mathbb{R}^3$ with boundary circles unknotted and unlinked. We identify "time evolution" corresponding to evolution to larger values of $f = t$. The generic level set is a 1-manifold. The basic idea is to break up the time evolution into elementary steps given by the basic data of the Frobenius algebra and then prove that the result is the same for two such functions by studying how the sequence of elementary steps changes when we

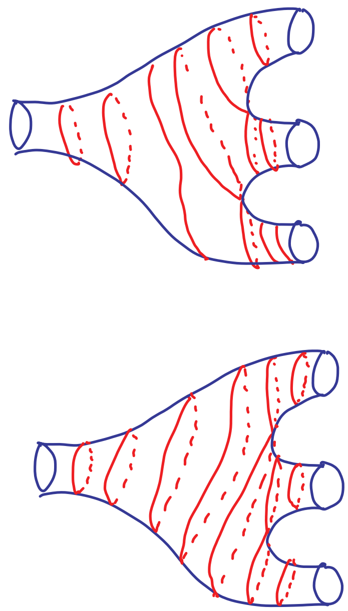


Figure 17: Two different Morse functions define different decompositions of the same surface.

connect the two functions by a path of functions.

In order to describe this "path of functions" with any precision we need a tiny digression into singularity theory.

8.1.1 A Little Singularity Theory

We will be considering the space of $\mathcal{C}^\infty$ real-valued functions on a bordism X.

The space of functions $\mathcal{C}^\infty(M_1, M_2)$ from a closed smooth manifold M_1 to another manifold M_2 will be endowed with the *Whitney topology*.

There is a basis for the Whitney topology of the form:

$$\mathcal{U}(f, (U, \phi), (V, \psi), K, \epsilon) \tag{8.5}$$

where $f \in \mathcal{C}^\infty(M_1, M_2)$, $\epsilon > 0$, (U, ϕ) is a smooth chart on M_1, $K \subset U$ is a compact subset, and (V, ψ) is a chart on M_2. Relative to the charts (U, ϕ) and (V, ψ) any function $g \in \mathcal{C}^\infty(M_1, M_2)$ can be represented as a smooth map

$$\tilde{g} : \tilde{U} \subset \mathbb{R}^{\dim M_1} \to \mathbb{R}^{\dim M_2} \tag{8.6}$$

Then the open set $\mathcal{U}(f,(U,\phi),(V,\psi),K,\epsilon)$ consists of all those functions g such that

$$\sup_{x\in\tilde{K}}|D^\alpha g^j(x) - D^\alpha f^j(x)| < \epsilon \qquad \forall\alpha, 1\le j\le \dim M_2 \tag{8.7}$$

Here $\alpha=(i_1,\dots,i_s)$ is a multi-index with $1\le i\le \dim M_1$ and

$$D^\alpha = \frac{\partial}{\partial x^{i_1}}\cdots\frac{\partial}{\partial x^{i_s}} \tag{8.8}$$

Now, a function $f: M\to\mathbb{R}$ is called a *Morse function* if it has finitely many *critical points* p: That is, points where $df(p)=0$ and, moreover, the Hessian at p

$$H_{ij}(p) = \frac{\partial^2 f}{\partial x^i \partial x^j} \tag{8.9}$$

is nondegenerate in any coordinate system near p. (Since p is a critical point the Hessian transforms by $H\to SHS^{tr}$ for a nondegenerate matrix S under coordinate transformation.) In addition we require that the *critical values* $c_p = f(p)$ are all distinct.

We now consider Morse functions on a bordism X from Y_0 to Y_1. A Morse function on such a bordism is said to be *excellent* if f is constant on Y_0 and Y_1 and the critical values can be ordered so that

$$a_0 = f(Y_0) < c_1 < \cdots < c_N < a_1 = f(Y_1) \tag{8.10}$$

We will use excellent Morse functions to give our time-slicings. An excellent Morse function is said to be *elementary* if it contains precisely one critical point. Thus, in two dimensions it will contain a maximum, a minimum, or a saddle.

If we give X an elementary Morse function then the corresponding morphism $F(X)$ in the topological field theory is:

1. Maximum: Trace
2. Minimum: Unit
3. Saddle: Multiplication or comultiplication.

One can prove [4] that if f is an excellent Morse function on a bordism X then X can be decomposed into elementary bordisms: We can write X as a gluing of $X_1,\dots,X_M$ such that f restricted to each X_α is an elementary Morse function. In this way an excellent Morse function gives a definite series of algebraic operations in computing the amplitude $F(X)$.

♣Need to say this is essentially unique. ♣

Now we can ask our question more precisely: Given two excellent Morse functions f_{-1} and f_{+1} will they produce the same amplitude?

Now, suppose we have a continuous (in the Whitney topology) path of f_t of excellent Morse functions. Then the sequence of algebraic operations is unchanged, and hence the amplitude $F(X)$ remains constant.

In the Whitney topology the space of excellent Morse functions in $\mathcal{C}^\infty(X,\mathbb{R})$ is open and dense, but is <u>disconnected</u>. Thus, if we have a generic path f_t of $\mathcal{C}^\infty$ functions from f_{-1}

[4] Apparently, the proofs for the following statements can be found in J. Cerf, "La stratification naturelle des espaces de fonctions differentiables reeles et la theorem de la pseudoisotopie," Publ. Math. IHES **39**(1970).

to f_{+1} there will be some values of t for which f_t fails to be an excellent Morse function. There will be generically codimension one walls where the functions are not excellent.

We define a function to be *good* (but not necessarily excellent) if it is either excellent or excellent except at one or two critical points such that:

1. Type A: If there is one point then at this point f is not Morse and in local coordinates is of the form $\pm y^2 + x^3$.

2. Type B: If there are two critical points they are Morse but have the same critical value.

Now the crucial point which follows from Cerf theory is that the space of excellent and good functions is <u>connected</u>. In fact the space of good functions is an infinite-dimensional manifolds and the space of functions which are good but not excellent is a real codimension one submanifold. A generic path of good functions f_t between two excellent functions will cross these walls in a finite set of points, that is, f_t will be excellent except at a finite set of critical values t_i.

8.1.2 Proof Of The Sewing Theorem

We now show that, given the axioms of a Frobenius algebra, if we evolve f_s through functions of Type A or Type B then the resulting map $F(\Sigma)$ is unchanged.

Note that reversing the time direction of the bordism changes an operator by its adjoint with respect to the Frobenius inner product. This reduces the number of cases we must check by a factor of two.

♣This picture violates the convention of ingoing on left and outgoing on right. ♣

For Type A: We may assume that for small s the family of functions is (near the critical point) $f_s = y^2 + x^3 + sx$. Comparing evolutions for $s = 0-$ and $s = 0+$ we obtain Figure 18. This encodes the axiom for the unit and trace.

For Type B, it helps to note first that if Φ and Φ' are two linear maps then $\Phi \otimes 1$ and $1 \otimes \Phi'$ commute. Geometrically this means that we need only consider the cases where the level set containing the two critical points is connected. Thus they both have Morse index 1.

Since both critical points have index 1 they map 2 circles to 1 circle, or 1 circle to 2 circles. Thus the only evolutions we can have are

$$\begin{aligned} &1 \to 2 \to 1 \\ &1 \to 2 \to 3 \\ &3 \to 2 \to 1 \\ &2 \to 1 \to 2 \\ &2 \to 3 \to 2 \end{aligned} \tag{8.11}$$

Moreover, the contour lines of the two degenerate critical points must be one of the two cases in 19.

In the first case $1 \to 2 \to 1$ there is only one possible cobordism, so there is nothing to check.

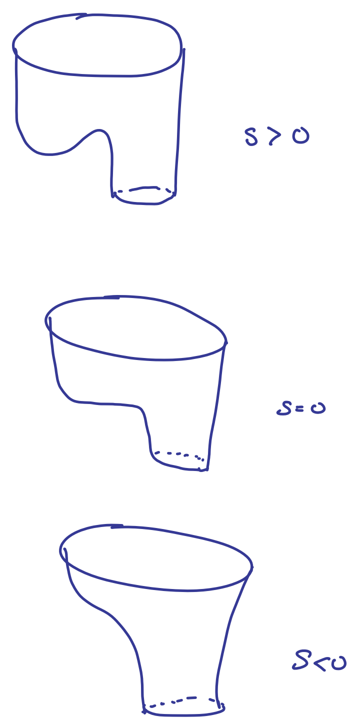


Figure 18: The different time evolutions for $y^2 + x^3 + sx$ as s moves through $s = 0$.

In the next case one circle maps to 3 circles. This is shown in Figure 17. A small perturbation leads to one of the two cuttings shown in that figure. So this is covered by the associativity axiom.

♣This picture violates the convention of ingoing on left and outgoing on right. ♣

In the second case, we have two circles mapping to two circles. That is, we are cutting the 4-holed sphere in different ways. The only algebraic maps which these bordisms can lead to are

$$\mathcal{C} \otimes \mathcal{C} \to \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$$

and

$$\mathcal{C} \otimes \mathcal{C} \to \mathcal{C} \otimes \mathcal{C} \otimes \mathcal{C} \to \mathcal{C} \otimes \mathcal{C}$$

These maps correspond to the decompositions shown in Figure 20. Algebraically they read:

$$\phi \otimes \phi' \mapsto \phi\phi' \mapsto \sum \phi\phi'\phi^\mu \otimes \phi_\mu$$

and

$$\phi \otimes \phi' \mapsto \sum \phi\phi^\mu \otimes \phi_\mu \otimes \phi' \mapsto \sum \phi\phi^\mu \otimes \phi_\mu\phi'$$

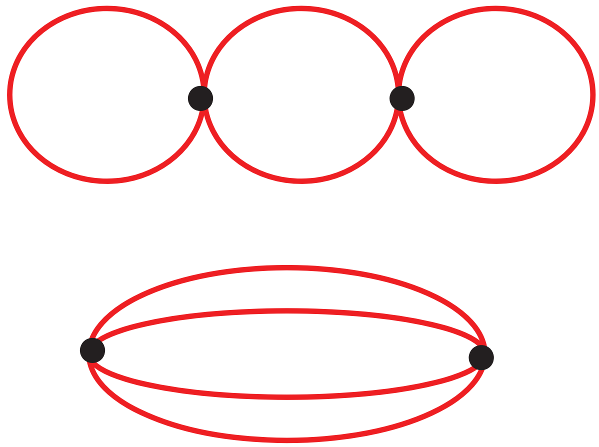


Figure 19: Two ways to connnect simultaneously occuring saddle points.

respectively, where $\{\phi_\mu\}$ and $\{\phi^\mu\}$ are dual bases of $\mathcal{C}$ such that $\theta_{\mathcal{C}}(\phi^\mu\phi_\nu) = \delta^\mu_\nu$. These two maps are equal because of the identity

$$\sum \phi'\phi_\mu \otimes \phi^\mu = \sum \phi_\mu \otimes \phi^\mu\phi', \tag{8.12}$$

Equation (8.12) holds in any Frobenius algebra (commutative or not) because the inner product of each side with $\phi^\nu \otimes \phi_\lambda$ is $\theta_{\mathcal{C}}(\phi^\nu\phi'\phi_\lambda) = \theta_{\mathcal{C}}(\phi_\lambda\phi^\nu\phi')$. ♠

Exercise

Show in a d dimensional TFT given by

$$F : \mathrm{Bord}^{SO}_{\langle d-1,d\rangle} \to \mathrm{VECT}_\kappa \tag{8.13}$$

the spaces $\mathcal{H}(S^{d-1})$ are commutative Frobenius algebras over κ.

b.) Show that there is an action of this Frobenius algebra on $\mathcal{H}(\Sigma)$ where Σ is any oriented $(d-1)$-manifold

---

8.2 Remarks On The Unoriented Case

In the case of an unoriented TFT

$$F : \mathrm{Bord}_{\langle d-1,d\rangle} \to \mathrm{VECT}_\kappa \tag{8.14}$$

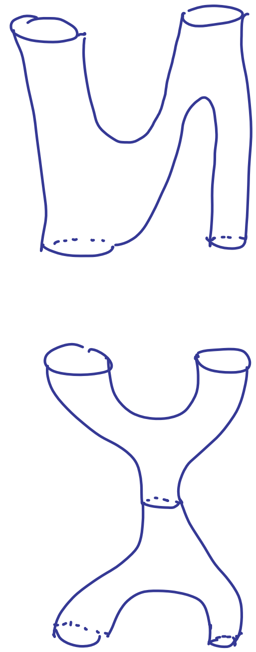


Figure 20: One or three circles in the intermediate channel of a degenerate Morse function.

it is still true that $\mathcal{C} = F(S^{d-1})$ is a commutative Frobenius algebra since an orientation was not used in the derivation of the Frobenius structure. But in the unoriented case $\mathcal{C}$ has *extra* structure. [5] There are two new ingredients:

1. Now orientation reversing diffeomorphisms act on $\mathcal{C}$. By the stable homeomorphism conjecture, the group of orientation preserving homeomorphisms of the sphere is connected. [6]

[5] We are following here a nice exposition in V. Turaev and P. Turner, "Unoriented topological quantum field theory and link homology," arXiv:math/0506229; Algebr. Geom. Topol. 6 (2006) 1069-1093. The same axioms were worked out a few years earlier in unpublished work by Ilka Brunner and myself.

[6] This is by no means a trivial statement. A homeomorphism $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *stable* if it is the identity on some open set of $\mathbb{R}^n$. The *stable homeomorphism conjecture* is the conjecture that every orientation-preserving homeomorphism $f : \mathbb{R}^n \to \mathbb{R}^n$ is a composition of a finite number of stable homeomorphisms. It is known to be true for all $n > 0$, although the proof is not simple, especially for $n = 4$. Now, an orientation-preserving homeomorphism $f : S^n \to S^n$ can be composed with a stable one so that it has a fixed point, and then from stereographic projection from that point we get a homeomorphism $\mathbb{R}^n \to \mathbb{R}^n$. Now we claim that a stable homeomorphism can be isotoped to the identity. Clearly we can

Therefore, up to isotopy there is only one orientation reversing diffeomorphism. This induces a transformation $\Omega : \mathcal{C} \to \mathcal{C}$ which squares to one.

2. For d even $\mathbb{RP}^d$ is unorientable, so in addition to the identity element - coming from S^d minus a d-dimensional ball we have a new element from $X : \emptyset^{d-1} \to S^{d-1}$ given by $\mathbb{RP}^d$ minus a d-dimensional ball. This is the state $\psi_X := F(X)(1)$. We will call it c (for "crosscap state").

The main new constraints coming from elementary bordisms are:

1. Ω is an isomorphism of Frobenius algebras. In particular

$$\Omega(\phi_1) \cdot \Omega(\phi_2) = \Omega(\phi_1 \cdot \phi_2) \tag{8.16}$$

while $\Omega \circ \iota = \iota$ and $\theta \circ \Omega = \theta$.

2. If d is even then

$$\Omega(c\phi) = c\phi \qquad \forall \phi \in \mathcal{C} \tag{8.17}$$

3. If $d = 2$ then

$$m(\Omega \otimes \mathrm{Id})(\Delta(1)) = c^2 \tag{8.18}$$

To prove these we note for the first statement that the orientation reversing transformation given by flipping the sign of one (any) coordinate in the unit ball preserves the ball. This shows that Ω preserves the unit and counit. Similarly, if we consider the multiplication to be given by a large unit ball centered on the origin with two small balls symmetric under $x^1 \to -x^1$ cut out then (combining with commutativity) this shows that $\Omega \circ m \circ (\Omega \otimes \Omega) = m$ but now use the fact that Ω is an involution.

For the second, we regard $\mathbb{RP}^d$ as the unit sphere in $\mathbb{R}^{d+1}$ modulo $x \to -x$. Now remove two disjoint d-dimensional balls B_1, B_2 with boundaries Σ_1 and Σ_2. Choosing a fundamental domain $x^{d+1} \geq 0$ we arrive at the picture in Figure 21. As explained in the figure caption if we regard this as a bordism $X : \Sigma_1 \to \Sigma_2$ then the image of a state ϕ on Σ_1 is $c\phi$ at Σ_2. On the other hand, since $\mathbb{RP}^d$ is unorientable the orientation reversing transformation $x \to -x$ on S^d descends to the identity, and is in the same component as an isotopy that preserves B_1 but takes B_2 to its image under $x \to -x$. But this homeomorphism acts as 1 on Σ_1 and reverses orientation on Σ_2. Therefore the diagram also produces $\Omega(c\phi)$ and hence

$$\Omega(c\phi) = c\phi \tag{8.19}$$

---

assume that it takes the unit ball to the unit ball and it is the identity outside the (open) unit ball. Now we use the Alexander trick (see Wikipedia article):

$$J(x,t) = \begin{cases} tf(x/t) & 0 \leq |x| < t \\ x & t \leq |x| \leq 1 \end{cases} \tag{8.15}$$

As t goes from 1 to 0 we get an isotopy of f to the identity.

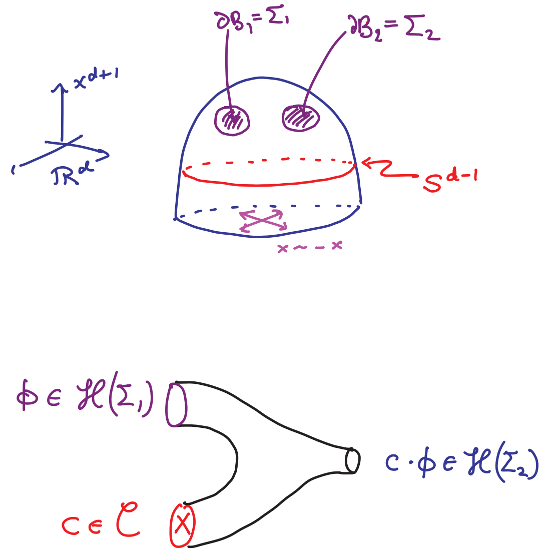


Figure 21: The top figure describes a bordism from $\Sigma_1 \cong S^{d-1}$ to $\Sigma_2 \cong S^{d-1}$ obtained by cutting out two balls from $\mathbb{RP}^d$. We represent $\mathbb{RP}^d$ as the unit sphere in $\mathbb{R}^{d+1}$ modulo $x \to -x$ and take a fundamental domain with $x^{d+1} \geq 0$. The boundary at $x^{d+1} = 0$ still has an identification and is a copy of $\mathbb{RP}^{d-1}$. A neighborhood of this copy of $\mathbb{RP}^{d-1}$ has a boundary indicated by the red circle. That boundary is a copy of S^{d-1}. (Thus, for example, in $d = 2$, the neighborhood of an $S^1 \subset S^2$ is a strip whose boundary consists of <u>two</u> circles, but for this copy of $S^1 = \mathbb{RP}^1$ in $\mathbb{RP}^2$ the neighborhood is a copy of the Mobius strip, and the boundary of that neighborhood is a <u>single</u> circle.) The S^{d-1} on the boundary of the neighborhood of $\mathbb{RP}^{d-1}$ can be thought of as carrying the crosscap state. Therefore if the input state at Σ_1 is ϕ then the output state at Σ_2 is $c\phi$.

For the third property recall that the Klein bottle is equivalent to two connected sums of $\mathbb{RP}^2$. Therefore, we if cut out a disk we get c^2. On the other hand, from the presentation of the Klein bottle as a circle bundle over the circle we see that cutting out a disk must produce the twisted handle operator (see the next subsection for the description of the handle operator):

$$\tilde{H} = \sum_\mu \Omega(\phi^\mu)\phi_\mu = c^2 \tag{8.20}$$

Note that, if we combine (8.19) with (8.20) we find that if H is the handle-adding state

$H = m(\Delta(1))$ then

$$Hc = c^3 \tag{8.21}$$

This reflects the famous fact that in the classification of Riemann surfaces, on an unorientable surface one can always turn a handle into two extra cross-caps.

Turaev and Turner prove an analog of the sewing theorem for the unoriented case with $d = 2$: The general unoriented $d = 2$ TFT is given by a commutative Frobenius algebra with the above extra structure.

9. Computing Amplitudes

One of the pleasant properties in 2d TFT is that one can immediately write down all the "amplitudes" in the theory.

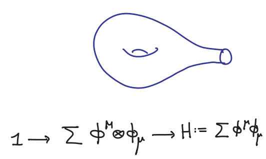


Figure 22: The characteristic element.

The key is to introduce the "characteristic element" H. This is defined by **??** which corresponds to the map

$$1 \to \sum (-1)^{\phi_\mu} \phi_\mu \otimes \phi^\mu \to \sum \phi^\mu \phi_\mu \tag{9.1}$$

and thus it is given by

$$H = \sum_\mu \phi^\mu \cdot \phi_\mu \tag{9.2}$$

where ϕ_μ is any basis for $\mathcal{C}$ and ϕ^μ is a dual basis wrt the trace $\theta(\phi_\mu \phi^\nu) = \delta_\mu^\nu$. (with these choices the formulae hold in the $\mathbb{Z}_2$-graded case).

Exercise

a.) Show that H is independent of the choice of basis.

b.) Compute $\theta(H)$
c.) Show that in the $\mathbb{Z}_2$-graded case $\theta(H) = STr(1) = S\dim(\mathcal{C})$

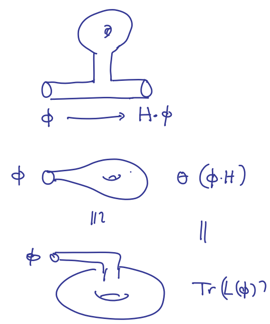


Figure 23: H is a handle-adding operator.

In the left (=right) regular representation, H is a "handle-adding operator" because of Figure 23.

Note that we have

$$\theta(\phi H) = \mathrm{Tr}_{\mathcal{C}}\Big(L(\phi)\Big) \tag{9.3}$$

where the trace is in the regular representation.

One can now immediately write down all amplitudes in the theory. In particular, for a genus g surface with have

$$Z(M_g) = \theta(H^g) = \mathrm{Tr}_{\mathcal{C}}(L(H))^{g-1} \tag{9.4}$$

Exercise *The general amplitude*

Write similar formula for "matrix elements" with arbitrary ingoing and outgoing states, that is, write the amplitude for $\phi_1 \otimes \cdots \phi_n$ to evolve into $\phi'_1 \otimes \cdots \phi'_m$. [7]

9.1 Summing over topologies

In quantum gravity one must face the issue of whether in the path integral over metrics ("geometries") one should also sum over topologies, and how to weight them.

[7] *Answer*: The image of $\phi_1 \otimes \cdots \otimes \phi_n$ for n ingoing circles on a genus g surface is $\sum_{\nu_1,\dots,\nu_m} \theta(\phi_1 \cdots \phi_n H^g \phi^{\nu_1} \cdots \phi^{\nu_m}) \phi_{\nu_1} \otimes \cdots \phi_{\nu_m}$ if there are m outgoing circles.

In string theory, we have a theory of quantum gravity in two dimensions coupled to a sigma model embedding the string into spacetime. In this case one is definitely obliged to sum over topologies. This sum is the analog of the Feynman diagram expansion of field theory in the spacetime.

With this motivation, let us ask what the sum over topologies might look like in our baby theory. We want to understand

$$Z_{\text{string}} = Z(M_0) + Z(M_1) + \cdots \tag{9.5}$$

Evidently, from our general formula, we can write the sum formally as

$$Z_{\text{string}} = \theta\left(\frac{1}{1-H}\right) \tag{9.6}$$

Suppose we have a Frobenius algebra (A,θ). Let us now define a new Frobenius algebra by $\tilde{\theta} := \lambda^{-2}\theta$. Then it is easy to check that

$$\tilde{H} = \lambda^2 H \tag{9.7}$$

and hence the new genus g amplitude is

$$\tilde{Z}(M_g) = \lambda^{-2+2g} Z(M_g) = \lambda^{-\chi(M_g)} Z(M_g) \tag{9.8}$$

Thus, we can interpret the scale of θ as the string coupling constant. If we define a theory by $\hat{\theta}(1) = 1$ then we have

$$Z_{\text{string}} = g^{-2}\hat{\theta}\left(\frac{1}{1-g^2\hat{H}}\right) \tag{9.9}$$

where $\theta(1) = g^{-2}$ is the string coupling constant.

It is unusual that perturbation theory is convergent – usually perturbation series are only asymptotic series. Since the series is convergent we can analytically continue to all values of g^2, thus defining the nonperturbative sum.

9.2 Semisimple algebras

In general this is all there is to it. However, one can say a little more when the algebra $(\mathcal{C},\theta)$ is *semisimple.*

The most useful characterization of semisimplicity is the following. The structure constants [8]

$$\phi_\mu \phi_\nu = N_{\mu\nu}^{\ \ \lambda} \phi_\lambda \tag{9.10}$$

defines a set of matrices via the left-regular representation, $L(\phi_\mu)$, with matrix elements $N_{\mu\nu}^{\ \ \lambda}$. Since $\mathcal{C}$ is commutative these are commuting matrices. Then:

Definition : $\mathcal{C}$ is semi-simple iff the matrices $L(\phi_\mu)$ are simultaneously diagonalizable.

[8] The structure constants $N_{\mu\nu}^{\ \ \lambda}$ need not be integral. But in many interesting examples there is a basis for the algebra in which they are in fact integral.

Thus in the semisimple case we can find a matrix S such that:

$$N_{\mu\nu}{}^{\lambda} = \sum_x S_{\nu}{}^{x} \Lambda_x^{(\mu)} (S^{-1})_x{}^{\lambda} \tag{9.11}$$

where $\Lambda_x^{(\mu)}$ are the different eigenvalues of $L(\phi_\mu)$.

Now choose a basis such with the index μ running over values $\mu = 0, \dots, n$, and take $\phi_0 = 1_{\mathcal{C}}$, the multiplicative identity. Putting $\mu = 0$ in equation (9.11) leads to a trivial identity, but putting $\nu = 0$ and using $N_{\mu\nu}^{\lambda} = N_{\nu\mu}^{\lambda}$ so that $N_{\mu 0}^{\lambda} = \delta_{\mu}{}^{\lambda}$ we see that $S_0{}^x \Lambda_x^{(\mu)} = S_{\mu}{}^x$ since this is the matrix inverse of S^{-1}. So:

$$S_0{}^x \Lambda_x^{(\mu)} = S_{\mu}{}^{x} \qquad \text{no sum on x} \tag{9.12}$$

Plugging back into (9.11) we get:

$$N_{\mu\nu}{}^{\lambda} = \sum_x \frac{S_{\nu}{}^x S_{\mu}{}^x (S^{-1})_x{}^{\lambda}}{S_0{}^x} \tag{9.13}$$

Note that $\theta(\phi_\mu \phi_\nu \phi_\lambda) = N_{\mu\nu}^{\lambda'} Q_{\lambda'\lambda} := N_{\mu\nu\lambda}$ is totally symmetric on μ, ν, λ. Suppose we further restrict attention to a basis $\{\phi_\mu\}$ so that $\theta(\phi_\mu) = \delta_{\mu,0}$. Then taking the trace of (9.10) we learn that $Q_{\mu\nu} = N_{\mu\nu}^{0}$ and then (9.13) gives

$$Q_{\mu\nu} = \sum_x S_{\nu}{}^x S_{\mu}{}^x \frac{(S^{-1})_x{}^0}{S_0{}^x} \tag{9.14}$$

so that

$$N_{\mu\nu\lambda} = \sum_x S_{\mu}{}^x S_{\nu}{}^x S_{\lambda}{}^x \frac{(S^{-1})_x{}^0}{(S_0{}^x)^2} \tag{9.15}$$

If we form the linear combinations

$$\epsilon_x = \sum_{\mu} S_0{}^x (S^{-1})_x{}^{\mu} \phi_\mu \tag{9.16}$$

then the ϵ_x serve as a set of basic idempotents, that is,

$$\begin{aligned} \mathcal{C} &= \oplus_x \mathbb{C} \epsilon_x \\ \epsilon_x \epsilon_y &= \delta_{x,y} \epsilon_y \end{aligned} \tag{9.17}$$

Moreover, if we choose the natural normalization $\theta(\phi_\mu) = \delta_{\mu,0}$ then

$$\theta_x := \theta_{\mathcal{C}}(\epsilon_x) = S_0{}^x (S^{-1})_x{}^0 \tag{9.18}$$

Here θ_i are some nonzero complex numbers. The unordered set $\{\theta_x\}$ is the only invariant of a finite dimensional commutative semisimple Frobenius algebra.

Note that in this case the characteristic element is simply

$$H = \sum_x \frac{1}{\theta_x} \epsilon_x \tag{9.19}$$

and hence the vacuum amplitude on a genus $g \geq 0$ surface is

$$Z_g = \sum_x \theta_x^{1-g} \tag{9.20}$$

Remarks:

1. If we consider the sum over all genera then the sum only converges when $|\theta_x| > 1$ (with conditional convergence on the unit circle), in which case it is:

$$Z_{\text{string}} = \sum_x \frac{\theta_x}{1-\theta_x^{-1}} \tag{9.21}$$

2. Note that nothing has fixed the overall normalization of the matrix S at this point. In some cases S will be unitary so that $(S^{-1})_x{}^0 = (S_0{}^x)^*$. Moreover, if the matrix elements $S_0{}^x$ can be taken to be real then we have a nice simplification of (9.15):

$$N_{\mu\nu\lambda} = \sum_x \frac{S_\mu{}^x S_\nu{}^x S_\lambda{}^x}{S_0{}^x} \tag{9.22}$$

This is how the Verlinde formula is usually stated.

Exercise

Show that the eigenvalues $\Lambda_x^{(\rho)}$ satisfy the algebra

$$\Lambda_x^{(\mu)} \Lambda_x^{(\nu)} = \sum_\lambda N_{\mu\nu}^\lambda \Lambda_x^{(\lambda)} \tag{9.23}$$

Exercise

A natural question in field theory is whether the vacuum amplitudes of a theory completely determine all the amplitudes in the theory.

Investigate this for the case of a semisimple 2d TFT.

---

10. Some examples of commutative Frobenius algebras arising in physical problems

10.1 Example 1: Finite Group Theory

Let G be a finite group. The space of complex-valued functions $C[G]$ is a C^* algebra (see below) with the obvious product given by pointwise multiplication

$$f_1 \cdot f_2(g) := f_1(g) f_2(g) \tag{10.1}$$

Let $\mathcal{C}$ be the subspace of class functions, that is, functions such that

$$f(hgh^{-1}) = f(g) \qquad \forall g, h \in G \tag{10.2}$$

This is the space of functions on the the (finite) set of conjugacy classes of G.

There are two natural bases of functions for $\mathcal{C}$. One makes it clear that $\mathcal{C}$ is a Frobenius algebra in a natural way, and the other makes it clear that this Frobenius algebra is semisimple.

The first natural basis for the space of class functions is given by the characters of the distinct irreps χ_μ, μ labels the distinct irreps of G.

Under the pointwise product

$$\chi_\mu \chi_\nu = \sum_\lambda N_{\mu\nu}^\lambda \chi_\lambda \tag{10.3}$$

where $N_{\mu\nu}^\lambda$ are the *fusion coefficients*, (they are also known as “Littlewood-Richardson coefficients”). They are determined by the Clebsch-Gordon series

$$T^\mu \otimes T^\nu = \oplus_\lambda N_{\mu\nu}^\lambda T^\lambda \tag{10.4}$$

and are nonnegative integers. The natural trace is

$$\theta(\chi_\mu) = \delta_{\mu,0} \tag{10.5}$$

where $\chi_0 = 1$ corresponds to the identity representation. Since for every rep μ there is a rep μ^* with $\chi_\mu \chi_{\mu^*} = \chi_0 + \cdots$, we conclude $\langle f, g\rangle = \theta(fg)$ is nondegenerate, and hence that $\mathcal{C}$ is indeed a Frobenius algebra.

Another natural basis of class functions are the delta functions on conjugacy classes:

$$\begin{aligned} \delta_C(g) &= 1 \qquad g \in C \\ &= 0 \qquad g \notin C \end{aligned} \tag{10.6}$$

where C is a conjugacy class. Note that in this basis the pointwise product is diagonal. Thus it is clear that $\mathcal{C}$ is semi-simple.

We can of course expand one basis in terms of another:

$$\chi_\mu = \sum_i \chi_\mu(C_i) \delta_{C_i} \tag{10.7}$$

Now recall a standard result from group representation theory: the orthogonality relations on the characters of the irreducible representations:

$$\frac{1}{|G|}\sum_g \chi_\mu(g)\chi_\nu(g^{-1}) = \delta_{\mu,\nu} \tag{10.8}$$

Since G is finite we can, WLOG, assume the representation T^μ is unitary. There fore the matrix

$$S_{i\mu} = \sqrt{\frac{m_i}{|G|}}\chi_\mu(C_i) \tag{10.9}$$

where m_i is the order of the class $|C_i|$, is a unitary matrix.

Now we have:

$$\chi_\mu = \sum_i \sqrt{\frac{|G|}{m_i}} S_{i\mu}\delta_{C_i} \tag{10.10}$$

and therefore since multiplication is diagonal in the basis δ_{C_i}, $S_{i\mu}$ is the matrix which diagonalizes the fusion rules in the character basis.

Now, using (9.18) we compute

$$\theta_x = |S_{0x}|^2 = \frac{(\mathrm{dim}V_x)^2}{|G|} \tag{10.11}$$

and hence the partition function on a compact Riemann surface of genus g is

$$Z_g = |G|^{1-g}\sum_x \frac{1}{(\mathrm{dim}V_x)^{2g-2}} \tag{10.12}$$

where the sum runs over irreducible representations of G. The first factor is relatively uninteresting (it can be absorbed in the scale of the string coupling) but the second is interesting.

What geometrical object is the sum in (10.12) *counting?*

We will answer this question in a few lectures.

Exercise

a.) Show that the center of the group algebra $C[G]$ with the convolution product is $\mathcal{C}$, the space of class functions.

b.) Show that the matrix $S_{i\mu}$ is a kind of Fourier transform between these two product structures on $\mathcal{C}$. Note that the basis of characters of irreps χ_μ diagonalize the convolution product:

$$\chi_\mu * \chi_\nu = \frac{\delta_{\mu\nu}}{n_\nu}\chi_\nu \tag{10.13}$$

c.) Show that the invariants θ_x of this Frobenius algebra are given by

$$\theta(\epsilon_\mu) = \frac{(\mathrm{dim}V_\mu)^2}{|G|} \tag{10.14}$$

10.2 Example 2: Loop Groups And The Fusion Ring Of A Rational Conformal Field Theory

What happens if we replace the group G of the previous example with a general compact Lie group?

If G is not finite, but is compact there will still be a complete set of finite-dimensional unitary representations (Peter-Weyl theorem). But now there will be infinitely many representations, hence the space of class functions is infinite-dimensional, and there will be continuous families of conjugacy classes.

For example, for $SU(2) \cong S^3$ the conjugacy classes are two-dimensional spheres inside the group. If we parametrize group elements $g \in SU(2)$ as

$$g = \cos\chi 1 + \mathrm{i}\sin\chi \hat{n}\cdot\vec{\sigma} \tag{10.15}$$

where $0 \le \chi \le \pi$ and $\hat{n} \in S^2$ is in the unit two-sphere then the distinct conjugacy classes are labeled by the continuous parameter χ. For $0 < \chi < \pi$ these conjugacy classes are spheres.

Remarkably, then, if we consider instead the *loop group*: [9]

$$LG := \mathrm{Map}(S^1, G) \tag{10.16}$$

then a certain class of representations of a central extension of LG behaves much more like the finite-dimensional case, at least when G itself is a finite-dimensional compact group.

We now describe this central extension in a bit more detail.

10.2.1 Central Extension Of Loop Algebras

Suppose $\mathfrak{g}$ is a finite dimensional Lie algebra. We can associate to it an infinite dimensional Lie algebra whose elements are maps

$$f : S^1 \to \mathfrak{g} \tag{10.17}$$

The set of all such maps is itself a Lie algebra for we can define

$$[f_1, f_2](z) := [f_1(z), f_2(z)] \tag{10.18}$$

This infinite dimensional Lie algebra is known as the loop algebra $L\mathfrak{g}$.

Loop algebras admit a very interesting central extension. At this point it is useful to take $\mathfrak{g}$ to be a simple Lie algebra. In this case it is an easy result that, up to isomorphism, there is the unique nontrivial central extension:

$$0 \to \mathbb{C} \to \widetilde{L\mathfrak{g}} \to L\mathfrak{g} \to 0 \tag{10.19}$$

Elements of $\widetilde{L\mathfrak{g}}$ are pairs (f, ξ) where f is a map, and $\xi \in \mathbb{C}$. The bracket of the central extension is

$$[(f_1, \xi_1), (f_2, \xi_2)] := ([f_1, f_2], \omega(f_1, f_2)) \tag{10.20}$$

[9] We will be vague about the precise class of maps. It should be some completion of the group with matrix elements taking values in Laurent polynomial in $z = e^{i\theta}$.

where the two-cocycle ω is defined by

$$\omega(f_1, f_2) := \frac{1}{2\pi i} \oint \mathrm{Tr}\left(f_1'(z) f_2(z)\right) dz \tag{10.21}$$

where in the integral we use an ad-invariant bilinear form $(X, Y) := \mathrm{Tr}(XY)$. Recall that for $\mathfrak{g}$ simple all such forms are the same up to scale, consistent with the uniqueness of the central extension.

It is useful to write this out in terms of a basis. Using the bilinear form on $\mathfrak{g}$ we define

$$\mathrm{Tr}(T^a T^b) = g^{ab} \tag{10.22}$$

and we define $K := (0, 1) \in \widetilde{L\mathfrak{g}}$. Then define the loop $T^a_n := z^n T^a$, with $n \in \mathbb{Z}$ and compute:

$$\begin{aligned} [T^a_n, T^b_m] &= f^{ab}_c T^c_{n+m} + n g^{ab} \delta_{n+m,0} K \\ [K, T^a_n] &= 0 \end{aligned} \tag{10.23}$$

This is the way the algebras are often written in the physics literature. (Note that in a representation, if K is represented by a single number k then its value can be absorbed into the normalization of the Killing form.)

Due to the central element K the natural Ad-invariant form of $\widetilde{L\mathfrak{g}}$:

$$(f_1, f_2) = \frac{1}{2\pi \mathrm{i}} \oint \mathrm{Tr}(f_1(z) f_2(z)) \frac{dz}{z} \tag{10.24}$$

is degenerate. It turns out to be very useful to extend the affine Lie algebra by adding one more generator, L_0, so that K is still central but L_0 has commutation relations

$$[L_0, f(z)] = z \frac{d}{dz} f(z) \tag{10.25}$$

One can check that the Jacobi relations still hold. The resulting Lie algebra $\widehat{L\mathfrak{g}} := \widetilde{L\mathfrak{g}} \oplus \mathbb{C} L_0$ is known as a *Kac-Moody algebra*.

The Cartan subalgebra of $\widehat{L\mathfrak{g}}$ is now

$$\mathbb{C} L_0 \oplus \mathbb{C} K \oplus \mathfrak{t} \tag{10.26}$$

where $\mathfrak{t}$ is spanned by the constant maps into the Cartan subalgebra of $\mathfrak{g}$. If one defines the bilinear form on the Lie algebra so that on the subalgebra spanned by L_0, K it is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{10.27}$$

then there is a nondegenerate Ad-invariant form on the $\widehat{L\mathfrak{g}}$. (The crucial constraint that leads to this definition is obtained by taking the inner product of L_0 with the equations (10.23).)

Remark: In fact, we can extend the algebra to a semidirect product with the whole Virasoro algebra.

Of course the loop algebra $L\mathfrak{g}$ is the Lie algebra of LG, so now we can ask when it exponentiates to give a well defined infinite-dimensional group.

10.2.2 Central Extensions Of Loop Groups

Under some circumstances, the central extension $\widetilde{L\mathfrak{g}}$ of the loop algebra can be exponentiated to a central extension $\widetilde{LG}$ of the loop group.

The precise theorem is: [10]

Theorem If G is compact, simple, and simply connected then the Lie algebra extension

$$0 \to \mathbb{R} \to \widetilde{L\mathfrak{g}} \to L\mathfrak{g} \to 0, \tag{10.28}$$

where $\mathfrak{g}$ is the real Lie algebra of G, defined by the cocycle ω corresponds to a group extension

$$1 \to U(1) \to \widetilde{LG} \to LG \to 1 \tag{10.29}$$

iff the differential form $\omega/(2\pi)$ represents an integral cohomology class on LG.

Writing the cocycle explicitly for the central extension of the group is not trivial. In this section we will show a clever construction by which the Wess-Zumino term is used to construct the central extension of LG.

10.2.3 The Wess-Zumino Term

Consider a sigma model of maps $g : \mathcal{S}_d \to G$ where G is a Lie group and $\mathcal{S}_d$ is a d-dimensional (pseudo-) Riemannian spacetime. The standard sigma model action for this theory is

$$\frac{f^2}{4} \int_{\mathcal{S}_d} \mathrm{Tr}(g^{-1}dg) \wedge *(g^{-1}dg) \tag{10.30}$$

where f is a coupling constant and $*$ is the Hodge star operator.

Consideration of anomalies in gauge theories led Wess and Zumino to introduce a very interesting term in the sigma model action [50] in the case of the four-dimensional sigma model. Its proper conceptual formulation and physical consequences were subsequently beautifully clarified in a series of papers by Witten [51, 52]. We will write it here for arbitrary even spacetime dimension $d = 2n$.

Let $\Theta = g^{-1}dg$ be the Maurer-Cartan form on G. Then $\mathrm{Tr}\Theta^{2n+1}$ is closed. If the rank of G is suitably larger than n (our main application is $n = 1$ and this will always be true) then it represents a nonzero cohomology class and for suitable normalization c_n

$$x_{2n+1} = [c_n \mathrm{Tr}\Theta^{2n+1}] \tag{10.31}$$

is a DeRham cohomology class that generates the integral lattice in $H^{2n+1}_{DR}(G)$.

Let $g : \mathcal{S}_{2n} \to G$ be a sigma-model field, and let us consider a *closed* spacetime so that $\partial\mathcal{S}_{2n} = 0$. Physically this is relevant even for fields on $\mathbb{R}^{2n}$ if we require that the fields

[10] For the proof of this theorem see the classic text by A. Pressley and G. Segal, *Loop Groups*, Oxford Mathematical Monographs. See Theorem 4.4.1.

approach 1 at spatial and temporal infinity. In that case, we can consider the the field to be defined on S^{2n}. [11]

There are several slightly different approaches one can take to define the Wess-Zumino term. One way to do it is to note that the image

$$g(\mathcal{S}_{2n}) \subset G \tag{10.32}$$

is a $(2n)$-cycle inside G which varies continuously with G. Now, if $H_{2n}(G;\mathbb{Z}) = 0$ (as is often the case[12]) we can fill in the image of spacetime with an oriented chain $\mathcal{B}_{2n+1}(g)$:

$$\partial\mathcal{B}_{2n+1}(g) = g(\mathcal{S}_{2n}). \tag{10.33}$$

This chain also varies continuously with g. Note however that there can be different choices of the chain $\mathcal{B}_{2n+1}(g)$.

Example $\mathcal{S} = S^2$, $G = SU(2) \cong S^3$, the map g takes $\mathcal{S}$ to the equator. Then we can use the upper hemisphere D_+.

We define the Wess-Zumino term to be:

$$WZ(g) := 2\pi k \int_{\mathcal{B}_{2n+1}(g)} \omega_{2n+1} \tag{10.34}$$

where k is a real, coupling constant with dimensions of $\hbar$. The WZW (Wess-Zumino-Witten) theory is the nonlinear sigma model with Minkowski-space action

$$\frac{f^2}{4}\int_{\mathcal{S}_d} \mathrm{Tr}(g^{-1}dg)\wedge *(g^{-1}dg) + WZ(g) \tag{10.35}$$

Now, at first the definition (10.34) seems absurd. There are two immediate problems:

• It appears to be an action for field configurations in $2n+1$ dimensions.

• It appears to depend on the choice of bounding chain $\mathcal{B}_{2n+1}$, and the constraint (10.33) leaves infinitely many choices for $\mathcal{B}_{2n+1}$.

Let us first address point (1.) Although the definition of the WZ term uses a $2n+1$ dimensional field configuration, the *variation* of the action only depends on the fields on the $2n$ dimensional boundary ∂B_{2n+1}, and hence, the action is in fact local! See the exercise below for some details on how to vary the action and derive the equations of motion.

Therefore, we find for the variation of the WZ term:

$$\delta WZ(g) = 2\pi k c_n(2n+1)\int_{\mathcal{S}_{2n}} \mathrm{Tr}(g^{-1}\delta g)(g^{-1}dg)^{2n} \tag{10.36}$$

Remarkably, even though the definition of the WZ action involves an extension into one higher dimension, this is a local action in the sense that its variation under local changes in

[11] The generalization to the case when spacetime has a boundary is very interesting. In that case $\exp[\mathrm{i}WZ]$ should be regarded as a section of a line bundle.

[12] For example $H_2(G;\mathbb{Z}) = 0$ always for a compact simple simply connected Lie group. $\pi_4(G) = 0$ for all compact simple simply connected groups except $\pi_4(USp(2n)) = \mathbb{Z}_2$. LIST $H_4(G;\mathbb{Z})$

the field $g(x)$ is a local density on spacetime! It's value might depend on subtle topological questions, but the variation is local.

Therefore, the equations of motion of the WZW theory are local partial differential equations:

$$-\frac{f^2}{2}d\big(*g^{-1}dg\big) + 2\pi k c_n(2n+1)(g^{-1}dg)^{2n} = 0 \tag{10.37}$$

Figure 24: Two slightly different $(2n+1)$-chains $\mathcal{B}$ and $\mathcal{B}'$ in G bounding the same $2n$-cycle $g(\mathcal{S}_{2n})$.

Now let us address the second point - the dependence on the choice of bounding chain $\mathcal{B}_{2n+1}(g)$. For a fixed g we can of course smoothly deform the chain to get a second chain as in 24. The difference $\mathcal{B}' - \mathcal{B}$ is a small closed $2n+1$ cycle in G which is, moreover, homologous to zero, so $\mathcal{B}' - \mathcal{B} = \partial\mathcal{Z}$ where $\mathcal{Z}$ is a $(2n+2)$-chain. But now, by Stokes' theorem:

$$\int_{\mathcal{B}'} \omega_{2n+1} - \int_{\mathcal{B}} \omega_{2n+1} = \int_{\mathcal{Z}} d\omega_{2n+1} = 0 \tag{10.38}$$

Figure 25: Two different $(2n+1)$-chains in G bounding the same $2n$-cycle $g(\mathcal{S}_{2n})$.

Thus $WZ(g)$ does not change under small deformations of $\mathcal{B}$.

However, it can happen that $\mathcal{B}'$ and $\mathcal{B}$ are not small deformations of each other as in 25. In general if $\mathcal{B}, \mathcal{B}'$ are two oriented chains with

$$\partial\mathcal{B} = \Sigma_{2n} \tag{10.39}$$

and

$$\partial\mathcal{B}' = \Sigma_{2n} \tag{10.40}$$

then

$$\mathcal{B} \cup -\mathcal{B}' = \Xi_{2n+1} \tag{10.41}$$

is a closed oriented $(2n+1)$-cycle. Therefore,

$$\int_{\mathcal{B}(g)} \omega_{2n+1} = \int_{\mathcal{B}'(g)} \omega_{2n+1} + \int_{\Xi_{2n+1}} \omega_{2n+1} \tag{10.42}$$

and hence, if the periods $\int_{\Xi_{2n+1}} \omega_{2n+1}$ are nonzero then *the expression* $WZ(g)$ *is not well-defined as a real number!*

This might seem disturbing, but, the cycle Ξ_{2n+1} defines an integral homology class, and hence the periods of ω_{2n+1} are quantized. Therefore, the ambiguity in the definition of $WZ(g)$ is an additive *quantized* shift of the form $2\pi kN$ where N is an integer. Put differently,

$$WZ(g)\mathrm{mod} 2\pi k\mathbb{Z} \tag{10.43}$$

is well-defined. The quantized ambiguity cannot vary under small variations of g. Thus, $WZ(g)$ is still a local action, and the equations of motion are still local.

Note that the situation here is very similar to our discussion of the action for general quantization of a symplectic manifold when the symplectic form has nontrivial periods.

The situation in quantum mechanics is a little more subtle, since in quantum mechanics one works directly with the action, and not just the equations of motion. However, in quantum mechanics the action only enters through $\exp[\frac{i}{\hbar}S]$, and therefore all that must really be well-defined is the expression

$$\exp[\frac{i}{\hbar}WZ(g)] \tag{10.44}$$

What is the ambiguity in (10.44) ? We see that it is just

$$\exp[2\pi i\frac{k}{\hbar}\int_{\Xi_{2n+1}} \omega_{2n+1}] \tag{10.45}$$

Therefore, if $k = \kappa\hbar$, where κ is an integer, then the

$$\exp[\frac{i}{\hbar}WZ(g)] \tag{10.46}$$

in the path integral is a well-defined $U(1)$-valued function on the space of fields $Map[\mathcal{S}_{2n}, G]$. Assuming we have a well-defined measure on the space of fields, there is no harm including this expression in the measure.

Thus, the coupling constant k *must be quantized for a mathematically well-defined measure in the quantum mechanical path integral.* This is one of the most beautiful examples of a topological quantization of a coupling constant.

We will usually set $\hbar = 1$. Thus, large k corresponds to the semiclassical limit.

Remarks:

1. *Integral normalization.* Here are some relevant facts. It can be shown [13] that for all compact, simple, connected, and simply connected groups G:

$$x_3 = \left[\frac{1}{48\pi^2 h}\mathrm{Tr}_{adj}(g^{-1}dg)^3\right] \tag{10.47}$$

generates the integral cohomology lattice in $H^3_{DR}(G)$, where h is the dual Coxeter number. In particular, for $SU(2)$ we can take

$$x_3 = \left[\frac{1}{24\pi^2}\mathrm{Tr}_2(g^{-1}dg)^3\right] \tag{10.48}$$

It follows that for $SU(N)$ we can take

$$x_3 = \left[\frac{1}{24\pi^2}\mathrm{Tr}_N(g^{-1}dg)^3\right] \tag{10.49}$$

to generate the integral cohomoloy lattice in $H^3_{DR}(G)$.

2. Here is another way to define the Wess-Zumino term. For each connected component $\mathcal{C}_\alpha$ of the fieldspace $Map(\mathcal{S}_{2n}, G) = \amalg_\alpha \mathcal{C}_\alpha$ we choose a "basepoint" field configuration $g_0^{(\alpha)} : \mathcal{S}_{2n} \to G$. If $\mathcal{S}_{2n}$ is contractible there is only one component and we can choose g_0 to be the constant map (say with image $1 \in G$). In general for field configurations $g \in \mathcal{C}_\alpha$ we choose a smooth homotopy $g(x,s)$, $0 \le s \le 1$ from $g_0^{(\alpha)}(x)$ at $s=0$ to $g(x)$ at $s=1$. Now we view the interpolation as a field in $2n+1$ dimensions, that is, as a map of the cylinder $\widehat{g} : I \times \mathcal{S}_{2n} \to G$. We can then define

$$WZ(g; g_0) := 2\pi k \int_{I \times \mathcal{S}_{2n}} \widehat{g}^*(\omega_{2n+1}) \tag{10.50}$$

3. The value of $WZ(g; g_0)$ depends on the choice of g_0 and on the interpolation, but only "locally," in the following sense: Suppose we have a continuous family of maps $\tilde{g}^\tau : \mathcal{S}_{2n} \to G$ in the connected component $\mathcal{C}_\alpha$. Then we find a continuous family of extensions $\tilde{g}^\tau_s : B_{2n+1} \to G$ such that $g_0^\tau(x) = g_0(x)$ for all τ. Then, letting $B = I \times \mathcal{S}_{2n}$ we have:

$$\boxed{\frac{\partial}{\partial\tau}\mathrm{Tr}(\tilde{g}^{-1}d_B\tilde{g})^{2n+1} = d_B\left[(2n+1)\mathrm{Tr}\tilde{g}^{-1}\frac{\partial\tilde{g}}{\partial\tau}(\tilde{g}^{-1}d_B\tilde{g})^{2n}\right]} \tag{10.51}$$

Proof: We know that the Maurer-Cartan form pulled back to $I \times B_{2n+1}$ is closed, so

$$(d_B + \delta)\mathrm{Tr}(\tilde{g}^{-1}(d_B+\delta)\tilde{g})^{2n+1} = 0 \tag{10.52}$$

where $\delta = d\tau\frac{\partial}{\partial\tau}$. Now forms on the product space can be decomposed into type (a,b) with a-forms along I and b-forms along B_{2n+1}. The component of (10.52) of type $(1, 2n+1)$ is

$$\delta\mathrm{Tr}(\tilde{g}^{-1}d_B\tilde{g})^{2n+1} + (2n+1)d_B\left[\mathrm{Tr}(\tilde{g}^{-1}\delta\tilde{g})(\tilde{g}^{-1}d_B\tilde{g})^{2n}\right] = 0 \tag{10.53}$$

[13] Ref: Mimura and Toda, *Topology of Lie Groups*, Translations of Math. Monographs **91**; R. Bott, Bull. Soc. Math. France **84**(1956) 251.

pulling out the $d\tau$ gives our identity. It is now an easy matter to show that the variation of the WZ term defined as in (10.50) only depends on the variation g_s^τ at $s=1$.

Exercise

a.) Calculate $\mathrm{Tr}(g^{-1}dg)^3$ for $SU(2)$ in terms of Euler angles for the group, using the fundamental representation:

$$\mathrm{Tr}_{\mathbf{2}}(g^{-1}dg)^3 = -\frac{3}{2}d\psi \wedge \sin\theta d\theta \wedge d\phi \tag{10.54}$$

b.) Write this differential form as a locally exact form.

c.) Show that

$$\int_{SU(2)} \frac{1}{24\pi^2}\mathrm{Tr}_{\mathbf{2}}(g^{-1}dg)^3 = -1 \tag{10.55}$$

and thus conclude that the form is not globally exact. Compare with the general normalizations above.

d.) Now show that x_3 in equation (10.49) defines a nontrivial cohomology class for all $SU(N)$.

Exercise *The Polyakov-Wiegman formula*

Consider the WZ term in two spacetime dimensions.

a.) Show that

$$\mathrm{Tr}\big((g_1g_2)^{-1}d(g_1g_2)\big)^3 = \mathrm{Tr}\big(g_1^{-1}dg_1\big)^3 + \mathrm{Tr}\big(g_2^{-1}dg_2\big)^3 + 3d\Big[\mathrm{Tr}(dg_2g_2^{-1})(g_1^{-1}dg_1)\Big] \tag{10.56}$$

b.) Conclude that the WZ term satisfies:

$$WZ(g_1g_2) = WZ(g_1) + WZ(g_2) + 6\pi k c_1 \int \mathrm{Tr}(dg_2g_2^{-1})(g_1^{-1}dg_1) \tag{10.57}$$

Exercise *Variation Of The WZ Term*

Using the variational formula

$$\delta\big(g^{-1}\partial_\mu g\big) = \partial_\mu(g^{-1}\delta g) + \big[g^{-1}\partial_\mu g, g^{-1}\delta g\big] \tag{10.58}$$

We compute:

$$\frac{\partial}{\partial s}\mathrm{Tr}(g^{-1}dg)^{2n+1} = (2n+1)\mathrm{Tr}d\left(g^{-1}\frac{\partial g}{\partial s}\right)(g^{-1}dg)^{2n} \tag{10.59}$$

The second term, involving the commutator drops out.

Now we compare with the RHS

$$d\left[\mathrm{Tr}g^{-1}\frac{\partial g}{\partial s}(g^{-1}dg)^{2n}\right] = \mathrm{Tr}d\left(g^{-1}\frac{\partial g}{\partial s}\right)(g^{-1}dg)^{2n} + \mathrm{Tr}g^{-1}\frac{\partial g}{\partial s}\left[d\Theta\Theta^{2n-1} - \Theta d\Theta\Theta^{2n-2} \pm \cdots\right] \tag{10.60}$$

and using the Maurer-Cartan equation we find that the second group of terms cancel in pairs. ♠

10.2.4 Construction Of The Cocycle For LG

The trick is to consider the group of maps DG from the *disk* to the group G, i.e. we introduce $DG = Map(D,G)$ where D is the disk. Note that the subgroup D_1G of maps such that $g|_{\partial D} = 1$ is a normal subgroup and $DG/D_1G \cong LG$, and explicit isomorphism being given by the restriction map.

Now, in contrast to LG, it is easy to write a central extension $\widetilde{DG}$ of the group DG:

$$(g_1,\lambda_1)\cdot(g_2,\lambda_2) = (g_1g_2, \lambda_1\lambda_2 f(g_1,g_2)) \qquad g_i \in DG \tag{10.61}$$

where

$$f(g_1,g_2) = \exp\left[2\pi i(6\pi c_1 k)\int_D \mathrm{Tr}(dg_2 g_2^{-1})(g_1^{-1}dg_1)\right] \tag{10.62}$$

Note that we have written our Ad-invariant inner product $(\cdot,\cdot)_{\mathfrak{g}}$ in terms of a definite trace Tr in some representation. For $SU(N)$ with the trace in the N dimensional representation $c_1 = 1/(24\pi^2)$.

Exercise

a.) Check that (10.62) is indeed a group cocycle.

b.) Compute the corresponding central extension on the Lie algebra $D\mathfrak{g}$ and show that it is trivial when one of the elements vanishes on the boundary. Indeed, show that it is

$$24\pi^2 i c_1 k \oint_{S^1}\mathrm{Tr}\epsilon_1 d\epsilon_2 = ik\oint \mathrm{Tr}\epsilon_1 d\epsilon_2 \tag{10.63}$$

for $c_1 = 1/(24\pi^2)$.

Now, the beautiful observation is that, when g_1 and g_2 are equal to 1 on the boundary ∂D, we can consider them to define maps from $S^2 \to G$, and therefore we can define the WZ term. But, because of the identity we proved above:

$$WZ(g_1 g_2) = WZ(g_1) + WZ(g_2) + 6\pi k c_1 \int \mathrm{Tr}(dg_2 g_2^{-1})(g_1^{-1} dg_1) \tag{10.64}$$

the cocycle becomes a coboundary when restricted to the subgroup $D_1 G$. Therefore, the extension

$$1 \to U(1) \to \widetilde{DG} \to DG \to 1 \tag{10.65}$$

splits over the normal subgroup $D_1 G$, that is:

$$\psi : g \to (g, e^{iWZ(g)}) \qquad g \in D_1 G \tag{10.66}$$

is a group homomorphism from $D_1 G$ to $\widetilde{DG}$, and hence we can take a quotient

$$1 \to U(1) \to \widetilde{DG}/\psi(D_1 G) \to DG/D_1 G = LG \to 1 \tag{10.67}$$

to construct the loop group $\widetilde{LG} := \widetilde{DG}/\psi(D_1 G)$.

Finally, if we include L_0 then note that

$$\exp[\mathrm{i}\theta_0 L_0] g(\theta) \exp[-\mathrm{i}\theta_0 L_0] = g(\theta + \theta_0) \tag{10.68}$$

so L_0 generates rigid rotations of loops.

Remarks

1. The above construction of the central extension is due to J. Mickelsson. For a generalization to $Map(X, G)$ for arbitrary manifolds X see [32] and references therein.

2. At the Lie group level one can construct a semidirect product with the Virasoro group - the centrally extended diffeomorphism group of the circle.

3. The above presentation of the centrally extended loop group is very convenient for quantizing three-dimensional Chern-Simons theory on $D \times \mathbb{R}$. The group of gauge transformations is $D_1 G$. The flat gauge fields on the disk are parametrized by DG.

10.2.5 Integrable Highest Weight Representations

Let G be simple and compact. The centrally-extended loop group constructed above will be denoted simply G_k for $k \in \mathbb{Z}_+$. [14]

What can we say about the representations of G_k? Clearly there are many. For example, G_k has a homomorphic image LG so, choosing any representation $\rho : G \to \mathrm{Aut}(V)$ of G (for example, a finite-dimensional irreducible representation of G) and a point z_0 on the circle we can define the evaluation representation:

$$\widehat{\rho} : G_k \to \rho(g(z_0)) \tag{10.69}$$

[14] In general for a compact group it can be shown that the central charge should be regarded as an element of $H^4(BG; \mathbb{Z})$. For G simple, compact, and connected this cohomology group is isomorphic to $\mathbb{Z}$ and we can consider the central extension to be an integer.

with carrier space V. Note that these representations do not interact well with L_0, since L_0 translates $z_0 \to z_0 e^{i\theta}$.

It turns out that G_k has a finite set of irreducible representations with L_0 bounded below. These representations are naturally constructed as highest weight representations of the Lie algebra and are known as *integrable representations* because they extend from representations of the Lie algebra to the Lie group.

The integrable representations are graded by L_0 and the spectrum is bounded below. The lowest weight space under L_0 is itself an irreducible representation of G, and has a highest weight λ corresponding to an element of

$$\Lambda_{wt}/W \tag{10.70}$$

where W is the Weyl group and Λ_{wt} is the weight lattice. After making a choice of simple roots for $\mathfrak{g}$ the highest weight of an irreducible representation of G can be labeled by a dominant highest weight $\lambda \in \Lambda_{wt}$. Recall this means that

$$\lambda = \sum n_i \lambda^{(i)} \tag{10.71}$$

where $n_i \geq 0$ and $\lambda^{(i)}$ is a basis of fundamental weights dual to the simple roots.

Now for the case of G_k the irreducible representations are labeled by the quotient

$$\Lambda_{wt}/\widehat{W}^{(k)} \tag{10.72}$$

where $\widehat{W}^{(k)}$ is the level k affine Weyl group. It is a discrete crystallographic subgroup of the group of affine transformations of $\mathfrak{t}^\vee$. As a group it is isomorphic to the semidirect product of the Weyl group W with the translation group by the coroot lattice Λ_{crt} but we denote it by $\widehat{W}^{(k)}$ because the translations act by

$$\{\sigma|v\} : \lambda \mapsto \sigma(\lambda) + kv \qquad \sigma \in W, v \in \Lambda_{crt} \tag{10.73}$$

It is useful to know that this is a Coxeter group, generated by reflections. These include the Weyl reflections and the reflection in the hyperplane $(\lambda, \theta) = k$, where $k \in \mathbb{Z}_+$, θ is the highest root and we use a normalization of the Killing form so that $(\theta, \theta) = 2$. A fundamental chamber for this action in Λ_{wt} is the finite set of dominant weights satisfying:

$$(\lambda, \theta) \leq k \tag{10.74}$$

This condition is usually derived in conformal field theory by using unitarity and a null-vector.

Example 1 $G = SU(2)$. Then $\theta = \alpha$ and $\lambda = j\alpha$ where $j \in \frac{1}{2}\mathbb{Z}_+$ is known in physics as the spin. (Mathematicians normally would write $\lambda = n_1 \lambda^{(1)}$ where $\lambda^{(1)} = \frac{1}{2}\alpha_1$ is the fundamental weight. Thus $n_1 = 2j$ is twice the spin.) The irreducible representation has of $SU(2)$ with weight $\lambda = j\alpha_1$ has dimension $2j+1$. The lattice Λ_{wt} is isomorphic to $\mathbb{Z}$. The hyperplane in $\mathfrak{t}^\vee \cong \mathbb{R}$ is $(x\alpha, \alpha) = k$ or $x = k/2$. So reflection in the hyperplane takes $j \to \frac{k}{2} - j$. Therefore a fundamental domain for the affine Weyl group is:

$$0 \leq j \leq \frac{k}{2} \tag{10.75}$$

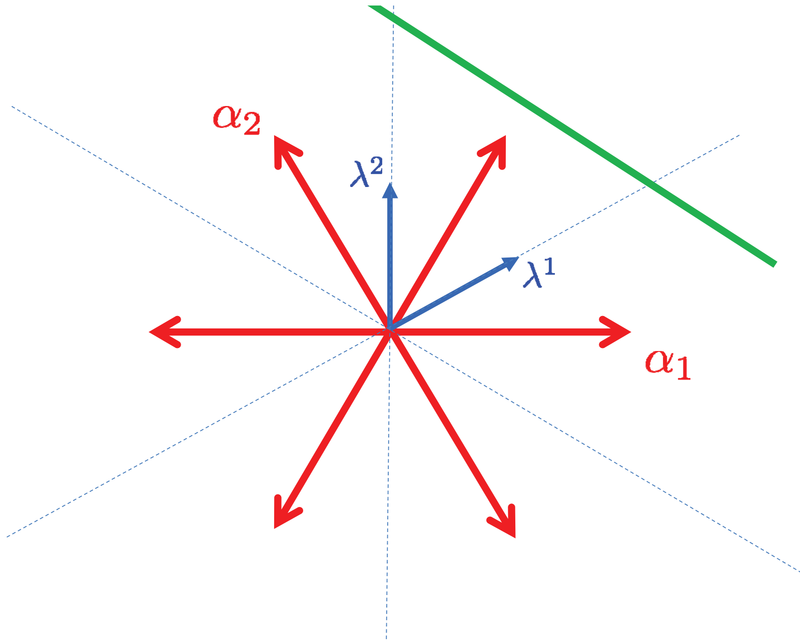


Figure 26: The root and weight lattice for $SU(3)$. A standard set of simple roots α_1, α_2 is shown along with fundamental weights λ^1, λ^2. The fundamental Weyl chamber is the positive cone spanned by these two weights. The highest root is $\theta = \alpha_1 + \alpha_2$. The heavy green line is the line $(\lambda, \theta) = k$ for some positive integer k. The affine Weyl chamber is the region $(\lambda, \theta) \le k$ and the integrable weights at level k is the intersection of the weight lattice with the fundamental Weyl chamber.

Note that $\widehat{W}^k$ in this case is isomorphic to $\mathbb{Z} \rtimes \mathbb{Z}_2$, the infinite dihedral group.

Example 2 $G = SU(3)$. For $SU(3)$ we can choose two simple roots. The standard choice is

$$\begin{aligned} \alpha_1 &= (\sqrt{2}, 0) \\ \alpha_2 &= (-\sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}}) \end{aligned} \tag{10.76}$$

$$\begin{aligned} \lambda^1 &= \frac{2}{3}\alpha_1 + \frac{1}{3}\alpha_2 \\ \lambda^2 &= \frac{1}{3}\alpha_1 + \frac{2}{3}\alpha_2 \end{aligned} \tag{10.77}$$

Now $\theta = \alpha_1 + \alpha_2$. The integrable weights at level k are $n_1\lambda^1 + n_2\lambda^2$ with $n_i \in \mathbb{Z}_+$ and $n_1 + n_2 \le k$.

The integrable highest weight representations $L(\lambda)$ turn out to be objects in a tensor category. The tensor product is <u>not</u> symmetric. Note that it is <u>not</u> obvious how to take a tensor product of two representations $L(\lambda)$ and $L(\lambda')$ to get a representation of the KM algebra with the <u>same</u> value of k. [15] The way to do this is to use conformal field theory.

[15] For the same reason one does not want to multiply characters.

The key observation is that if we consider the two-dimensional WZW model with $f^2 = 12\pi k c_1$ then the equation of motion in Minkowski space is:

$$\partial_+ \left(g^{-1} \partial_- g \right) = 0 \tag{10.78}$$

where $\partial_\pm$ are derivatives wrt light-cone variables $x^\pm = x^0 \pm x^1$. Equivalently, we may write:

$$\partial_- \left(\partial_+ g g^{-1} \right) = 0 \tag{10.79}$$

so the classical theory on $\mathbb{M}^{1,1}$ has a symmetry of $\mathrm{Map}(\mathbb{R}, G) \times \mathrm{Map}(\mathbb{R}, G)$ with a left-action on solutions by

$$(h_L, h_R) : g \mapsto h_L(x^+) g(x^+, x^-) (h_R(x^-))^{-1} \tag{10.80}$$

On a cylinder $S^1 \times \mathbb{R}$ this becomes a product $LG \times LG$. In the quantum theory this symmetry survives when LG is replaced by G_k. The theory splits into a theory of left- and right-movers. It is a conformal field theory. This was shown in detail in [19, 28, 51]. Two textbooks that treat this material in detail are:

1. P. Di Francesco, P. Mathieu, and D. Senechal, *Conformal Field Theory*
2. J. Fuchs, *Affine Lie Algebras And Quantum Groups*

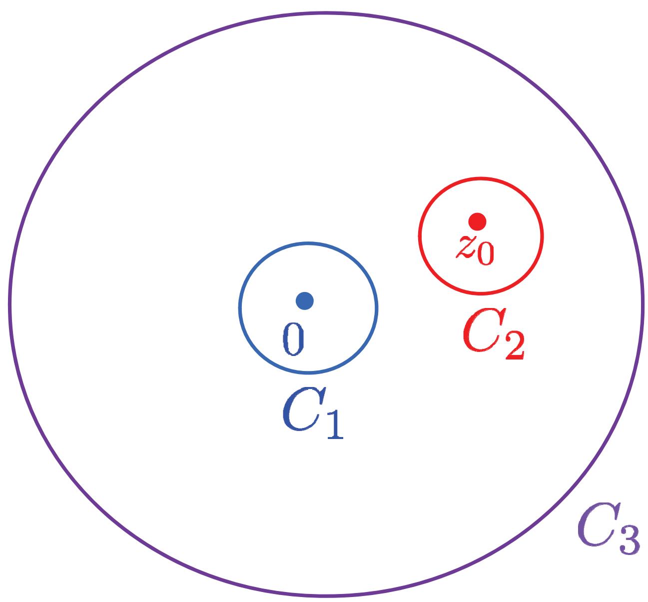


Figure 27: Three CFT state spaces are associated with the circles C_1, C_2, and C_3 and are associated with radial quantization around $z = 0, z_0, 0$, respectively.

The tensor product can be thought of as follows. (We follow the description from [35], equation (2.5). Rigorous descriptions of the tensor product using vertex operator algebra theory are given in [26, 27].)

We can form a current:

$$J^a(z) = \sum T^a_n z^{-n-1} dz \tag{10.81}$$

where we now analytically continue z to the complex plane - regarded as the Euclidean worldsheet of a 2d Euclidean QFT for the WZW model. Note that

$$T^a_n = \oint z^n J^a(z) \tag{10.82}$$

There is a state-operator correspondence: The insertion of a local operator $\Phi(z)$ at a point z on the plane produces a state in the Hilbert space of radial quantization centered on that point.

To give a tensor product we need a comultiplication $\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ where $\mathcal{A}$ is the algebra of local observables.

We imagine one Hilbert space of states on a small circle C_1 centered at $z = 0$, a second circle C_2 centered at $z = z_0$, and a third on a larger circle C_3 centered at $z = 0$ but encircling z_0. See Figure 27.

If local operators creating states in representations L_λ and L_μ are inserted at $z = 0$ and $z = z_0$ then the resulting state on the circle C_3 will have an action of the current with modes

$$\begin{aligned} \Delta_{0,z_0}(T^a_n) &= \oint_{C_3} z^n J^a(z) \\ &= \left(\oint_{C_1} z^n J^a(z)\right) \otimes 1 + 1 \otimes \left(\oint_{C_2} z^n J^a(z)\right) \end{aligned} \tag{10.83}$$

In the first line we have written an operator acting on the space of states on the circle C_3. (Think of it as the outgoing state space in a pair of pants diagram.) The next line is a contour deformation (since $J^a(z)$ is a holomorphic current) to give an action on the space of states on the circles C_1 and C_2. Since there are two ingoing states on the pair of pants we have a tensor product of state spaces. The interesting term is $\oint_{C_2} z^n J^a(z) dz$. When acting on the Hilbert space obtained by radial quantization centered at z_0 we should expand the current as

$$J^a(z) = \sum_{m \in \mathbb{Z}} (z - z_0)^{-m-1} J^a_m(z_0) d(z - z_0) \tag{10.84}$$

but

$$\oint_{C_2} z^n (z - z_0)^{-m-1} dz = \begin{cases} 0 & m \leq -1 \\ \binom{n}{m} z_0^{n-m} & m \geq 0 \end{cases} \tag{10.85}$$

and hence

$$\Delta_{0,z_0}(T^a_n) = T^a_n \otimes 1 + 1 \otimes \left(\sum_{k=0}^{\infty} \binom{n}{k} z_0^{n-k} T^a_k(z_0) \right) \tag{10.86}$$

Now, the fusion rules for multiplication, with this tensor product, of the the simple objects (that is, the irreducible representations of G_k) turn out to define a semisimple

Frobenius algebra: [16]

$$L(\mu) \otimes_{0,z_0} L(\nu) \cong \oplus_\lambda N_{\mu\nu}^{\lambda} L(\lambda). \tag{10.87}$$

Therefore, there is a matrix S that diagonalizes these rules.

The remarkable statement of E. Verlinde is that this matrix S can be taken to be the same matrix as the modular S-matrix for the characters of the representations [49]. To explain Verlinde's statement we define the characters by

$$\chi_\mu(\tau) = \mathrm{Tr}_{L(\mu)} q^{L_0 - c/24} \tag{10.88}$$

where $q = e^{2\pi i\tau}$ is in the unit disk (so τ is in the upper half-plane) and

$$c = \frac{k \mathrm{dim} G}{k+h} \tag{10.89}$$

is the central charge of the Virasoro algebra. Then it turns out that $\chi_\mu(\tau)$ are vector-valued modular functions. In particular

$$\chi_\mu(-1/\tau) = S_{\mu\nu}\chi_\nu(\tau) \tag{10.90}$$

Verlinde's observation was that this modular S-matrix diagonalizes the WZW fusion rules. It was proven in [14, 34, 35]. Kac-Peterson derived a formula for their transformation for $S_{\mu\nu}$.

Example Introduce the level k theta function defined by

$$\Theta_{\mu,k}(z,\tau) := \sum_{n\in\mathbb{Z}} q^{k(n+\mu/(2k))^2} y^{(\mu+2kn)} = \sum_{\ell=\mu \mathrm{mod} 2k} q^{\ell^2/(4k)} y^\ell \tag{10.91}$$

with $y = e^{2\pi i z}$. Now introduce the character:

$$\chi_j^k(z,\tau) := \mathrm{Tr}_{V(j)} q^{L_0 - c/24} e^{2\pi i z (2J_0^3)} \tag{10.92}$$

This might look unnatural from the point of view of Lie algebra theory, but it is well-motivated by physics: We are subtracting the groundstate energy. Then a special case of the Weyl-Kac character formula is:

$$\begin{aligned} \chi_j^k(z,\tau) &:= \mathrm{tr}\, q^{L_0 - c/24} e^{2\pi i z (2J_0^3)} \\ &= \frac{\Theta_{2j+1,k+2}(z,\tau) - \Theta_{-2j-1,k+2}(z,\tau)}{\Theta_{1,2}(z,\tau) - \Theta_{-1,2}(z,\tau)} \\ &= q^{(\ell+1)^2/(4(k+2)) - 1/8} \chi_j(y) + \cdots \end{aligned} \tag{10.93}$$

where $0 \le j \le k/2$ and j is half-integral.

Now the key transformation law of level k theta functions (easily derived using the Poisson summation formula) is

$$\Theta_{\mu,k}(-\omega/\tau, -1/\tau) = (-i\tau)^{1/2} e^{2\pi i k\omega^2/\tau} \sum_{\nu=0}^{2k-1} \frac{1}{\sqrt{2k}} e^{2\pi i \frac{\mu\nu}{2k}} \Theta_{\nu,k}(\omega,\tau) \tag{10.94}$$

[16] The conceptual reason for this is that one can gauge the G symmetry of the WZW model to produce the G/G model. This is a 2d TFT.

From this transformation law one can derive the S-matrix for $SU(2)_k$. It is

$$S_{jj'} = \sqrt{\frac{2}{k+2}} \sin \frac{\pi(2j+1)(2j'+1)}{k+2} \tag{10.95}$$

In general, labeling the irreducible highest weight representations of G_k by the dominant weight of the representation of G at the lowest eigenvalue of L_0 we have the eigenvalues:

$$\Lambda_\nu^{(\mu)} = \frac{S_\mu^{\ \nu}}{S_0^{\ \nu}} = \mathrm{ch}_\mu \left(2\pi \frac{\nu+\rho}{k+h} \right) \tag{10.96}$$

where μ, ν are dominant weights, ρ is the Weyl vector, [17] h is the dual coxeter number, and we have used the Killing form, normalized so that $(\theta, \theta) = 2$ to identify $\mathfrak{t}^\vee \cong \mathfrak{t}$ and thereby regard $\nu + \rho$ as an element of $\mathfrak{t}$.

♣Give the specialization of this formula to $SU(2)_k$. ♣

Using equation (10.96) it is possible to express the CFT fusion rules $N_{\mu\nu}^{\lambda}$ in terms of the Littlewood-Richardson coefficients $\bar{N}_{\mu\nu}^{\lambda}$ of the finite-dimensional group:

$$\mathrm{ch}_\mu \mathrm{ch}_\nu = \sum_{\lambda \in \Lambda_{wt}^+} \bar{N}_{\mu\nu}^{\lambda} \mathrm{ch}_\lambda \tag{10.97}$$

We know that, in general

$$\Lambda_x^{(\mu)} \Lambda_x^{(\nu)} = \sum_\lambda N_{\mu\nu}^{\lambda} \Lambda_x^{(\lambda)} \tag{10.98}$$

for semisimple Frobenius algebras. Evaluating (10.97) on the special conjugacy classes $2\pi(\lambda+\rho)/(k+h)$ and using some simple manipulations [18] one obtains:

$$N_{\mu\nu}^{\lambda} = \sum_{w \in \widehat{W}^k, w\cdot\lambda \in \Lambda_{wt}^+} \mathrm{sign}(w) \bar{N}_{\mu\nu}^{w\cdot\lambda} \tag{10.99}$$

(The sign of w is defined since $\widehat{W}^k$ is a Coxeter group. It is ± 1 according to whether the group element is a product of an even/odd number of reflections.)

For example, for $SU(2)_k$ the ordinary Clebsch-Gordon rules $\bar{N}_{jj'}^{j''}$ give

$$[j] \otimes [j'] = [|j-j'|] \oplus [|j-j'|+1] \oplus \cdots \oplus [j+j'] \tag{10.100}$$

However, if $j + j' > k/2$ then there will be an affine Weyl reflection around $j = k/2$. Each weight larger than $k - j - j'$ will have a reflected image larger than $k/2$ and these will cancel in pairs. In this way we get:

$$N_{jj'}^{j''} = \begin{cases} 1 & |j-j'| \le j'' \le \min\{j+j', k-j-j'\} \& j+j'+j'' \in \mathbb{Z} \\ 0 & \text{else} \end{cases} \tag{10.101}$$

♣Say something about tetrahedra ♣

Remarks

[17] The Weyl vector is half the sum of positive roots. It is equal to the sum of fundamental weights.

[18] See Di Francesco et. al. Section 16.2.1 or Fuchs, Section 5.5

1. In general, the characters are given by the Weyl-Kac character formula. Just as the Weyl character formula can be written

$$\frac{\sum_{w\in W} e^{w(\lambda+\rho)-\rho}}{\sum_{w\in W} e^{w(\rho)-\rho}} \tag{10.102}$$

the Weyl-Kac character formula can be written in the identical form, where we replace the sum over the Weyl group by the sum over the affine Weyl group and λ, ρ are replaced by suitable affine weights. We recognize the structure of the sum over the affine Weyl group in equation (**??**): The theta functions come from the sum over $\mathbb{Z}$ and the difference of the theta functions comes from the nontrivial reflection in the Weyl group of $SU(2)$.

2. It turns out that the representation theory of G_k is closely related to that of the corresponding quantum group when q is a suitable root of unity:

$$q = \exp(\frac{2\pi \mathrm{i}}{k+h}) \tag{10.103}$$

See the book by Fuchs for a detailed exposition.

3. There is a generalization of the above story to a much wider class of two-dimensional conformal field theories known as "rational conformal field theories."

Exercise

a.) Find the invariants θ_x for the Frobenius algebra defined by $N_{jj'}^{j''}$.

b.) Note that since $N_{jj'}{}^{j''}$ are integers, $Z(\Sigma_g)$ is an integer, a surprising fact when viewed as (9.20). This is a special case of the famous Verlinde formula.

Exercise

a.) Show that the product of theta functions of level k and k', as functions of z can be expanded in terms of theta functions of level $k+k'$. Thus, taking a direct sum of the span of the level k theta function defines a graded ring.

b.) Show that the characters $\chi_j^k(z,\tau)$ can be expanded in level k theta functions.

This is another way to see that the standard tensor products of representations of $SU(2)_k$ will not produce a representation of $SU(2)_k$.

10.3 Example 3: The Cohomology Of A Compact Oriented Manifold

The following example shows that

- Some natural Frobenius algebras are not semi-simple.
- The nondegeneracy of θ can amount to a deep theorem.

Let M be a compact oriented manifold. Consider $H^*_{DR}(M)$ (or $H^*(M;\mathbb{Q})$). We claim that this defines a graded commutative Frobenius algebra.

The multiplication is the wedge product. The trace is given by the integral:

$$\omega \to \theta(\omega) = \int_{[M]} \omega \tag{10.104}$$

The (deep) theorem of Poincare duality can be formulated as the statement that the quadratic form defined by θ, namely,

$$(\omega_1, \omega_2) \to \int_M \omega_1 \wedge \omega_2 \tag{10.105}$$

is *nondegenerate*. (It is crucial here that M be *compact* and that M is a *manifold*.) Thus, Poincaré duality says that this algebra is a Frobenius algebra.

Let ω_μ be a basis, and ω^μ a dual basis so that

$$\int \omega_\mu \wedge \omega^\nu = \delta_\mu{}^\nu \tag{10.106}$$

If ω_μ is a k-form, then ω^μ is an $(n-k)$-form.

Let us compute the characteristic element:

$$H = \sum_\mu \omega^\mu \wedge \omega_\mu \tag{10.107}$$

This is an n-form. Since it is a form of positive degree it is not invertible, so this algebra is not semisimple.

Let $\mathrm{vol}\,(M)$ be the integral generator of $H^n(M)$ corresponding to the orientation. Then from (10.106) we get

$$\omega^\mu \wedge \omega_\mu = (-1)^{\deg \omega_\mu \deg \omega^\mu} \mathrm{vol}\,(M) \tag{10.108}$$

so

$$H = \chi(M)\mathrm{vol}\,(M) \tag{10.109}$$

is given by the Euler character.

Another way to see the non-semisimplicity is to note that there is a "conserved charge," namely the degree of the form. Since the ring is finite, anything with positive charge must be nilpotent. But nilpotent matrices cannot be diagonalized.

Remarks:

1. A path integral that leads one to consider the above Frobenius algebra is $N = 1$ supersymmetric quantum mechanics with target a Riemannian manifold M. The quantum mechanics of a particle moving on a Riemannian manifold M is described by the action:

$$S = \int dt \frac{1}{2} G_{\mu\nu}(x(t)) \dot{x}^\mu \dot{x}^\nu \tag{10.110}$$

The Hilbert space of the theory is $L^2(M)$, the space of L^2 functions defined by the Riemannian metric. The Hamiltonian is classically

$$H = \frac{1}{2} p_\mu G^{\mu\nu} p_\nu \tag{10.111}$$

and in the quantum theory this becomes the standard Laplacian

$$\Delta = \frac{1}{\sqrt{g}} \partial_\mu \sqrt{g} G^{\mu\nu} \partial_\nu \tag{10.112}$$

acting on $L^2(M)$. When we make the theory supersymmetric The Hilbert space of the full theory is naturally isomorphic to the DeRham complex $\Omega^*(M)$ and one of the supersymmetry operators Q acts as the exterior derivative under this identification. The Hamiltonian is the Hodge Laplacian $(d + d^\dagger)^2$. Restricting to a "BPS sector" or "supersymmetric sector" or "topological sector" of states annihilated by $d + d^\dagger$ is restricting to the Harmonic forms. By a standard theorem this space of harmonic forms is isomorphic as a vector space with the DeRham cohomology. In general "BPS sectors" or "supersymmetry preserving sectors" of a theory are related to topological field theory. We will be more precise about that later.

2. Rational Homotopy Theory. The cohomology associates to a manifold a "differential graded algebra," (with $d = 0$). According to a famous theorem of Sullivan and Quillen, such DGA's characterize manifolds up to rational homotopy type, i.e. they determine $\pi_i(M) \otimes \mathbb{Q}$. ***** More detail here ****

10.4 Example 4: Landau-Ginzburg theory

An important example of Frobenius algebras in string theory are provided by 2-dimensional $N = 2$ supersymmetric Landau-Ginzburg theories. They provide moreover a nice set of examples for comparing the semisimple and non-semisimple cases.

Once again, one begins with a physical theory - a $d = 2$ $(2, 2)$ supersymmetric quantum field theory and restricts to a "topological sector" of the theory provided by a "topological twist." In the simplest case, the quantum field theory has as target space a linear space of fields $X_1, \dots, X_n$ (which we will just regard as coordinates on $\mathbb{C}^n$) and a holomorphic function $W(X_1, \dots, X_n)$ known as the "superpotential." [19]

Two-dimensional field theories with $(2, 2)$ supersymmetries actually admit two kinds of topological twists called "A-twists" and "B-twists." In the "B-twisted model" the chiral

[19] The X_i are actually "chiral superfields" and the theory also requires a nonholomorphic function $K(X_i, \bar{X}_i)$ known as the "Kahler potential."

superfields satisfy an algebra which turns out to be the polynomial algebra factored by the Jacobian ideal:

$$\mathcal{C} = \mathbb{C}[X_i]/(\partial_i W) \tag{10.113}$$

The Frobenius structure is defined by a residue integral in general. In the one-variable case we define ♣Give the general formula. ♣

$$\theta(\phi) := \mathrm{Res}_{X=\infty}\frac{\phi(X)}{W'(X)} \tag{10.114}$$

The vacua of the theory correspond to the critical points of W:

$$\frac{\partial W}{\partial X^i}|_{\vec{X}_0} = 0 \tag{10.115}$$

The critical points are said to be of Morse type is the matrix of second derivatives:

$$\frac{\partial^2 W}{\partial X^i \partial X^j}|_{\vec{X}_0} \tag{10.116}$$

is nondegenerate. Physically Morse critical points correspond to massive theories, while nonMorse critical points renormalize to nontrivial 2d CFT's in the infrared. Note that W is holomorphic, so these definitions are analogous to, but different from the definitions we gave above for a real Morse function.

If the critical points of W are all Morse then the algebra (10.113) is semisimple. Indeed, if all the critical points are Morse then the trace is easily written in terms of the critical points p_a as

$$\theta(\phi) = \sum_{dW(p_a)=0} \frac{\phi(p_a)}{\det(\partial_i\partial_j W|_{p_a})} \tag{10.117}$$

Example $W = \frac{1}{n+1}X^{n+1} - qX$. The critical points are at $\omega^j q^{1/n}$ where ω is a primitive n^{th} root of unity. Clearly, W'' is nonzero there.

10.5 Example 5: Quantum cohomology

Let X be a Kähler manifold. One can formulate the two-dimensional supersymmetric sigma model with X as target space. It is a theory of maps

$$\phi : \Sigma \to X \tag{10.118}$$

with action

$$S = \int_\Sigma (d\phi, *d\phi) + \cdots \tag{10.119}$$

For these models one can define an "A-twisted" topological field theory and the local operators are in 1-1 correspondence with $H^*_{DR}(X)$. However, the correlation functions of Q-invariant operators define a Frobenius algebra that is a *deformation* of the Frobenius algebra structure we saw above.

This is known as *quantum cohomology.*

Example Let $X = \mathbb{CP}^n$. The cohomology ring has a generator $x \in H^2(X)$ and is the algebra $\mathbb{C}[x]/(x^{n+1})$. When we consider the A-model with target space X there is a local operator $\mathcal{O}_x$ that generates the ring of Q-invariant local operators and this ring can be shown to be

$$\mathbb{C}[x]/(x^{n+1} - q) \tag{10.120}$$

where $q = e^{-A}$ and A is the Kähler class of X. For $q \neq 0$ this Frobenius algebra is semisimple.

It is thought that for general non-Calabi-Yau target spaces the quantum cohomology ring is semisimple.

11. Emergent Spacetime

It is very unusual to have a space of quantum states be an *algebra*. (We will stress this with a little review of quantum mechanics in the next section.)

♣Comment some more on how there is a state-operator correspondence in conformal field theory. ♣

What should we make of the fact that the space of states in a 2D TFT is an algebra? At least in the semisimple case there is a very nice answer. A beautiful theorem of Gelfand tells us that one can naturally associate a space to any commutative C^*-algebra. In this section we will describe that in somewhat heuristic terms. Then we will return to it in more formal terms.

11.1 The algebra of functions on a topological space X

Consider a topological space X. Let us begin to transcribe topological/geometric concepts into algebraic concepts.

Consider

$$\mathcal{C}(X) := \{f : X \to \mathbb{C} : f \quad \text{is} \quad \text{continuous}\} \tag{11.1}$$

What are the algebraic structures of $\mathcal{C}(X)$?

- $\mathcal{C}(X)$ is clearly a *vector space* over $\mathbb{C}$.
- Moreover, $\mathcal{C}(X)$ is an *algebra*: you can multiply functions:

$$(f_1 \cdot f_2)(x) := f_1(x) f_2(x) \tag{11.2}$$

Note that it is a commutative associative algebra.

It is interesting to study the representations of $\mathcal{C}(X)$. Because $\mathcal{C}(X)$ is commutative, one should study its 1-dimensional representations. Indeed, there is an obvious source of such representations given by the *evaluation map*. Given a point $x_0 \in X$, we define $\mathbf{ev}_{x_0}$:

$$\mathbf{ev}_{x_0}(f) := f(x_0) \tag{11.3}$$

This is a one-dimensional representation of $\mathcal{C}(X)$.

Recall that *characters* of an algebra are the algebra homomorphisms to $\mathbb{C}$, that is

$$\chi(fg) = \chi(f)\chi(g) \tag{11.4}$$

For a commutative algebra these coincide with the irreducible representations. The evaluation map is a character, and, it turns out, it is the only kind of character we can construct. This is reasonable: A linear functional on $\mathcal{C}(X)$ should depend linearly on $f(x)$ at every x and should therefore be some kind of sum $\sum_{x\in X} f(x)\mathbf{ev}_x$ (making this precise when the sum is infinite is the domain of functional analysis) but this will only be a character if $f(x)$ is only supported at a point. So we could think of X instead as the space of characters of $\mathcal{C}(X)$. This is a way of "algebraicizing X."

There is another way to think about X algebraically which is more obviously a formulation involving just the "internal structure" of $\mathcal{C}(X)$. If χ is a character then from (11.4) it follows that the kernel of χ is an *ideal* $I(\chi) \subset \mathcal{C}(X)$. For the character given by the evaluation map, we have the associated ideal:

$$\ker \mathbf{ev}_{x_0} := \mathbf{m}_{x_0} := \{f : f(x_0) = 0\} \tag{11.5}$$

In fact, $\mathbf{m}_{x_0}$ is a *maximal ideal.* This means there is no nontrivial ideal which contains $\mathbf{m}_{x_0}$ as a proper subset. In the present example the claim is easily verified. If $\mathbf{m}_{x_0} \subset I$ for some strictly larger ideal I then I must contain a function h such that $h(x_0) \neq 0$. But this means that the function 1 is in I since we may write:

$$h - (h - h(x_0)\mathbf{1}) = h(x_0)\mathbf{1} \tag{11.6}$$

and since $(h - h(x_0)\mathbf{1}) \in \mathbf{m}_{x_0}$ the LHS is in I. Since $h(x_0) \neq 0$, we can divide by it, and we conclude $\mathbf{1} \in I$. Since I is an ideal $I = \mathcal{C}(X)$.

Exercise *Identifying characters and maximal ideals*

a.) Show that

$$0 \to \mathbf{m}_{x_0} \to C(X) \to \mathbb{C} \to 0 \tag{11.7}$$

b.) When $I \subset A$ is an ideal in an algebra A then we can define an algebra A/I. Explain why, in this case $\mathcal{C}(X)/\mathbf{m}_{x_0}$ is in fact a field.

In a similar way we can algebraicize maps between topological spaces. If

$$\phi : X \to Y \tag{11.8}$$

is a continuous map then the "pullback map"

$$\phi^* : \mathcal{C}(Y) \to \mathcal{C}(X) \tag{11.9}$$

is defined by $f \to \phi^*(f)$ whose values are, by definition

$$\phi^*(f)(x) := f(\phi(x)) \tag{11.10}$$

The key algebraic structure here is: *ϕ is a homomorphism of algebras.*

Now the key idea of the algebraic approach is that, in some sense the maximal ideals are in 1-1 correspondence with the points of X, and the algebra homomorphisms $\mathcal{C}(Y) \to \mathcal{C}(X)$ are in 1-1 correspondence with the maps $\phi : X \to Y$. In order to make this work and to have some control on notions of topology we need some algebraic notion corresponding to the fact that we have continuous functions. (Nothing we said above relied on the fact that we were talking about continuous functions.) The solution to this problem is to consider commutative C^* algebras.

To get an idea of what a commutative C^* algebra is note that we can make $C(X)$ into a normed vector space by defining

$$\| f \| := \sup_{x\in X} |f(x)| \tag{11.11}$$

Moreover, there is a natural $\mathbb{C}$-antilinear map $f \to f^*$ such that

$$\| f f^* \| = \| f \|^2 \tag{11.12}$$

We develop these ideas more rigorously below. For the moment suffice it to say that given a commutative $C*$ algebra one can turn the space of characters, or equivalently, the space of ideals into a Hausdorff topological space. If the algebra has a unit it is compact. Before going into the general theory first let us see how it works in the fairly trivial context of 2D TFT.

11.2 Application To 2D TFT

Let us now apply these ideas to 2D TFT. We have seen that it is equivalent to a finite-dimensional commutative Frobenius algebra.

If $\mathcal{C}$ is semisimple, then it is isomorphic to $\oplus_x \mathbb{C}\epsilon_x$. This is a unital C^* algebra, and as we have just learned we can associate to it a compact topological space. In fact, in this case the space is just a finite disjoint set of points corresponding to the idempotents.

Thus the "target space physics" in this example is the following: Spacetime consists of a finite disjoint set of 0-dimensional disconnected "universes." Each basic idempotent ϵ_x corresponds to a point x. The only physical information is in $\theta_x = \theta(\epsilon_x) = Z(S^2)$, for that universe. We should interpret this as the string coupling, since the contribution of x to $Z_{\rm string}$ is

$$Z_{\rm string} = \sum_x g_x^{-2} \frac{1}{1 - g_x^2} \tag{11.13}$$

where $\theta_x = g_x^{-2}$.

We will further justify this interpretation when we consider open-closed theory and boundary conditions.

Remarks:

1. When $\mathcal{C}$ is not semisimple, the spacetime interpretation is not so straightforward.

2. Two-dimensional conformal field theory is a nice generalization of 2D TFT. In this case there is a state-operator correspondence, so that the space of states $\mathcal{H}(S^1)$ assigned to a circle is also an algebra. It is related to the mathematical theory of vertex operator algebras. In some cases, e.g. the level k WZW theory for a compact simple group G, a subsector of the operator product algebra approaches the algebra of functions on G as $k \to \infty$. To be more precise, if λ is in the affine Weyl chamber then the vertex operators corresponding to the states in $L(\lambda) \otimes \widetilde{L(\lambda)}$ at the lowest value of $L_0 + \tilde{L}_0$ correspond operators

$$\Phi^{\lambda}_{\mu,\tilde{\mu}}(z_0) \tag{11.14}$$

where z_0 is the operator-instertion point. These can be described in the WZW path integral as insertions of matrix elements of $\rho_\lambda(g(z_0))$ where ρ_λ is the representation of the finite-dimensional group G. These operators have $L_0 = \tilde{L}_0$ eigenvalue

$$\Delta_\lambda = \frac{(\lambda, \lambda + 2\rho)}{2(k+h)} \tag{11.15}$$

(So for $SU(2)_k$ a primary field of spin j has $\Delta_j = j(j+1)/(k+2)$.) ♣CHECK! ♣

The operator product expansion of these operators takes the form:

$$\Phi^{\lambda_1}_{\mu_1,\tilde{\mu}_1}(z_1)\Phi^{\lambda_2}_{\mu_2,\tilde{\mu}_2}(z_2) \sim \sum_{\lambda_3,\mu_3,\tilde{\mu}_3} \left| (z_1 - z_2)^{\Delta_{\lambda_3} - \Delta_{\lambda_1} - \Delta_{\lambda_2}} \right|^2 C_{123}\, \Phi^{\lambda_3}_{\mu_3,\tilde{\mu}_3}(z_2)\,(1 + \mathcal{O}(z_{12}, \bar{z}_{12})) \tag{11.16}$$

The OPE coefficients C_{123}, depend on $\lambda_i, \mu_i, \tilde{\mu}_i$ as well as k. They have a good $k \to \infty$ limit and in fact approach the usual structure constants for the multiplication of functions in $L^2(G)$. Since the weights $\Delta_\lambda \to 0$ for $k \to \infty$ at fixed λ if we take $k \to \infty$ and $z_1 \to z_2$ the OPE algebra becomes the commutative algebra of functions on G.

This, and other examples, gives a hint that vertex operator algebras provide a kind of generalization of geometry and topology. When we consider the chiral vertex operators, or - closely related - vertex operators on boundaries of Riemann surfaces then we find that the algebras are noncommutative.

There are many many papers exploring this idea. Two examples are [16, 45]

Exercise
What is the analog of (11.13) in the unoriented case?

---

12. Quantum Mechanics And C^* Algebras

This section is somewhat outside the main line of development of this chapter. Nevertheless we have included it for several reasons:

1. The material is of intrinsic interest.
2. The material will be of use in the more algebraic description of bundle theory.
3. We give a complete and rigorous proof of the Gelfand theorem, and put it into some context.

For the functional analysis I will follow:

1. N.P. Landsman, "Lecture Notes on C*-Algebras, Hilbert C*-modules, and Quantum Mechanics," arXiv:math-ph/9807030
2. Reed and Simon, *Methods of Modern Mathematical Physics*, especially, vol. I.
3. Rudin
4. Murphy
5. Varilly
6. Wegge-Olsen

12.1 Banach Algebras

The proper notion of continuity is captured by the notion of a "C^* algebra." In order to appreciate this concept we need to take a few steps back and review some standard functional analysis.

Definition. A vector space V is a *normed vector space* if there is a function $v \to \parallel v \parallel \in \mathbb{R}_+$ such that:

1. $\parallel v_1 + v_2 \parallel \leq \parallel v_1 \parallel + \parallel v_2 \parallel$.
2. $\parallel \alpha v \parallel = |\alpha| \parallel v \parallel$, where α is a scalar.
3. $\parallel v \parallel > 0$ if $v \neq 0$.

The key point about a normed vector space is that we can define a notion of open sets, and therefore continuity, etc. Thus, for example:

1. A sequence $\{v_n\} \subset V$ converges to $v \in V$ if

$$\lim_{n\to\infty} \parallel v_n - v \parallel = 0 \tag{12.1}$$

2. A *Cauchy sequence* $\{v_n\} \subset V$ is a sequence so that for all $\epsilon > 0$ there is an N so that for all $n, m > N$ we have $\parallel v_n - v_m \parallel < \epsilon$.
3. Every convergent sequence is a Cauchy sequence, but the converse is false.

Definition. A normed vector space is *complete* if every Cauchy sequence in V converges to some vector $v \in V$. Such a normed vector space is called a *Banach space*.

Note that it follows from the triangle inequality that for all v_1, v_2 in any normed vector space

$$| \parallel v_1 \parallel - \parallel v_2 \parallel | \leq \parallel v_1 - v_2 \parallel \tag{12.2}$$

Therefore, if $\{v_n\}$ converges to $v \in V$, then

$$\lim_{n\to\infty} \parallel v_n \parallel = \parallel v \parallel \tag{12.3}$$

Do not confuse a Banach space with a Hilbert space. To define the latter we first recall that an *inner product space* is a complex or real vector space V with a sesquilinear inner product: A map $V \times V \to \kappa$ (where the ground field is $\mathbb{R}$ or $\mathbb{C}$) that is linear in the second variable and antilinear in the first. The inner product is positive definite when $(v, v) = 0$ iff $v = 0$. Given a positive definite inner product we can define the structure of a normed vector space on V via

$$\parallel v \parallel := \sqrt{(v, v)}. \tag{12.4}$$

The slightly nontrivial point to check is the triangle inequality and this follows from the Cauchy-Schwarz inequality.

Definition A *Hilbert space* is a positive definite inner product space which is complete in the norm (12.4).

Thus, a Hilbert space structure on V defines a Banach space structure on V. We will soon see examples of Banach spaces that are not Hilbert spaces.

Suppose that $\mathcal{V}_1$ and $\mathcal{V}_2$ are two normed vector spaces. A linear operator $T : \mathcal{V}_1 \to \mathcal{V}_2$ is said to be *bounded* if

$$\parallel T \parallel := \sup_{v \in \mathcal{V}_1, v \neq 0} \frac{\parallel Tv \parallel_2}{\parallel v \parallel_1} < \infty \tag{12.5}$$

where the subscripts on the norms on the RHS remind us that we use the norms in $\mathcal{V}_1$ and $\mathcal{V}_2$, respectively. We will usually drop them to keep the equations from getting too busy.

In particular, if T is a bounded operator then for all v:

$$\parallel Tv \parallel \leq \parallel T \parallel \cdot \parallel v \parallel . \tag{12.6}$$

Note that if T_1, T_2 are bounded then so is $\alpha T_1 + \beta T_2$. Let $\mathcal{L}(\mathcal{V}_1, \mathcal{V}_2)$ be the vector space of all bounded operators. We claim that $T \to \parallel T \parallel$ is itself a norm on $\mathcal{L}(\mathcal{V}_1, \mathcal{V}_2)$. It is called the *operator norm*. To check this we must show for example that $\parallel T_1 + T_2 \parallel \leq \parallel T_1 \parallel + \parallel T_2 \parallel$. This holds because

$$\begin{aligned} \parallel T_1 + T_2 \parallel &:= \sup_{v \in \mathcal{V}_1, v \neq 0} \frac{\parallel (T_1 + T_2)(v) \parallel_2}{\parallel v \parallel_1} \\ &\leq \sup_{v \in \mathcal{V}_1, v \neq 0} \frac{\parallel T_1 v \parallel_2}{\parallel v \parallel_1} + \sup_{v \in \mathcal{V}_1, v \neq 0} \frac{\parallel T_2 v \parallel_2}{\parallel v \parallel_1} \\ &= \parallel T_1 \parallel + \parallel T_2 \parallel \end{aligned} \tag{12.7}$$

Proposition: If $\mathcal{V}_1, \mathcal{V}_2$ are normed linear spaces and $\mathcal{V}_2$ is a Banach space then $\mathcal{L}(\mathcal{V}_1, \mathcal{V}_2)$ is itself a Banach space.

Proof: Suppose $\{T_n\}$ is a Cauchy sequence in the operator norm. That is $\forall \epsilon > 0$ $\exists N$ such that $n, m > N$ implies $\parallel T_n - T_m \parallel < \epsilon$. Then for all $v \in \mathcal{V}_1$ $\{T_n(v)\} \subset \mathcal{V}_2$ is a Cauchy sequence and therefore has a limit because $\mathcal{V}_2$ is a Banach space. We call the limit $T(v)$, and it is easy to prove that $v \mapsto T(v)$ is a linear operator. We claim that T is a bounded operator. To prove this note that for all $v \neq 0$

$$
\begin{aligned}
\frac{\parallel T(v) \parallel_2}{\parallel v \parallel_1} &= \lim_{n\to\infty} \frac{\parallel T_n(v) \parallel_2}{\parallel v \parallel_1} \\
&\leq \lim_{n\to\infty} \parallel T_n \parallel
\end{aligned}
\tag{12.8}
$$

But we know that $\{\parallel T_n \parallel\}$ is a Cauchy sequence of real numbers, by (12.2). It therefore converges, so T is a bounded operator. Taking the supremum over all $v \neq 0$ in (12.8) we see that the limit is $\parallel T \parallel$.

Finally, we need to show that $T_n \to T$ in the operator norm. But

$$
\parallel (T - T_n)v \parallel = \lim_{m\to\infty} \parallel (T_m - T_n)v \parallel \leq \parallel v \parallel \lim_{m\to\infty} \parallel T_m - T_n \parallel \tag{12.9}
$$

so

$$
\parallel T - T_n \parallel \leq \lim_{m\to\infty} \parallel T_m - T_n \parallel \tag{12.10}
$$

which suffices to show that $\lim_{n\to\infty} \parallel T - T_n \parallel = 0$. ♠

There are two useful things we can immediately conclude from this Proposition. First, we can make the following important definition:

Definition: A *functional* on a Banach space $\mathcal{B}$ is a linear map $\ell : \mathcal{B} \to \mathbb{C}$ that is a bounded operator. The *dual Banach space* is the Banach space $\mathcal{B}^\vee := \mathcal{L}(\mathcal{B}, \mathbb{C})$ in the operator norm.

Second, it follows that if $\mathcal{B}$ is a Banach space then $\mathfrak{B} := \mathcal{L}(\mathcal{B}, \mathcal{B})$ is also an a Banach space. On the other hand, it is also an algebra. The one thing we need to check is that the product of operators is again bounded. But this follows because

$$
\parallel T_1 T_2 v \parallel \leq \parallel T_1 \parallel \cdot \parallel T_2 v \parallel \leq \parallel T_1 \parallel \cdot \parallel T_2 \parallel \cdot \parallel v \parallel \tag{12.11}
$$

Thus, $T_1 T_2$ is indeed bounded, and moreover:

$$
\parallel T_1 T_2 \parallel \leq \parallel T_1 \parallel \cdot \parallel T_2 \parallel \tag{12.12}
$$

This motivates the

Definition: A Banach space $\mathfrak{A}$ is called a *Banach algebra* if it has an algebra structure such that for all $a_1, a_2 \in \mathfrak{A}$:

$$
\parallel a_1 a_2 \parallel \leq \parallel a_1 \parallel \cdot \parallel a_2 \parallel \tag{12.13}
$$

Remarks:

1. Note that a bounded operator is always continuous in the norm topologies on $\mathcal{V}_1, \mathcal{V}_2$. Indeed, for all $\epsilon > 0$ if $\parallel v - v_0 \parallel \leq \delta := \epsilon / \parallel T \parallel$ then $\parallel T(v) - T(v_0) \parallel \leq \epsilon$. In a similar way, in a Banach algebra the multiplication $\mathfrak{B} \times \mathfrak{B} \to \mathfrak{B}$ is continuous in each variable.

2. If $v \in \mathcal{B}$ and $\ell(v) = 0$ for all $\ell \in \mathcal{B}^*$ then $v = 0$. This is not completely trivial and is a consequence of the Hahn-Banach theorem. Recall that the HB theorem says that if $\mathcal{B}_0 \subset \mathcal{B}$ is a linear subspace and $\ell_0 : \mathcal{B}_0 \to \mathbb{C}$ is a bounded linear functional then there is an extension to $\ell : \mathcal{B} \to \mathbb{C}$ with $\parallel \ell_0 \parallel = \parallel \ell \parallel$. (This theorem is nontrivial. For a proof see any textbook on functional analysis.) For $v \neq 0$ we can take $\mathcal{B}_0$ to be the line through v and define $\ell_0(\lambda v) := \lambda$. Then the extension ℓ clearly has the property that $\ell(v) \neq 0$.

3. If $\mathcal{H}$ is a Hilbert space then $\mathcal{B}(\mathcal{H}) := \mathcal{L}(\mathcal{H}, \mathcal{H})$ is a good example of a Banach space that is not a Hilbert space. One might try to define a sesquilinear form by $(T_1, T_2) = \mathrm{Tr}_{\mathcal{H}} T_1^\dagger T_2$ but this will not converge in general.

12.2 C^* Algebras

Definition: Let $\mathcal{A}$ be an algebra over $\mathbb{C}$. Then it is a $*$-*algebra* if there is a $\mathbb{C}$-antilinear involution $\mathcal{A} \to \mathcal{A}$ denoted $a \to a^*$ such that $(ab)^* = b^* a^*$.

Example: $\mathcal{B}(\mathcal{H})$ is a good example of a $*$ algebra, where $*$ is the usual Hermitian adjoint. Now we can note a nice way that $*$ interacts with the norm in this case:

$$\begin{aligned} \parallel Tv \parallel^2 &= (Tv, Tv) \\ &= (v, T^* T v) \\ &= |(v, T^* T v)| \\ &\leq \parallel v \parallel \parallel T^* T v \parallel \qquad\qquad \text{Cauchy} - \text{Schwarz} \\ &\leq \parallel v \parallel^2 \parallel T^* T \parallel \end{aligned} \tag{12.14}$$

It follows that

$$\parallel T \parallel^2 \leq \parallel T^* T \parallel \tag{12.15}$$

but we also know from (12.12) that

$$\parallel T^* T \parallel \leq \parallel T^* \parallel \parallel T \parallel \tag{12.16}$$

so

$$\parallel T \parallel^2 \leq \parallel T^* T \parallel \leq \parallel T^* \parallel \parallel T \parallel \tag{12.17}$$

and hence

$$\parallel T \parallel \leq \parallel T^* \parallel \tag{12.18}$$

But now replacing $T \to T^*$ and using the fact that $(T^*)^* = T$ we get

$$\| T^* \| = \| T \| \tag{12.19}$$

and from (12.17)

$$\| T^*T \| = \| T \|^2 \tag{12.20}$$

this is known as the C^*-identity. This discussion motivates the:

Definition. A C^*-algebra $\mathfrak{A}$ is a Banach algebra that is also a $*$-algebra and the two structures are compatible in the sense that for all $a \in \mathfrak{A}$:

$$\| a^*a \| = \| a \|^2 \tag{12.21}$$

We have just shown above that for a Hilbert space $\mathcal{H}$, the algebra of bounded operators $\mathcal{B}(\mathcal{H})$ is a C^*-algebra. Therefore, any norm closed subalgebra is a C^*-algebra. This gives lots of examples.

Definition: A *morphism of C^*-algebras* $\mathfrak{A}_1$ and $\mathfrak{A}_2$ is a $\mathbb{C}$-linear map $\varphi : \mathfrak{A}_1 \to \mathfrak{A}_2$ such that

1. $\varphi(aa') = \varphi(a)\varphi(a')$
2. $\varphi(a^*) = (\varphi(a))^*$

♣For isomorphism do we need to demand φ^{-1} is continuous, or is this automatic? ♣

Definition: A *representation of a C^*-algebra* $\mathfrak{A}$ is a morphism $\varphi : \mathfrak{A} \to \mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. It is *faithful* if the morphism is injective.

A nontrivial theorem of Gelfand and Naimark says that up to isomorphism, the <u>only</u> examples are subalgebras of $\mathcal{B}(\mathcal{H})$, for some $\mathcal{H}$: Every C^* algebra is isomorphic to a norm-closed self-adjoint subalgebra of the algebra $\mathcal{B}(\mathcal{H})$ of bounded operators on some Hilbert space This theorem in turn relies on a famous construction of Gelfand-Neimark-Segal. We will explain the GNS construction below.

♣EXPLAIN THIS THEOREM? ♣

Examples Let Θ^{ij} be a $2n \times 2n$ constant, antisymmetric, nondegenerate matrix. One algebra, the "algebra of functions on the noncommutative torus" is defined by taking $2n$ unitary operators U_i:

$$U_i U_i^* = U_i^* U_i = 1 \tag{12.22}$$

with the added relation

$$U_i U_j = \exp[\mathrm{i}\Theta^{ij}] U_j U_i \tag{12.23}$$

Another related algebra is the deformation of the algebra of functions on $\mathbb{R}^{2n}$ called the $*$-product or the Moyal product. [20] The latter is defined via the formula:

[20] According to Wikipedia it was introduced earlier by Groenewald.

$$\left(f_1 *_\Theta f_2\right)(x) := \exp\left[\frac{\mathrm{i}}{2}\Theta^{ij}\frac{\partial}{\partial x_1^i}\frac{\partial}{\partial x_2^j}\right]\left(f_1(x_1)f_2(x_2)\right)\Big|_{x_1=x_2=x} \tag{12.24}$$

We will discuss the associated C^*-algebras and the related physics in Sections ***** below.

We will also need the

Definition If $\mathfrak{A}$ is a C^*-algebra then an element $a \in \mathfrak{A}$ is *self-adjoint* if $a^* = a$.

12.3 Units In Banach Algebras

Definition: A *unit* in a Banach-algebra, denoted $\mathbf{1}$, is a multiplicative unit for the algebra structure such that $\| \mathbf{1} \| = 1$.

If a Banach algebra $\mathfrak{B}$ does not have a unit we can always embed it in another Banach algebra $\mathfrak{B}_{\mathbf{1}}$ constructed as follows: As a vector space

$$\mathfrak{B}_{\mathbf{1}} := \mathfrak{B} \oplus \mathbb{C} \tag{12.25}$$

with the algebra structure

$$(a \oplus \lambda\mathbf{1}) \cdot (a' \oplus \lambda'\mathbf{1}) := (aa' + a\lambda' + a'\lambda) \oplus \lambda\lambda'\mathbf{1} \tag{12.26}$$

The norm is, by definition:

$$\| a \oplus \lambda\mathbf{1} \| := \| a \| + |\lambda| \tag{12.27}$$

It is easy to see that $\mathfrak{B}_{\mathbf{1}}$ is a Banach space. Moreover, a simple application of the triangle inequality and (12.13) for $\mathfrak{B}$ shows that this satisfies (12.13):

$$\begin{aligned} \| (a_1 \oplus \lambda_1\mathbf{1})(a_2 \oplus \lambda_2\mathbf{1}) \| &= \| (a_1a_2 + \lambda_1a_2 + \lambda_2a_1) \oplus \lambda_1\lambda_2\mathbf{1} \| \\ &\le \| a_1a_2 + \lambda_1a_2 + \lambda_2a_1 \| + |\lambda_1| \cdot |\lambda_2| \\ &\le \| a_1 \| \| a_2 \| + |\lambda_1| \, \| a_2 \| + |\lambda_2| \, \| a_1 \| + |\lambda_1| \cdot |\lambda_2| \\ &= \| a_1 \oplus \lambda_1\mathbf{1} \| \cdot \| a_2 \oplus \lambda_2\mathbf{1} \| \end{aligned} \tag{12.28}$$

and hence $\mathfrak{B}_{\mathbf{1}}$ is a Banach algebra.

1. The Banach algebra $\mathfrak{B}_{\mathbf{1}}$ is called the *unitization* of $\mathfrak{B}$. $\mathfrak{B}$ is isometrically embedded in $\mathfrak{B}_{\mathbf{1}}$.

2. There are other ways of adding units.

3. The above definition of the norm does <u>not</u> satisfy the C^* identity for C^*-algebras. We will have to work harder to define a unitization of a nonunital C^*-algebra.

12.4 The Spectrum Of An Element $a \in \mathfrak{B}$

Let $\mathfrak{B}$ be a unital Banach algebra. We define the *spectrum* of an element $a \in \mathfrak{B}$ to be the subset of the complex plane:

$$\sigma(a) := \{z \in \mathbb{C} | a - z\mathbf{1} \quad \text{not} \quad \text{invertible}\} \tag{12.29}$$

The *resolvent* is then defined to be the complement of the spectrum. We will denote it by $\mathcal{R}(a)$.

Theorem The spectrum $\sigma(a)$ is a nonempty, compact subset of the disk of radius $\| a \|$.

Proof:

First, let us show that the spectrum $\sigma(a)$ is nonempty:

For $a \in \mathfrak{B}$ let $\mathcal{R}(a) = \mathbb{C} - \sigma(a)$ denote the resolvent. Define a function

$$g : \mathcal{R}(a) \to \mathfrak{B} \tag{12.30}$$

by

$$g(z) = \frac{1}{z - a} \tag{12.31}$$

Then it is not hard to show that

$$\lim_{z \to \infty} \| g(z) \| = 0 \tag{12.32}$$

Therefore, for any $\ell \in \mathfrak{B}^\vee$, the function $g_\ell : \rho(a) \to \mathbb{C}$ defined by

$$g_\ell(z) := \ell(g(z)) \tag{12.33}$$

is holomorphic and

$$\lim_{z \to \infty} g_\ell(z) = 0 \tag{12.34}$$

Now, if the spectrum $\sigma(a)$ were the empty set then $\mathcal{R}(a) = \mathbb{C}$ and g_ℓ would be entire, but by Liouville's theorem it would have to vanish. Since this argument works for all $\ell \in \mathfrak{B}^\vee$ it follows from the remark 2 in section **** that $g(z) = 0$. Since this is clearly false, we conclude that the spectrum must be nonempty.

Next, let us show that the resolvent $\mathcal{R}(a)$ is open, and hence $\sigma(a)$ is closed. If $z_0 \in \mathcal{R}(a)$ then for

$$|z - z_0| < \frac{1}{\| (a - z_0)^{-1} \|} \tag{12.35}$$

the sum

$$\frac{1}{z_0 - a} \sum_{k=0}^{\infty} \left(\frac{z_0 - z}{z_0 - a} \right)^k \tag{12.36}$$

converges in the norm topology, and it must converge to [21]

$$\frac{1}{z_0 - a}\sum_{k=0}^{\infty}\left(\frac{z_0 - z}{z_0 - a}\right)^k = \frac{1}{z - a} \tag{12.37}$$

Thus, $\mathcal{R}(a)$ is open.

Now we show that the spectrum is bounded:

If $|z| > \parallel a \parallel$ then we can say that

$$\frac{1}{z}\sum_{k=0}^{\infty}\left(\frac{a}{z}\right)^k \tag{12.38}$$

converges in the norm topology. Again it must equal $1/(z-a)$, and hence $z \in \rho(a)$. Therefore $\sigma(a)$ is contained in the disk of radius $\parallel a \parallel$. ♠

Definition The *spectral radius* of $a \in \mathfrak{B}$ is

$$r(a) := \sup\{|z| : z \in \sigma(a)\} \tag{12.39}$$

Theorem[Gelfand's Formula]

$$r(a) = \lim_{n\to\infty} \parallel a^n \parallel^{1/n} \tag{12.40}$$

Proof: [22]

If $R > r(a)$ then we can use the Cauchy formula to say

$$a^n = \frac{1}{2\pi \mathrm{i}}\oint_{C_R} \frac{z^n}{z-a} dz \tag{12.41}$$

where C_R is a circle of radius R. Since $z \to \frac{1}{z-a}$ is continuous for z in the resolvent and C_R is compact

$$M(R) = \sup_\theta \parallel \frac{1}{Re^{\mathrm{i}\theta} - a} \parallel < \infty \tag{12.42}$$

so

$$\parallel a^n \parallel \leq R^n (RM(R)) \tag{12.43}$$

and hence

$$\lim_{n\to\infty} \parallel a^n \parallel^{1/n} \leq R \tag{12.44}$$

for all $R > r(a)$ so

$$\lim_{n\to\infty} \parallel a^n \parallel^{1/n} \leq r(a) \tag{12.45}$$

[21] This uses the basic fact - easily proved - that if $\parallel A \parallel < 1$ then $\sum_{k=0}^{\infty} A^k$ converges in the norm topology to $(1-A)^{-1}$.

[22] I am skipping several details. See Rudin, Theorem 10.13 or Landsman, Proposition 2.2.7. I am also being slightly sloppy by replacing lim sup and lim inf by lim

On the other hand, note that

$$z^n - a^n = (z-a)(z^{n-1} + \cdots + a^{n-1}) \tag{12.46}$$

so if $(z-a)$ is not invertible then $z^n - a^n$ is not invertible. Therefore $z \in \sigma(a)$ implies $z^n \in \sigma(a^n)$. But recall that the spectrum of an operator is in the disk whose radius is the norm of that operator. Therefore $|z^n| \leq \parallel a^n \parallel$. Therefore:

$$r(a) \leq \lim_{n\to\infty} \parallel a^n \parallel^{1/n} \tag{12.47}$$

Putting together (12.45) and (12.47) gives the result ♠

Remarks:

1. Note $\parallel ab \parallel \leq \parallel a \parallel \cdot \parallel b \parallel$ implies that $r(a) \leq \parallel a \parallel$.

2. The spectrum of a nilpotent operator is $\{0\}$, because by (12.40) the spectral radius is zero. Thus, an operator can be nonzero and have zero spectral radius. This also shows that we really can have $r(a) < \parallel a \parallel$.

3. If $\mathfrak{B}$ is not unital we define the spectrum of $a \in \mathfrak{B}$ to be the spectrum of $a \oplus 0$ in the unitization $\mathfrak{B}_{\mathbf{1}}$.

Exercise

Show that if a is an element of the C^*-algebra $M_n(\mathbb{C})$ of $n \times n$ complex matrices, where n is a positive integer, then $\sigma(a)$ coincides with the zeroes of the characteristic polynomial of a.

Exercise *Gelfand-Mazur Theorem*

Show that if a unital $\mathfrak{B}$ algebra is a division algebra, that is, if every element $a \neq 0$ is invertible, then $\mathfrak{B} \cong \mathbb{C}$. [23]

---

[23] *Answer*: For all a there exists a complex number z_a so that $a - z_a\mathbf{1}$ is not invertible, since the spectrum is never empty. But if $\mathfrak{B}$ is a division algebra and $a - z_a\mathbf{1}$ is not invertible then $a = z_a\mathbf{1}$. So the isomorphism is $a \to z_a$. Moreover, $\parallel a \parallel = \parallel z_a\mathbf{1} \parallel = |z_a|$ so the isomorphism is an isometry.

12.5 Commutative Banach Algebras

12.5.1 Characters And $\mathrm{Spec}(\mathfrak{A})$

Definition. A *character* on a commutative Banach algebra $\mathfrak{A}$ is a nonzero linear map

$$\chi : \mathfrak{A} \to \mathbb{C} \tag{12.48}$$

such that

$$\chi(aa') = \chi(a)\chi(a') \tag{12.49}$$

That is, it is a homomorphism of algebras. We denote the set of all characters by $\mathrm{Spec}(\mathfrak{A})$. It is sometimes called the *structure space* of $\mathfrak{A}$.

Some simple consequences of this definition are, first, that if $\mathfrak{A}$ is unital then:

$$\chi(\mathbf{1}) = \mathbf{1} \tag{12.50}$$

because if $\chi \neq 0$ then for some a, $\chi(a) \neq 0$ and therefore $\chi(a) = \chi(\mathbf{1}a) = \chi(\mathbf{1})\chi(a)$. Next, for <u>any</u> z with $|z| > \parallel a \parallel$ we know that $z\mathbf{1} - a$ is invertible, but then $z - \chi(a)$ must be invertible. Therefore $|\chi(a)| \neq r$ for all $r > \parallel a \parallel$, and hence

$$|\chi(a)| \leq \parallel a \parallel \tag{12.51}$$

and in particular $\parallel \chi \parallel = 1$. Therefore $\mathrm{Spec}(\mathfrak{A})$ consists of bounded operators, so $\mathrm{Spec}(\mathfrak{A}) \subset \mathfrak{A}^{\vee}$. In fact it is in the unit "sphere" of elements of norm 1.

Example 1: If $\mathfrak{A} = \mathbb{C}$ then there is exactly one character since $\chi(1) = 1$ and by linearity $\chi(z) = z$.

Example 2: If $\mathfrak{A} = \mathbb{C} \oplus \cdots \oplus \mathbb{C}$ with n summands then there are exactly n characters: χ_i, $1 \leq i \leq n$ vanishes on all the summands except the i^{th} summand, on which $\chi(z) = z$.

Remark: Now we can see why the definition is only interesting for <u>commutative</u> Banach algebras: Consider the matrix algebra $M_n(\mathbb{C})$ for $n > 1$. For each $1 \leq i \leq n$ we can restrict to the subalgebra $\mathfrak{A}_i \cong \mathbb{C}$ of diagonal matrices with diagonal entry $= 1$ for $j \neq i$. Clearly this takes $\chi(a) = a_{ii}$. By multiplicativity it follows that on diagonal matrices $\chi(a) = \mathrm{det}a$. But if S is invertible then $\chi(S^{-1}) = \chi(S)^{-1}$ and hence on the subset of diagonalizable matrices $\chi(a) = \det(a)$. But now this is not linear! So there are <u>no</u> characters on $M_n(\mathbb{C})$ for $n > 1$.

12.5.2 Ideals And Maximal Ideals

Next, let us note the relation of $\mathrm{Spec}(\mathfrak{A})$ to maximal ideals.

Definition In any Banach algebra $\mathfrak{A}$, commutative or not, an *ideal* $\mathfrak{I} \subset \mathfrak{A}$ is a norm-closed linear subspace that is a two-sided ideal in the sense of algebra. It is called a proper ideal if $\mathfrak{I} \neq \mathfrak{A}$. It is called a *maximal ideal* if it is not a proper subset of any proper ideal.

If $\mathfrak{J}$ is any ideal in any Banach algebra $\mathfrak{A}$ then the quotient space $\mathfrak{A}/\mathfrak{J}$ is an algebra:

$$(a+\mathfrak{J})(a'+\mathfrak{J}) := aa' + \mathfrak{J} \tag{12.52}$$

and moreover $\mathfrak{A}/\mathfrak{J}$ can be given a norm:

$$\parallel a+\mathfrak{J} \parallel := \inf_{j\in\mathfrak{J}} \parallel a+j \parallel \tag{12.53}$$

The triangle inequality follows immediately from the definition.

One can prove (for details see Landsman pp. 19-20):

1. $\mathfrak{A}/\mathfrak{J}$ is a Banach space.
2. $\mathfrak{A}/\mathfrak{J}$ is a Banach algebra.

Now let us consider the case that $\mathfrak{A}$ is a unital commutative Banach algebra. Then the kernel of a character $\ker(\chi) := \{a|\chi(a) = 0\}$ is a maximal ideal.

Conversely, if $\mathfrak{J} \subset \mathfrak{A}$ is a maximal ideal in a commutative Banach space then it is the kernel of some character. To prove this note that since $\mathfrak{J}$ is maximal there must be some $b \neq 0$ which is not in $\mathfrak{J}$. Then there must exist an $a \in \mathfrak{A}$ and a $j \in \mathfrak{J}$ so that

$$\mathbf{1} = ba + j \tag{12.54}$$

(Using commutativity of $\mathfrak{A}$ one shows $\{ba + j|a \in \mathfrak{A}, j \in \mathfrak{J}\}$ is an ideal containing both b and $\mathfrak{J}$. Since $\mathfrak{A}$ is maximal it must be all of $\mathfrak{A}$, and therefore contains $\mathbf{1}$.) Let $\pi : \mathfrak{A} \to \mathfrak{A}/\mathfrak{J}$ be the projection. Then

$$\mathbf{1} = \pi(b)\pi(a) \tag{12.55}$$

Therefore $\mathfrak{A}/\mathfrak{J}$ is a Banach algebra where all nonzero elements are invertible. Therefore by the exercise above it is isomorphic to $\mathbb{C}$. Using this isomorphism we can consider the projection $\pi : \mathfrak{A} \to \mathfrak{A}/\mathfrak{J} \cong \mathbb{C}$ is the desired character: $\mathfrak{J} = \ker(\pi)$. So all maximal ideals are kernels of characters.

12.5.3 The Gelfand Transform

Let V be a topological vector space, and $V^\vee$ the dual space. Then there is a natural map $V \to (V^\vee)^\vee$. Given $v \in V$ we define $\hat{v} \in (V^\vee)^\vee$ to be the map

$$\hat{v} : \ell \mapsto \ell(v) \tag{12.56}$$

Notice that the image of V in $(V^\vee)^\vee$ separates points: This means that if $\ell_1 \neq \ell_2$ then there is a $\hat{v}$ so that $\hat{v}(\ell_1) \neq \hat{v}(\ell_2)$.

Definition: If $\mathcal{B}$ is a Banach space then the $w*$-topology on $\mathcal{B}^\vee$ is the weakest topology so that, for all $b \in \mathcal{B}$, the map $\widehat{b} : \mathcal{B}^\vee \to \mathbb{C}$ sending $\ell \to \ell(b)$ is continuous on $\mathcal{B}^\vee$. In this topology a sequence $\{\ell_n\}$ converges to $\ell \in \mathcal{B}^\vee$ iff $\lim_{n\to\infty} \ell_n(b) = \ell(b)$ for all $b \in \mathcal{B}$, and it is the weakest topology with that property.

One key fact we need is

Theorem: [Banach-Alaoglu] If $\mathcal{B}$ is a Banach space then in the $w*$-topology the unit ball in $\mathcal{B}^\vee$ is compact.

Proof: For every $b \in B$ the disk

$$\mathbb{D}_b = \{\lambda \in \mathbb{C} : |\lambda| \leq \parallel b \parallel\} \tag{12.57}$$

is clearly compact, and hence

$$\mathbb{D} := \prod_{b \in \mathcal{B}} \mathbb{D}_b \tag{12.58}$$

is compact in the product topology. [24] An element of this product is the same thing as a function $f : \mathcal{B} \to \mathbb{C}$ for some function. Of course $|f(b)| \leq \parallel b \parallel$, by definition. Therefore the unit disk $\mathcal{B}_1^\vee \subset \mathbb{D}$. It is the subset of functions that happen to be linear. Now suppose $\{f_n\}$ is a sequence of <u>linear</u> functions in $\mathbb{D}$, that is a sequence in $\mathcal{B}_1^\vee$. Then, for all b, we have a sequence of complex numbers $f_n(b)$ in a compact disk. This sequence must converge to a complex number of modulus less or equal to $\parallel b \parallel$. Call the result $f(b)$. But then

$$\begin{aligned} f(\alpha_1 b_1 + \alpha_2 b_2) &= \lim_{n\to\infty} f_n(\alpha_1 b_1 + \alpha_2 b_2) \\ &= \lim_{n\to\infty} \alpha_1 f_n(b_1) + \alpha_2 f(b_2) \\ &= \alpha_1 f(b_1) + \alpha_2 f(b_2) \end{aligned} \tag{12.59}$$

Therefore the map $b \mapsto f(b)$ is linear. Therefore $\mathcal{B}_1^\vee$ is closed in the $w*$-topology and it is obviously bounded. So it is compact. ♠

Now let us apply this to a commutative Banach algebra $\mathfrak{A}$. The $w*$-topology is a topology on $\mathfrak{A}^\vee$ and induces a topology on $\mathrm{Spec}(\mathfrak{A})$. This topology has a basis of open sets labeled by $a \in \mathfrak{A}$ and open sets $\mathcal{O} \subset \mathbb{C}$:

$$\mathcal{U}_{a,\mathcal{O}} := \{\chi \in \mathrm{Spec}(\mathfrak{A}) : \chi(a) \in \mathcal{O}\} \tag{12.60}$$

Theorem Let $\mathfrak{A}$ be a unital commutative Banach algebra. Then $\mathrm{Spec}(\mathfrak{A})$ in the $w*$-topology is a compact Hausdorff topological space.

Proof: $\mathrm{Spec}(\mathfrak{A})$ is Hausdorff because continuous functions separate points. In more detail: suppose $\chi_1 \neq \chi_2$. Then there must be an $a \in \mathfrak{A}$ so that $\chi_1(a) \neq \chi_2(a)$. Choose disjoint open sets $\mathcal{O}_1$ and $\mathcal{O}_2$ in the complex plane that contain $\chi_1(a)$ and $\chi_2(a)$, respectively. Then the open sets $\mathcal{U}_{a,\mathcal{O}_1}$ and $\mathcal{U}_{a,\mathcal{O}_2}$ separate χ_1 and χ_2.

Now we know that $\parallel \chi \parallel = 1$. So $\mathrm{Spec}(\mathfrak{A}) \subset \mathfrak{A}_1^\vee$. Since $\mathfrak{A}_1^\vee$ is compact in the $w*$-topology we need only show that $\mathrm{Spec}(\mathfrak{A})$ is closed. So we need to show that every sequence $\{\chi_n\}$

[24] The product topology on any product of topological spaces $\prod_\alpha X_\alpha$ is generated by the sets $p_\alpha^{-1}(U_\alpha)$ where $p_\alpha : X \to X_\alpha$ is the projection and $U_\alpha \subset X_\alpha$ is open in X_α. The fact that an arbitrary product of compact sets is compact is not at all trivial! It is known as *Tychonoff's theorem*. See any textbook on general topology for a discussion.

in $\mathrm{Spec}(\mathfrak{A})$ that converges, in the $w*$-topology, to some $\chi \in \mathfrak{A}^\vee$, in fact converges to an element of $\mathrm{Spec}(\mathfrak{A})$. That is we need to show that $\mathrm{Spec}(\mathfrak{A})$ is a closed subset of $\mathfrak{A}_1^\vee$.

Suppose $\{\chi_n\}$ is a sequence of characters that converges, in $\mathfrak{A}^\vee$, in the $w*$-topology. That means that for all $a \in \mathfrak{A}$ we have $\chi_n(a) \to \chi(a)$. What we must show is that $a \mapsto \chi(a)$ is in fact a character. To show this note that, for all n:

$$\begin{aligned}
|\chi(aa') - \chi(a)\chi(a')| &= |\chi(aa') - \chi_n(aa') + \chi_n(a)\chi_n(a') - \chi(a)\chi(a')| \\
&\leq |\chi(aa') - \chi_n(aa')| + |\chi_n(a)\chi_n(a') - \chi(a)\chi(a')| \\
&= |\chi(aa') - \chi_n(aa')| + |(\chi_n(a) - \chi(a))\chi_n(a') + \chi(a)(\chi_n(a') - \chi(a')| \\
&\leq |\chi(aa') - \chi_n(aa')| + |(\chi_n(a) - \chi(a))| \parallel a' \parallel + |(\chi_n(a') - \chi(a')| \parallel a \parallel
\end{aligned} \tag{12.61}$$

Now we take the $n \to \infty$ limit to see that $\chi(aa') = \chi(a)\chi(a')$. Thus, $\mathrm{Spec}(\mathfrak{A})$ is $w*$-closed. ♠

Definition The *Gelfand transform* of a commutative Banach algebra (not necessarily unital) is the map

$$\mathcal{G} : \mathfrak{A} \to C(\mathrm{Spec}(\mathfrak{A})) \tag{12.62}$$

given simply by $\mathcal{G}(a) = \hat{a} \in \mathfrak{A}^\vee$. Note that in the $w*$-topology on $\mathrm{Spec}(\mathfrak{A})$, (inherited from the $w*$ topology on $\mathfrak{A}^\vee$) the function $\hat{a}$ is continuous.

Now for any compact Hausdorff topological space X the set of continuous $\mathbb{C}$-valued functions $C(X)$ can be given the structure of a unital commutative Banach algebra: The norm is defined by:

$$\parallel f \parallel := \sup_{x\in X}|f(x)| \tag{12.63}$$

The norm clearly satisfies the requisite properties, and completeness is a standard property of continuous functions. (We use here that X is compact.) Moreover, $\mathcal{C}(X)$ is in fact a C^* algebra with the norm-preserving involution $f \to f^*$ being just complex conjugation. The unit is, of course, the constant function $x \to 1$. (Note that nothing in the above discussion uses the property that X is Hausdorff.)

The spectrum of any continuous function $f \in C(X)$ is its set of values (because a nowhere zero function on $C(X)$ can be inverted). For the continuous function $\mathcal{G}(a)$ on $\mathrm{Spec}(\mathfrak{A})$ the set of values is $\{\chi(a)|\chi \in \mathrm{Spec}(\mathfrak{A})\}$. Rather nicely this coincides with the spectrum of a itself:

$$\sigma(a) = \{\chi(a)|\chi \in \mathrm{Spec}(\mathfrak{A})\} \tag{12.64}$$

To prove (12.64) note that the resolvent of a is the set of complex numbers so that $z - a \in G(\mathfrak{A})$ where $G(\mathfrak{A})$ is the group of invertible elements of $\mathfrak{A}$. Now, if $b \in G(\mathfrak{A})$ then for all χ,

$$\chi(b)\chi(b^{-1}) = \chi(\mathbf{1}) = 1 \tag{12.65}$$

On the other hand, if $b \notin G(\mathfrak{A})$ then it is in the proper ideal $\{ab|a \in \mathfrak{A}\}$ (it is proper, because it does not contain $\mathbf{1}$) and hence in some maximal ideal. But every maximal ideal is the kernel of some $\chi \in \mathrm{Spec}(\mathfrak{A})$. So there is a χ with $\chi(b) = 0$. Therefore

$$b \in G(\mathfrak{A}) \quad \Leftrightarrow \quad \chi(b) \neq 0 \quad \forall \chi \in \mathrm{Spec}(\mathfrak{A}) \tag{12.66}$$

and in particular,

$$z \in \mathcal{R}(a) \quad \Leftrightarrow \quad z - a \in G(\mathfrak{A}) \quad \Leftrightarrow \quad \chi(a) \neq z \quad \forall \chi \in \mathrm{Spec}(\mathfrak{A}) \tag{12.67}$$

Since the spectrum is the complement of the resolvent we get (12.64).

Theorem Let $\mathfrak{A}$ be a unital commutative Banach algebra

1. $\mathcal{G} : \mathfrak{A} \to C(\mathrm{Spec}(\mathfrak{A}))$ is a homomorphism of C^*-algebras.
2. $\mathcal{G}$ is a contraction:

$$\parallel \mathcal{G}(a) \parallel \leq \parallel a \parallel \tag{12.68}$$

Proof: The fact that $\mathcal{G}$ is a homomorphism is easy. For all $\chi \in \mathrm{Spec}(\mathfrak{A})$:

$$\mathcal{G}(a_1 a_2)(\chi) = \chi(a_1 a_2) = \chi(a_1)\chi(a_2) = \mathcal{G}(a_1)(\chi)\mathcal{G}(a_2)(\chi) = (\mathcal{G}(a_1) \cdot \mathcal{G}(a_2))(\chi) \tag{12.69}$$

For the second note that

$$\begin{aligned} \parallel \mathcal{G}(a) \parallel &:= \sup\{|\mathcal{G}(a)(\chi)| : \chi \in \mathrm{Spec}(\mathfrak{A})\} \\ &= \sup\{|\chi(a)| : \chi \in \mathrm{Spec}(\mathfrak{A})\} \\ &= \sup\{|z| : z \in \sigma(a)\} \\ &:= r(a) \leq \parallel a \parallel . \end{aligned} \tag{12.70}$$

Remark Note that the above proves the nice result that $\parallel \mathcal{G}(a) \parallel = r(a)$. For Banach algebras, as opposed to C^* algebras, $\mathcal{G}$ can really be a contraction and not an isometry. We will see an example below.

Exercise

Let $f \in C(X)$. Show that the spectrum of f is the same as the set of values of f in the complex plane.

12.5.4 Commutative C^* Algebras

We are now finally ready to state and prove Gelfand's theorem, in the unital case:

Theorem Let $\mathfrak{A}$ be a unital commutative C^* algebra. Then the Gelfand transform defines an isometric isomorphism

$$\mathcal{G} : \mathfrak{A} \to C(\mathrm{Spec}(\mathfrak{A})) \tag{12.71}$$

and moreover, if $\mathfrak{A}$ is isomorphic to $C(X)$ for any topological space then X is homeomorphic to $\mathrm{Spec}(\mathfrak{A})$.

Proof: We need to show four things:

1. $\mathcal{G}(a^*) = \mathcal{G}(a)^*$
2. $\mathcal{G}$ is an isometry.
3. $\mathcal{G}$ is surjective.
4. The uniqueness statement.

For 1, since $\mathcal{G}$ is $\mathbb{C}$-linear it suffices to show that $\mathcal{G}(a)$ is real-valued if a is self-adjoint. To show this, suppose that $a^* = a$ and let χ be any character and set

$$\chi(a) = \alpha + \mathrm{i}\beta \tag{12.72}$$

Then, for any $t \in \mathbb{R}$ we have

$$\chi(a - (\alpha + \mathrm{i}t)\mathbf{1}) = \mathrm{i}(\beta - t) \tag{12.73}$$

and hence

$$\begin{aligned}
\beta^2 - 2\beta t + t^2 &= |\chi(a - (\alpha + \mathrm{i}t)\mathbf{1})|^2 \\
&\leq \parallel a - (\alpha + \mathrm{i}t)\mathbf{1} \parallel^2 && \text{since} \quad \parallel \chi \parallel = 1 \\
&= \parallel (a - (\alpha + \mathrm{i}t)\mathbf{1}))(a^* - (\alpha - \mathrm{i}t)\mathbf{1}) \parallel && C^* - \text{identity} \\
&= \parallel (a - \alpha\mathbf{1})^2 + t^2\mathbf{1} \parallel \\
&\leq \parallel a - \alpha\mathbf{1} \parallel^2 + t^2 && \text{triangle inequality}
\end{aligned} \tag{12.74}$$

So

$$\beta^2 - 2\beta t \leq \parallel a - \alpha\mathbf{1} \parallel^2 \qquad \forall t \in \mathbb{R} \tag{12.75}$$

and this implies $\beta = 0$. This proves 1.

Now for 2, note that if a is self-adjoint then $\parallel a^2 \parallel = \parallel a \parallel^2$ by the C^*-identity and therefore $\parallel a^{2^n} \parallel = \parallel a \parallel^{2^n}$ so by Gelfand's formula for the spectral radius

$$r(a) = \parallel a \parallel \tag{12.76}$$

However, we have already seen that $r(a) = \parallel \mathcal{G}(a) \parallel$, and hence $\mathcal{G}$ is an isometry on self-adjoint elements. But now the result for general elements is easy:

$$\begin{aligned}
\parallel a \parallel^2 &= \parallel a^* a \parallel && C^* - \text{identity} \\
&= \parallel \mathcal{G}(a^* a) \parallel \\
&= \parallel (\mathcal{G}(a)^* \mathcal{G}(a) \parallel \\
&= \parallel \mathcal{G}(a) \parallel^2 && C^* - \text{identity}
\end{aligned} \tag{12.77}$$

Finally, for 3, we need the

Stone-Weierstrass theorem: If X is a compact Hausdorff space then any C^*-subalgebra of $C(X)$ which separates points and contains $\mathbf{1}$ must be equal to all of $C(X)$. For a proof see J. Conway, Theorem 8.1, p.145 or Reed-Simon, Theorems IV.8-9.

Now just note that the image of $\mathcal{G}$ inside $C(\mathrm{Spec}(\mathfrak{A}))$ separates points, (because $\hat{a}$ separates points), and is a C^*-subalgebra, being the image of a morphism of C^*-algebras.

Thus we have established that

$$\mathfrak{A} \cong C(\mathrm{Spec}(\mathfrak{A})) \tag{12.78}$$

as an isometric isomorphism of C^*-algebras. The fourth and last thing we need to show is that for any compact Hausdorff space X

$$X \cong \mathrm{Spec}(C(X)) \tag{12.79}$$

as a homeomorphism of topological spaces.

To prove (12.79) we consider the evaluation map

$$\mathbf{ev} : X \to \mathrm{Spec}(C(X)) \tag{12.80}$$

A. Because X is Hausdorff we can use Urysohn's lemma from topology to conclude that there is a continuous function separating points. This means that **ev** is injective.

B. To prove that **ev** is surjective we use the identification of $\mathrm{Spec}(C(X))$ with the space of maximal ideals in $C(X)$. Suppose $\mathfrak{I} \subset C(X)$ is a maximal ideal that is not of the form $\ker(\mathbf{ev}_x)$ for some x. That means that for all x there must exist some continuous function $f^{(x)} \in \mathfrak{I}$ with $f^{(x)}(x) \neq 0$. Since $f^{(x_0)}$ is continuous the set $\mathcal{O}^{(x_0)} \subset X$ where $f^{(x_0)}(x) \neq 0$ is an open set, and clearly the $\mathcal{O}^{(x_0)}$ form an open cover of X. Since X is compact there is a finite cover $\{\mathcal{O}^{(x_1)}, \dots, \mathcal{O}^{(x_n)}\}$ so

$$g = \sum_{i=1}^{n} |f^{(x_i)}|^2 \tag{12.81}$$

is everywhere positive, and hence invertible. On the other hand, all the $f^{(x_i)} \in \mathfrak{I}$ and hence $g \in \mathfrak{I}$. But since g is invertible $\mathfrak{I} = C(X)$. This is a contradiction since maximal ideals are proper, by definition. Therefore **ev** is surjective.

C. It is a tautology that $\mathcal{G}(f) \circ \mathbf{ev} = f$. Using the fact that the Gelfand topology is the weakest topology so that $\mathcal{G}(f)$ is continuous we find that the topology on $\mathrm{Spec}(C(X))$ is the same as the original topology on X. (More details in Landsman, Theorem 2.4.1.) ♠

In informal terms, we can go back and forth between a compact topological space and a unital C^* algebra:

$$\begin{aligned} \mathrm{Spec}(C(X)) &\cong X \\ C(\mathrm{Spec}(\mathfrak{A})) &\cong \mathfrak{A} \end{aligned} \tag{12.82}$$

where $\cong$ means homeomorphism of topological spaces in the first line, and isomorphism of C^* algebras in the second line. Moreover, morphisms in one category map nicely to morphisms in the other. This is a nice example of a deep theorem which can be stated succinctly as an equivalence of categories.

Remark: As noted above, even when X is not Hausdorff (but still compact) $C(X)$ makes sense as a C^*-algebra. What happens in this case is that $\mathbf{ev} : X \to \mathrm{Spec}(C(X))$ fails to be injective.

12.5.5 Application of Gelfand's Theorem: The Spectrum Of A Self-Adjoint Element Of $\mathfrak{A}$

A corollary of Gelfand's theorem, proven below, is that if a is self-adjoint then $\sigma(a)$ is a subset of $\mathbb{R}$.

Quite generally, given an element a in a C^*-algebra $\mathfrak{A}$, we can form $C^*(a, \mathbf{1})$, the smallest algebra containing a and $\mathbf{1}$. We form monomials using a and a^*, take finite linear combinations of these, and take the norm closure.

If a and a^* commute, then this algebra is a commutative C^*-algebra, and we can use Gelfand's theorem. Such operators are said to be *normal*: $aa^* = a^*a$. In particular, if a is self-adjoint $a^* = a$, then $C^*(a, \mathbf{1})$ is the closure of the space of polynomials in a.

Theorem 2.5.1 of Landsman explains that for a self-adjoint element $a \in \mathfrak{A}$ the spectrum of a as an element of $\mathfrak{A}$ is the same as the spectrum of a as an element of the commutative C^*-algebra $C^*(1, a)$. In particular the spectrum of the <u>element</u> $a \in \mathfrak{A}$ is homeomorphic to the spectrum of the <u>algebra</u> $C^*(a, \mathbf{1})$. Let us write the homeomorphism as

$$\psi : \sigma(a) \to \mathrm{Spec}(C^*(a, \mathbf{1})) \tag{12.83}$$

One can show that the homeomorphism is such that the Gelfand transform continuous map pulled back to $\sigma(a)$:

$$\psi^*(\mathcal{G}(a)) : \sigma(a) \to \mathbb{C} \tag{12.84}$$

is nothing but the natural inclusion of $\sigma(a) \subset \mathbb{C}$.

We proved above that the Gelfand transform $\mathcal{G}(a)$ of a self-adjoint operator takes real values on the spectrum of a commutative C^*-algebra. So $\sigma(a) \subset \mathbb{R}$.

All of the above statements are proved carefully in Landsman's notes.

If follows that if a is self-adjoint and $f : \sigma(a) \to \mathbb{C}$ is any <u>continuous</u> function then $f(a) \in \mathfrak{A}$ makes sense since it certainly makes sense in $C(\mathrm{Spec}(C^*(a, \mathbf{1})))$ and

$$\| f(a) \| = \| f \|_\infty \tag{12.85}$$

In particular, for f the identity, i.e. the embedding $\sigma(a) \hookrightarrow \mathbb{C}$, we conclude that for self-adjoint a:

$$\| a \| = r(a) \tag{12.86}$$

Therefore, for all a, not necessarily self-adjoint, using the C^*-identity we get

$$\| a \| = \sqrt{r(a^*a)} \tag{12.87}$$

Since the spectral radius is defined purely algebraically, without using the norm, this shows that the norm on a C^* algebra is unique.

12.5.6 Compactness and noncompactness

<u>Motivating Example</u>: The following example (Landsman, p.23) illustrates that the Gelfand transform can be a contraction for Banach algebras that are not C^* algebras, and also shows that if the Banach algebra is non-unital then $\mathrm{Spec}(\mathfrak{B})$ is not compact.

The commutative Banach algebra is $\mathfrak{B} = L^1(\mathbb{R})$, the space of complex-valued functions on $\mathbb{R}$ such that

$$\| f \|_1 := \int_{\mathbb{R}} |f(x)| dx < \infty \tag{12.88}$$

This can be shown to be a Banach algebra with respect to the convolution product:

$$(f_1 * f_2)(x) := \int_{\mathbb{R}} f_1(x-y) f_2(y) dy \tag{12.89}$$

There is no unit in $\mathfrak{B}$ since a unit would have to be a Dirac delta function, which is not in $\mathfrak{B}$. Standard functional analysis (see Reed-Simon) shows that $\mathfrak{B}^\vee \cong L^\infty(\mathbb{R})$, (the Banach space of measurable functions, bounded a.e. and identified if they agree a.e.). The isomorphism maps a linear functional ℓ to the bounded function $\widehat{\ell}(x)$ where:

$$\ell(f) = \int_{\mathbb{R}} f(x) \widehat{\ell}(x) dx \tag{12.90}$$

Now if we want ℓ to be a <u>character</u>:

$$\ell(f_1 * f_2) = \ell(f_1)\ell(f_2) \tag{12.91}$$

then an easy computation shows this is true iff

$$\widehat{\ell}(x_1 + x_2) = \widehat{\ell}(x_1)\widehat{\ell}(x_2) \tag{12.92}$$

for almost all x_1, x_2. This is enough to show that $\widehat{\ell}(x) = e^{\mathrm{i}px}$ for some $p \in \mathbb{R}$. Let us call the character χ_p. Then the definition of the Gelfand transform yields the Fourier transform:

$$\mathcal{G}(f)(\chi_p) = \chi_p(f) = \int_{\mathbb{R}} f(x) e^{\mathrm{i}px} dx \tag{12.93}$$

and indeed the Fourier transform of a convolution product of functions is the pointwise product of Fourier transforms.

Now, it is a consequence of Fourier analysis that

1. $\| \mathcal{G}(f) \| < \| f \|_1$. So the Gelfand transform is strictly a contraction, as promised.
2. For all f, $\lim_{p\to\infty} \mathcal{G}(f)(\chi_p) = 0$. This is the Riemann-Lebesgue lemma and suggests the way to generalize Gelfand's theorem to nonunital algebras.

Remark: This example generalizes nicely to $L^1(G)$ for any locally compact (see below) abelian group. See Section VII.9, especially Theorem 9.6 of Conway: Using the convolution product $L^1(G)$ is a Banach algebra. Its spectrum is isomorphic to the set of continuous homomorphisms $\chi : G \to \mathbb{C}$. This set of characters, usually denoted $\hat{G}$ is itself a locally compact abelian group known as the dual group. We can therefore repeat the construction, and Pontryagin duality says that the dual of $\widehat{G}$ is isomorphic to G. See Kirillov, *Elements of representation theory* for a detailed discussion. In addition to the example $\widehat{\mathbb{R}} \cong \mathbb{R}$ just discussed we have $\widehat{U(1)} \cong \mathbb{Z}$ and $\widehat{\mathbb{Z}/n\mathbb{Z}} \cong \mathbb{Z}/n\mathbb{Z}$.

Recall that one can define a notion of a *locally compact* topological space. If X is Hausdorff this can be defined as a space in which every point has a compact neighborhood, or, equivalently, as a space such that for every point $x \in X$ and every neighborhood $\mathcal{U}$ of x there is a compact neighborhood K of x with $K \subset \mathcal{U}$.

Definition If X is locally compact and Hausdorff we say that f *vanishes at infinity* if, for all $\epsilon > 0$ there is a compact set $K \subset X$ so that $|f| < \epsilon$ on $X - K$.

One can check that for X locally compact the algebra $C_0(X)$ is still a C^*-algebra with the sup-norm:

$$\parallel f \parallel := \sup_{x \in X} |f(x)| \tag{12.94}$$

However, note that if X is noncompact then we have lost the unit element. The function $x \mapsto 1$ certainly doesn't vanish at infinity! Thus $C_0(X)$ is nonunital.

♣Is it true that $C_0(X) = C(X)$ for X compact? Landsman says so but I think this is wrong. ♣

If we replace "compact" with "locally compact but noncompact" on the topological side, and "nonunital" on the algebraic side the Gelfand theorem still holds:

Theorem Let $\mathfrak{A}$ be a non-unital commutative C^* algebra. Then $\mathrm{Spec}(\mathfrak{A})$ is a locally compact but noncompact Hausdorff topological space. The Gelfand transform defines an isometric isomorphism

$$\mathcal{G} : \mathfrak{A} \rightarrow C_0(\mathrm{Spec}(\mathfrak{A})) \tag{12.95}$$

and moreover, if $\mathfrak{A}$ is isomorphic to $C_0(X)$ for any noncompact but locally compact Hausdorff topological space then X is homeomorphic to $\mathrm{Spec}(\mathfrak{A})$.

Proof: See the references: Conway, Landsman, Murphy, Rudin,...

It is often of interest to take a noncompact space and "compactify" it - that is, to find another topological space $\bar{X}$ together with an embedding of X in $\bar{X}$ as an open dense subspace. Compactifications are often very important in physics. Sometimes, for topological arguments we would like to compactify a spacetime. Often we need to compactify moduli spaces of various kinds: instantons, holomorphic bundles or sheaves, spaces of holomorphic maps, Riemann surfaces, super-Riemann surfaces, etc.

A noncompact topological space X can have many different compactifications. We can add one point, or we can add many points. We can add one point and define the 1-point compactification X^+ as follows:

As a set, $X^+ = X \cup \infty$, where ∞ is called the "point at infinity." The open sets of X^+ are then

a.) The open sets of X

b.) Sets of the form $X^+ - K$ where $K \subset X$ is compact.

Example: $(\mathbb{R}^n)^+ = S^n$.

Among the many compactifications of X there is a "maximal" one - the Stone-Čech compactification, denoted βX. It is the largest in the sense that every continuous $f : X \to Z$, where Z is compact and Hausdorff factors through an extension from $\beta X \to Z$. Moreover, for every Hausdorff compactification $\bar{X}$ of X there is a surjective continuous map $\beta X \to \bar{X}$ that restricts to the identity on X. See textbooks on topology for the construction of βX, e.g. Munkres Section 5-3. The space βX can be "large" and "wild." It is generally not used in physics.

In the world of C^*-algebras, if $\mathfrak{A}$ is any C^*-algebra we can embed it into a unital C^*-algebra $\tilde{\mathfrak{A}}$ as follows. [25] Embed $\mathfrak{A}$ into the Banach algebra of bounded operators $\mathcal{L}(\mathfrak{A}, \mathfrak{A})$ by the left-regular representation: $\iota : \mathfrak{A} \to \mathcal{L}(\mathfrak{A}, \mathfrak{A})$ where

$$\iota(a) : a' \mapsto aa' \tag{12.96}$$

This is an isometric embedding $\| \iota(a) \| = \| a \|$. [26] Now we define

$$\tilde{\mathfrak{A}} := \iota(\mathfrak{A}) + \mathbb{C} \cdot \mathbf{1} \tag{12.97}$$

where $\mathbf{1} \in \mathcal{L}(\mathfrak{A}, \mathfrak{A})$ is the unit operator. Note well that we used a + sign and not a $\oplus$ sign! This might or might not be a direct sum. If $\mathfrak{A}$ is unital then $\iota(\mathbf{1}) = \mathbf{1}$ and hence the sum is not direct and $\mathfrak{A} \cong \tilde{\mathfrak{A}}$. If $\mathfrak{A}$ is not unital then we have a direct sum, and $\tilde{\mathfrak{A}} \cong \mathfrak{A}_{\mathbf{1}}$, at least as algebras, where $\mathfrak{A}_{\mathbf{1}}$ was defined for Banach algebras above. Indeed in $\tilde{\mathfrak{A}}$ we have the multiplication:

$$(\iota(a) + \lambda\mathbf{1})(\iota(b) + \mu\mathbf{1}) := \iota(ab + \lambda b + \mu a) + \lambda\mu\mathbf{1} \tag{12.98}$$

The $*$-involution is just $(\iota(a) + \lambda\mathbf{1})^* := \iota(a^*) + \lambda^*\mathbf{1}$. However, the norm inherited from $\mathcal{L}(\mathfrak{A}, \mathfrak{A})$ differs from the norm on $\mathfrak{A}_{\mathbf{1}}$ used above for Banach algebras. The new norm does not obviously satisfy the C^* identity since $\mathcal{L}(\mathfrak{A}, \mathfrak{A})$ has no $*$-involution and is just a Banach algebra. Nevertheless, we can show that the norm on $\tilde{\mathfrak{A}}$ indeed is a C^* norm as follows:

Let $x := \iota(a) + \lambda\mathbf{1}$. Then, by the definition of the operator norm, for every $\epsilon > 0$ we can find a $b \in \mathfrak{A}$ with $\| b \| \leq 1$ so that

$$\begin{aligned}
\| x \|^2 &= \| \iota(a) + \lambda\mathbf{1} \|^2 \\
&\leq \| (\iota(a) + \lambda\mathbf{1})(b) \|^2 + \epsilon \\
&= \| (ab + \lambda b)^*(ab + \lambda b) \|^2 + \epsilon \qquad\qquad C^* - \text{identity} \\
&\leq \| b^* \| \cdot \| (\iota(a) + \lambda\mathbf{1})^*(\iota(a) + \lambda\mathbf{1})(b) \|^2 + \epsilon \\
&\leq \| x^* x \| + \epsilon
\end{aligned} \tag{12.99}$$

So, as in our proof of the C^*-identity for $\mathcal{B}(\mathcal{H})$, we obtain $\| x \|^2 \leq \| x^* x \| \leq \| x^* \| \| x \|$, and as we saw, this is enough to prove the C^*-identity.

The definition of $\tilde{\mathfrak{A}}$ is justified by the nice property:

[25] The following discussion is taken from Wegge-Olsen, ch. 2

[26] $\| \iota(a)(a') \| \leq \| a \| \| a' \|$ so $\| \iota(a) \| \leq \| a \|$. On the other hand by the definition of the operator norm $\| \iota(a)(a^* / \| a \|) \| \leq \| \iota(a) \|$, and by the C^* identity we get $\| a \| \leq \| \iota(a) \|$.

$$C(X^+) = \widetilde{C_0(X)} \tag{12.100}$$

Indeed, note that Spec($\mathfrak{A}$) is embedded in Spec($\tilde{\mathfrak{A}}$) by $\chi \mapsto \tilde{\chi}$ where we can define

$$\tilde{\chi}(\iota(a) + \lambda \mathbf{1}) := \chi(a) + \lambda \tag{12.101}$$

But now there is one more "point"

$$\chi_\infty(\iota(a) + \lambda \mathbf{1}) := \lambda \tag{12.102}$$

One can show that this has the expected properties of a compactification: See Wegge-Olsen, ch. 2.

Remark: There are other ways of unitizing $\mathcal{C}_0(X)$ and similarly other ways of compactifying X. In fact, they are in 1-1 correspondence. See Chapter 2 of Wegge-Olsen. For X locally compact but noncompact $C_b(X)$, the C^*-algebra of *bounded* functions corresponds to the Stone-Čech compactification βX. Note that there are a lot of ways a sequence of bounded functions can "go to infinity." Algebraically, this corresponds to taking the "multiplier algebra" of $\mathfrak{A}$. One embeds $\mathfrak{A} \subset \mathcal{B}(\mathcal{H})$, as guaranteed by Gelfand-Naimark and then $M(\mathfrak{A})$ is the algebra of elements $b \in \mathcal{B}(\mathcal{H})$ such that $b\mathfrak{A} \subset \mathfrak{A}$ and $\mathfrak{A}b \subset \mathfrak{A}$. See Wegge-Olsen ch. 2 for more details.

Exercise

An *essential ideal* is $J \subset A$ is an ideal which intersects every other ideal in A.

Show that $\bar{X}$ is a compactification of X iff $\mathcal{C}(\bar{X})$ is a unital C^* algebra containing $C_0(X)$ as an essential ideal.

12.6 Noncommutative Topology: The C^*-Algebra Dictionary

Continuing along the above lines one can set up a dictionary between topological properties of a locally compact Hausdorff space X and algebraic properties of its $\mathbb{C}^*$ algebra of functions: [27]

[27] This table is taken from Wegge-Olsen and Varilly et. al.

Locally compact Hausdorff topological space	C^* -algebra
point	maximal ideal
open subset	ideal
open dense subset	essential ideal
closed subset	quotient
connected	no nontrivial idempotents
compact	unital
compactification	unitization
one-point compactification	$\tilde{\mathfrak{A}}$
Stone-Čech compactification	$M(\mathfrak{A})$
continuous proper map	homomorphism
homeomorphism	automorphism
measure	positive functional

This suggests a generalization of topology to "noncommutative topology" where one interprets theorems in the theory of general, noncommutative C^* algebras as statements about topology. When one adds extra structure one gets a notion of noncommutative geometry.

Remarks

1. A natural question at this point is whether one can similarly encode, algebraically, smooth structures, metric structures and so forth. For smooth structures we have Proposition 1, p.207 of [12]: Let M be a smooth compact manifold and $\mathcal{A} = C^\infty(M)$ the algebra of infinitely differentiable functions. Then there is an isomorphism of the Hochschild cohomology $H^k(\mathcal{A}, \mathcal{A}^*)$ with the space of k-dimensional DeRham "currents" (currents in the sense of smooth functionals dual to the DeRham complex):
$$\langle \mathcal{D}_\varphi, f_0 df_1 \wedge \cdots \wedge df_k \rangle = \frac{1}{k!} \sum_{\sigma \in S_k} \epsilon(\sigma) \varphi(f_{\sigma(1)}, \ldots, f_{\sigma(k)})(f_0) \tag{12.103}$$

2. Similarly, according to Connes (see [12], chapter VI) the metric structure, at least on a spin manifold, can be encoded into a "Fredholm module" using a Dirac operator. The geodesic distance between two points x_1, x_2 is, roughly speaking the supremum of $|f(x_1) - f(x_2)|$ over all functions for which $\| [D, f] \|$ makes sense, where D is the Dirac operator.

♣Compare: You can't hear the shape of a drum. Also why not use Dirac coupled to the the canonical bundle of Clifford algebras and drop the spin condition? ♣

For details see:

1. A. Connes, *Noncommutative Geometry.* [12] 1. J.M. Gracia-Bondia, J.C. Varilly, and H. Figueroa, *Elements of Noncommutative Geometry*, Birkhauser, 2001

2. N.E. Wegge-Olsen, "K-Theory and C*-Algebras: A Friendly Approach," Oxford
3. A. Connes, Noncommutative Geometry

12.6.1 Hopf Algebras And Quantum Groups

One place where the general philosophy of replacing a space by its algebra of functions has been extremely influential is in the subject of Hopf algebras and quantum groups. We comment briefly on this, following the beautiful introduction to V. Drinfeld's 1986 ICM address. Indeed Drinfeld, similarly to Connes, proposes that in some sense the category of "quantum spaces" should be dual to the category of "noncommutative algebras." The Devil is in the details.

To begin let us consider some structures which the "algebra of functions on a topological group G" will possess. (In this section I will not be careful about questions of analysis, hence the quotation marks.)

Let G be a topological group and $A = Fun(G)$ be some suitable algebra of κ-valued functions, where κ is a field. First of all, let us note that A is an algebra by pointwise multiplication, just as for $C(X)$ in our discussion above. We will denote it by $\mu : A\otimes A \to A$ and of course it is associative:

$$\begin{tikzcd} & A \otimes A \arrow[dr, "\mu"] & \\ A \otimes A \otimes A \arrow[ur, "\mu\otimes Id"] \arrow[dr, "Id\otimes\mu"'] & & A \\ & A \otimes A \arrow[ur, "\mu"'] & \end{tikzcd} \tag{12.104}$$

But now there will be extra structure on A arising from the fact that the group G has extra structure. In particular there is a multiplication

$$m : G \times G \to G \tag{12.105}$$

on the group G. This induces a *comultiplication* on A:

$$\Delta : A \to A \otimes A \tag{12.106}$$

defined by identifying $A \otimes A$ with the algebra of functions on $G \times G$ and declaring $\Delta(f)(g_1, g_2) := f(g_1 g_2)$. Now group multiplication is associative:

$$\begin{tikzcd} & G \times G \arrow[dr, "m"] & \\ G \times G \times G \arrow[ur, "m\times Id"] \arrow[dr, "Id\times m"'] & & G \\ & G \times G \arrow[ur, "m"'] & \end{tikzcd} \tag{12.107}$$

The induced diagram on A reverses all arrows and is the property of *coassociativity* of Δ:

$$\begin{tikzcd} & A \otimes A \arrow[dr, "\Delta\otimes Id"] & \\ A \arrow[ur, "\Delta"] \arrow[dr, "\Delta"'] & & A \otimes A \otimes A \\ & A \otimes A \arrow[ur, "Id\otimes\Delta"'] & \end{tikzcd} \tag{12.108}$$

This makes A a *coassociative coalgebra.*

Now, the next group axiom postulates a unit 1_G. The dual of this is the *counit* $\varepsilon : A \to \kappa$ (where κ is the ground field). For $A = Fun(G)$ we would define $\varepsilon(f) := f(1_G)$. We have two diagrams expressing the properties of the unit:

$$\begin{tikzcd} G \arrow[r, "Id"] \arrow[d, "(Id,1_G)"'] & G \\ G \times G \arrow[ur, "m"'] & \end{tikzcd} \qquad \begin{tikzcd} G \arrow[r, "Id"] \arrow[d, "(1_G,Id)"'] & G \\ G \times G \arrow[ur, "m"'] & \end{tikzcd} \tag{12.109}$$

The dual diagrams give the property of the counit:

$$\begin{tikzcd} A \arrow[r, "Id\otimes 1_\kappa"] \arrow[dr, "\Delta"'] & A \otimes \kappa \\ & A \otimes A \arrow[u, "Id\otimes\varepsilon"'] \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "1_\kappa\otimes Id"] \arrow[dr, "\Delta"'] & \kappa \otimes A \\ & A \otimes A \arrow[u, "\varepsilon\otimes Id"'] \end{tikzcd} \tag{12.110}$$

The final group axiom states that every group element has an inverse. If we say that $\mathcal{I} : G \to G$ is the map $g \mapsto g^{-1}$ then we can define a dual operation $S : A \to A$ by $S(f) = f \circ \mathcal{I}$. The linear operator S is known as the *antipode.* Now the group axiom is the pair of diagrams:

$$\begin{tikzcd} G \arrow[r, "(Id,\mathcal{I})"] \arrow[dr] & G \times G \arrow[r, "m"] & G \\ & \{1_G\} \arrow[ur] & \end{tikzcd} \qquad \begin{tikzcd} G \arrow[r, "(\mathcal{I},Id)"] \arrow[dr] & G \times G \arrow[r, "m"] & G \\ & \{1_G\} \arrow[ur] & \end{tikzcd} \tag{12.111}$$

Dually we get

$$\begin{tikzcd} A \arrow[r, "\Delta"] \arrow[drr, "\varepsilon"'] & A \otimes A \arrow[r, "Id\otimes S"] & A \otimes A \arrow[r, "\mu"] & A \\ & & \kappa \arrow[ur] & \end{tikzcd} \tag{12.112}$$

and a second diagram with $S \otimes Id$.

This motivates the general definition:

Definition A unital algebra A over κ equipped with multiplication μ, comultiplication Δ, counit ε, and antipode S satisfying equations (12.104),(12.108), (12.110),(12.112), is called a *Hopf algebra.*

We stress that this is a general concept. The algebra of functions on a group can be given a Hopf algebra structure, but, as we shall soon see, this is not the general Hopf algebra.

Remark: Now let us note that, quite generally, if A is a Hopf algebra then the vector space dual $A^\vee := \mathrm{Hom}_\kappa(A,\kappa)$ is also a Hopf algebra. First, let us define the product

$$\mu^\vee : A^\vee \otimes A^\vee \to A^\vee \tag{12.113}$$

If ℓ_1, ℓ_2 are two linear functionals then we define their product $\mu^\vee(\ell_1 \otimes \ell_2)$ by declaring that the value on a is obtained from forming $\ell_1 \otimes \ell_2(\Delta(a))$ and then using the multiplication $\kappa \otimes \kappa \to \kappa$. In more detail, suppose a_i is a linear basis for A, and suppose

$$\Delta(a_i) = \sum_{j,k} \Delta_i^{jk} a_j \otimes a_k \tag{12.114}$$

where $\Delta_i^{jk} \in \kappa$ Then $\mu^\vee(\ell_1 \otimes \ell_2) \in A^\vee$ is defined by

$$\mu^\vee(\ell_1 \otimes \ell_2)(a_i) := \sum_{j,k} \Delta_i^{jk} \ell_1(a_j)\ell_2(a_k). \tag{12.115}$$

Similarly, the dual comultiplication on $A^\vee$ is defined by

$$\Delta^\vee(\ell)(a_1 \otimes a_2) := \ell(\mu(a_1 \otimes a_2)) \tag{12.116}$$

The dual counit is

$$\varepsilon_{A^\vee}(\ell) := \ell(1_A) \tag{12.117}$$

and the dual antipode is simply

$$S_{A^\vee}(\ell)(a) := \ell(S(a)) \tag{12.118}$$

We leave it to the reader to check that $(\mu^\vee, \Delta^\vee, \varepsilon_{A^\vee}, S_{A^\vee})$ in fact define a Hopf algebra structure on $A^\vee$.

Applying the above remark to our example of $A = Fun(G)$ we obtain the group algebra $A^\vee = \kappa[G]$. At least formally, this can be viewed as the linear span of $\mathbf{ev}_g : A \to \kappa$ given by $\mathbf{ev}_g(f) = f(g)$. Now the multiplication on $\kappa[G]$ is:

$$\mu^\vee(\mathbf{ev}_{g_1} \otimes \mathbf{ev}_{g_2}) = \mathbf{ev}_{g_1 g_2} \tag{12.119}$$

while the comultiplication is:

$$\Delta^\vee(\mathbf{ev}_g)(f_1 \otimes f_2) = f_1(g) f_2(g) \tag{12.120}$$

and hence

$$\Delta^\vee(\mathbf{ev}_g) = \mathbf{ev}_g \otimes \mathbf{ev}_g \tag{12.121}$$

The counit is

$$\varepsilon^{\vee}(\mathbf{ev}_g) = 1 \qquad \forall g \in G \tag{12.122}$$

and the antipode is

$$S(\mathbf{ev}_g) = \mathbf{ev}_{g^{-1}} \tag{12.123}$$

Now, rather confusingly, $A^{\vee}$ as a vector space can also be identified with an algebra of κ-valued functions on the group G since we can write the general element as

$$\sum_{g\in G} f_g \mathbf{ev}_g \tag{12.124}$$

and $g \mapsto f_g$ is a function on the group. [28] However, viewed this way, the product $\mu^{\vee}$ is the convolution product:

$$\mu^{\vee}(f_1 \otimes f_2)(g) = \int_G f_1(h) f_2(gh^{-1}) dh \tag{12.125}$$

where dh is a Haar measure of volume one, while $\Delta^{\vee}$ takes $g \mapsto f_g$ to a function on $G \times G$ given by

$$\Delta^{\vee}(f)(g_1, g_2) = f_{g_1} \delta_{g_1, g_2} \tag{12.126}$$

In general a Hopf algebra B is said to be "cocommutative" if $\sigma \circ \Delta = \Delta$ where $\sigma : B \otimes B \to B \otimes B$ is the permutation operator.

♣Write commutative diagram ♣

The above two examples $A = Fun(G)$ with the pointwise product and $A^{\vee} = \kappa[G]$ with the convolution product have one property that does not hold for general Hopf algebras: A is commutative and $A^{\vee}$ is cocommutative.

Of course, while $A = Fun(G)$ is commutative, it is not co-commutative when G is noncommutative. Dually, $\kappa[G]$ is not commutative, when G is noncommutative, but it is always cocommutative.

There are other examples of natural Hopf algebras associated to Lie algebras and groups:

1. If G is a compact simple Lie group then its DeRham cohomology $H^*_{DR}(G)$ is a Hopf algebra. The comultiplication is define by the pullback m^* dual to group multiplication. This works because

$$\mathrm{Tr}\left((g_1 g_2)^{-1} d(g_1 g_2)\right)^{2n+1} = \mathrm{Tr}\left(g_1^{-1} dg_1\right)^{2n+1} + \mathrm{Tr}\left(g_2^{-1} dg_2\right)^{2n+1} + dB \tag{12.127}$$

   where B is a trace of a differential form that is a polynomial in d, $g_1^{\pm 1}$ and $g_2^{\pm 1}$. Passing to cohomology we see that m^* indeed defines a co-associative co-multiplication and in fact the traces of powers of Maurer-Cartan forms (which generate H^*_{DR}) are "primitive elements":

$$\Delta(x) = x \otimes 1 + 1 \otimes x \tag{12.128}$$

   Multiplication is the usual cup product of cohomology classes, so this example is (graded) commutative as well as co-commutative.

[28] We will not be careful here about the precise class of functions. For example, if we consider finite sums and G is a continuous group then the relevant functions would be discontinuous, and would only be nonzero at a finite set of points.

2. If $\mathfrak{g}$ is a Lie algebra then the universal enveloping algebra $U(\mathfrak{g})$ is noncommutative, if $\mathfrak{g}$ is noncommutative. It is a Hopf algebra with comultiplication

$$\Delta(x) := x \otimes 1 + 1 \otimes x \tag{12.129}$$

and is therefore cocommutative.

What about examples that are neither commutative nor co-commutative? In his very influential ICM address Drinfeld argued that the most natural source of examples of Hopf algebras that are neither commutative nor cocommutative is to be found in the theory of quantum inverse scattering and factorizable S-matrices. This was one of the major impulses to the modern theory of quantum groups.

Example The simplest nontrivial example is $U_q(\mathfrak{sl}(2))$. It is the unital algebra generated by e, f and K where K is invertible so there is a K^{-1} with $KK^{-1} = 1$. A very standard set of generators of $\mathfrak{sl}(2)$ are e, f, h with

$$\begin{aligned} [h, e] &= 2e \\ [h, f] &= -2f \\ [e, f] &= h \end{aligned} \tag{12.130}$$

The match to the standard basis in physics with $[J^i, J^j] = \epsilon^{ijk} J^k$ is

$$\begin{aligned} h &\to 2\mathrm{i} J^3 \\ e &\to 2\mathrm{i}\left(\frac{J^1 + \mathrm{i}J^2}{2}\right) \\ f &\to 2\mathrm{i}\left(\frac{J^1 - \mathrm{i}J^2}{2}\right) \end{aligned} \tag{12.131}$$

and the standard physics representation $J^a \to -\frac{\mathrm{i}}{2}\sigma^a$ becomes

$$\begin{aligned} h &\to \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ e &\to \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ f &\to \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \end{aligned} \tag{12.132}$$

In any case, the algebra $U_q(\mathfrak{sl}(2))$ is generated by e, f, K, K^{-1} with relations:

$$\begin{aligned} KK^{-1} &= K^{-1}K = 1 \\ KeK^{-1} &= q^2 e \\ KfK^{-1} &= q^{-2} f \\ [e, f] &= \frac{K - K^{-1}}{q - q^{-1}} \end{aligned} \tag{12.133}$$

If we consider $q = e^{\epsilon}$ and $K = e^{\epsilon h}$ then we can recognize the formal $\epsilon \to 0$ limit of these equations as the a deformation of the $\mathfrak{sl}(2)$ Lie algebra. The algebra $U_q(\mathfrak{sl}(2))$ arises rather naturally in the 1+1-dimensional sine-Gordon model. Similar deformations of all the simple Lie algebras can be given. One uses a basis of Serre generators $\{e_i, f_i, h_i\}_{i=1,\dots,r}$, exponentiates $K_i = e^{\epsilon h_i}$ and deforms the standard defining relations by:

$$\begin{aligned} [h_i, e_j] &= A_{ij} e_j \\ [h_i, f_j] &= -A_{ij} f_j \\ [e_i, f_j] &= \delta_{ij} \frac{K_i - K_i^{-1}}{q - q^{-1}} \end{aligned} \tag{12.134}$$

together with a fairly complicated deformation of the Chevalley-Serre relations on e_i, f_j:

$$\begin{aligned} Ad(e_i)^{1-A_{ij}}(e_j) &= 0 \\ Ad(f_i)^{1-A_{ij}}(f_j) &= 0 \end{aligned} \tag{12.135}$$

As the simplest example of how this can arise in an integrable quantum field theory we consider the 1+1 dimensional sine-Gordon model, following [5, 33]. This is a theory of a single real scalar field Φ with action proportional to

$$S = \int d^2x \left(\partial_z \Phi \partial_{\bar{z}} \Phi + \lambda \cos \beta \Phi \right) \tag{12.136}$$

where couplings λ and β determine the mass and interactions and play an important role in the theory. One defines non-locally-related fields [29]

$$\begin{aligned} \phi(x,t) &= \Phi(x,t) + \int_{-\infty}^{x} \partial_t \Phi(y,t) dy \\ \tilde{\phi}(x,t) &= \Phi(x,t) - \int_{-\infty}^{x} \partial_t \Phi(y,t) dy \end{aligned} \tag{12.137}$$

Then form operators related to the creation of solitons:

$$\begin{aligned} J_\pm &= e^{\pm i \mathfrak{a} \phi} & \tilde{J}_\pm &= e^{\mp i \mathfrak{a} \tilde{\phi}} \\ H_\pm &= \lambda \frac{\mathfrak{c}}{\mathfrak{b}} e^{\pm i (\mathfrak{b} \phi + \mathfrak{c} \tilde{\phi})} & \tilde{H}_\pm &= \lambda \frac{\mathfrak{c}}{\mathfrak{b}} e^{\mp i (\mathfrak{b} \tilde{\phi} + \mathfrak{c} \phi)} \end{aligned} \tag{12.138}$$

where $\mathfrak{a}, \mathfrak{d}, \mathfrak{c}$ are all simple real rational functions of β, given in [5, 33]. We then form five nonlocal, but conserved, charges:

$$\begin{aligned} I_\pm &= \int_{-\infty}^{+\infty} (J_\pm + H_\pm) dx \\ \tilde{I}_\pm &= \int_{-\infty}^{+\infty} (\tilde{J}_\pm + \tilde{H}_\pm) dx \\ T &= \frac{\beta}{2\pi} \int_{-\infty}^{+\infty} \partial_x \Phi dx \end{aligned} \tag{12.139}$$

[29] This is a type of duality transformation. Note that $\partial_x \phi$ is a self-dual combination of $d\Phi$ etc.

The last of these is known as the "topological charge" and is only nonzero in soliton sectors of the theory. Then it turns out that if we define $h_1 = -h_0 = T$, and

$$\begin{aligned} e_0 &= \mathfrak{z} e^{T/2} I_- & f_0 &= \mathfrak{z} e^{T/2} \tilde{I}_+ \\ e_1 &= \mathfrak{z} e^{-T/2} I_+ & f_1 &= \mathfrak{z} e^{-T/2} \tilde{I}_- \end{aligned} \tag{12.140}$$

then indeed, the operators in the QFT satisfy the quantum group relations for $U_q(\widehat{\mathfrak{sl}}(2))$, whose Cartan matrix is

$$A_{ij} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \tag{12.141}$$

provided the normalization $\mathfrak{z}$ is a suitable function of the couplings and

$$q = e^{-2\pi \mathrm{i}/\beta^2} \tag{12.142}$$

12.7 The Irrational (And Rational) Rotation Algebras

The irrational rotation algebra, also known as the algebra of functions on the noncommutative torus, is a C^* algebra that arises in many contexts in physics.

12.7.1 Definition

The algebra is the unital C^* algebra generated by U, V with the relations

$$\begin{aligned} UU^* &= U^*U = \mathbf{1} \\ VV^* &= V^*V = \mathbf{1} \\ UV &= e^{2\pi \mathrm{i}\theta} VU \end{aligned} \tag{12.143}$$

We denote this C^*-algebra by $\mathcal{A}_\theta$.

Note that, for all integers $n, m \in \mathbb{Z}$,

$$V^n U^m = e^{-2\pi \mathrm{i} nm\theta} U^m V^n \tag{12.144}$$

so that all monomials can be "normal-ordered." For example, if we decide to put powers of U on the left and V on the right then we would write

$$U^{m_1} V^{n_1} U^{m_2} V^{n_2} \cdots U^{m_k} V^{n_k} = e^{-2\pi \mathrm{i} L\theta} U^M V^N \tag{12.145}$$

where $M = \sum m_i$, $N = \sum n_i$ and $L = \sum_{1 \le i < j \le k} n_i m_j$. (Note that if $m_1 = 0$ then the monomial begins with a power of V, and if $n_k = 0$ it ends with a power of U, so the above is the general monomial.)

Another very useful point of view is that we consider the C^* algebra generated by unitary elements $W(\vec{n})$ associated to vectors in a symplectic lattice $\Lambda \cong \mathbb{Z} \oplus \mathbb{Z}$ with symplectic form $\omega(e_1, e_2) = \theta$ and multiplication rule

$$W(\vec{n}_1) W(\vec{n}_2) = e^{\mathrm{i}\pi\omega(\vec{n}_1, \vec{n}_2)} W(\vec{n}_1 + \vec{n}_2) \tag{12.146}$$

so that $U = W(e_1)$ and $V = W(e_2)$. Of course, this definition easily generalizes to higher dimensional symplectic lattices to define the algebras of noncommutative tori.

However we look at the multiplciation , the general element in the algebra can be written as

$$\sum_{m,n\in\mathbb{Z}} a_{m,n} U^m V^n \tag{12.147}$$

with $a_{m,n} \in \mathbb{C}$ falling off sufficiently rapidly with $m, n \to \infty$. (It certainly includes the Schwarz space $\mathcal{S}(\mathbb{Z}^2)$ of functions decreasing more rapidly than any polynomial in m, n.)

The norm is defined by considering all representations π in Hilbert space of the subalgebra of $\mathcal{A}_\theta$ consisting of polynomials in U, V and taking

$$\parallel a \parallel = \sup_\rho \{\parallel \rho(a) \parallel\} \tag{12.148}$$

where ρ runs over all representations. Of course, the C^* equation implies that

$$\parallel U \parallel = \parallel V \parallel = 1, \tag{12.149}$$

and indeed the norm of any monomial must be one.

These algebras have been much studied by mathematicians and physicists. Here are some notable structural results:

1. A *trace* on a C^* algebra is a map $\tau : \mathfrak{A} \to \mathbb{C}$ such that $\tau(\mathbf{1}) = 1$, $\tau(a^* a) \geq 0$ and $\tau(ab) = \tau(ba)$. When θ is irrational the algebra $\mathcal{A}_\theta$ has a unique trace:

$$\tau\left(\sum_{m,n\in\mathbb{Z}} a_{m,n} U^m V^n\right) = a_{0,0} \tag{12.150}$$

Here is the basic idea: Note that by cyclicity $\tau(UV) = \tau(VU)$, but by the defining relation and linearity of τ, $\tau(VU) = e^{-2\pi i\theta}\tau(UV)$. Therefore $\tau(UV)$ must vanish when $e^{2\pi i\theta} \neq 1$. Similar arguments apply to other monomials, provided θ is irrational. The following is a streamlined version of this argument: Consider the group $U(1) \times U(1)$ acting as automorphisms on $\mathcal{A}_\theta$ via

$$\begin{aligned} \alpha_{z_1,z_2}(U) &= z_1 U \\ \alpha_{z_2,z_2}(V) &= z_2 V \end{aligned} \tag{12.151}$$

Note that if $(z_1, z_2) = (e^{2\pi i n_1\theta}, e^{2\pi i n_2\theta})$ where $(n_1, n_2) \in \mathbb{Z}^2$ then the automorphism α_{z_1,z_2} is an inner automorphism. Now, for any cyclic trace we have

$$\tau(aba^{-1}) = \tau(b) \tag{12.152}$$

and hence $\tau(\alpha_{z_1,z_2}(a)) = \tau(a)$ for $(z_1, z_2) = (e^{2\pi i n_1\theta}, e^{2\pi i n_2\theta})$. But now for a fixed $a \in \mathcal{A}_\theta$ consider the set

$$\{(z_1, z_2) \in U(1) \times U(1) | \tau(\alpha_{z_1,z_2}(a)) = \tau(a)\} \tag{12.153}$$

One can show that the map $\alpha : U(1) \times U(1) \to \mathrm{Aut}(\mathcal{A}_\theta)$ is continuous so this is a closed subset of $U(1) \times U(1)$. On the other hand, it contains the set of elements

$(z_1, z_2) = (e^{2\pi i n_1 \theta}, e^{2\pi i n_2 \theta})$. For θ irrational this set is dense, and hence, for theta irrational we have that $\tau(\alpha_{z_1,z_2}(a)) = \tau(a)$ for <u>all</u> (z_1, z_2). But now we can say that

$$\begin{aligned} \tau(a) &= \oint \oint \tau(\alpha_{z_1,z_2}(a)) \frac{dz_1}{z_1} \frac{dz_2}{z_2} \\ &= \tau \left(\oint \oint \alpha_{z_1,z_2}(a) \frac{dz_1}{z_1} \frac{dz_2}{z_2} \right) \\ &= \tau(a_{0,0}\mathbf{1}) \\ &= a_{0,0} \end{aligned} \tag{12.154}$$

When $\theta \in \mathbb{Q}$ there can be many traces.

2. We can now prove that, so long as θ is irrational the algebra $\mathcal{A}_\theta$ is simple, that is, it has no proper nonzero two-sided ideals. Suppose that $\mathfrak{I} \subset \mathcal{A}_\theta$ is a nonzero ideal and let $a \in \mathfrak{I}$. In the expansion $a = \sum_{m,n} a_{m,n} U^m V^n$ we can WLOG assume that $a_{0,0} \neq 0$. The reason is that at least one coefficent $a_{m,n} \neq 0$ so by multiplying by a suitable monomial we get a nonzero element in $\mathfrak{I}$ with $a_{0,0} \neq 0$. Now, $\mathfrak{I}$ must be preserved by inner automorphisms so for any nonzero $a \in \mathfrak{I}$ we must have $\alpha_{z_1,z_2}(a) \in \mathfrak{I}$ if $(z_1, z_2) = (e^{2\pi i n_1 \theta}, e^{2\pi i n_2 \theta})$. But again

$$\{(z_1, z_2) \in U(1) \times U(1) | \alpha_{z_1,z_2}(a) \in \mathfrak{I}\} \tag{12.155}$$

must be a closed subset of $U(1) \times U(1)$ and hence must be all of $U(1) \times U(1)$. But then

$$\oint \oint \alpha_{z_1,z_2}(a) \frac{dz_1}{z_1} \frac{dz_2}{z_2} \in \mathfrak{I} \tag{12.156}$$

and hence $a_{0,0}\mathbf{1} \in \mathfrak{I}$. As we said, we can assume $a_{0,0} \neq 0$ and hence $\mathbf{1} \in \mathfrak{I}$ and hence $\mathfrak{I} = \mathcal{A}_\theta$. When $\theta \in \mathbb{Q}$ the algebra $\mathcal{A}_\theta$ is not simple (in the technical sense).

3. It is natural to ask when the algebras $\mathcal{A}_\theta$ are isomorphic. The main result is described in, for examples, [10, 43]:

Theorem

a.) Assume θ_1, θ_2 are irrational and $\theta_1, \theta_2 \in (0, \frac{1}{2})$. Then, if $\mathcal{A}_{\theta_1}$ is isomorphic to $\mathcal{A}_{\theta_2}$ it follows that $\theta_1 = \theta_2$.

b.) If θ is irrational with fractional part $\{\theta\}$ then let $\bar{\theta} = \{\theta\}$ or $\bar{\theta} = 1 - \{\theta\}$, depending on which is in $(0, 1/2)$. Then $\mathcal{A}_\theta$ is isomorphic to $\mathcal{A}_{\bar{\theta}}$.

Put differently, the moduli space of isomorphism classes of $\mathcal{A}_\theta$ for θ irrational is the space of orbits of the dihedral group generated by $\theta \to \theta + 1$ and $\theta \to -\theta$, acting on $\mathbb{R} - \mathbb{Q}$.

There is a very useful notion of "equivalence" of C^* algebras known as *Morita equivalence.* Roughly speaking, two algebras are "Morita equivalent" if their representation theories are "the same." (More technically: $\mathcal{A}_1$ and $\mathcal{A}_2$ are Morita equivalent if their is an equivalence of their categories of left-modules.) Then Rieffel's paper shows that

$\mathcal{A}_{\theta_1}$ and $\mathcal{A}_{\theta_2}$ are (strongly) Morita equivalent iff θ_1 is in the $GL(2,\mathbb{Z})$ orbit of θ_2 under the action

$$\theta \to \frac{a\theta + b}{c\theta + d}. \tag{12.157}$$

Remarks

1. The irrational rotation algebra was one of the main examples that led to A. Connes' formulation of noncommutative <u>geometry</u> as opposed to noncommutative <u>topology</u>. See [10, 11] and the book by Connes.
2. There is a very nice physical understanding of the above statement about Morita equivalence using T-duality in string theory, but it requires a little bit of background on the Moyal star product. See Section §12.8 below.

Exercise *$SL(2,\mathbb{Z})$ action on $\mathcal{A}_\theta$*

Show that the group $SL(2,\mathbb{Z})$ acts as a group of automorphisms of $\mathcal{A}_\theta$. Let

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \tag{12.158}$$

Show that

$$\begin{aligned} \alpha_A &: U \mapsto e^{-i\pi ac} U^a V^c \\ \alpha_A &: V \mapsto e^{-i\pi bd} U^b V^d \end{aligned} \tag{12.159}$$

is an automorphism of $\mathcal{A}_\theta$

♣Check it is a homomorphism into the group of automorphisms. ♣

Exercise *Noncommutative Binomial Theorem*

Suppose that we consider the noncommutative ring with generators u, v, q so that $uv = qvu$, while $qu = uq$ and $qv = vq$.

Show that [30]

$$(u+v)^n = \sum_{k=0}^{n} \binom{n}{k}_q u^{n-k} v^k \tag{12.161}$$

where

$$\binom{n}{k}_q := \frac{[n]_q}{[k]_q [n-k]_q} \tag{12.162}$$

[30] *Hint*: Prove that

$$\binom{n}{k}_q = \binom{n-1}{k-1}_q + q^k \binom{n-1}{k}_q \tag{12.160}$$

and imitate the usual inductive proof. Note that this shows that $\binom{n}{k}_q$ is in fact a polynomial in q.

and for a nonzero integer m

$$[m]_q := (1 - q^m)(1 - q^{m-1}) \cdots (1 - q) \tag{12.163}$$

while $[0]_q := 1$.

12.7.2 Realization In $\mathcal{B}(\mathcal{H})$

The Gelfand-Naimark construction assures us that there is a faithful representation of $\mathcal{A}_\theta$ on Hilbert space. We now describe a few such realizations, with comments on how they arise in physics.

One realization is in the quantum mechanics of a particle on the real line: $\mathcal{H} = L^2(\mathbb{R})$. Recall that in QM we introduce the Heisenberg algebra:

$$[\hat{q}, \hat{p}] = i\hbar \tag{12.164}$$

If $\psi \in L^2(\mathbb{R})$ we can represent the Heisenberg algebra:

$$\begin{aligned} (\hat{q} \cdot \psi)(q) &= q\psi(q) \\ (\hat{p} \cdot \psi)(q) &= -i\hbar \frac{d}{dq}\psi(q) \end{aligned} \tag{12.165}$$

Now, let us consider the unitary operators

$$\begin{aligned} U(\alpha) &:= \exp[i\alpha\hat{p}] \\ V(\alpha) &:= \exp[i\alpha\hat{q}] \end{aligned} \tag{12.166}$$

where $\alpha \in \mathbb{R}$. Of course $U(\alpha_1)U(\alpha_2) = U(\alpha_1 + \alpha_2)$ and similarly for $V(\alpha)$ so, separately, the group of operators $U(\alpha)$ is isomorphic to $\mathbb{R}$ as is the group of operators $V(\alpha)$. However, one can show in a number of ways that:

$$U(\alpha)V(\beta) = e^{i\hbar\alpha\beta}V(\beta)U(\alpha) \tag{12.167}$$

Physicists get an amazing amount of mileage out of this one equation.

One consequence of (12.167) is that the group generated by the operators $U(\alpha)$ and $V(\alpha)$ for $\alpha \in \mathbb{R}$, which we'll denote $\mathrm{Heis}(\mathbb{R} \times \mathbb{R})$ fits in a central extension:

$$1 \to U(1) \to \mathrm{Heis}(\mathbb{R} \times \mathbb{R}) \to \mathbb{R} \times \mathbb{R} \to 1 \tag{12.168}$$

By choosing any fixed α_0, β_0 so that $\hbar\alpha_0\beta_0 = 2\pi\theta$ we can restrict this extension to the subgroup generated by U_0, V_0. Then the group algebra of this subgroup is a morphism of C^* algebras to a subalgebra of $\mathcal{B}(\mathcal{H})$, with $\mathcal{H} = L^2(\mathbb{R})$.

In fact we can be more concrete and take the action on the Schwarz space $\mathcal{S}(\mathbb{R})$ of functions of rapid decrease to be:

$$\begin{aligned} (Uf)(q) &= f(q+1) \\ (Vf)(q) &= e^{2\pi\mathrm{i}\theta q} f(q) \end{aligned} \tag{12.169}$$

Then we extend to operators on $L^2(\mathbb{R})$ and take the norm-closure of the algebra generated by these.

Note, incidentally, that there is a pair of operators $\tilde{U}$, $\tilde{V}$ acting by

$$\begin{aligned}(\tilde{U}f)(q) &= f(q+\frac{1}{\theta})\\(\tilde{V}f)(q) &= e^{2\pi \mathrm{i} q}f(q)\end{aligned} \tag{12.170}$$

and these satisfy:

$$\tilde{U}\tilde{V} = e^{2\pi \mathrm{i}/\theta}\tilde{V}\tilde{U} \tag{12.171}$$

It is easy to check that both of $\tilde{U},\tilde{V}$ commute with both of U,V. In fact, for θ irrational, the commutant of the C^* algebra $\mathcal{A}_\theta$ in $\mathcal{B}(L^2(\mathbb{R}))$ is the algebra $\mathcal{A}_{1/\theta}$ generated by $\tilde{U},\tilde{V}$.

For a slightly different realization consider instead the Hilbert space $\mathcal{H} = L^2(S^1)$, and let us regard S^1 as $\mathbb{R}/\mathbb{Z}$ with parameter $t \sim t+1$. Then the algebra generated by $\tilde{V}$ above generates the commutative algebra of multiplication by suitably smooth functions, while the algebra generated by $\tilde{U}$ above generates irrational rotations of the circle, if θ is irrational. This is why $\mathcal{A}_\theta$ is called the irrational rotation algebra.

12.7.3 Electrons Confined To Two-Dimensions In A Magnetic Field

The irrational rotation algebra comes up in many different physical contexts. One significant example is in the system of an electron confined to a two-dimensional plane x_1, x_2. This can in fact be done in the laboratory with devices similar to transistors - so called heterostructures. Some electrons are confined to a slab of thickness ~ 100Å [22, 40].

We are going to stress the "magnetic translation group" below, so it is worthwhile recalling how translations are implemented in quantum mechanics. Usually in quantum mechanics translations in x_1, x_2 by a_1, a_2 are represented by

$$T_1(a_1) = e^{ia_1p_1/\hbar} \qquad\qquad T_2(a_2) = e^{ia_2p_2/\hbar} \tag{12.172}$$

with $T_1(a_1)T_2(a_2) = T_2(a_2)T_1(a_1)$, because $[p_1,p_2] = 0$.

Let us begin by reviewing the quantum mechanics of a charged particle of mass m that is charged with charge q under a $U(1)$ gauge field.

In general the path integral for a charged particle moving in Minkowski space in the presence of a $u(1)$ gauge potential A is:

$$\int [dx^\mu(s)]\exp\left[\frac{i}{\hbar}\int_{\mathcal{D}}\left\{mc\sqrt{-\frac{dx^\mu}{ds}\frac{dx_\mu}{ds}} + qA_\mu(x(s))\frac{dx^\mu}{ds}\right\}ds\right] \tag{12.173}$$

where the integral is over all maps from some one-dimensional domain $\mathcal{D}$ to Minkowski space $\mathbb{M}^{1,d-1}$ of dimension d. The second term in the action is the parallel transport due to the connection. In the adiabatic limit of slow motion around a loop γ the second term leads to the Aharonov-Bohm phase

$$AB(\gamma) := \exp[2\pi \mathrm{i} q\Phi/h], \tag{12.174}$$

where $\Phi = \oint_\gamma A$ is the enclosed flux. (For a charged particle moving in Minkowski space in the absence of magnetic monopoles Φ is well-defined. In the presence of singularities, or monopoles, or on spacetimes that are not simply connected only the phase is well-defined once the charge q is suitably quantized.) For an electron this can be written as

$$AB(\gamma) = \exp[2\pi \mathrm{i} \Phi/\Phi_0], \tag{12.175}$$

where $\Phi_0 = h/e$ is known as the *magnetic flux quantum.* [31]

More invariantly: The electron wavefunction is an L^2 section of a complex line bundle $\mathcal{L}$ over spacetime. The line bundle has a connnection ∇. Here we have taken the case of a trivialized line bundle, so sections are just complex-valued functions on spacetime. The connection on the associated principal $U(1)$ bundle over spacetime is $\nabla = d + \mathbf{A}$ where, in our case, $\mathbf{A}$ is a globally defined one-form valued in the Lie algebra of the structure group, $U(1)$. We identify this Lie algebra with the imaginary complex numbers $\mathfrak{u}(1) \cong \mathrm{i}\mathbb{R}$. Acting on sections of the associated line bundle $\mathbf{A}$ is in the representation of $U(1)$ of "charge one." That is, the defining representation. Hence we identify the "math" connection $\mathbf{A}$ with the "physics" gauge field A appearing in the above Hamiltonian by $\mathbf{A} = -\frac{\mathrm{i}e}{\hbar} A$. Note that our normalization of A absorbs the speed of light, compared to standard physics textbooks which use the gauge invariant momenta $p - eA/c$. Finally, here we have ignored electron spin for simplicity.

Now suppose that there is a magnetic field B perpendicular to a plane $\mathbb{R}^2$ with coordinates x_1, x_2, as in the quantum Hall effect. The Hamiltonian for a free electron in the presence of the magnetic field can be easily derived from (12.173) and is:

$$H = \frac{1}{2m}\left((p_1 - eA_1)^2 + (p_2 - eA_2)^2\right) \tag{12.176}$$

where e is the charge of the electron and m is the (effective) mass of the electron.

In two dimensions we have:

$$F_{12} = \partial_1 A_2 - \partial_2 A_1 = B \tag{12.177}$$

Now assume that B is <u>constant</u>. Then A_1, A_2 are affine-linear in the coordinates x_1, x_2. We can choose a gauge so that this can be put in the form:

$$H = \frac{1}{2m}\left((p_1 + \frac{eBx_2}{2})^2 + (p_2 - \frac{eBx_1}{2})^2\right) = \frac{1}{2m}(\tilde{p}_1^2 + \tilde{p}_2^2) \tag{12.178}$$

where the gauge invariant momenta are $\tilde{p}_i := p_i - eA_i$ are

$$\begin{aligned} \tilde{p}_1 &= p_1 + \frac{eB}{2} x_2 \\ \tilde{p}_2 &= p_2 - \frac{eB}{2} x_1 \end{aligned} \tag{12.179}$$

[31] One should be careful about a factor of two here since in superconductivity the condensing field has charge $2e$ and hence the official definition of the term "flux quantum" used, for example, by NIST is $\Phi_0 = h/2e$, half the value we use.

Note that p_1, p_2 *do not commute with* H. This is hardly surprising since H is no longer translation invariant. Moreover, $\tilde{p}_i$ do not commute with the Hamiltonian. Rather:

$$[\tilde{p}_1, \tilde{p}_2] = \mathrm{i}\hbar e B \tag{12.180}$$

These are, up to a constant, just the usual Heisenberg relations - but now for two "momenta"! Nevertheless, by the Stone-von Neumann theorem there is a unique representation up to unitary equivalence.

Equation (12.180) immediately leads us to a diagonalization of the Hamiltonian (12.178). One effective way to diagonalize the Hamiltonian is to introduce complex coordinates $z = x_1 + \mathrm{i}x_2$. Then define

$$\begin{aligned}
\tilde{p}_1 - \mathrm{i}\tilde{p}_2 &= -2\mathrm{i}\hbar(\partial_z - \beta\bar{z}) := -2\mathrm{i}\hbar A_z \\
\tilde{p}_1 + \mathrm{i}\tilde{p}_2 &= -2\mathrm{i}\hbar(\partial_{\bar{z}} + \beta z) := -2\mathrm{i}\hbar A_{\bar{z}} \\
\partial_z &= \frac{1}{2}(\partial_1 - \mathrm{i}\partial_2) \\
\partial_{\bar{z}} &= \frac{1}{2}(\partial_1 + \mathrm{i}\partial_2) \\
A_z &:= \partial_z - \beta\bar{z} \\
A_{\bar{z}} &:= \partial_{\bar{z}} + \beta z
\end{aligned} \tag{12.181}$$

where we defined

$$\beta := \frac{eB}{4\hbar}. \tag{12.182}$$

Note that $A_{\bar{z}} = -(A_z)^\dagger$ and

$$[A_z, A_{\bar{z}}] = 2\beta \tag{12.183}$$

Moreover, the Hamiltonian can be written in three equivalent ways:

$$\begin{aligned}
2mH &= \frac{1}{2}\left[(\tilde{p}_1 + \mathrm{i}\tilde{p}_2)(\tilde{p}_1 - \mathrm{i}\tilde{p}_2) + (\tilde{p}_1 - \mathrm{i}\tilde{p}_2)(\tilde{p}_1 + \mathrm{i}\tilde{p}_2)\right] \\
&= (\tilde{p}_1 + \mathrm{i}\tilde{p}_2)(\tilde{p}_1 - \mathrm{i}\tilde{p}_2) + \frac{1}{2}[(\tilde{p}_1 - \mathrm{i}\tilde{p}_2), (\tilde{p}_1 + \mathrm{i}\tilde{p}_2)] \\
&= (\tilde{p}_1 - \mathrm{i}\tilde{p}_2)(\tilde{p}_1 + \mathrm{i}\tilde{p}_2) + \frac{1}{2}[(\tilde{p}_1 + \mathrm{i}\tilde{p}_2), (\tilde{p}_1 - \mathrm{i}\tilde{p}_2)]
\end{aligned} \tag{12.184}$$

The solution for the spectrum depends on the sign of β:

If $\beta > 0$ *we use:*

$$2mH = 4\hbar^2 (A_{\bar{z}})^\dagger A_{\bar{z}} + \hbar e B \tag{12.185}$$

so the Hamiltonian is a sum of two positive semidefinite terms. The groundstates satisfy $A_{\bar{z}}\psi = 0$. The general solution to this equation is

$$\psi = f(z)\exp(-\beta z\bar{z}) \tag{12.186}$$

where $f(z)$ is an entire function such that ψ is square-normalizable. This would seem to imply an infinite ground-state degeneracy, and that would be unphysical. We will address this in a moment.

If $\beta < 0$ we should write instead:

$$2mH = 4\hbar^2 (A_z)^\dagger A_z - \hbar e B \tag{12.187}$$

and again the Hamiltonian is a sum of two positive semidefinite terms. The groundstates must solve the differential equation $A_z\psi = 0$ and this has the general solution

$$\psi = f(\bar{z})\exp(\beta z\bar{z}) \tag{12.188}$$

The two cases are related by a parity transformation. Since $[A_{\bar{z}}, (A_{\bar{z}})^\dagger] = 2\beta$ and $[A_z, (A_z)^\dagger] = -2\beta$ we immediately obtain that the spectrum of H is

$$\sigma(H) = \{(2N+1)\frac{|\hbar e B|}{2m} : N = 0, 1, 2, \dots \} \tag{12.189}$$

The eigenvalues of H can be, and often are, written as

$$(N + \frac{1}{2})\hbar\omega_c \tag{12.190}$$

where $\omega_c = eB/m$ is the classical frequency of an electron in a circular orbit in a magnetic field. The degenerate levels of energy eigenvalues are known as *Landau levels* after Lev Landau who first derived them in 1930. [32] The energy scale here is tiny: [33]

$$\frac{|\hbar e B|}{2m} = 5.80223 \times 10^{-5} eV \;\cdot\; \left(\frac{cB}{\text{Tesla}}\right) \tag{12.191}$$

where we used the standard mass of the electron $m = m_e$. Actually, it is important to note that in real materials the parameter m that one should use here is the *effective mass*, and this can be as small as $m = 0.07 m_e$ [40], thus increasing the energy by an order of magnitude.

♣Compare Zeeman energy with LL energy in 1 Tesla field. Idealized. Then real world with effective mass and effective g-factor. ♣

Of course, in nature the ground state won't be infinitely degenerate. In a finite-size system boundary conditions will lead to a finite number of states. We now estimate the number of possible states in the LLL in a finite size system. It is useful to introduce the *magnetic length*

$$\ell := \sqrt{\frac{\hbar}{eB}} = \frac{257\text{Å}}{\sqrt{cB/\text{Tesla}}} \tag{12.192}$$

so that

$$\beta = \frac{1}{4\ell^2} \tag{12.193}$$

[32]Landau in 1930, of course, did not know about transistors, much less $GaAs - AlAs$ hererostructures. He solved the problem of an electron in a constant 3-dimensional magnetic field. But the eigenstates just have a planewave for motion in the direction parallel to B.

[33]Recall that in our units cB has units of Tesla. The strength of the earth's magnetic field is about 30 microTesla, a refrigerator magnet is about 5 milliTesla, an NMR medical device uses a few Tesla, and the magnets at the LHC are 8 Tesla. The record on earth is 33.8 T and some stars produce magnetic fields on the order of 10^{11}T.

Suppose we have a droplet of 2d electrons of radius R. We should ask how many independent states we can fit into this droplet. Let us assume $\beta > 0$, for definiteness. Then a complete basis of L^2 wavefunctions on the plane of the form (12.186) is

$$\psi_n = z^n e^{-\beta|z|^2} \qquad\qquad n = 0, 1, 2, \dots \tag{12.194}$$

Recall that the probability distribution is proportional to $|\psi_n|^2$. It is just a function of $r = |z|$, and as a function of r it has a maximum at $r = r(n)$ where

$$r(n)^2 = \frac{n}{2\beta} = 2\ell^2 n \tag{12.195}$$

Therefore, we can get a rough estimate of the number of independent states in the LLL in a droplet of size R by setting

$$r(n_{\rm max}) = R \tag{12.196}$$

This gives

$$n_{\rm max} = 2\beta R^2 = \frac{R^2}{2\ell^2} = \frac{\Phi}{\Phi_0} \tag{12.197}$$

Here $\Phi = \pi R^2 B$ is the flux through the droplet. Meanwhile

$$\Phi_0 = h/e \tag{12.198}$$

is the *magnetic flux quantum* for a single electron. (See footnote above.)

Note that we obtain the higher Landau levels by acting with $(A_{\bar z})^\dagger$ or $(A_z)^\dagger$ depending on the sign of β. For polynomial or exponential $f(z)$ these just add one power of z or $\bar z$ to the prefactor. Therefore, for any fixed Landau level, in the large area limit, the same argument applies, and we find an equal number of possible states in each level.

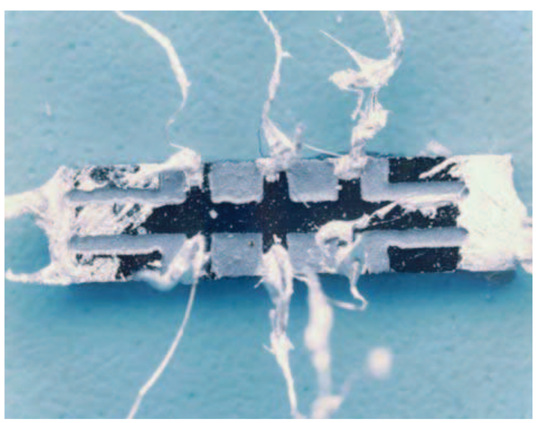


Figure 28: The sample used to discover the fractional quantum Hall effect. Courtesy of Ady Stern.

Remarks:

1. The numerical value for Φ_0 is about $4 \times 10^{-15} Weber$ and a $Weber$ is one Tesla times one square meter. So for a magnetic field of strength one Tesla we have

$$\frac{\Phi}{\Phi_0} = 2.5 \times 10^{14} \cdot \left(\frac{\text{Area}}{\text{meter}^2}\right) \tag{12.199}$$

In an area one square millimeter (a rather large scale for small devices) this works out to about 2.5×10^8, a large number. A picture of the sample used to discover the FQHE is shown in Figure 28.

2. One can usefully rederive this same formula in a different gauge. See the exercise below.

3. In a physical sample there is not one, but many electrons. The Hilbert space of the N electrons is the N^{th} antisymmetric product of $\mathcal{H}$, the Hilbert space of one electron. The groundstate is obtained by "filling" the lowest energy eigenstates compatible with the Pauli exclusion principle. For the highly degenerate LL case there are many ways to do this. In the case that we neglect interactions, for N electrons in the LLL the resulting wavefunction of the positions of N particles is

$$\Psi = const.\psi_0 \wedge \psi_1 \wedge \cdots \wedge \psi_N \tag{12.200}$$

where const. is a normalization constant. This can be thought of as a totally antisymmetric function of N positions, and by the Vandermonde formula it is:

$$\Psi(\vec{x}_1, \dots, \vec{x}_N) = const. \prod_{1 \le i < j \le N} (z_i - z_j) e^{-\sum_i |z_i|^2/4\ell^2} \tag{12.201}$$

4. In the multiparticle case we can define a notion of "filling fraction," always denoted ν:

$$\nu := \frac{\text{Density of electrons}}{\text{Density of fluxquanta}} \tag{12.202}$$

In our simple-minded setup this is just:

$$\nu = \frac{N}{\Phi/\Phi_0} = 1 \tag{12.203}$$

In the famous "fractional quantum Hall effect" the filling fraction ν, which can be measured from the Hall conductivity $\sigma_{xy} = \nu \frac{e^2}{h}$, turns out to be fractional. In this case the electron-electron interactions become important. Laughlin guessed a good approximation for <u>the</u> interacting groundstate. (At least for some cases.) Laughlin's guess is to take the many-body wavefunction to be of the form

$$\Psi = const. \prod_{1 \le i < j \le N} (z_i - z_j)^k e^{-\sum_i |z_i|^2/4\ell^2} \tag{12.204}$$

where k is a positive odd integer. In general, if we consider a many body state of the form:

$$\Psi = const.P(z_1, \dots, z_N) e^{-\sum_i |z_i|^2/4\ell^2} \tag{12.205}$$

where P is a translationally invariant odd polynomial of $z_1, \dots, z_N$ then we can compute the filling fraction as follows. Suppose the order of the polynomial (as a function, say, of z_1) is M. Then, reasoning as above, the state fills a circle of radius

$$R^2 = 2M\ell^2 \tag{12.206}$$

But then

$$\nu = \frac{N}{\pi R^2 B/\Phi_0} = \frac{N}{M} \tag{12.207}$$

So, for example, for the above Laughlin wavefunction

$$\nu = \frac{N}{kN-1} \to \frac{1}{k} \tag{12.208}$$

The first observed FQHE state had $\nu = 1/3$. Subsequently, many other rational values of ν have been observed. See [22, 40].

♣Explain more. Give more references. ♣

Exercise *Landau Levels In Landau Gauge*

a.) Show that by a gauge transformation we can take the electron Hamiltonian to be

$$H = \frac{1}{2m}(p_x^2 + (p_y - eBx)^2) \tag{12.209}$$

b.) Show that if $\psi(x,y) = e^{\mathrm{i}ky/\hbar}\psi_k(x)$ then the eigenvalue equation is that of a harmonic oscillator with the center of the potential at $x_0 = -k\ell^2$.

c.) Suppose the electrons are on a cylinder of radius R and length L where both R and L are very large compared to ℓ. Show that the number of groundstates is Φ/Φ_0.

d.) Note that the eigenfunctions of the form $\psi(x,y) = e^{\mathrm{i}ky/\hbar}\psi_k(x)$ are well-localized in x, but not in y. We could of course choose a different Landau gauge in which the eigenfunctions are well-localized in y but not in x. Why are these choices compatible?

Exercise *Coherent States*

Assume $\beta > 0$. Consider the "coherent state wavefunctions":

$$\psi_{\bar{v}} = \exp[-\beta z\bar{z} + \bar{v}z] \tag{12.210}$$

where $\bar{v}$ is any complex number.

a.) Show that these are L^2-normalizable groundstate wavefunctions.

b.) Show that they are an overcomplete basis.

c.) Express the wavefunctions ψ_n above in terms of the $\psi_{\bar{v}}$.

d.) Show that these are "minimum uncertainty" wavefunctions.

♣Should go after we mention "space is noncommutative" ♣

12.7.4 Magnetic Translation Group

As we stressed above, ordinary translations do not commute with the Hamiltonian. Nevertheless we can define the *magnetic translation operators*:

$$\pi_1 := p_1 - \frac{eBx_2}{2} \qquad\qquad \pi_2 := p_2 + \frac{eBx_1}{2} \tag{12.211}$$

Compare this carefully with the definitions of $\tilde{p}_i$. Note the relative signs! These operators satisfy $[\pi_i, \tilde{p}_j] = 0$. In particular they are translation-like operators that commute with the Hamiltonian: $[\pi_i, H] = 0$. Hence the name.

While they are called "translation operators" note that they do not commute:

$$[\pi_1, \pi_2] = -i\hbar eB \tag{12.212}$$

The "magnetic translation group" is generated by the operators

$$\begin{aligned} U(a_1) &= \exp[ia_1\pi_1/\hbar] \\ V(a_2) &= \exp[ia_2\pi_2/\hbar] \end{aligned} \tag{12.213}$$

The operators $U(a_1), V(a_2)$ satisfy the relations:

$$U(a_1)V(a_2) = \exp[\mathrm{i}eBa_1a_2/\hbar]V(a_2)U(a_1) \tag{12.214}$$

Imagine now that we have some rectangular lattice in the plane $\mathbb{R}^2$:

$$\Lambda = \{n_1a_1\hat{x} + n_2a_2\hat{y} | n_1, n_2 \in \mathbb{Z}\} \tag{12.215}$$

and - for some reason - we only consider translations by lattice vectors then the algebra generated by $U(a_1), V(a_2)$ is again $\mathcal{A}_\theta$ for $\theta = eBa_1a_2/h = \Phi/\Phi_0$ where $\Phi = a_1a_2B$ is the flux through the unit cell.

Although we could also obtain an irrational rotation algebra by exponentiating $\tilde{p}_i$, $i = 1, 2$, this is less relevant to the physics because such operators do not commute with the Hamiltonian.

In the above realization there is a major conceptual change in the physical interpretation of the algebra $\mathcal{A}_\theta$. In our first realization U, V were translation operators on *phase space* with coordinates (q, p). For the case of the electron in the uniform magnetic field $U(a_1), V(a_2)$ are "translation operators" in the physical space in which the electron moves.

Put another way, let Π be the projection operator from the Hilbert space $L^2(\mathbb{R}^2)$ to the LLL. For definiteness assume that $\beta > 0$. Note that

$$\Pi A_{\bar{z}} \Pi = 0 \tag{12.216}$$

But this means that in the LLL we can replace z by $-\frac{1}{\beta}\partial_{\bar{z}}$. But that in turn means that in the LLL

$$[\Pi z \Pi, \Pi \bar{z} \Pi] = -\frac{1}{\beta} \tag{12.217}$$

In other words

$$[\Pi x_1 \Pi, \Pi x_2 \Pi] = -2\mathrm{i}\frac{\hbar}{eB} = -2\mathrm{i}\ell^2 \tag{12.218}$$

and in this sense "space is noncommutative." We will come back to that in section ****

♣Uncertainty principle applied to this gives nice explanation of the filling factor. Explain this. ♣

12.7.5 The Algebra $\mathcal{A}_\theta$ For θ Rational

Notice that when $\theta = p/q$ is a rational number (with p, q positive and relatively prime) then from (12.144) it follows that the subalgebra generated by U^q and V^q is central in $\mathcal{A}_\theta$. Therefore, in an irreducible representation they will be represented by phases. Therefore, if there is a cyclic vector it will generate a q-dimensional vector space, so the representations are q-dimensional. As an example of such a representation we can consider the vector space of complex-valued functions on the cyclic group $\mathbb{Z}/q\mathbb{Z}$:

$$\begin{aligned} (Uf)(\bar{j}) &= f(\bar{j}+1) \\ (Vf)(\bar{j}) &= e^{2\pi i\theta\bar{j}} f(\bar{j}) \end{aligned} \tag{12.219}$$

where $\bar{j} \in \mathbb{Z}/q\mathbb{Z}$. We can choose a basis by using a fundamental domain $\bar{j} \in \{0, 1, \dots, q-1\}$ and delta-functions $\delta_{\bar{j}}$. Then, relative to this basis, U and V are represented by $q \times q$ matrices known as the famous "clock and shift operators"

$$\mathbf{u} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{12.220}$$

$$\mathbf{v} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \omega & 0 & 0 & \cdots & 0 \\ 0 & 0 & \omega^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \omega^{q-1} \end{pmatrix} \tag{12.221}$$

where $\omega = e^{2\pi i \frac{p}{q}}$. One should check directly that indeed

$$\mathbf{u}\mathbf{v} = \omega \mathbf{v}\mathbf{u} \tag{12.222}$$

Moreover

$$\mathbf{u}^q = \mathbf{v}^q = \mathbf{1}_{q\times q}. \tag{12.223}$$

Using the standard Hermitian structure on $\mathbb{C}^q$ this is a unitary representation.

Now, given a representation ρ of any algebra $\mathcal{A}$ and an automorphism α of $\mathcal{A}$ we can always get another (possibly the same) representation by considering

$$\rho \to \rho \circ \alpha := \rho_\alpha \tag{12.224}$$

Now recall that $\mathcal{A}_\theta$ has a canonical group of automorphisms α_{z_1,z_2}, isomorphic to $U(1) \times U(1)$. If we twist by this automorphism then the representation $\mathbf{u} = \rho(U)$ and $\mathbf{v} = \rho(V)$ becomes:

$$\begin{aligned} \rho_{z_1,z_2}(U) &= z_1 \mathbf{u} \\ \rho_{z_1,z_2}(V) &= z_2 \mathbf{v} \end{aligned} \tag{12.225}$$

Theorem: If $\theta = p/q$ is rational then the most general irreducible representation of $\mathcal{A}_\theta$ is of the form ρ_{z_1,z_2} for $(z_1, z_2) \in U(1) \times U(1)$. Moreover two such representations are isomorphic iff there are integers n_1, n_2 such that

$$z_1' = \omega^{n_1} z_1 \qquad\qquad z_2' = \omega^{n_2} z_2 \tag{12.226}$$

In other words, the irreducible representations can be identified with a torus, and this torus can be viewed as the quotient of the torus of automorphisms $U(1) \times U(1)$ by $\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$.
Proof:

1. As remarked above, U^q and V^q generate an abelian subalgebra and should be represented by phases in an irrep. Choosing a q^{th} root of these phases gives a representation of the form ρ_{z_1,z_2}.

2. Now, since α_{z_1,z_2} form a group of automorphisms is suffices to consider the representations equivalent to the original representation ρ. Note that conjugation by $S = U^{\ell_1} V^{\ell_2}$ transforms

$$\begin{aligned} \mathbf{u} &\to e^{-2\pi i \theta \ell_1} \mathbf{u} \\ \mathbf{v} &\to e^{-2\pi i \theta \ell_2} \mathbf{v} \end{aligned} \tag{12.227}$$

so in (12.226) we can always find a suitable ℓ_1, ℓ_2 to render the two representations equivalent.

3. These are the only isomorphisms between representations, because if $z_1 = e^{2\pi i \varphi_1}$ and $z_2 = e^{2\pi i \varphi_2}$ with $0 \le \varphi_1, \varphi_2 < \frac{1}{q}$ then $\rho_{z_1,z_2}(U)$ and/or $\rho_{z_1,z_2}(V)$ has an inequivalent spectrum from $\mathbf{u}$ and $\mathbf{v}$. ♠

Remarks

1. The set of matrices $\{\omega^k \mathbf{u}^n \mathbf{v}^m\}$ forms a group. It is isomorphic to a finite Heisenberg group. The finite Heisenberg groups can be defined as an extension of $\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$ by $\mathbb{Z}/q$ using the cocycle

$$f\left((\omega_1^s, \omega_2^t), (\omega_1^{s'}, \omega_2^{t'})\right) := \omega_3^{st'} \tag{12.228}$$

Where $\omega_1, \omega_2, \omega_3$ are all just ω but the subscript distinguishes the different conceptual roles they play. The resulting group $\mathrm{Heis}(\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z})$ sits in an exact sequence:

$$1 \to \mathbb{Z}/q\mathbb{Z} \to \mathrm{Heis}(\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}) \to \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z} \to 1 \tag{12.229}$$

This is a finite group of order q^3. The irreducible representations are all of dimension q and hence there should be q distinct irreps. These are obtained by replacing $\omega \to \omega^k$ in the above clock and shift operators.

2. Let W be the complex vector space of functions $f : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$. It is isomorphic to $\mathbb{C}^q$. The *finite Fourier transform* is the unitary transformation $\mathcal{F} : W \to W$, defined by

$$\mathcal{F}(f)(\bar{j}) := \frac{1}{\sqrt{q}} \sum_{\bar{k}} e^{2\pi \mathrm{i} \frac{\bar{j}\bar{k}}{q}} f(\bar{k}) \tag{12.230}$$

Note that $\mathcal{F}$ is unitary and $\mathcal{F}^4 = 1$. Then, for $p = 1$ we have ♣check sign ♣

$$\mathcal{F}\mathbf{u}\mathcal{F}^{-1} = \mathbf{v}^{-1} \tag{12.231}$$

3. Now consider the algebra generated by the matrices $\mathbf{u}, \mathbf{v}$. It is a subalgebra of the full matrix algebra $M_q(\mathbb{C})$, and we claim it is the full algebra $M_q(\mathbb{C})$. One way to show this is to note that the matrices $\mathbf{u}^\ell \mathbf{v}^s$ with $1 \le \ell, s \le q$ are linearly independent. Suppose that there are coefficients $a_{\ell,s}$ such that

$$0 = \sum_{s,\ell=1}^{q} a_{\ell,s} \mathbf{u}^\ell \mathbf{v}^s = \sum_{\ell=1}^{q} \mathbf{u}^\ell D(z_\ell) \tag{12.232}$$

where $D(z_\ell)$ is a diagonal matrix with diagonal entries $z_\ell^{(0)}, z_\ell^{(1)}, \dots, z_\ell^{(q-1)}$ with $z_\ell^{(k)} = \sum_s a_{\ell,s}\omega^{ks}$. Now, if this sum is zero then it is zero acting on the elementary basis e_i, $0 \le i \le q-1$. Observe that this means $z_\ell^{(i)} = 0$ and hence $a_{\ell,s} = 0$. Actually, this argument shows more: Call the "ℓ^{th} shifted diagonal the nonzero entries of $\mathbf{u}^\ell$. Then the ℓ^{th} shifted diagonal of $\sum_{s,\ell=1}^{q} a_{\ell,s}\mathbf{u}^\ell \mathbf{v}^s$ is up to a factor of $\sqrt{q}$ the finite Fourier transform of the functions $a_{\ell,s}$, as a function of s. ♣Say this more precisely so that it is useful. ♣

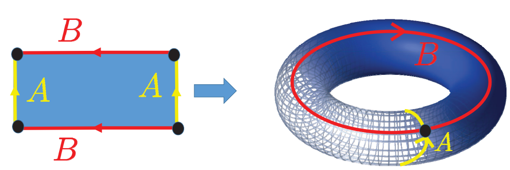


Figure 29: Left: A fundamental domain for $\mathbb{R}^2/\Lambda$. Right: The identification gives a torus.

12.7.6 Two-Dimensional Electrons On A Torus In A Magnetic Field

♣ALL EQUATIONS BELOW NEED TO BE RECHECKED ♣

Another, and mathematically very interesting, way of putting the electron in a finite system is to attempt to impose periodic boundary conditions. That is, we imagine an electron confined to a torus with a uniform magnetic field B perpendicular to the torus. For simplicity we will consider the torus to be obtained by identifying $x_1 \sim x_1 + a_1$ and $x_2 \sim x_2 + a_2$. That is, it is obtained from the quotient $\mathbb{R}^2/\mathbb{Z}^2$.

A fundamental domain for the action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ can be pictured as a rectangle as in Figure 29. The torus is obtained by identifying opposite sides as in the figure.

We immediately encounter a problem: The Hamiltonian (12.178) appears to be nonsense when x_1, x_2 are to have these periodic identifications. The same will be true of (12.176) no matter what gauge we choose, since the components of A must be linear in the x_i. But it is not nonsense. We must change the geometrical interpretation of the quantities a little bit.

Now note that the Aharonov-Bohm phase for transporting an electron around a loop going around the edge of the rectangle is

$$\exp[2\pi \mathrm{i}\Phi/\Phi_0] = \exp[2\pi \mathrm{i} a_1 a_2 B/(h/e)] \tag{12.233}$$

On the other hand, the phase around this loop must be one: This loops is of the form $ABA^{-1}B^{-1}$, and the holonomy associated the parallel transport along A, B is abelian (i.e. just a phase) so that holonomy along A cancels the holonomy along A^{-1} and similarly for B.

♣Very poor notation. Change A, B for the cycles. ♣

Since the phase (12.233) must be one we obtain a quantization condition:

$$\frac{\Phi}{\Phi_0} = k \qquad\qquad k \in \mathbb{Z}. \tag{12.234}$$

Equivalently we can write

$$\beta a_1 a_2 = \frac{\pi k}{2} \tag{12.235}$$

Note that if we write

$$A = \beta(z d\bar{z} - \bar{z} dz) \tag{12.236}$$

then $F = dA = 2\beta dz \wedge d\bar{z}$ makes sense on the torus and we find:

$$\int_T \frac{F}{2\pi \mathrm{i}} = \frac{(2\beta)(-2\mathrm{i})(a_1 a_2)}{2\pi \mathrm{i}} = -k \tag{12.237}$$

Now, let us note that with this quantization the magnetic translation operators $U(a_1)$ and $V(a_2)$ commute. They also commute with the Hamiltonian, so that they can be simultaneously diagonalized. Call $x := x_1$ and $y := x_2$. Then

$$U(a_1) = \exp\left[a_1 \partial_x - 2\mathrm{i} a_1 \beta y\right] \tag{12.238}$$

so

$$U(a_1)\psi = \psi \qquad \Rightarrow \qquad \psi(x + n_1 a_1, y) = e^{2\mathrm{i} n_1 a_1 \beta y}\psi(x, y) \qquad \forall n_1 \in \mathbb{Z} \tag{12.239}$$

Similarly:

$$V(a_2) = \exp\left[a_2 \partial_y + 2\mathrm{i} a_2 \beta x\right] \tag{12.240}$$

and hence

$$V(a_2)\psi = \psi \qquad \Rightarrow \qquad \psi(x, y + n_2 a_2) = e^{-2\mathrm{i} n_2 a_2 \beta x}\psi(x,y) \qquad \forall n_2 \in \mathbb{Z} \tag{12.241}$$

Altogether we have, for all integers n_1, n_2:

$$\begin{aligned}\psi(x + n_1 a_1, y + n_2 a_2) &= e^{\mathrm{i}\pi k n_1 n_2} e^{\mathrm{i}2\beta(n_1 a_1 y - n_2 a_2 x)}\psi(x,y) \\ &= e^{\mathrm{i}\pi k n_1 n_2} e^{\mathrm{i}2\beta\omega(\lambda,\vec{x})}\psi(x,y)\end{aligned} \tag{12.242}$$

where in the second line $\lambda = n_1 a_1 \hat{x} + n_2 a_2 \hat{y} \in \Lambda$ and ω is the symplectic form on $\mathbb{R}^2$ with $\omega(\hat{x}, \hat{y}) = 1$. That is:

$$\omega(\vec{x}, \vec{y}) = x_1 y_2 - x_2 y_1 \tag{12.243}$$

Note that (12.242) is consistent because $U(a_1)$ and $V(a_2)$ commute.

We can view (12.242) as the statement that we must make a gauge transformation when translating by lattice vectors. If we suitably transform the gauge field A in (12.176) and (12.178) then the Hamiltonian H makes perfectly good sense for an electron on the torus.

In more geometrical terms we can say the following: The meaning of equation (12.242) is that ψ is not a function but a section of a complex line bundle L over the torus. Roughly speaking, the prefactor on the right defines a set of transition functions $g(x, y)$. The gauge field transforms by

$$(d + \mathbf{A}') = g^{-1}(d + \mathbf{A})g \tag{12.244}$$

and $d + \mathbf{A}$ defines a connection on the line bundle: It is a prescription for defining a notion of parallel transport along paths in the torus. Once A and ψ are interpreted this way, the Hamiltonian makes perfect sense. Viewed this way, the generalization to electrons on an arbitrary compact Riemannian two-manifold is straightforward.

This is a key conceptual step: The geometrical nature of the "wavefunction" has changed: It is more properly regarded as a section of a complex line bundle.

To be a little more precise, and to give a preview of some things to be discussed later we will give a nice description of the complex line bundle L and its connection as an equivariant bundle with connection over the homogeneous space $T = X/G$ with $X = \mathbb{R}^2$ and $G = \mathbb{Z}^2$ as follows.

Consider, quite generally a right G-space and a representation W of G. The *equivariant vector bundle over* X/G is the quotient

$$E := (X \times W)/G \tag{12.245}$$

and we take the quotient using the right G-action on $X \times W$:

$$(x, \psi) \sim (x \cdot g, \rho(g^{-1})\psi) \tag{12.246}$$

In our example $X = \mathbb{R}^2$, $G = \mathbb{Z}^2$ and where W is a one-dimensional complex representation of G (so $W \cong \mathbb{C}$, as a complex vector space). If we let σ_1, σ_2 be generators of the first and second summands of $\mathbb{Z} \oplus \mathbb{Z}$ then the identifications are generated by the actions

$$\sigma_1 : (\vec{x}, \psi) \mapsto (\vec{x} + a_1 \hat{x}, e^{\mathrm{i}2\beta a_1 y}\psi) \tag{12.247}$$

$$\sigma_2 : (\vec{x}, \psi) \mapsto (\vec{x} + a_2\hat{y}, e^{-\mathrm{i}2\beta a_2 x}\psi) \tag{12.248}$$

or, in general the $\mathbb{Z}^2$ action is:

$$\vec{n} : (\vec{x}, \psi) \to (\vec{x} + \vec{R}, e^{2\mathrm{i}\beta\omega(\vec{R},\vec{x})}\psi) \tag{12.249}$$

where ω is the standard symplectic form on $\mathbb{R}^2$.

Returning to the general situation, the quotient space E is the *total space of a complex vector bundle over* $B = X/G$. Note that there is a projection map

$$\pi : E \to X/G \tag{12.250}$$

given by $\pi([(x,\psi)]) := [x]$. The fiber of the map π, i.e. the inverse image $\pi^{-1}([x])$ has a natural vector space structure and, as a vector space is isomorphic to W. A right-inverse s to π,

$$s : X/G \to W \tag{12.251}$$

that is, such that $\pi(s([x])) = [x]$ is called a *section* of the bundle. Sections always exist, by the axiom of choice. But the existence of sections with special properties is not guaranteed. We say a section is nonzero at $[x]$ if it is of the form $[(x,\psi)]$ with ψ a nonzero vector in W. When W is a one-dimensional vector space the bundle E is called a *line bundle*. In this case a continuous nowhere-vanishing section allows us to define a "bundle isomorphism" of E with the trivial bundle $X/G \times W$. There can be topological obstructions to the existence of such continuous nowhere-vanishing sections.

In general to give a section of E is to give an *equivariant function*

$$\psi : X \to W \tag{12.252}$$

That is, one which satisfies

$$\psi(xg) = \rho(g^{-1})\psi(x) \tag{12.253}$$

To see this just note that a section must be of the form $s([x]) = [(x, \psi(x))]$ for some association $x \to \psi(x)$. However, on the one hand,

$$\begin{aligned} s([x]) &= [(x, \psi(x))] \\ &= [(xg, \rho(g^{-1})\psi(x))] \end{aligned} \tag{12.254}$$

but on the other hand,

$$\begin{aligned} s([x]) &= s([xg]) \\ &= [(xg, \psi(xg))] \end{aligned} \tag{12.255}$$

so we must have (12.253). In our example, this reproduces the condition (12.242).

Now consider the *connection*. In general, a connection is a rule for "lifting paths" from the base $\gamma : [0,1] \to B = X/G$ to paths in the total space $\tilde{\gamma} : [0,1] \to L$.

Quite generally, if $\pi : E \to B$ is a fiber bundle then a *connection* is a path-lifting rule so that:

1. Given $\gamma : [0,1] \to B$ and a choice of lift $\tilde{\gamma}(0) \in \pi^{-1}(\gamma(0))$, that is, a choice of point in the fiber above $\gamma(0)$ there is a unique path $\tilde{\gamma} : [0,1] \to E$ such that it is a "lift," meaning:
$$\pi \circ \tilde{\gamma} = \gamma \tag{12.256}$$
2. The map $\pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(1))$ is compatible with the structure group. (For vector bundles this means that the map is a linear transformation and that the action of $GL(n,\mathbb{R})$ on the fibers $\pi^{-1}(\gamma(1))$ and $\pi^{-1}(\gamma(0))$ are related by conjugation.)
3. Moreover, the path satisfies a nice composition property: If γ_1, γ_2 are composable paths in X/G then let $\gamma_2 * \gamma_1$ denote the path $[0,1] \to B$ by running first γ_1 and then γ_2. Then if we take the initial lift of $\tilde{\gamma}_2$ to be $\tilde{\gamma}_1(1)$ the unique lifted path of $\gamma_2 * \gamma_1$ starting at $\tilde{\gamma}_1(0)$ is just $\tilde{\gamma}_2 * \tilde{\gamma}_1$.

We have phrased things this way so that we have in fact given the definition of a connection on an arbitrary fiber bundle $\pi : E \to B$.

In our special example we have a line bundle over a torus. Given $\gamma : [0,1] \to \mathbb{R}^2/\mathbb{Z}^2$ we can first lift it to a path $\hat{\gamma} : [0,1] \to \mathbb{R}^2$ (there is a unique connection on $\pi : \mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ since this is a covering by a discrete group). Then, if $\hat{\gamma}(0) = \vec{x}_0$ and we choose $\tilde{\gamma}(0) = [\vec{x}_0, \psi_0]$. Then the lifted path is
$$\tilde{\gamma}(t) = [(\hat{\gamma}(t), e^{\mathrm{i}\int_{\hat{\gamma}_t} A}\psi_0)] \tag{12.257}$$

♣Check sign ♣

Remarks

1. It can be shown that the data of a smooth connection is equivalent to giving a collection of one-forms, valued in endomorphisms of the fiber, defined on open sets in B where the bundle can be trivialized. In physics these are the "gauge fields."
2. In general, given a complex vector bundle $\pi : E \to B$, one can associate a set of integral cohomology classes $c_i(E) \in H^{2i}(B;\mathbb{Z})$ which measure - to some extent - the degree to which E is "twisted." For example if they are nonzero then certain fields of linearly independent sections do not exist. In our case the only possible Chern class is $c_1(L) \in H^2(T;\mathbb{Z})$. The image of this class in DeRham cohomology has a representative given by $F/(2\pi\mathrm{i})$, where F is the curvature of any connection on L.

Now let us actually construct explicit wavefunctions (12.242) that are in the ground-state of the Hamiltonian. These will be the analog of LLL wavefunctions on the torus. To do this we step back and consider a more general problem:

Suppose a group G has a right-action on a space X. Suppose that we have what is known as a *cocycle*, [34] namely, a function $\xi : X \times G \to \mathbb{C}^*$ so that
$$\xi(x; g_1)\xi(x \cdot g_1; g_2) = \xi(x; g_1 g_2) \tag{12.258}$$

[34] This is the defining equation for a cocycle in $H^1_G(X;\mathbb{C}^*)$.

Then we can construct a left-action of G on the function space $\mathrm{Map}(X,\mathbb{C})$:

$$(g\cdot f)(x) := \xi(x;g)f(x\cdot g) \tag{12.259}$$

We typically like to find "automorphic functions" which are fixed points of this action:

$$g\cdot f = f \tag{12.260}$$

In some cases we can construct such functions by an averaging procedure. Indeed, suppose G is discrete and h is an arbitrary function. Then form the average:

$$f(x) := \sum_{g\in G}\xi(x;g)h(x\cdot g) \tag{12.261}$$

A short computation shows that - formally - the function is indeed invariant: $g\cdot f = f$. Of course if G is infinite one must check that the sum converges (and this can be quite nontrivial). If G is continuous and has a left-invariant Haar measure then the same construction can be used.

Now let us apply this to our situation, where $X=\mathbb{R}^2$, $G=\mathbb{Z}^2$ and equation (12.242) tells us to take

$$\xi(\vec{x};\vec{n}) := e^{-\mathrm{i}\pi k n_1 n_2}e^{-2\mathrm{i}\beta(n_1a_1y-n_2a_2x)} \tag{12.262}$$

We can apply the averaging procedure to the coherent state wavefunctions in the LLL:

$$h(\vec{x}) = \exp[-\beta z\bar{z}+\bar{v}z] \tag{12.263}$$

where $\bar{v}$ is a complex number. Here we are assuming that $\beta>0$ and hence $k>0$, so the series is convergent. If $\beta<0$ and hence $k<0$ then we change the "seed coherent state" to

$$h(\vec{x}) = \exp[\beta z\bar{z}+v\bar{z}] \tag{12.264}$$

In either case, the series is absolutely convergent.

After a little bit of algebra the series can be written (taking the case $\beta>0$ from now on):

$$\psi_{\bar{v}} = e^{-\beta|z|^2+\bar{v}z}\sum_{\omega\in\Lambda}e^{-\beta|\omega|^2+\omega\bar{v}-2\beta\bar{\omega}z}e^{-\mathrm{i}\pi k n_1 n_2} \tag{12.265}$$

where the sum is over vectors $\omega\in\Lambda$ where $\Lambda\subset\mathbb{C}$ is a lattice and

$$\omega = n_1a_1+\mathrm{i}n_2a_2 \tag{12.266}$$

with $n_1,n_2\in\mathbb{Z}$. Note that the factor $\xi(\vec{x};\vec{n})$ has made the averaged wavefunction a holmorphic function of z up to the overall factor of $\exp[-\beta z\bar{z}]$.

Just as on the plane, by varying $\bar{v}$ we obtain an overcomplete set of groundstates, and the linear span of these will be the full space of groundstates on the torus. It is not immediately obvious from (12.265) what the dimension of the space is. In order to find that we should perform a Poisson resummation on the sum over n_2. After a little algebra the sum can be written very elegantly in the form

$$\psi_{\bar{v}} = ce^{-\beta|z|^2}\sum_{\mu=1}^{k}\Psi_\mu^h(z)\tilde{\Psi}_\mu^h(\bar{v}) \tag{12.267}$$

where c is a constant, $\Psi^h_\mu(z)$ are k linearly independent holomorphic functions of z (and $\tilde{\Psi}^h_\mu(\bar{v})$ are likewise k linearly independent holomorphic functions of $\bar{v}$). It follows that the space of groundstates is k-dimensional.

It is useful and interesting to derive explicit formulae for Ψ_μ and $\tilde{\Psi}_\mu$. The sum (12.265) fits in a very general story explained in Appendix A. The resulting wavefunctions can be expressed in terms of level $\kappa = k/2$ theta functions. We met them above when discussing the characters in the $SU(2)_\kappa$ WZW model.

Let us collect a few facts about theta functions: Recall that

$$\Theta_{\mu,\kappa}(z,\tau) := \sum_{n\in\mathbb{Z}} q^{\kappa(n+\mu/(2\kappa))^2} y^{(\mu+2\kappa n)} = \sum_{\ell=\mu \mathrm{mod} 2\kappa} q^{\ell^2/(4\kappa)} y^{\ell} \tag{12.268}$$

with $q = e^{2\pi i\tau}$ and $y = e^{2\pi i z}$. Here μ is an integer and κ is a positive half-integer (i.e. in $\frac{1}{2}\mathbb{Z}_+$). Note that if we shift $\mu \to \mu + 2\kappa s$, where s is any integer, then $\Theta_{\mu,\kappa}$ is unchanged. Often people take μ to be in the fundamental domain $-\kappa < \mu \leq \kappa$, but one should generally regard μ as an element of $\mathbb{Z}/2\kappa\mathbb{Z}$. As functions of z these functions are doubly-quasiperiodic:

$$\begin{aligned} \Theta_{\mu,\kappa}(z+\nu,\tau) &= \Theta_{\mu,\kappa}(z,\tau) \\ \Theta_{\mu,\kappa}(z+\nu\tau,\tau) &= e^{-2\pi i\kappa\nu^2\tau - 4\pi i\kappa\nu z}\Theta_{\mu,\kappa}(z,\tau) \end{aligned} \tag{12.269}$$

Here ν is any integer. Note that the theta functions transform the same way for all $\mu \in \mathbb{Z}/2\kappa\mathbb{Z}$. We will explain more about the geometrical meaning of these theta functions below.

We now can rewrite (12.267) as

$$\bar{\psi}_v = \sqrt{\frac{2}{k}\mathrm{Im}\tau}\sum_{\mu=1}^{k}\Psi_\mu(z,\bar{z})\tilde{\Psi}^h_\mu(\bar{v}) \tag{12.270}$$

$$\Psi_\mu(z,\bar{z}) = e^{-\frac{\pi k}{2}\frac{|z|^2+z^2}{a_1 a_2}}\Theta_{\mu,k/2}(\delta,\tau) \tag{12.271}$$

$$\tilde{\Psi}_\mu(\bar{v}) = e^{-\frac{a_1 a_2}{2\pi k}\bar{v}^2}\Theta_{-\mu,k/2}(\tilde{\delta},\tilde{\tau}) \tag{12.272}$$

with

$$\delta = iz/a_2 \qquad \tilde{\delta} = -i\bar{v}a_2/(\pi k) \qquad \tau = i\frac{a_1}{a_2} \qquad \tilde{\tau} = -\bar{\tau} \tag{12.273}$$

Now, we know that $U(a_1)$ and $V(a_2)$ act as the identity transformation on the groundstate wavefunctions. However, it turns out that the translations by the "k-torsion points" act nicely on the wavefunctions. [35] A short computation shows that

$$\begin{aligned} U(\frac{a_1}{k})\Psi_\mu &= \Psi_{\mu+1} \\ V(\frac{a_2}{k})\Psi_\mu &= e^{2\pi i\frac{\mu}{k}}\Psi_\mu \end{aligned} \tag{12.274}$$

[35] Of course $\mathbb{R}^2/\mathbb{Z}_2$ is itself an abelian group. k-torsion points are elements of this group whose k^{th} is the identity element.

Note that we consider $\mu \in \mathbb{Z}/k\mathbb{Z}$ in these formulae. Thus, the magnetic translations by the k-torsion points of the torus acts on the space of groundstates to give the irreducible representation of the finite Heisenberg group.

Remarks

1. Let us conclude with some remarks on the geometrical interpretation of the level κ theta functions $\Theta_{\mu,\kappa}(z,\tau)$. The torus $\mathbb{R}^2/\Lambda$ inherits a natural complex structure. Indeed
$$E_\tau = \mathbb{C}/(\mathbb{Z}\oplus\tau\mathbb{Z}) \tag{12.275}$$
is naturally a complex manifold. We can therefore consider *holomorphic bundles* over E_τ. The transformation equations (12.269) can be viewed as defining a holomorphic line bundle $\mathcal{L}$ over E_τ. The basic case is $\kappa = 1/2$. There is then only one theta function:
$$\Theta(z,\tau) = \sum_{n\in\mathbb{Z}} e^{\mathrm{i}\pi\tau n^2 + 2\pi \mathrm{i} n z} \tag{12.276}$$
As before, one construction of the holomorphic line bundle is as a quotient
$$\mathcal{L} = (\mathbb{C}\times V)/\mathbb{Z}\times\mathbb{Z} \tag{12.277}$$
where, again $V \cong \mathbb{C}$ is a one-dimensional vector space and now the generators of $\mathbb{Z}\times\mathbb{Z}$ act by
$$\begin{aligned} \sigma_1 &: (z,\psi) \mapsto (z+1,\psi) \\ \sigma_2 &: (z,\psi) \mapsto (z+\tau, g(z;\tau)\psi) \\ g(z;\tau) &:= e^{-\mathrm{i}\pi\tau - 2\pi \mathrm{i} z} \end{aligned} \tag{12.278}$$
where $\psi \in V$ is any vector. As before we have a map
$$\pi : \mathcal{L} \to E_\tau \tag{12.279}$$
defined by $\pi([z,\psi]) := [z]$, but now the big difference from before is that this is a holomorphic map.

2. We can take tensor products of line bundles. In terms of our quotient construction we choose different one-dimensional representations $V_n \cong \mathbb{C}$ and $\mathcal{L}^{\otimes n}$ is the holomorphic line bundle defined by
$$\begin{aligned} \sigma_1 &: (z,\psi) \mapsto (z+1,\psi) \\ \sigma_2 &: (z,\psi) \mapsto (z+\tau, g(z;\tau)^n\psi) \end{aligned} \tag{12.280}$$
There will be holomorphic sections for $n \geq 0$ and not for $n < 0$. The holomorphic sections for $n > 0$ form a vector space, and one basis for this vector space are the level κ theta functions with $\kappa = n/2$.

3. Note that we can put an Hermitian metric on $\mathcal{L}^{\otimes n}$. When we used the unitary transition functions, as with (12.242) the norm square $|\psi(x,y)|^2$ was doubly-periodic and descended to a well-defined function on the torus. However, if we consider the relation of the ground-state wavefunctions to the theta functions in equation (12.271) the prefactor is not a periodic function. If we are considering sections of a holomorphic line bundle and we want to assign a length-square to them which descends to a well-defined function on the torus we must multiply by some kind of quadatric exponent.

   If $s([z])$ is any section of $\mathcal{L}^{\otimes n}$ then we can express it in the form $[(z,\psi(z))]$ with $\psi(z) \in \mathbb{C}$. Then we can define

$$\parallel s([z]) \parallel^2 := e^{-2\pi n \frac{(\mathrm{Im} z)^2}{\mathrm{Im}\tau}} |\psi(z)|^2 \tag{12.281}$$

   The reader should check that different representatives give the same quantity on the RHS.

4. Given an Hermitian holomorphic line bundle there is a natural connection on it, and the corresponding curvature has an elegant formula:

$$\mathcal{R} = \frac{1}{2\pi}\partial\bar{\partial}\log \parallel s \parallel^2 \tag{12.282}$$

   where we can choose any holomorphic section s. In our case we get

$$\mathcal{R} = n\frac{dx \wedge dy}{\mathrm{Im}\tau} \tag{12.283}$$

5. As a final remark on theta functions, for a fixed κ consider the map form $\mathbb{C}$ to $\mathbb{C}^{2\kappa}$ given by

$$z \mapsto (\Theta_{\mu_0+1,\kappa}(z,\tau), \Theta_{\mu_0+2,\kappa}(z,\tau), \dots, \Theta_{\mu_0+2\kappa,\kappa}(z,\tau)) \tag{12.284}$$

   The zeroes of theta functions are well-understood, and it can be shown that for sufficiently large κ ($\kappa \geq 2$ will suffice) the vector on the RHS never vanishes. Therefore the map descends to a well-defined map into projective space $\mathbb{CP}^{2\kappa-1}$:

$$z \mapsto [\Theta_{\mu_0+1,\kappa}(z,\tau) : \Theta_{\mu_0+2,\kappa}(z,\tau) : \cdots : \Theta_{\mu_0+2\kappa,\kappa}(z,\tau)] \tag{12.285}$$

   Moreover, note from the quasiperiodic behavior of $\Theta_{\mu,\kappa}(z,\tau)$ that the transformation law under $z \to z+1$ and $z \to z+\tau$ is independent of μ. Therefore the map further descends to a holomoprhic map

$$E_\tau \to \mathbb{CP}^{2\kappa-1} \tag{12.286}$$

   For $\kappa \geq 2$ this map can be shown to be an embedding. Thus $\mathcal{L}$ is an example of an *ample line bundle* meaning that sections of some positive power of it defines an embedding in projective space. Furthermore, the theta functions satisfy a host of quartic polynomial relations known as *Riemann identities*. Specializing these identities appropriately gives explicit polynomial equations defining the embedded E_τ. For example, the torus E_τ can be realized as an intersection of two quadrics in $\mathbb{CP}^3$. For details on the above claims see [38], p.11 et. seq.

♣Check that this lower bound is correct! ♣

♣Should give more details here. ♣

12.7.7 Band Theory

Now we consider a different case of "electrons" in a zero electromagnetic field, but in the presence of a periodic potential $V(\vec{x})$. We put the quotation marks because we will ignore spin and electron-electron interactions.

The typical situation here is that in a crystalline structure in $\mathbb{E}^n$, where $\mathbb{E}^n$ is n-dimensional affine Euclidean space the atoms are arranged in a *crystal* $C \subset \mathbb{E}^n$. This simply means that there is a lattice $\Lambda \subset \mathbb{R}^n$ and there is a subset $C \subset \mathbb{E}^n$ invariant under translation by Λ.

We first explain the conventional viewpoint on band structure:

Choose an origin and identify $\mathbb{E}^n \cong V \cong \mathbb{R}^n$. Consider a single spinless electron propagating in $\mathbb{R}^n$ and interacting in some way with a crystal C. The Schrödinger Hamiltonian is the operator on $\mathcal{H} = L^2(\mathbb{R}^n)$ given by

$$H = -\frac{\hbar^2}{2m}\nabla^2 + U(x) \tag{12.287}$$

and the potential energy $U(\vec{x})$ is invariant under translations by the lattice Λ. Now, the abelian group Λ acts unitarily on $\mathcal{H}$ via $\rho(\lambda) = \exp[\mathrm{i}\langle \hat{p}, \lambda\rangle]$ where $\hat{p}$ is the usual momentum operator, $\lambda \in \Lambda$ and we denote the pairing $V^\vee \times V \to \mathbb{R}$ by $\langle\cdot,\cdot\rangle$. This group commutes with H and hence we expect to decompose the Hilbert space $\mathcal{H}$ as a direct sum over "isotypical components":

Recall that quite generally, if $\mathcal{H}$ is a completely reducible representation of a group G and G has a list of distinct irreps $\{R_\alpha\}$ then the decomposition

$$\mathcal{H} \cong \oplus_\alpha D_\alpha \otimes R_\alpha \tag{12.288}$$

is called the "isotypical decomposition." The group acts as $\oplus_\alpha 1 \otimes \rho_\alpha(g)$ on the RHS. The sum $\oplus_\alpha$ might well be a direct integral. The summands $D_\alpha \otimes R_\alpha$ are called the "isotypical components" and can be characterized invariantly as the image of the canonical evaluation map:

$$\mathrm{Hom}_G(V_\alpha, \mathcal{H}) \otimes V_\alpha \to \mathcal{H}. \tag{12.289}$$

The Hamiltonian is block diagonalized in this decomposition, that is, it acts separately on each isotypical component. In fact, it is of the form $\oplus_\alpha H_\alpha \otimes 1$.

So, let us consider the set of unitary irreducible representations of Λ. Since Λ is Abelian they are all one-dimensional, and in fact, they form a group - the Pontryagin dual group. The set of characters is in fact a manifold that can be identified with a torus as follows. Given any vector $k \in V^\vee := \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$ we can form an irrep:

$$\chi_{\bar{k}} : \lambda \mapsto e^{2\pi \mathrm{i}\langle k,\lambda\rangle} \in U(1) \tag{12.290}$$

and all irreps can be so represented. Of course, if we shift k by an element K of $\Lambda^\vee := \mathrm{Hom}_{\mathbb{Z}}(\Lambda, \mathbb{Z})$ then k and $k + K$ define the same irrep. In this way we can identify the space of unitary irreps of Λ with the torus

$$T^\vee := V^\vee/\Lambda^\vee \tag{12.291}$$

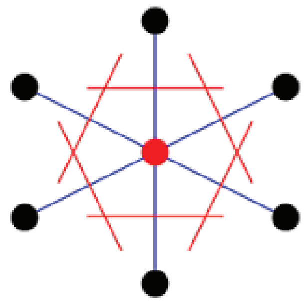


Figure 30: Constructing a Wigner-Seitz (or Voronoi) cell for the triangular lattice. The cells are regular hexagons. Figure from Wikipedia.

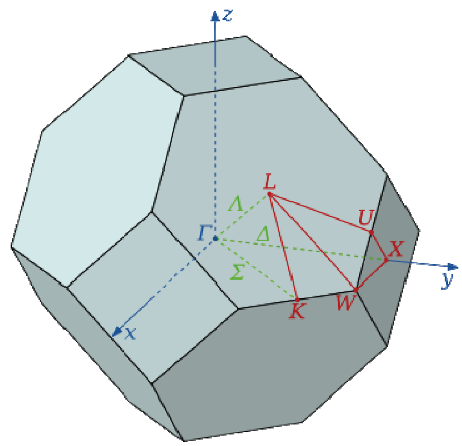


Figure 31: A Wigner-Seitz (or Voronoi) cell for the cubic lattice in $\mathbb{R}^3$. Figure from Wikipedia.

The torus $T^\vee$ is known in solid state physics as the *Brillouin torus* . Elements $k \in V^\vee$ are called *reciprocal vectors* and lattice vectors $K \in \mathrm{Hom}(\Lambda, \mathbb{Z})$ are called *reciprocal lattice vectors*.

In general, given an embedded lattice $\Lambda \subset \mathbb{R}^n$ we can use the metric to produce a canonical (i.e. basis-independent) set of fundamental domains, for the Λ-action on $\mathbb{R}^n$ by translation. These are known as *Voronoi cells* in mathematics and as *Wigner-Seitz cells* in physics. Choose any lattice point $v \in \Lambda$ and take $\bar{\mathcal{F}}$ to be the set of all points in $\mathbb{R}^n$ which are closer to v than to any other point. (If the points are equidistant to another lattice point we include them in the closure $\bar{\mathcal{F}}$. The Wigner-Seitz cell in $V^\vee$ for the reciprocal lattice is known in solid state physics as the *Brillouin zone*. Note that there is a clear algorithm for constructing $\bar{\mathcal{F}}$: Starting with v we look at all other points $v' \in \Lambda$. We consider the hyperplane perpendicular to the line between v and v' and take the intersection of all the half-planes containing v. It is also worth remarking that the concept of Voronoi cell does not require a lattice and applies to any collection of points, indeed, any collection of subsets of $\mathbb{R}^n$. The case of a triangular lattice is shown in Figure 30 and for a cubic lattice in $\mathbb{R}^3$ it is shown in Figure 31. In solid state physics the Brillouin zone is centered on $K = 0$ and the origin is always denoted as Γ.

Given the above group-theoretic facts the isotypical decomposition of the Hilbert space should be something like:

$$\mathcal{H} = \int_{T^\vee} d\bar{k} \mathcal{H}_{\bar{k}} \tag{12.292}$$

we will be somewhat more precise about this formula below.

Wavefunctions which transform under translation by Λ with a definite character $\chi_{\bar{k}}$ are known as "Bloch waves" and can always be written in the form

$$\psi(x) = e^{2\pi \mathrm{i}\langle k,x\rangle}\wp(x) \tag{12.293}$$

where k is some (any) lift of $\bar{k}$ to $V^\vee$ and $\wp(x)$ is a *periodic* function:

$$\wp(x+\lambda) = \wp(x) \qquad\qquad \forall \lambda \in \Lambda \tag{12.294}$$

Let us call this a "Bloch decomposition of ψ." Of course, there is some ambiguity in this decomposition. If we shift the lift $k \to k + K$, with $K \in \Lambda^\vee$ and simultaneously change

$$\wp(x) \to e^{-2\pi \mathrm{i}\langle K,x\rangle}\wp(x) \tag{12.295}$$

then the result is a different Bloch decomposition of the same wavefunction.

If we substitute a Bloch wavefunction into the eigenvalue equation for the Schrödinger Hamiltonian we obtain

$$H_k \wp = E \wp \tag{12.296}$$

where

$$H_k = \frac{\hbar^2}{2m}\left(-\mathrm{i}\nabla + 2\pi k\right)^2 + U(x) \tag{12.297}$$

Note that H_k is acting on <u>periodic</u> functions $\wp(x)$. These can equally well be considered as functions on the quotient torus

$$T := V/\Lambda \tag{12.298}$$

Viewed that way, we can take $\wp \in L^2(T)$ and H_k is an elliptic self-adjoint operator with a discrete spectrum bounded below and not above (provided U is bounded below).

Thus, we can find Bloch wavefunctions which are (formally) eigenfunctions of the Hamiltonian. We can only say this formally because of course the Bloch waves can never be in $L^2(\mathbb{R}^n)$. Nevertheless, they are very useful. Indeed, while this might seem like a mathematical drawback it is physically important: In quantum mechanics the electron wave can scatter coherently off the crystal without degrading. This would not make sense with a particle picture of electrons.

Now the spectrum $\sigma(H_k)$ of H_k acting on $L^2(T)$ actually only depends on the projection $\bar{k}$ of k to the Brillouin torus $T^\vee$, since we can conjugate H_k with the unitary operator

$e^{2\pi \mathrm{i}\langle K,x\rangle}$ for any $K \in \Lambda^\vee$. We will therefore denote the spectrum as $\mathcal{S}_{\bar{k}}$. Since it is discrete and bounded below we can write

$$\mathcal{S}_{\bar{k}} = \{E_n(\bar{k})\}_{n=0}^{\infty} \tag{12.299}$$

and, for a fixed $\bar{k}$, we can choose to label the eigenvalues so that we have an ordering:

$$E_0(\bar{k}) \le E_1(\bar{k}) \le \cdots \tag{12.300}$$

Now, it can be shown that the $E_n(\bar{k})$ are piecewise smooth functions of $\bar{k}$. They are called *energy bands.* These bands can intersect, and then $E_n(\bar{k})$ can be smoothly continued through the intersections (but then of course the ordering changes). If there is an n so that there is no intersection of $E_n(\bar{k})$ and $E_{n+1}(\bar{k})$ for any $\bar{k} \in T^\vee$ and moreover

$$\max_{\bar{k}\in T^\vee} E_n(\bar{k}) < \min_{\bar{k}\in T^\vee} E_{n+1}(\bar{k}) \tag{12.301}$$

we say there is a *band gap.* (If the maximum and minimum are attained at the same $\bar{k}$ it is called a *direct gap* otherwise it is an *indirect gap.*)

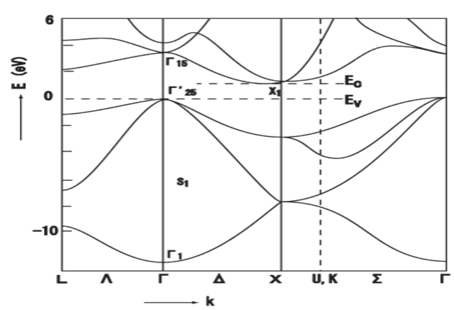


Figure 32: Band structure for silicon.

Remarks: *A little solid state physics*

1. In order to visualize the dependence of the energy eigenvalues on $\bar{k}$ and study the so-called *band structure* physicists typically choose a line in the Brillouin zone and plot $\mathcal{S}_{\bar{k}}$ along that line. See for example Figure 32. (For the two-dimensional case one can attempt to draw the energy surfaces.)

2. Crystal structures with band gaps can support materials that are *insulators.* The heuristic picture explaining this is the following: In the many-body Hilbert space $\Lambda^N\mathcal{H}$ the groundstate is obtained by filling up the lowest energy eigenvalues. If all the bands up to $E_n(\bar{k})$ are filled and none of the bands are filled for $m > n$ then applying a small electric field will not generate a current because not enough energy is available for electrons to make the transition from the n^{th} to the $(n+1)^{th}$ band.

We would now like to rephrase the above a bit more geometrically. Aside from being elegant, it will address two problems:

1. There is no global parametrization of Bloch wavefunctions on $T^\vee$.
2. The Bloch wavefunctions are not in $\mathcal{H} = L^2(\mathbb{E}^n)$, the physical Hilbert space.

It is useful to introduce the *Poincaré line bundle*, a complex line bundle

$$L_{\mathcal{P}} \to T^\vee \times T \tag{12.302}$$

where

$$T = V/\Lambda \qquad\qquad T^\vee = V^\vee/\Lambda^\vee \tag{12.303}$$

Again, we will present it as a homogeneous vector bundle, similarly to what we did for the electrons on a torus.

The total space is

$$L_{\mathcal{P}} := \big(V \times V^\vee \times W\big) / \big(\Lambda \times \Lambda^\vee\big) \tag{12.304}$$

where $W \cong \mathbb{C}$ is a one-dimensional irrep of $\Lambda \times \Lambda^\vee$ and the group action is

$$(\lambda, K) : (x, k; \psi) \mapsto (x + \lambda, k + K; e^{2\pi \mathrm{i}\langle k,\lambda\rangle}\psi) \tag{12.305}$$

Note that if we choose a particular character $\bar{k} \in T^\vee$ then we have a map

$$\begin{tikzcd} & L_{\mathcal{P}} \arrow[d] \\ T \arrow[r, "\iota_{\bar{k}}"'] & T \times T^\vee \end{tikzcd} \tag{12.306}$$

The pullback by $\iota_{\bar{k}}$ of $L_{\mathcal{P}}$ defines a line bundle over T:

$$L_{\bar{k}} := \iota_{\bar{k}}^*(L_{\mathcal{P}}) \tag{12.307}$$

Explicitly, $L_{\bar{k}} = (V \times W)/\Lambda$ with the identification:

$$(x; \psi) \sim (x + \lambda; \chi_{\bar{k}}(\lambda)\psi) \qquad \forall \lambda \in \Lambda \tag{12.308}$$

Thought of as equivariant functions $\psi : V \to W$, the sections of $L_{\bar{k}}$ are just Bloch waves with character $\chi_{\bar{k}}$:

$$\psi(x + \lambda) = \chi_{\bar{k}}(\lambda)\psi(x) \tag{12.309}$$

Now we claim that the physical Hilbert space $\mathcal{H} = L^2(V; W) \cong L^2(\mathbb{R}^n)$ can be identified with $L^2(T \times T^\vee; \mathcal{L}_{\mathcal{P}})$. The L^2 condition on the latter means that, for an equivariant function $\Psi(k, x)$ we have the norm-square:

$$\parallel \Psi \parallel^2 = \int_{T^\vee \times T} d\bar{k} dx |\Psi(k, x)|^2 \tag{12.310}$$

(note that thanks to the equivariance $|\Psi(k, x)|^2$ descends to a well-defined function on $T^\vee \times T$.

If we have an L^2 function $\psi(x)$ then we can form a family, parametrized by $\bar{k} \in T^\vee$, of equivariant functions by averaging:

$$\Psi(\bar{k}, x) := \sum_{\lambda \in \Lambda} \chi_{\bar{k}}(\lambda)^{-1} \psi(x + \lambda) \tag{12.311}$$

(and this sum will certainly converge for $\psi(x)$ in the Schwarz space of functions of rapid decrease on V, and these are dense in $\in L^2$) while conversely given a family of equivariant functions $\Psi(\bar{k}, x)$, defining L^2 sections of $L_{\bar{k}} \to T$ we can form

$$\psi(x) = \int_{T^\vee} \Psi(\bar{k}, x) d\bar{k} \tag{12.312}$$

Note that

$$\psi(x + \lambda) = \int_{T^\vee} \chi_{\bar{k}}(\lambda) \Psi(\bar{k}, x) d\bar{k} \tag{12.313}$$

so for $\lambda \to \infty$ this goes to zero. In fact the averaged function will be in $L^2(V; W)$. To see this write

$$\begin{aligned} \int_V dx |\psi(x)|^2 &= \int_{V/\Lambda} dx \sum_\lambda \int_{T^\vee \times T^\vee} \Psi(\bar{k}_1, x + \lambda)^* \Psi(\bar{k}_2, x + \lambda) \\ &= \int_{V/\Lambda} dx \int_{T^\vee \times T^\vee} \Psi(\bar{k}_1, x)^* \Psi(\bar{k}_2, x) \sum_\lambda \chi_{\bar{k}_1}(\lambda)^* \chi_{\bar{k}_2}(\lambda) \\ &= \int_{V/\Lambda} dx \int_{T^\vee \times T^\vee} \Psi(\bar{k}_1, x)^* \Psi(\bar{k}_2, x) \delta(\bar{k}_1 - \bar{k}_2) \\ &= \int_{T \times T^\vee} dx d\bar{k} |\Psi(\bar{k}, x)|^2 \end{aligned} \tag{12.314}$$

Finally, we observe a very general result about bundles over product spaces:

Proposition Suppose we have a vector bundle over a product of two spaces $E \to X \times Y$. Then, for each x let

$$\mathcal{E}_x := L^2(Y; \iota_x^*(E)) \tag{12.315}$$

For each $x \in X$, $\mathcal{E}_x$ is a Hilbert space. With suitable operator topologies these form a continuous family of Hilbert spaces and in this way we get a bundle of Hilbert spaces over X:

$$\pi : \mathcal{E} \to X \tag{12.316}$$

whose fiber at x is just $\mathcal{E}_x$. Then

$$L^2(X \times Y; E) \cong L^2(X; \mathcal{E}) \tag{12.317}$$

For a proof see Appendix D of [17]. There is an issue here about what topology to use on the group $U(\mathcal{H})$ of unitary transformations on Hilbert space when requiring that transition functions be "continuous." See Appendix D of and Appendix A of [1].

Applied to our present example, we have

$$\mathcal{E}_{\bar{k}} = L^2(T; L_{\bar{k}}) \tag{12.318}$$

As $\bar{k}$ varies over $T^\vee$ the Hilbert spaces fit into a Hilbert bundle $\mathcal{E} \to T^\vee$, and we can identify

$$\begin{aligned} \mathcal{H} := L^2(V) &\cong L^2(T \times T^\vee; L_{\mathcal{P}}) \\ &\cong L^2(T^\vee; \mathcal{E}) \end{aligned} \tag{12.319}$$

Remarks:

1. It can be shown that the unitary group on Hilbert space $U(\mathcal{H})$ is contractible in both the compact-open and in the norm topologies. Consequently, all bundles of Hilbert spaces are trivializable. However, they might not come with a natural trivialization, so it would be a mistake to assume every Hilbert bundle is of the form $X \times \mathcal{H}$. The present case is an example of such a situation.

2. The bundle $\mathcal{E} \to T^\vee$ carries a canonical family of flat connections labeled by $\bar{x}_0 \in T$. In order to see this it suffices to describe the parallel transport along straight-line paths in $T^\vee$. The most general such path is the projection of a straight line path in $V^\vee$. That is, consider the path in $V^\vee$:

$$\hat{\gamma}(t) = k(t) = k_i + t\Delta k, \qquad 0 \le t \le 1 \tag{12.320}$$

where $\Delta k = k_f - k_i$. Then the projected path

$$t \mapsto \gamma(t) := [k(t)] := \overline{k(t)} \in V^\vee/\Lambda^\vee \tag{12.321}$$

is a straightline path from $\bar{k}_i$ to $\bar{k}_f$. Of course, if we change $k_f \to k_f + K$ with $K \in \Lambda^\vee$ we get another straightline path from $\bar{k}_i$ to $\bar{k}_f$: It might wrap around a nontrivial loop in $T^\vee$ several times (specified by K) before coming to an end on $\bar{k}_f$. The paths with fixed endpoints are a torsor for $\Lambda^\vee$. In particular, if we consider closed paths with $\bar{k}_i = \bar{k}_f$ we see that

$$\pi_1(T^\vee, \bar{k}_0) \cong \Lambda^\vee \tag{12.322}$$

for any basepoint $\bar{k}_0$. Now let us describe the parallel transport. For any $\psi_{\bar{k}} \in \mathcal{E}_{\bar{k}}$ we have to say how it is parallel transported along $\overline{k(t)}$. We interpret the fibers of $\mathcal{E}$ as spaces of quasiperiodic functions satisfying (12.309) and define a family of quasiperiodic functions

$$(U_{x_0}(t) \cdot \psi_{\bar{k}_i})(x) := e^{2\pi i t \Delta k \cdot (x - x_0)} \psi_{\bar{k}_i}(x) \tag{12.323}$$

where x_0 is a lift of $\bar{x}_0 \in T$. Note that for each t the resulting function has the quasiperiodicity (12.309) determined by $\overline{k(t)}$:

$$\begin{aligned} (U_{x_0}(t) \cdot \psi_{\bar{k}_i})(x + \lambda) &:= e^{2\pi \mathrm{i} t \Delta k \cdot (x + \lambda - x_0)} \psi_{\bar{k}_i}(x + \lambda) \\ &= \left(e^{2\pi \mathrm{i} t \Delta k \cdot \lambda} e^{2\pi \mathrm{i} k_i \cdot \lambda} \right) e^{2\pi \mathrm{i} t \Delta k \cdot (x - x_0)} \psi_{\bar{k}_i}(x) \\ &= \chi_{\overline{k(t)}}(\lambda) (U_{x_0}(t) \cdot \psi_{\bar{k}_i})(x) \end{aligned} \tag{12.324}$$

Computation of the parallel transport around contractible loops shows that the connection is a "flat connection": The parallel transport around homotopically trivial loops in T is always one. (Equivalently, the curvature is zero.) However, the holonomy around the closed loop γ in homotopy class $K \in \Lambda^\vee$ (using (12.322)) is clearly multiplication by the *periodic* function (put $t = 1$ and $\Delta k = K$ in equation (12.323)

$$h_{x_0}(\gamma) = e^{2\pi i \langle K, (x - x_0) \rangle} \tag{12.325}$$

and is nontrivial. Note that it only depends on the projection $\bar{x}_0$ of x_0 to T, and so the isomorphism class of the connection only depends on $\bar{x}_0$.

3. As we remarked above, for insulators, the groundstate electron wavefunction distinguishes a finite-dimensional sub-bundle of $\pi : \mathcal{E} \to X$ spannned by the electron wavefunctions in the filled bands: If bands 1 to n are "filled" then we define a projection operator $P(\bar{k})$ to be the projector

$$P(\bar{k}) = \theta(E_{n+1}(\bar{k}) - H) \tag{12.326}$$

acting on $\mathcal{E}_{\bar{k}}$. Here

$$\theta(x) := \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}. \tag{12.327}$$

and we recall that it makes sense to apply a measurable function to an operator. ("Borel functional calculus.") This defines a finite-dimensional vector bundle $\mathcal{F} \to T^\vee$. [36] Moreover, given a connection on $\mathcal{E}$, $\mathcal{F}$ contains a canonical connection: Whenever we have a bundle $E \to X$ with a family of projection operators $P(x)$ on E if E has a connection then the sub-bundle whose fibers are $F_x = P(x) E_x$ also inherits a connection, known as a projected connection: If we have a path $\gamma(t)$ in X and we lift $\gamma(0)$ so that $\tilde{\gamma}(0) \in F_{x_0}$ then we use the lifted path $\tilde{\gamma}_E(t)$ in E and then project it to F:

$$\tilde{\gamma}_F(t) := P(\gamma(t)) \tilde{\gamma}_E(t) \tag{12.328}$$

[36] In fact, as we will see below, a fundamental result is that every vector bundle is the image of a continuous family of projection operators acting on a trivial bundle. This important theorem is known as the *Serre-Swan theorem.* To prove it let us, WLOG assume that E has an Hermitian metric and a reduction of structure group to $U(n)$. The theorem applies when the base manifold has a finite cover with a partition of unity so that the bundle is trivializable on each open set in the cover. (If X is a smooth compact manifold such a cover always exists.) Choose such a finite cover $\{U_\alpha\}$ for the base X with a partition of unity $\{\lambda_\alpha\}$. Then, for each U_α choose a unitary basis $s_i^{(\alpha)}$ of sections of E on U_α. Then, while $s_i^{(\alpha)}$ are only locally defined, the sections $\lambda_\alpha s_i^{(\alpha)}$ are globally defined. Moreover, at any $x \in X$ the span of these vectors is the fiber E_x. Now let V be the span of this set of section $\{\lambda_\alpha s_i^{(\alpha)}\}_{i,\alpha}$ as a vector subspace of $\Gamma(E)$. Then we consider the trivial bundle $X \times V$ and let the projector be $P(x) = \sum_\alpha \lambda_\alpha(x) s_i^{(\alpha)}(x) (s_i^{(\alpha)})^\dagger(x)$. It is not difficult to show that $P(x)^2 = P(x)$. The operator $s_i^{(\alpha)}(x)(s_i^{(\alpha)})^\dagger(x)$ operates on a section of E by evaluating at x and contracting with $s_i^{(\alpha)}(x)$ using the Hermitian metric. The image is the fiber of E at x.

Since, in our case $\mathcal{E}$ has a canonical family of flat connections, it follows that $\mathcal{F}$ has a canonical family of projected (in general not flat) connections. [37]

A beautiful development in condensed matter physics in the past 8 years has been the discovery that the topology of the vector bundle $\mathcal{F}$ has direct physical implications. Moreover, the projection of the canonical connection mentioned above is known as the *Berry connection.* Some physically observable quantities involve various integrals of quantities associated with the Berry connection. Furthermore, it turns out that there are also more subtle torsion invariants associated with the K-theory class of the bundle that also have physically observable consequences.

4. Similarly, $L_{\mathcal{P}}$ carries a nontrivial flat connection. In fact, $T^{\vee}$ is the moduli space of flat line bundles over T, so $\mathcal{L}_{\mathcal{P}}$ is an example of a *universal bundle*: At a value of the modulus $\bar{k}$ it is the bundle parametrized by $\bar{k}$. ♣Need to clarify this remark ♣

5. COMMENT ON FOURIER-MUKAI TRANSFORM

12.7.8 Crystallographic Symmetry And Point Group Equivariance

When discussing band structure physicists often make use of more symmetry than the group of lattice translations of a crystal. As any visit to a museum will make plane, many crystals have much more symmetry. How are these symmetries realized in the above geometrical context?

Let us put this question into a broader context. Suppose that we have a group action G on X as well as a fiber bundle

$$\pi : E \to X \tag{12.329}$$

We say that the symmetry *lifts* if there is a group action G on E which commutes with the projection π.

$$\pi(\tilde{\mu}(g, e)) = \mu(g, \pi(e)) \tag{12.330}$$

where $\mu : G \times X \to X$ is the group action on the base, and $\tilde{\mu} : G \times E \to E$ is the lifted group action on the total space.

In commutative diagrams we write:

$$\begin{tikzcd} E \arrow[r, "\tilde{\mu}(g,\cdot)"] \arrow[d, "\pi"] & E \arrow[d, "\pi"] \\ X \arrow[r, "\mu(g,\cdot)"] & X \end{tikzcd} \tag{12.331}$$

Definition A bundle $\pi : E \to X$ together with a group action by G compatible with the projection (i.e. such that the group action lifts a G-action on X) is called an *equivariant vector bundle.*

[37] A difference in our discussion from the standard discussion in the condensed matter literature is that in the latter it is assumed that there is a natural trivialization of the Hilbert bundle and that the projected connection is the projection of the trivial connection. Consequently condensed matter physicists speak of "the" Berry connection. In fact, the projected connection is the projection of a nontrivial, but flat connection, and moreover there is actually a family of such connections labeled by T.

Example: Consider a vector bundle over a single point: $E \to X$, where $X = \{x_0\}$ is a point. The fiber is a vector space E_{x_0}. Then the only possible group action on x_0 is the trivial one. A lift of this group action is a representation of G on the fiber E_{x_0}.

It can very well happen that a group might not lift, but a central extension of the group will lift. For example, consider the magnetic monopole bundle, the subbundle of $S^2 \times \mathbb{C}^2$ given by the continuous family of projection operators:

$$P(\hat{x}) = \frac{1}{2}(1 + \hat{x} \cdot \vec{\sigma}) \tag{12.332}$$

That is, the fiber is:

$$\mathcal{L}_{\hat{x}} = P(\hat{x})(\hat{x} \times \mathbb{C}^2) \tag{12.333}$$

The group $SO(3)$ acts by rotations on S^2 in the standard way: $R : \hat{x} \to \hat{x}'$. But this group does <u>not</u> lift to $\mathcal{L}$. Rather, the central extension $SU(2)$ lifts. Indeed given $u \in SU(2)$ we can say that

$$u^{-1}\hat{x} \cdot \vec{\sigma} u = (R(u)\hat{x}) \cdot \vec{\sigma} = \hat{x}' \cdot \vec{\sigma} \tag{12.334}$$

So

$$u^{-1}P(\hat{x})u = P(\hat{x}') \tag{12.335}$$

and hence the lifted action is

$$u : \psi_{\hat{x}} \in L_{\hat{x}} \mapsto u^{-1}\psi_{\hat{x}} \tag{12.336}$$

Now, let us consider a crystal $C \subset \mathbb{E}^d$. To describe its symmetries we start with the affine Euclidean group of isometries of $\mathbb{E}^d$. It fits in an exact sequence:

$$1 \to \mathbb{R}^d \to \mathrm{Euc}(d) \to O(d) \to 1 \tag{12.337}$$

This sequence splits, i.e. $\mathrm{Euc}(d)$ is isomorphic to a semidirect product $\mathbb{R}^d \rtimes O(d)$. But there is no natural isomorphism of splitting. The rotation-reflections $O(d)$ do *not* act naturally on affine space. In order to define such an action one needs to *choose an origin* of the affine space and thus identify it with $\mathbb{R}^d$.

If we do choose an origin then we can identify $\mathbb{E}^d \cong \mathbb{R}^d$ and then to a pair $R \in O(d)$ and $v \in \mathbb{R}^d$ we can associate the isometry: [38]

$$\{R|v\} : x \mapsto Rx + v \tag{12.338}$$

In this notation -known as the *Seitz notation* - the group multiplication law is

$$\{R_1|v_1\}\{R_2|v_2\} = \{R_1R_2|v_1 + R_1v_2\} \tag{12.339}$$

which makes clear that

[38] Logically, since we operate with R first and then translate by v the notation should have been $\{v|R\}$, but unfortunately the notation used here is the standard one.

1. There is a nontrivial automorphism used to construct the semidirect product: $O(d)$:

$$\{R|v\}\{1|w\}\{R|v\}^{-1} = \{1|Rw\} \tag{12.340}$$

and $\pi : \{R|v\} \to R$ is a surjective homomorphism $\mathrm{Euc}(d) \to O(d)$.

2. Thus, although $\mathbb{R}^d$ is abelian, the extension is *not* a central extension.

3. On the other hand, having chosen an origin, the sequence is split. We can choose a splitting $s : O(d) \to \mathrm{Euc}(d)$ by

$$s : R \mapsto \{R|0\} \tag{12.341}$$

By definition, a *crystallographic group* is a group that is isomorphic to the subgroup of $\mathrm{Euc}(d)$ preserving a crystal $C \subset \mathbb{E}^d$. Therefore, there is an exact sequence:

$$1 \to \Lambda \to G(C) \to P(C) \to 1 \tag{12.342}$$

where Λ is the lattice of translations preserving C and $P(C) \subset O(n)$ is known as the *point group*. In general this sequence does <u>not</u> split:

Example: In $d = 2$ consider

$$C = \mathbb{Z}^2 \cup \left(\mathbb{Z}^2 + (\delta, \frac{1}{2})\right) \tag{12.343}$$

with $0 < \delta < \frac{1}{2}$ then $G(C)$ is not split. Indeed the group is generated by

$$g_1 : (x^1, x^2) \to (-x^1 + \delta, x^2 + \frac{1}{2}) \tag{12.344}$$

and

$$g_2 : (x^1, x^2) \to (x^1, -x^2) \tag{12.345}$$

Then $\Lambda = \mathbb{Z}^2$, and $P(C) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ is generated by the projections of g_1, g_2. However, no lift of $\pi(g_1)$ will square to one because g_1^2 is a translation by $(0, 1)$.

In general, when there is no splitting, given a rotation-reflection $R \in P(C)$ any lift $\{R|v\} \in G(C)$ must be accompanied by a translation by a vector v that is <u>not</u> in the lattice Λ. In solid state physics this is known as a *non-symmorphic lattice*.

Now, of course, there is a right action of $\mathrm{Euc}(d)$ on $L^2(V)$:

$$\left(\psi \cdot \{R|v\}\right)(x) := \psi(Rx + v) \tag{12.346}$$

So restricting to $G(C) \subset \mathrm{Euc}(d)$, we learn that $L^2(V)$ is a $G(C)$-representation. Now we can ask: Is the isomorphism

$$L^2(V) \cong L^2(T^\vee; \mathcal{E}) \tag{12.347}$$

an isomorphism of $G(C)$ representations? This turns out to be subtle.

Let us first note a very general fact about group theory. Suppose we have an extension:

$$1 \to N \to G \to Q \to 1 \tag{12.348}$$

Let X be the space of irreps of N. Then Q acts on X as follows:

Suppose $\rho_W : N \to \mathrm{Aut}(W)$ is an irrep of N with carrier space W. Then we can twist it by an element $q \in Q$ by choosing a section $s(q) \in G$ and defining

$$\rho_W^{s,q}(n) := \rho_W(s(q) n s(q)^{-1}) \tag{12.349}$$

Note that since N is normal in G this makes sense: $s(q)ns(q)^{-1} \in N$, and $\rho_W^{s,q}$ is a new representation of N on the vector space W. It clearly depends on the choice of section s, but the isomorphism class of of $\rho_W^{s,q}$ does not depend on the choice of section s. In general, the isomorphism class of $\rho_W^{s,q}$ is distinct from that of ρ_W. Denoting the isomorphism class by $[W]$ for brevity, we have a well-defined map on X for each $q \in Q$:

$$[W] \to [W] \cdot q. \tag{12.350}$$

It is not difficult to show (exercise!!) that this is in fact a well-defined right-action:

$$([W] \cdot q_1) \cdot q_2 = [W] \cdot q_1 q_2 \tag{12.351}$$

Let us apply the above general remarks to our case. By the general remarks, $P(C)$ must act on $T^\vee$. We compute the action as follows. Given a rotation-reflection $R \in P(C) \subset O(d)$ our section has the form

$$s : R \in P(C) \mapsto \{R|v_R\} \in G(C) \tag{12.352}$$

We stress that this is just a section. There is absolutely no claim that it is a group homomorphism, and in general no such homomorphism exists. Nevertheless, we can compute the change of character:

$$\begin{aligned}
\chi_{\bar{k}}^R(\lambda) &= \chi_{\bar{k}}\left(\{R|v_R\}\{1|\lambda\}\{R|v_R\}^{-1}\right) \\
&= \chi_{\bar{k}}\left(\{1|R\lambda\}\right) \\
&= \chi_{\bar{k}'}(\lambda)
\end{aligned} \tag{12.353}$$

where $\bar{k}'$ is defined as follows: Given $\bar{k}$, choose a lift $k \in V^\vee$. Then $R^{tr}k = k'$ and we then project k' to $\bar{k}'$. The action $\bar{k} \to \bar{k}'$ does not depend on the choice of lift since R^t takes $\Lambda^\vee \to \Lambda^\vee$.

Note that if we identify $V^\vee \cong V$ using a Euclidean metric then the pairing $\langle k, x\rangle$ becomes the Euclidean metric so $\langle Rk, Rx\rangle = \langle k, x\rangle$ and $R^{tr} = R^{-1}$. We will in fact do this henceforth so that we can write R instead of $R^{tr,-1}$.

So, $P(C)$ indeed acts on $T^\vee$. Now let us try to lift this action to the Hilbert bundle $\mathcal{E} \to T^\vee$. We think of sections as equivariant functions $\Psi : T^\vee \times V \to \mathbb{C}$. Then

$$(\Psi \cdot R)(\bar{k}, x) := \Psi(R\bar{k}, Rx + v_R) \tag{12.354}$$

Recall that restricting $\Psi(\bar{k}, x)$ to a fixed $\bar{k}$ gives an equivariant function $V \to \mathbb{C}$ defining a section of $L_{\bar{k}} \to T$. Such an equivariant function is a vector in the fiber $\mathcal{E}_{\bar{k}}$ of the Hilbert bundle $\mathcal{E}$ over $\bar{k} \in T^\vee$. Therefore for the above formula to make sense $R \in P(C)$ should take a vector in the fiber $\mathcal{E}_{\bar{k}}$ to a vector in the fiber $\mathcal{E}_{R^{-1}\bar{k}}$. Let us check this indeed the case:

$$\begin{aligned}(\Psi\cdot R)\left(R^{-1}\bar{k}, x+\lambda\right) &= \Psi(\bar{k}, R(x+\lambda)+v_R)\\ &= \Psi(\bar{k}, R(x+\lambda)+v_R)\\ &= e^{2\pi i\langle k, R\lambda\rangle}\Psi(\bar{k}, Rx+v_R)\\ &= e^{2\pi i\langle R^{-1}k,\lambda\rangle}(\Psi\cdot R)\left(R^{-1}\bar{k}, x\right)\\ &= \chi_{R^{-1}\bar{k}}(\lambda)\,(\Psi\cdot R)\left(R^{-1}\bar{k}, x\right)\end{aligned} \tag{12.355}$$

So, indeed, for each element $R\in P(C)$ we can define a bundle map on $\mathcal{E}$ that correctly covers the $P(C)$ action on the base $T^\vee$. However, it is by no means clear that it defines a group action. It is not obvious that it satisfies the group law! Why not? There can be trouble if $s: P(C)\to G(C)$ is not a splitting, that is, if the crystal is non-symmorphic.

Note that the group law in Euc(d), and hence in $G(C)$ can be written:

$$\{R_1|v_1\}\{R_2|v_2\} = \{R_1R_2|R_1v_2+v_1\} \tag{12.356}$$

Now apply that to the elements $s(R)=\{R|v_R\}$. Trivially, we have:

$$\{R_1|v_{R_1}\}\{R_2|v_{R_2}\} = \{R_1R_2|R_1v_{R_2}+v_{R_1}\} \tag{12.357}$$

However, in general we cannot choose $v_{R_1R_2}$ to be equal to $R_1v_{R_2}+v_{R_1}$. However:

$$\{R_1R_2|v_{R_1R_2}\}^{-1}\{R_1|v_{R_1}\}\{R_2|v_{R_2}\} = \{1|(R_1R_2)^{-1}\left(R_1v_{R_2}+v_{R_1}-v_{R_1R_2}\right)\} \tag{12.358}$$

must preserve the crystal. Since it is a pure translation it must be a translation in Λ. Since $R\in P(C)$ preserves Λ we must have a lattice vector $\lambda(R_1,R_2)\in\Lambda$ for every pair $R_1,R_2\in P(C)$ so that

$$R_1v_{R_2}+v_{R_1} = v_{R_1R_2}+\lambda(R_1,R_2) \tag{12.359}$$

It is now a small computation to show that, for any equivariant function $\Psi(\bar{k},x)$ (writing it as a right-action):

$$\Psi\cdot R_1\cdot R_2 = e^{2\pi i\langle R_1R_2k,\lambda(R_1,R_2)\rangle}\Psi\cdot R_1R_2 \tag{12.360}$$

If the phase factor $e^{2\pi i\langle R_1R_2k,\lambda(R_1,R_2)\rangle}$ is nontrivial and there is no choice of the v_R that makes it trivial then we do not have a $G(C)$ equivariant bundle!

In order to explore further what is happening let us note that there will be special points $\bar{k}$ of $T^\vee$, known as "orbifold points" or "high-symmetry points" where a nontrivial subgroup of $P(\bar{k},C)\subset P(C)$ will stabilize $\bar{k}$. Then (12.360), evaluated at such a point $\bar{k}$, and for $R_1,R_2\in P(\bar{k},C)$ shows that actually it is a projective representation $\tilde{P}(\bar{k},C)$ that acts on the fiber. The cocycle extending $\tilde{P}(\bar{k},C)$ is

$$c(R_1,R_2) = \chi_{\bar{k}}((R_1R_2)^{-1}\lambda(R_1,R_2)) \tag{12.361}$$

Of course, the cocycle is ambiguous up to a group commutator. If $\tilde{P}(\bar{k},C)$ is abelian then we can form the gauge invariant quantity:

$$s(R_1,R_2) = \chi_{\bar{k}}((R_1R_2)^{-1}\left(\lambda(R_1,R_2)-\lambda(R_2,R_1)\right)) \tag{12.362}$$

So, is there a projective representation $\tilde{G}(C)$ acting on $\mathcal{E} \to T^\vee$? No! At the central point Γ, i.e. the projection to $T^\vee$ of $k = 0$ the entire point group is unextended. However, it can very well happen that at other points there is a nontrivial central extension.

Example: Consider the example (12.343) above. Then $P(C)$ is $\mathbb{Z}_2 \times \mathbb{Z}_2 \subset O(2)$ generated by

$$R_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad R_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{12.363}$$

so at

$$\bar{k} = \overline{(0, \frac{1}{2})} \qquad \& \qquad \bar{k} = \overline{(\frac{1}{2}, \frac{1}{2})} \tag{12.364}$$

the fixed point group is the entire group $P(C)$. Now we should clearly choose $v_{R_1} + (\delta, \frac{1}{2})$ and $v_{R_2} = 0$. If we lift $R_1 R_2$ to $g_1 g_2 = \{-1|(\delta, \frac{1}{2})\}$ then we would take $v_{R_1 R_2} = (\delta, \frac{1}{2})$. If we lift $R_1 R_2 = R_2 R_1$ to $g_2 g_1 = \{-1|(\delta, -\frac{1}{2})\}$ then we would take $v_{R_1 R_2} = (\delta, -\frac{1}{2})$. We need to make a definite choice and we will take $v_{R_1 R_2} = (\delta, \frac{1}{2})$. Then we compute:

$$\lambda(R_1, R_2) = 0 \qquad \& \qquad \lambda(R_2, R_1) = (0, -1) \tag{12.365}$$

Therefore,

$$s(R_1, R_2) = \chi_{\bar{k}}((0, 1)) = -1 \tag{12.366}$$

The central extension is D_4.

The mathematical structure we have discovered is a generalization of an equivariant vector bundle, known as a *twisted equivariant bundle*, where the adjective "twisted" is used in sense of "twisted K-theory," and not "topologically twisted." We will return to "twisted" bundles and "twisted K-theory" later. For more details see [17].

12.7.9 Electron In A Periodic Potential And A Magnetic Field

♣Also introduce magnetic translation group here and explain some uses of it. ♣

Now we combine both a magnetic field and periodic potential. For simplicity we consider a two-dimensional square lattice.

There are now two competing length scales: The length scale of the lattice a, and the magnetic length scale,

$$\ell = \sqrt{\frac{\hbar}{eB}} \tag{12.367}$$

equivalently there are two competing time scales of the problem: The inverse of the cyclotron energy m/eB and the time-scale for motion in a periodic lattice. For an electron with momentum $p = 2\pi\hbar/a$, where a is the lattice spacing, and energy $p^2/2m$ this would be a time scale $2ma^2/h$. The ratio of these timescales is just

$$\frac{(2ma^2/h)}{m/eB} = 2\frac{eBa^2}{h} = 2\frac{\Phi}{\Phi_0}. \tag{12.368}$$

It turns out that the spectrum of the Hamiltonian is an exquisitely sensitive function of this ratio of scales.

Suppose therefore there is a periodic potential $U(x)$ invariant under translation by Λ. The Hamiltonian is:

$$H = \frac{1}{2m}(\tilde{p}_1^2 + \tilde{p}_2^2) + U(x) \tag{12.369}$$

Finding the spectrum of such a Hamiltonian is a very difficult and subtle problem.

An important approximation to this physics problem is known as a *tight-binding approximation.* We imagine that the electron is confined to the lattice sites by a strong binding potential, but can hop from one site to another. When it hops it picks up a phase from the parallel transport with the Maxwell gauge field.

The resulting model is based on an infinite product of complex Clifford algebras $\mathbb{C}\ell_1$ where the factors are thought of as a Clifford algebra attached to each vertex of the lattice Λ. Denote the generators by $a(\lambda), a(\lambda)^\dagger$. They are standard fermionic creation and annihilation operators:

$$\begin{aligned} \{a(\lambda), a(\lambda')^\dagger\} &= \delta_{\lambda,\lambda'} \\ \{a(\lambda), a(\lambda')\} &= 0 \\ \{a(\lambda)^\dagger, a(\lambda')^\dagger\} &= 0 \end{aligned} \tag{12.370}$$

The form of the tight-binding Hamiltonian is

$$H_{tb} = \sum_e t(e) a(\lambda_f)^\dagger U(e) a(\lambda_i) \tag{12.371}$$

where the sum is over all the edges of the lattice connecting two neighboring lattice points. The edges are oriented and can carry either orientation. The oriented edge goes from λ_i to λ_f. The amplitude $t(e)$ represents a tunneling, or hopping amplitude for an electron to move from site λ_i to site λ_f. WLOG we can take it to be real and positive. Meanwhile $U(e) = e^{\mathrm{i}\varphi(e)}$ is a phase, representing the parallel transport by the connection.

The adjective "hopping" refers to the following picture: Up to isomorphism there is one Clifford module for $\mathbb{C}\ell_1$. It has a canonical ordered basis $\{|0\rangle, |1\rangle\}$ and with respect to this basis

$$a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad\qquad a^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{12.372}$$

The state $|0\rangle$ represents no electron and $|1\rangle$ represents one electron. Therefore, if there is an electron at site λ_i, the operator $a(\lambda_f)^\dagger a(\lambda_i)$ annihilates an electron at λ_i and creates one at λ_f. Therefore, the electron has "hopped," with amplitude $t(e)e^{\mathrm{i}\varphi(e)}$, from λ_i to λ_f.

The phase $U(e) = e^{\mathrm{i}\varphi(e)}$ models the phase from parallel transport along a path from λ_i to λ_f in an electromagnetic field in the continuum model. The Hamiltonian H must be Hermitian and this requires that if $-e$ is the edge with opposite orientation to e then

$$e^{\mathrm{i}\varphi(-e)} = e^{-\mathrm{i}\varphi(e)} \tag{12.373}$$

Of course, one can make a phase redefinition. These are gauge transformations, defined at the vertices of the lattice:

$$a(\lambda) \to e^{\mathrm{i}\theta(\lambda)} a(\lambda) \tag{12.374}$$

and this will change

$$e^{\mathrm{i}\varphi(e)} \to e^{-\mathrm{i}\theta(\lambda_f)} e^{\mathrm{i}\varphi(e)} e^{\mathrm{i}\theta(\lambda_i)} \tag{12.375}$$

Note that the holonomy around closed loops is unchanged. All closed loops can be decomposed into closed loops around unit cells of the lattice, so the only gauge invariant information is the holonomy around the unit cells. If we consider similar small loops for a connection on a manifold then we are measuring the curvature in the neighborhood of the loop.

Let us now assume that all the hopping amplitudes are the same in the horizontal direction. Denote them by t_1. Similarly, assume that all the hopping amplitudes are the same in the vertical direction and denote them by t_2. Furthermore let $\lambda = m\hat{x} + n\hat{y}$, and denote the corresponding fermionic oscillators by $a_{m,n}, a_{m,n}^{\dagger}$. In this notation the Hamiltonian is:

$$H_{tb} = \sum_{m,n\in\mathbb{Z}} \left(t_1 e^{\mathrm{i}\varphi_1(m,n)} a_{m,n}^{\dagger} a_{m-1,n} + h.c. \right) + \left(t_2 e^{\mathrm{i}\varphi_2(m,n)} a_{m,n}^{\dagger} a_{m,n-1} + h.c. \right) \tag{12.376}$$

We next make a further simplifying assumption: We take the holonomy around each unit cell to be the same. This is the discrete approximation to a uniform magnetic field. Then we can choose a gauge (the discrete analog of Landau gauge) so that

$$H_{tb} = \sum_{\lambda} t_1 \left(a_{m,n}^{\dagger} a_{m-1,n} + h.c. \right) + t_2 \left(e^{2\pi \mathrm{i}\phi m} a_{m,n}^{\dagger} a_{m,n-1} + h.c. \right) \tag{12.377}$$

where ϕ is a constant. Note that the holonomy around a unit cell in the counterclockwise direction is

$$e^{2\pi \mathrm{i}\phi(m+1)} e^{-2\pi \mathrm{i}\phi m} = e^{2\pi \mathrm{i}\phi} \tag{12.378}$$

Comparing with the continuum expressions we see that

$$\phi = \frac{\Phi}{\Phi_0} \tag{12.379}$$

where Φ is the magnetic flux through a unit cell.

Now consider

$$\Psi = \sum \psi(m,n) a_{m,n}^{\dagger} |0\rangle \tag{12.380}$$

This represents a single electron propagating through the lattice. $\psi(m,n)$ is the amplitude for the electron to be at the site $\lambda = m\hat{x} + n\hat{y}$. Ψ will be an eigenstate of the Hamiltonian if:

$$t_1 \left(\psi(m-1,n) + \psi(m+1,n)\right) + t_2 \left(e^{2\pi \mathrm{i}\phi m} \psi(m,n-1) + e^{-2\pi \mathrm{i}\phi m} \psi(m,n+1) \right) = E\psi(m,n). \tag{12.381}$$

Since the Schrödinger operator is translation invariant in n we can introduce the discrete analog of Bloch waves:

$$\psi(m,n) = e^{2\pi \mathrm{i} k_y n} \wp(m) \tag{12.382}$$

where $k_y \sim k_y + 1$. Then the eigenvalue equation becomes

$$t_1 \left(\wp(m+1) + \wp(m-1)\right) + t_2 \left(e^{2\pi \mathrm{i}(\phi m - k_y)} + e^{-2\pi \mathrm{i}(\phi m - k_y)} \right) \wp(m) = E\wp(m) \tag{12.383}$$

Now we introduce translation and multiplication operators U and V on functions on $\ell^2(\mathbb{Z})$:

$$\begin{aligned}(U\wp)(m) &= \wp(m+1)\\ (V\wp)(m) &= e^{2\pi\mathrm{i}\phi m}\wp(m)\end{aligned} \tag{12.384}$$

If $t_1 \neq 0$ we can measure energy in units of t_1. Setting $\mu = t_2/t_1$ our Hamiltonian has been reduced to

$$H = U + U^\dagger + \mu\left(zV + z^*V^\dagger\right) \tag{12.385}$$

with $z = e^{-2\pi\mathrm{i}k_y}$. For $\mu \neq 0$ this is the simplest nontrivial self-adjoint element of $\mathcal{A}_\theta$. It is known as an *almost Mathieu operator* and the special case of $\mu = 1$ is known as the *Harper* or *Hofstadter* or *Azbel* Hamiltonian (or some linear combination of these names). Finding the spectrum of this Hamiltonian is extremely nontrivial - it has been the subject of a great deal of work by mathematical physicists and condensed matter physicists going back at least to Onsager and, presumably, Mathieu.

Let us make some simple immediate observations:

1. It is a self-adjoint element of a C^*-algebra. So its spectrum is a compact subset of $\mathbb{R}$.

2. Moreover the spectral radius $r(H) = \parallel H \parallel$ and by the triangle inequality it is clear that

$$\parallel H \parallel \leq 2 + 2|\mu| \tag{12.386}$$

3. For $\mu = 1$ and $z = 1$ H is invariant under Fourier transform. Implying certain symmetries of the spectrum.

A major conjecture in this subject is the so-called "ten-martini problem," so named by Mark Kac and Barry Simon. It was proven to be true in [2]:

Theorem: If $\mu \neq 0$ and ϕ is irrational then the spectrum (which is then independent of k_y and only depends on $\phi \mathrm{mod} 1$) is a Cantor set.

Recall that a *Cantor set* is a topological space homeomorphic to the subset of $[0,1]$ obtained by successively removing the open middle thirds of intervals. It is an uncountable subset of $[0,1]$ of Lebesgue measure zero and is a compact space. This shows that the spectrum can be highly nontrivial!

One can get a lot of insight by considering the problem for the case when $\phi = p/q$ is rational. In this case the one-dimensional Schrödinger equation (12.383) is periodic in $m \to m+q$. Therefore, we can again introduce Bloch waves:

$$\wp(m) = e^{2\pi\mathrm{i}k_x m}\tilde{\wp}(m) \tag{12.387}$$

where $\tilde{\wp}(m+q) = \tilde{\wp}(m)$. Therefore, the Brillouin torus is $k_x \sim k_x + \frac{1}{q}$. Equation (12.383) now becomes

$$t_1\left(e^{2\pi\mathrm{i}k_x}\tilde{\wp}(m+1) + e^{-2\pi\mathrm{i}k_x}\tilde{\wp}(m-1)\right) + t_2\left(e^{2\pi\mathrm{i}(\phi m - k_y)} + e^{-2\pi\mathrm{i}(\phi m - k_y)}\right)\tilde{\wp}(m) = E\tilde{\wp}(m) \tag{12.388}$$

so, at fixed (k_x, k_y) the eigenvalue problem is reduced to finding the eigenvalues of the $q \times q$ matrix:

$$t_1 \left(z_1 \mathbf{u} + z_1^* \mathbf{u}^\dagger \right) + t_2 \left(z_2 \mathbf{v} + z_2^* \mathbf{v}^\dagger \right) \tag{12.389}$$

with $z_1 = e^{2\pi \mathrm{i} k_x}$ and $z_2 = e^{-2\pi \mathrm{i} k_y}$, where $\mathbf{u}$ and $\mathbf{v}$ are the shift and clock operators.

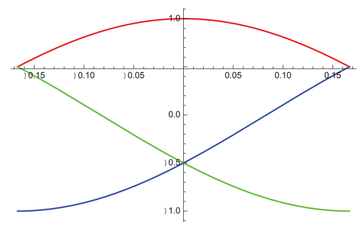


Figure 33: The simple band structure for the case $t_2 = 0$, $2t_1 = 1$ and $q = 3$. The bands are shown in Brillouin zone $-\frac{1}{6} \leq k_x \leq \frac{1}{6}$.

Before describing the spectrum in general let us look at three easy special cases:

1. If $t_1 = 0$ then since $\mathbf{v}$ is already diagonal we clearly have q bands as a function of k_y:

$$E_n = 2t_2 \cos(2\pi(\phi n - k_y)) \qquad n = 0, 1, \ldots, q-1 \tag{12.390}$$

Of course, if $t_2 = 0$ then, since $\mathbf{u}$ is the Fourier transform of $\mathbf{v}$ there is an analogous story with q bands

$$E_n = 2t_1 \cos(2\pi(\phi n + k_x)) \qquad n = 0, 1, \ldots, q-1 \tag{12.391}$$

If we take the union of the spectrum over the Brillouin torus we simply get an interval $[-2t_1, 2t_1]$ or $[-2t_2, 2t_2]$, respectively. See, for example, Figure 33.

2. Now suppose that ϕ is an integer, so $\phi \sim 0$. Then $\mathbf{v} = 1$ and we have, again rather trivially,

$$E_n = 2t_1 \cos(2\pi(\frac{n}{q} + k_x)) + 2t_2 \cos(2\pi k_y) \qquad n = 0, 1, \ldots, q-1 \tag{12.392}$$

so the union of the spectrum over the Brillouin torus is the full interval $[-2(t_1+t_2), 2(t_1+t_2)]$.

Now suppose that the fractional part of ϕ is nonzero and $t_1 t_2 \neq 0$. WLOG we can measure energies in units of t_1 so we can consider the spectrum of

$$h^{\phi,\mu}_{z_1,z_2} := z_1\mathbf{u} + z_1^*\mathbf{u}^\dagger + \mu\left(z_2\mathbf{v} + z_2^*\mathbf{v}^\dagger\right) \tag{12.393}$$

where $\mu = t_2/t_1$.

To find a criterion for finding the eigenvalues of (12.393) write the eigenvalue problem as

$$\begin{pmatrix} \tilde{\wp}(\bar{m}+\bar{1}) \\ \tilde{\wp}(\bar{m}) \end{pmatrix} = A(\bar{m})\begin{pmatrix} \tilde{\wp}(\bar{m}) \\ \tilde{\wp}(\bar{m}-\bar{1}) \end{pmatrix} \tag{12.394}$$

where we consider $\tilde{\wp}(\bar{m})$ as a function on $\mathbb{Z}/q\mathbb{Z}$ and

$$A(\bar{m}) = \begin{pmatrix} z_1^{-1}(E-\mu(z_2\omega^m + z_2^{-1}\omega^{-m})) & -z_1^{-2} \\ 1 & 0 \end{pmatrix} \tag{12.395}$$

with $\omega = e^{2\pi i\phi}$. Then, iterating the recursion relation (12.394) once around the circle we must have

$$\begin{pmatrix} \tilde{\wp}(\bar{q}+\bar{1}) \\ \tilde{\wp}(\bar{q}) \end{pmatrix} = A(\bar{q})A(\bar{q}-\bar{1})\cdots A(\bar{1})\begin{pmatrix} \tilde{\wp}(\bar{1}) \\ \tilde{\wp}(\bar{0}) \end{pmatrix} \tag{12.396}$$

But $\tilde{\wp}(\bar{q}+\bar{1}) = \tilde{\wp}(\bar{1})$ and $\tilde{\wp}(\bar{q}) = \tilde{\wp}(\bar{0})$. It follows that $A(\bar{q})A(\bar{q}-\bar{1})\cdots A(\bar{1})$ must have eigenvalue 1. However, its determinant is clearly z_1^{-2q}. Therefore, the quantization condition on the energies is

$$\mathrm{Tr}\left(A(\bar{q})A(\bar{q}-\bar{1})\cdots A(\bar{1})\right) = 1 + z_1^{-2q} \tag{12.397}$$

This is a little more elegant if we multiply by z_1^q and observe that $z_1 A(\bar{m})$ is conjugate to

$$B(\bar{m}) = \begin{pmatrix} E-\mu(z_2\omega^m + z_2^{-1}\omega^{-m}) & -1 \\ 1 & 0 \end{pmatrix}. \tag{12.398}$$

Therefore the eigenvalue equation is

$$\mathrm{Tr}\left(B(\bar{q})B(\bar{q}-1)\cdots B(\bar{1})\right) = z_1^q + z_1^{-q} \tag{12.399}$$

Note that the LHS is a monic polynomial in E of degree q that is independent of z_1.

Of course, this polynomial equation must be the same as the characteristic equation

$$\det(E\mathbf{1}_{q\times q} - h^{\phi,\mu}_{z_1,z_2}) = 0 \tag{12.400}$$

On the other hand, we know that

$$\mathcal{F}h^{\phi,\mu}_{z_1,z_2}\mathcal{F}^{-1} = \mu h^{\phi,1/\mu}_{z_2^{-1},z_1^{-1}} \tag{12.401}$$

Therefore it must also be true that the characteristic equation is of the form

$$P(E) = \mu^q(z_2^q + z_2^{-q}) \tag{12.402}$$

where $P(E)$ is monic polynomial in E of degree q and independent of z_2. Therefore, the characteristic equation for the energy eigenvalues must be of the form:

$$\begin{aligned}\det(x\mathbf{1}_{q\times q} - h^{\phi,\mu}_{z_1,z_2}) &= P_{\phi,\mu}(x) - \Big(z_1^q + z_1^{-q} + \mu^q(z_2^q + z_2^{-q})\Big) \\ &:= P_{\phi,\mu}(x) - f(\theta)\end{aligned} \tag{12.403}$$

where $P_{\phi,\mu}(x)$ is a polynomial independent of both z_1, z_2. Some examples are given in [30, 29] but it remains somewhat mysterious. We also write $(z_1, z_2) = e^{\mathrm{i}\theta_1}, e^{\mathrm{i}\theta_2})$ and $\theta = (\theta_1, \theta_2)$.

For a fixed θ there will be q roots of the equation $P_{\phi,\mu}(x) = f(\theta)$, $\{E_1(\theta), \dots, E_q(\theta)\}$ and as θ varies over the torus they will define bands. The r^{th} band will be $E_r(\theta) \in [E^r_{\min}, E^r_{\max}]$ where $E^r_{\min}, E^r_{\max}$ are critical values of the function $E_r(\theta)$ on the torus. The band functions are real-analytic functions of θ so that we can differentiate $P_{\phi,\mu}(E_r(\theta)) = f(\theta)$ to get:

$$P'_{\phi,\mu}(E_r(\theta))E'_r(\theta) = f'(\theta) \tag{12.404}$$

If two bands do not touch $P'_{\phi,\mu}(E_r(\theta)) \neq 0$ and hence critical points of $E'_r(\theta)$ are critical points of $f(\theta)$. For each critical point θ_a of $f(\theta)$ we can then find the bands by looking at the inverse images E_{crit}

$$P_{\phi,\mu}(E_{\mathrm{crit}}) = f(\theta_a) \tag{12.405}$$

In this way one can plot the bands.

Let $\mathrm{Sp}(\phi)$ be the union of the roots as (z_1, z_2) varies over the Brillouin torus. Recall that all the roots are real. So $\mathrm{Sp}(\phi)$ is just the inverse image under $P_{\phi,\mu}(x)$ of the interval $[-4, 4]$. There are q disjoint bands separated by $(q-1)$ open intervals - called "gaps," with one exception: When q is even the central gap closes. In [25] Hofstadter had the bright idea of plotting (for the case $\mu = 1$) these energy bands as a function of ϕ. The resulting figure is the famous "Hofstadter butterfly," reproduced in Figure 34.

Remarks

1. There is an interesting fractal structure in the bandstructure. It was first suggested by Azbel [3]. Roughly speaking if we write the continued fraction expansion for ϕ:

$$\phi = [N_1, N_2, N_3, \dots] = \frac{1}{N_1 + [N_2, N_3, \dots]} \tag{12.406}$$

   Then N_1 bands split into N_2 subbands split into N_3 subsubbands etc.

2. The labeling of the gaps is related to some interesting mathematics and physics. On the mathematical side one should consider the K-theory of the C^*-algebra $K_0(\mathcal{A}_\phi)$. We will discuss what this means below, but for now suffice it to say that we define equivalence classes on projection operators in (roughly speaking) $\mathcal{A}_\phi$. [39] Then we

[39] More precisely, in matrix algebras over $\mathcal{A}_\phi$, $M_N(\mathcal{A}_\phi)$ where N can be arbitrarily large. We can define equivalences of projectors $P_i \in M_{n_i}(\mathcal{A})$ by declaring them to be equivalent if there exists k_i and $v \in M_N(\mathcal{A})$ so that $vv^* = \mathrm{Diag}\{P_1, \mathbf{1}_{k_1}\}$ and $v^*v = \mathrm{Diag}\{P_2, \mathbf{1}_{k_2}\}$. The abelian group structure is then $[P_1] + [P_2] := [\mathrm{Diag}\{P_1, P_2\}]$.

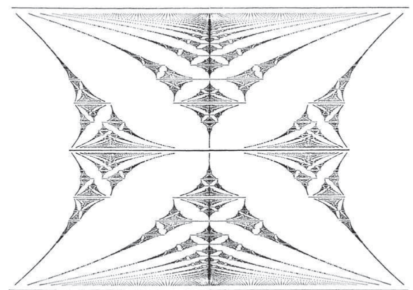


Figure 34: Figure 1 from Hoftstadter's paper. Energy bands are plotted horizontally and sit inside $[-4, 4]$. The vertical axis is ϕ, ranging from 0 to 1. Hofstadter plotted the bands for rational values of ϕ with $q \leq 50$.

define an Abelian group structure on these equivalence classes. The result is the K-theory of the operator algebra. The K-theory for $\mathcal{A}_\phi$ has been computed and for ϕ irrational it turns out to be isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$:

$$K_0(\mathcal{A}_\phi) = \mathbb{Z}[1] + \mathbb{Z}[P_\phi] \tag{12.407}$$

where 1 is the unit in $\mathcal{A}_\phi$ (certainly a projector!) and P_ϕ is a very non-obvious projector known as the Powers-Rieffel projector. It satisfies $\tau(P_\phi) = \phi$. Now suppose that ε is in a spectral gap, and let $P(\varepsilon)$ the the spectral projection of the self-adjoint operator H associated with the Borel set $(-\infty, \varepsilon]$. (See the discussion of the spectral theorem below.) Then one can label the gap by the integer n such that

$$[P(\varepsilon)] = m[1] + n[P_\phi] \tag{12.408}$$

Applying τ to this formula we obtain

♣Need to explain my τ is defined on the K-theory class. ♣

$$n\phi = \tau(P(\varepsilon))\mathrm{mod}1 \tag{12.409}$$

This determines n uniquely, if ϕ is irrational. Now, $\tau(P)$ for a projector P is a kind of regularized dimension, so in the case of rational ϕ, $\tau(P(\varepsilon)) = r$ should be the number of the gap in a successive labeling of the gaps in the direction of positive energy with

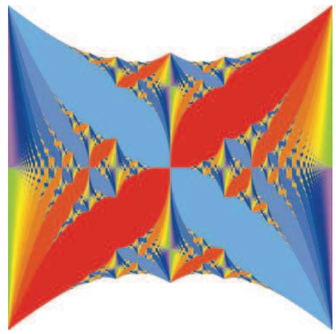


Figure 35: Gaps in the Hofstadter spectrum are colored in this figure, taken from [39]. Note the figure has been rotated by 90 degrees relative to the previous figure: Now ϕ is plotted on the horizontal axis. Different colors label gaps with fixed value of so that $\tau(P(\varepsilon)) = n\phi \text{mod} 1$.

the gap including $-\infty$ labeled by $r = 0$ and the gap including $+\infty$ labeled by $r = q$. Therefore, when $\phi = p/q$ is rational the equation becomes

$$n\frac{p}{q} = r \text{mod} 1 \tag{12.410}$$

that is we have a Diophantine equation

$$np - qs = r \tag{12.411}$$

It was shown in the extremely important paper by TKNN [48] that n has the interpretation of the quantized Hall conductance. Gaps of fixed values of n are plotted in Figure 35.

For more about this see [4, 39].

3. Is the above phenomena experimentally observable? Hoftstadter's original paper noted that $\phi = \Phi/\Phi_0$ of order one would require unreasonably large magnetic fields in traditional solid state setups. For example, the lattice constant for diamond at $T = 300K$ is about $4\text{Å} = 4 \times 10^{-10}m$. Recalling that $\Phi_0 \sim 4 \times 10^{-15}\text{Tesla} \cdot \text{meter}^2$ we see that $\phi = 1$ requires a magnetic field of about $25,000$ Tesla. However, in two different ways this obstacle has been overcome in the past two years. Some groups have used tricks with graphene and claimed to find experimental confirmation.

Another set of experiments makes use of one of the very interesting experimental developments of the past 20 years - the realization of BEC and optical lattices. The Hamiltonian (12.377) is used to describe fermionic atoms in an optical lattice. Typical lattice lengths are that of optical wavelengths, hence thousands of Angstroms. Hence, in optical lattices one can obtain values of ϕ that would normally require $10^4 - 10^5$ Tesla in conventional solid state setups. (In the optical lattices one does not literally use a magnetic field. Rather one uses a "synthetic magnetic field.") ♣ADD MANY REFS ♣

Exercise *Symmetries of the butterfly*
Observe that the Hofstadter butterfly has a four-fold symmetry.
Why is that?

12.7.10 The Effective Topological Field Theory

Describe σ_{xy} and QHE.

Laughlin argument

Abelian CS.

12.8 Deforming The Algebra Of Functions On $\mathbb{R}^{2n}$

12.8.1 The Moyal (or $*$) Product

As a second example of interesting algebras realized as operator algebras we consider a deformation of the algebra of functions on $\mathbb{R}^{2n}$. To begin with we will work quite formally, and then state the precise class of functions at the end (as well as a precise definition of the meaning of "deformation of the algebra").

Let Θ^{ij} be a $2n \times 2n$ constant, antisymmetric, nondegenerate matrix. One can then define the "$*$-product" for multiplying two functions on $\mathbb{R}^{2n}$ via the formula: [40]

$$(f_1 *_\Theta f_2)(x) := \exp\left[\frac{\mathrm{i}}{2}\Theta^{ij}\frac{\partial}{\partial x_1^i}\frac{\partial}{\partial x_2^j}\right](f_1(x_1)f_2(x_2))\Big|_{x_1=x_2=x} \tag{12.412}$$

For examples, let us compute:

$$[x^i, x^j] := x^i *_\Theta x^j - x^j *_\Theta x^i = \mathrm{i}\Theta^{ij} \tag{12.413}$$

It is also very useful to introduce the plane-wave function:

$$e_k(x) := e^{\mathrm{i}kx} \tag{12.414}$$

Then one easily computes:

$$e_{k_1} *_\Theta e_{k_2} = e^{-\frac{\mathrm{i}}{2}k_{1,i}\Theta^{ij}k_{2,j}} e_{k_1+k_2} \tag{12.415}$$

[40] This is often referred to as the "Moyal product" although according to Wikipedia it was introduced earlier by Groenewald.

This defines an algebra structure on a suitable space of functions on $\mathbb{R}^{2n}$. We will call it $\mathcal{A}(\mathbb{R}^{2n}_{\Theta})$, once we specify a suitable class of functions below. (This class will certainly include the Schwarz functions $\mathcal{S}(\mathbb{R}^{2n})$ of exponentially rapid decrease.)

Here are some properties of this algebra. They are most easily proven after we have related it to Weyl quantization:

1. The algebra is noncommutative as is clear from (12.413).

2. The algebra is associative.

3. If we formally expand in Θ^{ij} then the algebra is a deformation of the usual commutative algebra structure of functions on $\mathbb{R}^{2n}$:

$$f_1 *_\Theta f_2 = f_1 \cdot f_2 + \frac{\mathrm{i}}{2}\Theta^{ij}\frac{\partial f_1}{\partial x^i}\frac{\partial f_2}{\partial x^j} - \frac{1}{8}\Theta^{ij}\Theta^{kl}\frac{\partial^2 f_1}{\partial x^i \partial x^k}\frac{\partial^2 f_2}{\partial x^j \partial x^l} + \cdots \tag{12.416}$$

4. If we include the unit function $\mathbf{1}(x) = 1$ in the algebra then it acts as a unit

$$\mathbf{1} * f = f * \mathbf{1} = f \tag{12.417}$$

5. The algebra has a trace:

$$\tau(f) = \mathcal{N}\int_{\mathbb{R}^{2n}} f(x) d^{2n}x \tag{12.418}$$

(where $\mathcal{N}$ can be any normalization constant) such that

$$\tau(f_1 *_\Theta f_2) = \tau(f_2 *_\Theta f_1) \tag{12.419}$$

(Note that if we include $\mathbf{1}$ in the algebra its trace is not finite.)

Note that we can define a derivation of $\mathcal{A}(\mathbb{R}^{2n}_{\Theta})$:

$$\partial_i f := -\mathrm{i}\Theta^{-1}_{ij}(x^j * f - f * x^j) = -\mathrm{i}[\Theta^{-1}_{ij}x^j, f] \tag{12.420}$$

Note that these "derivatives" do not actually commute:

$$[\partial_i, \partial_j] = \mathrm{i}\Theta^{-1}_{ij} \tag{12.421}$$

Note that the planewave e_k is an eigenstate of ∂_i:

$$\partial_i e_k = \mathrm{i}k_i e_k \tag{12.422}$$

Using this one can check that

$$\begin{aligned} (e_k * f * e_{-k})(x) &= \left(e^{-k_i\Theta^{ij}\partial_j}\right) f(x) \\ &= f(x^i - \Theta^{ij}k_j) \end{aligned} \tag{12.423}$$

12.8.2 The Dipole Model

We would like to interpret the formula (12.423) in heuristic physical terms. Set $n = 1$. Then equation (12.421) should remind us of a charged particle in two dimensions with a transverse magnetic field. Consider two 2D charged particles of mass m but of opposite electric charges moving in a constant magnetic field B with a harmonic potential between them. The action is

♣Ref. Bigatti-Susskind. Sheik-Jabbari. ♣

$$S = \int \left[\frac{1}{2} m \left(\| \dot{x}_1 \|^2 + \| \dot{x}_2 \|^2 \right) - V(x_1 - x_2) \right] dt + \oint_{\gamma_1} eA - \oint_{\gamma_2} eA \tag{12.424}$$

Let us choose symmetric gauge $A = \frac{1}{2} B \epsilon_{\mu\nu} x^\mu dx^\nu$, and - very importantly - consider the large B limit. Formally we take the mass $m \to 0$. Then the action becomes [41]

$$S = \int \left[\frac{B}{2} \epsilon_{\mu\nu} \left(\dot{x}_1^\mu x_1^\nu - \dot{x}_2^\mu x_2^\nu \right) - V(x_1 - x_2) \right] dt \tag{12.425}$$

Now, we assume that the interaction potential V is such that it binds the two particles into a composite system. Let us change variables to the center-of-mass and relative degrees of freedom:

$$X^\mu := \frac{1}{2}(x_1^\mu + x_2^\mu) \qquad\qquad \Delta^\mu := \frac{1}{2}(x_1^\mu - x_2^\mu) \tag{12.426}$$

so that the action becomes:

$$S = \int \left[2B \epsilon_{\mu\nu} \dot{X}^\mu \Delta^\nu - V(\Delta) + \frac{B}{2} \frac{d}{dt} (\epsilon_{\mu\nu} x_1^\mu x_2^\nu) \right] dt \tag{12.427}$$

Ignoring the total derivative term we see that the center-of-mass momentum is

$$\boxed{P_\mu = 2B \epsilon_{\mu\nu} \Delta^\nu} \tag{12.428}$$

Thus, the spatial extent of the dipole depends on the center-of-mass momentum. Therefore, the dipole as a single system will appear to have nonlocal interactions. If particle 1 interacts with an external potential $U(x_1)$ then, in terms of the center of mass and relative coordinates we have

$$U(x_1^\mu) = U(X^\mu + \Delta^\mu) = U(X^\mu - \frac{1}{2B} \epsilon^{\mu\nu} P_\mu) \tag{12.429}$$

We usually describe particles in momentum eigenstates by plane waves e^{ikx} and compute transition amplitudes (in the Born approximation) by taking Fourier transforms of the potential. The similarity of (12.429) to (12.423) suggests that we can think of the noncommutative plane waves as representing such low mass dipoles in a magnetic field.

Remark: Equation (12.428) is an absolutely crazy equation from the point of view of a physicist: Usually we think of momentum and length as inversely related - as in the

[41] This is really a specialization of an Abelian Chern-Simons action, if we drop the harmonic potential.

uncertainty principle. This leads to the usual picture of "UV" (for "ultraviolet") being related to short distances and high energies - the two limits are thought of as equivalent. Similarly, "IR" (for "infrared") is related to long distances and low energies. Again, the two limits are thought of as equivalent. However (12.428) relates a momentum P_μ linearly to a length Δ^μ. This, ultimately, leads to strange effects of "UV-IR mixing" in noncommutative field theory.

12.8.3 The Weyl Transform

We now relate the algebra $\mathcal{A}(\mathbb{R}^{2n}_\Theta)$ to standard formulae in quantum mechanics.

Since Θ^{ij} is nondegenerate we can make a linear change of variables so that it has the form

$$\Theta^{ij} = \hbar \begin{pmatrix} 0 & 1_{n\times n} \\ -1_{n\times n} & 0 \end{pmatrix} \tag{12.430}$$

Let us call the coordinates in this basis

$$(x^1, \dots, x^{2n}) = (q^a, p_a) = (q^1, \dots, q^n, p_1, \dots, p_n) \tag{12.431}$$

Then we have

$$q^a * p_b - p_b * q^a = [q^a, p_b] = \mathrm{i}\hbar\delta^a{}_b \tag{12.432}$$

Now we can appeal to the Stone-von Neumann theorem. Up to unitary isomorphism the unique unitary irreducible representation of this algebra on a Hilbert space is $\mathcal{H} := L^2(\mathbb{R}^n)$ with

$$(\hat{Q}^a\psi)(q) = q^a\psi \qquad a = 1, \dots n \tag{12.433}$$

$$(\hat{P}_a\psi)(q) = -\mathrm{i}\hbar\frac{\partial}{\partial q^a} \qquad a = 1, \dots n \tag{12.434}$$

This gives a nice perspective on the algebra $\mathcal{A}(\mathbb{R}^{2n}_\Theta)$.

For a function $f(q,p)$ on $\mathbb{R}^{2n}$ we define the *Weyl transform* to be the linear operator on $\mathcal{H}$ defined by

$$\mathbf{Weyl}(f) := \int_{\mathbb{R}^{2n}} \check{f}(u,v) S(u,v) \frac{d^n u d^n v}{(2\pi)^n} \tag{12.435}$$

$$\check{f}(u,v) := \int_{\mathbb{R}^{2n}} f(q,p) e^{\mathrm{i}(u\cdot q + v\cdot p)} \frac{d^n q d^n p}{(2\pi)^n} \tag{12.436}$$

$$S(u,v) := e^{-\mathrm{i}(u\widehat{Q} + v\widehat{P})} \tag{12.437}$$

where $\widehat{Q} = \rho(q)$ and $\widehat{P} = \rho(p)$ are operators on $\mathcal{H}$.

Note that if f is real then $\check{f}(u,v)^* = \check{f}(-u,-v)$ and hence,

$$\mathbf{Weyl}(f)^\dagger = \mathbf{Weyl}(f) \tag{12.438}$$

so when f is real $\mathbf{Weyl}(f)$ is, a symmetric operator, and, least formally, it is self-adjoint. (If f is such that $\mathbf{Weyl}(f)$ is a bounded operator then it is certainly self-adjoint.)

Note that for a plane-wave:

$$f(q,p) = e^{\mathrm{i}(\alpha q + \beta p)} \tag{12.439}$$

we have

$$\check{f}(u,v) = (2\pi)^n \delta(u+\alpha)\delta(v+\beta) \tag{12.440}$$

and hence

$$\mathbf{Weyl}(f) = S(-\alpha,-\beta) = e^{\mathrm{i}(\alpha\hat{Q}+\beta\hat{P})} \tag{12.441}$$

that is:

$$\boxed{\mathbf{Weyl}(e^{\mathrm{i}(\alpha q+\beta p)}) = e^{\mathrm{i}(\alpha\hat{Q}+\beta\hat{P})}} \tag{12.442}$$

Comparing with (12.415) we now deduce the key fact that (for the above skew-diagonalized Θ)

$$\mathbf{Weyl}(f_1 *_\Theta f_2) = \mathbf{Weyl}(f_1)\ \mathbf{Weyl}(f_2) \tag{12.443}$$

Thus, the Moyal product on $\mathbb{R}^{2n}$, interpreted as a symplectic manifold, is nothing but standard quantum mechanics in disguise.

<u>The Weyl transform of real analytic functions</u>: By expanding the exponentials on both sides of (12.442), using linearity, and matching powers of $\alpha^n\beta^m$ we deduce that the Weyl transform of a polynomial in q,p is the total symmetrization of that polynomial. Thus, for example:

$$\begin{aligned}
\mathbf{Weyl}(q) &= \hat{Q} \\
\mathbf{Weyl}(p) &= \hat{P} \\
\mathbf{Weyl}(qp) &= \frac{1}{2}(\hat{P}\hat{Q}+\hat{Q}\hat{P}) \\
\mathbf{Weyl}(q^2p) &= \frac{1}{3}(\hat{Q}^2\hat{P}+\hat{Q}\hat{P}\hat{Q}+\hat{P}\hat{Q}^2)
\end{aligned} \tag{12.444}$$

and so forth.

The same result can be obtained, somewhat more tediously, as follows: To compute $\mathbf{Weyl}(p)$ we compute the distribution

$$\check{f}(u,v) = -2\pi\mathrm{i}\delta(u)\delta'(v) \tag{12.445}$$

so

$$\mathbf{Weyl}(p) = \mathrm{i}\int \left(\delta(u)\delta(v)\partial_v S(u,v)\right) du dv = \hat{P} \tag{12.446}$$

and similarly $\mathbf{Weyl}(q) = \hat{Q}$. Alternatively, one can simply take derivatives of (12.442) with respect to α,β and then set them to zero. Now consider $f(q,p) = qp$. So: let us check:

$$\check{f}(u,v) = -2\pi\delta'(u)\delta'(v) \tag{12.447}$$

So

$$\mathbf{Weyl}(f) := -\int_{\mathbb{R}^{2n}} \delta(u)\delta(v)\partial_u\partial_v(S(u,v))dudv \tag{12.448}$$

We have to be careful about computing the derivatives of $S(u,v)$ before evaluating the delta functions. One way to do this is to write

$$S(u,v) = e^{\frac{\mathrm{i}}{2}uv\hbar}e^{-\mathrm{i}u\hat{Q}}e^{-\mathrm{i}v\hat{P}} \tag{12.449}$$

Now, thanks to the delta-functions, the two derivatives act on the c-number prefactor or on the operator to give

$$\mathbf{Weyl}(f) = -\frac{\mathrm{i}}{2}\hbar + \hat{Q}\hat{P} = \frac{1}{2}(\hat{Q}\hat{P} + \hat{P}\hat{Q}) \tag{12.450}$$

as expected. For a general real-analytic function we need the result for $f(q,p) = q^n p^m$. Now

$$\check{f} = (2\pi)(-\mathrm{i})^{n+m}\delta^{(n)}(u)\delta^{(m)}(v) \tag{12.451}$$

$$\mathbf{Weyl}(f) = (\mathrm{i})^{n+m}\int \delta(u)\delta(v)\left(\partial_u^n\partial_v^m S(u,v)\right) dudv \tag{12.452}$$

Now, to compute the derivatives we have to be a bit more careful. Recall that for <u>any</u> family of operators $\mathcal{O}(t)$ we have

$$\frac{d}{dt}e^{\mathcal{O}(t)} = \int_0^1 e^{(1-s)\mathcal{O}(t)}\frac{d}{dt}\mathcal{O}(t)e^{s\mathcal{O}(t)}ds \tag{12.453}$$

Remarks

1. It is useful to note that

$$\mathrm{Tr}_{\mathcal{H}} S(u',v')S(u,v)^{-1} = \hbar^{-n}e^{\frac{\mathrm{i}\hbar}{2}(u'v'-uv)}\delta(u-u')\delta(v-v') \tag{12.454}$$

This is easily proved by inserting complete sets of states and recalling (12.462). Therefore

$$\mathrm{Tr}_{\mathcal{H}}\left(\ \mathbf{Weyl}(f)S(u,v)^{-1}\right) = (2\pi\hbar)^{-n}\check{f}(u,v) \tag{12.455}$$

Therefore

$$\int (f_1(q,p))^* f_2(q,p)\frac{d^nqd^np}{(4\pi^2\hbar)^n} = \mathrm{Tr}\ \mathbf{Weyl}(f_1)^\dagger\ \mathbf{Weyl}(f_2) \tag{12.456}$$

Note that from (12.456) it follows that $f \in L^2(\mathbb{R}^{2n})$ maps to the space of Hilbert-Schmidt operators. The space of Hilbert-Schmidt operators is a Hilbert space, and hence complete <u>in the Hilbert-Schmidt norm</u>. Thus one rigorous way of defining $\mathcal{A}(\mathbb{R}^{2n}_\Theta)$ would be to take the completion of the Schwarz space in the Hilbert-Schmidt norm. The resulting algebra is the space of Hilbert-Schmidt operators.

♣Say what happens if you take the norm-closure of the Hilbert-Schmidt operators. $\parallel T \parallel \leq \parallel T \parallel_2$ so the norm closure will be bigger. ♣

2. Note too that evaluation of (12.455) at $u = v = 0$ gives the trace

$$\mathrm{Tr}_{\mathcal{H}}\left(\ \mathbf{Weyl}(f)\right) = \int_{\mathbb{R}^{2n}} f(q,p)\frac{d^nqd^np}{(4\pi^2\hbar)^n} = \tau(f) \tag{12.457}$$

Exercise
Compute the Moyal product of two Gaussian wavepackets.

Exercise

Show that in the position space representation

♣Need to check this formula. ♣

$$(\mathbf{Weyl}(f)\psi)(q) = \int K(q,q')\psi(q')d^n q' \tag{12.458}$$

$$\begin{aligned} K(q,q') &= \int f(\frac{q+q'}{2},p)e^{-\mathrm{i}\frac{p}{\hbar}(q-q')}\frac{d^n p}{(2\pi\hbar)^n} \\ &= \int \check{f}(v,\frac{q-q'}{2})e^{-\mathrm{i}\frac{v}{2}(q+q')}\frac{d^n v}{(2\pi\hbar)^n} \end{aligned} \tag{12.459}$$

12.8.4 The Wigner Function

The inverse of a Weyl transform defines the *Wigner function* associated to a linear operator $\widehat{T} \in \mathcal{L}(\mathcal{B})$.

$$\mathbf{Wigner}(\mathbf{Weyl}(f)) = f \tag{12.460}$$

The explicit formula is that $\mathbf{Wigner}(\widehat{T})$ is the function on phase space:

$$\mathbf{Wigner}(\widehat{T})(q,p) := \int \langle q - \frac{1}{2}v|\widehat{T}|q+\frac{1}{2}v\rangle e^{\mathrm{i}\frac{p}{\hbar}v}d^n v \tag{12.461}$$

To prove (12.464) note that it suffices to prove it for $\hat{T} = e^{\mathrm{i}(\alpha\hat{Q}+\beta\hat{P})}$. Now, recall that

$$\langle q|p\rangle = \frac{1}{\sqrt{(2\pi\hbar)^n}}e^{\mathrm{i}qp/\hbar} \tag{12.462}$$

Then compute:

$$\begin{aligned} \int \langle q-\frac{1}{2}v|e^{\mathrm{i}(\alpha\hat{Q}+\beta\hat{P})}|q+\frac{1}{2}v\rangle e^{\mathrm{i}\frac{p}{\hbar}v}d^n v &= \int e^{\mathrm{i}\alpha\beta\hbar/2}\langle q-\frac{1}{2}v|e^{\mathrm{i}\alpha\hat{Q}}e^{\mathrm{i}\beta\hat{P}}|q+\frac{1}{2}v\rangle e^{\mathrm{i}\frac{p}{\hbar}v}d^n v \\ &= \int e^{\mathrm{i}\alpha\beta\hbar/2}\langle q-\frac{1}{2}v|e^{\mathrm{i}\alpha\hat{Q}}|p'\rangle\langle p'|e^{\mathrm{i}\beta\hat{P}}|q+\frac{1}{2}v\rangle e^{\mathrm{i}\frac{p}{\hbar}v}d^n v d^n p' \\ &= (2\pi\hbar)^{-n}\int e^{\mathrm{i}\alpha\beta\hbar/2}e^{\mathrm{i}\alpha(q-\frac{1}{2}v)+\mathrm{i}pv/\hbar}e^{\mathrm{i}p'(\beta-v/\hbar)}d^n v d^n p' \\ &= e^{\mathrm{i}(\alpha q+\beta p)} \end{aligned} \tag{12.463}$$

So the equation is true for plane-waves. But this is enough by linearity and Fourier transform. Of course, an analogous argument also gives

$$\mathbf{Wigner}(\widehat{T})(q,p) := \int \langle p+\frac{1}{2}u|\widehat{T}|p-\frac{1}{2}u\rangle e^{\mathrm{i}\frac{q}{\hbar}u}d^n u \tag{12.464}$$

Wigner introduced his transform to attempt to associate a classical probability distribution on phase space with a quantum state. In order to do this we should apply the Wigner transform to a density matrix (i.e. a quantum state). In this case the operator

$\widehat{T} = \rho$ is a positive self-adjoint operator of trace one. Thus the Wigner function will be real. For example, for a pure state, a line through $\psi \in \mathcal{H}$ we have a rank one projector

$$\rho = P_\psi = |\psi\rangle\langle\psi| \tag{12.465}$$

(we assume $\langle\psi|\psi\rangle = 1$). Then the corresponding Wigner function is

$$\varpi_\psi(q,p) = \int \psi^*(q - \frac{1}{2}v)\psi(q + \frac{1}{2}v)e^{\mathrm{i}pv/\hbar}dv \tag{12.466}$$

Note that

$$\int dp\varpi_\psi(q,p) = \psi^*(q)\psi(q) \tag{12.467}$$

and one can show that similarly

$$\int dq\varpi_\psi(q,p) = \hat{\psi}^*(p)\hat{\psi}(p) \tag{12.468}$$

Indeed, if we have any two states then the overlap is

$$|\langle\psi_1|\psi_2\rangle|^2 = \mathrm{Tr}_{\mathcal{H}} P_{\psi_1} P_{\psi_2} = \int P_{\psi_1}(q,p)P_{\psi_2}(q,p)dqdp \tag{12.469}$$

But this means that if ψ_1 and ψ_2 are orthogonal then

$$\int P_{\psi_1}(q,p)P_{\psi_2}(q,p)dqdp = 0 \tag{12.470}$$

Hence we conclude that $\varpi_\psi(q,p)$ cannot always be positive and do not represent true probability distributions.

It is amusing to work out the Wigner functions in some simple cases. For the harmonic oscillator groundstate $P = |0\rangle\langle 0|$

$$\varpi_\psi(q,p) \sim \exp[-p^2 - x^2] \tag{12.471}$$

♣Should also do first excited state ♣

Moreover, if ψ is <u>any</u> state and we evolve it as a function of time using the harmonic oscillator Hamiltonian $H = \frac{1}{2}(p^2 + x^2)$ then, remarkably, we have

$$\varpi_{\psi(t)}(q,p) = \varpi_\psi(q(t),p(t)) \tag{12.472}$$

where $(q(t),p(t))$ is the classical trajectory in phase space. For some nice pictures of Wigner functions of various standard quantum states see http://www.iqst.ca/quantech/wiggalery.php.

Remark: The Wigner functions satisfy some peculiar properties [8]. Apply the definition (12.466) for a normalized function ψ. Now apply the Cauchy-Schwarz inequality to obtain

$$|\varpi_\psi(q,p)| \leq 2 \tag{12.473}$$

This is another feature we would not expect from a general probability distribution on phase space. Moreover, if we consider the value at the origin then

$$\varpi_\psi(0,0) = \pm 2 \tag{12.474}$$

if $\psi(q)$ is an even or odd function of q, respectively. Note that this is true even if ψ represents far separated wavepackets whose value at $q = 0$ is nearly zero.

****

SHOULD DO THE EXAMPLES OF COHERENT STATES AND SQUEEZED STATES. FIGURE OF H.O. TIME EVOLUTION OF COHERENT STATE. (COHERENT STATE OPERATOR AS TRANSLATION OPERATOR IN PHASE SPACE.

****

12.8.5 Field Theory On A Noncommutative Space

Since fields are functions on a spacetime, it is natural to try to generalize field theory on spacetime to field theory on a spacetime like $\mathbb{R}^{2n}_{\Theta}$. One can take one's favorite Lagrangian field theory and every time a product of fields

$$\Phi_1(x)\Phi_2(x) \tag{12.475}$$

appears in the action density, one simply replaces it by the Moyal product

$$\Phi_1(x) * \Phi_2(x) \tag{12.476}$$

Then one "integrates" using a trace τ on the algebra. For the Moyal algebra this is just an ordinary integral over $\mathbb{R}^{2n}$. The resulting theories share many of the characteristics of ordinary field theory. This is extremely surprising! In general if one introduces derivative interactions with arbitrarily high numbers of derivatives then the resulting field theory is very badly behaved: The Cauchy problem does not make sense and the quantum perturbation theory is badly behaved. Remarkably, if one introduces nonlocality in the controlled way given by the Moycal product, the resulting theories are relatively well-behaved. There is a curious "mixing between IR and UV."

See the review [15] for more details. ♣Should say more here. ♣

12.9 Relation To Open String Theory

We now give a perspective on Moyal quantization following from string theory. This perspective also gives significant insight into the fact that there is an isomorphism of C^* algebras $\mathcal{A}_{\theta_1}$ and $\mathcal{A}_{\theta_2}$ when θ_1 and θ_2 are related by integer fractional linear transformations.

The physical interpretation uses the theory of open strings moving in a target space with constant metric and B-field. In a certain limit where the string length goes to zero. The open string vertex operator algebra of a brane wrapped on a torus approaches the algebra $\mathcal{A}_\theta$, in a way analogous to the degeneration of the VOA of the WZW model for $k \to \infty$. See, for example, [9, 46, 47] and the review [15] and references therein. We mostly follow the definitive version from Seiberg and Witten.

12.9.1 String Theory In A p-nutshell

We consider a bosonic perturbative string theory. It is based on maps

$$x : \Sigma \to \mathcal{X} \tag{12.477}$$

where the *worldsheet* Σ is a two-dimensional surface equipped with a metric, and the *target space* $\mathcal{X}$, is equipped with a metric g, and other geometrical structures. For our purposes we can consider it to be equipped with a globally defined two-form b. (More generally, b is part of a "gerbe connection.") There is an important issue of the signature of the worldsheet and target space metrics. If we think about strings propagating in spacetime then both the worldsheet and target space metrics should have Lorentzian signature metrics. However, it is technically much more convenient to take the target space to have Euclidean signature, and we will do so. We begin with Lorentzian signature worldsheet metric and then Wick rotate to Euclidean signature worldsheet metric.

Remark: In string theory one integrates over the space of Riemann surfaces and over maps. The integration over Riemann surfaces includes a sum over topologies. When the worldsheet has Lorentzian signature this leads to singularities, so in string perturbation theory a Euclidean signature is always assumed. When the target space is Lorentzian and has nontrivial time-dependence in the geometry many new subtleties arise that the subject is not completely understood.

We will focus on *oriented* string theory. Thus Σ will be assumed to have an orientation. The string action, entering the path integral as $e^{\mathrm{i}S_L/\hbar}$ is:

♣And we were not careful about that orientation. For self-duality $d\sigma \wedge d\tau$ is best with Lorentzian metric $-d\tau^2 + d\sigma^2$. ♣

$$S_L[h, x; g, b] = -\frac{1}{4\pi\ell_s^2}\int_\Sigma \Big[(dx, *dx) - x^*(b)\Big] \tag{12.478}$$

where the Hodge star uses the metric on Σ and $(\cdot,\cdot)$ is contraction in the metric on $\mathcal{X}$. Upon analytic continuation to Euclidean signature worldsheet metric the Euclidean action, entering the path integral as $e^{-S_E/\hbar}$ is:

$$S_E[h, x; g, b] = \frac{1}{4\pi\ell_s^2}\int_\Sigma \Big[(dx, *dx) - \mathrm{i}x^*(b)\Big] \tag{12.479}$$

where the worldsheet metric is positive definite and oriented (so that we can integrate $x^*(b)$). With positive definite target space metric the path integral is at least formally convergent. Written in terms of local coordinates the action is:

$$S_E[h, x; g, b] = \frac{1}{4\pi\ell_s^2}\int_\Sigma \Big[\sqrt{h}h^{ab}g_{ij}(x(\xi))\partial_a x^i \partial_b x^j - \mathrm{i}b_{ij}(x(\xi))\epsilon^{ab}\partial_a x^i \partial_b x^j\Big] d^2\xi \tag{12.480}$$

Remarks:

1. The parameter ℓ_s determines a fundamental length-scale, called the *string length.* In much of the string theory literature one finds instead

$$\ell_s^2 = \alpha' \tag{12.481}$$

There is much evidence that, in string theory, the nature of space and time changes dramatically at length scales on the order of ℓ_s. Note that $1/\ell_s^2$ has units of $[\text{Length}]^{-2}$. Since we have set $\hbar = c = 1$ this is the same as [Energy]/[Length]. Indeed, $T \sim 1/\ell_s^2$ has the interpretation of the *string tension*.

2. Classically, the action only depends on the conformal structure, and this is also true quantum mechanically, if $\dim\mathcal{X} = 26$ and the metric and b-field are flat. We will not need to worry about the quantum Weyl anomaly for the following discussion, so we leave the dimension D of the target unspecified. This is fine for tree level string theory on flat worldsheets.

♣Should say more here and mention Mumford isomorphism. ♣

We now specialize considerably to the case of a target space

$$\mathcal{X} = \mathbb{R}^{2n} \times \mathbb{R}^{D-2n} \tag{12.482}$$

We assume a constant metric and b-field, and moreover we assume that they respect the product structure. We write the metric as:

$$g = \sum_{1\leq i,j\leq 2n} g_{ij} dx^i \otimes dx^j + \sum_{2n<i,j} g_{ij} dx^i \otimes dx^j \tag{12.483}$$

and the constant b-field as

$$b = \sum_{1\leq i,j\leq 2n} b_{ij} dx^i \wedge dx^j . \tag{12.484}$$

Remark: The theory we are studying is a special case of the WZW model where the Lie group target space is just the abelian group $\mathbb{R}^D$. In this case the WZ term is just the b-field action, and we have the freedom to add a flat gerbe connection. The string tension ℓ_s^{-2} plays the role of the level k.

First consider the Lorentz-signature worldsheet $h = -d\tau^2 + d\sigma^2$ where Σ is the strip $\mathbb{R} \times [0,\pi]$. The action is:

$$S_L = \frac{1}{4\pi\ell_s^2} \int_{\mathbb{R}} d\tau \int_0^{\pi} d\sigma \left[g_{ij}(\partial_\tau x^i \partial_\tau x^j - \partial_\sigma x^i \partial_\sigma x^j) + 2b_{ij}\partial_\tau x^i \partial_\sigma x^j \right] \tag{12.485}$$

The equation of motion is simply the free wave equation

♣Here we make a choice of orientation. ♣

$$(-\partial_\tau^2 + \partial_\sigma^2) x^i = 0 \tag{12.486}$$

The quantization of the free quantum field theory defined by the action (12.485) is relatively straightfoward and well-known textbook material. We review it briefly. First we compute the momentum density of the string:

$$p_i(\sigma) := \frac{\delta S}{\delta \dot{X}^i(\sigma)} = \frac{1}{2\pi\ell_s^2} \left(g_{ij}\partial_\tau x^j + b_{ij}\partial_\sigma x^j \right) \tag{12.487}$$

The phase space thus consists of maps $(x, P) : [0, \pi] \rightarrow T^*\mathcal{X}$. The naive symplectic structure is, formally:

$$\omega = \int_0^\pi \delta x^i(\sigma) \wedge \delta p_i(\sigma) d\sigma \tag{12.488}$$

but the presence of the b-field can alter this naive formula.

♣Give the correct *omega*. Derive it from first principles. Explain relation to "second class constraints." ♣

Now we come to the important consideration of boundary conditions. For *closed strings* we would take $x^i(\sigma)$ and $p_i(\sigma)$ to be maps from a spatial circle S^1. For the open string, by conformal invariance we can, and will take the spatial domain to be the interval $[0, \pi]$.

In general perturbation theory, for open strings we take Σ to be a Riemann surface with boundary. The boundary values of the fields must lie in a Lagrangian subspace of phase space defined (for Euclidean signature worldsheet) by

$$\delta x^i \left(g_{ij} \partial_n x^j + \mathrm{i} b_{ij} \partial_{\|} x^j \right) |_{\delta\Sigma} = 0 \tag{12.489}$$

Here ∂_n and $\partial_{\|}$ are normal and tangential derivatives on the boundary and there is a sum over $i = 1, \ldots, D$.

Without trying to find the most general boundary condition, the most obvious choice is to set $\delta x^i|_\partial = 0$ or $\left(g_{ij} \partial_n x^j + \mathrm{i} b_{ij} \partial_{\|} x^j \right) |_\partial = 0$. We can make different choices for different values of i and different choices for different connected components of the boundary. There are clearly many choices one can make here. They correspond to very different physical situations.

We will make the choice that

$$\left(g_{ij} \partial_\sigma x^j + b_{ij} \partial_\tau x^j \right) |_{\sigma=0,\pi} = 0 \qquad\qquad i \le 2n \tag{12.490}$$

and

$$\delta x^i|_{\sigma=0,\pi} = 0 \qquad\qquad 2n < i \tag{12.491}$$

The physical interpretation is that there is a $2n$-dimensional hyperplane, defined by $x^i = x^i_0$, on which the open string endpoints are confined. This hyperplane should be thought of as dynamical: It can wiggle and have waves moving on it: It is a "brane" with $2n$ spacetime dimensions. [42]

♣More on this. FIGURE ♣

One very nice viewpoint on phase space is that it is the space of (gauge invariant) solutions to the equations of motion [13]. In this free field theory it is straightforward to give the general solution to the equations of motion and find linear coordinates on it.

We will now say a little about the quantization. Focussing on the fields x^i with $i \le 2n$. We will parametrize phase space by finding the general solution of the equation of motion: We can separate the solution into the zero- and nonzero-Fourier modes: $x^i = x^i_{\rm lin} + x^i_{\rm osc}$ where

$$x^i_{\rm lin} = x^i_0 + L^i_1 \tau + L^i_2 \sigma \tag{12.492}$$

and

$$x^i_{\rm osc} = \sum_{n \neq 0} x^i_{(n)}(\sigma) e^{in\tau} \tag{12.493}$$

[42] The physics terminology originated from the joke that branes with p spatial dimensions should be called *p-branes*. Thus, a 0-brane is a particle, a 1-brane is a string, a 2-brane is a (dynamical) surface, and so on.

where $x^i_n(\sigma)$ is a linear combination of $\cos(n\sigma)$ and $\sin(n\sigma)$. Recall that the Schrödinger equation is $\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = H\Psi$. Thus, $n > 0$ corresponds to negative energies and $n < 0$ corresponds to positive energies.

To obtain the linear piece we need two equations for L_1 and L_2. The boundary conditions determine

$$g_{ij}L_2^j + b_{ij}L_1^j = 0 \tag{12.494}$$

To get the second equation, define

$$p_i := \int_0^\pi p_i(\sigma)d\sigma = \frac{1}{2\ell_s^2}(g_{ij}L_1^j + b_{ij}L_2^j) \tag{12.495}$$

we now have two linear equations for L_1 and L_2 and we find:

$$x^i_{\mathrm{lin}} = x^i_0 + 2\ell_s^2(G^{ij}p_j\tau - (g^{-1}bG^{-1})^{ij}p_j\sigma) \tag{12.496}$$

Where it is convenient to define a matrix G_{ij} by

$$G := g - bg^{-1}b \tag{12.497}$$

or, with indices:

$$G_{ij} = g_{ij} - b_{ik}g^{kl}b_{lj} \tag{12.498}$$

Then G^{ij} are the matrix elements of the inverse G^{-1}. Note that

$$G^{ij} = \left(\frac{1}{g+b}\right)^{ij}_{\text{symmetric}} \tag{12.499}$$

Similarly, the oscillator piece is determined by the boundary conditions to have the form:

$$x^i_{(n)}(\sigma) = \ell_s\left(\mathrm{i}\delta^i{}_j\cos(n\sigma) + (g^{-1}b)^i{}_j\sin(n\sigma)\right)\frac{\alpha^j_n}{n} \tag{12.500}$$

The on-shell momentum density now turns out to be

$$p_i(\sigma,\tau) = p_i - \frac{1}{2\ell_s}\sum_{n\neq 0}G_{ij}\alpha^j_n\cos(n\sigma)e^{\mathrm{i}n\tau} \tag{12.501}$$

These equations give a complete set of linear coordinates on the subspace of phase space determined by the boundary conditions.

With a suitable symplectic form the quantization is given by:

♣Need to explain this symplectic form more thoroughly. ♣

$$[x^i_0, p_j] = \mathrm{i}\delta^i{}_j \qquad\qquad [\alpha^i_n, \alpha^j_m] = G^{ij}n\delta_{n+m,0} \tag{12.502}$$

The quantization of the harmonic oscillators is then standard. One chooses the representation based on the vacuum line with $\alpha^i_n|0\rangle = 0$ for $n > 0$. Thus, the free one-string Hilbert space is just

$$\mathcal{H}_{1-\mathrm{string}} = L^2(\mathbb{R}^{2n})\otimes\mathcal{F} \tag{12.503}$$

where $\mathcal{F}$ is the Fock space for an infinite set of harmonic oscillators $\alpha_{\pm n}$ of frequency n.

Remarks

1. The normalization of the amplitude using α_n^j turns out to be very convenient in the quantization. The fluctuations are order one. So the oscillator modes are fluctuating on the order of a string length. Meanwhile, for small string length, the momentum of the oscillator modes is large: These oscillations cost a lot of energy.

2. Note that there is a very nice intuitive interpretation of (12.504). Let us write the equation as

$$\begin{aligned} g_{ij}x^j_{(n)}(\sigma,\tau) &= \frac{\mathrm{i}\ell_s}{2}\left((g+b)_{ij}e^{\mathrm{i}n(\tau-\sigma)} + (g-b)_{ij}e^{\mathrm{i}n(\tau+\sigma)}\right)\frac{\alpha_n^j}{n} \\ &= \frac{\mathrm{i}\ell_s}{2}(g+b)_{ij}\left(\delta^j{}_k e^{\mathrm{i}n(\tau-\sigma)} + (\mathcal{R})^j{}_k e^{\mathrm{i}n(\tau+\sigma)}\right)\frac{\alpha_n^k}{n} \end{aligned} \tag{12.504}$$

   Written this way it is apparent that the left- and right-moving waves $e^{\mathrm{i}n(\tau\pm\pm\sigma)}$ are reflected with a matrix

$$\mathcal{R} = (g+b)^{-1}(g-b) \tag{12.505}$$

   One can check that, with respect to the metric g, this is an orthogonal matrix. That is:

$$\mathcal{R}^{tr}g\mathcal{R} = g \tag{12.506}$$

   One could diagonalize $g^{1/2}\mathcal{R}g^{-1/2}$. In that basis left and right-moving waves are reflected with a phase.

The formulation in terms of oscillators is conceptually important but can become computationally very messy. It is better to use Green's functions on Euclidean worldsheets. We Wick rotate the strip worldsheet to Euclidean signature with coordinate $w = \sigma + \mathrm{i}\tau$. Next we make a conformal transformation to the upper half-plane:

$$z = e^{\mathrm{i}(\sigma+\mathrm{i}\tau)} \tag{12.507}$$

Thus, strings in the far past are near $z = 0$ and time-ordering becomes radial ordering. The equal τ spatial slices become semicircles as in Figure 36.

The Euclidean action is

$$S = \frac{1}{4\pi\ell_s^2}\int_\Sigma \left(h^{ab}g_{ij}\partial_a x^i\partial_b x^j - \mathrm{i}b_{ij}dx^i\wedge dx^j\right) \tag{12.508}$$

Here h_{ab} is a Riemannian metric on Σ.

The boundary conditions are, once again:

$$\left(g_{ij}\partial_n x^j + \mathrm{i}b_{ij}\partial_t x^j\right)|_{\delta\Sigma} = 0 \tag{12.509}$$

♣Need to give some excuses for the fact that real fields cannot solve this boundary condition. ♣

When $\sigma = 0, \pi$, τ is a real coordinate along the boundary of the strip. Correspondingly $\tilde\tau = \pm\exp[-\tau]$ is a real coordinate along the positive and negative real axis. For simplicity of notation we henceforth denote

$$\tilde\tau \to \tau. \tag{12.510}$$

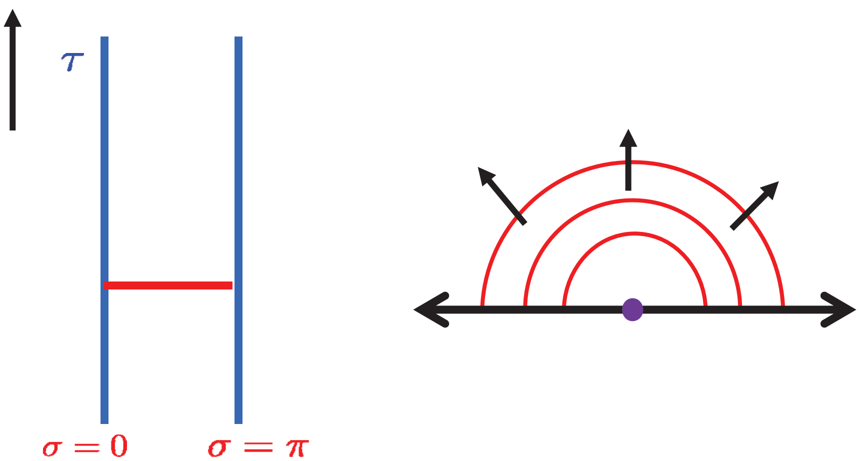


Figure 36: Left: An open string worldsheet represented as a strip: The spatial coordinate σ runs from 0 to π. The time direction τ is upward. The Euclidean worldsheet has the same picture but now we view the strip as a domain in the complex plane with $w = \sigma + i\tau$. A conformal transformation of this domain $z = \exp[iw]$ maps it to the upper half-plane. The dot represents $z = 0$. This is the infinite past on the string worldsheet. The red semicircles are equal time slices.

We want to consider the OPE of the open string vertex operators $V_k(\tau) =: e^{ikx(\tau)}$:. Here the normal-ordering symbols refer to that relevant to the oscillators $\alpha^i_{\pm n}$ acting on $\mathcal{H}_{1-\mathrm{string}}$. In this free field theory all we need is the Green's function for the x-field restricted to the boundary since

$$V_{k_1}(\tau_1)V_{k_2}(\tau_2) = e^{-k_{1,i}k_{2,j}\langle x^i(\tau_1)x^j(\tau_2)\rangle} : e^{ik_{1,i}x^i(\tau_1)+ik_{1,i}x^i(\tau_1)} : \tag{12.511}$$

where the colons denote normal-ordering symbols, and we are working in radial quantization around $z = 0$.

Now, to compute the Green's function $\langle x^i(\tau_1)x^j(\tau_2)\rangle$ we will compute the Green's function $\langle x^i(z_1)x^j(z_2)\rangle$ and then take the boundary values $z_1 = \tau_1$ and $z_2 = \tau_2$. The latter Green's function is the solution of the equations of motion:

$$\partial_1\bar{\partial}_1\langle x^i(z_1)x^j(z_2)\rangle = \partial_2\bar{\partial}_2\langle x^i(z_1)x^j(z_2)\rangle = 0 \tag{12.512}$$

with logarithmic singularity at $z_1 = z_2$, at most log growth at infinity and satisfying the boundary conditions (12.509), which in our special case become

$$\left(g_{ij}(\partial - \bar{\partial})x^j + b_{ij}(\partial + \bar{\partial})x^j\right)|_{\partial\Sigma} = 0 \tag{12.513}$$

We therefore have

$$\langle x^i(z_1)x^j(z_2)\rangle = A_1^{ij}\log(z_1-z_2)+A_2^{ij}\log(z_1-\bar z_2)+A_3^{ij}\log(\bar z_1-z_2)+A_4^{ij}\log(\bar z_1-\bar z_2)+A_5^{ij} \tag{12.514}$$

where A_α^{ij} are matrix functions of g_{ij} and b_{ij}, constant as functions of z_1, z_2. These matrices are determined by the boundary conditions. Solving and taking the boundary value of the propagator we get:

$$\langle x^i(\tau)x^j(\tau')\rangle = -\frac{\ell_s^2 G^{ij}}{2\pi}\log(\tau-\tau')^2 + \frac{\mathrm{i}}{2}\Theta^{ij}\mathrm{sign}(\tau-\tau') + A_5^{ij} \tag{12.515}$$

and

$$\begin{aligned} G^{ij} &= \left(\frac{1}{g+b}\right)^{ij}_{\mathrm{symm}} \\ \Theta^{ij} &= \ell_s^2\left(\frac{1}{g+b}\right)^{ij}_{\mathrm{anti-symm}} \end{aligned} \tag{12.516}$$

The matrix A_5^{ij} will not play an important role and can be dropped.

Then, applying (12.511) we find that for $\tau_1 > \tau_2$:

$$V_{k_1}(\tau_1)V_{k_2}(\tau_2) \sim (\tau_1-\tau_2)^{\ell_s^2 G^{ij}k_{1,i}k_{2,j}/\pi}e^{-\frac{\mathrm{i}}{2}\Theta^{ij}k_{1,i}k_{2,j}}V_{k_1+k_2}(\tau_2) + \mathcal{O}(\tau_1-\tau_2) \tag{12.517}$$

and for $\tau_1 < \tau_2$:

$$V_{k_1}(\tau_1)V_{k_2}(\tau_2) \sim (\tau_1-\tau_2)^{\ell_s^2 G^{ij}k_{1,i}k_{2,j}/\pi}e^{\frac{\mathrm{i}}{2}\Theta^{ij}k_{1,i}k_{2,j}}V_{k_1+k_2}(\tau_2) + \mathcal{O}(\tau_1-\tau_2) \tag{12.518}$$

The prefactor depending on G^{ij} determines "anomalous dimensions." It is part of the fairly complicated story of vertex operator algebras. However, one can take a limit in which that story reduces to the Moyal product of functions. In order to have ordinary functions V_k should have dimension zero and therefore we want a limit so that

$$\ell_s^2 G^{ij} \to 0 \tag{12.519}$$

while Θ^{ij} remains order one.

Definition [*Seiberg-Witten limit*]. Let B_{ij} be $2n\times 2n$ antisymmetric and invertible. Introduce $\ell_0, g_{0,ij}$ and set:

$$\ell_s^2 = \epsilon^{1/2}\ell_0^2 \tag{12.520}$$

$$g_{ij} = \epsilon g_{0,ij} \tag{12.521}$$

$$b_{ij} := \ell_s^2 B_{ij} \tag{12.522}$$

The *Seiberg-Witten limit* is defined by the limit $\epsilon\to 0$ holding $g_{0,ij}, \ell_0^2$ and B_{ij} fixed.

In the SW limit we can expand the expressions for G^{ij} and Θ^{ij} as series in g/B:

$$(g+b)^{-1} = b^{-1} - b^{-1}gb^{-1} + \cdots \tag{12.523}$$

we find that in the SW limit:

$$G^{ij} \to -(B^{-1} g_0 B^{-1})^{ij} \qquad\qquad \Theta^{ij} = (B^{-1})^{ij} \tag{12.524}$$

(Note that G^{ij} is positive definite, and the minus sign is required so that the expression is positive definite.)

In this limit $V_k(\tau)$ has zero conformal dimension, and there is no singularity in the OPE. This means:

1. We can now safely take $\tau_1 \to \tau_2$, suppressing the contribution to the OPE of all the oscillator modes. The resulting algebra is independent of τ and is isomorphic to the Moyal algebra.

2. The can represent ordinary functions on the target space manifold $\mathcal{X}$. More precisely, In string field theory the operators $V_k(\tau)$ multiply the momentum modes of a field on spacetime:

$$\Psi = \int dk T(k) V_k(\tau) \tag{12.525}$$

   where Ψ is the string field. By deducing the couplings of this field theory from the string S-matrix we will discover it is a noncommutative field theory. Actually, when this is done more properly, what one finds is noncommutative gauge theory on space time. (It is the gauge theory modes have have zero mass and survive the SW limit.)

There is a very nice connection to the dipole picture of Section §12.8.2 above. In the SW limit the string action becomes

$$S = -\frac{\mathrm{i}}{4\pi} \int_\Sigma x^*(B) = -\frac{\mathrm{i}}{4\pi} \int_\Sigma B_{ij} dx^i \wedge dx^j \tag{12.526}$$

This is a topological field theory. The action is (locally) a total derivative. Therefore, for the case of a strip worldsheet $\mathbb{R} \times [0, \pi]$ we have

$$S = \frac{\mathrm{i}}{4\pi} \int_{\mathbb{R}} B_{ij} x_2^i \frac{dx_2^j}{d\tau} d\tau - \frac{\mathrm{i}}{4\pi} \int_{\mathbb{R}} B_{ij} x_1^i \frac{dx_1^j}{d\tau} d\tau \tag{12.527}$$

where $x_2^i(\tau) = x^i(\sigma = \pi, \tau)$ and $x_1^i(\tau) = x^i(\sigma = 0, \tau)$. This is the essential part of the action used in the dipole picture. Note that in the SW limit the string tension $T \sim \ell_s^{-2}$ has gone to infinity. This suppresses all the oscillator modes, since exciting them requires an infinite amount of energy, so we should think of the endpoints of the string as being connected by a rigid rod.

Remark: It is just a happy coincidence that the traditional notation B_{ij} for the string theory B-field coincides with the traditional notation B for the magnetic component of the electromagnetic fieldstrength. The two geometric quantities are, *a priori* quite distinct. In general B_{ij} is really a local connection for a gerbe. In electromagnetism B is a particular component of a fieldstrength.

12.9.2 Toroidal Compactification

Now we consider the target space of the string theory to be

$$\mathcal{X} = T^{2n} \times \mathbb{R}^{D-2n} \tag{12.528}$$

For simplicity we will form the torus by identifying coordinates

$$x^i \sim x^i + 2\pi R, \qquad i = 1, \dots, 2n \tag{12.529}$$

Not surprisingly, the vertex operator algebra $V_k(\tau)$, for k in the cotangent space of T^{2n} becomes the noncommutative torus algebra in the limit (12.520), (12.521).

To see this we use the dipole picture described above. The action (12.527) is first order in time derivatives, and hence should be considered an action on phase space with symplectic form

$$\omega = \frac{1}{4\pi} B_{ij} dx_2^i dx_2^j - \frac{1}{4\pi} B_{ij} dx_1^i dx_1^j \tag{12.530}$$

Now, both x_i^i and x_2^i are periodic coordinates on T^{2n}. But this does not mean that the phase space is $T^{2n} \times T^{2n}$ with the above symplectic form, because we must remember that there is a string that connects the two points!

In fact, the phase space is the space of morphisms of the *fundamental groupoid* of T^{2n}. This is the category whose objects are points of T^{2n} and whose morphisms are homotopy classes of paths between two points. In our example, given two points there is a string connecting them. Physically, since the tension of the string goes to infinity, the string should be the minimal length path in its homotopy class, i.e., a straight line on the universal cover.

Alternatively, if one simply starts from the action (12.527) then there is a gauge invariance $x^i(\sigma,\tau) \to x^i(\sigma,\tau) + \delta x^i(\sigma,\tau)$ with $\delta x^i(\sigma,\tau)|_{\sigma=0,\pi} = 0$. The reason is that the action is the pullback of a closed 2-form on $\mathcal{X}$. So the variation vanishes by Stokes' theorem. In the case where the worldsheet has a boundary $\partial\Sigma \neq \emptyset$ we must require that the variation of x^i on the boundary vanishes. Again, the conclusion is the same: Since the two points $x_1, x_2 \in T^{2n}$ are connected by a string the phase space is the space of morphisms of the fundamental groupoid.

Viewing $x_1, x_2 \in \mathbb{R}^{2n}$ as elements of the universal cover we can make a change of variables

$$\begin{aligned} x_1^i &= y^i + \frac{1}{2}\Delta^i \\ x_2^i &= y^i - \frac{1}{2}\Delta^i \end{aligned} \tag{12.531}$$

Note that under the Deck transformations:

$$\begin{aligned} x_1^i &\to x_1^i + 2\pi n^i R, \\ x_2^i &\to x_2^i + 2\pi n^i R, \end{aligned} \tag{12.532}$$

(where we use the <u>same</u> integer n^i when transforming both x_1^i and x_2^i) the center of mass transforms as

$$y^i \to y^i + 2\pi n^i R \tag{12.533}$$

so it descends to a point on the torus, while the difference Δ^i, which measures the homotopy class of the geodesic on T^{2n} defined by the straight-line path between x_1 and x_2 on $\mathbb{R}^{2n}$, is invariant under the Deck transformation.

In these coordinates the symplectic form is

$$\omega = \frac{B_{ij}}{2\pi} d\Delta^i dy^j \tag{12.534}$$

This is a symplectic form on T^*T^{2n}. The Poisson brackets are

$$\{\Delta^i, y^j\} = 2\pi (B^{-1})^{ij} = \Theta^{ij} \tag{12.535}$$

and upon quantization we have

$$\Delta^i = -2\pi \mathrm{i} \Theta^{ij} \frac{\partial}{\partial y^j} \tag{12.536}$$

Now, classically, the operators associated with one endpoint of the string are

$$U_i = \exp[\mathrm{i} x_1^i / R] \tag{12.537}$$

and they generate the algebra of functions on T^{2n}. Upon quantization we have

$$U_i U_j = \exp[\frac{2\pi \mathrm{i}}{R^2} \Theta^{ij}] U_j U_i \tag{12.538}$$

and we get the noncommutative torus algebra. Note that the other end of the string describes functions

$$\tilde{U}_i = \exp[\mathrm{i} x_2^i / R] \tag{12.539}$$

and upon quantization these satisfy

$$\tilde{U}_i \tilde{U}_j = \exp[-\frac{2\pi \mathrm{i}}{R^2} \Theta^{ij}] \tilde{U}_j \tilde{U}_i \tag{12.540}$$

(Do not confuse this with the fact that the commutant of the noncommutative torus algebra in $L^2(\mathbb{R})$ with parameter θ is the algebra with parameter $1/\theta$.)

♣We can also arrive at this conclusion using crossed products and orbifolds. ♣

12.9.3 Closed Strings And T-Duality

It is good to begin by recalling a few aspects of electric-magnetic duality in 1+1 dimensions:

Recall that given an orientation on a finite-dimensional vector space V of dimension n with metric the Hodge dual satisfies

$$\omega * \omega = \parallel \omega \parallel^2 \mathrm{vol} \tag{12.541}$$

where ω is a p-form in $\Lambda^p V$,

$$\parallel \omega \parallel^2 = \frac{1}{p!} g^{\mu_1 \nu_1} \cdots g^{\mu_p \nu_p} \omega_{\mu_1 \dots \mu_p} \omega_{\nu_1 \dots \nu_p} \tag{12.542}$$

is the induced norm on $\Lambda^p V$ and vol is the volume form

$$\mathrm{vol} = e_1 \wedge \dots e_n \tag{12.543}$$

for an oriented orthonormal basis of V. We have

$$* : \Lambda^p V \to \Lambda^{n-p} V \tag{12.544}$$

and

$$*^2 = (\mathrm{signdet} g) \cdot (-1)^{p(n-p)} \tag{12.545}$$

For the Minkowski metric on the worldsheet, $-d\tau^2 + d\sigma^2$ with orientation $d\sigma \wedge d\tau$ the action of Hodge $*$ on one-forms is determined by

♣Actually, the opposite orientation was assumed in the action (12.485) above. ♣

$$\begin{aligned} *d\tau &= d\sigma \\ *d\sigma &= d\tau \end{aligned} \tag{12.546}$$

Now suppose that we have a one-form abelian fieldstrength $F \in \Omega^1(\Sigma)$. In the absence of sources it satisfies the Bianchi identity and equation of motion

$$\begin{aligned} dF &= 0 \\ d * F &= 0 \end{aligned} \tag{12.547}$$

For a worldsheet $\Sigma = \mathbb{R} \times \mathcal{D}$, where $\mathcal{D}$ is a spatial domain (a circle for closed strings, an interval for open strings) we can solve, at least locally:

$$F = dx \tag{12.548}$$

and then x satisfies the wave equation. Note that the self dual equations imply

$$\begin{aligned} F = *F \qquad &\Rightarrow \qquad x = x_L(\tau + \sigma) \\ F = - * F \qquad &\Rightarrow \qquad x = x_R(\tau - \sigma) \end{aligned} \tag{12.549}$$

Every fieldstrength can be decomposed into self-dual and anti-self-dual parts, and correspondingly the general solution of the wave equation is

$$x = x_L(\tau + \sigma) + x_R(\tau - \sigma) \tag{12.550}$$

That is, it is a sum of left- and right-moving waves. Now the *electromagnetic dual fieldstrength* is defined to be

$$F_D := *F \tag{12.551}$$

At the level of the 0-form potential we can define the dual coordinate x_D by:

$$dx_D := *dx \tag{12.552}$$

so

$$\begin{aligned} (x_D)'_L &= x'_L \\ (x_D)'_R &= -x'_R \end{aligned} \tag{12.553}$$

We can choose the constant of integration so that

$$\begin{aligned} (x_D)_L &= x_L \\ (x_D)_R &= -x_R \end{aligned} \tag{12.554}$$

Thus, electric-magnetic duality maps left-movers to left-movers, and right-movers to right-movers, but with an important sign flip on the right-movers. Note that if $\mathcal{D}$ has a boundary, say at $\sigma = 0$ then:

1. <u>Dirichlet boundary conditions</u> reflect left-movers into right-movers with a minus sign:

$$\partial_\tau \left(x_L(\tau+\sigma) + x_R(\tau-\sigma) \right) |_{\sigma=0} \Rightarrow x'_L(\tau) = -x'_R(\tau) \tag{12.555}$$

2. <u>Neumann boundary conditions</u> reflect left-movers into right-movers with a plus sign:

$$\partial_\sigma \left(x_L(\tau+\sigma) + x_R(\tau-\sigma) \right) |_{\sigma=0} \Rightarrow x'_L(\tau) = +x'_R(\tau) \tag{12.556}$$

Therefore, electric-magnetic duality exchanges Dirichlet and Neumann boundary conditions

Now let us consider the case of <u>closed</u> strings with a target space $\mathcal{X} = S^1$. We have

$$x : \mathbb{R} \times S^1 \to \mathcal{X} \tag{12.557}$$

Do not confuse the worldsheet spatial circle with the target space circle $\mathcal{X}$. We take the metric $dx \otimes dx$ on the target but impose boundary conditions

$$x \sim x + 2\pi R \tag{12.558}$$

Put differently (this will be useful in the discussion of the general case) we consider

$$\mathcal{X} = \mathbb{R}/(2\pi R\mathbb{Z}) \tag{12.559}$$

the metric is induced from $dx \otimes dx$ on $\mathbb{R}$ and is equivalent to the metric on a circle of radius R.

The general solution of the equation of motion looks like

$$x = x_0 + \frac{1}{2}(\ell_s^2 p + wR)(\tau+\sigma) + \frac{1}{2}(\ell_s^2 p - wR)(\tau-\sigma) + x_{\rm osc}. \tag{12.560}$$

We briefly review the quantization of $x_{\rm osc}$, in a more general context, below. The "zero-mode part" requires a bit more discussion.

The momentum density is

$$p(\tau,\sigma) = \frac{\delta S}{\delta \dot{x}} = \frac{1}{2\pi \ell_s^2}\dot{x}(\tau,\sigma) \tag{12.561}$$

The zeromode of $p(\tau,\sigma)$ as a function of σ is, by definition, denoted p. Quantization gives $[\hat{p}, \hat{x}_0] = -{\rm i}$, so $\exp[{\rm i}(2\pi R)\hat{p}] = 1$ and the eigenvalues of $\hat{p}$ are quantized to be of the form n/R where $n \in \mathbb{Z}$. Similarly the integral around the worldsheet must be of the form

$$\oint_{S^1} dx = \oint_{S^1} \partial_\sigma x d\sigma = 2\pi R w \tag{12.562}$$

where $w \in \mathbb{Z}$. Of course, the expansion (12.560) is written on the universal cover $\mathbb{R}$ of the target space, and w is interpreted as the winding number of the map $x : S^1 \to \mathcal{X}$ at fixed τ.

It is useful to rewrite (12.560) as:

$$x = x_0 + \frac{1}{2}\ell_s^2 p_L(\tau+\sigma) + \frac{1}{2}\ell_s^2 p_R(\tau-\sigma) + x_{\rm osc} \tag{12.563}$$

$$\begin{aligned} \ell_s p_L &= n\frac{\ell_s}{R} + w\frac{R}{\ell_s} \\ \ell_s p_R &= n\frac{\ell_s}{R} - w\frac{R}{\ell_s} \end{aligned} \tag{12.564}$$

Note that under electric-magnetic duality we have

$$\begin{aligned} n_D &= w \\ w_D &= n \\ RR_D &= \ell_s^2 \end{aligned} \tag{12.565}$$

Moreover, the Hamiltonian of the theory is

$$\begin{aligned} H &= \frac{1}{4\pi\ell_s^2}\oint\left[(2\pi\ell_s^2 p(\sigma))^2 + (\partial_\sigma x)^2\right]d\sigma \\ &= \frac{1}{4}\left[(\ell_s p_L)^2 + (\ell_s p_R)^2\right] + H_{\rm osc} \\ &= \frac{1}{2}\left[n^2\left(\frac{\ell_s}{R}\right)^2 + w^2\left(\frac{R}{\ell_s}\right)^2\right] + H_{\rm osc} \end{aligned} \tag{12.566}$$

From these equations we see that the theory is completely invariant under

$$\frac{R}{\ell_s} \to \frac{\ell_s}{R}. \tag{12.567}$$

This quantum equivalence of two different quantum field theories in $1+1$ dimensions is the first example of a duality known as *T-duality*. It is electro-magnetic duality in $1+1$ dimensions and exchanges momentum and winding modes, or equivalently, electrically and magnetically charged states. It is a simple demonstration of the general claim that the nature of spacetime changes dramatically at the string scale. In ordinary QFT on the target space circle $\mathcal{X}$ the theories at small and large values of R are completely different. For example, the spectrum of the Hamiltonian is completely different.

Remark: At the fixed point of the duality transformation, $R = \ell_s$, we should expect something special to happen. Indeed this is the case, the theory generically has symmetry under the centrally-extended loop groups $\widehat{LU(1)} \times \widehat{LU(1)}$. These are associated with the holomorphic currents $-\mathrm{i}\partial x(z)$ and $-\mathrm{i}\bar\partial x(\bar z)$. However, at $R = \ell_s$ the vertex operator algebra has a larger symmetry group $\widehat{LSU(2)} \times \widehat{LSU(2)}$ (with $k = 1$). In general in the Gaussian model we have the symmetry $\mathrm{vir} \oplus \widetilde{\mathrm{vir}}$, with the expressions

$$L_0 = \frac{1}{4}\left(\frac{n}{r} + mr\right)^2 + \sum_{n>0}\alpha_{-n}\alpha_n \tag{12.568}$$

$$\tilde{L}_0 = \frac{1}{4}\left(\frac{n}{r} - mr\right)^2 + \sum_{n>0} \tilde{\alpha}_{-n}\tilde{\alpha}_n \tag{12.569}$$

where $r := R/\ell_s$. There are primary operators associated with the exponentials:

$$V_{\mathbf{p}} =: e^{\mathrm{i}p_L x_L}(z) \otimes e^{\mathrm{i}p_R x_R}(\bar{z}) : \tag{12.570}$$

Note that when $r^2 = p/q$ it is possible to choose integers n, m so that the operator is purely holomorphic or anti-holomorphic. In particular, at $r = 1$, if we choose $n = m = \pm 1$ we get the purely holomorphic dimension one current

$$e^{\pm \mathrm{i} 2 x_L/\ell_s}(z) \tag{12.571}$$

and similarly $n = -m = \pm 1$ gives purely anti-holomorphic dimension one currents. The operators $-\mathrm{i}\partial x(z), e^{\pm \mathrm{i} 2 x_L/\ell_s}(z)$ have an OPE corresponding to the currents of a level one $su(2)$ affine Lie algebra. This is known in mathematics as the Frenkel-Kac-Segal construction. More generally, when $r^2 = p/q$ there are holomorphic vertex operators

$$\exp[\mathrm{i}2\sqrt{pq}\, x_L(z)/\ell_s] \tag{12.572}$$

of $\Delta = pq$. These are holomorphic higher spin currents and lead to extra symmetries. One consequence of these higher spin symmetries is that there are nice holomorphic factorizations of the partition functions and correlation functions of the theory. These are examples of *rational conformal field theories* (and the name actually originates from these examples).

Let us now generalize these equations to the case that $\mathcal{X} = \mathbb{R}^d/(2\pi\Lambda)$ where Λ is an embedded lattice and $\mathbb{R}^d$ is equipped with constant metric g_{ij} and B-field b_{ij}. We can write, once again:

$$x^i = x_0^i + \frac{1}{2}\ell_s^2 p_L^i(\tau + \sigma) + \frac{1}{2}\ell_s^2 p_R^i(\tau - \sigma) + x^i_{\mathrm{osc}} \qquad i = 1, \dots, d \tag{12.573}$$

but now

$$p_i(\tau, \sigma) = \frac{\delta S_L}{\delta \dot{x}^i} = \frac{1}{2\pi\ell_s^2}\left(g_{ij}\partial_\tau x^j + b_{ij}\partial_\sigma x^j\right) \tag{12.574}$$

From the viewpoint of electric-magnetic duality we have a gauge group $U(1)^d$ with magnetic charges

$$\begin{aligned} \mathbf{m}^i &:= \oint_{S^1} dx^i = \pi\ell_s^2(p_L^i - p_R^i) \\ \mathbf{e}^i &:= \oint_{S^1} *dx^i = \pi\ell_s^2(p_L^i + p_R^i) \end{aligned} \tag{12.575}$$

Once again, there will be quantization of the electric and magnetic charges. The magnetic quantization is easiest since $\mathbf{m}^i$ must be a vector in $2\pi\Lambda$.

We can choose a basis of the form $\ell_s e_a$, $a = 1, \dots, d$, for Λ so that e_a are dimensionless. The components of the vectors will be ${e_a}^i$. Therefore we have the magnetic quantization conditions:

$$\frac{1}{2}\ell_s(p_L^i - p_R^i) = w^a {e_a}^i \tag{12.576}$$

where w^a, $a = 1, \dots, d$, is a vector of winding integers.

Similarly, the momenta conjugate to the zeromode x_0^i is

$$\begin{aligned} p_i &= \frac{1}{2\pi\ell_s^2}\oint \left(g_{ij}\partial_\tau x^j + b_{ij}\partial_\sigma x^j\right) d\sigma \\ &= \frac{1}{2\pi\ell_s^2}\left(g_{ij}\mathbf{e}^j + b_{ij}\mathbf{m}^j\right) \\ &= \frac{1}{2}\left(g_{ij}(p_L + p_R)^j + b_{ij}(p_L - p_R)^j\right) \end{aligned} \tag{12.577}$$

Again upon quantization we have that

$$\exp[\mathrm{i}2\pi\ell_s {e_a}^i \hat{p}_i] = 1 \tag{12.578}$$

on the free one-string Hilbert space. This quantizes p_i to be vectors in the dual lattice $\Lambda^\vee$. More precisely, we assume the lattice has full rank so $\det({e_a}^i) \neq 0$, so ${e_a}^i$ has an inverse matrix ${e_i}^a$

$${e_a}^i {e_i}^b = {\delta_a}^b \qquad {e_i}^a {e_a}^j = {\delta_i}^j \tag{12.579}$$

Then $\hat{p}_i$ must have a spectrum of the form

$$\ell_s^{-1} n_a {e_i}^a \tag{12.580}$$

where the n_a are integers for $a = 1, \dots, d$.

Putting together (12.576) and (12.580) we get the quantization of $(\ell_s p_L^i, \ell_s p_R^i)$:

$$\begin{aligned} \frac{1}{2}(\ell_s p_L - \ell_s p_R)^i &= w^a {e_a}^i \\ \frac{1}{2}\left(g_{ij}(\ell_s p_L + \ell_s p_R)^j + b_{ij}(\ell_s p_L - \ell_s p_R)^j\right) &= n_a {e_i}^a \end{aligned} \tag{12.581}$$

where n_a, w^a are integers. We are going to put (12.581) in a more beautiful mathematical form, but first we would like to make a side remark about the relation to an analogous effect in $3+1$-dimensional Yang-Mills theory.

Remark: Analog in magnetic monopole theory: The equations (12.577) and (12.580) are closely related to a phenomenon in magnetic monopole theory known as the *Witten effect*. We consider Yang-Mills-Higgs theory where the YM action has a θ-term:

$$-\frac{1}{g^2}\int (F, *F) + \frac{\theta}{8\pi^2}\int (F, F) - \frac{1}{g^2}\int (D\Phi, D\Phi) + \cdots \tag{12.582}$$

Then if the Higgs field spontaneously breaks the gauge group to a maximal torus at infinity there can be magnetic charges. These are measured by

$$\frac{1}{2\pi}\int_{S^2_\infty} F = \mathbf{m} \tag{12.583}$$

where S^2_∞ is a spatial sphere at infinity, and $\mathbf{m} \in \Lambda_{cw}$ is quantized to be in the "coweight lattice" inside the Cartan subalgebra $\mathfrak{t}$. Similarly, the electric charge is defined to be

$$\frac{2}{g^2}\int_{S^2_\infty} *F = \mathbf{e} \tag{12.584}$$

Because of the θ term, the momentum conjugate to $A_i(x)$ is not the electric field but rather

$$\Pi^i(x) = \frac{1}{g^2}E^i(x) + \frac{\theta}{4\pi^2}\epsilon^{ijk}F_{jk} \tag{12.585}$$

where $E^i := F_{0i}$. Upon quantization we have operators

$$[\widehat{\Pi}^i(x), \hat{A}_j(y)] = -\mathrm{i}\delta^i{}_j\delta^{(3)}(x-y) \tag{12.586}$$

The momentum translates the gauge fields, as usual, and in particular gauge transformations by $\epsilon : \mathbb{R}^3 \to \mathfrak{g}$ are implemented by

$$Q(\epsilon) = \mathrm{i}\int_{\mathbb{R}^3}(\epsilon, D_i\widehat{\Pi}^i) \tag{12.587}$$

Gauge transformations are implemented by ϵ such that $\lim_{x\to\infty}\epsilon(x) = 0$, and for these transformations $Q(\epsilon)$ acts as zero on the physical Hilbert space. But there are also global gauge transformations with $\lim_{x\to\infty}\epsilon(x) = H \in \mathfrak{t}$. In general these do not act trivially on the Hilbert space. However, this must generate a representation of the unbroken gauge group $T \subset G$. Therefore, if H is in the co-character lattice, so that $\exp[2\pi H] = 1$ then the quantum operator $\exp[2\pi Q(H)]$ must act as the identity on the Hilbert space. This means that the operator

$$\lim_{r\to\infty}\int_{S^2}\hat{n}^i\Pi^i(x)r^2\sin\theta d\theta d\phi \tag{12.588}$$

should exist in the quantum theory, and moreover it will have quantized eigenvalues. Specifically, it must be in the integral dual of the cocharacter lattice $\Lambda_G \subset \mathfrak{t}$ with respect to the quadratic form given by the Killing metric used to define the action. (It is isomorphic to the character lattice of G.) Call this quantized momentum $\hat{\gamma}_e$. From equation (12.585) we learn that

$$\mathbf{e} \; = \hat{\gamma}_e + \frac{\theta}{2\pi}\mathbf{m} \tag{12.589}$$

That is, in the presence of a magnetic monopole, the physical electric charge has a fractional part proportional to $\frac{\theta}{2\pi}$. Monopoles with electric charge are known as *dyons*. We learn that in the presence of a generic theta angle all monopoles are in fact dyons. This is quite analogous to the fact that, in the presence of a generic B-field strings with nonzero winding number must carry momentum.

Let us now return to our closed string theory with target space $\mathcal{X} = T^d$ and interpret the facts about the quantization (12.581) of $(\ell_s p^i_L, \ell_s p^i_R)$. Rearranging those equations slightly we find:

$$\begin{aligned} \ell_s p^i_L &= g^{ij}e_j{}^a n_a + w^a e_a{}^i + w^a e_a{}^j b_{jk}g^{ki} \\ \ell_s p^i_R &= g^{ij}e_j{}^a n_a - w^a e_a{}^i + w^a e_a{}^j b_{jk}g^{ki} \end{aligned} \tag{12.590}$$

This equation can be more usefully written in matrix form

$$\frac{1}{\sqrt{2}}\begin{pmatrix}\ell_s p_L \\ \ell_s p_R\end{pmatrix} = \mathcal{E}\begin{pmatrix} n \\ w\end{pmatrix} \tag{12.591}$$

Here $\mathcal{E}$ is a $2d \times 2d$ matrix. It can be written in block form

$$\mathcal{E} = \begin{pmatrix}\mathcal{E}_{11} & \mathcal{E}_{12} \\ \mathcal{E}_{21} & \mathcal{E}_{22}\end{pmatrix} \tag{12.592}$$

where the $d \times d$ blocks have matrix elements

$$\begin{aligned}(\mathcal{E}_{11})^{ia} &= (\mathcal{E}_{21})^{ia} = \frac{1}{\sqrt{2}} g^{ij} e_j{}^a \\ (\mathcal{E}_{12})^i{}_a &= \frac{1}{\sqrt{2}}\left(e_a{}^i + e_a{}^j b_{jk} g^{ki}\right) \\ (\mathcal{E}_{22})^i{}_a &= -\frac{1}{\sqrt{2}}\left(e_a{}^i - e_a{}^j b_{jk} g^{ki}\right)\end{aligned} \tag{12.593}$$

Now, define the two quadratic forms of signature (d,d):

$$Q_0 := \begin{pmatrix} g_{ij} & 0 \\ 0 & -g_{ij}\end{pmatrix} \qquad\qquad Q := \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0\end{pmatrix} \tag{12.594}$$

then a small, but important, computation shows that:

$$\boxed{\mathcal{E}^{\mathrm{tr}} Q_0 \mathcal{E} = Q} \tag{12.595}$$

One way to read equation (12.595) is that it states that the columns of $\mathcal{E}$ define a collection of vectors in $\mathbb{R}^{2d}$. This space is equipped with signature (d,d) metric Q_0, and we denote that space by $\mathbb{R}^{d;d}$. Thus, the columns of $\mathcal{E}$ generate a rank $2d$ embedded lattice $\Gamma \subset \mathbb{R}^{d;d}$ with Gram matrix Q.

In the classification of integral lattices, (equivalently, in the classification of integral symmetric quadratic forms), those with Gram matrix equivalent to Q are unique. These are the even unimodular lattices of signature (d,d) and their equivalence class is denoted $II^{d,d}$. It is a direct sum of the basic case $II^{1,1}$. A standard model for $II^{1,1}$ is $\mathbb{Z}e \oplus \mathbb{Z}f$ where the two generators have inner products:

$$Q(e,e) = Q(f,f) = 0 \qquad\qquad Q(e,f) = Q(f,e) = 1 \tag{12.596}$$

Thus, the set of quantized zeromodes

$$\boldsymbol{p} := \frac{1}{\sqrt{2}}\begin{pmatrix}\ell_s p_L \\ \ell_s p_R\end{pmatrix} \tag{12.597}$$

given by the integral span of the columns of $\mathcal{E}$ *is an embedding of a standard copy of* $II^{d,d}$ *into the space* $\mathbb{R}^{d;d}$.

Note that the inner product Q_0 on $\mathbb{R}^{d;d}$ expressed in terms of $\mathbf{p}$ is

$$\mathbf{p}^2 = \frac{1}{2}\left(\ell_s p_L^i g_{ij} \ell_s p_L^j - \ell_s p_R^i g_{ij} \ell_s p_R^j\right) \tag{12.598}$$

Now let us briefly review the (straightforward) treatment of the oscillators. The general solution of the equation of motion is

$$x^i_{\mathrm{osc}} = \mathrm{i}\frac{\ell_s}{\sqrt{2}} \sum_{n\neq 0} \left(\frac{\alpha_n^i}{n} e^{\mathrm{i}n(\tau+\sigma)} + \frac{\tilde{\alpha}_n^i}{n} e^{\mathrm{i}n(\tau-\sigma)} \right) \tag{12.599}$$

Reality imposes $(\alpha_n^i)^* = \alpha_{-n}^i$ and similarly for $\tilde{\alpha}_n^i$. Thus, the $(\alpha_n^i, \tilde{\alpha}_n^i)$ for $n > 0$ are (complex) coordinates on phase space. We can therefore determine $p_i(\tau,\sigma)$ from the above equations as another function on phase space.

Now, the natural symplectic form on phase space is

$$\omega = \oint_{S^1} d\sigma \delta p_i(\tau,\sigma) \wedge \delta x^i(\tau,\sigma) \tag{12.600}$$

Substituting the expressions in terms of the oscillators, doing the σ-integral, and taking into account numerous cancellations one finally arrives at:

$$\omega = \delta p_i \wedge \delta x_0^i - \mathrm{i}\sum_{n>0} \left(\frac{\delta\alpha_n^i g_{ij} \delta\alpha_{-n}^j}{n} + \frac{\delta\tilde{\alpha}_n^i g_{ij} \delta\tilde{\alpha}_{-n}^j}{n} \right) \tag{12.601}$$

Now recall that for the harmonic oscillator with frequence ω we have the standard complex coordinates on phase space

$$\begin{aligned} a &= \frac{p - \mathrm{i}q}{\sqrt{2}\omega} \\ \bar{a} &= \frac{p + \mathrm{i}q}{\sqrt{2}\omega} \end{aligned} \tag{12.602}$$

so that

$$\delta p \wedge \delta q = -\mathrm{i}\frac{\delta a \wedge \delta \bar{a}}{\omega} \tag{12.603}$$

It follows that in the quantum theory the oscillators satisfy:

$$\begin{aligned} [\alpha_n^i, \alpha_m^j] &= g^{ij} n \delta_{n+m,0} \\ [\tilde{\alpha}_n^i, \tilde{\alpha}_m^j] &= g^{ij} n \delta_{n+m,0} \end{aligned} \tag{12.604}$$

with all other commutators vanishing. Note that we use here the closed string metric g_{ij}. The representation is the standard Fock space

$$\mathcal{F} \otimes \tilde{\mathcal{F}} \tag{12.605}$$

with the vacuum line annihilated by the negative energy oscillators (that is, annihilated by $\alpha_n^i, \tilde{\alpha}_n^i$ for $n > 0$).

Putting together the zeromodes and the oscillators the Hilbert space of the CFT can be viewed as

$$\mathcal{H}_{1-closedstring} = \mathbb{C}[\Gamma] \otimes \mathcal{F} \otimes \tilde{\mathcal{F}} = \oplus_{\mathbf{p}\in\Gamma} \mathcal{F} \otimes \tilde{\mathcal{F}} \otimes V_{\mathbf{p}} \tag{12.606}$$

where $V_{\mathbf{p}}$ is the vacuum line for the Heisenberg representation of the oscillator modes determined by a vector in $\mathbf{p} \in \Gamma$. In the language of 2d CFT it is the state created from the $SL(2,\mathbb{R}) \times \widetilde{SL(2,\mathbb{R})}$-invariant vacuum by the vertex operator $\exp[\mathrm{i}p_{L,i}x_L^i + \mathrm{i}p_{R,i}x_R^i]$.

Now let us discuss the moduli space of conformal field theories with target space $\mathcal{X} = T^d$. The matrix $\mathcal{E}$ is the essential piece of data that determines the entire CFT. For example, as we have just explained, using $\mathcal{E}$ we can determine the lattice Γ of zero-modes. So, to begin, we have a family of conformal field theories over the space $\mathfrak{E}$ given by the space of real matrices of the form (12.592) and (12.593). (Note one can easily recover g_{ij} and b_{ij} separately from this matrix.) However, there is some redundancy in this matrix.

For example, the Hamiltonian of the theory is

$$H = \frac{\ell_s^2}{4}(p_L^i g_{ij} p_L^j + p_R^i g_{ij} p_R^j) + H_{\mathrm{osc}} \tag{12.607}$$

$$H_{\mathrm{osc}} = \sum_{n>0} \left(\alpha_{-n}^i g_{ij} \alpha_n^j + \tilde{\alpha}_{-n}^i g_{ij} \tilde{\alpha}_n^j \right) \tag{12.608}$$

This can be written as

$$H = L_0 + \tilde{L}_0 \tag{12.609}$$

$$L_0 = \frac{\ell_s^2}{4} p_L^i g_{ij} p_L^j + \sum_{n>0} \alpha_{-n}^i g_{ij} \alpha_n^j \tag{12.610}$$

$$\tilde{L}_0 = \frac{\ell_s^2}{4} p_R^i g_{ij} p_R^j + \sum_{n>0} \tilde{\alpha}_{-n}^i g_{ij} \tilde{\alpha}_n^j \tag{12.611}$$

The spectrum of H_{osc} is just $N + \tilde{N}$, where $N, \tilde{N} \in \mathbb{Z}_+$ with degeneracy $p_d(N)p_d(\tilde{N})$. That is

$$\mathrm{Tr}_{\mathcal{F}\otimes\tilde{\mathcal{F}}} q^{L_0^{\mathrm{osc}}} \bar{q}^{\tilde{L}_0^{\mathrm{osc}}} = \frac{(q\bar{q})^{d/24}}{\eta^d \bar{\eta}^d} \tag{12.612}$$

The spectrum of the Hamiltonian for zeromodes only depends on

$$\begin{pmatrix} \ell_s p_L & \ell_s p_R \end{pmatrix} \begin{pmatrix} g_{ij} & 0 \\ 0 & +g_{ij} \end{pmatrix} \begin{pmatrix} \ell_s p_L \\ \ell_s p_R \end{pmatrix} = \begin{pmatrix} n_a & w^a \end{pmatrix} \begin{pmatrix} \tilde{g}^{ab} & -(\tilde{g}^{-1}\tilde{b})^a{}_b \\ (\tilde{b}\tilde{g}^{-1})_a{}^b & \tilde{g}_{ab} - (\tilde{b}\tilde{g}^{-1}\tilde{b})_{ab} \end{pmatrix} \begin{pmatrix} n_b \\ w^b \end{pmatrix} \tag{12.613}$$

where we define

$$\tilde{g}_{ab} := e_a{}^i e_b{}^j g_{ij} \qquad \tilde{b}_{ab} := e_a{}^i e_b{}^j b_{ij}. \tag{12.614}$$

Therefore the spectrum of the Hamiltonian only depends on the projection of $\mathcal{E}$ under

$$\pi : \mathfrak{E} \to \left(O_{\mathbb{R}}(g) \times O_{\mathbb{R}}(g)\right) \backslash \mathfrak{E} := \mathfrak{B} \tag{12.615}$$

where

$$O_{\mathbb{R}}(g) := \{\alpha \in GL(d,\mathbb{R}) | \alpha^{\mathrm{tr}} g \alpha = g\} \tag{12.616}$$

We can identify the coset space $\mathfrak{B}$ with the space of $d \times d$ real matrices with positive definite real part. An explicit map to this space is

$$\mathcal{E} \to (\mathcal{E}_{11})^{-1}\mathcal{E}_{12} = e_a^{\ j}(g_{jk} - b_{jk})e_b^{\ k} = \tilde{g}_{ab} - \tilde{b}_{ab} := E_{ab} \tag{12.617}$$

♣Usually E is defined with $g+b$. Change conventions? ♣

Not only the spectrum of the Hamiltonian but the entire CFT actually descends from a family over $\mathfrak{E}$ to a family over $\mathfrak{B}$. This is especially obvious if we recall that we are identifying

$$x \sim x + 2\pi\Lambda \tag{12.618}$$

and hence we can change coordinates to

$$x^i = \ell_s \xi^a e_a^{\ i} \tag{12.619}$$

so that the fields $\xi^a(\tau,\sigma)$ are dimensionless have all have periodicity

$$\xi^a \sim \xi^a + 2\pi \tag{12.620}$$

In terms of these fields we can write the action as

$$S_L = \frac{1}{4\pi}\int_\Sigma d\tau d\sigma(\partial_\tau - \partial_\sigma)\xi^a E_{ab}(\partial_\tau + \partial_\sigma)\xi^b = \frac{1}{4\pi}\int_\Sigma d\tau d\sigma(\partial_\tau + \partial_\sigma)\xi^a E^{\rm tr}_{ab}(\partial_\tau - \partial_\sigma)\xi^b \tag{12.621}$$

There is another source of redundancy in $\mathcal{E}$. We can construct the lattice of zeromodes from an embedding of $II^{d,d}$ into $\mathbb{R}^{d;d}$. However, what matters in the construction of the theory is not the choice of basis vectors $e_a \in \Gamma$, but just the lattice Γ itself. We do not change the lattice of zeromodes by changing basis! An integral change of basis on e_a is obtained by right-multiplication of $\mathcal{E}$ by an invertible integral matrix $\mathfrak{d} \in GL(2d,\mathbb{Z})$. However this change of basis must preserve (12.595) and hence we must have $\mathfrak{d} \in O_{\mathbb{Z}}(Q)$, that is, $\mathfrak{d}$ must be in the group of integral matrices such that $\mathfrak{d}^{\rm tr}Q\mathfrak{d} = Q$. We can write $\mathfrak{d}$ in block form as

$$\mathfrak{d} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in O_{\mathbb{Z}}(Q) \tag{12.622}$$

where $\mathfrak{d} \in O_{\mathbb{Z}}(Q)$ iff $\alpha,\beta,\gamma,\delta \in M_d(\mathbb{Z})$ satisfy:

$$\begin{aligned} \alpha^{\rm tr}\delta + \gamma^{\rm tr}\beta &= \mathbf{1} \\ \alpha^{\rm tr}\gamma + \gamma^{\rm tr}\alpha &= 0 \\ \delta^{\rm tr}\beta + \beta^{\rm tr}\delta &= 0 \end{aligned} \tag{12.623}$$

(It is often useful to note that (12.623) holds iff

$$\begin{aligned} \delta\alpha^{\rm tr} + \gamma\beta^{\rm tr} &= \mathbf{1} \\ \alpha\beta^{\rm tr} + \beta\alpha^{\rm tr} &= 0 \\ \delta\gamma^{\rm tr} + \gamma\delta^{\rm tr} &= 0.) \end{aligned} \tag{12.624}$$

The space of embedded lattices is

$$\mathfrak{L} := \mathfrak{E}/O_{\mathbb{Z}}(Q) \tag{12.625}$$

Thus, there are two sources of redundancy and the space of CFT's descends from a family over $\mathfrak{E}$ in two ways:

$$\begin{tikzcd} & \mathfrak{E} \arrow[dl] \arrow[dr] & \\ \mathfrak{B} & & \mathfrak{L} \end{tikzcd} \tag{12.626}$$

where $\mathfrak{B}$ is the moduli space of classical sigma models: They are determined by E_{ab}. $\mathfrak{L}$ is the moduli space of embeddings of the unique even unimodular lattice of signature (d, d) into $\mathbb{R}^{d;d}$.

These two sources of redundancy in $\mathcal{E}$ are independent of each other so that in fact the family of CFT's descends to

$$\begin{tikzcd} & \mathfrak{E} \arrow[dl] \arrow[dr] & \\ \mathfrak{B} \arrow[dr] & & \mathfrak{L} \arrow[dl] \\ & \mathcal{N}_{d,d} & \end{tikzcd} \tag{12.627}$$

where

$$\mathcal{N}_{d,d} := (O_{\mathbb{R}}(g) \times O_{\mathbb{R}}(g)) \backslash \mathfrak{E}/O_{\mathbb{Z}}(Q) \tag{12.628}$$

This is the moduli space of 2d CFT's with target space $\mathcal{X} = T^d$ with flat metric and B-field. In the string theory literature it is known as *Narain moduli space*.

If we look at two points of $\mathfrak{B}$ related by the right action of $\mathfrak{d} \in O_{\mathbb{Z}}(Q)$ then we get very different background data $\mathcal{E}$ and $\tilde{\mathcal{E}}$:

$$\tilde{\mathcal{E}} = \mathcal{E} \cdot \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{12.629}$$

Under this transformation E transforms by the fractional linear transformation:

$$\begin{aligned} \tilde{E} &= \tilde{\mathcal{E}}_{11}^{-1} \tilde{\mathcal{E}}_{12} \\ &= (\mathcal{E}_{11}\alpha + \mathcal{E}_{12}\gamma)^{-1}(\mathcal{E}_{11}\beta + \mathcal{E}_{12}\delta) \\ &= (\alpha + E\gamma)^{-1}(\beta + E\delta) \end{aligned} \tag{12.630}$$

This is the famous formula for T-duality transformations.

Finally, we claim that the space $\mathfrak{E}$ is essentially a real orthogonal group for a form of signature (d, d). To facilitate the proof we first make a small simplification: Actually it is redundant to introduce both a family of metrics g_{ij} on $\mathbb{R}^d$ and a family of embedded lattices $2\pi\Lambda \subset \mathbb{R}^d$ if we want to discuss the family of metrics on the torus T^d. So we simplify our formulae, without loss of generality, by choosing linear coordinates on the universal cover $\mathbb{R}^d$ so that $g_{ij} = \delta_{ij}$. [43] Then

$$\tilde{g}_{ab} := e_a{}^i e_b{}^i \tag{12.631}$$

is the (dimensionless) Gram matrix for the embedded lattice $\Lambda \subset \mathbb{R}^d$. Now

♣Use a better notation than $\tilde{g}_{ab}$. ♣

$$Q_0 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \tag{12.632}$$

and if we let

$$\mathcal{S} := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{pmatrix} \tag{12.633}$$

then $\mathcal{S} = \mathcal{S}^{-1} = \mathcal{S}^{\rm tr}$ and

$$\mathcal{S}^{tr} Q_0 \mathcal{S} = Q \tag{12.634}$$

(After all $\frac{1}{2}(\sigma^3 + \sigma^1)\sigma^3(\sigma^3 + \sigma^1) = \frac{1}{2}[\sigma^1, \sigma^3]\sigma^3 = \sigma^1$.)

A second way to read equation (12.595) is that it says that, up to a left- or right-multiplication by an invertible matrix, $\mathcal{E}$ is in a real orthogonal group of signature (d, d). Indeed, for any $\mathcal{E} \in \mathfrak{E}$ we have

$$\mathcal{E}\mathcal{S} \in O_{\mathbb{R}}(Q_0) \tag{12.635}$$

$$\mathcal{S}\mathcal{E} \in O_{\mathbb{R}}(Q) \tag{12.636}$$

Conversely, given one embedding of $II^{d,d}$ into $\mathbb{R}^{d;d}$ any rotation by $O_{\mathbb{R}}(Q_0)$ gives another embedding, so we can in fact identify $\mathfrak{E}$ with the entire orthogonal group, not just a subset. Thus, we finally obtain the standard form for the Narain moduli space as a double-coset:

$$\mathcal{N}_{d,d} := (O(d) \times O(d)) \backslash \mathfrak{E} / O_{\mathbb{Z}}(Q) \cong (O(d) \times O(d)) \backslash O_{\mathbb{R}}(Q) / O_{\mathbb{Z}}(Q) \tag{12.637}$$

Example 1: For $d = 1$, we have a 1×1 matrix $e_a{}^i = r := R/\ell_s$ so that $e_a{}^i = 1/r$. Therefore

$$\mathcal{E} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{r} & r \\ \frac{1}{r} & -r \end{pmatrix} \tag{12.638}$$

so there are two spanning vectors from the columns:

$$e = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{r} \\ \frac{1}{r} \end{pmatrix} \qquad\qquad f = \frac{1}{\sqrt{2}} \begin{pmatrix} r \\ -r \end{pmatrix} \tag{12.639}$$

Using the Lorentz metric on $\mathbb{R}^{1;1}$ we compute

$$e \cdot e = f \cdot f = 0 \qquad\qquad e \cdot f = f \cdot e = 1 \tag{12.640}$$

[43] We did not do this for the case of open strings on $\mathbb{R}^{2n}$ above because we wanted to consider a family of backgrounds determined by the SW limit, and it is more convenient to think of a family of metrics g_{ij} rather than a family of coordinate transformations. Now, however, our family will be given by the family of lattices Λ so also keeping g_{ij} is an unnecessary complication.

thus confirming that $\mathcal{E}$ describes an embedding of $II^{1,1}$ into $\mathbb{R}^{1;1}$. Moreover, $O_{\mathbb{Z}}(Q) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. After all, if we drop the integrality condition, we are talking about the Lorentz group in $1+1$ dimensions, and it has four connected components.

$$O_{\mathbb{Z}}(Q) = \{\pm \mathbf{1}, \pm\sigma^1\} \tag{12.641}$$

The element $-\mathbf{1}$ does not act effectively on $\mathfrak{B}$, but σ^1 takes

$$\frac{1}{\sqrt{2}}\begin{pmatrix} \frac{1}{r} & r \\ \frac{1}{r} & -r \end{pmatrix} \to \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{1}{r} & r \\ \frac{1}{r} & -r \end{pmatrix}\sigma^1 = \frac{1}{\sqrt{2}}\begin{pmatrix} r & \frac{1}{r} \\ -r & \frac{1}{r} \end{pmatrix} \tag{12.642}$$

and by an $O_{\mathbb{R}}(1) \times O_{\mathbb{R}}(1)$ transformation we can map this to:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} r & \frac{1}{r} \\ -r & \frac{1}{r} \end{pmatrix} \to \sigma^3\frac{1}{\sqrt{2}}\begin{pmatrix} r & \frac{1}{r} \\ -r & \frac{1}{r} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} r & \frac{1}{r} \\ r & -\frac{1}{r} \end{pmatrix} \tag{12.643}$$

so the net result is equivalent to $r \to 1/r$ and we simply recover the T-duality transformation of the simple introductory discussion above.

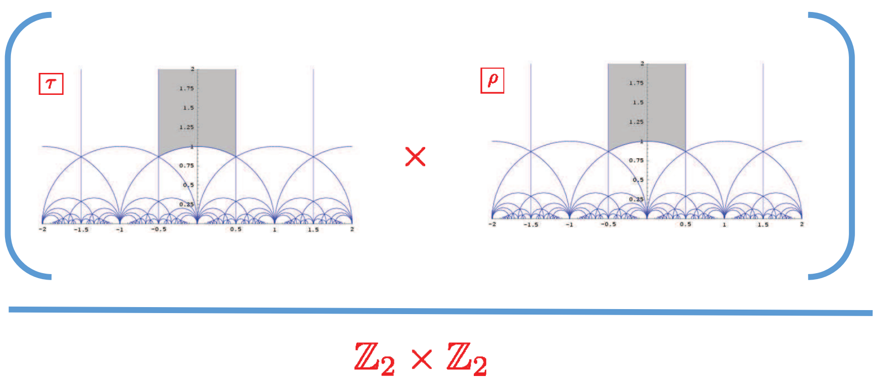


Figure 37: A picture of the Narain moduli space $\mathcal{N}_{2,2}$ for two-dimensional toroidal compactifications of the closed string.

Example 2: Already for $d = 2$ the story is much richer. To begin, recall that the vector space of $M_2(\mathbb{R})$ of 2×2 real matrices is isomorphic to $\mathbb{R}^4$ and has a natural quadratic form of signature $(2,2)$ given by the determinant. Therefore, left-action by $SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$ preserves this quadratic form. In this way we derive the exact sequence:

$$1 \to \mathbb{Z}_2 \to SL(2,\mathbb{R}) \times SL(2,\mathbb{R}) \xrightarrow{\psi} O^0_{\mathbb{R}}(Q) \to 1 \tag{12.644}$$

where the superscript indicates the connected component of 1 and the kernel is the group generated by $(-1,-1)$. To be more explicit suppose

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 \tag{12.645}$$

then define

$$\mathfrak{X}(x) = \begin{pmatrix} x_1 & x_2 \\ -x_4 & x_3 \end{pmatrix} \tag{12.646}$$

so that

$$2\mathrm{det}\mathfrak{X} = 2(x_1x_3 + x_2x_4) = x^{tr}Qx \tag{12.647}$$

Then we define the projection

$$\psi : SL(2,\mathbb{R}) \times SL(2,\mathbb{R}) \to O^0_{\mathbb{R}}(Q) \tag{12.648}$$

by

$$A\mathfrak{X}(x)B^{tr} = \mathfrak{X}(\psi(A,B) \cdot x) \tag{12.649}$$

Explicitly, if

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R}) \tag{12.650}$$

then

$$\psi(A,1) = \begin{pmatrix} a & 0 & 0 & -b \\ 0 & a & b & 0 \\ 0 & c & d & 0 \\ -c & 0 & 0 & d \end{pmatrix} \tag{12.651}$$

$$\psi(1,A) = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & d & -c \\ 0 & 0 & -b & a \end{pmatrix} \tag{12.652}$$

Now for $E = \tilde{g} - \tilde{b}$ define

$$\rho = -\tilde{b}_{12} + \mathrm{i}\sqrt{\mathrm{det}\tilde{g}} \tag{12.653}$$

$$\tau = \frac{\tilde{g}_{12} + \mathrm{i}\sqrt{\mathrm{det}\tilde{g}}}{\tilde{g}_{22}} \tag{12.654}$$

These have the interpretation of the "complexified Kähler class" and complex structure of the target space torus T^2. Note that, importantly, both τ and ρ have positive imaginary part. A set of generators for $O_{\mathbb{Z}}(Q)$ is given in Appendix B. Using these one can show that the group action on $(\tau,\rho) \in \mathcal{H}\times\mathcal{H}$ is generated by

$$(\tau,\rho) \to \left(\frac{a\tau+b}{c\tau+d},\rho\right) \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \tag{12.655}$$

$$(\tau,\rho) \to \left(\tau,\frac{a\rho+b}{c\rho+d}\right) \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \tag{12.656}$$

$$(\tau,\rho) \to (\rho,\tau) \tag{12.657}$$

$$(\tau,\rho) \to (-\bar{\tau},-\bar{\rho}) \tag{12.658}$$

Thus, $\mathcal{N}_{2,2}$ is isomorphic to $(\mathcal{F}\times\mathcal{F})/(\mathbb{Z}_2\times\mathbb{Z}_2)$ where $\mathcal{F}$ is a fundamental domain for the $PSL(2,\mathbb{Z})$ action on $\mathcal{H}$. The transformation (12.657) is the mirror symmetry transformation for T^2, considered as a one-dimensional Calabi-Yau manifold.

Remarks

1. Relation to another common convention. The fractional linear transformation of T-duality is often written as $E \to (\alpha E+\beta)(\gamma E+\delta)^{-1}$. On can obtain this particular form by using simple redefinitions. In particular, let

$$E' = E^{tr} \tag{12.659}$$

(i.e. we exchange $b \to -b$) and let

$$\mathfrak{d}' := \mathfrak{d}^{-1} = \begin{pmatrix} \delta^{\rm tr} & \beta^{\rm tr} \\ \gamma^{\rm tr} & \alpha^{\rm tr} \end{pmatrix} := \begin{pmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{pmatrix} \tag{12.660}$$

then $E \to (\alpha+E\gamma)^{-1}(\beta+E\delta)$ is equivalent to

$$E' \to (\alpha' E'+\beta')(\gamma' E'+\delta')^{-1} \tag{12.661}$$

2. Open strings and the SW limit. Let us now return to the open string. Recall that in the SW limit the open string field theory becomes noncommutative field theory on a Moyal space, and in the case of toroidal compactification we argued that it becomes noncommutative field theory for the noncommutative torus. What happens to E in the Seiberg-Witten limit? At this point we set $d=2n$ and reinstate g_{ij}, since we once again want to speak of a family of spacetime metrics on $\mathbb{R}^{2n}$. Recall the SW limit is defined by

$$\begin{aligned} \ell_s^2 &= \epsilon^{1/2}\ell_0^2 \\ g_{ij} &= \epsilon g_{0,ij} \\ b_{ij} &= \ell_s^2 B_{ij} \end{aligned} \tag{12.662}$$

with $\epsilon \to 0$ holding ℓ_0, g_0, B_{ij} fixed. To define the limit in toroidal compactification we also have to say that dimensionless vectors $e_a^{\ i}$ are of the form

$$e_a^{\ i} = \frac{R}{\ell_s}\bar{e}_a^{\ i} \qquad\qquad e_i^{\ a} = \frac{\ell_s}{R}\bar{e}_i^{\ a} \tag{12.663}$$

and we will define the Seiberg-Witten limit by holding R and $\bar{e}_a^{\ i}$, $\bar{e}_i^{\ a}$ fixed. Recall that in this limit the open string metric behaves like:

$$\begin{aligned} G^{ij} &\to -(B^{-1}g_0B^{-1})^{ij} \\ G_{ij} &\to -(Bg_0^{-1}B)_{ij} \end{aligned} \tag{12.664}$$

Now, we want to check that the limit is compatible with T-duality. First of all

$$E_{ab} = e_a^{\ j}(g_{jk} - b_{jk})e_b^{\ k} \to -R^2\bar{e}_a^{\ j}\bar{e}_b^{\ k}B_{jk} \tag{12.665}$$

becomes antisymmetric. Now recall that T-duality acts as fractional linear transformations on E_{ab}. It is not immediately obvious that the T-dual transform is antisymmetric. Nevertheless, using the conditions (12.624) one can readily show that if E_{ab} is antisymmetric then

$$\tilde{E} = (\alpha + E\gamma)^{-1}(\beta + E\delta) \tag{12.666}$$

is also antisymmetric. This suggests that it is consistent to send the closed string metric to zero, in spite of the naive expectation that a transformation taking $r \to 1/r$ would be inconsistent with the SW limit. Indeed, recall that the inverse of the open string metric is

$$G^{ab} = \left(\frac{1}{E}\right)_{\text{symmetric}} \tag{12.667}$$

so this goes to zero since E becomes antisymmetric. That is compatible with

$$G^{ab} = e_i^{\ a}e_j^{\ b}G^{ij} \tag{12.668}$$

and the fact that G^{ij} has an order one SW limit. Thus the inverse open string metric goes to zero in lattice units. This means that $G_{ab} \to \infty$. Again, let us check compatibility with T-duality: Under (12.666) we have

$$\begin{aligned} \tilde{G}^{ab} &= \frac{1}{2}\left(\frac{1}{\tilde{E}} + \frac{1}{\tilde{E}^{\text{tr}}}\right) \\ &= \frac{1}{2}(\beta + E\delta)^{-1}(E + E^{tr})(\beta + E\delta)^{-1,\text{tr}} \\ &= \frac{1}{2}(\beta + E\delta)^{-1}E(E^{-1} + E^{-1,\text{tr}})E^{\text{tr}}(\beta + E\delta)^{-1,\text{tr}} \\ &= \frac{1}{2}(\delta + E^{-1}\beta)^{-1}(E^{-1} + E^{-1,\text{tr}})(\delta + E^{-1}\beta)^{-1,\text{tr}} \\ &= (\delta + E^{-1}\beta)^{-1}G^{-1}(\delta + E^{-1}\beta)^{-1,\text{tr}} \end{aligned} \tag{12.669}$$

Therefore, $G^{ab} \to 0$ in any duality frame so long as $\delta + E^{-1}\beta$ is invertible. (However, $\delta + E^{-1}\beta$ is not always invertible. In fact, if E is a rational antisymmetric matrix then

$E\delta+\beta$ will fail to be invertible for some duality transformations. See the next remark.) Note that line two of the above equation also shows that $\tilde{G}^{-1} = (\beta + E\delta)^{-1}\tilde{g}(\beta + E\delta)^{-1,\mathrm{tr}}$. Sice $\tilde{g}_{ab} = e_a^{\ i} e_b^{\ \tilde{j}} g_{ij} = \epsilon^{1/2}\tilde{g}_{0,ab} \to 0$ there is no contradiction. Therefore, in any duality frame we will get a noncommutative algebra of functions as the limit of the vertex operator algebra. [44] Recalling that B_{ij} controls the noncommutativity parameter of the noncommutative torus algebra we have arrived at a very elegant physical explanation of Rieffel's theorem on the isomorphism of C^*-algebras!

3. Ergodic Action Of T-Duality On The Boundary Of Narain Moduli Space. Using the property that $\tilde{g}_{ab}$ is positive definite one can show that the action of the T-duality group on $\mathfrak{B}$ is properly discontinuous. [45] However, the SW limit takes E to the "boundary" of $\mathfrak{B}$. Typically, arithmetic groups $G_{\mathbb{Z}}$ have an ergodic action on the boundary of noncompact domains of the form $K\backslash G$. For example, $SL(2,\mathbb{Z})$ acts properly discontinuously on $SO(2)\backslash SL(2,\mathbb{R}) \cong \mathcal{H}$, but on the boundary of $\mathcal{H}$, namely the real line: $\mathbb{R} = \partial\overline{\mathcal{H}}$, the action

$$x \to \frac{ax+b}{cx+d} \tag{12.670}$$

is not properly discontinuous. For example the rational numbers form one dense orbit. Now consider the SW limit in the case of $d=2$. Of course τ is conformally invariant and doesn't vary with ϵ, but

$$\rho \to \tilde{B}_{0,12} \tag{12.671}$$

so ρ reduces to a real number. The duality group action is therefore ergodic. More generally, if we define a "boundary" of $\mathcal{B}$ by allowing $\tilde{g}_{ab}$ to become degenerate (in particular to become zero) then the action of $O_{\mathbb{Z}}(Q)$ on this boundary will be ergodic. The SW limit takes $\tilde{g}_{ab} \to 0$ so E becomes an element of the boundary. This is also why the action on the open string metric is not always well-defined: To get a well-defined action on the boundary we must add points at infinity, just as in the $SL(2,\mathbb{Z})$ action on the boundary of the upper half-plane: To get a well-defined action one must add the point at infinity and consider the boundary to be $\mathbb{RP}^1$.

Note that in the SW limit the bundle of CFT's becomes a bundle over a noncommutative manifold.

The above facts are also related to the fact that toroidal compactification on Lorentzian signature target space tori also has an ergodic action of the T-duality group [36]. Indeed, for a two-dimensional torus the conformal classes of Lorentzian signature is defined by the foliation by left- and right-moving lightrays. These describe a pair of ♣say how ♣

[44] One should also check that the string coupling does not blow up. See [47], p.68 for this.

[45] Recall that the action of a discrete topological group G on a topological space X is *properly discontinuous* if the map $G\times X \to X\times X$ defined by $(g,x) \to (g\cdot x, x)$ is proper. If X is locally compact then an equivalent statement is that for every compact $K \subset X$ the set of g with $g\cdot K \cap K \neq \emptyset$ is finite. If a discrete group has a properly discontinuous action on a Hausdorff manifold then the quotient is a Hausdorff orbifold.

points $\tau_\pm$ on the boundary of the upper half-plane and, as we remarked, the action of $SL(2,\mathbb{Z})$ on such pairs of points is ergodic. (In fact, the idea that there would be a role for noncommutative geometry in toroidal compactification was predicted in [36] based on the fact that the boundary of Narain moduli space is a noncommutative manifold. This should not be confused with Seiberg-Witten's statement that the target space torus, for a fixed point path in Narain moduli space (defined by the SW limit) becomes a noncommutative manifold.)

4. Full proof of T-duality. The main claim above about the Narain moduli space being $\mathcal{N}_{d,d}$ was only partially justified by the arguments above. We only checked that the spectrum of the Hamiltonian on the circle descends to this space. For the full conformal field theory one really needs to show that there is an isomorphism between the vertex operator algebras. This is actually not difficult since it is implemented by interpreting $O(d,d;\mathbb{R})$ as a group of Bogoliubov transformations mixing left-movers with right-movers. The result is that there is actually an equivariant bundle of conformal field theories over $\mathfrak{B}$ descending to a bundle over $\mathcal{N}_{d,d}$. In fact, this bundle of CFT's comes equipped with a natural equivariant connection. For details see [36, 41].

5. Rational Conformal Field Theories. Let

$$\pi_L : \mathbb{R}^{d;d} \to \mathbb{R}^{d;0} \qquad\qquad \pi_R : \mathbb{R}^{d;d} \to \mathbb{R}^{0;d} \tag{12.672}$$

be the projections to the positive definite and negative definite subspaces of $\mathbb{R}^{d;d}$. In general the projection of the zeromode lattice Γ is a dense set of points. For example, in the $d=1$ case

$$\pi_L(e) = \frac{r}{\sqrt{2}} \qquad \pi_L(f) = \frac{r^{-1}}{\sqrt{2}} \tag{12.673}$$

Unless $r^2 \in \mathbb{Q}$ the projection of these two vectors will generate a dense subgroup of the real line. However, when E_{ab} is a matrix of rational numbers the left and right projections are, separately, crystals in $\mathbb{R}^d$. To see this we define:

$$\tilde{p}_{L,a} := \tilde{g}_{ab} e_i{}^b \ell_s p_L^i \qquad\qquad \tilde{p}_{R,a} := \tilde{g}_{ab} e_i{}^b \ell_s p_R^i \tag{12.674}$$

so that

$$\tilde{g}^{ab}\tilde{p}_{L,a}\tilde{p}_{L,b} = (\ell_s p_L^i) g_{ij} (\ell_s p_R^j) \tag{12.675}$$

and

$$\frac{1}{\sqrt{2}}\begin{pmatrix} \tilde{p}_{L,a} \\ \tilde{p}_{R,a} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbf{1} & E \\ \mathbf{1} & -E^{\mathrm{tr}} \end{pmatrix}\begin{pmatrix} n_a \\ w^a \end{pmatrix} = \tilde{\mathcal{E}}\begin{pmatrix} n_a \\ w^a \end{pmatrix} \tag{12.676}$$

where $\tilde{\mathcal{E}}^{\mathrm{tr}} Q_0 \tilde{\mathcal{E}} = Q$ but now with

$$Q_0 \to \begin{pmatrix} \tilde{g}^{ab} & 0 \\ 0 & -\tilde{g}^{ab} \end{pmatrix} \tag{12.677}$$

Now, we claim that if E_{ab} is a matrix of rational numbers then there is a rank d lattice of "purely left-moving vectors" in Γ, that is, vectors that project to zero under π_R. The equation $\pi_R(\mathbf{p}) = 0$ is equivalent to the system of Diophantine equations:

$$\tilde{p}_{R,a} = n_a - (E^{\rm tr})_{ab} w^b = 0 \qquad (12.678)$$

for integers n_a, w^b. For such solutions of (12.678) we have

$$\tilde{p}_{L,a} = 2\tilde{g}_{ab} w^b \qquad (12.679)$$

The set of solutions $(n_a, w^a) \in \mathbb{Z}^{2d}$ of (12.678) forms a subgroup. It is a finitely generated torsion free abelian group and the rank is at most d since the n_a are determined from the w^a. In fact the rank is exactly d since, if we write the matrix elements as fractions in lowest terms $E_{ab} = p_{ab}/q_{ab}$ then, taking all the w^b to be divisible by the LCM, N of the q_{ab} we can solve the equations for integers n_a. That is, if we take all $w^a = N\bar{w}^a$ with arbitrary integers $\bar{w}^a$ we get a sub-lattice of the solution lattice. Let us denote by $\Gamma_L \subset \mathbb{R}^{d;0}$ the lattice of purely left-moving vectors $\mathbf{p}$. The lattice Γ_L is even, but in general is not unimodular. [46] To see that it is even we use (12.598) to write

$$\mathbf{p}^2 = \frac{1}{2}\tilde{p}_{L,a}\tilde{g}^{ab}\tilde{p}_{L,b} = 2w^a \tilde{g}_{ab}\tilde{w}^b \in 2N\mathbb{Z}. \qquad (12.680)$$

In an entirely analogous way, there is also a rank d lattice $\Gamma_R \subset \mathbb{R}^{0;d}$ of purely rightmoving momenta. Therefore $\Gamma_{LR} := \Gamma_L \oplus \Gamma_R$ is a finite index sublattice of Γ. The index grows with N and is a highly discontinuous function on $\mathcal{N}_{d,d}$. We can write $\Gamma = \amalg_s(\Gamma_{LR} + \mathbf{p}_s)$ for a finite set of glue vectors $\mathbf{p}_s$. The significance of these points for CFT is that there are integer spin purely (anti-)holomorphic vertex operators:

$$\exp[\mathrm{i}\tilde{p}_{L,a}\xi^a_L](z) \otimes 1 \qquad\qquad 1 \otimes \exp[\mathrm{i}\tilde{p}_{R,a}\xi^a_R](\bar{z}) \qquad (12.681)$$

that "enhance" the $\mathfrak{u}(1)_L^{\oplus d} \oplus \mathfrak{u}(1)_R^{\oplus d}$ current algebra that is present for all the CFT's parametrized by $\mathcal{N}_{d,d}$. These "enhancing" vertex operators correspond to states in the CFT with

$$L_0 = \frac{1}{4}\tilde{p}_{L,a}\tilde{g}^{ab}\tilde{p}_{L,b} = w^a\tilde{g}_{ab}w^b \qquad\qquad \tilde{L}_0 = 0 \qquad (12.682)$$

$$L_0 = 0 \qquad\qquad \tilde{L}_0 = \frac{1}{4}\tilde{p}_{R,a}\tilde{g}^{ab}\tilde{p}_{R,b} = w^a\tilde{g}_{ab}w^b \qquad (12.683)$$

respectively. Even though $\tilde{g}_{ab}$ is a rational number these conformal dimensions will be integral.

6. Remarks on the genus one partition function and RCFT. The torus partition function is

$$Z = (\eta\bar{\eta})^{-d} \sum_{\mathbf{p}\in\Gamma} q^{\frac{1}{4}(\ell_s p_L)^2} \bar{q}^{\frac{1}{4}(\ell_s p_R)^2} := \frac{\Theta_\Gamma(\tau,\bar{\tau})}{\eta^d\bar{\eta}^d} \qquad (12.684)$$

[46] It can happen that Γ has crystallographic symmetries but the projections π_L, π_R are not crystals, but rather *quasicrystals*. For a nice example see [24].

The numerator is known as a *Siegel-Narain theta function.*

In general if $\Lambda \subset \mathbb{R}^{b_+;b_-}$ is an embedded lattice in a pseudo-Euclidean space with signature $(+1^{b_+}, -1^{b_-})$ then we can define the general Siegel-Narain theta function

$$\Theta_\Lambda(\tau,\bar\tau;\alpha,\beta;\xi) := \exp[\frac{\pi}{2y}(\xi_+^2 - \xi_-^2)] \sum_{\lambda\in\Lambda} \exp\left\{ i\pi\tau(\lambda+\beta)_+^2 + i\pi\bar\tau(\lambda+\beta)_-^2 + 2\pi i(\lambda+\beta,\xi) - 2\pi i(\lambda + \frac{1}{2}\beta,\alpha) \right\} \tag{12.685}$$

where $\mathrm{Im}\tau = y$, and $\lambda_\pm$ is the projection of a vector into the positive (resp. negative)-definite subspaces. The main transformation law is:

$$\Theta_\Lambda(-1/\tau,-1/\bar\tau;\alpha,\beta;\frac{\xi_+}{\tau}+\frac{\xi_-}{\bar\tau}) = \sqrt{\frac{1}{|\mathcal{D}|}}(-i\tau)^{b_+/2}(i\bar\tau)^{b_-/2}\Theta_{\Lambda^*}(\tau,\bar\tau;\beta,-\alpha;\xi) \tag{12.686}$$

where Λ^* is the dual lattice, and $\mathcal{D} = \Lambda^*/\Lambda$ is a finite abelian group known as the *discriminant group.* Equation (12.686) can be proven straightforwardly by using the Poisson summation formula.

In the case of toroidal bosonic string compactification Γ is even and unimodular. It then follows from (12.686) that the partition function (12.684) is modular invariant.

In the case when E_{ab} is a matrix of rational numbers, that is, for rational conformal field theories, Z can be written as a finite sum of holomorphic times anti-holomorphic "conformal blocks":

$$Z = \sum_{s=1}^{N} Z_s(\tau)\tilde{Z}_s(\bar\tau) \tag{12.687}$$

for some integer N. the Siegel-Narain theta function factorizes as

$$\Theta_\Gamma = \sum_s \Theta_{\Gamma_L+\mathbf{p}_{s,L}}(\tau)\Theta_{\Gamma_R+\mathbf{p}_{s,R}}(\bar\tau) \tag{12.688}$$

and

$$Z_s(\tau) = \frac{\Theta_{\Gamma_L+\mathbf{p}_{s,L}}(\tau)}{\eta^d} \tag{12.689}$$

transform in a finite-dimensional unitary representation of $SL(2,\mathbb{Z})$. The $\tilde{Z}_s(\bar\tau)$ transform in a dual representation so that $Z(\tau,\bar\tau)$ is modular invariant.

♣Say this in more detail. ♣

The $Z_s(\tau)$ are examples of *conformal blocks* of an RCFT. In general, the correlation functions of all operators on all Riemann surfaces factorize into finite sums of holomorphic times anti-holomorphic objects. The vector space of these objects provides a finite-dimensional projective representation of the mapping class group of a punctured Riemann surface. These finite-dimensional spaces can also be identified with the space of states of a $2+1$ dimensional TQFT.

7. Crystallographic symmetries and orbifold points. At the rational points, where $\pi_L(\Gamma)$ and $\pi_R(\Gamma)$ are crystals these crystals can have nontrivial crystallographic symmetries.

That is, there can be elements of $O_{\mathbb{R}}(g) \times O_{\mathbb{R}}(g)$ that are equivalent to change of basis:

$$R\mathcal{E} = \mathcal{E}\mathfrak{d} \tag{12.690}$$

where $R \in O_{\mathbb{R}}(g) \times O_{\mathbb{R}}(g)$ and $\mathfrak{d} \in O_{\mathbb{Z}}(Q)$. The set of R's for which this holds will be a finite group $\mathcal{P}$ because they also form the point group for the lattice $\Gamma_L \oplus \Gamma_R$ with positive definite signature. At points where there is nontrivial crystallographic symmetry $\mathcal{N}_{d,d}$ has orbifold singularities. We have different crystallographic groups $\mathcal{P}$ at different points. $\mathcal{N}_{d,d}$ is a good example of a moduli *stack*. For example for $d = 1$, $\mathcal{N}_{1,1} = \mathbb{R}_+/\mathbb{Z}_2$ and there is a single $\mathbb{Z}_2$ orbifold point at $r = 1$. For $d = 2$ the story is much richer. The quotient $\mathcal{H}/SL(2,\mathbb{Z})$ already has $\mathbb{Z}_2$ and $\mathbb{Z}_3$ orbifold points at $\tau = \mathrm{i}, e^{\mathrm{i}\pi/3}$. [47] Using these points for τ or ρ we construct 4 components of complex codimension one orbifold loci in $\mathcal{N}_{2,2}$. In addition there is a locus of $\mathbb{Z}_2$ singularities along $\tau = \rho$ due to the mirror symmetry transformation.

8. A Special Example With Special Crystallographic Symmetry. At any point with purely left- and/or right-acting crystallographic symmetries we have orbifold points. As a dramatic example, consider the case of $d = 24$. One very interesting point is

$$\Gamma = (\Lambda_{\mathrm{Leech}}, 0) \oplus (0, \Lambda_{\mathrm{Leech}}) \tag{12.691}$$

where Λ_{Leech} is the famous *Leech lattice*. It is the unique even unimodular integral lattice of rank 24 with no vectors of length-squared $= 2$. Now, crystallographic symmetries in $O_{\mathbb{R}}(g) \times O_{\mathbb{R}}(g)$ are equivalent to a product of Conway groups.

9. Points with enhanced level one affine Lie algebra symmetry. An important set of special points in $\mathcal{N}_{d,d}$ can be constructed as follows. Let $\mathfrak{g}$ be a semi-simple Lie algebra of rank d where the summands are simply-laced: A_n, D_n or E_n. Embed the weight lattice $\Lambda_{wt}(\mathfrak{g}) \subset \mathbb{R}^d$ so that the root lattice is generated by root vectors of square-length two. Consider the rank $2d$ embedded lattice

$$\Gamma(\mathfrak{g}) \subset (\Lambda_{wt}(\mathfrak{g}); 0) \oplus (0; \Lambda_{wt}(\mathfrak{g})) \subset \mathbb{R}^{d;d} \tag{12.692}$$

where $\Gamma(\mathfrak{g})$ is defined by the condition that it contains the vectors $\mathbf{p}$ such that $p_L - p_R \in \Lambda_{rt}(\mathfrak{g})$. We claim that $\Gamma(\mathfrak{g}) \subset \mathbb{R}^{d;d}$ is an even unimodular lattice and hence defines a point in $\mathcal{N}_{d,d}$. We will call these $\mathfrak{g}$-points. To see that $\Gamma(\mathfrak{g})$ is an even lattice note that a general vector is of the form $(p_L; p_R) = (q_L + \lambda; q_R + \lambda)$ where q_L, q_R are in the root lattice and λ is one of the fundamental weights. Then Γ is even because: [48]

$$\mathbf{p}^2 = q_L^2 - q_R^2 + 2(q_L - q_R) \cdot \lambda \in 2\mathbb{Z} \tag{12.693}$$

$\Gamma(\mathfrak{g})$ is also unimodular. Perhaps the best way to see that is to make a modular transformation on the Siegel-Narain theta function. According to our formula above

[47] It is in fact the moduli stack of elliptic curves.

[48] It is convenient to take $\ell_s = \sqrt{2}$ and the standard Euclidean metric for g_{ij}.

(with $\xi = \alpha = 0$):

$$\begin{aligned}\Theta_{\Lambda_{rt}+\lambda}(-1/\tau) &= \frac{1}{\sqrt{|\mathcal{Z}|}}(-i\tau)^{d/2}\sum_{\tilde\lambda\in\Lambda_{wt}} e^{\mathrm{i}\pi\tau\tilde\lambda^2-2\pi\mathrm{i}(\tilde\lambda,\lambda)} \\ &= \frac{1}{\sqrt{|\mathcal{Z}|}}(-i\tau)^{d/2}\sum_{[\tilde\lambda]\in\Lambda_{wt}/\Lambda_{rt}}\sum_{q\in\Lambda_{rt}} e^{\mathrm{i}\pi\tau(q+\tilde\lambda)^2-2\pi\mathrm{i}(q+\tilde\lambda,\lambda)}\end{aligned} \tag{12.694}$$

where $\mathcal{Z} = \Lambda_{wt}/\Lambda_{rt}$ is isomorphic to the center of the simply connected Lie group G with Lie algebra $\mathfrak{g}$. We choose a set of fundamental weights λ_s representing the elements of this group and define:

$$Z_s(\tau) = \frac{\Theta_{\Lambda_{rt}+\lambda_s}}{\eta^d} \tag{12.695}$$

Then the modular group representation is generated by

$$Z_s(-1/\tau) = \frac{1}{\sqrt{|\mathcal{Z}|}}\sum_{s'} e^{-2\pi\mathrm{i}(\lambda_s,\lambda_{s'})} Z_{s'}(\tau) \tag{12.696}$$

$$Z_s(\tau+1) = e^{2\pi\mathrm{i}(\frac{1}{2}(\lambda_s,\lambda_s)-d/24)} Z_s(\tau) \tag{12.697}$$

It is worthwhile checking that we really do get a representation of the modular group. The generator S is represented by a finite Fourier transform, so in indeed $S^2=-1$ is satisfied. The harder relation to check is

$$(ST)^3 = S^2 \tag{12.698}$$

This should be written as

$$S^{-1}TS = T^{-1}S^{-1}T^{-1} \tag{12.699}$$

Then the relation is easily checked using the Gauss-Milgram formula: If $q(x)$ is a quadratic refinement on the discriminant group of an integral lattice with $q(0)=0$ then

$$\frac{1}{\sqrt{|\mathcal{D}|}}\sum_{x\in\mathcal{D}} e^{2\pi\mathrm{i}q(x)} = e^{2\pi\mathrm{i}\sigma/8} \tag{12.700}$$

Applying this to the present case we have

$$\sum_{\lambda_s} e^{\mathrm{i}\pi(\lambda_s,\lambda_s)} = \sqrt{|\mathcal{Z}|}e^{2\pi\mathrm{i}d/8} \tag{12.701}$$

from which it is easy to check the nontrivial relation of the modular group. We conclude that $\sum_s Z_s(\tau)Z_s(\bar\tau)$ is modular invariant and hence $\Gamma(\mathfrak{g})$ is unimodular.

♣Explain where we used the assumption that $\mathfrak{g}$ is simply laced. ♣

The $\mathfrak{g}$-points define orbifold points of $\mathcal{N}_{d,d}$ since they are fixed points under the left-action on $\mathfrak{L}$ of $(w_1;w_2)\in O_{\mathbb{R}}(g)\times O_{\mathbb{R}}(g)$ where $w_i\in W(\mathfrak{g})$. For example,

$$\begin{aligned}(w_1;1):(q_L+\lambda;q_R+\lambda) &\mapsto (w_1q_L+w_1\lambda;q_R+\lambda) \\ &\mapsto (w_1q_L+q_L'+\lambda;q_R+\lambda)\end{aligned} \tag{12.702}$$

since, for any element of the weight lattice and for any element of the Weyl group $\lambda - w \cdot \lambda \in \Lambda_{rt}$. This follows since for Weyl reflections:

$$\sigma_\alpha \cdot \lambda - \lambda = -2\frac{(\alpha,\lambda)}{(\alpha,\alpha)}\alpha \in \Lambda_{rt} \tag{12.703}$$

At these points the conformal field theory has enhanced Lie algebra symmetry, generalizing the discussion we gave above for $d = 1$, $R = \ell_s$. Among the holomorphic vertex operators are

$$\exp[\mathrm{i}\alpha \cdot \xi_L](z) \otimes 1 \qquad\qquad 1 \otimes \exp[\mathrm{i}\alpha \cdot \xi_R](\bar{z}) \tag{12.704}$$

where $\alpha \in \Lambda_{rt}(\mathfrak{g})$. These operators have conformal dimensions $(1,0)$ and $(0,1)$, respectively and combine with the $\mathfrak{u}(1)$ currents to form a symmetry

$$\mathfrak{g}^{(1)}_{k=1} \oplus \mathfrak{g}^{(1)}_{k=1} \tag{12.705}$$

That is, at these points the conformal field theory becomes the level $k = 1$ WZW model for G.

10. <u>A Tale Of Two Field Theories</u>. In order to appreciate better the significance of the $\mathfrak{g}$-points we have just identified we need to recall briefly some basic ideas about string theory. This is a theory that simultaneously describes <u>two</u> field theories. One is quantum gravity in two-dimensions. One couples a CFT to 2-dimensional quantum gravity and integrates over all topologies and metrics. On the other hand, when the CFT is a nonlinear sigma model with target space $\mathcal{X}$, the amplitudes of string theory can also be interpreted as S-matrix amplitudes for scattering of particles on $\mathcal{X}$. One can then attempt to associate a quantum field theory on the target space $\mathcal{X}$ which likewise describes the same S-matrix amplitudes. Such a target space QFT is known as a *string field theory*. We will denote it as $QFT(\mathcal{X})$. Skipping over an enormous number of subtleties the basic ideas for writing down a string field theory are the following:

   On-shell particles are described, at tree level by the Lie algebra cohomology for the Virasoro algebra with values in the Virasoro module provided by the CFT. In order for the differential to be well-defined the total central charge of the CFT must be $c = 26$. Knowing the particles we can write a quadratic action $\sim \int \Psi Q \Psi$ where Ψ, the string field is a general element of the closed string Hilbert space. The interactions are then obtained by computing correlation functions of Ψ and integrating over suitable portions of the moduli space of Riemann surfaces.

11. <u>Spacetime Yang-Mills Theory</u>. Suppose the target space is $\mathcal{X} = \mathbb{R}^D \times T^d$, and we have a constant metric and B-field on T^d.

   At generic points on $\mathcal{N}_{d,d}$ the target space theory contains a Yang-Mills theory with abelian gauge group $U(1)^{2d}$. The vertex operators:

$$\int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}k\hat{y}(z,\bar{z})} \otimes (\partial x^i(z) \otimes \epsilon_\mu \bar{\partial}\hat{y}^\mu_R(\bar{z}))A_{i,\epsilon}(k) \in \mathcal{H}_{\text{closed string}} \tag{12.706}$$

and

$$\int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}k\hat{y}(z,\bar{z})} \otimes (\epsilon_\mu \partial \hat{y}^\mu_L(z) \otimes \bar{\partial} x^i_R(\bar{z}))\tilde{A}_{i,\epsilon}(k) \in \mathcal{H}_{\text{closed string}} \tag{12.707}$$

where $\hat{y}(z,\bar{z})$ are the coordinates of the string describing the mapping of $\Sigma \to \mathbb{R}^D$, while $\epsilon \in T^*\mathbb{R}^D$ is a cotangent vector on $\mathbb{R}^D$ and the coefficients in the expansion $A_{i,\epsilon}(k), \tilde{A}_{i,\epsilon}(k)$ define Fourier modes of <u>spacetime fields</u> $A_{i,\mu}(y), \tilde{A}_{i,\mu}(y)$ on the space $y \in \mathbb{R}^D$ with polarization ϵ:

$$\epsilon^\mu A_{i,\mu}(y) := \int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}ky} A_{i,\epsilon}(k) \qquad i = 1, \dots, d \tag{12.708}$$

$$\epsilon^\mu \tilde{A}_{i,\mu}(y) := \int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}ky} A_{i,\epsilon}(k) \qquad i = 1, \dots, d \tag{12.709}$$

By computing string theory scattering amplitudes with such vertex operators we learn that the QFT of these fields on $\mathbb{R}^D$ is just a generalized Maxwell theory with gauge group $U(1)^d \times U(1)^d$ and variable coupling constants:

$$\int_{\mathbb{R}^D} \sum_{I,J=1}^{2d} \frac{1}{e^2_{IJ}(n)} F_I * F_J \subset S_{QFT(\mathcal{X})} \tag{12.710}$$

where the coupling constants $e^2_{IJ}(n)$ are functions on the Narain moduli space: $n \in \mathcal{N}_{d,d}$.

The notation $S \subset S_{QFT(\mathcal{X})}$ means that we can write the action $S_{QFT(\mathcal{X})}$ as a sum of terms, linearly independent under field redefinition, such that S is one of the terms.

However, at the points on $\mathcal{N}_{d,d}$ defined by $\Gamma(\mathfrak{g})$ we can also form the vertex operators:

$$\Psi = \int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}k\hat{y}} \otimes (\exp[\mathrm{i}\alpha \cdot \xi_L](z) \otimes \epsilon \partial \hat{y}_R(\bar{z}))) E_{\alpha,\epsilon}(k) \tag{12.711}$$

where α is a root vector, so $\alpha^2 = 2$, and the coefficients $E_\alpha(k)$ define Fourier components of <u>spacetime fields</u> on the space $\mathbb{R}^D$ that we can call $A^\alpha_\mu(y)$. By computing string theory scattering amplitudes with such vertex operators we learn that the QFT of these fields on $\mathbb{R}^D$ is a <u>nonabelian</u> Yang-Mills theory with the $A^\alpha(y)$ playing the role of Fourier modes of "W-bosons." The gauge group of the theory at these points will be enhanced to $G \times G$ where G is an exponentiated form of $\mathfrak{g}$. Historically, this mechanism for producing nonabelian Yang-Mills gauge fields on $\mathcal{X}$ was extremely important in the development of techniques for constructing string theories intended to describe nature. A modification of this mechanism was used in the construction of the heterotic string, for example.

Another important conceptual lesson we learn from this is that the duality transformations are generalizations of gauge transformations in string theory, since the Weyl transformations are part of the gauge group of the target space Yang-Mills theory. For more details about these special points in $\mathcal{N}_{d,d}$ see [21].

♣Should explain that this shows we actually have a moduli <u>stack</u>. Also, describe the Zamolodchikov metric on this space? ♣

12. Low Energy Effective Couplings As Automorphic Forms For $O_{\mathbb{Z}}(Q)$. In addition to the spacetime gauge fields mentioned above there are also d^2 massless scalar fields associated with the vertex operators

$$\Psi = \int_{\hat{\mathbb{R}}^D} dk e^{\mathrm{i}k\cdot\hat{y}} \varphi^{ij}(k) \partial x_L^i(z) \otimes \partial x_R^j(\bar{z}) \tag{12.712}$$

where as usual the spacetime field is

$$\varphi^{ij}(y) = \int_{\hat{\mathbb{R}}^D} dk \varphi^{ij}(k) e^{\mathrm{i}ky} \tag{12.713}$$

These are fields in a nonlinear sigma model with $\mathbb{R}^D$ as the domain and $\mathcal{N}_{d,d}$ as the target:

$$\varphi : \mathbb{R}^D \to \mathcal{N}_{d,d} \tag{12.714}$$

The field φ^{ab} is a spacetime variation of the data E_{ab}. The Lagrangian for these fields is a nonlinear sigma model with action

$$\int_{\mathbb{R}^D} dy g_{ab,cd}(n) \partial_\mu \varphi^{ab} \partial^\mu \varphi^{cd} \subset S_{QFT(\mathcal{X})} \tag{12.715}$$

The couplings $g_{ab,cd}(n)$ are derived from a metric on $\mathcal{N}_{d,d}$ known as the *Zamolodchikov metric*. It is essentially just the homogeneous space matric on the double coset $O(d) \times O(d) \backslash O(d,d;\mathbb{R})/O(d,d;\mathbb{Z})$. In general, the effective coupling constants of the spacetime theory will be interesting automorphic forms for $O(d,d;\mathbb{Z})$. ♣Say more? ♣

13. Special (generating) elements of the T-duality group. Some special elements of the T-duality group $O_{\mathbb{Z}}(Q)$ have simple physical interpretations. Let us return to the version of the action given in (12.621): We can of course make a change of coordinates

$$\xi^a \to \tilde{\xi}^a = \alpha^a{}_b \xi^b \tag{12.716}$$

where, because of the periodicities $\xi^a \sim \xi^a + 2\pi$ we must have $\alpha \in GL(d,\mathbb{Z})$. This transformation clearly takes:

$$E \to \alpha^{tr} E \alpha \tag{12.717}$$

and corresponds to duality transformations in $O_{\mathbb{Z}}(Q)$ of the form:

$$\mathfrak{d} = \begin{pmatrix} \alpha^{tr,-1} & 0 \\ 0 & \alpha \end{pmatrix} \tag{12.718}$$

Similarly, an "obvious" T-duality transformation corresponds to $E \to E + \beta$, where β is an antisymmetric matrix of integers. Under this transformation the action changes by

$$S \to S + \frac{\mathrm{i}}{2\pi} \int \beta_{ab} d\xi^a \wedge d\xi^b \tag{12.719}$$

but the periods of $\beta_{ab} d\xi^a \wedge d\xi^b$ on $\mathcal{X}$ are in $(2\pi)^2\mathbb{Z}$, and hence there is no effect on the theory. These transformations are often called "B-field shifts." These correspond to elements of $O_{\mathbb{Z}}(Q)$ of the form

$$\mathfrak{d} = \begin{pmatrix} \mathbf{1} & \beta \\ 0 & \mathbf{1} \end{pmatrix} \tag{12.720}$$

Thus the couplings (or spacetime fields in $QFT(\mathcal{X})$) $\tilde{b}_{ab}$ are periodic variables.

Finally, a less-obvious set of transformations are given by

$$\mathfrak{d}_i = \begin{pmatrix} \mathbf{1} - e_{ii} & e_{ii} \\ e_{ii} & \mathbf{1} - e_{ii} \end{pmatrix} \tag{12.721}$$

for $i = 1, \ldots, d$. In the case of a square torus with zero B-field we have

$$E = \mathrm{Diag}\{r_1^2, \ldots, r_d^2\} \tag{12.722}$$

and $\mathfrak{d}_i$ takes $r_i \to 1/r_i$ holding the other radii fixed.

In Appendix B we show how to construct a set of generators for $O_{\mathbb{Z}}(Q)$. A corollary of this discussion shows that (12.718), (12.720), and (12.721) generate all of $O_{\mathbb{Z}}(Q)$. So if one wants to prove duality symmetry directly from a path integral it suffices to show the theory is invariant under these transformations. The invariance under (12.718) and (12.720) follows immediately from the action. The duality transformations (12.721) are much less obvious, and we will give a path integral argument for them in Section 12.9.4. In fact they are special cases of a more general transformation known as *Buscher duality.*

♣Comment on the fact that the duality group in superstring theory is actually $Pin(d, d; \mathbb{Z})$. This arises from the action in Ramond sectors. In the spacetime theory, the duality action must be defined on spacetime spinors. Note that just the $O(d) \times O(d)$ subgroup of $O(d, d)$ must be lifted, and therefore the entire group must be lifted to a double cover. ♣

14. SYZ Picture Of Mirror Symmetry.

12.9.4 Relation to electric-magnetic duality

One nice way to see that the partition functions of T-dual CFT's with target T^d are the same is to view T-duality as electromagnetic duality and implement the transformation in the path integral. In fact there is a significant generalization, known as *Buscher duality* [6] to a larger class of sigma models. The idea to relate it to electro-magnetic duality is in [44]. We will follow that discussion with a very slight improvement in the treatment of some global issues.

The proper context for the duality is that of general sigma models of the form:

$$S_E = \frac{1}{4\pi\ell_s^2} \int_\Sigma \Big[g_{ij} dx^i \wedge * dx^j - \mathrm{i} b_{ij} dx^i \wedge dx^j + 2\Phi(x)\mathfrak{e} \Big]. \tag{12.723}$$

Here g_{ij}, b_{ij} can be functions of x^i, and we have added another term. The last term has the Euler density of the worldsheet metric $\mathfrak{e} := \sqrt{h}\mathcal{R}^{(2)}(h)$ so that $\int_\Sigma \mathfrak{e} = 2\pi\chi(\Sigma)$ is the Euler character of Σ. The conjugate field $\Phi(x)$ on spacetime $\mathcal{X}$ is known as the "dilaton." Note that if $\Phi(x) = \Phi_0$ is constant then the partition function with Euclidean signature worldsheet is weighted by $\exp[-(\Phi_0/\ell_s^2)\chi(\Sigma)]$ so that the closed string coupling constant is $\exp[\Phi_0/\ell_s^2]$.

In order to implement Buscher duality we must assume there is a $U(1)$ action on $\mathcal{X}$, acting without fixed points, and acting as a symmetry of the background fields (g, b, Φ). What this means in practice is that we can choose (locally) target space coordinates (x^0, x^s) on $\mathcal{X}$ with $s = 2, \ldots, d$ so that g_{ij}, b_{ij}, and Φ are independent of x_0. Therefore we can write the Euclidean action as:

♣Strange convention. If s runs from 2 to d then the periodic coordinate should be x^1. Or perhaps use another letter emphasizing it is an angle? ♣

$$S_E = \frac{1}{4\pi\ell_s^2} \int_\Sigma \left[g_{00} dx^0 \wedge * dx^0 + 2 g_{0s} dx^0 \wedge * dx^s + g_{st} dx^s \wedge * dx^t \right] - \mathrm{i} \left[2 b_{0s} dx^0 \wedge dx^s + b_{st} dx^s \wedge dx^t \right] + 2\Phi\mathfrak{c} \tag{12.724}$$

where all background fields are functions of x^s, but not functions of x^0.

Now, there can be winding modes for the field x^0 and we can normalize it so that

$$\varphi := e^{\mathrm{i}x^0/\ell_s} \tag{12.725}$$

is single-valued on Σ.

A good way to discuss electromagnetic duality (for all generalized abelian gauge theories in all dimensions) is to introduce differential cohomology. In our extremely simple case the relevant differential cohomology group is just

$$\check{H}^1(\Sigma) := \mathrm{Map}(\Sigma, U(1)) \tag{12.726}$$

(We will assume the maps are differentiable.) Note that this is indeed an abelian group. To an element $\varphi \in \check{H}^1(\Sigma)$ we can associate

1. The *fieldstrength*
$$F(\varphi) := \frac{1}{2\pi\mathrm{i}} \varphi^{-1} d\varphi \in \Omega^1_{\mathbb{Z}}(\Sigma) \tag{12.727}$$
Here $\Omega^1_{\mathbb{Z}}(\Sigma)$ is the abelian group of 1-forms on Σ that have integral periods. Note that a differential form with integral periods is necessarily closed.

2. The *characteristic class*
$$a(\varphi) := \varphi^*(\omega) \in H^1(\Sigma; \mathbb{Z}) \tag{12.728}$$
where $\omega \in H^1(U(1); \mathbb{Z})$ is a generator. The characteristic class measures the winding numbers of φ around the various cycles in Σ.

This differential cohomology group, like all differential cohomology groups, fits in two (compatible) exact sequences:

$$0 \to H^0(\Sigma; \mathbb{R}/\mathbb{Z}) \to \check{H}^1(\Sigma) \xrightarrow{F} \Omega^1_{\mathbb{Z}}(\Sigma) \to 0 \tag{12.729}$$

$$0 \to \Omega^0(\Sigma)/\Omega^0_{\mathbb{Z}}(\Sigma) \to \check{H}^1(\Sigma) \xrightarrow{a} H^1(\Sigma; \mathbb{Z}) \to 0 \tag{12.730}$$

Here $H^0(\Sigma; \mathbb{R}/\mathbb{Z})$, the *flat fields* with zero fieldstrength are just the constant maps. This group is just a copy of $U(1)$. Meanwhile $\Omega^0_{\mathbb{Z}}(\Sigma)$ are the functions with integral "periods." These are just the constant functions given by an integer.

One can noncanonically write the abelian group $\check{H}^1(\Sigma)$ as a product of three groups:

$$\check{H}^1(\Sigma) = U(1) \times \mathbb{Z}^{b_1(\Sigma)} \times V \tag{12.731}$$

where V is an infinite-dimensional vector space. For example, if we choose a metric on Σ then we can decompose F into a harmonic piece and an exact form:

$$F = \omega + d\phi \tag{12.732}$$

where $\omega \in \mathcal{H}^1(\Sigma)$ is in the real vector space (of dimension $b_1(\Sigma)$) of harmonic forms and ϕ is a globally well-defined map $\phi : \Sigma \to \mathbb{R}$. Then we can take $U(1) = \ker(d)/\mathbb{Z}$, to circle of gauge inequivalent constant modes, and V is $(\ker d)^\perp$. Then, since the periods of F are integral ω actually lies in the full rank lattice $\mathcal{H}^1_{\mathbb{Z}}(\Sigma) \subset \mathcal{H}^1(\Sigma)$, and if we choose a basis for $H_1(\Sigma;\mathbb{Z})$ the we can use the periods to define an isomorphism $\mathcal{H}^1_{\mathbb{Z}}(\Sigma) \cong \mathbb{Z}^{b_1(\Sigma)}$.

Now we can write the action in terms of the fieldstrength $F = F(\varphi)$ as:

$$\begin{aligned} S_E &= S_E^1 + S_E^2 \\ &= \int_\Sigma \pi g_{00} F \wedge *F + g_{0s} F \wedge *d(x^s/\ell_s) - \mathrm{i} b_{0s} F \wedge d(x^s/\ell_s) \\ &+ \frac{1}{4\pi \ell_s^2} \int_\Sigma \Big[g_{st} dx^s * dx^t - \mathrm{i} b_{st} dx^s \wedge dx^t + 2\Phi \mathfrak{e} \Big] \end{aligned} \tag{12.733}$$

We have split the action into a piece S_E^1 that depends on x^0 and an action S_E^2 that does not depend on x^0. We are going to focus on the path integral over x^0, holding the remaining coordinates x^s fixed. Therefore, in doing this path integral we can treat the couplings as constant. The path integral measure is a rather formal object that we will denote $\mu(\varphi)$. Formally it is the Riemannian volume element on function space induced from the metric:

$$\parallel \delta x^0 \parallel^2_{g_{00}} := \int_\Sigma g_{00} \delta x^0 \wedge * \delta x^0 \tag{12.734}$$

We will just denote it as $\mu(\varphi)$. One important aspect of this measure is that it is a translationally invariant measure on the group $\check{H}^1(\Sigma)$.

The next step is to gauge the $U(1)$ symmetry. Thus we replace $F(\varphi)$ by

$$\mathcal{F} := F(\varphi) - A = d\left(\frac{x^0}{2\pi \ell_s}\right) - A \tag{12.735}$$

where $A \in \Omega^1(\Sigma)$ is a one-form. Conceptually, it is a one-form on a principal $U(1)$ bundle over Σ. However, this bundle is trivial so we can consider it to be a globally well-defined form. (See Remark 3 below for more about this.) We follow the physics convention and take A to be real. and we will integrate over A.

Next, we consider the following path integral:

$$\begin{aligned} \mathcal{I} = \mathcal{N}^{-1} & \int_{\check{H}^1(\Sigma)} \mu(\varphi) \int_{\check{H}^1(\Sigma)} \mu(\varphi_D) \int_{\Omega^1(\Sigma)} \mu(A) \\ & \exp\left[- \int_\Sigma [\pi g_{00} \mathcal{F} \wedge * \mathcal{F} + g_{0s} \mathcal{F} \wedge * d\xi^s - \mathrm{i} b_{0s} \mathcal{F} \wedge d\xi^s] + 2\pi \mathrm{i} \int_\Sigma A \wedge F_D \right] \end{aligned} \tag{12.736}$$

Here

1. $\varphi_D \in \check{H}^1(\Sigma)$ will be the dual field with fieldstrength $F_D = F(\varphi_D)$. The measure $\mu(\varphi_D)$ is defined as before, but with $g_{00} \to g_{00}^{-1}$.
2. We defined $\xi^s = x^s/\ell_s$ to simplify the notation.
3. The measure on $\mu(A)$ is defined formally using the metric $\parallel \delta A \parallel^2 = \int_\Sigma \delta A \wedge *\delta A$.
4. The normalization factor is, formally, given by

$$\mathcal{N} = \mathrm{vol}_{\,g_{00}}(\mathcal{H}^1_{\mathbb{Z}}) g_{00}^{-1/2} \tag{12.737}$$

   where $\mathcal{H}^1_{\mathbb{Z}}$ is the lattice of harmonic one-forms on Σ with integral periods, and again the volume form formally follows from the metric with g_{00} included, as above. The reason for this strange factor will become apparent from the derivation below. The volume of the lattice can be more rigorously treated by introducing a Gaussian suppression factor and using the volume of the divergence $\epsilon^{-b_1(\Sigma)/2}$ as $\epsilon \to 0$.

Now the idea is to do the path integral $\mathcal{I}$ in two different ways and thereby obtain a duality transformation.

The first evaluation does the integral over φ_D and then the integral over A. The result is the original path integral over x^0 with action S^1_E. The second path integral does the Gaussian integral over A. Then does the integral over φ. The result is the dual path integral over x^0_D. We now explain this in great detail:

First we do the integral over φ_D. We have

$$\begin{aligned}\int_{\check{H}^1(\Sigma)} \mu(\varphi_D) e^{2\pi \mathrm{i} \int_\Sigma A \wedge F_D} &= g_{00}^{-1/2} \delta(A^{nh}) \sum_{\omega \in \mathcal{H}^1_{\mathbb{Z}}} e^{2\pi \mathrm{i} \int A^h \wedge \omega} \\ &= g_{00}^{-1/2} \delta(A^{nh}) \sum_{\omega \in \mathcal{H}^1_{\mathbb{Z}}} \delta(A^h - \omega)\end{aligned} \tag{12.738}$$

Here we used the worldsheet metric to give our noncanonical decomposition of $\check{H}^1(\Sigma)$ as well as the orthogonal decomposition $A = A^h + A^{nh}$ into its harmonic and non-harmonic part. The factor $g_{00}^{-1/2}$ is the volume of the flat fields.

Next we can easily do the path integral over A by evaluating the δ-functions. The result is a sum over $\mathcal{H}^1_{\mathbb{Z}}$ with $\mathcal{F} = F(\varphi) - \omega$. But, precisely because ω has integral periods, and because $\mu(\varphi)$ is translation invariant we can shift away ω in each term of the sum. The result is

$$\mathcal{I} = \mathcal{N}^{-1} g_{00}^{-1/2} \mathrm{vol}_{\,g_{00}}(\mathcal{H}^1_{\mathbb{Z}}) \int_{\check{H}^1(\Sigma)} e^{-S^1_E} = \int_{\check{H}^1(\Sigma)} e^{-S^1_E} \tag{12.739}$$

That is, thanks to our choice of normalization in (12.737), $\mathcal{I}$ is just the x^0 path integral.

Now we turn to the second evaluation of $\mathcal{I}$. Now we first do the integral over A. This is just a Gaussian integral. For a fixed φ we can shift A so that $\mathcal{F}$ is just $-A$. To do the Gaussian integral most efficiently proceed as follows:

Pick complex coordinates on Σ so that $*dz = -\mathrm{i}dz$ and $*d\bar{z} = +\mathrm{i}d\bar{z}$. (This coincides with the standard orientation on the plane for $z = x^1 + \mathrm{i}x^2$.) Then decompose:

$$A = \alpha dz + \bar{\alpha} d\bar{z} \tag{12.740}$$

$$F_D = f_D dz + \bar{f}_D d\bar{z} \tag{12.741}$$

$$d\xi^s = \partial\xi^s dz + \bar{\partial}\xi^s d\bar{z} \tag{12.742}$$

Next, write out the action separating out $\mathrm{i}dz \wedge d\bar{z}$. Next find the stationary point with respect to α and $\bar{\alpha}$:

$$\begin{aligned} \alpha_* &= -g_{00}^{-1}(f_D - (g_{0s} + b_{0s})\partial(\xi^s/2\pi)) \\ \bar{\alpha}_* &= g_{00}^{-1}(\bar{f}_D - (g_{0s} - b_{0s})\bar{\partial}(\xi^s/2\pi)) \end{aligned} \tag{12.743}$$

Now, substitute back into the action. After some algebra one should find:

$$\begin{aligned} \tilde{S}_E^1 &= \int_\Sigma \left[\frac{\pi}{g_{00}} F_D \wedge *F_D - \frac{b_{0s}}{g_{00}} F_D \wedge *d\xi^s + \mathrm{i}\frac{g_{0s}}{g_{00}} F_D \wedge d\xi^s \right] \\ &- \frac{1}{4\pi\ell_s^2} \int_\Sigma \left[\frac{g_{t0}g_{0s} + b_{t0}b_{0s}}{g_{00}} dx^t \wedge *dx^s + \mathrm{i}\frac{b_{t0}g_{0s} + g_{t0}b_{0s}}{g_{00}} dx^t \wedge dx^s \right] \end{aligned} \tag{12.744}$$

Putting this together with S_E^2 we obtain the "Buscher rules":

$$\begin{aligned} \tilde{g}_{00} &= \frac{1}{g_{00}} \\ \tilde{g}_{0s} &= -\frac{b_{0s}}{g_{00}} \\ \tilde{b}_{0s} &= -\frac{g_{0s}}{g_{00}} \\ \tilde{g}_{st} &= g_{st} - \frac{g_{t0}g_{0s} + b_{t0}b_{0s}}{g_{00}} \\ \tilde{b}_{st} &= g_{st} - \frac{b_{t0}g_{0s} + g_{t0}b_{0s}}{g_{00}} \end{aligned} \tag{12.745}$$

The integral over A also produces a "one-loop determinant" which is, formally:

$$\left(\frac{1}{\sqrt{g_{00}}} \right)^{\dim\Omega^1(\Sigma)} \tag{12.746}$$

Now, the result of the Gaussian integral on A is independent of φ so we can now do the integral over φ to get an overall normalization constant:

$$\begin{aligned} \mathcal{N}^{-1} \left(\frac{1}{\sqrt{g_{00}}} \right)^{\dim\Omega^1(\Sigma)} \mathrm{vol}_{\,g_{00}}(\check{H}^1) &= g_{00}^{1/2} \left(\frac{1}{\sqrt{g_{00}}} \right)^{\dim\Omega^1(\Sigma)} \frac{\mathrm{vol}_{\,g_{00}}(\check{H}^1)}{\mathrm{vol}_{\,g_{00}}(\mathcal{H}^1_{\mathbb{Z}})} \\ &= g_{00}^{1/2} \left(\frac{1}{\sqrt{g_{00}}} \right)^{\dim\Omega^1(\Sigma)} \frac{g_{00}^{1/2}\mathrm{vol}_{\,g_{00}}(\mathcal{H}^1_{\mathbb{Z}})\mathrm{vol}_{\,g_{00}}(\Omega^1_{nh})}{\mathrm{vol}_{\,g_{00}}(\mathcal{H}^1_{\mathbb{Z}})} \\ &= g_{00} \left(\frac{1}{\sqrt{g_{00}}} \right)^{\dim\Omega^1(\Sigma)} (\sqrt{g_{00}})^{\dim\Omega^1_{nh}} \\ &= g_{00} g_{00}^{-\frac{1}{2}\dim\mathcal{H}^1} \\ &= g_{00}^{\frac{1}{2}\chi(\Sigma)} \end{aligned} \tag{12.747}$$

Thus, the string coupling constant also changes:

$$\tilde{g}_{\text{string}}^{-\chi(\Sigma)} = g_{\text{string}}^{-\chi(\Sigma)} g_{00}^{\frac{1}{2}\chi(\Sigma)} \tag{12.748}$$

♣Clearly not mathematically rigorous. Comment on how to make it ok by Gaussian regularization of the volumes. ♣

A more conceptual version of this equation is:

$$\tilde{g}_{\text{string}}^{-2}\sqrt{\tilde{g}_{00}} = g_{\text{string}}^{-2}\sqrt{g_{00}} \tag{12.749}$$

(See remark **** below for an explanation of why it is more conceptual.) Thus, the dilaton field shifts:

$$\tilde{\Phi} = \Phi - \frac{\ell_s^2}{2}\log g_{00} \tag{12.750}$$

Remarks:

1. Applying this to the special case where $\mathcal{X} = T^d$ and g_{ij}, b_{ij} are constant, we have d independent $U(1)$ isometries, and we have given a path integral derivation of the dualities $\mathfrak{d}_i$ mentioned above, as promised.

2. Buscher's original argument was based on a symmetry of the beta functions for the sigma model. To leading order, the beta functions are [7]:

$$\begin{aligned} \beta^{\Phi} &= \frac{d-26}{48\pi^2} + \frac{\ell_s^2}{16\pi^2}\left(4(\nabla\Phi)^2 - 4\nabla^2\Phi - \mathcal{R} + \frac{1}{12}H^2\right) + \mathcal{O}(\ell_s^4/L^4) \\ \beta^{G} &= \mathcal{R}_{\mu\nu} - \frac{1}{4}H_{\mu}{}^{\lambda\rho}H_{\nu\lambda\rho} + 2\nabla_\mu\nabla_\nu(\Phi/\ell_s^2) + \mathcal{O}(\ell_s^2/L^2) \\ \beta^{H} &= \nabla^{\lambda}H_{\lambda\mu\nu} - 2(\nabla_\lambda(\Phi/\ell_s^2))H^{\lambda}{}_{\mu\nu} + \mathcal{O}(\ell_s^2/L^2) \end{aligned} \tag{12.751}$$

These are the first terms in an expansion in ℓ_s/L where L is a typical length scale of the target space. When the sigma model also has supersymmetry many of the higher corrections vanish. For (2,2) worldsheet supersymmetry this is how one derives the Calabi-Yau condition (at $H = 0$, and subject to some important subtleties). Buscher observed that for targets with $U(1)$ isometries his eponymous transformation rules are a symmetry of the fixed point equations.

3. We have interpreted $\varphi = \exp[\mathrm{i}x^0/\ell_s]$ to be a $U(1)$-valued function on Σ. However, in the presence of vertex operators this geometric interpretation can change. In order to illustrate it suffices to take $\mathcal{X} = \mathbb{R}/2\pi\mathbb{Z}$ with $x \sim x + 2\pi R$. If we insert a vertex operator

$$\exp[\mathrm{i}px(\mathcal{P})] \tag{12.752}$$

at a point $\mathcal{P}$ on Σ then the Euclidean signature path integral becomes

$$\int [dx] e^{-\frac{\mathrm{i}}{2\pi\ell_s^2}\int_\Sigma \partial x\bar{\partial}x + \mathrm{i}px(\mathcal{P})} \tag{12.753}$$

In this free field theory we can shift by the stationary point (solution of the equations of motion). The stationary point is given by

$$\partial\bar{\partial}x = -\pi\ell_s^2 p\delta(\mathcal{P}) \tag{12.754}$$

where $\delta(\mathcal{P})$ is a $(1,1)$ Dirac delta form with support at $\mathcal{P}\in\Sigma$. So we can shift

$$x(\mathcal{Q}) = -\pi\ell_s^2 pG(\mathcal{Q},\mathcal{P}) + x_{quantum} \tag{12.755}$$

where $G(\mathcal{Q},\mathcal{P})$ is Green's function for $\partial\bar{\partial}$.

In the special case $\Sigma = \mathbb{C}$ with Euclidean metric the following formulae are helpful:

$$\begin{aligned}(\partial_x^2+\partial_y^2)\log r^2 &= 4\pi\delta^{(2)}(0)\\ \partial_z\partial_{\bar{z}}\log|z|^2 &= \pi\delta^{(2)}(0)\\ \partial_z\frac{1}{\bar{z}} = \partial_{\bar{z}}\frac{1}{z} &= \pi\delta^{(2)}(0)\end{aligned} \tag{12.756}$$

Now $F = dx = (\partial+\bar{\partial})x$ so $*F = \mathrm{i}(-\partial+\bar{\partial})x$ so the classical equations of motion become

$$\begin{aligned}dF &= 0\\ d*F &= 2\mathrm{i}\partial\bar{\partial}x = -2\pi\ell_s^2 p\delta(\mathcal{P}) = j_e\end{aligned} \tag{12.757}$$

where j_e is the electric current, supported at $\mathcal{P}$.

One the other hand, in the dual picture the insertion of the vertex operator (12.752) is handled quite differently. Now the classical equations of motion for the dual field are:

$$\begin{aligned}d*F_D &= 0\\ dF_D &= j_e\end{aligned} \tag{12.758}$$

In this interpretation, if $j_e\neq 0$ then we cannot have $F_D = dx_D$ for a smooth field x_D. Nevertheless, there are two (related) geometrical interpretations of x_D:

For the first interpretation let us recall that in general if (P,∇) is a principal $U(1)$ bundle with connection over any manifold M and if $s:\mathcal{U}\to P$ is any local section defined over $\mathcal{U}\subset M$ then ∇s is a 1-form valued in P. Therefore $s^{-1}\nabla s$ is a locally defined 1-form. It is only defined in the region $\mathcal{U}$ where the section s is defined. This is just the one-form of the connection relative to the trivialization of P defined by s. Then $d(s^{-1}\nabla s)$ is the curvature of the connection. As opposed to the connection one-form, the curvature one-form can be globally defined: If we choose another patch $\mathcal{U}'$ and section s' we will produce the same curvature form on the overlap $\mathcal{U}\cap\mathcal{U}'$. Therefore in the presence of j_e we should view $\exp[\mathrm{i}x_D]$ as a locally trivializing section of a principal $U(1)$ bundle with curvature. Thus, as we have seen several times in this course *the geometrical nature of the field has changed.*

Now, since F_D must be globally well-defined the Gauss law implies that

$$\int_\Sigma j_e = 0 \tag{12.759}$$

in order for there to be nonzero correlators. Therefore, if we interpret x_D as local trivializing sections of a $U(1)$ bundle $P \to \Sigma$ that bundle must be trivializable. However, there is no natural trivialization.

Back in the original frame it is natural to allow for both magnetic and electric currents:

$$\begin{aligned} dF &= j_m \\ d * F &= j_e \end{aligned} \tag{12.760}$$

Indeed, we have already done so! These correspond to the states with general momentum and winding number. To convert states into vertex operators we should use the exponential map to convert a cylinder to a punctured complex plane:

$$z = e^{\mathrm{i}(\sigma + \mathrm{i}\tau)} \tag{12.761}$$

Recall that on the cylinder we had the on-shell expansion (12.573), and for a Euclidean signature worldsheet the Wick rotated version is:

$$x^i = x_0^i + \frac{1}{2}\ell_s^2 p_L^i(\mathrm{i}\tau + \sigma) + \frac{1}{2}\ell_s^2 p_R^i(\mathrm{i}\tau - \sigma) + x^i_{\mathrm{osc}} \qquad i = 1, \dots, d \tag{12.762}$$

so that

$$x = x_0 - \frac{\mathrm{i}\ell_s^2}{2}\left(p_L \log z + p_R \log \bar{z}\right) + x_{osc} \tag{12.763}$$

so that

$$F = dx = -\frac{1}{2}\mathrm{i}\ell_s^2\left(p_L \frac{dz}{z} + p_R\frac{d\bar{z}}{\bar{z}}\right) + dx_{osc} \tag{12.764}$$

Now recall that

$$d\left(\frac{d\bar{z}}{\bar{z}}\right) = \partial_z\left(\frac{1}{\bar{z}}\right) dz d\bar{z} = \pi\delta^{(2)}(0) dz d\bar{z} \tag{12.765}$$

$$d\left(\frac{dz}{z}\right) = -\partial_{\bar{z}}\left(\frac{1}{z}\right) dz d\bar{z} = -\pi\delta^{(2)}(0) dz d\bar{z} \tag{12.766}$$

Therefore

$$\begin{aligned} dF &= \frac{\mathrm{i}\pi\ell_s^2}{2}\left(p_L - p_R\right)\delta^{(2)}(0) dz d\bar{z} \\ d * F &= \frac{\pi\ell_s^2}{2}\left(p_L + p_R\right)\delta^{(2)}(0) dz d\bar{z} \end{aligned} \tag{12.767}$$

For our second interpretation we see from (12.763) that when $p_L - p_R \neq 0$ the field x is not single-valued in the neighborhood of the vertex operator. This is just what we expect from a vertex operator for a winding mode: It is an example of a "disorder operator" Nevertheless, the physically relevant quantity is the "fieldstrength" (in the sense of differential cohomology) $F = dx$. This is single-valued, but it is singular. In general, disorder operators, or "defect" such as "'t Hooft operators," "monopole operators," and so on can be incorporated into a path integral by specifying a singularity of the fields in the path integral.

4. The exchange of g_{0s} with b_{0s} in the Buscher rules has dramatic consequences. Suppose the target space $\mathcal{X}$ is a nontrivial circle bundle $\pi : \mathcal{X} \to \bar{\mathcal{X}}$, so the fibers of π are copies of $U(1)$ and there is a fixed-point free right action on $\mathcal{X}$ with $\pi(p \cdot e^{i\theta}) = \pi(p)$. We assume moreover that this right $U(1)$ action is an isometry of the metric and B-field on $\mathcal{X}$. Therefore the metric has the form:

$$ds^2 = \bar{g}_{st} dx^s dx^t + \varpi(x^s)\Theta^2 \tag{12.768}$$

where Θ is a connection on the circle bundle and can be locally written as $\Theta = dx^0 + A$. The function $\varpi(x^s)$ is a function known as a *warp factor*. Expanding this out we see that

$$\begin{aligned} g_{00} &= \varpi \\ g_{0s} &= \varpi A_s \\ g_{st} &= \bar{g}_{st} + \varpi(x^s) A_s A_t \end{aligned} \tag{12.769}$$

Although the local one-forms $A_s dx^s$ are not globally defined in general (and cannot be globally defined on topologically nontrivial circle bundles) the field-strength $F = d(A_s dx^s)$ is a globally well-defined 2-form. Now (for simplicity) assume $b_{ij} = 0$ and consider the dual model. According to the Buscher rules the dual picture now has

$$\begin{aligned} \tilde{g}_{0s} &= 0 \\ \tilde{b}_{0s} &= A_s \end{aligned} \tag{12.770}$$

This is remarkable: T-duality has "untwisted" the circle bundle. The dual geometry is now globally $\bar{\mathcal{X}} \times S^1$. Moreover, the B-field is only locally defined: We have a topologically nontrivial gerbe connection, and $\tilde{H} = d\tilde{b} = F \wedge d\tilde{x}^0$.

♣Write out the example of level k SU(2) WZW model using Hopf fibration. ♣

5. COMMENT ON WHAT HAPPENS WHEN $U(1)$ HAS FIXED POINTS.

6. EXPLAIN THAT (12.749) means that the Planck length and Einstein frame metric are T-duality invariant.

7. COMMENT A BIT on the generalization of differential cohomology to ℓ-forms for $\ell > 1$.

12.10 Deformations Of Algebras And Hochschild Cohomology

Above we noted that the Moyal product is a formal deformation of the commutative algebra of functions on $\mathbb{R}^{2n}$.

We can formalize this notion as follows:

Definition: Let A be an associative algebra over a field κ. Then a formal deformation of the algebra is an associative algebra structure on the algebra $A[[t]]$ over the ring formal series $\kappa[[t]]$ such that:

$$\mu(a,b) = \mu_0(a,b) + \mu_1(a,b)t + \mu_2(a,b)t^2 + \cdots \tag{12.771}$$

where μ_0 is the original algebra structure on A.

Note that this definition implies

1. μ is bilinear over $\kappa[[t]]$.
2. μ is associative so

$$\mu(\mu(a,b),c) - \mu(a,\mu(b,c)) = 0 \tag{12.772}$$

Expanding out in t this gives a lot of complicated equations on the μ_n. Of course the zeroth order equation is satisfied since, by assumption, μ_0 is associative. If we write $\mu_0(a,b) = ab$ for simplicity then the first order equation is

$$a\mu_1(b,c) - \mu_1(ab,c) + \mu_1(a,bc) - \mu_1(a,b)c = 0 \tag{12.773}$$

and similarly, the higher order equations are

$$a\mu_n(b,c) - \mu_n(ab,c) + \mu_n(a,bc) - \mu_n(a,b)c = \sum_{j=1}^{n-1} \Big\{ \mu_j(a,\mu_k(b,c)) - \mu_j(\mu_k(a,b),c) \Big\} \tag{12.774}$$

where in the sum $j+k=n$.

For a nontrivial first order deformation we must solve equation (12.773), but we must also be sure that the deformation can't be undone by a simple redefinition. This leads to the

Definition Two formal deformations μ and $\tilde{\mu}$ of an associative algebra A are *equivalent* if there is a $\kappa[[t]]$-linear map

$$F : A[[t]] \to A[[t]] \tag{12.775}$$

such that for $a \in A$,

$$F(a) = a + \sum_{n=1}^{\infty} t^n f_n(a) \tag{12.776}$$

(so the $f_n : A \to A$ are themselves linear) and such that

$$F(\tilde{\mu}(a,b)) = \mu(F(a),F(b)) \tag{12.777}$$

In particular, a first order deformation $\tilde{\mu}_1$ will be equivalent to μ_1 if there exists a linear map $f_1 : A \to A$ such that

$$\tilde{\mu}_1(a,b) = \mu_1(a,b) + (af_1(b) - f_1(ab) + f_1(a)b) \tag{12.778}$$

Thus, we can phrase the problem of finding first-order deformations as a cohomology problem: Define

$$C^n(A,A) := \mathrm{Hom}_\kappa(A^{\otimes n}, A) \tag{12.779}$$

The reason for the two arguments of C^n will be explained below. Note that our deformations are $\mu_n \in C^2(A,A)$. Define a map

$$d : C^2(A,A) \to C^3(A,A) \tag{12.780}$$

by saying that if $m \in C^2(A,A)$ then

$$dm(a,b,c) := am(b,c) - m(ab,c) + m(a,bc) - m(a,b)c \tag{12.781}$$

So our first order deformation must satisfy $d\mu_1 = 0$. Similarly, define a map

$$d : C^1(A,A) \to C^2(A,A) \tag{12.782}$$

by saying that if $f \in C^1(A,A)$ then

$$df(a,b) := af(b) - f(ab) + f(a)b \tag{12.783}$$

One easily checks that $d(df) = 0$:

$$\begin{aligned} d(df)(a,b,c) &= adf(b,c) - df(ab,c) + df(a,bc) - df(a,b)c \\ &= a(bf(c) - f(bc) + f(b)c) \\ &- (abf(c) - f(abc) + f(ab)c) \\ &+ af(bc) - f(abc) + f(a)bc) \\ &- (af(b) - f(ab) + f(a)b)c \\ &= 0 \end{aligned} \tag{12.784}$$

So $\tilde{\mu}_1 \sim \mu_1$ if there is an f_1 with

$$\tilde{\mu}_1 = \mu_1 + df_1 \tag{12.785}$$

and we conclude:

Equivalence classes of first order deformations of an associative algebra are given by cohomology classes in

$$H^2(A,A) := \ker[d : C^2 \to C^3]/\mathrm{im}[d : C^1 \to C^2]. \tag{12.786}$$

Moreover, if we define $\Delta\mu := \sum_{n=1}^\infty t^n \mu_n$ so that $\mu = \mu_0 + \Delta\mu$, then the full deformation equation can be written as a kind of Maurer-Cartan equation:

$$d(\Delta\mu) - \frac{1}{2}[\Delta\mu, \Delta\mu] = 0 \tag{12.787}$$

where, for $\delta_1, \delta_2 \in C^2$ we define $[\delta_1, \delta_2] \in C^3$ by the formula:

$$\begin{aligned} [\delta_1, \delta_1](a,b,c) &:= \delta_1(\delta_2(a,b),c) - \delta_1(a, \delta_2(b,c)) \\ &+ \delta_2(\delta_1(a,b),c) - \delta_2(a, \delta_1(b,c)) \end{aligned} \tag{12.788}$$

There is a larger algebraic structure here:

Definition The *Hochschild chain complex of an associative algebra A with values in the bimodule A* is the graded vector space

$$C^\bullet(A,A) := \oplus_{n=0}^{\infty} C^n(A,A) \tag{12.789}$$

$$C^n(A,A) := \mathrm{Hom}_\kappa(A^{\otimes n}, A) \tag{12.790}$$

(where $C^0(A,A)$ is naturally isomorphic to A via $f(1) = a \in A$) with differential

$$d : C^n(A,A) \to C^{n+1}(A,A) \qquad\qquad n \geq 0 \tag{12.791}$$

$$\begin{aligned} df(a_1,\dots,a_{n+1}) = a_1 \cdot f(a_2,\dots,a_{n+1}) \\ + \sum_{i=1}^{n}(-1)^i f(a_1,\dots,a_{i-1},a_i a_{i+1},\dots,a_{n+1}) \\ + (-1)^{n+1} f(a_1,\dots,a_n)\cdot a_{n+1} \end{aligned} \tag{12.792}$$

The *Hochschild cohomology of A* is the cohomology of this differential. The first few cohomology groups have simple interpretations:

1. A zero-cochain f maps $\kappa \to A$ so that if we write $f(1) = b$ then $df : A \to A$ is the linear map:

   $$df(a) = ab - ba \tag{12.793}$$

   Thus $H^0(A,A) = Z(A)$ is the center of A.

2. If $f \in C^1(A,A)$, then $df = 0$ means f is a derivation:

   $$f(ab) = af(b) + bf(a) \tag{12.794}$$

   Of course, the commutator with b is always a derivation, therefore $H^1(A,A)$ is the quotient of the space of derivations of A by the space of inner derivations.

   Note: A covariant derivative is a derivation, and under gauge transformation it transforms by the addition of an inner derivation.

   ♣Explain more here! ♣

3. As we have seen $H^2(A,A)$ is isomorphism classes of first order deformations of the algebra structure.

There is another very interesting algebraic structure on $C^\bullet(A,A)$: It is a graded Lie algebra.

If $f \in C^n(A,A)$ and $g \in C^n(A,A)$ then define

$$f \circ g \in C^{n+m-1} \tag{12.795}$$

by

$$f \circ g(P) := \sum_{\mathrm{Pa}_3} (-1)^{|P_1|} f(P_1, g(P_2), P_3) \tag{12.796}$$

Here P is an ordered $(n+m-1)$-tuple of elements of A:

$$P = \{a_1, \cdots, a_{n+m-1}\} \tag{12.797}$$

and the sum is over ordered disjoint decompositions $P = P_1 \amalg P_2 \amalg P_3$, meaning that each of P_1, P_2, P_3 is ordered, and the ordering is inherited from P. Each of P_1, P_2, P_3 can be disjoint. Finally $|P_1|$ is the number of elements of P_1. Now, the *Gerstenhaber bracket* is defined as

$$[f, g] := f \circ g - (-1)^{(n-1)(m-1)} g \circ f \tag{12.798}$$

♣Sign here is from Takhtadjan. But it seems to simple. It differs from GMW Appendix A, but they have a mistake. ♣

Theorem: Give $C^\bullet(A, A)$ a grading so that $C^n(A, A)$ has grading $(n-1)$. Then the G-bracket satisfies the graded Jacobi identity:

$$(-1)^{|f_1||f_3|}[f_1, [f_2, f_3]] + (-1)^{|f_2||f_1|}[f_2, [f_3, f_1]] + (-1)^{|f_3||f_2|}[f_3, [f_1, f_2]] = 0. \tag{12.799}$$

where $|f|$ is the degree of f.

♣Need to give a proof ♣

In fact, restoring the notation μ_0 for the multiplication operator on A, and considering it as a 2-cochain, the Hochschild differential is just

$$df = [f, \mu_0] \tag{12.800}$$

♣Check sign ♣

Remarks

1. The above description of Hochschild cohomology is just the beginning of a much larger algebraic story. For one thing, if M is any bimodule for A then we can define Hochschild cohomology with coefficients in a bimodule $HH^\bullet(A; M)$. Now n-cochains are maps $\phi : A^{\otimes n} \to M$ and the formula for the differential (12.792) still makes sense and still squares to zero. There is also a dual theory of Hochschild homology, that plays an important role in noncommutative geometry.
2. INTERPRETATION OF H^1 IN TERMS OF COVARIANT DERIVATIVES
3. SHIFTING THE ZERO
4. It is natural to ask about the Hochschild cohomology of the Weyl algebra. See [20].
5. GENERALIZATION TO LINEAR CATEGORIES.
6. PUT INTO CONTEXT OF L_∞ ALGEBRAS

Exercise
Show that the operator d defined in (12.792) really is a <u>differential</u>, that is, that $d^2 = 0$.

12.10.1 Poisson Manifolds

Definition: A *Poisson algebra* is an associative algebra associated with a bracket: $\{\cdot,\cdot\}: A \otimes A \to A$ that is:

1. A Lie bracket, so $\{a,b\} = -\{b,a\}$ and the Jacobi identity is satisfied.

2. A derivation:
$$\{a,bc\} = \{a,b\}c + b\{a,c\} \tag{12.801}$$

If A is a <u>commutative</u> algebra, then a Hochschild cocycle of degree two defines a *Poisson algebra*. ♣Should do graded case? ♣

DEFINE POISSON MANIFOLDS

TWO EXAMPLES.

STATEMENT OF DEFORMATION QUANTIZATION PROBLEM

KONTSEVICH THEOREM

Observe SW limit of (**??**) is

$$S = -2\pi \mathrm{i} \int_{\Sigma} B_{ij} dx^i \wedge dx^j \tag{12.802}$$

Generalization: Hochschild cohomology of A-modules. (A infty?) categories.

FINISH THIS NEXT TIME. REFERENCES:

1. M. Kontsevich, arXiv:math/9904055
2. A. Cattaneo and G. Felder, arXiv:hep-th/0102208
3. A. Voronov, arXiv:math/0111009
4. L. Baulieu, A. Losev, and N. Nekrasov, arXiv:hep-th/0106042
5. A. Kapustin http://arxiv.org/abs/hep-th/0502212

(Also Nekrasov emails: Nov. 13, 2015; Kapustin email: Nov. 13, 2015; Witten email: Nov. 18, 2015)

12.11 C^*-Algebra Approach To Quantum Mechanics

Now return to our discussion of C^*-algebras. For technicalities on C^* algebras and functional analysis we mostly follow [31, 42].

12.11.1 Positive Elements And Maps For C^* Algebras

Definition: An element $a \in \mathfrak{A}$ in a C^* algebra $\mathfrak{A}$ is *positive* if $a = a^*$ and its spectrum is positive: $\sigma(a) \subset \mathbb{R}_+$. We write $a \geq 0$ and denote the set of positive elements by $\mathfrak{A}^+$.

Examples:

1. $\mathfrak{A} = \mathcal{B}(\mathcal{H})$. Then a is positive iff $(\psi, a\psi) \geq 0$ for all ψ.

2. $\mathfrak{A} = C_0(X)$. Then f is positive iff $f(x) \geq 0$ for all $x \in X$.

Theorem: $a \in \mathfrak{A}$ is positive iff $a = b^2$ for a self-adjoint element $b \in \mathfrak{A}$. Likewise a is positive iff $a = b^* b$ for some element $b \in \mathfrak{A}$.

Proof: The proof is immediate using the continuous functional calculus. If $\sigma(a) \subset \mathbb{R}^+$ then apply $f(x) = \sqrt{x}$ to a to get $b = f(a)$. ♠

Every self-adjoint element $a \in \mathfrak{A}$ can be written as $a = a_+ - a_-$. This follows again from the continuous functional calculus. Note that:

$$f(x) := \begin{cases} x & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{12.803}$$

is a positive continuous function. So we define $a_+ = f(a)$ and $a_- = f(-a)$. With this choice of $a_\pm$ we also see that in the decomposition we can take $a_\pm$ such that $\parallel a_\pm \parallel \leq \parallel a \parallel$. Therefore every element can be written as

$$a = a_+ - a_- + \mathrm{i}(b_+ - b_-) \tag{12.804}$$

where $a_\pm, b_\pm$ are all positive, and $\parallel a_\pm \parallel \leq \parallel a \parallel$ and $\parallel b_\pm \parallel \leq \parallel a \parallel$.

We can also speak of positive linear maps between C^* algebras:

Definition: A linear map $\varphi : \mathfrak{A} \to \mathfrak{B}$ between two C^* algebras is *positive* if $\varphi(a) \geq 0$ whenever $a \geq 0$.

It is not difficult to show tht positive maps are bounded, and therefore continuous. (Landsman 2.8.5). We stress that φ need only be a linear map and in applications it is usually <u>not</u> a morphism of C^* algebras. In fact, if φ is a morphism of C^* algebras then it is automatically positive since

$$\varphi(a^* a) = b^* b \tag{12.805}$$

where $b = \varphi(a)$.

There is a generalization of the notion of a positive map known as a *completely positive map*. We first observe that if $\mathfrak{A}$ is any C^* algebra then $M_n(\mathfrak{A})$, the $*$-algebra of $n \times n$ matrices over $\mathfrak{A}$ (where $*$ includes hermitian conjugation of the matrix). If $\mathcal{H}$ is a faithful representation of $\mathfrak{A}$ then $\mathcal{H} \otimes \mathbb{C}^n$ is a faithful representation of $M_n(\mathfrak{A})$ and we define the C^*-norm on $M_n(\mathfrak{A})$ by the operator norm on $\mathcal{H} \otimes \mathbb{C}^n$. This makes $M_n(\mathfrak{A})$ a C^*-algebra. (Recall the C^* norm is unique.)

Now, if $\varphi : \mathfrak{A} \to \mathfrak{B}$ is a linear map we say it is *completely positive* if for all $n \geq 1$ induced map $\varphi : M_n(\mathfrak{A}) \to M_n(\mathfrak{B})$ (defined by applying φ to the matrix elements) is positive.

♣Need to give an example of a map that is positive but not completely positive. ♣

12.11.2 States On A C^*-Algebra

Definition A *state* on a C^* algebra is a positive norm 1 linear map

$$\omega : \mathfrak{A} \to \mathbb{C}. \tag{12.806}$$

In the case that $\mathfrak{A}$ is a unital C^* algebra we could alternatively define a state ω as a positive map so that $\omega(\mathbf{1}) = 1$.

Note that since $\omega(a) \geq 0$ for $a \in \mathfrak{A}^+$, and since every self-adjoint element can be written as $a = a_+ - a_-$ with $a_\pm$ positive, it follows that $\omega(a) \in \mathbb{R}$ when a is self-adjoint. Therefore

$$\omega(a^*) = \omega(a)^*. \tag{12.807}$$

Example 1: A key example is obtained by taking a C^*-subalgebra $\mathfrak{A} \subset \mathcal{B}(\mathcal{H})$ and a rank one projection operator $P \in \mathfrak{A}$. Then

$$\omega_P(a) = \mathrm{Tr}_{\mathcal{H}}(Pa) = \frac{\langle\psi|a|\psi\rangle}{\langle\psi|\psi\rangle} \tag{12.808}$$

If ψ is any nonzero vector in $\mathcal{H}$ then there is a corresponding rank one projector to the line ℓ_ψ spanned by ψ:

$$P = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \tag{12.809}$$

Therefore for ψ e a nonzero vector in $\mathcal{H}$ we can define

$$\omega_\psi(a) := \frac{\langle\psi|a|\psi\rangle}{\langle\psi|\psi\rangle} \tag{12.810}$$

So these states are consequently sometimes called *vector states*. The corresponding state only depends on the line in $\mathcal{H}$ through ψ, that is

$$\omega_\psi = \omega_{z\psi} \tag{12.811}$$

for every $z \in \mathbb{C}^*$. In physics one hears the statement that the physical state is a "ray" in Hilbert space.

Example 2: More generally, if ρ is a positive traceclass operator of trace one, i.e. a density matrix, then

$$\omega(a) = \mathrm{Tr}_{\mathcal{H}}\rho a \tag{12.812}$$

is a state. For the ideal of compact operators $\mathcal{K}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ one can show that these are the only states. (Landsman 2.13.10.1). Warning! This is far from true for the C^*-algebra of all bounded operators $\mathcal{B}(\mathcal{H})$. See the very useful remarks at the beginning of section 2.13 of [31].

Example 3: Another key example arises for commutative C^* algebras $\mathfrak{A} = C_0(X)$ where X is a locally compact Hausdorff topological space. By the Riesz-Markov theorem there is a (complex regular Borel) measure μ on X so that any linear functional $\ell : C_0(X) \to \mathbb{C}$ is given by

$$\ell(f) = \int_X f d\mu \tag{12.813}$$

Moreover, $\parallel \ell \parallel = |\mu(X)|$. Therefore, the states on $C_0(X)$ are positive measures of total measure $\mu(X) = 1$.

Theorem: The space of states $\mathcal{S}(\mathfrak{A})$ is a compact convex set.

Proof: Use the $w*$ topology and apply the Banach-Alaoglu theorem.

Because it is convex we can define the *extremal points.* These are the states $\omega \in \mathcal{S}(\mathfrak{A})$ which cannot be written in the form

$$\omega = t\omega_1 + (1-t)\omega_2 \tag{12.814}$$

where $0 < t < 1$ and ω_1, ω_2 are states. An extremal state is known as a *pure state.* A state that is not a pure state is a *mixed state.*

Example 1: Vector states are always pure states.

Example 2: For the compact operators $\mathcal{K}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ the pure states are precisely the set of vector states (Landsman 2.13.10.1). However, one can show that for any C^* algebra $\mathfrak{A}$, given any self-adjoint element $a \in \mathfrak{A}$, and $E \in \sigma(a)$, there is a pure state ω_E on $\mathfrak{A}$ such that $\omega_E(a) = E$. If we choose a to have continuous spectrum and E to be in the continuous spectrum, then there is no normalizable eigenvector with eigenvalue E, so ω_E is not a vector state.

Example 3: If $\mathfrak{A} = C_0(X)$ where X is a locally compact Hausdorff topological space then the pure states of $\tilde{\mathfrak{A}}$ (where we must unitize if X is noncompact) are the Dirac measures supported at a point. δ_x. Therefore, the space of pure states is in fact homeomorphic to X^+, the one-point compactification.

♣More remarks on the case where X is noncompact. ♣

12.11.3 GNS Construction

Given a state ω on a C^*-algebra $\mathfrak{A}$ we construct a Hilbert space $\mathcal{H}_\omega$ and a C^*-morphism $\rho : \mathfrak{A} \to \mathcal{B}(\mathcal{H}_\omega)$.

A representation of $\mathfrak{A}$ on a Hilbert space $\mathcal{H}$ is *non-degenerate* if the only vector annihilated by $\pi(a)$ for all $a \in \mathfrak{A}$ is the zero vector. A representation is called *cyclic* if there is a vector Ψ_0 so that the closed subspace $\pi(\mathfrak{A})\Psi_0$ of $\mathcal{H}$ coincides with $\mathcal{H}$.

Now suppose that $\mathfrak{A}$ is unital, and $\omega \in \mathcal{S}(\mathfrak{A})$ is a state on $\mathfrak{A}$. We now construct a representation $\pi : \mathfrak{A} \to \mathcal{H}_\omega$ as follows:

We begin by using ω to define a sesquilinear form on $\mathfrak{A}$:

$$(a_1, a_2) := \omega(a_1^* a_2) \tag{12.815}$$

Note that the form is positive <u>semi</u>-definite:

$$(a, a) = \omega(a^* a) \geq 0 \tag{12.816}$$

but it might well have a nontrivial *radical*:

$$\begin{aligned} \mathfrak{N}_\omega &:= \{a \in \mathfrak{A} | (a, b) = 0 \qquad \forall b \in \mathfrak{A}\} \\ &= \{a \in \mathfrak{A} | (b, a) = 0 \qquad \forall b \in \mathfrak{A}\} \end{aligned} \tag{12.817}$$

The equality of the two sets follows because $\omega(a^*) = \omega(a)^* = 0$ for every $a \in \mathfrak{A}$.

Note that $\mathfrak{N}_\omega$ is clearly a left-ideal, for if $a \in \mathfrak{N}_\omega$ and $c \in \mathfrak{A}$ then $ca \in \mathfrak{N}_\omega$ because

$$(ca, b) = \omega((ca)^* b) = \omega(a^*(c^* b)) = 0 \tag{12.818}$$

There is another characterization of $\mathcal{N}_\omega$ which can be useful. Note that, putting $b = a$ in the first line of (12.817) we see that if $a \in \mathfrak{N}_\omega$ then $\omega(a^* a) = 0$. In fact, this is a sufficient condition, that is:

$$\mathfrak{N}_\omega = \{a \in \mathfrak{A} | \omega(a^* a) = 0\} \tag{12.819}$$

To prove this we first note that, quite generally, for any state ω on a C^* algebra we have the Cauchy-Schwarz inequality: [49]

$$|\omega(a^* b)|^2 \leq \omega(a^* a)\omega(b^* b) \tag{12.820}$$

Now, if $\omega(a^* a) = 0$ then for any b we have $|\omega(a^* b)|^2 \leq \omega(a^* a)\omega(b^* b) = 0$. This proves (12.819).

As we see from the examples of vector states on $\mathcal{B}(\mathcal{H})$, the radical might well be be nonzero! For ω_ψ the radical consists of all operators containing ψ in the kernel. If we choose ψ to be the first basis vector in an orthogonal basis then the matrix representation of $a \in \mathcal{N}_{\omega_\psi}$ has first column equal to zero. (Note this is obviously a left ideal.) Given this characterization it is clear that $\mathcal{B}(\mathcal{H})/\mathcal{N}_{\omega_\psi} \cong \mathcal{H}$.

The GNS representation is now defined by constructing a positive <u>definite</u> inner product on $\mathfrak{A}/\mathcal{N}_\omega$:

$$([a], [b]) := \omega(a^* b) \tag{12.821}$$

The expression (12.821) is well-defined because $\mathcal{N}_\omega$ is an ideal: If $n_1, n_2 \in \mathcal{N}_\omega$ then

$$\begin{aligned} \omega((a + n_1)^*(b + n_2)) &= \omega(a^* b) + \omega(n_1^* b) + \omega(a^* n_2) + \omega(n_1^* n_2) \\ &= \omega(a^* b) \end{aligned} \tag{12.822}$$

Now the nondegeneracy is immediate: If $(a, b) = 0$ for all b then $a \in \mathcal{N}_\omega$ by definition so $[a] = 0$. Actually, $\mathfrak{A}/\mathcal{N}_\omega$ with this inner product needs to be completed, and that defines the Hilbert space:

$$\mathcal{H}_\omega := \overline{\mathfrak{A}/\mathfrak{N}_\omega} \tag{12.823}$$

Now we define $\pi : \mathfrak{A} \to \mathcal{B}(\mathcal{H}_\omega)$ by

$$\pi_\omega(a)[b] := [ab] \tag{12.824}$$

Next one needs to show that this is a C^*-algebra morphism. Note that:

$$\begin{aligned} ([c], \pi_\omega(a)^*[b]) &:= (\pi_\omega(a)[c], [b]) \\ &= ([ac], [b]) \\ &= \omega((ac)^* b) \\ &= \omega(c^* a^* b) \\ &= ([c], [a^* b]) \\ &= ([c], \pi_\omega(a^*)[b]) \end{aligned} \tag{12.825}$$

[49] Proof: $\omega((a - zb)^*(a - zb)) \geq 0$ as a function of $z \in \mathbb{C}$. Thus we have a positive semidefinite quadratic form in the real and imaginary parts of z, and the Cauchy-Schwarz inequality is the corresponding condition on the discriminant.

Note that the Hilbert space $\mathcal{H}_\omega$ has a canonical vector

$$\Psi_\omega := [\mathbf{1}] \tag{12.826}$$

provided by the projection of the unit of $\mathfrak{A}$. By definition

$$\pi_\omega(\mathfrak{A})\Psi_\omega = \mathfrak{A}/\mathcal{N}_\omega \tag{12.827}$$

so, by definition, Ψ_ω is a cyclic vector since $\mathcal{H}_\omega$ is the completion of $\mathfrak{A}/\mathcal{N}_\omega$.

Example 1 If $\mathfrak{A} = C(X)$ then, as we have seen, a state is a positive measure μ on X. The corresponding Hilbert space is $\mathcal{H}_\mu = L^2(X, d\mu)$.

There is also a kind of converse to the GNS construction: If $\mathfrak{A}$ is represented on a Hilbert space $\mathcal{H}$ and Ψ_0 is a cyclic vector then we can construct the vector state associated with Ψ_0 and the corresponding GNS representation of $\mathfrak{A}$ is unitarily isomorphic to the original one. Moreover, $\mathcal{H}_\omega$ is irreducible iff ω is a pure state.

Now we are ready to state and prove the famous Gelfand-Neumark theorem:

Theorem If $\mathfrak{A}$ is a C^* algebra then there is a Hilbert space $\mathcal{H}_U$ and an injective morphism of C^* algebras $\pi_U : \mathfrak{A} \to \mathcal{H}_U$.

Proof: Recall that the space of states $\mathcal{S}(\mathfrak{A})$ is a compact space, and for each state ω we can construct a cyclic representation $\mathcal{H}_\omega$. We simply take

$$\mathcal{H}_U := \oplus_{\omega\in\mathcal{S}(\mathfrak{A})}\mathcal{H}_\omega \tag{12.828}$$

The proof is now trivial. If $\pi_U(a) = 0$ then $\omega(a^*a) = 0$ for all states ω and this implies $\parallel a^*a \parallel = 0$ so $\parallel a \parallel = 0$ by the C^*-identity and hence $a = 0$ by the definition of a norm. ♠

Actually, this definition of $\mathcal{H}_U$ is overdoing things a bit. It suffices to take a direct sum just over the pure states. That is:

$$\oplus_{\omega\in\mathcal{P}(\mathfrak{A})}\mathcal{H}_\omega \tag{12.829}$$

will also provide a faithful representation. In fact, we can do better: We can define to pure states to be equivalent iff their GNS representations are unitarily equivalent. Then

♣Should interpret as space of states of a Hilbert bundle. ♣

$$\mathfrak{A} \cong \oplus_{[\omega]}\pi_\omega(\mathfrak{A}) \tag{12.830}$$

where we sum over equivalence classes of pure states. In particular, the irreducible representations of a finite-dimensional C^* algebra are finite-dimensional and hence any finite-dimensional C^* algebra is a direct sum of a finite number of matrix algebras.

12.11.4 Operator Topologies

When we speak of operators on a Hilbert space $\mathcal{H}$ there are other topologies of operator algebras that can be defined. There are three topologies which one typically encounters, known as the strong, weak, and norm topologies. They are included in each other according to

$$\mathfrak{T}^{\text{weak}} \subset \mathfrak{T}^{\text{strong}} \subset \mathfrak{T}^{\text{norm}} \tag{12.831}$$

1. $\mathfrak{T}^{\rm weak}$: $\lim^{\rm weak}_{n\to\infty} a_n = a$ iff for all $\psi_1,\psi_2\in\mathcal{H}$:

$$\lim_{n\to\infty}(\psi_1,(a-a_n)\psi_2)=0 \tag{12.832}$$

2. $\mathfrak{T}^{\rm strong}$: $\lim^{\rm strong}_{n\to\infty} a_n = a$ iff for all $\psi\in\mathcal{H}$:

$$\lim_{n\to\infty} \parallel (a-a_n)\psi \parallel = 0 \tag{12.833}$$

3. $\mathfrak{T}^{\rm norm}$: $\lim^{\rm norm}_{n\to\infty} a_n = a$ iff
$$\lim_{n\to\infty} \parallel (a-a_n) \parallel = 0 \tag{12.834}$$

Note that equation (12.834) implies (12.833) by the definition of the operator norm, while (12.833) implies (12.832) by the Cauchy-Schwarz inequality. Therefore if a set $C\subset \mathcal{B}(\mathcal{H})$ is closed in the norm topology it is closed in the strong topology, and if it is closed in the strong topology it is also closed in the weak topology. Of course, a closed set is the complement in $\mathcal{B}(\mathcal{H})$ of an open set, and hence if $\mathcal{U}\subset\mathcal{B}(\mathcal{H})$ is open in the weak topology then it is open in the strong topology, and if it is open in the strong topology then it is also open in the strong topology. This establishes the inclusions (12.831).

GIVE BASIS OF OPEN SETS.

The inclusions (12.831) are proper inclusions. A standard set of examples (Reed and Simon p.184) is the following: Assume $\mathcal{H}$ is separable and choose an orthonormal basis $\{e_n\}_{n=1}^{\infty}$ (thus choosing an isomorphism $\mathcal{H}\cong\ell^2$).

1. Consider $N_n(\xi)=\frac{1}{n}\xi$. Clearly $N_n\to 0$ in the norm topology.

2. Now let S_n the the projector onto the orthogonal complement to the space spanned by the first n vectors $\{e_j\}_{j=1}^{n}$, that is:

   $$S_n(\xi)=(0,\dots 0,\xi_{n+1},\xi_{n+2},\dots) \tag{12.835}$$

   Then $\parallel S_n\xi \parallel^2=\sum_{n+1}^{\infty}|\xi_j|^2\to 0$ for every vector ξ, so $S_n\to 0$ in the strong topology, but it is also true that $\parallel S_n(e_j)\parallel = 1$ for $j>n$ so S_n does <u>not</u> converge to zero in the norm topology.

3. Now let W_n be the n^{th} iteration of the Hilbert hotel map:

   $$W_n(\xi)=(0,\dots,0,\xi_1,\xi_2,\dots) \tag{12.836}$$

   That is, $W_n: e_i\to e_{i+n}$. Then for any two vectors $\xi,\tilde\xi$ we have $(\xi,W_n\tilde\xi)=(S_n\xi,W_n\tilde\xi)$ so

   $$|(\xi,W_n\tilde\xi)|=|(S_n\xi,W_n\tilde\xi)|\le\parallel S_n(\xi)\parallel\parallel W_n\tilde\xi\parallel=\parallel S_n(\xi)\parallel\parallel\tilde\xi\parallel\to 0 \tag{12.837}$$

   So, $W_n\to 0$ in the weak operator topology. However, $\parallel W_n(\xi)\parallel=\parallel\xi\parallel$ so W_n does not go to zero in the strong operator topology.

12.11.5 Von Neumann Algebras And Measure Spaces

A *measurable space* is a set X together with a collection $\mathfrak{M}$ of subsets of X such that $X \in \mathfrak{M}$, and $\mathfrak{M}$ is closed under complements and countable unions. (It follows that $\emptyset \in \mathfrak{M}$ and $\mathfrak{M}$ is closed under countable intersections.) Such a collection of subsets of X has several names in the literature, among them *σ-algebra.* Elements of $\mathfrak{M}$ are called *measurable sets.*

A *morphism of measure spaces* $f : (X, \mathfrak{M}_X) \to (Y, \mathfrak{M}_Y)$ is a function $f : X \to Y$ so that if $S \in \mathfrak{M}_Y$ then $f^{-1}(S) \in \mathfrak{M}_X$. An isomorphism of measures spaces is a bijection $f : X \to Y$ so that f and f^{-1} are morphisms.

Remark It is interesting to compare a topological space with a measure space. Both are defined by collections of sets on a set X. If X is simultaneously endowed with a topology and a measure, that is, a topological space with a measure then measurable functions can be highly discontinous. For this reason there is no such thing as a "dimension" of a measure space. For example, $\mathbb{R}^n$ with the Euclidean measure are all equivalent as measure spaces!

♣Say more, and prove this. Discuss the "standard" measure space. ♣

By definition, a *von Neumann algebra* is a $*$-subalgebra of $\mathcal{B}(\mathcal{H})$ that is closed in the weak topology.

To give a good example of a von Neumann algebra, let $(X, \mathfrak{M}, \mu)$ be a measure space. We can then form a Hilbert space $\mathcal{H} = L^2(X, \mu)$. The space of bounded measurable functions on X, $L^\infty(X, \mu)$ is an algebra and acts on $\mathcal{H}$ as multiplication operators:

$$(M_f \psi)(x) = f(x)\psi(x) \tag{12.838}$$

It can be shown that this is an abelian von Neumann algebra.

In fact, there is a nice analog of Gelfand's theorem for von Neumann algebras:

Theorem: Commutative von Neumann algebras are in 1-1 correspondence with measure spaces. That is, if $\mathfrak{V}$ is a commutative von Neumann algebra then there is a measure space (X, μ) and an isomorphism of $\mathfrak{V} \cong L^\infty(X, \mu)$.

PROOF OR REFERENCE??

Remark: Since the weak topology is weaker than the norm topology a von Neumann algebra is automatically a C^* algebra. One might therefore wonder what Gelfand's theorem implies about commutative von Neumann algebras.

[Explain Graeme Segal's emails: Oct 5,6, 2015]

An absolutely central result is the following (Landsman 2.14.13):

♣Should at least state the double commutant theorem ♣

Theorem If $\mathcal{H}$ is a Hilbert space and $\mathfrak{M}$ is a $*$-subalgebra of $\mathcal{B}(\mathcal{H})$ then $\mathcal{M}$ is weakly-closed iff it is strongly-closed. Moreover, it is weakly-closed iff the double commutant is the original algebra:

$$\mathfrak{M}'' = \mathfrak{M} \tag{12.839}$$

Proof: First of all $\mathfrak{M} \subset \mathfrak{M}''$ trivially, since elements of $\mathfrak{M}$ define linear functionals on $\mathfrak{M}'$. Now, by definition of weak-closure it follows that for any subalgebra $\mathfrak{N} \subset \mathcal{B}(\mathcal{H})$ the linear dual $\mathfrak{N}'$ is always weakly-closed: For suppose that $\{a_n\}$ is a sequence (or more generally a net) of operators in $\mathfrak{N}'$ that is a Cauchy sequence in $\mathcal{B}(\mathcal{H})$ in the weak topology. Then for all $b \in \mathfrak{N}$,

$$\begin{aligned} 0 &= \lim_{n\to\infty} (\psi_1, [a_n, b]\psi_2) \qquad\qquad a_n \in \mathfrak{N}' \\ &= \lim_{n\to\infty} ((\psi_1, a_n b\psi_2) - (\psi_1, ba_n\psi_2)) \\ &= \lim_{n\to\infty} ((\psi_1, a_n b\psi_2) - (b^*\psi_1, a_n\psi_2)) \\ &= ((\psi_1, ab\psi_2) - (b^*\psi_1, a\psi_2)) \qquad\qquad \text{def. of weak closure} \\ &= (\psi_1, [a, b]\psi_2) \end{aligned} \tag{12.840}$$

Therefore, $\mathfrak{M}'' = (\mathfrak{M}')'$ is automatically weakly closed, so if $\mathfrak{M} = \mathfrak{M}''$ then $\mathfrak{M}$ is weakly closed.

Since the weak topology is weaker than the strong topology, if $\mathfrak{M}$ is weakly closed then its complement is weakly open. But $\mathfrak{T}^{\text{weak}} \subset \mathfrak{T}^{\text{strong}}$, so the complement of $\mathfrak{M}$ is strongly open. Therefore $\mathfrak{M}$ is strongly closed.

Therefore we can close the loop of implications if we show that $\mathfrak{M}$ strongly closed implies $\mathfrak{M} = \mathfrak{M}''$. This is more nontrivial....

.... FINISH ♠

12.11.6 The Spectral Theorem

♣Need to give more proofs and examples in this section. ♣

Recall continuous functional calculus: If a is a self-adjoint element of $\mathfrak{A}$ then it generates an abelian C^* subalgebra $C^*(\mathbf{1}, a) \subset \mathfrak{A}$ and the spectrum of a, namely $\sigma(a)$, is the same considered as an element of either C^* algebra. Moreover, the topological space $\text{Spec}(C^*(\mathbf{1}, a))$ defined by Gelfand's theorem is isomorphic to the compact set $\sigma(a) \subset \mathbb{R}$. Therefore

$$C(\sigma(a)) \cong C^*(\mathbf{1}, a). \tag{12.841}$$

Moreover, under this isomorphism the Gelfand transform of a is just the inclusion of $\sigma(a) \hookrightarrow \mathbb{R}$. Applying the Stone-Weierstrass theorem to $\sigma(a)$ we learn that every continuous function on $\sigma(a)$ is uniformly approximated by polynomial functions. It follows that if $f : \sigma(a) \to \mathbb{C}$ is a <u>continuous</u> function then $f(a) \in \mathfrak{A}$ makes sense and $\sigma(f(a)) = f(\sigma(a))$. (The last statement is the "spectral mapping theorem")

The above "continuous functional calculus" can be extended to the "Borel functional calculus" as follows: Suppose now that we are given a self-adjoint element $a \subset \mathcal{B}(\mathcal{H})$ and a vector $\psi \in \mathcal{H}$. Then the map

$$f \to (\psi, f(a)\psi) \tag{12.842}$$

is a positive linear functional on the algebra of <u>continuous</u> functions $C(\sigma(a))$. By the Riesz-Markov theorem it follows that there is a positive measure $\mu_{a,\psi}$ such that

$$(\psi, f(a)\psi) = \int_{\sigma(a)} f(x) d\mu_{a,\psi}(x) \tag{12.843}$$

Now, if g is a <u>measurable</u> function on $\mathbb{R}$ then, for every $\psi \in \mathcal{H}$ we define:

$$(\psi, g(a)\psi) := \int_{\sigma(a)} g(x) d\mu_{a,\psi}(x) \tag{12.844}$$

Now, by the polarization identity we can recover $(\psi_1, g(a)\psi_2)$ for any two vectors ψ_1, ψ_2, and hence we have defined the operator $g(a)$. If g is bounded then $g(a) \in \mathcal{B}(\mathcal{H})$, since one can show

$$\parallel g(a) \parallel_{\mathcal{B}(\mathcal{H})} \leq \parallel g \parallel_\infty \tag{12.845}$$

Moreover, one can show that if $f_n \to f$ pointwise, and $\parallel f_n \parallel_\infty$ is bounded, then $f_n(a) \to f(a)$ in the strong topology. Thus we have a $*$-homomorphism

$$\Phi : \mathcal{B}(\mathcal{R}) \to \mathcal{B}(\mathcal{H}) \tag{12.846}$$

where $\mathcal{B}(\mathcal{R})$ is the $*$ algebra of bounded Borel measurable functions on $\mathbb{R}$, given by $\Phi(f) = f(a)$. The image of $\mathcal{B}(\mathcal{R})$ is the smallest C^* algebra containing a that is strongly closed. (It is therefore larger than $C^*(\mathbf{1}, a)$, which is norm closed.)

The main point of the extension to the Borel functional calculus is that we can now consider the characteristic function associated to any measurable set $E \subset \mathbb{R}$:

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases} \tag{12.847}$$

and it makes sense to speak of $\chi_E(a)$ for a bounded self-adjoint operator $a \in \mathcal{B}(\mathcal{H})$.

Example 1: For a finite dimensional Hilbert space $\mathcal{H}$ a self-adjoint operator a has a finite set of distinct eigenvalues $\{\lambda_i\}_{i=1}^n$ and there is a finite set of orthogonal projection operators P_i onto the eigenspace of eigenvalue λ_i. These projectors can be written as polynomials in a:

$$P_i = \frac{\prod_{j\neq i}(a - \lambda_j \mathbf{1})}{\prod_{j\neq i}(\lambda_i - \lambda_j)} \tag{12.848}$$

Then for a Borel subset $E \subset \mathbb{R}$ we have

$$\chi_E(a) = \sum_{\lambda_i \in E} P_i \tag{12.849}$$

In fact, this applies to infinite dimensions provided a has a discrete spectrum. If there is an infinite set of eigenvalues in E then the infinite sum converges in the strong topology.

Example 2: Suppose $a \in C(X)$. Then, as we have seen $\sigma(a) = \{a(x) | x \in X\}$. However, $\chi_E(a)$ cannot be in the C^*-algebra $C(X)$ because its Gelfand transform would correspond to the function $\chi_E(x)$ on $\mathbb{R}$, restricted to $\sigma(a)$. But this, in general is not a continuous function on $\sigma(a)$.

Definition: A *projection-valued measure* is a map

$$P : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(\mathcal{H}) \tag{12.850}$$

such that

1. $P(E)$ is an orthogonal projection operator for all $E \in \mathcal{B}(\mathbb{R})$.

2. $P(\emptyset) = 0$ and $P(\mathbb{R}) = 1$.

3. If $E = \amalg_{i=1}^{\infty}$ is a countable *disjoint* union of sets $E_i \in \mathcal{B}(\mathbb{R})$ then

$$P(E) = \mathrm{s} - \lim_{n\to\infty} \sum_{i=1}^{n} P(E_i) \tag{12.851}$$

   where the convergence is in the strong topology.

The Borel functional calculus tells us that to a self-adjoint operator $a \in \mathcal{B}(\mathcal{H})$ we have a corresponding projection-valued measure P_a. Then, given any vector $\psi \in \mathcal{H}$ we have an ordinary measure on $\mathbb{R}$ given by

$$E \mapsto (\psi, P_a(E)\psi) \tag{12.852}$$

Let us call this $\hat{\mu}_{a,\psi}$.

Recall that, via the Riesz-Markov theorem given ψ and a Borel measurable function g we had

$$(\psi, g(a)\psi) = \int_{\sigma(a)} g(x) d\mu_{a,\psi}(x) \tag{12.853}$$

Now, using the projection valued measure P_a we obtain an alternative expression:

$$(\psi, g(a)\psi) = \int_{\mathbb{R}} g(x) d\hat{\mu}_{a,\psi}(x) \tag{12.854}$$

This equation is the content of the spectral theorem: There is a one-one correspondence between projection valued measures in $\mathcal{B}(\mathcal{H})$ and bounded self-adjoint operators on $\mathcal{H}$.

****

EXPLAIN UNITARY EQUIVALENCE TO A SUM OF HILBERT SPACES ON WHICH a IS A MULTIPLICATION OPERATOR

****

12.11.7 States And Operators In Classical Mechanics

Classical mechanics is, by definition the study of symplectic manifolds (M, ω).

From the viewpoint of C^* algebra theory we naturally associate to it the algebra $\mathfrak{A} = C(M)$ for M compact and $C_0(M)$ for M noncompact.

Physical observables should be real-valued functions on M. These are clearly the self-adjoint elements of $\mathfrak{A}$.

What are states in classical mechanics? The standard viewpoint is that they are points in phase space.

For example, if we have a system of N interacting particles in $\mathbb{R}^D$ the corresponding symplectic manifold is

$$M := T^*\mathbb{R}^D = V \oplus V^\vee \tag{12.855}$$

where $V \cong \mathbb{R}^D$ and there is a canonical symplectic form ω based on the antisymmetric form:

$$\langle q_1 \oplus p_1, q_2 \oplus p_2 \rangle := p_1 \cdot q_2 - p_2 \cdot q_1 \tag{12.856}$$

In this context we would typically think of a state of a classical mechanical system as a specification of the coordinates and momenta, that is, a point in M.

However, to make the formulation of classical mechanics as parallel as possible with quantum mechanics we should broaden our notion of "physical states" to include states on the C^* algebra $\mathfrak{A} = C_0(M)$. As we saw above these correspond to probability measures on M. The pure states, corresponding to Dirac measures supported at points $p \in M$ are what are typically thought of as states in classical mechanics. The general states, in the sense of C^*-algebra theory might be considered "classical probability distributions."

Now note that we have a natural pairing of states and observables to the set $\mathbb{B}$ of Borel measures on the real line:

$$\mathcal{S} \times \mathcal{O} \to \mathbb{B} \tag{12.857}$$

The value of the measure (f, μ) on a measurable set $E \subset \mathbb{R}$ is defined by:

$$(f, \mu)(E) := \int_{f^{-1}(E)} f d\mu \tag{12.858}$$

A key point here is that the expectation value of f is $\int_X f d\mu$ and if $d\mu$ is a Dirac measure at some point $x \in \mathcal{M}$ then there is no variance, $\langle f^2 \rangle_{d\mu} = \langle f \rangle_{d\mu}^2$.

Finally, since $\mathcal{M}$ is symplectic there is a canonical Liouville measure $d\mu_{\text{Liouville}} = \frac{\omega^n}{n!}$ where ω is the symplectic form and given a state $d\mu$ we can define $d\mu(x) = \rho(x) d\mu_{\text{Liouville}}$. Then the classical analog of the Schrödinger equation is the Liouville equation

$$\frac{d\rho(x;t)}{dt} = -\{H, \rho\} \tag{12.859}$$

12.11.8 States And Operators In Quantum Mechanics

The essential part of quantum mechanics is the *Born rule*, a pairing of physical observables $\mathcal{O}$ and states ω to produce a probability distribution on the real line: $(\omega, \mathcal{O}) \in \mathbb{B}$. The value of $(\omega, \mathcal{O})$ evaluated on the Borel set $E \subset \mathbb{R}$ is interpreted as the probability that the observable $\mathcal{O}$ measured in the state ω will take values in the set E.

Now, in the C^*-algebra approach to quantum mechanics the central object is not a phase space, but a C^*-algebra $\mathfrak{A}$. So, to a physical system that we wish to describe, first and foremost we assign a C^* algebra. Then the self-adjoint elements $\mathfrak{A}_{\mathbb{R}}$ are meant to correspond to the (bounded) physical observables. The physical states are meant to correspond to the the states $\mathcal{S}(\mathfrak{A})$ in the sense of C^*-algebra theory.

In order to formulate the Born rule we need a Hilbert space, because there is no spectral theorem for abstract C^*-algebras. Rather we have a spectral theorem for bounded operators on a Hilbert space. Given a representation $\pi : \mathfrak{A} \to \mathcal{B}(\mathcal{H})$ and given a self-adjoint element $a \in \mathfrak{A}$ there is a corresponding projection-valued-measure $P_{\pi(a)}$ of operators on Hilbert space. The state on the C^* algebra maps to a trace-class positive operator ρ of trace one. Now we can state the Born rule: The pairing of state and observable is the

♣Does $\mathfrak{A}$ have to be in the compact operators? ♣

probability measure on $\mathbb{R}$ given by

$$(a,\omega)(E) := \mathrm{Tr}_{\mathcal{H}} P_{\pi(a)}(E)\rho \tag{12.860}$$

on Borel-measurable subsets $E \subset \mathbb{R}$.

An important special case arises when the physical system also has a version described in terms of classical mechanics, hence using a symplectic manifold (M,ω). Then we somehow want to assign quantum operators to functions on M, but they might no longer commute. Thus, at the minumum we want a map

$$Q : C_0(M) \to \mathcal{B}(\mathcal{H}) \tag{12.861}$$

However, Q is in general <u>not</u> a representation. In order to make good sense of probabilities, it should be a positive map of C^* algebras.

****

1. Example: $M = T * X$.
2. Example: M is Kähler with positive holomorphic line bundle.
3. Special case: Induced representations and the orbit method.
4. Semiclassical limits and coherent states.

****

13. Boundary conditions

Now let us enrich our theory by allowing the time-slices Y to be manifolds with boundary. There will be a set of boundary conditions $\mathcal{B}_0$, and we will attach an element of $\mathcal{B}_0$ to each boundary component of ∂Y.

A bordism X from Y_0 to Y_1 will thus have two kinds of boundaries:

$$\partial X = Y_0 \cup Y_1 \cup \partial_{\mathrm{cstr}} X \tag{13.1}$$

where $\partial_{\mathrm{cstr}} X$ is the time-evolution of the spatial boundaries. We will call this the "constrained boundary."

Figure 38: A general open-closed bordism.

In $d = 2$, in this enlarged geometric category the initial and final state-spaces are associated with circles, as before, and now also with intervals. The boundary of each interval carries a label $a, b, c, \dots$ from the set $\mathcal{B}_0$.

Figure 39: Morphism space for open strings: $\mathcal{O}_{ab}$.

Definition: We denote the space $\mathcal{O}_{ab}$ for the space associated to the interval $[0,1]$ with label b at 0 and a at 1.

In the theory of D-branes, the intervals are open strings ending on submanifolds of spacetime. That is why we call the time-evolution of these boundaries the "constrained boundaries" – because the ends are constrained to live in the D-brane worldvolume.

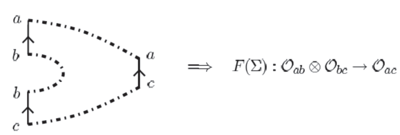


Figure 40: Basic bordism of open strings.

As in the closed case, the bordism $[0,1] \times [0,1]$ defines $P_{ab} : \mathcal{O}_{ab} \to \mathcal{O}_{ab}$, and we can assume WLOG that it is $P_{ab} = 1$.

Now consider the bordism in **??**. This clearly gives us a bilinear map

$$\mathcal{O}_{ab} \times \mathcal{O}_{bc} \to \mathcal{O}_{ac} \tag{13.2}$$

As in the closed case we see that these maps satisfy an associativity law. Moreover, as in the closed case, there is an element 1_a, defined by **??** which is an identity for the multiplication.

Comparing with the definition of a category we see that we should interpret $\mathcal{B}_0$ as the space of objects in a category $\mathcal{B}$, whose morphism spaces $Hom(b, a) = \mathcal{O}_{ab}$. Note that the morphism spaces are vector-spaces. This is the defining property of a $\mathbb{C}$-linear category. In fact, this category has a very special property. We also have the trace map **??**, and as we

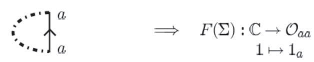


Figure 41: A disk defining an element $1_a \in \mathcal{O}_{aa}$

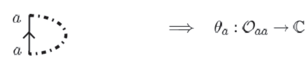


Figure 42: The trace element: $\theta_a : \mathcal{O}_{aa} \to \mathbb{C}$.

learn from considering the S-shaped bordism (the open string analog of 7. We learn that $\theta_a : \mathcal{O}_{aa} \to \mathbb{C}$ defines a nondegenerate inner product:

$$Q_a(\psi_1, \psi_2) = \theta_a(\psi_1 \psi_2) \tag{13.3}$$

Thus, the $\mathcal{O}_{aa}$ are Frobenius algebras.

Moreover, using the S-shaped bordism analogous to 7 we learn that $\mathcal{O}_{ab}$ is dual to $\mathcal{O}_{ba}$. In fact we have

$$\begin{aligned} \mathcal{O}_{ab} \otimes \mathcal{O}_{ba} \to \mathcal{O}_{aa} &\stackrel{\theta_a}{\to} \mathbb{C} \\ \mathcal{O}_{ba} \otimes \mathcal{O}_{ab} \to \mathcal{O}_{bb} &\stackrel{\theta_b}{\to} \mathbb{C} \end{aligned} \tag{13.4}$$

are perfect pairings with

$$\theta_a(\psi_1 \psi_2) = \theta_b(\psi_2 \psi_1) \tag{13.5}$$

for $\psi_1 \in \mathcal{O}_{ab}, \psi_2 \in \mathcal{O}_{ba}$.

Definition A *Frobenius category* is a $\mathbb{C}$-linear category in which there is a perfect pairing of $Hom(a, b)$ with $Hom(b, a)$ for all $a, b \in Ob(C)$ by a pairing which factorizes through the composition in either order.

Remark: It is important to note that the argument for commutativity fails in the open case: The algebras $\mathcal{O}_{aa}$ are in general *noncommutative*. This is an elementary but important point to emphasize: There is no natural ordering of small disks in a larger disk, but there *is* an ordering of points, or intervals, on a one-dimensional line.

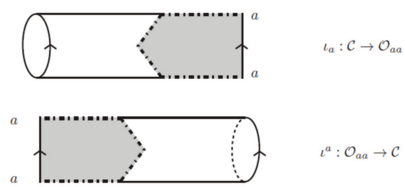


Figure 43: The open-closed transition maps

So, to give an open and closed TFT involves giving a Frobenius cateogry. But the open and closed strings must also be related to each other. The essential new information is a pair of linear maps

$$\begin{aligned} \iota_a &: \mathcal{C} \to \mathcal{O}_{aa} \\ \iota^a &: \mathcal{O}_{aa} \to \mathcal{C} \end{aligned} \tag{13.6}$$

corresponding to the open-closed string transitions of **??**.

By drawing pictures we can readily discover the following necessary algebraic conditions:

1. ι_a is an algebra homomorphism

$$\iota_a(\phi_1\phi_2) = \iota_a(\phi_1)\iota_a(\phi_2) \tag{13.7}$$

2. The identity is preserved

$$\iota_a(1_{\mathcal{C}}) = 1_a \tag{13.8}$$

3. Moreover, ι_a is central in the sense that

$$\iota_a(\phi)\psi = \psi\iota_b(\phi) \tag{13.9}$$

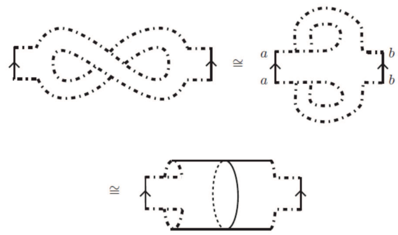


Figure 44: Factorization of the open string loop on closed string exchange. Also known as the "Cardy condition."

for all $\phi \in \mathcal{C}$ and $\psi \in \mathcal{O}_{ab}$

4. ι_a and ι^a are adjoints:

$$\theta_{\mathcal{C}}(\iota^a(\psi)\phi) = \theta_a(\psi\iota_a(\phi)) \tag{13.10}$$

for all $\psi \in \mathcal{O}_{aa}$.

5. The "Cardy conditions."[50] Define $\pi_b{}^a : \mathcal{O}_{aa} \to \mathcal{O}_{bb}$ as follows. Since $\mathcal{O}_{ab}$ and $\mathcal{O}_{ba}$ are in duality (using θ_a or θ_b), if we let ψ_μ be a basis for $\mathcal{O}_{ba}$ then there is a dual basis ψ^μ for $\mathcal{O}_{ab}$. Then we define

$$\pi_b{}^a(\psi) = \sum_\mu \psi_\mu \psi \psi^\mu, \tag{13.11}$$

and we have the "Cardy condition":

$$\pi_b{}^a = \iota_b \circ \iota^a. \tag{13.12}$$

This is illustrated in **??**.

Exercise

Draw pictures associated to the other algebraic conditions given above.

Theorem Open-Closed Sewing Theorem. The above conditions are the complete set of sewing constraints on the algebraic data.

[50]These are actually generalization of the conditions stated by Cardy. One recovers his conditions by taking the trace. Of course, the factorization of the double twist diagram in the closed string channel is an observation going back to the earliest days of string theory.

This is proved in the paper of Moore-Segal cited below.

Example: Let G be a finite group. There is a natural category $\mathcal{B}$ associated to G, namely the category of all finite-dimensional complex representations of G. We take the morphisms to be

$$\hom(V,V') = \mathrm{Hom}_G(V,V') \tag{13.13}$$

that is, morphisms are $\mathbb{C}$-linear transformations $V \to V'$ commuting with the G-actions, also known as intertwiners.

The algebra $\mathcal{O}_{VV} = \mathrm{End}_G(V)$ and the natural trace is

$$\theta_V(\psi) = \frac{1}{|G|}\mathrm{Tr}_V(\psi) \tag{13.14}$$

Let $\{V_\mu\}$ be a complete set of distinct irreps of G. Then, for G compact, all representations are completely decomposable into sums of irreps. More precisely, we have the *isotypical decomposition*

$$V \cong \oplus_\mu M_\mu \otimes V_\mu \tag{13.15}$$

Here M_μ are degeneracy spaces. They can be identified with

$$M_\mu := \hom(V_\mu, V) \tag{13.16}$$

Schur's lemma then tells us that

$$\mathrm{End}_G(V) \cong \oplus_\mu \mathrm{End}(M_\mu) \tag{13.17}$$

is a sum of matrix algebras.

What shall we take for the closed string algebra? The closed to open map ι_V must map to the center. If all the spaces M_μ are nonzero then the center of $\mathrm{End}_G(V)$ is just a direct sum of $\mathbb{C}$, one for each irrep, or equivalently, one for each conjugacy class. We can identify this space with the group algebra discussed above in the examples of closed theories. In particular, we can take $\mathcal{C}$ to be the algebra of class functions on G. Given such a function $f = \sum_g a_g g$ we define $\iota_V(f)$ to be $\sum_g a_g \rho(g)$, and if f is a class function then $\iota_V(f)$ will be central in $\mathrm{End}_G(V)$. Conversely $\iota^V(\Psi)$ is

$$\sum_\mu \mathrm{Tr}_V(P_\mu \Psi)\chi_\mu \tag{13.18}$$

where P_μ is the projector to the isotypical subspace for V_μ.

Exercise

Write out the full set of open-closed string data for the example of a finite group and check the sewing conditions.

Once again - we can ask what geometrical problem we are solving here. We will see the answer below.

---

14. Open and closed 2D TFT in the semisimple case: D-branes and vector bundles

As we saw in the closed case, if $\mathcal{C}$ is semisimple one can go further. In this case we can derive a spacetime $X = \mathrm{Spec}(\mathcal{C})$, and the data of a Frobenius algebra is given by the closed string coupling $\theta_x = g_x^{-2}$ on each connected component of spacetime.

In this section we address the question: *What is the "spacetime interpretation" of the open string sector?*

We first need two theorems from abstract algebra, which we just state:

Definition: A Frobenius algebra is *simple* if there are no nontrivial ideals. It is semisimple if it is a direct sum of simple algebras.

Wedderburn Theorem A semisimple (noncommutative) Frobenius algebra $\mathcal{O}$ is isomorphic to a direct sum of matrix algebras:

$$\mathcal{O} = \oplus_{i=1}^{N} \mathrm{Mat}_{n_i}(\mathbb{C}) \tag{14.1}$$

with $\theta = \oplus_i \theta_i \mathrm{Tr}_i$.

Theorem. A Frobenius algebra is semisimple iff the characteristic element $H = \sum \psi_\mu \psi^\mu$ is invertible.

In one direction this is obvious. One simply computes that if (14.1) is true then $H = \oplus_i \frac{n_i}{\theta_i} 1_{n_i}$ is clearly invertible. For the other direction we use a standard criterion for semisimplicity: An algebra is semisimple if the trace in the left-regular representation defines a nondegenerate quadratic form: $(\psi, \psi') \to \mathrm{Tr}\psi\psi'$. [See Lang's *Algebra*, for example.] Now one need only note that $\theta(H\psi) = \mathrm{Tr}\psi$ ♠

Now we have:

Theorem: If $\mathcal{C}$ is semisimple, then for any boundary condition a, $\mathcal{O} = \mathcal{O}_{aa}$ is semisimple and $\mathcal{O} = \mathrm{End}_{\mathcal{C}}(W)$ for some finite-dimensional representation W of $\mathcal{C}$.

Proof: If $\mathcal{C}$ is semisimple it is a direct sum of $\mathbb{C}$ for each spacetime point x. Fix a single point x and a boundary condition a. Then $\iota_a(\epsilon_x) = P_{x,a}$ are central projection operators and $\mathcal{O}_x = P_x \mathcal{O} P_x$ is an algebra over the Frobenius algebra $\mathcal{C}_x = \epsilon_x \mathbb{C}$. So we can work over a single spacetime point. Therefore, we must have $\iota^a(1_{\mathcal{O}_x}) = \alpha 1_{\mathcal{C}_x}$. In fact, $\alpha = \theta_a(1_{\mathcal{O}_x})/\theta_x$ from the adjoint relation. From the Cardy condition

$$\alpha 1_{\mathcal{O}_x} = \sum \psi_\mu \psi^\mu \tag{14.2}$$

where ψ_μ is a basis for $\mathcal{O}_x$. Applying θ_a we find $\alpha\theta_a(1_{\mathcal{O}_x}) = \dim\mathcal{O}_x$. Let us assume L that $\mathcal{O}_x$ is nonzero. Then α is nonzero. This says the characteristic element of $\mathcal{O}_x$ is invertible, and hence $\mathcal{O}_x$ is semisimple. By the Wedderburn theorem we conclude that $\mathcal{O}_x$ is a direct sum of matrix algebras:

$$\mathcal{O}_x = \oplus_{i=1}^{s} Mat_{n_i}(\mathbb{C}) \tag{14.3}$$

However, we can go further. By cyclicity, the trace θ_a must have the form

$$\theta_a(\psi) = \sum_{i=1}^{s} \theta_a^i \mathrm{Tr}(\psi^i) \qquad \psi = \oplus_{i=1}^{s} \psi^i \tag{14.4}$$

Now, going back to the Cardy condition one finds that in fact we must have $s = 1$; there can be only *one* summand in (14.4); that is, $\mathcal{O}_{aa}$ must be a <u>full</u> matrix algebra. ♠

Thus, the most general $\mathcal{O}_{aa}$ is obtained by choosing a vector space $W_{x,a}$ for each spacetime point x and

$$\mathcal{O}_{aa} = \oplus_x \mathrm{End}(W_{x,a}) \tag{14.5}$$

What we have discovered is that to a boundary condition a we can associated a *vector bundle over spacetime.* These are the D-branes in this 2D TFT. [51]

Exercise Show that if $\psi = \oplus_x \psi_x$ then

$$\begin{aligned} \theta_a(\psi) &= \sum_x \sqrt{\theta_x} \mathrm{Tr}(\psi_x) \\ \iota^a(\psi) &= \oplus_x \mathrm{Tr}(\psi_x) \frac{\varepsilon_x}{\sqrt{\theta_x}} \\ \pi_b{}^a(\psi_{aa}) &= \oplus_x \frac{1}{\sqrt{\theta_x}} \mathrm{Tr}_{W_{x,a}}(\psi_{x,aa}) P_{x,b} \end{aligned} \tag{14.6}$$

$$\mathcal{O}_{ab} \cong \oplus_x \mathrm{Hom}(W_{x,a}; W_{x,b}) \tag{14.7}$$

So, we have a complete answer to the category of boundary conditions in this simplest of all cases:

Theorem

• If $\mathcal{C}$ is semisimple, corresponding to a space-time X, then the category $\mathcal{B}$ of boundary conditions is equivalent to the category $\mathrm{Vect}(X)$ of vector bundles on X, by the inverse functors

$$\{W_x\} \mapsto \oplus W_x \otimes a_x, \tag{14.8}$$

$$a \mapsto \{\mathcal{O}_{a_x a}\}. \tag{14.9}$$

[51] Notice that the bundle is not unique, since we can always tensor with a line bundle $W_x \to W_x \otimes L_x$ where L_x is one-dimensional. This, ultimately, is the source of the B-field degree of freedom in string theory.

where $\mathcal{O}_{a_x a_x} \cong \mathbb{C}$ is suppored at x.

• The equivalence of $\mathcal{B}$ with $\mathrm{Vect}(X)$ is unique up to transformations $\mathrm{Vect}(X) \to \mathrm{Vect}(X)$ given by tensoring with a line bundle $L = \{L_x\}$ on X. • The Frobenius structure on $\mathcal{B}$ is determined by choosing a square-root $\{\sqrt{\theta}_x\}$ of the dilaton field. It is therefore unique up to multiplication by an element $\sigma \in \mathcal{C}$ such that $\sigma^2 = 1$.

Exercise *Boundary states*

Let $B_a := \iota^a(1_{\mathcal{O}_{aa}})$. This is known as the "boundary state" for boundary condition a.

a.) Show that the partition function for a genus g amplitude with h holes all with constrained boundaries with boundary condition a is given by

$$Z = \theta_{\mathcal{C}}(H^g B_a^h) \tag{14.10}$$

b.) Show that under a change of scale $\theta_{\mathcal{C}} \to \lambda^{-2}\theta_{\mathcal{C}}$ the boundary states scale as $B_a \to \lambda B_a$.

c.) Show that the closed string coupling is always the square of the open string coupling.

Exercise *Open Problem*

Generalize the above theorem to the unoriented case, and relate the classification of boundary conditions to KR theory of spacetime.

---

15. Closed strings from open strings

In string theory one usually thinks of specifying a spacetime manifold, then a metric on that manifold, then other closed-string data, and finally one starts "wrapping branes" around various cycles. This way of thinking puts the closed string on a more fundamental basis - one asks - for a given closed string background, what are the D-branes in that background? How do we classify them?

The above 2D TFT suggests a radically different point of view, which makes the open strings more fundamental, and the spacetime, and its closed strings a derived concept. There is evidence that this is indeed a more fundamental view from the Matrix theory approach to defining M-theory and from the AdS/CFT correspondence.

If one begins with a Frobenius category, one can try to derive the closed string algebra. In the semisimple case we might proceed by considering a "generic" boundary condition a and then taking the *center* of the algebra $\mathcal{O}_{aa}$. Generic means - *a postiori* - that $W_{a,x}$ is the nonzero vector space on every spacetime point x.

How does one generalize this idea? The essential point, as described in ref. 2 below is to take the "cyclic cohomology" of the Frobenius category. This will define for us a commutative Frobenius algebra, from which we can derive the spacetime. See reference 2 and references therein for further details. ♣Explain more about cyclic cohomology ♣

15.1 The Grothendieck group

In the category of boundary conditions we can always define direct sums of objects $a \oplus b$ as follows. For any two objects a and b we define a new object $a \oplus b$ by

$$\mathcal{O}_{a\oplus b,c} := \mathcal{O}_{ac} \oplus \mathcal{O}_{bc} \tag{15.1}$$

$$\mathcal{O}_{c,a\oplus b} := \mathcal{O}_{ca} \oplus \mathcal{O}_{cb}, \tag{15.2}$$

and hence

$$\mathcal{O}_{a\oplus b,a\oplus b} := \begin{pmatrix} \mathcal{O}_{aa} & \mathcal{O}_{ab} \\ \mathcal{O}_{ba} & \mathcal{O}_{bb} \end{pmatrix}, \tag{15.3}$$

with the composition laws using the above data and matrix multiplication. (This construction is known in operator algebra theory as the *linking algebra*.) Finally, the trace is

$$\theta_{a\oplus b} : \mathcal{O}_{a\oplus b,a\oplus b} \to \mathbb{C} \tag{15.4}$$

given by

$$\theta_{a\oplus b} \begin{pmatrix} \psi_{aa} & \psi_{ab} \\ \psi_{ba} & \psi_{bb} \end{pmatrix} = \theta_a(\psi_{aa}) + \theta_b(\psi_{bb}). \tag{15.5}$$

The new object is the direct sum of a and b in the enlarged category of boundary conditions.

Now, let us recall that a semigroup S is a set with an associative binary product. We asume it is commutative so we denote the product $a \oplus b$, because of the application below. There is no notion of a unit or an inverse, so S is not a group.

Nevertheless, - one can form a corresponding group $K(S)$ by manufacturing inverses as follows. $K(S)$ is defined to be the set of equivalence classes of pairs (a, b) where the equivalence relation is

$$(a, b) \sim (a', b') \qquad \Leftrightarrow \qquad \exists c \qquad a \oplus b' \oplus c = a' \oplus b \oplus c \tag{15.6}$$

(Note that in general $a \oplus c = b \oplus c$ does not imply $a = b$.)

Then the group law is

$$[(a, b)] + [(c, d)] := [(a \oplus c, b \oplus d)] \tag{15.7}$$

We can think of $[(a, b)]$ informally as $a - b$ and the equivalence relation comes from rearranging $a - b = a' - b'$.

This simple construction is known as the *Grothendieck construction*.

Example 1: Let $S = \mathbb{N}_+$ be the natural numbers $1, 2, 3, \ldots$. Then $K(S) = \mathbb{Z}$, the integers.

Example 2: Let S be isomorphism classes of finite dimensional vector spaces. Then $K(S)$ is the group of virtual vector spaces. These are in 1-1 correspondence with the integers.

The Grothendieck construction can be applied to the set of isomorphism classes of objects in our category $\mathcal{B}$ to define $K(\mathcal{B})$. Then applied to the category of vector bundles on a topological space X it defines $K^0(X)$.

In the semisimple case one can show that the recovery of the closed string sector from the open string sector amounts to

$$K(\mathcal{B}) \otimes \mathbb{C} \cong \mathcal{C}. \tag{15.8}$$

Exercise

a.) Show that the group law (15.7) is well-defined.

b.) If there is an infinite element in S, i.e. an element $\infty \in S$ such that $a \oplus \infty = \infty$ then $K(S) = 0$.

Exercise

Using $K(pt) \cong \mathbb{Z}$ prove equation (15.8).

---

16. Three Dimensions And Modular Tensor Categories

♣Put unitarity here, since this is where the concept is used in an important way. ♣

17. Other Generalizations

1. Homotopy Field Theory. Turaev's book. Equivariant theory: Coupling to a gauge bundle.
2. Invertible TFT's
3. Anomaly Field Theories
4. Field theories valued in other field theories, and "relative field theory"

18. Higher Categories, Locality, and extended objects

Finally, we would like to indicate one direction in which the above ideas continue to be developed in current research. See the papers by J. Lurie and A. Kapustin cited below for details.

One way to motivate this subject is to impose a greater degree of locality than we have thus far imposed.

Figure 45: Cutting a closed 2-fold into two pieces

Figure 46: Cutting a circle into two intervals.

We have shown how the notion of a functor allows us to compute partition functions $Z(M_2)$ of a compact two-manifold by splitting it up into pieces and associating algebraic data with the pieces.

For example, in 45 we cut a Riemann surface, and associate algebraic objects with the pieces. But now, we can ask if we can similarly learn about the value of $Z(Y)$, where Y is the cutting circle, by splitting $Y = S^1$ into intervals: Can we define a $Z(I^{\pm})$ and glue these to get a vectorspace $Z(Y)$ as in 46 ?

This becomes important if we want to evaluate $Z(M)$ in d-dimensional TFT for $d \geq 3$. Now there is no simple decomposition analogous to that for Riemann surfaces, in general. In general we would need to chop up M into *manifolds with corners*. A manifold with corners is a space which is locally like $\mathbb{R}^n \times \mathbb{R}_+^m$.

Figure 47: Hierarchies of structure in 2d and 3d TFT

So, let us look at the hierarchy we had in 2d TFT. See 47.

In the d=3 case there is a nice way to say what the *category* associated to S^1 would be: Letting Z denote the functor, we can define

$$\tilde{Z}(M) = Z(S^1 \times M) \tag{18.1}$$

and then $\tilde{Z}$ is a two-dimensional TFT. As we have just seen, for such things we associate a $\mathbb{C}$-linear category to a point. This will then be the category we associate with $Z(S^1)$: This is a baby version of the Kaluza-Klein idea.

Figure 48: Pictorial version of morphisms between morphisms

But what should we then assign to a point in the 3d TFT? The answer is some kind of 2-category. In a two category we have objects, morphisms, and morphisms between morphisms, which can be pictured as in 48. There are now lots of axioms and, in fact, mathematicians are not quite in agreement as to what is the best definition of an n-category.

Figure 49: A physical realization of higher morphisms using point and line defects in a boundary.

One nice way, advocated by Kapustin, of understanding the physical role of these higher categories is to introduce extended objects into the field theory. Let us define a *domain wall* to be some extended object which separates space into two components. Thus it is real codimension one in spacetime. Let us imagine that there are many kinds of domain walls, labeled by $A, B, \ldots$. A domain wall in which there is no space on one side is a boundary condition.

Now, within the domain wall there could be "sub"-domain walls. These might or might not be "bound" to the domain wall. If they are not, they would constitute real codimension two objects in spacetime. Now they will have labels $\alpha, \beta, \ldots$. If they can be fused then they can be viewed as "morphisms" between the "objects" $A, B, \ldots$ which are the domain walls.

Now, within the sub-domain walls we could have sub-sub-domain walls, labeled by $i, j, k \ldots$ and separating type α from β, etc. These would correspond to codimension three objects. Mathematically they could be interpreted as 2-morphisms.

The simplest example of a 2-category is the 2-category of algebras:

1. Objects = Algebras.

2. Morphisms = Bimodules. So, if A, B are algebras then a morphism is a left-A and right-B bimodule. These can be tensored to give composition of morphisms.

3. 2-Morphisms: Maps of bimodules.

19. References

There is an enormous literature on topological field theory. One of the key early papers is

1. M.F. Atiyah, "Topological quantum field theories." Inst. Hautes Etudes Sci. Publ. Math. No. 68 (1988), 175–186 (1989).

This paper was inspired by Witten's work together with Graeme Segal's axiomatization of conformal field theory, now available as:

2. G. Segal, "The Definition Of Conformal Field Theory," in *Topology, Geometry and Quantum Field Theory: Proceedings of the 2002 Oxford Symposium in Honour of the 60th Birthday of Graeme Segal*

For an introduction written more from a physicist's perspective see

We have also relied on lecture notes

3. Segal, Stanford notes, http://www.cgtp.duke.edu/ITP99/segal/

as well as the expository article:

4. G. Segal and G. Moore, "D-branes and K-theory in 2D topological field theory" hep-th/0609042. See also chapter 2 in *Mirror Symmetry II*, Clay Mathematics Institute Monograph, by P. Aspinwall et. al.

For a very meticulous can careful discussion of topological field theory in general see

5. V.G. Turaev, *Quantum Invariants Of Knots And 3-Manifolds*, De Gruyter, 2nd edition 2010.

The special case of d=3 TQFT and modular tensor categories is described in detail in

6. Bakalov and Kirillov.

7.

For a review aimed more at physicists see

8. D. Birmingham, M. Blau, M. Rakowski,m and G. Thompson, "Topological Field Theory," Phys.Rept. 209 (1991) 129-340

For bordism theory see the classic book

9. Milnor and Stasheff, *Characteristic Classes*, PUP

We have used some material from the very nice lecture notes of Dan Freed :

10. D. Freed, "Bordism Old And New,"

https://www.ma.utexas.edu/users/dafr/M392C-2012/index.html

For recent developments in higher category theory and locality see

11. A. Kapustin, "Topological Field Theory, Higher Categories, and Their Applications," arXiv:1004.2307 [math.QA]

12. J. Lurie, "On the classification of topological field theories," arXiv:0905.0465 [math.AT]

13. D. Freed, "The cobordism hypothesis"

A. Sums Over Symplectic Lattices And Theta Functions

A.1 Symplectic structures, complex structures, and metrics

A frequently recurring problem is how to express a sum of a gaussian function over a symplectic lattice in terms of theta functions.

Suppose $V_{\mathbb{Z}} \subset V_{\mathbb{R}}$ is a lattice, of rank $2N$, and suppose we have a symplectic form Ω, integral valued on $V_{\mathbb{Z}}$.

A complex structure J on $V_{\mathbb{R}}$ is compatible with Ω if

$$\Omega(Jv, Jw) = \Omega(v, w) \tag{A.1}$$

In this situation we can define a symmetric quadratic form:

$$g(v, w) := \Omega(Jv, w). \tag{A.2}$$

We now assume there is a symplectic basis [52] α^I, β_I for $V_{\mathbb{Z}}$, $I = 1, \dots, N$, such that

$$\begin{aligned} &\Omega(\alpha^I, \alpha^J) = \Omega(\beta_I, \beta_J) = 0 \\ &\Omega(\alpha^I, \beta_J) = \delta^I{}_J \end{aligned} \tag{A.3}$$

Now we choose a basis of vectors ζ^I of type $(0,1)$. We extend J $\mathbb{C}$-linearly to $V_{\mathbb{C}}$ so that, by definition

$$J \cdot \zeta^I = i\zeta^I \tag{A.4}$$

We can express the complex structure J in terms of the components of the period matrix. The latter is defined by choosing a basis ζ^I of vectors of type $(1,0)$ of the form:

$$\zeta^I := \alpha^I + \tau^{IJ}\beta_J \tag{A.5}$$

From $g(\zeta^I, \zeta^J) = g(\zeta^J, \zeta^I)$ we learn that τ^{IJ} is symmetric, and moreover g is of type $(1,1)$. Note that

$$g(\zeta^I, \bar{\zeta}^J) = 2\mathrm{Im}\tau^{IJ} \tag{A.6}$$

We can express the complex structure in terms of the period matrix as follows. The complex structure acts as:

$$\begin{aligned} J \cdot \alpha^I &= A_{I'}^{I}\alpha^{I'} + C^{I'I}\beta_{I'} \\ J \cdot \beta_I &= B_{I'I}\alpha^{I'} + D^{I'}{}_I\beta_{I'} \end{aligned} \tag{A.7}$$

We define components of vectors by

$$v = v_I^1\alpha^I + v_2^I\beta_I = \begin{pmatrix} v_I^1 & v_2^I \end{pmatrix} \begin{pmatrix} \alpha^I \\ \beta_I \end{pmatrix} \tag{A.8}$$

so that J acts on the components as the matrix

[52] We would like to relax this assumption and discuss what happens when the skew eigenvalues of Ω are other integers.

$$J = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{A.9}$$

Compatibility of the complex structure implies that this defines a symplectic matrix.

Equating real and imaginary parts of (A.4), using the definition (A.7) we find the matrix expression of J in the basis α^I, β_I:

$$J = \begin{pmatrix} -Y^{-1}X & Y^{-1} \\ -Y - XY^{-1}X & XY^{-1} \end{pmatrix} \tag{A.10}$$

One can check both $J^2 = -1$ and $J^{tr}\Omega J = \Omega$.

The metric g in the α, β basis is:

$$g(v,w) = \begin{pmatrix} v_I^1 & v_2^J \end{pmatrix} \begin{pmatrix} XY^{-1}X + Y & -XY^{-1} \\ -Y^{-1}X & Y^{-1} \end{pmatrix} \begin{pmatrix} w_I^1 \\ w_2^J \end{pmatrix} \tag{A.11}$$

It is useful to have formulae for the transformation from the integral symplectic basis to the complex basis.

$$\begin{pmatrix} \zeta^I \\ \bar{\zeta}^I \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ 1 & \bar{\tau} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{A.12}$$

has inverse:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{i}{2} \begin{pmatrix} \bar{\tau} & -\tau \\ -1 & 1 \end{pmatrix} Y_{IJ}^{-1} \begin{pmatrix} \zeta^J \\ \bar{\zeta}^J \end{pmatrix} \tag{A.13}$$

Thus the complex projections of (A.8) are:

$$\begin{aligned} v^{(1,0)} &= -\frac{i}{2}(v_2^J - v_I^1 \bar{\tau}^{IJ}) Y_{JK}^{-1} \zeta^K \\ v^{(0,1)} &= \frac{i}{2}(v_2^J - v_I^1 \tau^{IJ}) Y_{JK}^{-1} \bar{\zeta}^K \end{aligned} \tag{A.14}$$

Note that

$$v = v^{(1,0)} + v^{(0,1)} \tag{A.15}$$

A.2 Statement Of The Problem

We will now assume that g is positive definite, i.e. $\mathrm{Im}\tau$ is positive definite. The first problem is to express

$$S_0 := \sum_{\nu \in V_{\mathbb{Z}}} e^{-\frac{1}{2}\pi k g(\nu,\nu) + \Omega(\nu, \tilde{\ell})} \tag{A.16}$$

in terms of theta functions for the complex torus $V_{\mathbb{R}}/V_{\mathbb{Z}}$. We write

$$\tilde{l} = -l_I^2\alpha^I + l_1^I\beta_I \tag{A.17}$$

so that $\Omega(\nu, \tilde{l}) = n_I l_1^I + m^I l_I^2$.

An important variation on (A.16) is the following. Suppose that φ is a quadratic refinement of Ω, i.e.

$$\varphi(\nu_1 + \nu_2) = \varphi(\nu_1)\varphi(\nu_2)e^{i\pi k\Omega(\nu_1,\nu_2)} \tag{A.18}$$

Then we can define a twisted sum:

$$S_1 := \sum_{\nu \in V_{\mathbb{Z}}} \varphi(\nu) e^{-\frac{1}{2}\pi k g(\nu,\nu) + \Omega(\nu,\tilde{\ell})} \tag{A.19}$$

and again we would like to express this in terms of theta functions for $V_{\mathbb{R}}/V_{\mathbb{Z}}$. Note that for k even, there is no distinction between a twisted and untwisted sum.

A.3 Level κ Theta Functions

We define our level κ theta functions to be

$$\Theta_{\beta,\kappa}(\xi,\tau) = \sum_{s_I \in \mathbb{Z}} e^{2\pi i\kappa(s_I + \frac{1}{2\kappa}\beta_I)\tau^{IJ}(s_I + \frac{1}{2\kappa}\beta_I)} e^{2\pi i \xi^I(2\kappa s_I + \beta_I)} \tag{A.20}$$

MORE ABOUT THETA FUNCTIONS

Having chosen a symplectic basis the general quadratic refinement can be written as

$$\varphi(\nu) = e^{2\pi i(\theta^I \nu_I^1 + \phi_I \nu_2^I)} e^{i\pi k \nu_I^1 \nu_2^I} \tag{A.21}$$

Claim 1:

$$S_1 = \sqrt{\det \frac{2}{k} Y} e^Q \sum_{\beta \in (\mathbb{Z}/k\mathbb{Z})^N} \Theta_{\beta,k/2}(\delta^I, \tau^{IJ})\Theta_{-\beta,k/2}(\bar{\delta}^I, -\bar{\tau}^{IJ}) \tag{A.22}$$

where δ and $\tilde{\delta}$ are, essentially, the $(0,1)$ and $(1,0)$ components of ℓ, respectively. More precisely:

$$\begin{aligned} l_1^I + \tau^{IJ} l_J^2 + (2\pi i)(\theta^I + \tau^{IJ}\phi_J) &= 2\pi i k \delta^I \\ l_1^I + \bar{\tau}^{IJ} l_J^2 + (2\pi i)(\theta^I + \bar{\tau}^{IJ}\phi_J) &= 2\pi i k \bar{\delta}^I \\ l_1^I + \tau^{IJ} l_J^2 &= -2iY^{IJ}\tilde{l}_J^{(0,1)} \\ l_1^I + \bar{\tau}^{IJ} l_J^2 &= 2iY^{IJ}\tilde{l}_J^{(1,0)} \end{aligned} \tag{A.23}$$

and

$$Q = \frac{\pi k}{2}(\delta^I - \bar{\delta}^I) Y_{IJ}^{-1} (\delta^J - \bar{\delta}^J) \tag{A.24}$$

Claim 2:

$$S_0 = \sqrt{\det \frac{2}{k} Y} e^{\frac{1}{2\pi k}\ell^2_J Y^{JK} l^2_K} \sum_{\beta,\bar\beta} \Theta_{\beta,2k}(\frac{1}{4\pi i k}\psi^I, \tau^{IJ})\Theta_{\bar\beta,2k}(\frac{1}{4\pi i k}\bar\psi^I, -\bar\tau^{IJ}) \tag{A.25}$$

where ψ^I are defined in the proof below. We sum over a set of $4k$ characteristics $(\beta, \bar\beta)$ described below. They satisfy $2(\beta+\bar\beta)=0$ and $k(\beta-\bar\beta)=0$.

****

Stress that the original sums did NOT make use of a Lagrangian splitting and that there are many Lagrangian splittings related by $Sp(2N,\mathbb{Z})$. Derive the transformation laws on τ and theta functions from this.

****

A.4 Splitting instanton sums

Proof: The lattice splits as $\Lambda = \Lambda_1 \oplus \Lambda_2$. We write $\nu = n_I\alpha^I + m^I\beta_I$, with Λ_1 spanned by α^I. We do a PSF on m^I. By shifting the vector $\tilde l$ it suffices to consider the case $\theta^I = \phi_I = 0$. The main formula is then

$$\begin{aligned} S_\epsilon = &\sqrt{\det \frac{2}{k} Y} e^{\frac{1}{2\pi k}\ell^2_J Y^{JK} l^2_K} \sum_{p_L,p_R} \\ &\exp\Big\{ i\pi k (p_L)_I \tau^{IJ}(p_L)_J - i\pi k (p_R)_I \bar\tau^{IJ}(p_R)_J + (p_L)_I\psi^I + (p_R)_I\bar\psi^I \Big\} \end{aligned} \tag{A.26}$$

with

$$\begin{aligned} (p_L)_I &= \frac{1}{2}n_I + \frac{1}{k}(\tilde m_I + \frac{k\epsilon}{2}n_I) \\ (p_R)_I &= \frac{1}{2}n_I - \frac{1}{k}(\tilde m_I + \frac{k\epsilon}{2}n_I) \end{aligned} \tag{A.27}$$

and

$$\begin{aligned} \psi^I &= l_1^I + \tau^{IJ}l^2_J \\ \bar\psi^I &= l_1^I + \bar\tau^{IJ}l^2_J \end{aligned} \tag{A.28}$$

Now, we need to split the sum. We need to discuss the twisted and untisted cases separately.

For $\epsilon = 1$ we can write $\tilde m_I = \beta_I - k s_I$, where $s_I \in \mathbb{Z}$ are unconstrained integers and $\beta_I \in \{0,1,\dots,k-1\}$. Then the splitting is immediate, and we have

$$S_1 = \sqrt{\det \frac{2}{k} Y} e^{Q} \sum_{\beta\in(\mathbb{Z}/k\mathbb{Z})^N} \Theta_{\beta,k/2}(\frac{1}{2\pi i k}\psi^I, \tau^{IJ})\Theta_{-\beta,k/2}(\frac{1}{2\pi i k}\bar\psi^I, -\bar\tau^{IJ}) \tag{A.29}$$

$$
\begin{aligned}
\psi^I &= l_1^I + \tau^{IJ} l_J^2 \\
\bar{\psi}^I &= l_1^I + \bar{\tau}^{IJ} l_J^2 \\
l_1^I + \tau^{IJ} l_J^2 &= -2iY^{IJ} \tilde{l}_J^{(0,1)} \\
l_1^I + \bar{\tau}^{IJ} l_J^2 &= 2iY^{IJ} \tilde{l}_J^{(1,0)}
\end{aligned}
\tag{A.30}
$$

and

$$
Q = \frac{1}{2\pi k} l_J^2 Y^{JK} l_K^2 \tag{A.31}
$$

To recover the general case we shift $l_1^I \to l_1^I + 2\pi i \theta^I$, $l_I^2 \to l_I^2 + 2\pi i \phi_I$. this leads to (A.22) above.

For $\epsilon = 0$ we must work harder. We need to split $p_L = n/2 + m/k, p_R = n/2 - m/k$. We first write

$$
\begin{aligned}
n &= 2s' + \gamma & & s \in \mathbb{Z}, \gamma \in \{0,1\} \\
m &= kt' + \rho & & , t \in \mathbb{Z}, \rho \in \{0,1,\ldots,k-1\}
\end{aligned}
\tag{A.32}
$$

Then we decompose

$$
\begin{aligned}
& s' + t' = 2s + \zeta \\
& s' - t' = 2t + \zeta, \qquad s,t \in \mathbb{Z}, \zeta \in \{0,1\}
\end{aligned}
\tag{A.33}
$$

Now we have

$$
\begin{aligned}
\beta_I &= 2k\zeta_I + k\gamma_I + 2\rho_I \\
\bar{\beta}_I &= 2k\zeta_I + k\gamma_I - 2\rho_I
\end{aligned}
\tag{A.34}
$$

Note that these are, unfortunately, not defined modulo $4k$, but only modulo $2k$. Since we have chosen explicit fundamental domains above these equations still make sense, and define $2 \times 2 \times k = 4k$ distinct pairs $(\beta, \bar{\beta})$. ♠

Remarks:

• Finally, note that when k is even, the twisted sum is in fact equivalent to the untwisted sum. The way this comes about is as follows. One can express theta functions of one level in terms of those of another level. Decompose the sum over integers into a sum over integers relative to some modulus. For example, in

$$
\Theta_{\mu,k}(\omega,\tau) \equiv \sum_{n\in\mathbb{Z}} q^{k(n+\mu/(2k))^2} y^{(\mu+2kn)} \tag{A.35}
$$

we could write $n = \ell\Delta + \delta$, $0 \le \delta \le \Delta - 1$. In this way we show that

$$\Theta_{\mu,k}(\omega,\tau)=\sum_{\delta=0}^{\Delta-1}\Theta_{\Delta(\mu+2k\delta),k\Delta^2}(\omega/\Delta,\tau) \tag{A.36}$$

Thus, we can write a theta function of index k as a linear combination of theta functions of index $k\Delta^2$. When k is even we can arrange the sum on $\beta,\bar{\beta}$ in (A.25) to rewrite it as (A.22).

A.5 Geometrical Interpretation

$V_{\mathbb{R}}/V_{\mathbb{Z}}$ is a principally polarized variety. $\Theta_{\beta,\kappa}(\xi,\tau)$ is a section of a line bundle. If μ is a vector of integers:

$$\begin{aligned}\Theta_{\beta,\kappa}(\xi+\mu,\tau)&=\Theta_{\beta,\kappa}(\xi,\tau)\\ \Theta_{\beta,\kappa}(\xi+\tau\mu,\tau)&=e^{-2\pi i\kappa\mu_I\tau^{IJ}\mu_J-4\pi i\kappa\xi^I\mu_I}\Theta_{\beta,\kappa}(\xi,\tau)\end{aligned} \tag{A.37}$$

Therefore,

$$e^{-4\pi\kappa(\mathrm{Im}\xi^I)Y_{IJ}^{-1}(\mathrm{Im}\xi^J)}\Theta_{\beta,\kappa}(\xi,\tau)\overline{\Theta_{\beta',\kappa}(\xi,\tau)} \tag{A.38}$$

is invariant. Now we compute the representative of $c_l(\mathcal{L})$:

$$\omega=\frac{1}{2\pi i}\partial\bar{\partial}\log\parallel s\parallel^2=2\kappa dx^I dy_I \tag{A.39}$$

where $\xi^I=x^I+\tau^{IJ}y_J$.

B. Generators For $O_{\mathbb{Z}}(Q)$

The idea is to use induction on the rank d in $II^{d,d}$. We also use some tricks mentioned in Appendix F of [23] (and I thank Steve Miller for pointing out this reference as having the relevant tricks.)

Let U be our standard copy of $II^{1,1}$. Then we first establish the result for the duality group for $d=2$, that is, for $\mathrm{Aut}(U\oplus U)$. Recall we defined a homomorphism $\psi: SL(2,\mathbb{Z})\times SL(2,\mathbb{Z})\to O_{\mathbb{Z}}(Q)$ such that if

$$A=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in SL(2,\mathbb{Z}) \tag{B.1}$$

then

$$\psi(A,1)=\begin{pmatrix}a&0&0&-b\\0&a&b&0\\0&c&d&0\\-c&0&0&d\end{pmatrix} \tag{B.2}$$

$$\psi(1,A)=\begin{pmatrix}a&b&0&0\\c&d&0&0\\0&0&d&-c\\0&0&-b&a\end{pmatrix} \tag{B.3}$$

Recall also that $SL(2,\mathbb{Z})$ is generated by S and T with

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad\qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{B.4}$$

We begin by giving a minimal set of generators for $O_{\mathbb{Z}}(Q)$ for the case $d = 2$:

Proposition: For $d = 2$, $O_{\mathbb{Z}}(Q)$ is generated by $\psi(S,1), \psi(T,1), \psi(1,S), \psi(1,T)$ together with

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{B.5}$$

and

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{B.6}$$

and no proper subset of these six generators will generate the entire group.

Proof: We consider a generic element $\mathfrak{d}$ and try to reduce it to a diagonal form, using a reduction procedure that will also be useful for the case of general d.

A fundamental observation, valid for all d, is that every row and column of $\mathfrak{d}$ consists of a set of integers with $gcd = 1$. This follows since $\det(\mathfrak{d}) = \pm 1$.

A second very useful observation, again valid for all d is that the bottom row γ_{dj} and δ_{dj} of $\mathfrak{d}$ are orthogonal vectors. To show this note that from (12.624) we know that

$$\delta\gamma^{\rm tr} + \gamma\delta^{\rm tr} = 0 \tag{B.7}$$

Taking the dd matrix element says that we have orthogonal vectors:

$$\sum_{j=1}^{d} \delta_{dj}\gamma_{dj} = 0 \tag{B.8}$$

Now let us turn to $d = 2$. Our first goal is to bring the bottom row of $\mathfrak{d}$ to $(0,0,0,\pm 1)$.

The first step in setting the bottom row to $(0,0,0,\pm 1)$ is to set $\gamma_{2,1} = 0$. Using right-multiplication by $\psi(1,A)$ for a suitable A we can set $\gamma_{2,1} = 0$. The detailed argument for this is the following: Suppose $\gamma_{2,1}$ is nonzero. Then, if $\gamma_{2,2} = 0$ we right-multiply by $\psi(1,S)$. If $\gamma_{2,2}$ is also nonzero then $\gamma_{2,1} = gx$ and $\gamma_{2,2} = gy$ where g is $gcd(\gamma_{2,1}, \gamma_{2,2})$ and x, y are relatively prime nonzero integers. Therefore we right-multiply by $\psi(1,A)$ with $a = y$ and $c = -x$, we can always find a corresponding b, d to get $A \in SL(2,\mathbb{Z})$ because x, y are relatively prime. Thus we can set $\gamma_{2,1} = 0$.

But now by (B.8) we know that $\gamma_{2,2}\delta_{2,2} = 0$, so at least one of $\gamma_{2,2}, \delta_{2,2}$ is zero. We now consider cases:

1. If $\gamma_{2,2} = \delta_{2,2} = 0$ then we must have $\delta_{2,1} = \pm 1$. Then right-multiplication by $\psi(1, S)$ puts $\mathfrak{d}$ into the desired form.

2. If $\gamma_{2,2} = 0$ but $\delta_{2,2} \neq 0$ then we can multiply by $\psi(1, A)$ for a suitable A preserving $\gamma_{2,1} = \gamma_{2,2} = 0$ and setting $\delta_{2,1} = 0$. Then $\delta_{2,2} = \pm 1$. Now $\mathfrak{d}$ is in the desired form.

3. If $\gamma_{2,2} \neq 0$ and $\delta_{2,2} = 0$ then we use

$$\mathfrak{d} \to \mathfrak{d} \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} = \begin{pmatrix} \beta & \alpha \\ \delta & \gamma \end{pmatrix} \tag{B.9}$$

to exchange δ and γ and use the previous argument. Note that

$$\psi(1, S)\psi(S, 1) = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \tag{B.10}$$

(where we define S with $b = 1, c = -1$).

Now we consider the consequences of the group conditions (12.623) and (12.624). The following is valid for all d: Suppose that the bottom row of $\mathfrak{d}$ is of the form $(0, \ldots, 0, \pm 1$. That is $\gamma_{dj} = 0$ for all j and $\delta_{d,j} = 0$ for $j = 1, \ldots, d-1$. Consider the j, d matrix element of (B.7). Using (B.26) we learn that also

$$\gamma_{j,d} = 0 \tag{B.11}$$

Similarly, from

$$\delta\alpha^{\mathrm{tr}} + \gamma\beta^{\mathrm{tr}} = \mathbf{1} \tag{B.12}$$

taking the d, j matrix elements gives

$$\alpha_{j,d} = \delta_{dd}\mathbf{1}_{j,d} \tag{B.13}$$

Altogether we learned that if $\mathfrak{d}$ has a bottom row of the form $(0, \ldots, 0, \pm 1)$ then it must be an element of the "parabolic subgroup" defined by

$$\begin{aligned} \alpha_{j,d} &= \delta_{dd}\mathbf{1}_{j,d} \\ \gamma_{d,j} &= \gamma_{j,d} = 0 \\ \delta_{d,i} &= \delta_{dd}\mathbf{1}_{d,i} \end{aligned} \tag{B.14}$$

Now return to the case $d = 2$. We have reduced $\mathfrak{d}$ to the form:

$$\begin{pmatrix} * & 0 & * & * \\ * & \delta_{22} & * & * \\ * & 0 & * & * \\ 0 & 0 & 0 & \delta_{22} \end{pmatrix} \tag{B.15}$$

with $\delta_{22} = \pm 1$.

The next step is to try to set the last column of $\mathfrak{d}$ to the form $(0,0,0,\pm 1)$. In order to do this we first put $\beta_{1,2}=0$. If it is not already zero we can left-multiply by $\psi(T^{\pm\beta_{1,2}},1)$. [53] This does not disturb the condition that the bottom row is $(0,0,0,\delta_{22})$

Next we again apply the group conditions: In particular, the dd matrix element of

$$\delta^{\mathrm{tr}}\beta+\beta^{\mathrm{tr}}\delta=0 \tag{B.16}$$

Taking the dd matrix element shows that

$$\sum_{j=1}^{d}\delta_{jd}\beta_{jd}=0 \tag{B.17}$$

are orthogonal vectors. In the case of $d=2$, if $\beta_{1,2}=0$ then $\beta_{22}\delta_{22}=0$ but since $\delta_{22}=\pm 1$ we have $\beta_{22}=0$. Next we left-multiply by $\psi(1,A)$ (with $b=0$) to set $\delta_{1,2}=0$. Now we have achieved a column of the form $(0,0,0,\pm 1)$. Now again using the group conditions we find that $\beta_{2j}=0$ and $\alpha_{2j}=\mathbf{1}_{2,j}\delta_{2,2}$. We have therefore arrived at the form

$$\begin{pmatrix}\alpha_{11}&0&\beta_{11}&0\\0&\delta_{22}&0&0\\\gamma_{11}&0&\delta_{11}&0\\0&0&0&\delta_{22}\end{pmatrix} \tag{B.18}$$

Again we apply the group conditions and find there are two possibilities:

$$\begin{pmatrix}0&0&\beta_{11}&0\\0&\delta_{22}&0&0\\\beta_{11}&0&0&0\\0&0&0&\delta_{22}\end{pmatrix} \tag{B.19}$$

with $\beta_{11}=\pm 1$ and $\delta_{22}=\pm 1$ and

$$\begin{pmatrix}\delta_{11}&0&0&0\\0&\delta_{22}&0&0\\0&0&\delta_{11}&0\\0&0&0&\delta_{22}\end{pmatrix} \tag{B.20}$$

By multiplying by $-\mathbf{1}$ we reduce each of these possibilities from four to two cases. Thus we need to add two new generators

$$\begin{pmatrix}0&0&1&0\\0&1&0&0\\1&0&0&0\\0&0&0&1\end{pmatrix} \tag{B.21}$$

and

$$\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&-1\end{pmatrix} \tag{B.22}$$

[53] At this point we have used all the generators $\psi(S,1),\psi(T,1),\psi(1,S),\psi(1,T)$.

This completes the proof of our proposition ♠

Now that we have understood the case $d = 2$ completely we can solve the general case by induction. The inductive step follows essentially the procedure we used for reducing the $d = 2$ case.

We claim that the automorphism group for $U^{\oplus d}$ is generated by the subgroups acting as $\mathrm{Aut}(U_i \oplus U_j)$ while holding the other U_k, $k \neq i, j$ fixed generate the entire group. One proves this by induction as follows.

Consider a $2d \times 2d$ matrix $\mathfrak{d} \in O_{\mathbb{Z}}(Q)$. The bottom row of integers has gcd=1, since the determinant of $\mathfrak{d}$ is ± 1. Therefore, using embedded $SL(2,\mathbb{Z})$ subgroups of the $O(2,2;\mathbb{Z})$ groups $\mathrm{Aut}(U_i \oplus U_j)$ one can use right-multiplication to bring the bottom row of $\mathfrak{d}$ to the form $(0, 0, \ldots, 0, \delta_{dd})$ where $\delta_{dd} = \pm 1$. We describe in detail how to do this:

Recall the homomorphism $\psi : SL(2,\mathbb{Z}) \times SL(2,\mathbb{Z}) \to O_{\mathbb{Z}}(U \oplus U)$ defined in (12.649) above. Let ψ_{ij} denote the homomorphism into $\mathrm{Aut}(U_i \oplus U_j)$. Now, using right-multiplication successively by

$$\psi_{12}(1, A_1), \psi_{23}(1, A_2), \ldots \psi_{d-1,d}(1, A_{d-1}) \tag{B.23}$$

with suitable $SL(2,\mathbb{Z})$ matrices $A_1, A_2, \ldots, A_{d-1}$ we can set

$$\gamma_{d,1} = \gamma_{d,2} = \cdots = \gamma_{d,d-1} = 0 \tag{B.24}$$

Once again using right-multiplication by matrices of the form (B.23) up to $\psi_{d-2,d-1}$ we can, without disturbing the condition (B.24) also set

$$\delta_{d,1} = \delta_{d,2} = \cdots = \delta_{d,d-2} = 0 \tag{B.25}$$

Now, as in the $d = 2$ case we know that $\gamma_{dd}\delta_{dd} = 0$. So, at least one of γ_{dd} and δ_{dd} is zero and at least one of $\gamma_{dd}, \delta_{d,d-1}, \delta_{dd}$ is nonzero. As in the $d = 2$ example we have three cases to deal with. The same manipulations that we used there allow us to bring the bottom row to the form

$$\begin{aligned} \gamma_{d,i} &= 0 \\ \delta_{d,i} &= \delta_{dd}\mathbf{1}_{d,i} \end{aligned} \tag{B.26}$$

Moreover, $\delta_{dd} = \pm 1$.

Now for the next step in the reduction we multiply on the left by

$$\psi_{12}(1, A_1), \psi_{23}(1, A_2), \ldots \psi_{d-2,d-1}(1, A_{d-2}) \tag{B.27}$$

with suitable $SL(2,\mathbb{Z})$ matrices $A_1, A_2, \ldots, A_{d-2}$ to set

$$\beta_{i,d} = 0 \qquad i = 1, \ldots, d-2 \tag{B.28}$$

$$\delta_{i,d} = 0 \qquad i = 1, \ldots, d-2 \tag{B.29}$$

without disturbing the conditions (B.14). Now, finally, we can multiply on the left by $\psi_{d-1,d}(A, 1)$ to set $\beta_{d-1,d} = 0$ without disturbing the previous conditions. Then again we act on the left with $\psi_{d-1,d}(A', 1)$ without disturbing previous conditions to set $\delta_{d-1,d} = 0$.

Now we apply the group conditions

$$\delta^{\rm tr}\beta + \beta^{\rm tr}\delta = 0 \tag{B.30}$$

Taking the dd matrix element shows that $\beta_{dd} = 0$. Now taking the j,d matrix elements shows that $\beta_{dj} = 0$. Next we use

$$\alpha^{\rm tr}\delta + \gamma^{\rm tr}\beta = \mathbf{1} \tag{B.31}$$

Taking the j,d matrix elements and using the previous conditions shows that $\alpha_{d,j} = \delta_{dd}\mathbf{1}_{d,j}$.

We have now set to zero all matrix elements of α, β, δ with row or column index equal to d, except for $\alpha_{dd} = \delta_{dd} = \pm 1$. The remaining matrix elements define an automorphism of $U^{\oplus(d-1)}$ and therefore by induction we have established our claim.

♣There is probably a simpler inductive proof adapting the inductive proof that every automorphism of an integral lattice is a product of reflections. See Cassels for the latter. ♣

♣Another approach is to use the Steinberg theorem on generators of Chevalley groups using exponentials of roots. ♣

References

[1] M. Atiyah and G. Segal, "Twisted K-theory," math/0407054 [math-kt].

[2] A. Avila and S. Jitomirskaya, "The Ten Martini Problem," arXiv:math/0503363

[3] M. Ya. Azbel' , Sov. Phys. - JETP **19** 634 (1964)

[4] J. Bellissard, "Gap Labeling Theorems...." in *From Number Theory To Physics*, ...

[5] D. Bernard and A. Leclair, "Quantum group symmetries and nonlocal currents in 2-D QFT," Commun. Math. Phys. **142**, 99 (1991).

[6] T. H. Buscher, "Path Integral Derivation of Quantum Duality in Nonlinear Sigma Models," Phys. Lett. B **201**, 466 (1988). doi:10.1016/0370-2693(88)90602-8

[7] C. G. Callan, Jr., E. J. Martinec, M. J. Perry and D. Friedan, "Strings in Background Fields," Nucl. Phys. B **262**, 593 (1985). doi:10.1016/0550-3213(85)90506-1

[8] W.B. Case, "Wigner functions and Weyl transforms for pedestrians," Am. J. Phys. **76**, 2008, pp.937-946

[9] A. Connes, M. R. Douglas and A. S. Schwarz, "Noncommutative geometry and matrix theory: Compactification on tori," JHEP **9802**, 003 (1998) [hep-th/9711162].

[10] A. Connes, "C*-algebras and Differential Geometry," arXiv:hep-th/0101093

[11] A. Connes and M. Marcolli, "A Walk in the noncommutative garden," math/0601054 [math.QA].

[12] A. Connes, *Noncommutative Geometry*, Academic Press 1994

[13] C. Crnkovic and E. Witten, "Covariant Description Of Canonical Formalism In Geometrical Theories," In *Hawking, S.W. (ed.), Israel, W. (ed.): Three hundred years of gravitation*, 676-684 and Preprint - Crnkovic, C. (86,rec.Dec.) 13 p

[14] R. Dijkgraaf and E. P. Verlinde, "Modular Invariance and the Fusion Algebra," Nucl. Phys. Proc. Suppl. **5**, 87 (1988).

[15] M. R. Douglas and N. A. Nekrasov, "Noncommutative field theory," Rev. Mod. Phys. **73**, 977 (2001) [hep-th/0106048].

[16] G. Felder, J. Frohlich, J. Fuchs and C. Schweigert, "The Geometry of WZW branes," J. Geom. Phys. **34**, 162 (2000) [hep-th/9909030].

[17] D. S. Freed and G. W. Moore, "Twisted equivariant matter," Annales Henri Poincare **14**, 1927 (2013) [arXiv:1208.5055 [hep-th]].

[18] Michael H Freedman, Alexei Kitaev, Chetan Nayak, Johannes K Slingerland, Kevin Walker, Zhenghan Wang, "Universal manifold pairings and positivity," arXiv:math/0503054 [math.GT]

[19] D. Gepner and E. Witten, "String Theory on Group Manifolds," Nucl. Phys. B **278**, 493 (1986).

[20] M. Gerstenhaber and A. Giaquinto, "On the cohomology of the Weyl algebra, the quantum plane, and the q-Weyl algebra," J. Pure and Applied Algebra 218 (2014)879

[21] A. Giveon, M. Porrati and E. Rabinovici, "Target space duality in string theory," Phys. Rept. **244**, 77 (1994) doi:10.1016/0370-1573(94)90070-1 [hep-th/9401139].

[22] S. Girvin, "The Quantum Hall Effect: Novel Excitations and Broken Symmetries," arXiv:cond-mat/9907002 [cond-mat.mes-hall]

[23] M. B. Green, S. D. Miller and P. Vanhove, "Small representations, string instantons, and Fourier modes of Eisenstein series (with an appendix by D. Ciubotaru and P. Trapa)," arXiv:1111.2983 [hep-th].

[24] J. A. Harvey, G. W. Moore and C. Vafa, "Quasicrystalline Compactification," Nucl. Phys. B **304**, 269 (1988). doi:10.1016/0550-3213(88)90627-X

[25] D. R. Hofstadter, "Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields," Phys. Rev. B **14**, 2239 (1976).

[26] Y. Z. Huang and J. Lepowsky, "Tensor products of modules for a vertex operator algebra and vertex tensor categories," hep-th/9401119.

[27] Y. Z. Huang, "Vertex operator algebras, fusion rules and modular transformations," math/0502558 [math-qa].

[28] V. G. Knizhnik and A. B. Zamolodchikov, "Current Algebra and Wess-Zumino Model in Two-Dimensions," Nucl. Phys. B **247**, 83 (1984).

[29] Ch. Kreft and R. Seiler, "Models of the Hofstadter-type," Journal of Mathematical Physics **37** 5207 (1996)

[30] M. Lamoureux and J. Mingo, "On the Characteristic Polynomial of the Almost Mathieu Operator," Proc. Amer. Math. Soc, vol. 135 (2007), 3205–3215; arXiv:math/0606507 [math.OA]

[31] N.P. Landsman, "Lecture Notes on C*-Algebras, Hilbert C*-modules, and Quantum Mechanics," arXiv:math-ph/9807030

[32] A. Losev, G. W. Moore, N. Nekrasov and S. L. Shatashvili, "Central extensions of gauge groups revisited," hep-th/9511185.

[33] S. L. Lukyanov and A. B. Zamolodchikov, "Form-factors of soliton creating operators in the sine-Gordon model," Nucl. Phys. B **607**, 437 (2001) [hep-th/0102079].

[34] G. W. Moore and N. Seiberg, "Polynomial Equations for Rational Conformal Field Theories," Phys. Lett. B **212**, 451 (1988).

[35] G. W. Moore and N. Seiberg, "Classical and Quantum Conformal Field Theory," Commun. Math. Phys. **123**, 177 (1989).

[36] G. W. Moore, “Finite in all directions,” Yale Univ. New Haven - YCTP-P12-93 (93/05,rec.Jun.) 66 p. e: LANL hep-th/9305139 [hep-th/9305139].

[37] G. W. Moore and Y. Tachikawa, “On 2d TQFTs whose values are holomorphic symplectic varieties,” Proc. Symp. Pure Math. **85**, 191 (2012) [arXiv:1106.5698 [hep-th]].

[38] D. Mumford, *Tata Lectures on Theta I*, Birkhäuser, 1983

[39] D. Osadchy, J. Avron, “Hofstadter butterfly as Quantum phase diagram,” arXiv:math-ph/0101019

[40] R.E. Prange and S.M. Girvin, eds. *The Quantum Hall Effect*, second edition, Springer Verlag, p.28.

[41] K. Ranganathan, H. Sonoda and B. Zwiebach, “Connections on the state space over conformal field theories,” Nucl. Phys. B **414**, 405 (1994) doi:10.1016/0550-3213(94)90436-7 [hep-th/9304053].

[42] Reed and Simon, *Methods of Modern Mathematical Physics*, especially, vol. I.

[43] M.A. Rieffel, “C*-algebras associated with irrational rotations,” Pacific J. Math., 93 (1981) 415429.

[44] M. Rocek and E. P. Verlinde, “Duality, quotients, and currents,” Nucl. Phys. B **373**, 630 (1992) doi:10.1016/0550-3213(92)90269-H [hep-th/9110053].

[45] D. Roggenkamp and K. Wendland, “Decoding the geometry of conformal field theories,” Bulg. J. Phys. **35**, 139 (2008) [arXiv:0803.0657 [hep-th]].

[46] V. Schomerus, “D-branes and deformation quantization,” JHEP **9906**, 030 (1999) [hep-th/9903205].

[47] N. Seiberg and E. Witten, “String theory and noncommutative geometry,” JHEP **9909**, 032 (1999) [hep-th/9908142].

[48] D. J. Thouless, M. Kohmoto, M. P. Nightingale and M. den Nijs, “Quantized Hall Conductance in a Two-Dimensional Periodic Potential,” Phys. Rev. Lett. **49**, 405 (1982).

[49] E. P. Verlinde, “Fusion Rules and Modular Transformations in 2D Conformal Field Theory,” Nucl. Phys. B **300**, 360 (1988).

[50] J. Wess and B. Zumino, “Consequences of anomalous Ward identities,” Phys. Lett. B **37**, 95 (1971).

[51] E. Witten, “Nonabelian Bosonization in Two-Dimensions,” Commun. Math. Phys. **92**, 455 (1984).

[52] E. Witten, “Global Aspects of Current Algebra,” Nucl. Phys. B **223**, 422 (1983).