Preprint typeset in JHEP style - HYPER VERSION

Chapter 2: Linear Algebra User's Manual

Gregory W. Moore

Abstract: An overview of some of the finer points of linear algebra usually omitted in physics courses. April 27, 2018

-TOC-

Contents

---

1. Introduction

Linear algebra is of course very important in many areas of physics. Among them:

1. Tensor analysis - used in classical mechanics and general relativity.
2. The very formulation of quantum mechanics is based on linear algebra: The states in a physical system are described by “rays” in a projective Hilbert space, and physical observables are identified with Hermitian linear operators on Hilbert space.
3. The realization of symmetry in quantum mechanics is through representation theory of groups which relies heavily on linear algebra.

For this reason linear algebra is often taught in physics courses. The problem is that it is often mis-taught. Therefore we are going to make a quick review of basic notions stressing some points not usually emphasized in physics courses.

We also want to review the basic canonical forms into which various types matrices can be put. These are very useful when discussing various aspects of matrix groups.

For more information useful references are Herstein, Jacobsen, Lang, Eisenbud, *Commutative Algebra* Springer GTM 150, Atiyah and MacDonald, *Introduction to Commutative Algebra.* For an excellent terse summary of homological algebra consult S.I. Gelfand and Yu. I. Manin, *Homological Algebra.*

We will only touch briefly on some aspects of functional analysis - which is crucial to quantum mechanics. The standard reference for physicists is:

Reed and Simon, *Methods of Modern Mathematical Physics*, especially, vol. I.

2. Basic Definitions Of Algebraic Structures: Rings, Fields, Modules, Vector Spaces, And Algebras

2.1 Rings

In the previous chapter we talked about groups. We now overlay some extra structure on an abelian group R, with operation $+$ and identity 0, to define what is called a *ring.* The new structure is a *second* binary operation $(a, b) \to a \cdot b \in R$ on elements $a, b \in R$. We demand that this operation be associative, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, and that it is compatible with the pre-existing additive group law. To be precise, the two operations $+$ and $\cdot$ are compatible in the sense that there is a distributive law:

$$a \cdot (b + c) = a \cdot b + a \cdot c \tag{2.1}$$

$$(a + b) \cdot c = a \cdot c + b \cdot c \tag{2.2}$$

Remarks

1. A ring with a multiplicative unit 1_R such that $a \cdot 1_R = 1_R \cdot a = a$ is called a *unital ring* or a *ring with unit.* One needs to be careful about this because many authors will simply assume that "ring" means a "ring with unit." [1]

2. If $a \cdot b = b \cdot a$ then R is a *commutative ring.*

3. If R is any ring we can then form another ring, $M_n(R)$, the ring of $n \times n$ matrices with matrix elements in R. Even if R is a commutative ring, the ring $M_n(R)$ will be noncommutative in general if $n > 1$.

2.2 Fields

Definition: A commutative ring R such that $R^* = R - \{0\}$ is also an abelian group with respect to $\cdot$ is called a *field.*

Some important examples of rings which are not fields are

1. $\mathbb{Z}$.

[1] Some authors use the term "rng" – pronounced "rung" - for a ring possibly without a unit. We will not do that. Similarly, one can define a notion called a "rig" - which is a ring without negatives. That is, it is an abelian monoid with the operation $+$ and a compatible multiplication $\cdot$.

2. $\mathbb{Z}/N\mathbb{Z}$, when N is not prime.

3. If R is any ring then we can form the ring of polynomials with coefficients in R, denoted $R[x]$. Iterating this we obtain polynomial rings in several variables $R[x_1, \dots, x_n]$. Similarly, we can consider a ring of power series in x.

4. Let U be an open subset of the complex plane (or of $\mathbb{C}^n$) then we can consider the ring $\mathcal{O}(U)$ of holomorphic functions on U.

Some important examples of fields are

1. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$

2. $\mathbb{Z}/N\mathbb{Z}$, when $N = p$ is prime.

3. In fact, for every prime power $q = p^k$ there is a (unique, up to isomorphism) field $\mathbb{F}_q$ of q elements which is an extension of $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. This field is *not* to be confused with the ring $\mathbb{Z}/p^k\mathbb{Z}$. For example $R = \mathbb{Z}/4\mathbb{Z}$ is not a field with the usual ring multiplication. For example $2 \in R^*$ and $2 \cdot 2 = 0 \mathrm{mod} 4$. One way to represent the field $\mathbb{F}_4$ is as a set ♣Check uniqueness. ♣

$$\mathbb{F}_4 = \{0, 1, \omega, \bar{\omega}\} \tag{2.3}$$

with the relations

$$\begin{aligned} 0 + x &= x & \forall x \in \mathbb{F}_4 \\ x + x &= 0 & \forall x \in \mathbb{F}_4 \\ 1 + \omega &= \bar{\omega} \\ 1 + \bar{\omega} &= \omega \\ \omega + \bar{\omega} &= 1 \\ 0 \cdot x &= 0 & \forall x \in \mathbb{F}_4 \\ 1 \cdot x &= x & \forall x \in \mathbb{F}_4 \\ \omega \cdot \bar{\omega} &= 1 \\ \omega \cdot \omega &= \bar{\omega} \\ \bar{\omega} \cdot \bar{\omega} &= \omega \end{aligned} \tag{2.4}$$

Note that although $\omega^3 = \bar{\omega}^3 = 1$ you cannot identity ω with a complex number the third root of unity: This field is not a subfield of $\mathbb{C}$.

4. For the ring of polynomials $R[x]$ we can consider the associated field of fractions $p(x)/q(x)$ where $q(x)$ is nonzero. This is an example of "localization." ♣Check. R has to be a PID? ♣

5. The field of meromorphic functions on a complex variety.

2.3 Modules

Definition A *module* over a ring R is a set M with a multiplication $R \times M \to M$ such that for all $r, s \in R$ and $v, w \in M$:

1. M is an abelian group wrt $+$, called "vector addition."
2. $r(v + w) = rv + rw$
3. $r(sv) = (rs)v$
4. $(r + s)v = rv + sv$

Axioms 2,3,4 simply say that all the various operations on R and M are compatible in the natural way.

Remarks:

1. If the ring has a multiplicative unit 1_R then we require $1_R \cdot v = v$.
2. If the ring is noncommutative then one should distinguish between left and right modules. Above we have written the axioms for a left-module. For a right-module we have $(v \cdot r) \cdot s = v \cdot (rs)$.
3. There is an important generalization known as a *bimodule* over two rings R_1, R_2. A bimodule M is simultaneously a left R_1-module and a right R_2-module. A good example is the set of $n \times m$ matrices over a ring R, which is a bimodule over $R_1 = M_n(R)$ and $R_2 = M_m(R)$.
4. Any ring is a bimodule over itself. For a positive integer n we define the module R^n of n-tuples of elements of R by componentwise addition and multiplication in the obvious way. This is an R-bimodule and also an $M_n(R)$ bimodule.
5. In quantum field theory if we divide up a spatial domain into two parts with a codimension one subspace then the states localized near the division is a bimodule for operators localized on the left and the right of the partition.

Examples:

1. Any Abelian group is a $\mathbb{Z}$ module, and any $\mathbb{Z}$-module is just an Abelian group.
2. Meromorphic functions with a pole at some point z_0 in the complex plane with order $\leq n$. These form an important example of a module over the ring of holomorphic functions.

2.4 Vector Spaces

Recall that a *field* is a commutative ring such that $R^* = R - \{0\}$ is also an abelian group.

Let κ be a field. Then, by definition a *vector space over* κ is simply a κ-module. Written out in full, this means:

♣In the rest of the notes we need to change notation for a general field from k to κ since k is often also used as an integer or a momentum. ♣

Definition . V is a vector space over a field κ if for every $\alpha \in \kappa, v \in V$ there is an element $\alpha v \in V$ such that

1. V is an abelian group under $+$
2. $\alpha(v + w) = \alpha v + \alpha w$
3. $\alpha(\beta v) = (\alpha\beta)v$
4. $(\alpha + \beta)v = \alpha v + \beta v$
5. $1v = v$

for all $\alpha, \beta \in \kappa, v, w \in V$

For us, the field κ will almost always be $\kappa = \mathbb{R}$ or $\kappa = \mathbb{C}$. In addition to the well-worn examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ two other examples are

1. Recall example 2.9 of Chapter 1: If X is any set then the power set $\mathcal{P}(X)$ is an Abelian group with $Y_1 + Y_2 := (Y_1 - Y_2) \cup (Y_2 - Y_1)$. As we noted there, $2Y = \emptyset$. So, $\mathcal{P}(X)$ is actually a vector space over the field $\mathbb{F}_2$.

2.5 Algebras

♣Take material on ideals below and move it here. ♣

So far we have taken abelian groups and added binary operations $R \times R \to R$ to define a ring and $R \times M \to M$ to define a module. It remains to consider the case $M \times M \to M$. In this case, the module is known as an *algebra*.

Everything that follows can also be defined for modules over a ring but we will state the definitions for a vector space over a field.

♣Rewrite, and do everything for modules over a ring?? ♣

Definition An *algebra* over a field κ is a vector space A over κ with a notion of multiplication of two vectors

$$A \times A \to A \tag{2.5}$$

denoted:

$$a_1, a_2 \in A \to a_1 \odot a_2 \in A \tag{2.6}$$

which has a ring structure compatible with the scalar multiplication by the field. Concretely, this means we have axioms:

i.) $(a_1 + a_2) \odot a_3 = a_1 \odot a_3 + a_2 \odot a_3$

ii.) $a_1 \odot (a_2 + a_3) = a_1 \odot a_2 + a_1 \odot a_3$

iii.) $\alpha(a_1 \odot a_2) = (\alpha a_1) \odot a_2 = a_1 \odot (\alpha a_2), \qquad \forall \alpha \in \kappa.$

The algebra is *unital*, i.e., it has a unit, if $\exists 1_A \in A$ (*not* to be confused with the multiplicative unit $1 \in \kappa$ of the ground field) such that:

iv.) $1_A \odot a = a \odot 1_A = a$

In the case of rings we assumed associativity of the product. It turns out that this is too restrictive when working with algebras. If, in addition, the product of vectors satisfies:

$$(a_1 \odot a_2) \odot a_3 = a_1 \odot (a_2 \odot a_3) \tag{2.7}$$

for all $a_1, a_2, a_3 \in A$ then A is called an *associative algebra*.

Remark: We have used the heavy notation $\odot$ to denote the product of vectors in an algebra to stress that it is a new structure imposed on a vector space. But when working with algebras people will generally just write $a_1 a_2$ for the product. One should be careful here as it can (and will) happen that a given vector space can admit more than one interesting algebra product structure.

Example 1 $M_n(\kappa)$ is a vector space over κ of dimension n^2. It is also an associative algebra because matrix multiplication defines an algebraic structure of multiplication of the "vectors" in $M_n(\kappa)$.

Example 2 More generally, if A is a vector space over κ then $\mathrm{End}(A)$ is an associative algebra. (See next section for the definition of this notation.)

In general, a *nonassociative algebra* means a not-necessarily associative algebra. In any algebra we can introduce the *associator*

$$[a_1, a_2, a_3] := (a_1 \cdot a_2) \cdot a_3 - a_1 \cdot (a_2 \cdot a_3) \tag{2.8}$$

`eq:associator`

Note that it is trilinear. There are important examples of non-associative algebras such as Lie algebras and the octonions.

Definition A *Lie algebra* over a field κ is an algebra A over κ where the multiplication of vectors $a_1, a_2 \in A$, satisfies in addition the two conditions:

1. $\forall a_1, a_2 \in A$:

$$a_2 \odot a_1 = -a_1 \odot a_2 \tag{2.9}$$

`eq:LieDef-1`

2. $\forall a_1, a_2, a_3 \in A$:

$$((a_1 \odot a_2) \odot a_3) + ((a_3 \odot a_1) \odot a_2) + ((a_2 \odot a_3) \odot a_1) = 0 \tag{2.10}$$

`eq:LieDef-2`

This is known as the *Jacobi relation*.

Now, tradition demands that the product on a Lie algebra be denoted not as $a_1 \odot a_2$ but rather as $[a_1, a_2]$ where it is usually referred to as the *bracket*. So then the two defining conditions (2.9) and (2.10) are written as:

1. $\forall a_1, a_2 \in A$:
$$[a_2, a_1] = -[a_1, a_2] \tag{2.11}$$

2. $\forall a_1, a_2, a_3 \in A$:
$$[[a_1, a_2], a_3] + [[a_3, a_1], a_2] + [[a_2, a_3], a_1] = 0 \tag{2.12}$$

Remarks:

1. Note that we call $[a_1, a_2]$ the bracket and not the commutator. It might well not be possible to write $[a_1, a_2] = a_1 \odot a_2 - a_2 \odot a_1$ where $\odot$ is some other multiplication structure defined within A. Rather $[\cdot, \cdot] : A \times A \to A$ is just an abstract product satisfying the two rules (2.9) and (2.10). Let us give two examples to illustrate the point:

   - Note that the vector space $A \subset M_n(\kappa)$ of anti-symmetric matrices is not closed under normal matrix multiplication: If a_1 and a_2 are antisymmetric matrices then $(a_1 a_2)^{tr} = a_2^{tr} a_1^{tr} = (-a_2)(-a_1) = a_2 a_1$ and in general this is not $-a_1 a_2$. So, it is not an algebra under normal matrix multiplication. But if we define the bracket using normal matrix multiplication
$$[a_1, a_2] = a_1 a_2 - a_2 a_1 \tag{2.13}$$
where on the RHS $a_1 a_2$ means matrix multiplication. Since $[a_1, a_2]$ is an anti-symmetric matrix the product is closed within A. The Jacobi relation is then inherited from the associativity of matrix multiplication. This Lie algebra is sometimes denoted $\mathfrak{o}(n, \kappa)$. It is the Lie algebra of an orthogonal group.
   - Consider first order differential operators of $\mathcal{C}^\infty$ functions on the line. These will be written as $D = f(x)\frac{d}{dx} + g(x)$ for smooth functions $f(x), g(x)$. Then the ordinary composition of two differential operators $D_1 \circ D_2$ is a second order differential operator. Nevertheless if we take the difference:
$$\begin{aligned} [D_1, D_2] &= (f_1(x)\frac{d}{dx} + g_1(x))(f_2(x)\frac{d}{dx} + g_2(x)) - (f_2(x)\frac{d}{dx} + g_2(x))(f_1(x)\frac{d}{dx} + g_1(x)) \\ &= \left(f_1 f_2' - f_2 f_1'\right)(x)\frac{d}{dx} + (f_1 g_2' - f_2 g_1')(x) \end{aligned} \tag{2.14}$$
we get a first order differential operator. It is obviously anti-symmetric and one can check the Jacobi relation.

- In both these examples we embed the Lie algebra into a larger associative algebra where the bracket can be written as $[a_1, a_2] = a_1 \odot a_2 - a_2 \odot a_1$ with $\odot$ the algebra product that only closes within the larger algebra.

2. A Lie algebra is in general a nonassociative algebra. Indeed, using the Jacobi relation we can compute the associator as:

$$[a_1, a_2, a_3] = [[a_1, a_2], a_3] - [a_1, [a_2, a_3]] = [[a_1, a_2], a_3] + [[a_2, a_3], a_1] = -[[a_3, a_1], a_2] \tag{2.15}$$

and the RHS is, in general, nonzero.

3. Note that the vector space of $n \times n$ matrices over κ, that is, $M_n(\kappa)$ has two interesting algebra structures: One is matrix multiplication. It is associative. The other is a Lie algebra structure where the bracket is defined by the usual commutator. It is nonassociative. It is sometimes denoted $\mathfrak{gl}(n, \kappa)$, and such a notation would definitely imply a Lie algebra structure.

Exercise *Opposite Algebra*
If A is an algebra we can always define another algebra A^{opp} with the product

$$a_1 \odot^{\text{opp}} a_2 := a_2 \odot a_1 \tag{2.16}$$

a.) Show that $\odot^{\text{opp}}$ indeed defines the structure of an algebra on the set A.
b.) Consider the algebra $M_n(\kappa)$ where κ is a field. Is it isomorphic to its opposite algebra?
c.) Give an example of an algebra not isomorphic to its opposite algebra. ♣Need to provide an answer here. ♣

Exercise *Structure constants*
In general, if $\{v_i\}$ is a basis for the algebra then the *structure constants* are defined by

$$v_i \cdot v_j = \sum_k c_{ij}^k v_k \tag{2.17}$$

a.) Write out a basis and structure constants for the algebra $M_n(k)$.

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

$$L(a) \cdot b := ab \tag{2.18}$$

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

$$a_i a_j = c_{ij}{}^k a_k \tag{2.19}$$

define the matrix elements of the LRR:

$$(L(a_i))_j{}^k = c_{ij}^k \tag{2.20}$$

An algebra is said to be *semisimple* if these operators are diagonalizable.

b.) If A is an algebra, then it is a bimodule over $A \otimes A^o$ where A^o is the opposite algebra.

---

3. Linear Transformations

Definition

a.) A *linear transformation* or *linear operator* between two R modules is a map $T : M_1 \to M_2$ which is a group homomorphism with respect to $+$:

$$T(m + m') = T(m) + T(m') \tag{3.1}$$

and moreover such that $T(r \cdot m) = r \cdot T(m)$ for all $r \in R$, $m \in M_1$.

b.) T is an *isomorphism* if it is one-one and onto.

c.) The set of all linear transformations $T : M_1 \to M_2$ is denoted $\mathrm{Hom}(M_1, M_2)$, or $\mathrm{Hom}_R(M_1, M_2)$ when we wish to emphasize the underlying ring R.

There are some algebraic structures on spaces of linear transformations we should immediately take note of:

1. $\mathrm{Hom}_R(M_1, M_2)$ is an abelian group where the group operation is addition of linear operators: $T_1 + T_2$.

2. Moreover $\mathrm{Hom}_R(M_1, M_2)$ is an R-module provided that R is a commutative ring.

3. In particular, if V_1, V_2 are vector spaces over a field k then $\mathrm{Hom}(V_1, V_2)$ is itself a vector space over k.

4. If M is a module over a ring R then sometimes the notation

$$\mathrm{End}_R(M) := \mathrm{Hom}_R(M, M) \tag{3.2}$$

is used. In this case composition of linear transformations $T_1 \circ T_2$ defines a binary operation on $\mathrm{End}_R(M)$, and if R is commutative this is itself a ring because

$$T_1 \circ (T_2 + T_3) = T_1 \circ T_2 + T_1 \circ T_3 \tag{3.3}$$

and so forth.

5. In general if M is a module over a commutative ring R then $\mathrm{End}_R(M)$ is not a group wrt $\circ$, since inverses don't always exist. However we may define:

   Definition The set of *invertible* linear transformations of M, denoted $GL(M, R)$, is a group. If we have a vector space over a field k we generally write $GL(V)$.

   Example: For $R = \mathbb{Z}$ and $M = \mathbb{Z} \oplus \mathbb{Z}$, the group of invertible transformations is isomorphic to $GL(2, \mathbb{Z})$. [2]

A *representation* of an algebra A is a vector space V and a morphism of algebras $T : A \to \mathrm{End}(V)$. This means that

$$\begin{aligned} T(\alpha_1 a_1 + \alpha_2 a_2) &= \alpha_1 T(a_1) + \alpha_2 T(a_2) \\ T(a_1 \odot a_2) &= T(a_1) \odot T(a_2) \end{aligned} \tag{3.4}$$ eq:repalgebra

Remarks

1. We must be careful here about the algebra product $\odot$ being used since, as noted above, there are two interesting algebra structures on $\mathrm{End}(V)$ given by composition and by commutator. If we speak of a morphism of algebras what is usually meant by $\odot$ on the RHS of (3.4) is composition of linear transformations. However, if we are speaking of a representation of Lie algebras then we mean the commutator. So, for a Lie algebra a representation would satisfy

$$T([a_1, a_2]) = T(a_1) \circ T(a_2) - T(a_2) \circ T(a_1) \tag{3.5}$$

2. If we consider the algebra $M_n(\kappa)$ with matrix multiplication as the algebra product then a theorem states that the general representation is a direct sum (See Section **** below) of the fundamental, or defining representation $V_{fund} = \kappa^{\oplus n}$. That is, the general representation is

$$V_{fund} \oplus \cdots \oplus V_{fund} \tag{3.6}$$

   If we have m summands then $T(a)$ would be a block diagonal matrix with a on the diagonal m times. This leads to a concept called "Morita equivalence" of algebras: Technically, two algebras A, B are "Morita equivalent" if their categories of representations are equivalent categories. In practical terms often it just means that $A = M_n(B)$ or vice versa.

3. On the other hand, if we consider $M_n(\kappa)$ as a Lie algebra then the representation theory is much richer, and will be discussed in Chapter **** below.

---

[2] Warning: This is NOT the same as 2×2 matrices over $\mathbb{Z}$ with nonzero determinant!

Exercise

Let R be any ring. Show that if $\mathcal{M}$ is an $R[x]$-module then we can associate to it an R-module M together with a linear transformation $T : M \to M$.

b.) Conversely, show that if we are given an R-module M together with a linear transformation T then we can construct uniquely an $R[x]$ module $\mathcal{M}$.

Thus, $R[x]$-modules are in one-one correspondence with pair (M, T) where M is an R-module and $T \in \mathrm{End}_R(M)$.

---

4. Basis And Dimension

4.1 Linear Independence

Definition . Let M be a module over a ring R.

1. If $S \subset M$ is any subset of M the *linear span of* S is the set of finite linear combinations of vectors drawn from S:

$$L(S) := Span(S) := \{\sum r_i v_i : r_i \in R, v_i \in S\} \tag{4.1}$$ eq:linespan

$L(S)$ is the smallest submodule of M containing S. We also call S a *generating set* of $L(S)$.

2. A set of vectors $S \subset M$ is said to be linearly independent if for any finite sum of vectors in $L(S)$:

$$\sum_s \alpha_s v_s = 0 \qquad \Rightarrow \qquad \alpha_s = 0 \tag{4.2}$$

3. A linearly independent generating set S for a module M is called a *basis for* M. We will often denote a basis by a symbol like $\mathcal{B}$.

Remarks:

1. A basis $\mathcal{B}$ need not be a finite set. However, all sums above are finite sums. In particular, when we say that $\mathcal{B}$ generates V this means that every vector $m \in M$ can be written (uniquely) as a finite linear combination of vectors in $\mathcal{B}$.

2. To appreciate the need for restriction to a finite sums in the definitions above consider the vector space $\mathbb{R}^\infty$ of infinite tuples of real numbers $(x_1, x_2, \dots)$. (Equivalently, the vector space of all functions $f : \mathbb{Z}_+ \to \mathbb{R}$.) Infinite sums like

$$(1,1,1,\dots) - (2,2,2,\dots) + (3,3,3,\dots) - (4,4,4,\dots) + \cdots \tag{4.3}$$

are clearly ill-defined.

3. For a finite set $S = \{v_1, \dots, v_n\}$ we will also write

$$L(S) := \langle v_1, \dots, v_n \rangle \tag{4.4}$$

4.2 Free Modules

A module is called a *free module* if it has a basis. If the basis is finite the free module is isomorphic to R^n for some positive integer n.

Not all modules are free modules, e.g.

1. $\mathbb{Z}/n\mathbb{Z}$ is not a free $\mathbb{Z}$-module. Exercise: Explain why.

2. Fix a set of points $\{z_1, \dots, z_k\}$ in the complex plane and a set of integers $n_i \in \mathbb{Z}$ associated with those points. The set of holomorphic on $\mathbb{C} - \{z_1, \dots, z_k\}$ which have convergent Laurent expansions of the form

$$f(z) = \frac{a^i_{-n_i}}{(z - z_i)^{n_i}} + \frac{a^i_{-(n_i-1)}}{(z - z_i)^{n_i-1}} + \cdots \tag{4.5}$$

in the neighborhood of $z = z_i$, for all $i = 1, \dots, k$ is a module over the ring of holomorphic functions, but it is not a free module.

4.3 Vector Spaces

isVectorSpaces

One big simplification when working with vector spaces rather than modules is that they are always free modules. We should stress that this is <u>not</u> obvious! The statement is false for general modules over a ring, as we have seen above, and the proof requires the use of Zorn's lemma (which is equivalent to the axiom of choice).

Theorem 4.3.1:

a.) Every nonzero vector space V has a basis.

b.) Given any linearly independent set of vectors $S \subset V$ there is a basis $\mathcal{B}$ for V with $S \subset \mathcal{B}$.

Proof: Consider the collection $\mathcal{L}$ of linearly independent subsets $S \subset V$. If $V \neq 0$ then this collection is nonempty. Moreover, for every ascending chain of elements in $\mathcal{L}$:

$$S_1 \subset S_2 \subset \cdots \tag{4.6}$$

the union $\cup_i S_i$ is a set of linearly independent vectors and is hence in $\mathcal{L}$. We can then invoke Zorn's lemma to assert that there exists a maximal element $\mathcal{B} \subset \mathcal{L}$. That is, it is a linearly independent set of vectors not properly contained in any other element of $\mathcal{L}$.

We claim that $\mathcal{B}$ is a basis. To see this, consider the linear span $L(\mathcal{B}) \subset V$. If $L(\mathcal{B})$ is a proper subset of V there is a vector $v_* \in V - L(\mathcal{B})$. But then we claim that $\mathcal{B} \cup \{v_*\}$ is a linearly independent set of vectors. The reason is that if

$$\alpha_* v_* + \sum_{w \in \mathcal{B}} \beta_w w = 0 \tag{4.7}$$

(remember: all but finitely many $\beta_w = 0$ here) then if $\alpha_* = 0$ we must have $\beta_w = 0$ because $\mathcal{B}$ is a linearly independent set. But if $\alpha_* \neq 0$ then we can divide by it. (It is exactly at this point that we use the fact that we are working with a vector space over a field κ rather than a general modular over a ring R!!) Then we would have

$$v_* = -\sum_{w \in \mathcal{B}} \frac{\beta_w}{\alpha_*} w \tag{4.8}$$

but this contradicts the hypothesis that $v_* \notin L(\mathcal{B})$. Thus we conclude that $L(\mathcal{B}) = V$ and hence $\mathcal{B}$ is a basis.

To prove part (b) apply Zorn's lemma to the set of linearly independent sets containing a fixed linearly independent set S. ♠

Theorem 4.3.2: Let V be a vector space over a field κ. Then any two bases for V have the same cardinality.
Proof: See Lang, *Algebra*, ch. 3 Sec. 5. Again the proof explicitly uses the fact that you can divide by nonzero scalars.

By this theorem we know that if V has a finite basis $\{v_1, \ldots, v_n\}$ then any other basis has n elements. (The basic idea is to observe that for any linearly independent set of m elements we must have $m \leq n$, so two bases must have the same cardinality.)

We call this basis-invariant integer n the *dimension of* V:

$$n := \dim_\kappa V \tag{4.9}$$

the dimension of V. If there is no finite basis then V is infinite-dimensional.

Remarks

1. Note well that the notion of dimension refers to the ground field. If $\kappa_1 \subset \kappa_2$ then the notion of dimension over κ_1 and κ_2 will be different. For example, any vector space over $\kappa = \mathbb{C}$ is, *a fortiori* also a vector space over $\kappa = \mathbb{R}$. Let us call it $V_{\mathbb{R}}$. It is the same set, but now the vector space structure on this Abelian group is just defined by the action of real scalars. Then we will see that:

$$\dim_{\mathbb{R}} V = 2\dim_{\mathbb{C}} V \tag{4.10}$$

   We will come back to this important point in Section 9.

2. Any two finite dimensional vector spaces of the same dimension are isomorphic. However, it is in general not true that two infinite-dimensional vector spaces are isomorphic. However the above theorem states that if they have bases $\{v_i\}_{i \in I}$ and $\{w_\alpha\}_{\alpha \in I'}$ with a one-one map $I \to I'$ then they are isomorphic.

3. The only invariant of a finite dimensional vector space is its dimension. One way to say this is the following: Let **VECT** be the category of finite-dimensional vector spaces and linear transformations. Define another category **vect** whose objects are

the nonnegative integers $n = 0, 1, 2, \ldots$ and whose morphisms $\hom(n, m)$ are $m \times n$ matrices, with composition of morphisms given by matrix multiplication. (If n or m is zero there is a unique morphism with the properties of a zero matrix.) We claim that **VECT** and **vect** are equivalent categories. It is a good exercise to prove this.

4. Something which can be stated or proved without reference to a particular basis is often referred to as *natural* or *canonical* in mathematics. (We will use these terms interchangeably.) More generally, these terms imply that a mathematical construction does not make use of any extraneous information. Often in linear algebra, making a choice of basis is just such an extraneous piece of data. One of the cultural differences between physicists and mathematicians is that mathematicians often avoid making choices and strive for naturality. This can be a very good thing as it oftentimes happens that expressing a construction in a basis-dependent fashion obscures the underlying conceptual simplicity. On the other hand, insisting on not using a basis can sometimes lead to obscurity. We will try to strike a balance.

5. One of the many good reasons to insist on natural constructions is that these will work well when we consider continuous families of vectors spaces (that is, when we consider vector bundles). Statements which are basis-dependent will tend not to have analogs for vector bundles, whereas natural constructions easily generalize to vector bundles.

   For those to whom "vector bundle" is a new concept a good, nontrivial, and ubiquitous example is the following: [3] Consider the family of projection operators $P_\pm(\hat{x}) : \mathbb{C}^2 \to \mathbb{C}^2$ labeled by a point $\hat{x}$ in the unit sphere in three dimensions: $\hat{x} \in S^2 \subset \mathbb{R}^3$. We take them to be

   $$P_\pm(\hat{x}) = \frac{1}{2}(1 \pm \hat{x} \cdot \vec{\sigma}) \tag{4.11}$$

   The images $L_{\pm,\hat{x}}$ of $P_\pm(\hat{x})$ are one-dimensional subspaces of $\mathbb{C}^2$. So, explicitly:

   $$L_{\pm,\hat{x}} := \{P_\pm v | v \in \mathbb{C}^2\} \tag{4.12}$$ eq:SpinLines

   For those who know about spin, if we think of $\mathbb{C}^2$ as a q-bit consisting of a spin-half particle then $L_{\pm,\hat{x}}$ is the line in which the particle spins along $\hat{x}$ (for the + case) and along $-\hat{x}$ (for the − case).

   This is a good example of a "family of vector spaces." More generally, if we have a family of projection operators $P(s)$ acting on some fixed vector space V and depending on some control parameters s valued in some manifold then we have a family of vector spaces

   $$E_s = P(s)[V] = \mathrm{Im}(P(s)) \tag{4.13}$$

   parametrized by that manifold. If the family of projection operators depends "continuously" on s, (note that you need a topology on the space of projectors to make mathematical sense of that) then our family of vector spaces is a *vector bundle*.

[3] Some terms such as "projection operator" are only described below, so the reader might wish to return to this - important! - remark later.

Returning to (4.12), since they are one-dimensional subspaces we can certainly say that, for every $\hat{x} \in S^2$ there are isomorphisms

$$\psi_{\pm}(\hat{x}) : L_{\pm,\hat{x}} \to \mathbb{C} \tag{4.14}$$

However, methods of topology can be used to prove rigorously that there is no continuous family of such isomorphisms. Morally speaking, if there had been a natural family of isomorphisms one would have expected it to be continuous.

6. When V is infinite-dimensional there are different notions of what is meant by a "basis." The notion we have defined above is known as a *Hamel basis.* In a Hamel basis we define "generating" and "linear independence" using finite sums. In a vector space there is no *a priori* notion of infinite sums of vectors, because defining such infinite sums requires a notion of convergence. If V has more structure, for example, if it is a Banach space, or a Hilbert space (see below) then there is a notion of convergence and we can speak of other notions of basis where we allow convergent infinite sums. These include *Schauder basis, Haar basis, ...* In the most important case of a Hilbert space, an *orthonormal basis* is a maximal orthonormal set. Every Hilbert space has an orthonormal basis, and one can write every vector in Hilbert space as an infinite (convergent!) linear combination of orthonormal vectors. See

   K. E. Smith, http://www.math.lsa.umich.edu/∼kesmith/infinite.pdf

   for a nice discussion of the issues involved.

4.4 Linear Operators And Matrices

Let V, W be finite dimensional vector spaces over k. Given a linear operator $T : V \to W$ and ordered bases $\{v_1, \dots, v_m\}$ for V and $\{w_1, \dots w_n\}$ for W, we may associate a matrix $M \in Mat_{n\times m}(k)$ to T:

$$Tv_i = \sum_{s=1}^{n} M_{si} w_s \qquad\qquad i = 1, \dots, m \tag{4.15}$$ eq:mtrx

Note! A matrix depends on a *choice* of basis. The same linear transformation can look very different in different bases. A particularly interesting example is,

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \sim \begin{pmatrix} x & y \\ z & -x \end{pmatrix} \tag{4.16}$$

whenever $x^2 + yz = 0$ and $(x, y, z) \neq 0$.

In general if we change bases

$$\begin{aligned} \tilde{w}_s &= \sum_{t=1}^{n} (g_2)_{ts} w_t \\ \tilde{v}_j &= \sum_{j=1}^{m} (g_1)_{ji} v_j \end{aligned} \tag{4.17}$$

then with respect to the new bases the same linear transformation is expressed by the new matrix:

$$\tilde{M} = g_2^{-1} M g_1. \tag{4.18}$$

With the choice of indices in (4.15) *composition* of linear transformations corresponds to *matrix multiplication.* If $T_1 : V_1 \to V_2$ and $T_2 : V_2 \to V_3$ and we choose ordered bases

$$\begin{aligned} &\{v_i\}_{i=1,\dots,d_1} \\ &\{w_s\}_{s=1,\dots,d_2} \\ &\{u_x\}_{x=1,\dots,d_3} \end{aligned} \tag{4.19}$$

then

$$(T_2 \circ T_1) v_i = \sum_{x=1}^{d_3} (M_2 M_1)_{xi} u_x \qquad\qquad i = 1, \dots, d_1 \tag{4.20}$$

where

$$(M_2 M_1)_{xi} = \sum_{s=1}^{d_2} (M_2)_{xs} (M_1)_{si} \tag{4.21}$$

Remarks

1. Left- vs. Right conventions: One could compose linear transformations as $T_1 T_2$ and then all the indices would be transposed... ♣explain more clearly ♣

2. *Change of coordinates and active vs. passive transformations.* Given a *choice* of basis $\{v_i\}$ of an n-dimensional vector space V there is a canonically determined system of *coordinates* on V defined by

$$v = \sum_i v_i x^i = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{4.22}$$

   If we apply a linear transformation $T : V \to V$ then we can think of the transformation in two ways:

   A.) We can say that when we move the vector $v \mapsto T(v)$ the coordinates of the vector change according to

$$T(v) = \sum_i T(v_i) x^i = \sum_i v_i \tilde{x}^i \tag{4.23}$$

   with

$$\begin{pmatrix} \tilde{x}^1 \\ \vdots \\ \tilde{x}^n \end{pmatrix} = \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \dots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{4.24}$$

B.) On the other hand, we could say that the transformation defines a new basis $\tilde{v}_i = T(v_i)$ with

$$\begin{pmatrix} \tilde{v}_1 & \cdots & \tilde{v}_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \cdots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{pmatrix} \tag{4.25}$$

This leads to the *passive* viewpoint: We could describe linear transformations by saying that they are just changing basis. In this description the vector does not change, but only its coordinate description changes. In the passive view the new coordinates of the *same* vector are gotten from the old by

$$v = \sum_i \tilde{v}_i y^i = \sum_i v_i x^i \tag{4.26}$$

and hence the change of coordinates is:

$$\vec{y} = M^{-1}\vec{x} \tag{4.27}$$

Exercise

Let V_1 and V_2 be finite dimensional vector spaces over κ. Show that

$$\dim_\kappa \mathrm{Hom}(V_1, V_2) = (\dim_\kappa V_1)(\dim_\kappa V_2) \tag{4.28}$$

Exercise

If $n = \dim V < \infty$ then $GL(V)$ is isomorphic to the group $GL(n, \kappa)$ of invertible $n \times n$ matrices over κ, but it is not canonically isomorphic.

Exercise

A one-dimensional vector space L (also sometimes referred to as a *line*) is isomorphic to, but not canonically isomorphic to, the one-dimensional vector space κ.

Show that, nevertheless, the one-dimensional vector space $\mathrm{Hom}(L, L)$ is indeed <u>canonically</u> isomorphic to the vector space κ.

4.5 Determinant And Trace

Two important quantities associated with a linear transformation $T : V \to V$ are the trace and determinant. To define these we choose any ordered basis for V so we can define a matrix M_{ij} relative to that basis. Then we can define:

$$\begin{aligned} \mathrm{tr}\,(T) &:= \mathrm{tr}\,(M) := \sum_{i=1}^{n} M_{ii} \\ \det(T) &:= \det(M) := \frac{1}{n!}\sum_{\sigma\in S_n} \epsilon_{\sigma(1)\cdots\sigma(n)} M_{1\sigma(1)}\cdots M_{n\sigma(n)} \end{aligned} \tag{4.29}$$

eq:trdeta

Then we note that if we change basis we have $M \to g^{-1}Mg$ for $g \in GL(n,\kappa)$ and the above expressions remain invariant.

This is a good example where it is simplest to choose a basis and define the quantity, even though it is canonically associated to the linear transformation.

Remark: Note well that this only applies to $T : V \to V$. If $T : V \to W$ is a linear transformation between different vector spaces - *even if they have the same dimension!* - then the definition of the determinant is more subtle. See Section §24 below.

Exercise

Prove that the expressions on the RHS of (4.29) are basis independent so that the equations make sense.

Exercise *Standard identities*

Prove:

1.) $\mathrm{tr}\,(M+N) = \mathrm{tr}\,(M) + \mathrm{tr}\,(N)$

2.) $\mathrm{tr}\,(MN) = \mathrm{tr}\,(NM)$

3.) $\mathrm{tr}\,(SMS^{-1}) = \mathrm{tr}\,M$, $S \in GL(n,k)$.

4.) $\det(MN) = \det M \det N$

5.) $\det(SMS^{-1}) = \det M$

What can you say about $\det(M+N)$?

Exercise *Laplace expansion in complementary minors*

Show that the determinant of an $n\times n$ matrix $A = (a_{ij})$ can be written as

$$\det A = \sum_{H} \epsilon^{HK} b_H c_K \tag{4.30}$$

Here we fix an integer p, H runs over disjoint subsets of order p, $H = \{h_1, \ldots, h_p\}$ and K is the complementary set so that $H \amalg K = \{1, \ldots, n\}$. Then we set

$$b_H := \det(a_{i,h_j})_{1 \le i,j \le p} \tag{4.31}$$

$$c_L = \det(a_{j,k_l})_{p+1 \le j \le n, 1 \le l \le q} \tag{4.32}$$

$p + q = n$ and

$$\epsilon^{HK} = \mathrm{sign}\begin{pmatrix} 1 & 2 & \cdots & p & p+1 & \cdots & n \\ h_1 & h_2 & \cdots & h_p & k_1 & \cdots & k_q \end{pmatrix} \tag{4.33}$$

---

5. New Vector Spaces from Old Ones

5.1 Direct sum

subsec:dirsum

Given two modules V, W over a ring we can form the direct sum. As a set we have:

$$V \oplus W := \{(v, w) : v \in V, w \in W\} \tag{5.1}$$

eq:ds

while the module structure is defined by:

$$\alpha(v_1, w_1) + \beta(v_2, w_2) = (\alpha v_1 + \beta v_2, \alpha w_1 + \beta w_2) \tag{5.2}$$

eq:dsii

valid for all $\alpha, \beta \in R, v_1, v_2 \in V, w_1, w_2 \in W$. We sometimes denote (v, w) by $v \oplus w$. In particular, if the ring is a field these constructions apply to vector spaces.

If V, W are finite dimensional vector spaces then:

$$\dim(V \oplus W) = \dim V + \dim W \tag{5.3}$$

Example 5.1.1:

$$\mathbb{R}^n \oplus \mathbb{R}^m \cong \mathbb{R}^{n+m} \tag{5.4}$$

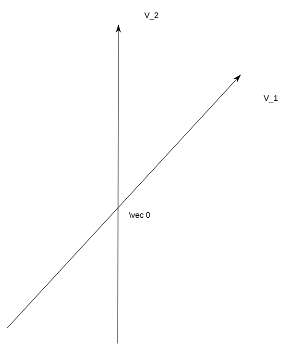

Figure 1: $\mathbb{R}^2$ is the direct sum of two one-dimensional subspaces V_1 and V_2. fig:drctsum

Similarly for operators: With $T_1 : V_1 \to W_1, T_2 : V_2 \to W_2$ we define

$$(T_1 \oplus T_2)(v \oplus w) := T_1(v) \oplus T_2(w) \tag{5.5}$$

for $v \in V_1$ and $w \in W_1$.

Suppose we choose ordered bases:

1. $\{v_1^{(1)}, \dots, v_{n_1}^{(1)}\}$ for V_1
2. $\{v_1^{(2)}, \dots, v_{n_2}^{(2)}\}$ for V_2
3. $\{w_1^{(1)}, \dots, w_{m_1}^{(1)}\}$ for W_1
4. $\{w_1^{(2)}, \dots, w_{m_2}^{(2)}\}$ for W_2

Then, we have matrix representations M_1 and M_2 of T_1 and T_2, respectively. Among the various bases for $V_1 \oplus V_2$ and $W_1 \oplus W_2$ it is natural to choose the ordered bases:

1. $\{v_1^{(1)} \oplus 0, \dots, v_{n_1}^{(1)} \oplus 0, 0 \oplus v_1^{(2)}, \dots, 0 \oplus v_{n_2}^{(2)}\}$ for $V_1 \oplus V_2$
2. $\{w_1^{(1)} \oplus 0, \dots, w_{m_1}^{(1)} \oplus 0, 0 \oplus w_1^{(2)}, \dots, 0 \oplus w_{m_2}^{(2)}\}$ for $W_1 \oplus W_2$

With respect to these ordered bases the matrix of $T_1 \oplus T_2$ will be block diagonal:

$$\begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \tag{5.6}$$

eq:DirectSumMa

But there are, of course, other choices of bases one could make.

Remarks:

1. *Internal and external direct sum.* What we have defined above is sometimes known as *external direct sum.* If $V_1, V_2 \subset V$ are linear subspaces of V, then $V_1 + V_2$ makes sense as a linear subspace of V. If $V_1 + V_2 = V$ and in addition $v_1 + v_2 = 0$ implies $v_1 = 0$ and $v_2 = 0$, that is, if $V_1 \cap V_2 = \{0\}$ then we say that V is the *internal direct sum of* V_1 *and* V_2. In this case every vector $v \in V$ has a canonical decomposition $v = v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$ and hence there is a canonical isomorphism $V \cong V_1 \oplus V_2$. Therefore, we will in general not distinguish carefully between internal and external direct sum. Note well, however, that it can very well happen that $V_1 + V_2 = V$ and yet $V_1 \cap V_2$ is a nonzero subspace. In this case V is most definitely *not* a direct sum!

2. *Subtracting vector spaces in the Grothendieck group.* Since we can "add" vector spaces with the direct sum it is a natural question whether we can also "subtract" vector spaces. There is indeed such a notion, but one must treat it with care. The Grothendieck group can be defined for any monoid. It can then be applied to the monoid of vector spaces. If M is a commutative monoid so we have an additive

operator $m_1 + m_2$ and a 0, but no inverses then we can formally introduce inverses as follows: We consider the set of pairs (m_1, m_2) with the equivalence relation that $(m_1, m_2) \sim (m_3, m_4)$ if

$$m_1 + m_4 = m_2 + m_3 \tag{5.7}$$

Morally speaking, the equivalence class of the pair $[(m_1, m_2)]$ can be thought of as the difference $m_1 - m_2$ and we can no add

$$[(m_1, m_2)] + [(m_1', m_2')] = [(m_1 + m_1', m_2 + m_2')] \tag{5.8}$$

But now, $[(m, m)] = [(0, 0,)]$ so $[(m_2, m_1)]$ is the inverse of $[(m_1, m_2)]$ so now we have produced an Abelian group $K(M)$ associated to the monoid. For example, if we apply this to the monoid of nonnegative integers then we recover the Abelian group of all integers.

This idea can be generalized to suitable categories. One setting is that of an *additive category* $\mathcal{C}$. What this means is that the morphism spaces $\hom(x_1, x_2)$ are Abelian groups and the composition law is bi-additive, meaning the composition of morhphisms

♣Need to check: Do you need an abelian category to define $K(\mathcal{C})$. More generally we can define $K(\mathcal{C})$ for an exact cagetory. Is there a problem with just using additive categories? Definition here seems to make sense. ♣

$$\hom(x, y) \times hom(y, z) \to \hom(x, z) \tag{5.9}$$

is "bi-additive" i.e. distributive: $(f+g) \circ h = f \circ h + g \circ h$. In an additive category we also have a zero object 0 that has the property that $\hom(0, x)$ and $\hom(x, 0)$ are the trivial Abelian group. Finally, there is a notion of direct sum of objects $x_1 \oplus x_2$, that is, given two objects we can produce a new object $x_1 \oplus x_2$ together with distinguished morphisms $\iota_1 \in \hom(x_1, x_1 \oplus x_2)$ and $\iota_2 \in \hom(x_2, x_1 \oplus x_2)$ so that

$$\hom(z, x_1) \times \hom(z, x_2) \to \hom(z, x_1 \oplus x_2) \tag{5.10}$$

defined by $(f, g) \mapsto \iota_1 \circ f + \iota_2 \circ f$ is an isomorphism of Abelian groups.

The category **VECT** of finite-dimensional vector spaces and linear transformations or the category $\mathrm{Mod}(R)$ of modules over a ring R are good examples of additive categories. In such a category we can define the *Grothendieck group* $K(\mathcal{C})$. It is the abelian group of pairs of objects $(x, y) \in Obj(\mathcal{C}) \times Obj(\mathcal{C})$ subject to the relation

$$(w, x) \sim (y, z) \tag{5.11}$$

if there exists an object u so that the object $w \oplus z \oplus u$ is isomorphic to $y \oplus x \oplus u$. In the case of **VECT** we can regard $[(V_1, V_2)] \in K(\mathbf{VECT})$ as the formal difference $V_1 - V_2$. It is then a good exercise to show that, as Abelian groups $K(\mathbf{VECT}) \cong \mathbb{Z}$. (This is again essentially the statement that the only invariant of a finite-dimensional vector space is its dimension.) However, when we apply this construction to continuous families of vector spaces parametrized by topological spaces we obtain a very nontrivial mathematical subject known as *K-theory*.

In physics one way such *virtual vector spaces* arise is in $\mathbb{Z}_2$-graded or super-linear algebra. See Section §23 below. Given a supervector space $V^0 \oplus V^1$ it is natural to associate to it the virtual vector space $V^0 - V^1$ (but you cannot go the other way - why not?). Some important constructions only depend on the virtual vector space (or, more generally, the virtual vector bundle, when working with families).

Exercise

Construct an example of a vector space V and proper subspaces $V_1 \subset V$ and $V_2 \subset V$ such that

$$V_1 + V_2 = V \tag{5.12}$$

but $V_1 \cap V_2$ is not the zero vector space. Rather, it is a vector space of positive dimension. In this case V is not the internal direct sum of V_1 and V_2.

Exercise

1. $\mathrm{Tr}(T_1 \oplus T_2) = \mathrm{Tr}(T_1) + \mathrm{Tr}(T_2)$
2. $\det(T_1 \oplus T_2) = \det(T_1)\det(T_2)$

Exercise

a.) Show that there are natural isomorphisms

$$V \oplus W \cong W \oplus V \tag{5.13}$$

$$(U \oplus V) \oplus W \cong U \oplus (V \oplus W) \tag{5.14}$$

b.) Suppose I is some set, not necessarily ordered, and we have a family of vector spaces V_i indexed by $i \in I$. One can give a definition of the vector space:

$$\oplus_{i \in I} V_i \tag{5.15}$$

but to do so in general one should use the restricted product so that an element is a collection $\{v_i\}$ of vectors $v_i \in V_i$ where in the Cartesion product all but finitely many v_i are zero. Define a vector space structure on this set.

Exercise

How would the matrix in (5.6) change if we used bases:

1. $\{v_1^{(1)} \oplus 0, \dots, v_{n_1}^{(1)} \oplus 0, 0 \oplus v_1^{(2)}, \dots, 0 \oplus v_{n_2}^{(2)}\}$ for $V_1 \oplus V_2$
2. $\{0 \oplus w_1^{(2)}, \dots, 0 \oplus w_{m_2}^{(2)}, w_1^{(1)} \oplus 0, \dots, w_{m_1}^{(1)} \oplus 0\}$ for $W_1 \oplus W_2$

5.2 Quotient Space

$W \subset V$ is a vector subspace then

$$V/W \tag{5.16}$$

is the space of equivalence classes $[v]$ where $v_1 \sim v_2$ if $v_1 - v_2 \in W$. This is the quotient of abelian groups. It becomes a vector space when we add the rule

$$\alpha(w + W) := \alpha w + W. \tag{5.17}$$

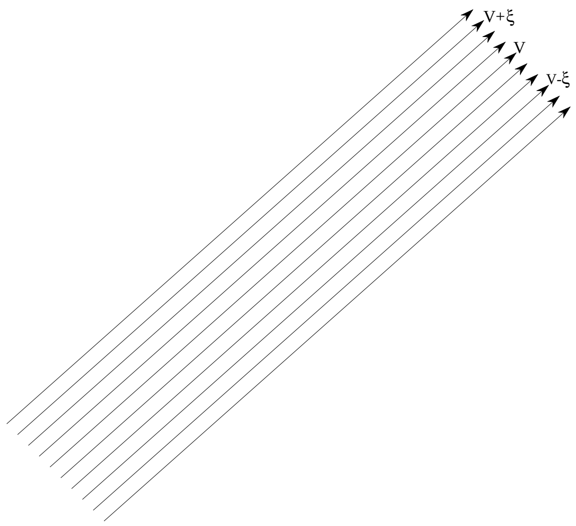

Figure 2: Vectors in the quotient space $\mathbb{R}^2/V$, $[v] = V + v$. The quotient space is the moduli space of lines parallel to V. fig:quotspace

Claim: V/W is also a vector space. If V, W are finite dimensional then

$$\dim(V/W) = \dim V - \dim W \tag{5.18}$$

We define a *complementary subspace* [4] to W to be another subspace $W' \subset V$ so that V is the internal direct sum of W and W'. Recall that this means that every $v \in V$ can be *uniquely* written as $v = w + w'$ with $w \in W$ and $w' \in W'$ so that $V \cong W \oplus W'$. It follows from Theorem 4.3.1 that a complementary subspace to W always exists and moreover there is an isomorphism:

$$V/W \cong W' \tag{5.19}$$ eq:vwwprime

Note that given $W \subset V$ there is a canonical vector space V/W, but there is no unique choice of W' so the isomorphism (5.19) cannot be natural.

Warning! One should not confuse (as is often done) V/W with a subspace of V.

---

[4] Please note, it is not a "complimentary subspace." A "complimentary subspace" might praise your appearance, or accompany snacks on an airplane flight.

Exercise *Practice with quotient spaces*

a.) $V = \mathbb{R}^2$, $W = \{(\alpha_1 t, \alpha_2 t) : t \in \mathbb{R}\}$. If $\alpha_1 \neq 0$ then we can identify $V/W \cong \mathbb{R}$ via $s \to (0, s) + W$. Show that the inverse transformation is given by $(v_1, v_2) \to s$ where $s = (\alpha_1 v_2 - \alpha_2 v_1)/\alpha_1$. What happens if $\alpha_1 = 0$?

b.) If $V = \mathbb{R}^n$ and $W \cong \mathbb{R}^m$, $m < n$ with $W = \{(x_1, \dots, x_m, 0, \dots, 0)\}$, when is $v_1 + W = v_2 + W$?

Exercise

Suppose $T : V_1 \to V_2$ is a linear transformation and $W_1 \subset V_1$ and $W_2 \subset V_2$ are linear subspaces. Under what conditions does T descend to a linear transformation

$$\tilde{T} : V_1/W_1 \to V_2/W_2 \qquad ? \tag{5.20}$$

The precise meaning of "descend to" is that $\tilde{T}$ fits in the commutative diagram

$$\begin{tikzcd} V_1 \arrow[r, "T"] \arrow[d, "\pi_1"] & V_2 \arrow[d, "\pi_2"] \\ V_1/W_1 \arrow[r, "\tilde{T}"] & V_2/W_2 \end{tikzcd} \tag{5.21}$$

5.3 Tensor Product

The (technically) natural definition is a little sophisticated, (see, e.g., Lang's book *Algebra*, and remark 3 below), but for practical purposes we can describe it in terms of bases:

Let $\{v_i\}$ be a basis for V, and $\{w_s\}$ be a basis for W. Then $V \otimes W$ is the vector space spanned by $v_i \otimes w_s$ subject to the rules:

$$\begin{aligned} (\alpha v + \alpha' v') \otimes w &= \alpha(v \otimes w) + \alpha'(v' \otimes w) \\ v \otimes (\alpha w + \alpha' w') &= \alpha(v \otimes w) + \alpha'(v \otimes w') \end{aligned} \tag{5.22}$$ eq:dspp

for all $v, v' \in V$ and $w, w' \in W$ and all scalars α, α'.

If V, W are finite dimensional then so is $V \otimes W$ and

$$\dim(V \otimes W) = (\dim V)(\dim W) \tag{5.23}$$

In particular:

$$\mathbb{R}^n \otimes \mathbb{R}^m \cong \mathbb{R}^{nm} \tag{5.24}$$

We can similarly discuss the tensor product of operators: With $T_1 : V_1 \to W_1, T_2 : V_2 \to W_2$ we define

$$(T_1 \otimes T_2)(v_i \otimes w_j) := (T_1(v_i)) \otimes (T_2(w_j)) \tag{5.25}$$ eq:deftprd

Remarks

1. Examples where the tensor product occurs in physics are in quantum systems with several independent degrees of freedom. In general two distinct systems with Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ have statespace $\mathcal{H}_1 \otimes \mathcal{H}_2$. For example, consider a system of N spin $\frac{1}{2}$ particles. The Hilbert space is

$$\mathcal{H} = \overbrace{\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}^{\text{N times}} \qquad \dim_{\mathbb{C}} \mathcal{H} = 2^N \tag{5.26}$$

2. In Quantum Field Theory the implementation of this idea for the Hilbert space associated with a spatial manifold, when the spatial manifold is divided in two parts by a domain wall, is highly nontrivial due to UV divergences.

3. In spacetimes with nontrivial causal structure the laws of quantum mechanics become difficult to understand. If two systems are separated by a horizon, should we take the tensor product of the Hilbert spaces of these systems? Questions like this quickly lead to a lot of interesting puzzles.

4. *Defining Tensor Products Of Modules Over Rings*

   The tensor product of modules over a ring R can always be defined, even if the modules are not free and do not admit a basis. If we have no available basis then the above low-brow approach will not work. One needs to use a somewhat more abstract definition in terms of a "universal property."

   Consider <u>any</u> *bilinear* mapping $f : V \times W \to U$ for <u>any</u> R-module U. Then the characterizing (or "universal") property of a tensor product is that this map f "factors uniquely through a map from the tensor product." That means, there is

   (a) A bilinear map $\pi : V \times W \to V \otimes_R W$

   (b) A unique linear map $f' : V \otimes_R W \to U$ such that

$$f = \pi \circ f' \tag{5.27}$$

   In terms of commutative diagrams

$$\begin{tikzcd} V \times W \arrow[r, "\pi"] \arrow[dr, "f"'] & V \otimes_R W \arrow[d, "f'"] \\ & U \end{tikzcd} \tag{5.28}$$

   One then proves that, if a module $V \otimes_R W$ satisfying this property exists then it is unique up to unique isomorphism. Hence this property is in fact a defining property of the tensor product: This is the "natural" definition one finds in math books.

   The above property defines the tensor product, but does not prove that such a thing exists. To construct the tensor product one considers the free R-module generated by objects $v \times w$ where $v \in V, w \in W$ and takes the quotient by the submodule generated by all vectors of the form:

$$(v_1 + v_2) \times w - v_1 \times w - v_2 \times w \tag{5.29}$$

$$v \times (w_1 + w_2) - v \times w_1 - v \times w_2 \tag{5.30}$$

$$\alpha(v \times w) - (\alpha v) \times w \tag{5.31}$$

$$\alpha(v \times w) - v \times (\alpha w) \tag{5.32}$$

The projection of a vector $v \times w$ in this (incredibly huge!) module into the quotient module are denoted by $v \otimes w$.

An important aspect of this natural definition is that it allows us to define the tensor product $\otimes_{i \in I} V_i$ of a family of vector spaces labeled by a not necessarily ordered (but finite) set I.

Exercise

Given any three vector spaces U, V, W over a field κ show that there are natural isomorphisms:

a.) $V \otimes W \cong W \otimes V$

b.) $(V \otimes W) \otimes U \cong V \otimes (W \otimes U)$

c.) $U \otimes (V \oplus W) \cong U \otimes V \oplus U \otimes W$

d.) If $T_1 \in \mathrm{End}(V_1)$ and $T_2 \in \mathrm{End}(V_2)$ then under the isomorphism $V_1 \otimes V_2 \cong V_2 \otimes V_1$ the linear transformation $T_1 \otimes T_2$ is mapped to $T_2 \otimes T_1$.

e.) Show that $T_1 \otimes 1$ commutes with $1 \otimes T_2$.

Exercise *Practice with the $\otimes$ product*

1. $\mathrm{Tr}(T_1 \otimes T_2) = \mathrm{Tr}(T_1) \cdot \mathrm{Tr}(T_2)$
2. $\det(T_1 \otimes T_2) = (\det T_1)^{\dim V_2} \cdot (\det T_2)^{\dim V_1}$

Exercise *Matrices For Tensor Products Of Operators*

Let V, W be modules over a commutative ring R (or vector spaces over a field κ). Suppose that a linear transformation $T^{(1)} : V \to V$ has matrix $A_1 \in M_n(R)$, with matrix elements $(A_1)_{ij}$, $1 \le i, j \le n$ with respect to an ordered basis $\{v_1, \dots, v_n\}$ and $T^{(2)} : W \to W$ has matrix $A_2 \in M_m(R)$ with matrix elements $(A_2)_{ab}$, $1 \le a, b \le m$ with respect to an ordered basis $\{w_1, \dots, w_m\}$.

a.) Show that if we use the ("lexicographically") ordered basis

$$v_1 \otimes w_1, v_1 \otimes w_2, \ldots, v_1 \otimes w_m, v_2 \otimes w_1, \ldots, v_2 \otimes w_m, \ldots, v_n \otimes w_1, \ldots, v_n \otimes w_m \tag{5.33}$$

Then the matrix for the operator $T^{(1)} \otimes T^{(2)}$ may be obtained from the following rule: Take the $n \times n$ matrix for $T^{(1)}$. Replace each of the matrix elements $(A_1)_{ij}$ by the $m \times m$ matrix

$$(A_1)_{ij} \to ((A_1)_{ij}(A_2)_{ab})_{1 \le a,b \le m} = \begin{pmatrix} (A_1)_{ij}(A_2)_{11} & \cdots & (A_1)_{ij}(A_2)_{1m} \\ \vdots & \vdots & \vdots \\ (A_1)_{ij}(A_2)_{m1} & \cdots & (A_1)_{ij}(A_2)_{mm} \end{pmatrix} \tag{5.34}$$

The result is an element of the ring

$$M_n(M_m(R)) \cong M_{nm}(R) \tag{5.35}$$

b.) Show that if we instead use the ordered basis

$$v_1 \otimes w_1, v_2 \otimes w_1, \ldots, v_n \otimes w_1, v_1 \otimes w_2, \ldots, v_n \otimes w_2, \ldots, v_1 \otimes w_m, \ldots, v_n \otimes w_m \tag{5.36}$$

Then to compute the matrix for the same linear transformation, $T^{(1)} \otimes T^{(2)}$, we would instead start with the matrix $A^{(2)}$ and replace each matrix element by inserting that matrix element times the matrix $A^{(1)}$. The net result is an element of The result is an element of the ring

$$M_m(M_n(R)) \cong M_{mn}(R) \tag{5.37}$$

Since $M_{nm}(R) = M_{mn}(R)$ we can compare to the expression in (a) and it will in general be different.

c.) Using the two conventions of parts (a) and (b) compute

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \otimes \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \tag{5.38}$$

as a 4×4 matrix.

Exercise *Tensor product of modules*

The above universal definition also applies to tensor products of modules over a ring R:

$$M \otimes_R N \tag{5.39}$$

where it can be very important to specify the ring R. In particular we have the very crucial relation that $\forall m \in M, r \in R, n \in N$:

$$m \cdot r \otimes n = m \otimes r \cdot n \tag{5.40}$$

where we have written M as a right-module and N as a left module so that the expression works even for noncommutative rings. [5]

These tensor products have some features which might be surprising if one is only familiar with the vector space example.

a.) $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Z}_n = 0$

b.) $\mathbb{Z}_n \otimes_{\mathbb{Z}} \mathbb{Z}_m = \mathbb{Z}_{(n,m)}$

In particular $\mathbb{Z}_p \otimes_{\mathbb{Z}} \mathbb{Z}_q$ is the 0-module if p, q are relatively prime.

5.4 Dual Space

Consider the *linear functionals* $Hom(V, \kappa)$. This is a vector space known as the *dual space*.

This vector space is often denoted $\check{V}$, or by $V^\vee$, or by V^*. Since V^* is sometimes used for the complex conjugate of a complex vector space we will generally use the more neutral symbol $V^\vee$

♣Eliminate the use of V^*. This is too confusing. ♣

One can prove that $\dim V^\vee = \dim V$ so if V is finite dimensional then V and $V^\vee$ must be isomorphic. However, there is no natural isomorphism between them. If we choose a basis $\{v_i\}$ for V then we can define linear functionals ℓ_i by the requirement that

$$\ell_i(v_j) = \delta_{ij} \qquad \forall v_j \tag{5.41}$$

and then we extend by linearity to compute ℓ_i evaluated on linear combinations of v_j. The linear functionals ℓ_i form a basis for $V^\vee$ and it is called *dual basis* for $V^\vee$ with respect to the v_i. Sometimes we will call the dual basis $\hat{v}^i$ or $v_i^\vee$.

Remarks:

1. It is important to stress that there is no *natural* isomorphism between V and $V^\vee$. Once one chooses a basis v_i for V then there is indeed a naturally associated dual basis $\ell_i = \hat{v}^i = v_i^\vee$ for $V^\vee$ and then both vector spaces are isomorphic to $\kappa^{\dim V}$. The lack of a natural isomorphism means that when we consider vector spaces in families, or add further structure, then it can very well be that V and $V^\vee$ become nonisomorphic. For example, if $\pi : E \to M$ is a vector bundle over a manifold then there is a canonically associated vector bundle $\pi^\vee : E^\vee \to M$ where we take the dual space of the fibers of E. In general $E^\vee$ and E are nonisomorphic vector bundles. To be more specific, the two (one for $+$ and one for $-$) families of vector spaces $L_{\pm,\vec{x}}$ in equation ******* above define vector bundles $\pm_\pm : L_\pm \to S^2$. The dual bundle to L_+ is not isomorphic to L_+. So that this means is that, even though there is indeed a family of isomorphisms
$$\psi(\hat{x}) : L_{+,\vec{x}} \cong L^\vee_{+,\vec{x}} \tag{5.42}$$
(because both sides are one dimensional vector spaces!) there is in fact no continuous family of isomorphisms.

---

[5] In general a left-module for a ring R is naturally isomorphic to a right-module for the opposite ring $R^{\rm opp}$. When R is commutative R and $R^{\rm opp}$ are also naturally isomorphic.

2. *Notation.* There is also a notion of a complex conjugate of a complex vector space which should not be confused with $V^\vee$ (the latter is defined for a vector space over any field κ). We will denote the complex conjugate, defined in Section §9 below, by $\overline{V}$. Similarly, the dual operator $T^\vee$ below is not to be confused with Hermitian adjoint. The latter is only defined for inner product spaces, and will be denoted $T^\dagger$. Note, however, that for complex numbers we will occasionally use z^* for the complex conjugate.

♣For consistency you really should use $\bar{z}$ for the complex conjugate of a complex number z everywhere. ♣

3. If
$$T : V \to W \tag{5.43}$$
is a linear transformation between two vector spaces then there is a canonical dual linear transformation
$$T^\vee : W^\vee \to V^\vee \tag{5.44}$$
To define it, suppose $\ell \in W^\vee$. Then we define $T^\vee(\ell)$ by saying how it acts on a vector $v \in V$. The formula is:
$$(T^\vee(\ell))(v) := \ell(T(v)) \tag{5.45}$$

4. Note especially that dualization "reverses arrows": If
$$V \xrightarrow{T} W \tag{5.46}$$
then
$$V^\vee \xleftarrow{T^\vee} W^\vee \tag{5.47}$$
This is a general principle: If there is a commutative diagram of linear transformations and vector spaces then dualization reverses all arrows.

♣Need exercise relating $(A_1 \otimes A_2)^{tr}$ to $A_1^{tr} \otimes A_2^{tr}$. ♣

Exercise
If V is finite dimensional show that
$$(V^\vee)^\vee \cong V \tag{5.48}$$
and in fact, this is a natural isomorphism (you do not need to choose a basis to define it).

For this reason the dual pairing between V and $V^\vee$ is often written as $\langle \ell, v\rangle$ to emphasize the symmetry between the two factors.

Exercise *Matrix of* $T^\vee$
a.) Suppose $T : V \to W$ is a linear transformation, and we choose a basis $\{v_i\}$ for V and $\{w_s\}$ for W, so that the matrix of the linear transformation is M_{si}.

Show that the matrix of $T^\vee : W^\vee \to V^\vee$ with respect to the dual bases to $\{v_i\}$ and $\{w_s\}$ is the transpose:

$$(M^{tr})_{is} := M_{si} \tag{5.49}$$

b.) Suppose that v_i and v'_i are two bases for a vector space V and are related by $v'_i = \sum_j S_{ji} v_j$. Show that the corresponding dual bases ℓ_i and ℓ'_i are related by $\ell'_i = \sum_j \hat{S}_{ji} \ell_j$ where

$$\hat{S} = S^{tr,-1} \tag{5.50}$$

c.) For those who know about vector bundles: If a bundle $\pi : E \to M$ has a coordinate chart with $\mathcal{U}_\alpha \subset M$ and transition functions $g^E_{\alpha\beta} : \mathcal{U}_\alpha \cap \mathcal{U}_\beta \to GL(n,\kappa)$ show that the dual bundle has transition functions $g^{E^\vee}_{\alpha\beta} = (g^E_{\alpha\beta})^{-1,tr}$.

Exercise

Prove that there are <u>canonical</u> isomorphisms:

$$\mathrm{Hom}(V,W) \cong V^\vee \otimes W. \tag{5.51}$$

eq:HomVW-iso

$$\mathrm{Hom}(V,W)^\vee \cong \mathrm{Hom}(W,V) \tag{5.52}$$

$$\mathrm{Hom}(V,W) \cong \mathrm{Hom}(W^\vee, V^\vee) \tag{5.53}$$

Although the isomorphism (5.51) is a natural isomorphism it is useful to say what it means in terms of bases: If

$$T v_i = \sum_s M_{si} w_s \tag{5.54}$$

then the corresponding element in $V^\vee \otimes W$ is

$$T = \sum_{i,s} M_{si} v_i^\vee \otimes w_s \tag{5.55}$$

Exercise

If U is any vector space and $W \subset U$ is a linear subspace then we can define

$$W^\perp := \{\ell | \ell(w) = 0 \quad \forall w \in W\} \subset U^\vee. \tag{5.56}$$

Show that
a.) $(W^\perp)^\perp = W$.
b.) $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$.
c.) There is a canonical isomorphism $(U/W)^\vee \cong W^\perp$.

Exercise *Flags*

♣Some good exercises can be extracted from the snakes paper. ♣

6. Tensor spaces

Given a vector space V we can form

$$V^{\otimes n} \equiv V \otimes V \otimes \cdots \otimes V \tag{6.1}$$

Elements of this vector space are called *tensors of rank n over V*.

Actually, we could also consider the dual space $V^\vee$ and consider more generally

$$V^{\otimes n} \otimes (V^\vee)^{\otimes m} \tag{6.2}$$

eq:mxdtens

Up to isomorphism, the order of the factors does not matter.

Elements of (6.2) are called *mixed tensors of type* (n, m). For example

$$\mathrm{End}(V) \cong V^\vee \otimes V \tag{6.3}$$

are mixed tensors of type $(1, 1)$.

Now, if we choose an ordered basis $\{v_i\}$ for V then we have a canonical dual basis for $V^\vee$ given by $\{\hat{v}^i\}$ with

$$\hat{v}^i(v_j) = \delta^i{}_j \tag{6.4}$$

Notice, we have introduced a convention of upper and lower indices which is very convenient when working with mixed tensors.

A typical mixed tensor can be expanded using the basis into its components:

$$T = \sum_{i_1, i_2, \dots, i_n, j_1, \dots, j_m} T^{i_1, i_2, \dots, i_n}{}_{j_1, \dots, j_m} v_{i_1} \otimes \cdots \otimes v_{i_n} \otimes \hat{v}^{j_1} \otimes \cdots \otimes \hat{v}^{j_m} \tag{6.5}$$

We will henceforth often assume the summation convention where repeated indices are automatically summed.

Recall that if we make a change of basis

$$w_i = g^j{}_i v_j \tag{6.6}$$

(sum over j understood) then the dual bases are related by

$$\hat{w}^i = \hat{g}_j{}^i \hat{w}^j \tag{6.7}$$

where the matrices are related by

$$\hat{g} = g^{tr, -1} \tag{6.8}$$

The fact that g has an upper and lower index makes good sense since it also defines an element of $\mathrm{End}(V)$ and hence the matrix elements are components of a tensor of type $(1, 1)$.

Under change of basis the (passive) change of components is given by

$$(T')^{i'_1, \dots, i'_n}{}_{j'_1, \dots, j'_m} = g^{i'_1}{}_{k_1} \cdots g^{i'_n}{}_{k_n} \hat{g}_{j'_1}{}^{\ell_1} \cdots \hat{g}_{j'_m}{}^{\ell_m} T^{k_1, \dots, k_n}{}_{\ell_1, \dots, \ell_m} \tag{6.9}$$

This is the standard transformation law for tensors.

♣Comment on "covariant" and "contravariant" indices. ♣

6.1 Totally Symmetric And Antisymmetric Tensors

There are two subspaces of $V^{\otimes n}$ which are of particular importance. Note that we have a homomorphism $\rho : S_n \to GL(V^{\otimes n})$ defined by

$$\rho(\sigma) : u_1 \otimes \cdots \otimes u_n \to u_{\sigma(1)} \otimes \cdots \otimes u_{\sigma(n)} \tag{6.10}$$

on any vector in $V^{\otimes n}$ of the form $u_1 \otimes \cdots \otimes u_n$ Then we extend by linearity to all vectors in the vector space. Note that with this definition

$$\rho(\sigma_1) \circ \rho(\sigma_2) = \rho(\sigma_1 \sigma_2) \tag{6.11}$$

Thus, $V^{\otimes n}$ is a *representation* of S_n in a natural way. This representation is *reducible*, meaning that there are invariant proper subspaces. (See Chapter four below for a systematic treatment.)

Two particularly important proper subspaces are:

$S^n(V)$: These are the *totally symmetric tensors*, i.e. the vectors invariant under S_n. This is the subspace where $T(\sigma) = 1$ for all $\sigma \in S_n$.

$\Lambda^n(V)$: antisymmetric tensors, transform into $v \to \pm v$ depending on whether the permutation is even or odd. These are also called n-forms. This is the subspace of vectors on which $T(\sigma)$ acts by ± 1 given by $T(\sigma) = \epsilon(\sigma)$.

Remarks

1. A basis for $\Lambda^n V$ can be defined using the extremely important wedge product construction. If v, w are two vectors we define

$$v \wedge w := \frac{1}{2!}(v \otimes w - w \otimes v) \tag{6.12}$$

If $v_1, \dots, v_n$ are any n vectors we define

$$v_1 \wedge \cdots \wedge v_n := \frac{1}{n!} \sum_{\sigma \in S_n} \epsilon(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)} \tag{6.13}$$

Thus, if we choose a basis w_i for V then $\Lambda^n V$ is spanned by $w_{i_1} \wedge \cdots \wedge w_{i_n}$. Note, in particular, that if $\mathrm{dim} V = d$ then $\Lambda^d(V)$ is one-dimensional.

2. In quantum mechanics if V is the Hilbert space of a single particle the Hilbert space of n identical particles is a subspace of $V^{\otimes n}$. In three space dimensions the only particles that appear in nature are bosons and fermions, which are states in $S^n(V)$ and $\Lambda^n(V)$, respectively. In this way, given a space V of one-particle states - interpreted as the span of creation operators $\{a_j^\dagger\}$ we form the bosonic Fock space

$$S^\bullet V = \mathbb{C} \oplus \oplus_{\ell=1}^\infty S^\ell V \tag{6.14}$$

when the $a_j^\dagger$'s commute or the Fermionic Fock space

$$\Lambda^\bullet V = \mathbb{C} \oplus \oplus_{\ell=1}^\infty \Lambda^\ell V \tag{6.15}$$

when they anticommute. (The latter terminates, if V is finite-dimensional).

Exercise *Decomposing* $V^{\otimes 2}$

a.) Show that $\Lambda^2(V)$ is given by linear combinations of vectors of the form $x\otimes y - y\otimes x$, and $S^2(V)$ is given by vectors of the form $x\otimes y + y\otimes x$.

Show that [6]

$$V^{\otimes 2} \cong S^2(V) \oplus \Lambda^2(V) \tag{6.16}$$

b.) If V_1, V_2 are two vector spaces show that

$$S^2(V_1 \oplus V_2) \cong S^2(V_1) \oplus S^2(V_2) \oplus V_1 \otimes V_2 \tag{6.17}$$

$$\Lambda^2(V_1 \oplus V_2) \cong \Lambda^2(V_1) \oplus \Lambda^2(V_2) \oplus V_1 \otimes V_2 \tag{6.18}$$

♣Incorporate some nice generalizations of this from CSS project with Jeff. ♣

Exercise *Counting dimensions*

Suppose V is finite dimensional of dimension $\dim V = d$. It is of interest to count the dimensions of $\Lambda^n(V)$ and $S^n(V)$.

a.) Show that

$$\sum_{n\geq 0} q^n \dim\Lambda^n(V) = (1+q)^d \tag{6.19}$$

and therefore

$$\dim\Lambda^n(V) = \binom{d}{n} \tag{6.20}$$

b.) Show that

$$\sum_{n\geq 0} q^n \dim S^n(V) = \frac{1}{(1-q)^d} \tag{6.21}$$

and therefore

$$\dim S^n(V) = \binom{n+d-1}{n} \tag{6.22}$$

Remark The formula for the number of n^{th}-rank symmetric tensors in d dimensions involves *multichoosing*. In combinatorics we denote

$$\left(\!\!\binom{d}{n}\!\!\right) := \binom{n+d-1}{n} \tag{6.23}$$

eq:d-multichoo

and say "d multichoose n." This is the number of ways of counting n objects from a set of d elements where repetition is allowed and order does not matter. One standard proof is the "method of stars and bars." The distinct components of the symmetric tensor are $T_{1,\dots 1,2,\dots,2,\dots,d,\dots,d}$. We can imagine $n+d-1$ slots, and we are to choose $d-1$ of these slots as the position of a bar. Then the number of unoccupied slots between bars, reading from

[6] *Answer*: $\Lambda^2(V)$ is the span of vectors of the form $x\otimes y - y\otimes x$, while $S^2(V)$ is the span of vectors of the form $x\otimes y + y\otimes x$. But we can write any $x\otimes y$ as a sum of two such vectors in an obvious way.

left to right, gives the number of 1's, 2's, ... d's. Thus, the independent components of a symmetric tensor are in 1-1 correspondence with a choice of $(d-1)$ slots out of a total of $n+d-1$ slots. This is the binomial coefficient (6.23). Note then that there is a pleasing symmetry between the antisymmetric and symmetric cases:

$$\dim\Lambda^n(V) = \binom{d}{n} = \frac{d(d-1)(d-2)\cdots(d-(n-1))}{n!} \tag{6.24}$$

$$\dim S^n(V) = \left(\!\binom{d}{n}\!\right) = \frac{d(d+1)(d+2)\cdots(d+(n-1))}{n!} \tag{6.25}$$

♣Does this mean that bosons in negative numbers of dimensions are fermions? ♣

Exercise

a.) Let $\dim V = 2$. Show that

$$\dim S^n(V) = n+1 \tag{6.26}$$

b.) Let $d = \dim V$. Show that

$$S^3(V) \oplus \Lambda^3(V) \tag{6.27}$$

is a subspace of $V^{\otimes 3}$ of codimension $\frac{2}{3}d(d^2-1)$ and hence has positive codimension for $d>1$.

Thus, there are more nontrivial representations of the symmetric group to account for. These will be discussed in Chapter 4.

6.2 Algebraic structures associated with tensors

:TensorAlgebra

There are a number of important algebraic structures associated with tensor spaces.

Definition For a vector space V over κ the *tensor algebra* TV is the $\mathbb{Z}$-graded algebra over κ with underlying vector space:

$$T^\bullet V := \kappa \oplus V \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus \cdots \tag{6.28}$$

and multiplication defined by using the tensor product to define the multiplication

$$V^{\otimes k} \times V^{\otimes \ell} \to V^{\otimes(k+\ell)} \tag{6.29}$$

and then extending by linearity.

Remarks

1. In concrete terms the algebra multiplication is defined by the very natural formula:

$$(v_1 \otimes \cdots \otimes v_k)\cdot(w_1 \otimes \cdots \otimes w_\ell) := v_1 \otimes \cdots \otimes v_k \otimes w_1 \otimes \cdots \otimes w_\ell \tag{6.30}$$

2. Note that we can define $V^{\otimes 0} := \kappa$ and $V^{\otimes n} = \{0\}$ when n is a negative integer making

$$T^{\bullet}V = \oplus_{\ell\in\mathbb{Z}} V^{\otimes \ell} \tag{6.31}$$

into a $\mathbb{Z}$-graded algebra. The vectors V are then regarded as having degree $= 1$.

Several quotients of the tensor algebra are of great importance in mathematics and physics.

♣This discussion of ideals really belongs in the section on algebras. Then you can give standard examples such as subalgebra of $\mathbb{Z}$ of integers divisible by n. etc. ♣

Quite generally, if A is an algebra and $B \subset A$ is a subalgebra we can ask if the vector space A/B admits an algebra structure. The only natural definition would be

$$(a + B) \odot (a' + B) := a \odot a' + B \tag{6.32}$$

However, there is a problem with this definition! It is not necessarily well-defined. The problem is very similar to trying to define a group structure on the set of cosets G/H of a subgroup $H \subset G$. Just as in that case, we can only give a well-defined product when $B \subset A$ satisfies a suitable condition. In the present case we need to know that, for all $a, a' \in A$ and for all $b, b' \in B$ then there must be a $b'' \in B$ so that

$$(a + b) \odot (a' + b') = a \odot a' + b'' \tag{6.33}$$

Taking various special cases this implies that for all $b \in B$ and all $a \in A$ we must have

$$a \odot b \in B \qquad \& \qquad b \odot a \in B \tag{6.34}$$

Such a subalgebra $B \subset A$ is known as an *ideal.* If we have a subset $S \subset A$ then $I(S)$, the ideal generated by S, also denoted simply (S) is the smallest ideal in A that contains S.

1. *Symmetric algebra*: This is the quotient of $T^{\bullet}V$ by the ideal generated by all tensors of the form $v \otimes w - w \otimes v$. It is denoted $S^{\bullet}V$, and as a vector space is

$$S^{\bullet}V = \oplus_{\ell=0}^{\infty} S^{\ell} V \tag{6.35}$$

If we denote the symmetrization of an elementary tensor simply by

$$v_1 \cdots v_k := \frac{1}{k!} \sum_{\sigma\in S_k} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \tag{6.36}$$

then the product is simply

$$(v_1 \cdots v_k) \cdot (w_1 \cdots w_\ell) = v_1 \cdots v_k w_1 \cdots w_\ell \tag{6.37}$$

It is also the free commutative algebra generated by V. Even when V is finite dimensional this is an infinite-dimensional algebra.

2. *Exterior algebra*: This is the quotient of $T^{\bullet}V$ by the ideal generated by all tensors of the form $v \otimes w + w \otimes v$ where $v, w \in V$. It is denoted $\Lambda^{\bullet}V$, and as a vector space is

$$\Lambda^{\bullet}V = \oplus_{\ell=0}^{\infty} \Lambda^{\ell} V \tag{6.38}$$

with product given by exterior product of forms

$$(v_1 \wedge \cdots \wedge v_k) \cdot (w_1 \wedge \cdots \wedge w_\ell) := v_1 \wedge \cdots \wedge v_k \wedge w_1 \wedge \cdots \wedge w_\ell \tag{6.39}$$

When V is finite dimensional this algebra is finite dimensional.

3. *Clifford algebra*: Let Q be a quadratic form on V. Then the Clifford algebra is the algebra defined by TV/I where I is the ideal generated by $v \otimes w + w \otimes v - Q(v, w)1$. This is an extremely important algebra in physics. For the moment let us just note the following elementary points: If we choose a basis $\{e_i\}$ for V and $Q(e_i, e_j) = Q_{ij}$ then we can think of $C\ell(Q)$ as the algebra over κ generated by the e_i subject to the relations:

$$e_i e_j + e_j e_i = Q_{ij} \tag{6.40}$$

If Q can be diagonalized then we can choose a basis and we write

$$e_i e_j + e_j e_i = q_i \delta_{ij} \tag{6.41}$$

If $Q = 0$ we obtain what is known as the *Grassmann algebra.* The other extreme is the case where Q is nondegenerate so all the q_i are all nonzero. (And this is indeed sometimes assumed when an someone speaks of a "Clifford algebra.") To go further it depends quite a bit one what field we are working with. If $\kappa = \mathbb{C}$ and the $q_i \neq 0$ we can change basis so that

$$e_i e_j + e_j e_i = 2\delta_{ij} \tag{6.42}$$

This is, of course, the algebra of "gamma matrices" and the Clifford algebras are intimately related with spinors and spin representations. If $\kappa = \mathbb{R}$ the best we can do is choose a basis so that

$$e_i e_j + e_j e_i = 2\eta_i \delta_{ij} \tag{6.43}$$

where $\eta_i \in \{\pm 1\}$. We note that if V is finite-dimensional then, as a vector space

$$C\ell(Q) \cong \oplus_{j=0}^{d} \Lambda^j V \tag{6.44}$$

but this isomorphism is completely false as algebras. (Why?) Clifford algebras are discussed in great detail in Chapter 10. For much more information about this crucial case see

1. The classic book: C. Chevalley, *The algebraic theory of spinors*

2. The chapters by P. Deligne in the AMS books on Strings and QFT for mathematicians.

3. Any number of online lecture notes. Including my own:

http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf (Chapter 13)

http://www.physics.rutgers.edu/∼gmoore/PiTP-LecturesA.pdf (Section 2.3)

4. *Universal enveloping algebra.* Let V be a Lie algebra. (See Chapter 8(?) below.) Then $U(V)$, the universal enveloping algebra is TV/I where I is the ideal generated by $v \otimes w - w \otimes v - [v, w]$.

Remarks

1. Explain coalgebra structures.

2. Explain about A_∞ and L_∞ algebra.

6.2.1 An Approach To Noncommutative Geometry

One useful way of thinking about the symmetric algebra is that $S^\bullet V^\vee$ is the algebra of polynomial functions on V. Note that there is a natural evaluation map

$$S^k(V^\vee) \times V \to \kappa \tag{6.45}$$

defining a polynomial function on V. To make this quite explicit choose a basis $\{v_i\}$ for V so that there is canonically a dual basis $\{\hat{v}^i\}$ for $V^\vee$. Then an element of $S^k(V^\vee)$ is given by a totally symmetric tensor $T_{i_1 \cdots i_k} \hat{v}^{i_1} \cdots \hat{v}^{i_k}$, and, when evaluated on a general element $x^i v_i$ of V we get the number

$$T_{i_1 \cdots i_k} x^{i_1} \cdots x^{i_k} \tag{6.46}$$

so the algebraic structure is just the multiplication of polynomials.

Now, quite generally, *derivation* of an algebra A is a linear map $D : A \to A$ which obeys the Leibniz rule

$$D(ab) = D(a)b + aD(b) \tag{6.47}$$

In differential geometry, derivations arise naturally from vector fields and indeed, the general derivation of the symmetric algebra $S^\bullet(V^\vee)$ is given by a vector field

$$D = \sum_i f^i(x) \frac{\partial}{\partial x^i} \tag{6.48}$$

Now these remarks give an entree into the subject of *noncommutative geometry*: We can also speak of derivations of the tensor algebra $T^\bullet V$. Given the geometrical interpretation of $S^\bullet V$ it is natural to consider these as functions on a *noncommutative manifold.* We could, for example, introduce formal variables x^i which do not commute and still consider functions of these. Then, vector fields on this noncommutative manifold would simply be derivations of the algebra. In general, if noncommutative geometry the name of the game is to replace geometrical concepts with equivalent algebraic concepts using commutative rings (or fields) and then generalize the algebraic concept to noncommutative rings.

Remarks:

1. A very mild form of noncommutative geometry is known as *supergeometry.* We discuss that subject in detail in Section §23 below. For now, note that our discussion of derivations and vector fields has a nice extension to the exterior algebras.

7. Kernel, Image, and Cokernel

A linear transformation between vector spaces (or R-modules)

$$T : V \to W \tag{7.1}$$

is a homomorphism of abelian groups and so, as noted before, there are automatically three canonical vector spaces:

$$\mathrm{ker}T := \{v : Tv = 0\} \subset V. \tag{7.2}$$

$$\mathrm{im}T := \{w : \exists v \qquad w = Tv\} \subset W. \tag{7.3}$$

and

$$\mathrm{im}T \cong V/\mathrm{ker}T \tag{7.4}$$

One defines exact sequences and short exact sequences exactly as for groups.

In linear algebra, since the target W is abelian we can make one new construction not available for general groups. We define the *cokernel* of T to be the vector space:

$$\mathrm{cok}T := W/\mathrm{im}T. \tag{7.5}$$

Remark: If V, W are inner product spaces then $\mathrm{cok}T \cong \mathrm{ker}T^{\dagger}$. (See definition and exercise below.)

Exercise
a.) If

$$0 \to V_1 \to V_2 \to V_3 \to 0 \tag{7.6}$$

is a short exact sequence of vector spaces, then $V_3 \cong V_2/V_1$.

b.) Show that a splitting of this sequence is equivalent to a choice of complementary subspace to V_1 in V_2.

Exercise *The vectorfield propagator*
Let $k_\mu, \mu = 1, \dots, d$ be a nonvanishing vector in $\mathbb{R}^d$ with its canonical Euclidean metric.
a.) Compute the rank and kernel of

$$M_{\mu\nu}(k) := k_\mu k_\nu - \delta_{\mu\nu}\vec{k}^2 \tag{7.7}$$

b.) Compute an inverse on the orthogonal complement (in the Euclidean metric) of the kernel.

Exercise
Show that if $T : V \to W$ is *any* linear operator then there is an exact sequence

$$0 \to \mathrm{ker}T \to V \stackrel{T}{\to} W \to \mathrm{cok}T \to 0 \tag{7.8}$$

7.1 The index of a linear operator

If $T : V \to W$ is any linear operator such that $\mathrm{ker}T$ and $\mathrm{cok}T$ are finite dimensional we define the *index of the operator* T to be:

$$\mathrm{Ind}(T) := \mathrm{dimcok}T - \mathrm{dimker}T \tag{7.9}$$

For V and W finite dimensional vector spaces you can easily show that

$$\mathrm{Ind}T = \mathrm{dim}W - \mathrm{dim}V \tag{7.10}$$

Notice that, from the LHS, we see that it does not depend on the details of T!

As an example, consider the *family* of linear operators $T_\lambda : v \to \lambda v$ for $\lambda \in \mathbb{C}$. Note that for $\lambda \neq 0$

$$\mathrm{ker}T = \{0\} \qquad \mathrm{cok}T = V/V \cong \{0\} \tag{7.11}$$

but for $\lambda = 0$

$$\mathrm{ker}T = \mathrm{cok}T = V \tag{7.12}$$

Both the kernel and cokernel change, but the index remains *invariant.*

As another example consider the family of operators $T_\lambda : \mathbb{C}^2 \to \mathbb{C}^2$ given in the standard basis by:

$$T_\lambda = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{7.13}$$

eq:fmly

Now, for $\lambda \neq 0$

$$\mathrm{ker}T_\lambda = \{0\} \qquad \mathrm{cok}T_\lambda = \mathbb{C}^2/\mathbb{C}^2 \cong \{0\} \tag{7.14}$$

but for $\lambda = 0$

$$\mathrm{ker}T_0 = \mathbb{C} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathrm{cok}T_0 = \mathbb{C}^2/\mathbb{C} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{7.15}$$

and again the index is invariant.

In infinite dimensions one must be more careful, but one can define the index of a *Fredholm operator* between two Hilbert spaces. The main theorem of index theory is then that if T_λ is a *continuous* family of Fredholm operators (in the norm operator topology) then the index is continuous, and hence, being an integer, is independent of λ.

The index of an operator plays an essential role in modern mathematical physics. To get an idea of why it is important, notice that the RHS of the above formula does not refer to the details of T, and yet provides some information on the number of zero eigenvalues of T and $T^\dagger$. It turns out that sometimes one can find formulae for the index of operators on Hilbert space. These are especially beautiful for problems arising from geometry, such as the Atiyah-Singer index theorem.

Exercise

Compute the index of the family of operators:

$$T_u = \begin{pmatrix} u & u & u^2 \\ \sin(u) & \sin(u) & \sin^2(u) \end{pmatrix} \tag{7.16}$$

Find special values of u where the kernel and cokernel jump.

Exercise

Consider the usual harmonic oscillator operators $a, a^\dagger$ acting on a separable Hilbert space. (See below.) Compute the index of a and $a^\dagger$.

Consider the families of operators $T = \lambda a$ and $T = \lambda a^\dagger$ for $\lambda \in \mathbb{C}$. Is the index invariant?

References on the Atiyah-Singer theorem:
For physicists: Eguchi, Gilkey, Hanson, Physics Reports; Nakahara.
For mathematicians: Michelson and Lawson, *Spin Geometry*

8. A Taste of Homological Algebra

A module M over a ring R is said to be $\mathbb{Z}$-*graded* if it is a direct sum over modules

$$M = \oplus_{n\in\mathbb{Z}} M^n \tag{8.1}$$

Physicists should think of the grading as a charge of some sort in some kind of "space of quantum states."

If the ring is $\mathbb{Z}$-graded then also

$$R = \oplus_{n\in\mathbb{Z}} R^n \tag{8.2}$$

and moreover $R^n \cdot R^m \subset R^{n+m}$. In this case it is understood that the ring action on the module is also $\mathbb{Z}$-graded so

$$R^n \times M^{n'} \to M^{n+n'} \tag{8.3}$$

A *cochain complex* is a $\mathbb{Z}$-graded module together with a linear operator usually called $d : M \to M$ or $Q : M \to M$ where $Q^2 = 0$ and Q is "of degree one." What this means is that Q increases the degree by one, so Q take M^n into M^{n+1}. Put differently, we can introduce an operator F which is multiplication by n on M^n (for example, in physics it might be "fermion number") and then we require that

$$[Q, F] = Q \tag{8.4}$$

A chain complex is usually indicated as a sequence of modules M^n and linear maps d^n:

$$\cdots \longrightarrow M^{n-1} \xrightarrow{d^{n-1}} M^n \xrightarrow{d^n} M^{n+1} \longrightarrow \cdots \tag{8.5}$$

with the crucial property that

$$\forall n \in \mathbb{Z} \qquad d^n d^{n-1} = 0 \tag{8.6}$$

The sequence might be infinite or finite on either side. If it is finite on both sides then $M^n = 0$ for all but finitely many n.

It is important to stress that the cochain complex is *not* an exact sequence. Indeed, we can characterize how much it differs from being an exact sequence by introducing its *cohomology*:

$$H(M, d) := \mathrm{ker}d/\mathrm{im}d \tag{8.7}$$

This definition makes sense, because $d^2 = 0$. Because the complex is $\mathbb{Z}$-graded, and d has degree one the cohomology is also $\mathbb{Z}$-graded and we can define

$$H^n(M, d) := \mathrm{ker}d^n/\mathrm{im}d^{n-1} \tag{8.8}$$

When one has a cochain complex one can always take the dual complex to obtain a *chain complex.* Now d has degree -1. Letting $M_n = \mathrm{Hom}(M^n, R)$ we have

$$\cdots \longleftarrow M_{n-1} \xleftarrow[d_n]{} M_n \xleftarrow[d_{n+1}]{} M_{n+1} \longleftarrow \cdots \tag{8.9}$$

and again $d_n d_{n+1} = 0$. The *homology* groups are ker$d$/im$d$ and are $\mathbb{Z}$-graded so

$$H_n(M_\bullet, d) = \mathrm{ker}d_n/\mathrm{im}d_{n+1} \tag{8.10}$$

8.1 The Euler-Poincaré principle

Suppose we have a finite cochain complex of vector spaces which we will assume begins at degree zero. Thus

$$0 \longrightarrow V^0 \xrightarrow{d^0} V^1 \xrightarrow{d^1} \cdots \xrightarrow{d^{n-1}} V^n \longrightarrow 0 \tag{8.11}$$

with $d^{j+1} d^j = 0$.

Let us compare the dimensions of the spaces in the cochain complex to the dimensions of the cohomology groups. We do this by introducing a *Poincare polynomial*

$$P_{V^\bullet}(t) = \sum_{j=0}^{n} t^j \mathrm{dim} V^j \tag{8.12}$$

$$P_{H^\bullet}(t) = \sum_{j=0}^{n} t^j \mathrm{dim} H^j \tag{8.13}$$

Now we claim that

$$P_{V^\bullet}(t) - P_{H^\bullet}(t) = (1+t)W(t) \tag{8.14}$$ eq:MorseIn

where $W(t)$ is a polynomial in t with positive integer coefficients.

Proof: Let h^j be the dimension of the cohomology. Then by choosing representatives and complementary spaces we can find subspaces $W^j \subset V^j$ of dimension w^j so that

$$\begin{aligned} \mathrm{dim} V^0 &= h^0 + w^0 \\ \mathrm{dim} V^1 &= h^1 + w^0 + w^1 \\ \mathrm{dim} V^2 &= h^2 + w^1 + w^2 \\ \vdots &\quad \vdots \\ \mathrm{dim} V^n &= h^n + w^{n-1} \end{aligned} \tag{8.15}$$

so $W(t) = w^0 + w^1 t + \cdots + w^{n-1} t^{n-1}$.

Putting $t = -1$ we have the beautiful *Euler-Poincaré principle*:

$$\sum (-1)^i \mathrm{dim} V^i = \sum (-1)^i \mathrm{dim} H^i \tag{8.16}$$ eq:EulerChar

This common integer is called the *Euler characteristic* of the complex.

Remark: These observations have their origin in topology and geometry. Equation (8.14) is related to *Morse inequalities*, while (8.16) is related, among many other things, to the fact that one can compute topological invariants of topological spaces by counting simplices in a simplicial decomposition.

8.2 Chain maps and chain homotopies

Chain map $f : C \to D$ of degree zero and commutes with d.

Diagram

Therefore there is an induced map $f^* : H(D) \to H(D)$. $f^*([c]) = [f(c)]$.

This defines a "morphism of complexes"

Homotopy of chain maps: $T : C \to D$ of degree -1 so that

$$[d, T] = f_1 - f_2 \tag{8.17}$$

Diagram.

If there is a chain homotopy then $f_1^* = f_2^*$.

Special case: If there is a chain homotopy with $[d, T] = 1$ then the cohomology is trivial. This can be a useful way to show that cohomology groups vanish.

8.3 Exact sequences of complexes

Define SES of complexes. Then give the LES and describe the connecting homomorphism.

8.4 Left- and right-exactness

Behavior of exact sequences under Hom and $\otimes$.

Definition of Ext and Tor.

9. Relations Between Real, Complex, And Quaternionic Vector Spaces

ec:RealComplex

Physical quantities are usually expressed in terms of real numbers. Thus, for example, we think of the space of electric and magnetic fields in terms of real numbers. In quantum field theory more generally one often works with real vector spaces of fields. On the other hand, quantum mechanics urges us to use the complex numbers. One could formulate quantum mechanics using only the real numbers, but it would be terribly awkward to do so. In physics and mathematics it is often important to have a firm grasp of the relation between complex and real structures on vector spaces, and this section explains that in excruciating detail.

9.1 Complex structure on a real vector space

Definition Let V be a real vector space. A *complex structure* on V is a linear map $I : V \to V$ such that $I^2 = -1$.

Choose a squareroot of -1 and denote it i. If V is a real vector space with a complex structure I, then we can define an associated complex vector space (V, I). We take (V, I) to be identical with V, *as sets*, but define the scalar multiplication of a complex number $z \in \mathbb{C}$ on a vector v by

$$z \cdot v := x \cdot v + I(y \cdot v) = x \cdot v + y \cdot I(v) \tag{9.1}$$

where $z = x + iy$ with $x, y \in \mathbb{R}$.

We note that for any nonzero vector $v \in V$, the vectors v and $I(v)$ are linearly independent over $\mathbb{R}$. To prove this, suppose that on the contrary there are nonzero scalars $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha v + \beta I(v) = 0 \tag{9.2}$$

Then applying I to this equation and using $I^2 = -1$ we get

$$\beta v - \alpha I(v) = 0 \tag{9.3}$$

but since $\alpha\beta \neq 0$ this implies $\beta = \lambda\alpha$ and $-\alpha = \lambda\beta$ for some real number λ and hence $-\alpha\beta = \lambda^2\alpha\beta$, but this is a contradiction. It follows that if V is finite dimensional then its dimension must be even and there is always a real basis for V in which I takes the form

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9.4}$$

where the blocks are $n \times n$ for an integer $n = \frac{1}{2}\dim_{\mathbb{R}} V$. That is, we have the simple:

Lemma If I is any $2n \times 2n$ *real* matrix which squares to -1_{2n} then there is $S \in GL(2n, \mathbb{R})$ such that

$$SIS^{-1} = I_0 := \begin{pmatrix} 0 & -1_n \\ 1_n & 0 \end{pmatrix} \tag{9.5}$$

eq:CanonCS

We leave the proof as an exercise below. Thus, there can be many different complex structures on a real vector space, because if I is a complex structure, then so is SIS^{-1}. However, they can all be related one to the other by a change of basis.

Note that our lemma also shows that if V is finite dimensional and has a complex structure then the dimension of the complex vector space (V, I) is:

$$\dim_{\mathbb{C}}(V, I) = \frac{1}{2} \dim_{\mathbb{R}} V \tag{9.6}$$

eq:half-dim

Note carefully that, while v and $I(v)$ are linearly independent in the real vector space V, they are *linearly dependent* in the complex vector space (V, I) since

$$iv + (-1) \cdot I(v) = 0 \tag{9.7}$$

Example Consider the real vector space $V = \mathbb{R}^2$. Let us choose

$$I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9.8}$$

Then multiplication of the complex scalar $z = x + iy$, with $x, y \in \mathbb{R}$ on a vector $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \in \mathbb{R}^2$ can be defined by:

$$(x + iy) \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} a_1 x - a_2 y \\ a_1 y + a_2 x \end{pmatrix} \tag{9.9}$$

eq:cplxstrone

By equation (9.6) this must be a one-complex dimensional vector space, so it should be isomorphic to $\mathbb{C}$ as a complex vector space. Indeed this is the case. Define $\Psi : (V, I) \to \mathbb{C}$ by

$$\Psi : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \mapsto a_1 + ia_2 \tag{9.10}$$

Then one can check (exercise!) that this is an isomorphism of complex vector spaces.

Quite generally, if I is a complex structure then so is $\tilde{I} = -I$. So what happens if we take our complex structure to be instead:

$$\tilde{I} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{9.11}$$

Now the rule for multiplication by a complex number in $(V, \tilde{I})$ is

$$(x + iy) \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} := \begin{pmatrix} a_1 x + a_2 y \\ -a_1 y + a_2 x \end{pmatrix} \tag{9.12}$$

eq:cplxstronep

Now one can check that $\tilde{\Psi} : (V, \tilde{I}) \to \mathbb{C}$ defined by

$$\tilde{\Psi} : \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \mapsto a_1 - i a_2 \tag{9.13}$$

is also an isomorphism of complex vector spaces. (Check carefully that $\tilde{\Psi}(z\vec{a}) = z\tilde{\Psi}(\vec{a})$.)

How are these two constructions related? Note that if we introduce the real linear operator

$$C := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{9.14}$$

then $C^2 = 1$ and

$$CIC^{-1} = CIC = -I \tag{9.15}$$

What can we say about the set of all complex structures on $\mathbb{R}^2$? It helps to distinguish those that are compatible with the Euclidean metric

$$\parallel \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \parallel^2 := a_1^2 + a_2^2 \tag{9.16}$$

from those that are not. "Compatible" means that $\parallel I(v) \parallel^2 = \parallel v \parallel^2$. In other words, I is an orthogonal transformation. One easily checks that, if S is an orthogonal transformation then

$$SI_0S^{-1} = SI_0S^{tr} = (\det S)I_0 \tag{9.17}$$

so there are precisely two complex structures compatible with the Euclidean norm: I_0 and $-I_0$.

More generally, if we speak of all complex structures then we invoke a theorem (easily proven - see remark below) that says that every $S \in GL(2, \mathbb{R})$ can be written as:

$$S = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\xi \sin(\theta) & \xi \cos(\theta) \end{pmatrix} \tag{9.18}$$ eq:KAN-GL2

with $x \in \mathbb{R}$, $\lambda_1, \lambda_2 > 0$, $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ and, finally, $\xi \in \{\pm 1\}$ is a sign that coincides with the sign of $\det(S)$ and tells us which of the two connected components of $GL(2, \mathbb{R})$ S sits in. We then compute that

$$\begin{aligned} SI_0S^{-1} &= \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \xi\lambda_1/\lambda_2 \\ -\xi\lambda_2/\lambda_1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -x\alpha & \alpha^{-1} + x^2\alpha \\ -\alpha & x\alpha \end{pmatrix} \end{aligned} \tag{9.19}$$

(where $\alpha = \xi\lambda_2/\lambda_1$ so $\alpha \in \mathbb{R}^*$ and $x \in \mathbb{R}$). So the space of complex structures (not necessarily compatible with the Euclidean metric) is a two-dimensional manifold with two connected components.

In section 9.6 below we describe the generalization of this result to the space complex structures on any finite dimensional real vector space.

Remark: The decomposition (9.18) is often very useful. Indeed, it generalizes to $GL(n,\mathbb{R})$, where it says we can write any $g \in GL(n,\mathbb{R})$ as $g = nak$ where $k \in O(n)$, a is diagonal with positive entries, and $n = 1+t$ where t is strictly upper triangular. This is just a statement of the Gram-Schmidt process. See section 21.2 below. This kind of decomposition generalizes to all algebraic Lie groups, and it is this more general statement that constitutes the "KAN theorem."

Exercise *Canonical form for a complex structure*
Prove equation (9.5) above. [7]

9.2 Real Structure On A Complex Vector Space

Given a complex vector space V can we produce a real vector space? Of course, by restriction of scalars, if V is complex, then it is also a real vector space, which we can call $V_{\mathbb{R}}$. V and $V_{\mathbb{R}}$ are the same as *sets* but in $V_{\mathbb{R}}$ the vectors v and iv, are linearly independent (they are clearly not linearly independent in V!). Thus:

$$\dim_{\mathbb{R}} V_{\mathbb{R}} = 2\dim_{\mathbb{C}} V. \tag{9.23}$$

There is another way we can get real vector spaces out of complex vector spaces. A *real structure* on a complex vector V space produces a different real vector space of half the real dimension of $V_{\mathbb{R}}$, that is, a vector space of real dimension equal to the complex dimension of V.

[7] *Answer*: Proceed by induction. If $V \neq 0$ then there is a nonzero vector $v \in V$ and hence $\langle v, I(v)\rangle \subset V$ is a nonzero two-dimensional subspace. If $\dim_{\mathbb{R}} V = 2$ we are done. If $\dim_{\mathbb{R}} V = 2n+2$ and we assume the result is true up to dimension $2n$ then we consider the quotient space

$$V/\langle v, I(v)\rangle \tag{9.20}$$

and we note that I descends to a map $\tilde{I}$ on this quotient space and by the induction hypothesis there is a basis $[w_1], \dots, [w_{2n}]$ for $V/\langle v, I(v)\rangle$ such that

$$\begin{aligned} \tilde{I}([w_i]) &= [w_{n+i}] \\ \tilde{I}([w_{n+i}]) &= -[w_i] \qquad i = 1, \dots, n \end{aligned} \tag{9.21}$$

Now, choosing specific representatives w_i we know there are scalars $\alpha_1^i, \dots, \beta_2^i$ such that

$$\begin{aligned} \tilde{I}(w_i) &= w_{n+i} + \alpha_1^i v + \beta_1^i I(v) \\ \tilde{I}([w_{n+i}]) &= -w_i + \alpha_2^i v + \beta_2^i I(v) \qquad i = 1, \dots, n \end{aligned} \tag{9.22}$$

and consistency of these equations with $I^2 = -1$ implies $\alpha_2^i = \beta_1^i$ and $\beta_2^i = -\alpha_1^i$. Now check that $\tilde{w}_i := w_i + \alpha_1^i I(v)$ and $\tilde{w}_{n+i} = w_{n+i} + \beta_1^i I(v)$ is the suitable basis in which I takes the desired form.

Definition An *antilinear* map $\mathcal{T} : V \to V$ on a complex vector space V satisfies

1. $\mathcal{T}(v + v') = \mathcal{T}(v) + \mathcal{T}(v')$,
2. $\mathcal{T}(\alpha v) = \alpha^* \mathcal{T}(v)$ where $\alpha \in \mathbb{C}$ and $v \in V$.

Note that $\mathcal{T}$ is a linear map on the underlying real vector space $V_{\mathbb{R}}$.

Definition Suppose V is a complex vector space. Then a *real structure* on V is an antilinear map $\mathcal{C} : V \to V$ such that $\mathcal{C}^2 = +1$.

If $\mathcal{C}$ is a real structure on a complex vector space V then we can define *real vectors* to be those such that

$$\mathcal{C}(v) = v \tag{9.24}$$

Let us call the set of such real vectors V_+. This set is a real vector space, but it is not a complex vector space, because $\mathcal{C}$ is antilinear. Indeed, if $\mathcal{C}(v) = +v$ then $\mathcal{C}(iv) = -iv$. If we let V_- be the *imaginary vectors*, for which $\mathcal{C}(v) = -v$ then we claim

$$V_{\mathbb{R}} = V_+ \oplus V_- \tag{9.25}$$

The proof is simply the isomorphism

$$v \mapsto \left(\frac{v + \mathcal{C}(v)}{2}\right) \oplus \left(\frac{v - \mathcal{C}(v)}{2}\right) \tag{9.26}$$

Moreover multiplication by i defines an isomorphism of real vector spaces: $V_+ \cong V_-$. Thus we have

$$\dim_{\mathbb{R}} V_+ = \dim_{\mathbb{C}} V \tag{9.27}$$

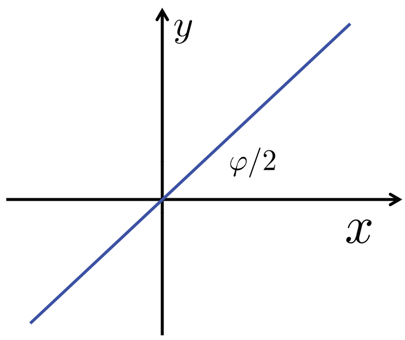

Figure 3: The real structure $\mathcal{C}$ has fixed vectors given by the blue line. This is a real vector space determined by the real structure $\mathcal{C}$.

fig:REALVECTOR

Example Take $V = \mathbb{C}$, and let $\varphi \in \mathbb{R}/2\pi\mathbb{Z}$ and define:

$$\mathcal{C} : x + iy \to e^{i\varphi}(x - iy) \tag{9.28}$$

The fixed vectors under $\mathcal{C}$ consist of the real line at angle $\varphi/2$ to the x-axis as shown in Figure 3. Note that the lines are not oriented so the ambiguity in the sign of $e^{i\varphi/2}$ does not matter.

Once again, note that there can be many distinct real structures on a given complex vector space. The space of such real structures is discussed in Section 9.6 below.

In general, if V is a finite dimensional complex vector space, and if we choose any basis (over $\mathbb{C}$) $\{v_i\}$ for V, then we can define a real structure:

$$\mathcal{C}(\sum_i z_i v_i) = \sum_i \bar{z}_i v_i \tag{9.29}$$

and thus

$$V_+ = \{\sum a_i v_i | a_i \in \mathbb{R}\} \tag{9.30}$$

Remark: Antilinear operators famously appear in quantum mechanics when dealing with symmetries which reverse the orientation of time. In condensed matter physics they also appear as "particle-hole symmetry operators." (This is a little confusing since in relativistic particle physics the charge conjugation operator is complex linear.)

Exercise *Antilinear maps from the real point of view*

Suppose W is a real vector space with complex structure I giving us a complex vector space (W, I).

Show that an antilinear map $\mathcal{T} : (W, I) \to (W, I)$ is the same thing as a real linear transformation $T : W \to W$ such that

$$TI + IT = 0 \tag{9.31}$$

9.2.1 Complex Conjugate Of A Complex Vector Space

There is another viewpoint on what a real structure is which can be very useful. If V is a complex vector space then we can, canonically, define another complex vector space $\bar{V}$. We begin by declaring $\bar{V}$ to be the same *set*. Thus, for every vector $v \in V$, the same vector, regarded as an element of $\bar{V}$ is simply written $\bar{v}$. However, $\bar{V}$ is different from V *as a complex vector space* because we alter the vector space structure by altering the rule for scalar multiplication by $\alpha \in \mathbb{C}$:

$$\alpha \cdot \bar{v} := \overline{\alpha^* \cdot v} \tag{9.32}$$ eq:conjvs

where α^* is the complex conjugate in $\mathbb{C}$.

Of course $\bar{\bar{V}} = V$.

Note that, given any $\mathbb{C}$-linear map $T : V \to W$ between complex vector spaces there is, canonically, a $\mathbb{C}$-linear map

$$\bar{T} : \bar{V} \to \bar{W} \tag{9.33}$$

defined by

$$\bar{T}(\bar{v}) := \overline{T(v)} \tag{9.34}$$

With the notion of $\bar{V}$ we can give an alternative definition of an *anti-linear map*: An anti-linear map $\mathcal{T} : V \to V$ is the same as a $\mathbb{C}$-linear map $T : V \to \bar{V}$, related by

$$\mathcal{T}(v) = \overline{T(v)} \tag{9.35}$$

Similarly, we can give an alternative definition of a real structure on a complex vector space V as a $\mathbb{C}$- *linear* map

$$C : V \to \bar{V} \tag{9.36}$$

such that $C\bar{C} = 1$ and $\bar{C}C = 1$, where $\bar{C} : \bar{V} \to V$ is canonically determined by C as above. In order to relate this to the previous viewpoint note that $\mathcal{C} : v \mapsto \bar{C}(\bar{v})$ is an antilinear transformation $V \to V$ which squares to 1.

Exercise

A linear transformation $T : V \to W$ between two complex vector spaces with real structures C_V and C_W commutes with the real structures if the diagram

$$\begin{array}{ccc} V & \stackrel{T}{\to} & W \\ \downarrow C_V & & \downarrow C_W \\ \bar{V} & \stackrel{\bar{T}}{\to} & \bar{W} \end{array} \tag{9.37}$$

commutes.

Show that in this situation T defines an $\mathbb{R}$-linear transformation on the underlying real vector spaces: $T_+ : V_+ \to W_+$.

Exercise *Complex conjugate from the real point of view*

Suppose W is a real vector space with complex structure I so that we can form the complex vector space (W, I). Show that

$$\overline{(W, I)} = (W, -I) \tag{9.38}$$

9.2.2 Complexification

If V is a real vector space then we can define its *complexification* $V_{\mathbb{C}}$ by putting a complex structure on $V \oplus V$. This is simply the real linear transformation

$$I : (v_1, v_2) \mapsto (-v_2, v_1) \tag{9.39}$$ eq:cplx-def-1

and clearly $I^2 = -1$. This complex vector space $(V \oplus V, I)$ is known as the *complexification of* V. Another way to define the complexification of V is to take

$$V_{\mathbb{C}} := V \otimes_{\mathbb{R}} \mathbb{C}. \tag{9.40}$$ eq:cplx-def-2

Note that we are taking a tensor product of vector spaces over $\mathbb{R}$ to get a real vector space, but there is a natural action of the complex numbers on these vectors:

$$z \cdot (v \otimes z') := v \otimes zz' \tag{9.41}$$

making $V_{\mathbb{C}}$ into a complex vector space. In an exercise below you show that these two definitions are equivalent.

Note that

$$\dim_{\mathbb{C}} V_{\mathbb{C}} = \dim_{\mathbb{R}} V \tag{9.42}$$

Note that $V_{\mathbb{C}}$ has a canonical real structure. Indeed

$$\overline{V_{\mathbb{C}}} = V \otimes_{\mathbb{R}} \overline{\mathbb{C}} \tag{9.43}$$

and we can define $C : V_{\mathbb{C}} \to \overline{V_{\mathbb{C}}}$ by setting

$$C : v \otimes 1 \mapsto v \otimes \bar{1} \tag{9.44}$$

and extending by $\mathbb{C}$-linearity. Thus

$$\begin{aligned} C(v \otimes z) &= C(z \cdot (v \otimes 1)) && \text{def of } V_{\mathbb{C}} \\ &= z \cdot C((v \otimes 1)) && \mathbb{C} - \text{linear extension} \\ &= z \cdot (v \otimes \bar{1}) \\ &= v \otimes \overline{z^*} && \text{definition of scalar action on} \bar{V}_{\mathbb{C}} \end{aligned} \tag{9.45}$$

Finally, it is interesting to ask what happens when one begins with a *complex* vector space V and then complexifies the underlying real space $V_{\mathbb{R}}$. If V is complex then we claim there is an isomorphism of complex vector spaces:

$$(V_{\mathbb{R}})_{\mathbb{C}} \cong V \oplus \bar{V} \tag{9.46}$$ eq:complxi

Proof: The vector space $(V_{\mathbb{R}})_{\mathbb{C}}$ is, by definition, generated by the space of pairs (v_1, v_2), $v_i \in V_{\mathbb{R}}$ with complex structure defined by $I : (v_1, v_2) \to (-v_2, v_1)$. Now we map:

$$\psi : (v_1, v_2) \mapsto (v_1 + iv_2) \oplus (v_1 - iv_2) \tag{9.47}$$

and compute

$$(x + Iy) \cdot (v_1, v_2) = (xv_1 - yv_2, xv_2 + yv_1) \tag{9.48}$$

so

$$\psi : z \cdot (v_1, v_2) \mapsto (x + iy) \cdot (v_1 + iv_2) \oplus (x - iy) \cdot (v_1 - iv_2) = z \cdot v + \bar{z} \cdot \bar{v} \tag{9.49}$$

Another way to look at (9.46) is as follows. Let V be a real vector space with complex structure I. Now consider $V_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$. There are now two ways of multiplying by a complex number $z = x + iy$: We can multiply the second factor $\mathbb{C}$ by z or we could operate on the first factor with $x + Iy$. We can decompose our space $V \otimes_{\mathbb{R}} \mathbb{C}$ into eigenspaces where $I = +i$ and $I = -i$ using the projection operators

$$P_{\pm} = \frac{1}{2}(1 \mp I \otimes i) \tag{9.50}$$

the images of these projection operators have complex structures, and are isomorphic to V and $\bar{V}$ as complex vector spaces, respectively.

Exercise *Equivalence of two definitions*

Show that the two definitions (9.39) and (9.40) define canonically isomorphic complex vector spaces.

Exercise

Show that

$$\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} = \mathbb{C} \oplus \mathbb{C} \tag{9.51}$$

$$\mathbb{C} \otimes_{\mathbb{C}} \mathbb{C} \cong \mathbb{C} \tag{9.52}$$

as algebras.

Exercise

Suppose V is a complex vector space with a real structure C and that V_+ is the real vector space of fixed points of C.

Show that, as complex vector spaces

$$V \cong V_+ \otimes_{\mathbb{R}} \mathbb{C}. \tag{9.53}$$

9.3 The Quaternions

Definition The *quaternion algebra* $\mathbb{H}$ is the algebra over $\mathbb{R}$ with generators $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$ satisfying the relations

$$\mathfrak{i}^2 = -1 \qquad \mathfrak{j}^2 = -1 \qquad \mathfrak{k}^2 = -1 \tag{9.54}$$

$$\mathfrak{k} = \mathfrak{i}\mathfrak{j} \tag{9.55}$$

Note that, as a consequence of these relations we have

$$\mathfrak{i}\mathfrak{j} + \mathfrak{j}\mathfrak{i} = \mathfrak{i}\mathfrak{k} + \mathfrak{k}\mathfrak{i} = \mathfrak{j}\mathfrak{k} + \mathfrak{k}\mathfrak{j} = 0 \tag{9.56}$$

The quaternions form a four-dimensional algebra over $\mathbb{R}$. As a vector space we can write

$$\mathbb{H} = \mathbb{R}\mathfrak{i} + \mathbb{R}\mathfrak{j} + \mathbb{R}\mathfrak{k} + \mathbb{R} \cong \mathbb{R}^4 \tag{9.57}$$

and in the middle we have an internal direct sum. The algebra is associative, but noncommutative. It has a rich and colorful history. See the remark below.

Note that if we denote a generic quaternion by

$$q = x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k} + x_4 \tag{9.58}$$

then we can define the conjugate quaternion by the equation

$$\bar{q} := -x_1\mathfrak{i} - x_2\mathfrak{j} - x_3\mathfrak{k} + x_4 \tag{9.59}$$

Now using the relations we compute

$$q\bar{q} = \bar{q}q = \sum_{\mu=1}^{4} x_\mu x_\mu \in \mathbb{R}_+ \tag{9.60}$$

This defines a norm on the quaternioni algebra:

$$\parallel q \parallel^2 := q\bar{q} = \bar{q}q \tag{9.61}$$

Note that the *unit quaternions*, i.e. those with $\parallel q \parallel = 1$ form a nonabelian group. (Exercise: Prove this!) In the exercises below you show that this group is isomorphic to $SU(2)$.

One fact about the quaternions that is often quite useful is the following. There is a left- and right-action of the quaternions on itself. If $\mathfrak{q}$ is a quaternion then we can define $L(\mathfrak{q}) : \mathbb{H} \to \mathbb{H}$ by

$$L(\mathfrak{q}) : \mathfrak{q}' \mapsto \mathfrak{q}\mathfrak{q}' \tag{9.62}$$

and similarly there is a right-action

$$R(\mathfrak{q}) : \mathfrak{q}' \mapsto \mathfrak{q}'\mathfrak{q} \tag{9.63}$$

The algebra of operators $L(\mathfrak{q})$ is isomorphic to $\mathbb{H}$ and the algebra of operators $R(\mathfrak{q})$ is isomorphic to $\mathbb{H}^{\rm opp}$, which in turn is isomorphic to $\mathbb{H}$ itself. Now $\mathbb{H}$ is a four-dimensional real vector space and $L(\mathfrak{q})$ and $R(\mathfrak{q})$ are commuting real-linear operators. Therefore there is an inclusion

$$\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\rm opp} \hookrightarrow {\rm End}_{\mathbb{R}}(\mathbb{H}) \cong {\rm End}(\mathbb{R}^4) \tag{9.64}$$

Since $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\rm opp}$ has real dimension 16 this is isomorphism of algebras over $\mathbb{R}$.

Remarks:

1. It is very interesting to compare $\mathbb{C}$ and $\mathbb{H}$.

   - We obtain $\mathbb{C}$ from $\mathbb{R}$ by introducing a squareroot of -1, $\mathfrak{i}$, but we obtain $\mathbb{H}$ from $\mathbb{R}$ by introducing three squareroots of -1, $\mathfrak{i}, \mathfrak{j}, \mathfrak{k}$.
   - $\mathbb{C}$ is commutative and associative, but $\mathbb{H}$ is noncommutative and associative.
   - There is a conjugation operation $z \to \bar{z}$ and $q \to \bar{q}$ so that $z\bar{z} \in \mathbb{R}_+$ and $q\bar{q} = \bar{q}q \in \mathbb{R}_+$ and is zero iff $z = 0$ or $q = 0$, respectively.
   - The set of all square-roots of -1 in $\mathbb{C}$ is $\{\pm\mathfrak{i}\}$ is a sphere S^0. The set of all square-roots of -1 in $\mathbb{H}$ is a sphere S^2.
   - The group of unit norm elements in $\mathbb{C}$ is the Abelian group $U(1)$. The group of unit norm elements in $\mathbb{H}$ is the Abelian group $SU(2)$.

2. *A little history*

   There is some colorful history associated with the quaternions which mathematicians are fond of relating. My sources for the following are

   1. E.T. Bell, *Men of Mathematics* and C.B. Boyer, *A History of Mathematics*.

   2. J. Baez, "The Octonions," Bull. Amer. Math. Soc. 39 (2002), 145-205. Also available at http://math.ucr.edu/home/baez/octonions/octonions.html

   The second of these is highly readable and informative.

   In 1833, W.R. Hamilton presented a paper in which, for the first time, complex numbers were explicitly identified with pairs (x, y) of real numbers. Hamilton stressed that the multiplication law of complex numbers could be written as:

$$(x, y)(u, v) = (xu - vy, vx + yu)$$

   and realized that this law could be interpreted in terms of rotations of vectors in the plane.

   Hamilton therefore tried to associate ordered triples (x, y, z) of real numbers to vectors in $\mathbb{R}^3$ and sought to discover a multiplication law which "expressed rotation" in $\mathbb{R}^3$. It seems he was trying to look for what we today call a normed division algebra. According to his account, for 10 years he struggled to define a multiplication law – that is, an algebra structure – on ordered 3-tuples of real numbers. I suspect most of his problem was that he didn't know what mathematical structure he was really searching for. This is a situation in which researchers in mathematics and physics often find themselves, and it can greatly confound and stall research. At least one of his stumbling blocks was the assumption that the algebra had to be commutative.

   Then finally – so the story goes – on October 16, 1843 he realized quite suddenly during a walk that if he *dropped the commutativity law* then he could write a consistent algebra in four dimensions, that is, he denoted $q = a + b\mathfrak{i} + c\mathfrak{j} + d\mathfrak{k}$ and realized that he should impose

$$\mathfrak{i}^2 = \mathfrak{j}^2 = \mathfrak{k}^2 = \mathfrak{i}\mathfrak{j}\mathfrak{k} \tag{9.65}$$

He already knew that $\mathfrak{i}^2 = -1$ was essential, so surely $\mathfrak{j}^2 = \mathfrak{k}^2 = -1$. Then $\mathfrak{i}^2 = \mathfrak{i}\mathfrak{j}\mathfrak{k}$ implies $\mathfrak{i} = \mathfrak{j}\mathfrak{k}$. But for this to square to one we need $\mathfrak{j}\mathfrak{k} = -\mathfrak{k}\mathfrak{j}$. Apparently he carved these equations into Brougham Bridge while his wife, and presumably, not the police, stood by. He lost no time, and then went on to announce his discovery to the Royal Irish Academy, the very same day. Hamilton would spend the rest of his life championing the quaternions as something of cosmic significance, basic to the structure of the universe and of foundational importance to physics. That is not the general attitude today: The quaternions fit very nicely into the structure of Clifford algebras, and they are particularly important in some aspects of four-dimensional geometry and the geometry of hyperkahler manifolds. However, in more general settings the vector analysis notation introduced by J. W. Gibbs at Yale has proved to be much more flexible and congenial to working with mathematical physics in general dimensions.

The day after his great discovery, Hamilton wrote a detailed letter to his friend John T. Graves explaining what he had found. Graves replied on October 26th, and in his letter he said:

" There is still something in the system which gravels me. I have not yet any clear views as to the extent to which we are at liberty arbitrarily to create imaginaries, and to endow them with supernatural properties. ... If with your alchemy you can make three pounds of gold, why should you stop there? "

By Christmass Graves had discovered the octonions, and in January 1844, had gone on to a general theory of "2^m-ions" but stopped upon running into an "unexpected hitch."

Later, again inspired by Hamilton's discovery of the quaternions, Arthur Cayley independently discovered the octonions in 1845.

See Chapter 7 below for a discussion of the octonions.

Exercise *Quaternionic Conjugation And The Opposite Quaternion Algebra*
a.) Show that quaternion conjugation is an anti-automorphism, that is:

$$\overline{q_1 q_2} = \overline{q_2}\ \overline{q_1} \tag{9.66}$$

b.) Show that $\mathbb{H}^{\rm opp} \cong \mathbb{H}$.

Exercise *Identities On Sums Of Squares*
a.) If z_1 and z_2 are complex numbers then

$$|z_1|^2 |z_2|^2 = |z_1 z_2|^2 \tag{9.67}$$

Show that this can be interpreted as an identity for sums of squares of real numbers x, y, s, t:

$$(x^2+y^2)(s^2+t^2) = (xs-yt)^2 + (ys+xt)^2 \tag{9.68}$$

b.) Show that

$$\| q_1 \|^2 \| q_2 \|^2 = \| q_1 q_2 \|^2 \tag{9.69}$$

and interpret this as an identity of the form

$$\left(\sum_{\mu=1}^{4}(x_\mu)^2\right)\left(\sum_{\mu=1}^{4}(y_\mu)^2\right) = \left(\sum_{\mu=1}^{4}(z_\mu)^2\right) \tag{9.70}$$

where z_μ is of the form

$$z_\mu = \sum_{\nu,\lambda=1}^{4} a_{\mu\nu\lambda} x_\nu y_\lambda \tag{9.71}$$

Find the explicit formula for z_μ.

Exercise *Quaternions And Rotations*

a.) Show that

$$\mathfrak{i} \to -\sqrt{-1}\sigma^1 \qquad \mathfrak{j} \to -\sqrt{-1}\sigma^2 \qquad \mathfrak{k} \to -\sqrt{-1}\sigma^3 \tag{9.72}$$

defines a set of 2×2 complex matrices satisfying the quaternion algebra. Under this mapping a quaternion q is identified with a 2×2 complex matrix

$$q \to \rho(q) = \begin{pmatrix} z & -\bar{w} \\ w & \bar{z} \end{pmatrix} \tag{9.73}$$ eq:2x2-quat

with $z = -\mathrm{i}(x_3 + \mathrm{i}x_4)$ and $w = -\mathrm{i}(x_1 + \mathrm{i}x_2)$.

b.) Show that the set of matrices in (9.73) may be characterized as the set of 2×2 complex matrices A so that

$$A^* = JAJ^{-1} \qquad\qquad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{9.74}$$

If we introduce the epsilon symbol $\epsilon^{\alpha\beta}$ which is totally antisymmetric and such that $\epsilon^{12} = 1$ (this is a choice) then we can write the condition as

$$(A_{\alpha\dot{\beta}})^* = \epsilon^{\alpha\gamma}\epsilon^{\dot{\beta}\dot{\delta}}A_{\gamma\dot{\delta}} \tag{9.75}$$

c.) Using the fact that $q\bar{q} = \bar{q}q$ is a real scalar show that the set of unit-norm elements in $\mathbb{H}$ is a group. Using the representation (9.73) show that this group is isomorphic to the group $SU(2)$.

d.) Define the *imaginary quaternions* to be the subspace of $\mathbb{H}$ of quaternions such that $\bar{q} = -q$. Denote this as $\Im(\mathbb{H})$. Show that $\Im(\mathbb{H}) \cong \mathbb{R}^3$, and show that $\Im(\mathbb{H})$ is in fact a real Lie algebra with

$$[q_1, q_2] := q_1 q_2 - q_2 q_1 \tag{9.76}$$

eq:ImagQuatLie

Using (9.73) identify this as the Lie algebra of $SU(2)$: $\Im(\mathbb{H}) \cong \mathfrak{su}(2)$. (Recall that $\mathfrak{su}(2)$ is the real Lie algebra of 2×2 traceless anti-Hermitian matrices.)

e.) Show that $\det(\rho(q)) = q\bar{q} = x_\mu x_\mu$ and use this to define a homomorphism

$$\rho : SU(2) \times SU(2) \to SO(4) \tag{9.77}$$

eq:DoubleCover

f.) Show that we have an exact sequence [8]

$$1 \to \mathbb{Z}_2 \to SU(2) \times SU(2) \to SO(4) \to 1 \tag{9.78}$$

where $\iota : \mathbb{Z}_2 \to SU(2) \times SU(2)$ takes the nontrivial element -1 to $(-1_2, -1_2)$.

g.) Show that under the homomorphism ρ the diagonal subgroup of $SU(2) \times SU(2)$ preserves the scalars and acts as the group $SO(3)$ of rotations of $\Im(\mathbb{H}) \cong \mathbb{R}^3$.

Exercise *Unitary Matrices Over Quaternions And Symplectic Groups*

a.) Show that the algebra $\mathrm{Mat}_n(\mathbb{H})$ of $n \times n$ matrices with quaternionic entries can be identified as the subalgebra of $\mathrm{Mat}_{2n}(\mathbb{C})$ of matrices A such that

$$A^* = JAJ^{-1} \qquad\qquad J = \begin{pmatrix} 0 & -1_n \\ 1_n & 0 \end{pmatrix} \tag{9.79}$$

b.) Show that the unitary group over $\mathbb{H}$:

$$U(n, \mathbb{H}) := \{u \in \mathrm{Mat}_n(\mathbb{H}) | u^\dagger u = 1\} \tag{9.80}$$

is isomorphic to

$$USp(2n) := \{u \in U(2n, \mathbb{C}) | u^* = JuJ^{-1}\} \tag{9.81}$$

To appreciate the notation show that matrices $u \in USp(2n)$ also satisfy

$$u^{tr} J u = J \tag{9.82}$$

which is the defining relation of $Sp(2n, \mathbb{C})$.

---

[8] *Answer*: We use the homomorphism $\rho : SU(2) \times SU(2) \to SO(4)$ defined by $\rho(q_1, q_2) : q \mapsto q_1 q \overline{q_2}$. To compute the kernel we search for pairs (q_1, q_2) of unit quaternions so that $q_1 q \overline{q_2} = q$ for all $q \in \mathbb{H}$. Applying this to $q = 1$ gives $q_1 = q_2$. Then applying it to $q = \mathfrak{i}, \mathfrak{j}, \mathfrak{k}$ we see that $q_1 = q_2$ must be a real scalar. The only such unit quaternions are $q_1 \in \{\pm 1\}$. To check the image is entirely in $SO(4)$ you can check this for diagonal matrices in $SU(2) \times SU(2)$ and then everything conjugate to these must be in $SO(4)$, so the image is a subgroup of $SO(4)$ of dimension six and therefore must be all of $SO(4)$.

Exercise *Complex structures on* $\mathbb{R}^4$

a.) Show that the complex structures on $\mathbb{R}^4$ compatible with the Euclidean metric can be identified as the maps

$$q \mapsto nq \qquad\qquad n^2 = -1 \tag{9.83}$$

OR

$$q \mapsto qn \qquad\qquad n^2 = -1 \tag{9.84}$$

b.) Use this to show that the space of such complex structures is $S^2 \amalg S^2$.

Exercise *Regular Representation*

♣This is too important to be an exercise, and is used heavily later. ♣

Compute the left and right regular representations of $\mathbb{H}$ on itself as follows: Choose a real basis for $\mathbb{H}$ with $v_1 = \mathfrak{i}, v_2 = \mathfrak{j}, v_3 = \mathfrak{k}, v_4 = 1$. Let $L(\mathfrak{q})$ denote left-multiplication by a quaternion $\mathfrak{q}$ and $R(\mathfrak{q})$ right-multiplciation by $\mathfrak{q}$. Then the representation matrices are:

$$L(\mathfrak{q})v_a := \mathfrak{q} \cdot v_a := L(\mathfrak{q})_{ba} v_b \tag{9.85}$$

$$R(\mathfrak{q})v_a := v_a \cdot \mathfrak{q} := R(\mathfrak{q})_{ba} v_b \tag{9.86}$$

a.) Show that:

$$L(\mathfrak{i}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{9.87}$$

$$L(\mathfrak{j}) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{9.88}$$

$$L(\mathfrak{k}) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{9.89}$$

$$R(\mathfrak{i}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \tag{9.90}$$

$$R(\mathfrak{j}) = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{9.91}$$

$$R(\mathfrak{k}) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{9.92}$$

b.) Show that these matrices generate the full 16-dimensional algebra $M_4(\mathbb{R})$. This is the content of the statement that

$$\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H}^{\mathrm{opp}} \cong \mathrm{End}(\mathbb{R}^4) \tag{9.93}$$

Exercise *'t Hooft symbols and the regular representation of* $\mathbb{H}$

The famous 't Hooft symbols, introduced by 't Hooft in his work on instantons in gauge theory are defined by

♣CHECK FACTORS OF TWO!!! ♣

$$\alpha^{\pm,i}_{\mu\nu} := \pm\delta_{i\mu}\delta_{\nu 4} \mp \delta_{i\nu}\delta_{\mu 4} + \epsilon_{i\mu\nu} \tag{9.94}$$

where $1 \le \mu, \nu \le 4$, $1 \le i \le 3$ and $\epsilon_{i\mu\nu}$ is understood to be zero if μ or ν is equal to 4. (Note: Some authors will use the notation $\eta^i_{\mu\nu}$, and some authors will use a different overall normalization.)

a.) Show that

$$\alpha^{+,1} = R(\mathfrak{i}) \qquad \alpha^{+,2} = R(\mathfrak{j}) \qquad \alpha^{+,3} = R(\mathfrak{k}) \tag{9.95}$$

$$\alpha^{-,1} = -L(\mathfrak{i}) \qquad \alpha^{-,2} = -L(\mathfrak{j}) \qquad \alpha^{-,3} = -L(\mathfrak{k}) \tag{9.96}$$

b.) Verify the relations

$$\begin{aligned} [\alpha^{\pm,i}, \alpha^{\pm,j}] &= -2\epsilon^{ijk}\alpha^{\pm,k} \\ [\alpha^{\pm,i}, \alpha^{\mp,j}] &= 0 \end{aligned} \tag{9.97}$$

c.) Let $\mathfrak{so}(4)$ of 4×4 denote the real Lie algebra of real anti-symmetric matrices. It is the Lie algebra of $SO(4)$. Show that it is of dimension 6 and that every element can be uniquely decomposed as $L(q_1) - R(q_2)$ where q_1, q_2 are imaginary quaternions.

d.) Show that the map

$$\Im(\mathbb{H}) \oplus \Im(\mathbb{H}) \to \mathfrak{so}(4) \tag{9.98}$$

$$q_1 \oplus q_2 \to L(q_1) - R(q_2) \tag{9.99}$$

defines an isomorphism of Lie algebras

$$\mathfrak{su}(2) \oplus \mathfrak{su}(2) \cong \mathfrak{so}(4) \tag{9.100}$$

eq:so4-isomorp

e.) Using (b) and (9.73) show that the above isomorphism of Lie algebras can be expressed as a mapping of generators

$$\begin{aligned} \frac{\sqrt{-1}}{2}\sigma^i \oplus 0 &\mapsto \frac{1}{2}\alpha^{+,i} \\ 0 \oplus \frac{\sqrt{-1}}{2}\sigma^i &\mapsto \frac{1}{2}\alpha^{-,i} \end{aligned} \tag{9.101}$$

f.) Now show that

$$\{\alpha^{\pm,i}, \alpha^{\pm,j}\} = -2\delta^{ij} \tag{9.102}$$

and deduce that:

$$\begin{aligned} \alpha^{+,i}\alpha^{+,j} &= -\delta^{ij} - \epsilon^{ijk}\alpha^{+,k} \\ \alpha^{-,i}\alpha^{-,j} &= -\delta^{ij} - \epsilon^{ijk}\alpha^{-,k} \end{aligned} \tag{9.103}$$

g.) Deduce that the inverse isomorphism to (9.100) is

$$T \mapsto \left(-\sqrt{-1}\mathrm{Tr}(\alpha^{+,i}T)\sigma^i\right) \oplus \left(-\sqrt{-1}\mathrm{Tr}(\alpha^{-,i}T)\sigma^i\right) \tag{9.104}$$

Exercise *Quaternions And (Anti-)Self-Duality*

a.) Introduce the epsilon tensor $\epsilon_{\mu\nu\lambda\rho}$ with $\epsilon_{1234} = +1$. Show that the rank-two antisymmetric tensors are *self-dual* and *anti-self-dual* in the sense that

$$\alpha^{+,i}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\alpha^{+,i}_{\lambda\rho} \tag{9.105}$$

$$\alpha^{-,i}_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}\alpha^{-,i}_{\lambda\rho} \tag{9.106}$$

b.) On $\mathfrak{so}(4)$, which as a vector space can be identified with the space of two-index anti-symmetric rank tensors define

$$(*T)_{\mu\nu} := \frac{1}{2}\epsilon_{\mu\nu\lambda\rho}T_{\lambda\rho} \tag{9.107}$$

Show that the linear transformation $* : \mathfrak{so}(4) \to \mathfrak{so}(4)$ satisfies $*^2 = 1$. Therefore

$$P^{\pm} = \frac{1}{2}(1 \pm *) \tag{9.108}$$

are projection operators. Interpret the isomorphism (9.100) as the decomposition into self-dual and anti-self-dual tensors.

c.) If T is an antisymmetric tensor with components $T_{\mu\nu}$ then a common notation is $P^{\pm}(T) = T^{\pm}$. Check that

$$\begin{aligned} T^{\pm}_{12} &= \frac{1}{2}(T_{12} \pm T_{34}) = \pm T^{\pm}_{34} \\ T^{\pm}_{13} &= \frac{1}{2}(T_{13} \pm T_{42}) = \pm T^{\pm}_{42} \\ T^{\pm}_{14} &= \frac{1}{2}(T_{14} \pm T_{23}) = \pm T^{\pm}_{23} \end{aligned} \tag{9.109}$$

We remark that the choice $\epsilon_{1234} = +1$, instead of $\epsilon_{1234} = -1$ is a choice of orientation on $\mathbb{R}^4$. A change of orientation exchanges self-dual and anti-self-dual.

d.) Show that $v^-_{\mu\nu} = v_\mu\overline{v_\nu} - v_\nu\overline{v_\mu}$ are anti-self-dual and $v^+_{\mu\nu} = \overline{v_\mu}v_\nu - \overline{v_\nu}v_\mu$ are self-dual.

Exercise *More about $SO(4)$ matrices*
Show that if $x_{1,2,3,4}$ are real then

$$L(x_4 + x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k}) = \begin{pmatrix} x_4 & -x_3 & x_2 & x_1 \\ x_3 & x_4 & -x_1 & x_2 \\ -x_2 & x_1 & x_4 & x_3 \\ -x_1 & -x_2 & -x_3 & x_4 \end{pmatrix} \tag{9.110}$$

eq:L-GenQuat

b.) Show that when $x_\mu x_\mu = 1$ the matrix $L(x_4 + x_1\mathfrak{i} + x_2\mathfrak{j} + x_3\mathfrak{k})$ is in $SO(4)$ rotation.
c.) Similarly

$$R(y_4 + y_1\mathfrak{i} + y_2\mathfrak{j} + y_3\mathfrak{k}) = \begin{pmatrix} y_4 & y_3 & -y_2 & y_1 \\ -y_3 & y_4 & y_1 & y_2 \\ y_2 & -y_1 & y_4 & y_3 \\ -y_1 & -y_2 & -y_3 & y_4 \end{pmatrix} \tag{9.111}$$

eq:R-GenQuat

is an $SO(4)$ rotation when $y_\mu y_\mu = 1$.
d.) Show that the general $SO(4)$ matrix is a product of these.
e.) Show that, in particular that if we identify $\mathfrak{k} = \mathrm{i}\sigma^3$ then

$$\rho(e^{\mathrm{i}\theta\sigma^3}, 1) = \begin{pmatrix} R(\theta) & 0 \\ 0 & R(-\theta) \end{pmatrix} \tag{9.112}$$

$$\rho(1, e^{\mathrm{i}\theta\sigma^3}) = \begin{pmatrix} R(\theta) & 0 \\ 0 & R(\theta) \end{pmatrix} \tag{9.113}$$

where ρ is the homomorphism defined in (9.77) and

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{9.114}$$

9.4 Quaternionic Structure On A Real Vector Space

Definition: A *quaternionic vector space* is a vector space V over $\kappa = \mathbb{R}$ together with three real linear operators $I, J, K \in \mathrm{End}(V)$ satisfying the quaternion relations. In other words, it is a real vector space which is a module for the quaternion algebra.

Example 1: Consider $\mathbb{H}^{\oplus n} \cong \mathbb{R}^{4n}$. Vectors are viewed as n-component column vectors with quaternion entries. Each quaternion is then viewed as a four-component real vector. The operators I, J, K are componentwise left-multiplication by $L(\mathfrak{i}), L(\mathfrak{j}), L(\mathfrak{k})$.

It is possible to put a quaternionic Hermitian structure on a quaternionic vector space and thereby define the quaternionic unitary group. Alternatively, we can define $U(n, \mathbb{H})$ as the group of $n \times n$ matrices over $\mathbb{H}$ such that $uu^\dagger = u^\dagger u = 1$. In order to define the conjugate-transpose matrix we use the quaternionic conjugation $q \to \bar{q}$ defined above.

Exercise *A natural sphere of complex structures*

Show that if V is a quaternionic vector space with complex structures I, J, K then there is a natural sphere of complex structures give by

$$\mathcal{I} = x_1 I + x_2 J + x_3 K \qquad x_1^2 + x_2^2 + x_3^2 = 1 \tag{9.115}$$

9.5 Quaternionic Structure On Complex Vector Space

Just as we can have a complex structure on a real vector space, so we can have a *quaternionic structure* on a complex vector space V. This is a $\mathbb{C}$-anti-linear operator K on V which squares to -1. Once we have $K^2 = -1$ we can combine with the operator I which is just multiplication by $\sqrt{-1}$, to produce $J = KI$ and then we can check the quaternion relations. The underlying real space $V_{\mathbb{R}}$ is then a quaternionic vector space.

Example 2: The canonical example is given by taking a complex vector space V and forming

$$W = V \oplus \bar{V} \tag{9.116}$$

The underlying real vector space $W_{\mathbb{R}}$ has quaternion actions:

$$I : (v_1, \overline{v_2}) \mapsto (iv_1, i\overline{v_2}) = (iv_1, -\overline{iv_2}) \tag{9.117}$$

$$J : (v_1, \overline{v_2}) \mapsto (-v_2, \overline{v_1}) \tag{9.118}$$

$$K : (v_1, \overline{v_2}) \mapsto (-iv_2, -\overline{iv_1}) \tag{9.119}$$

9.5.1 Complex Structure On Quaternionic Vector Space

Recall that a quaternionic vector space is a real vector space V with an action of the quaternions. So for every $q \in \mathbb{H}$ we have $T(q) \in \mathrm{End}_{\mathbb{R}}(V)$ such that

$$T(q_1)T(q_2) = T(q_1 q_2) \tag{9.120}$$

In other words, a representation of the real quaternion algebra.

If we think of V as a copy of $\mathbb{H}^n$ with the quaternionic action <u>left</u>-multiplication by q componentwise, so that

$$T(q) \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} := \begin{pmatrix} qq_1 \\ \vdots \\ qq_n \end{pmatrix} \tag{9.121}$$

then a complex structure would be a left-action by any $GL(n,\mathbb{H})$ conjugate of $T(\mathfrak{i})$. If we wish to preserve the norm, then it is a $U(n,\mathbb{H})$ conjugate of $T(\mathfrak{i})$.

A complex structure then describe an embedding of $\mathbb{C}^n$ into $\mathbb{H}^n$ so that we have an isomorphism of

$$\mathbb{H}^n \cong \mathbb{C}^n \oplus \mathbb{C}^n \tag{9.122}$$

as complex vector spaces.

9.5.2 Summary

To summarize we have described three basic structures we can put on vector spaces:

1. A *complex structure* on a real vector space W is a real linear map $I : W \to W$ with $I^2 = -1$. That is, a representation of the real algebra $\mathbb{C}$.
2. A *real structure* on a complex vector space V is a $\mathbb{C}$-anti-linear map $K : V \to V$ with $K^2 = +1$.
3. A *quaternionic structure* on a complex vector space V is a $\mathbb{C}$-anti-linear map $K : V \to V$ with $K^2 = -1$.
4. A *complex structure on a quaternionic vector space* V is a representation of the real algebra $\mathbb{H}$ with a complex structure commuting with the $\mathbb{H}$-action.

9.6 Spaces Of Real, Complex, Quaternionic Structures

-RCH-Structure

This section makes use of the "stabilizer-orbit theorem." See the beginning of section 3.

♣So, we should move this to chapter three as an application of the Stabilizer-Orbit theorem. ♣

We saw above that if V is a finite-dimensional real vector space with a complex structure then by an appropriate choice of basis we have an isomorphism $V \cong \mathbb{R}^{2n}$, for a suitable integer n, and $I = I_0$. So, choose an isomorphism of V with $\mathbb{R}^{2n}$ and identify with space of complex structures on V with those on $\mathbb{R}^{2n}$. Then the general complex structure on $\mathbb{R}^{2n}$ is of the form SI_0S^{-1} with $S \in GL(2n,\mathbb{R})$. In other words, there is a transitive action of $GL(2n,\mathbb{R})$ on the space of complex structures. We can then identify the space with a homogeneous space of $GL(2n,\mathbb{R})$ by computing the stabilizer of I_0. Now if

$$gI_0g^{-1} = I_0 \tag{9.123}$$

eq:Stab-I0

for some $g \in GL(2n,\mathbb{R})$ then we can write g in block-diagonal form

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{9.124}$$

and then the condition (9.123) is equivalent to $C = -B$ and $D = A$, so that

$$\mathrm{Stab}_{GL(2n,\mathbb{R}}(I_0) = \{g \in GL(2n,\mathbb{R}) | g = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}\} \tag{9.125}$$

we claim this subgroup of $GL(2n,\mathbb{R})$ is isomorphic to $GL(n,\mathbb{C})$. To see this simply note that if we introduce the matrix

$$S_0 := -\frac{1}{2\mathrm{i}}\begin{pmatrix} 1_n & 1_n \\ -\mathrm{i}1_n & \mathrm{i}1_n \end{pmatrix} \tag{9.126}$$

then if $g \in \mathrm{Stab}_{GL(2n,\mathbb{R})}(I_0)$ and we write it in block form we have

$$S_0^{-1} g S_0 = \begin{pmatrix} A - \mathrm{i}B & 0 \\ 0 & A + \mathrm{i}B \end{pmatrix} \tag{9.127}$$

thus the determinant is

$$\det(g) = |\det(A + \mathrm{i}B)|^2 \tag{9.128}$$

and since $\det(g) \neq 0$ we know that $\det(A + \mathrm{i}B) \neq 0$. Conversely, if we have a matrix $h \in GL(n,\mathbb{C})$ we can decompose it into its real and imaginary parts $h = A + \mathrm{i}B$ and embed into $GL(2n,\mathbb{R})$ via

$$h \mapsto \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \tag{9.129}$$

(We could change the sign of B in this embedding. The two embeddings differ by the complex conjugation automorphism of $GL(n,\mathbb{C})$.)

We conclude that the space of complex structures is a homogeneous space

$$\boxed{\mathrm{ComplexStr}(\mathbb{R}^{2n}) \cong GL(2n,\mathbb{R})/GL(n,\mathbb{C})} \tag{9.130}$$

We could demand that our complex structures are compatible with the Euclidean metric on $\mathbb{R}^{2n}$. Then the conjugation action by $O(2n)$ is transitive and the above embedding is an embedding of $U(n)$ into $O(2n)$ and

$$\boxed{\mathrm{CompatComplexStr}(\mathbb{R}^{2n}) \cong O(2n)/U(n)} \tag{9.131}$$

We now turn to the real structures on a finite-dimensional complex vector space. We can choose an isomorphism $V \cong \mathbb{C}^n$ and then the general real structure is related to

$$\mathcal{C}_0 : \sum_{i=1}^{n} z_i e_i \to \sum_{i=1}^{n} z_i^* e_i \tag{9.132}$$

by conjugation: $\mathcal{C} = g^{-1}\mathcal{C}_0 g$ with $g \in GL(n,\mathbb{C})$. The stabilizer of $\mathcal{C}_0$ is rather obviously $GL(n,\mathbb{R})$, which sits naturally in $GL(n,\mathbb{C})$ as a subgroup. Thus:

$$\boxed{\mathrm{RealStr}(\mathbb{C}^n) \cong GL(n,\mathbb{C})/GL(n,\mathbb{R})} \tag{9.133}$$

$$\boxed{\mathrm{CompatComplexStr}(\mathbb{C}^n) \cong U(n)/O(n)} \tag{9.134}$$

Let us now consider quaternionic structures on a complex vector space. Again, we can fix an isomorphism $V \cong \mathbb{C}^{2n}$ and the anti-linear operator

$$J_0 : \begin{pmatrix} v_1 \\ v_2^* \end{pmatrix} \to \begin{pmatrix} -v_2 \\ v_1^* \end{pmatrix} \tag{9.135}$$

where $v_1, v_2 \in \mathbb{C}^n$. Now we compute the stabilizer J_0, that is, the matrices $g \in GL(2n,\mathbb{C})$ such that $gJ_0g^{-1} = J_0$ acting on vectors in $\mathbb{C}^{2n}$. In terms of matrices this means:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} g \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = g^* \tag{9.136}$$

which works out to:

$$\mathrm{Stab}_{GL(2n,\mathbb{C})}(J_0) = \{g \in GL(2n,\mathbb{C}) | g = \begin{pmatrix} A & B \\ -B^* & A^* \end{pmatrix}\} \tag{9.137}$$

Recall our characterization of $n \times n$ matrices over the quaternions. It follows that this defines an embedding of $GL(n,\mathbb{H}) \to GL(2n,\mathbb{C})$. [9] So

$$\boxed{\mathrm{QuatStr}(\mathbb{C}^{2n}) \cong GL(2n,\mathbb{C})/GL(n,\mathbb{H})} \tag{9.138}$$

Putting the natural Hermitian structure on $\mathbb{C}^{2n}$ we could demand that quaternionic structures are compatible with this Hermitian structure. Then the conjugation action by $U(2n)$ is transitive and the above embedding is an embedding of $USp(2n)$ into $U(2n)$ and

$$\boxed{\mathrm{CompatQuatStr}(\mathbb{C}^{2n}) \cong U(2n)/USp(2n)} \tag{9.139}$$

Finally, in a similar way we find that the space of complex structures on a quaternionic vector space can be identified with

$$\boxed{\mathrm{CompatCmplxStr}(\mathbb{H}^n) \cong USp(2n)/U(n)} \tag{9.140}$$

Remarks

1. *Relation to Cartan involutions.* The above homogeneous spaces have an interesting relation to *Cartan involutions.* A Cartan involution [10] θ on a Lie algebra is a Lie algebra automorphism so that $\theta^2 = 1$. Decomposing into $\pm$ eigenspaces we have

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p} \tag{9.141}$$

where $\mathfrak{k} = \{X \in \mathfrak{g} | \theta(X) = X\}$ and $\mathfrak{p}$ is the -1 eigenspace and we have moreover

$$[\mathfrak{k},\mathfrak{k}] = \mathfrak{k} \qquad [\mathfrak{k},\mathfrak{p}] = \mathfrak{p} \qquad [\mathfrak{p},\mathfrak{p}] = \mathfrak{k} \tag{9.142}$$

At the group level we have an involution $\tau : G \to G$ so that at the identity element $d\tau = \theta$. Then if $K = \mathrm{Fix}(\tau)$ we have a diffeomorphism of G/K with the subset in G of "anti-fixed points":

$$G/K \cong \mathcal{O} := \{g \in G | \tau(g) = g^{-1}\} \tag{9.143}$$

[9] It is possible, but tricky to define the notion of a determinant of a matrix of quaternions. It is best to think of $GL(n,\mathbb{H})$ as a Lie group with Lie algebra $Mat_n(\mathbb{H})$, or in terms of $2n \times 2n$ matrices over $\mathbb{C}$, the group we have written explicitly.

[10] See Chapter **** below for much more detail.

The above structures, when made compatible with natural metrics are nice examples:

Complex structures on real vector spaces: $\mathbb{R}^{2n} \cong \mathbb{C}^n$. Moduli space:

$$O(2n)/U(n) \tag{9.144}$$

eq:ClassCartSp

This comes from $\tau(g) = I_0 g I_0^{-1}$ where I_0 is (9.5).

Real structures on complex vector spaces: $\mathbb{R}^n \hookrightarrow \mathbb{C}^n$. Moduli space

$$U(n)/O(n) \tag{9.145}$$

eq:ClassCartSp

This comes from $\tau(u) = u^*$.

Quaternionic structures on complex vector spaces: $\mathbb{C}^{2n} \cong \mathbb{H}^n$. Moduli space:

$$U(2n)/Sp(n) \tag{9.146}$$

eq:ClassCartSp

Viewing $Sp(n)$ as $USp(2n) := U(2n) \cap Sp(2n;\mathbb{C})$ we can use the involutive automorphism $\tau(g) = I_0^{-1} g^* I_0$ on $U(2n)$. The fixed points in $U(2n)$ are the group elements with $g I_0 g^{tr} = I_0$, but this is the defining equation of $Sp(2n,\mathbb{C})$.

Complex structures on quaternionic vector spaces: $\mathbb{C}^n \hookrightarrow \mathbb{H}^n$. Moduli space:

$$Sp(n)/U(n) \tag{9.147}$$

eq:ClassCartSp

Viewing $Sp(n)$ as unitary $n \times n$ matrices over the quaternions the involution is $\tau(g) = -\mathfrak{i} g \mathfrak{i}$, i.e. conjugation by the unit matrix times $\mathfrak{i}$.

When Cartan classified compact symmetric spaces he found the 10 infinite series of the form $O \times O/O$, $U \times U/U$, $Sp \times Sp/Sp$, $O/O \times O$, $U/U \times U$, $Sp/Sp \times Sp$ and the above for families. In addition there is a finite list of exceptional cases.

2. *The 10-fold way.* In condensed matter physics there is a very beautiful classification of ensembles of Hamiltonians with a given symmetry type known as the 10-fold way. It is closely related to the above families of Cartan symmetric spaces, as discovered by Altland and Zirnbauer. See, for example,

   http://www.physics.rutgers.edu/∼gmoore/695Fall2013/CHAPTER1-QUANTUMSYMMETRY-OCT5.pdf

   http://www.physics.rutgers.edu/∼gmoore/PiTP-LecturesA.pdf

Exercise *Quaternionic Structures On* $\mathbb{R}^4$

In an exercise above we showed that the space of Euclidean-metric-compatible quaternionic structures on $\mathbb{R}^4$ is $S^2 \amalg S^2$. Explain the relation of this to the coset $O(4)/U(2)$.

---

10. Some Canonical Forms For a Matrix Under Conjugation

10.1 What is a canonical form?

We are going to collect a compendium of theorems on special forms into which matrices can be put.

There are different ways one might wish to put matrices in a "canonical" or standard form.

In general we could consider multiplying our matrix by different linear transformations on the left and the right

$$A \to S_1 A S_2 \tag{10.1}$$

where S_1, S_2 are invertible.

On the other hand, if A is the matrix of a linear transformation $T : V \to V$ then change of bases leads to change of A by conjugation:

$$A \to SAS^{-1} \tag{10.2}$$

eq:congj

for invertible S.

On the other hand, if A is the matrix of a bilinear form on a vector space (see below) then the transformation will be of the form:

$$A \to SAS^{tr} \tag{10.3}$$

eq:trnwsp

and here we can divide the problem into the cases where A is symmetric or antisymmetric.

There is some further important fine print on the canonical form theorems: The theorems can be different depending on whether the matrix elements in

$$\mathbb{C} \supset \mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z} \tag{10.4}$$

Also, we could put restrictions on $S \in GL(n, \mathbb{C})$ and require the matrix elements of S to be in $\mathbb{R}, \mathbb{Q}, \mathbb{Z}$. Finally, we could require S to be unitary or orthogonal. As we put more restrictions on the problem the nature of the canonical forms changes.

Some useful references for some of these theorems:

1. Herstein, sec. 6.10,
2. Hoffman and Kunze, Linear Algebra

10.2 Rank

If $T : V \to W$ is a linear transformation between two vector spaces the dimension of the image is called the *rank* of T. If V and W are finite dimensional complex vector spaces it is the only invariant of T under change of basis on V and W:

Theorem 1 Consider any $n \times m$ matrix over a field κ, $A \in Mat_{n\times m}(\kappa)$. This has a left and right action by $GL(n, \kappa) \times GL(m, \kappa)$. By using this we can always bring A to the canonical form: Let r denote the rank, that is, the dimension of the image space. Then there exist $g_1 \in GL(n, \kappa)$ and $g_2 \in GL(m, \kappa)$ so that $g_1 A g_2^{-1}$ is of the form:

♣left- action and right-action not defined until next chapter. ♣

a.) If $r < n, m$:

$$\begin{pmatrix} 1_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{pmatrix} \tag{10.5}$$ eq:rankinv

b.) If $r = n < m$ then we write the matrix as

$$\begin{pmatrix} 1_n & 0_{n\times(m-n)} \end{pmatrix} \tag{10.6}$$

c.) If $r = m < n$ then we write the matrix as

$$\begin{pmatrix} 1_m \\ 0_{(n-m)\times n} \end{pmatrix} \tag{10.7}$$

d.) If $r = n = m$ then we have the identity matrix.

That is, the only invariant under arbitrary change of basis of domain and range is the rank.

The proof easily follows from the fact that any subspace $V' \subset V$ has a complementary vector space V'' so that $V' \oplus V'' \cong V$.

Exercise

Prove this. Choose an arbitrary basis for κ^n and κ^m and define an operator T using the matrix A. Now construct a new basis, beginning by choosing a basis for ker$T$.

Exercise

If M is the matrix of a rank 1 operator, in any basis, then it has the form

$$M_{i\alpha} = v_i w_\alpha \qquad i = 1, \dots, n; \alpha = 1, \dots, m \tag{10.8}$$

for some vectors v, w.

10.3 Eigenvalues and Eigenvectors

sec:ee

Now let us consider a *square matrix* $A \in Mat_{n\times n}(\kappa)$. Suppose moreover that it is the matrix of a linear transformation $T : V \to V$ expressed in some basis. If we are actually studying the operator T then we no longer have the luxury of using different transformations g_1, g_2 for change of basis on the domain and range. We must use the *same* invertible matrix S, and hence our matrix transforms by conjugation

$$A \to S^{-1}AS \tag{10.9}$$

This changes the classification problem dramatically.

When thinking about this problem it is useful to introduce the basic definition: If $T : V \to V$ is a linear operator and $v \in V$ is a <u>nonzero</u> vector so that

$$Tv = \lambda v \tag{10.10}$$

then λ is called an *eigenvalue* of T and v is called an *eigenvector*. A similar definition holds for a matrix. [11]

Example. The following matrix has two eigenvalues and two eigenvectors:

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{10.11}$$

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\lambda \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{10.12}$$

Note that the two equations can be neatly summarized as one by making the eigenvectors columns of a square matrix:

$$\begin{pmatrix} 0 & \lambda \\ \lambda & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix} \tag{10.13}$$

and so the matrix of eigenvectors diagonalizes our operator.

Generalizing the previous example, if $T \in End(V)$ and V has a basis $\{v_1, \dots, v_n\}$ of eigenvectors of T with eigenvalues λ_i. Then, wrt that basis, the associated matrix is diagonal:

$$\begin{pmatrix} \lambda_1 & 0 & 0 \cdots & 0 \\ 0 & \lambda_2 & 0 \cdots & 0 \\ \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 \cdots & \lambda_n \end{pmatrix} \tag{10.14}$$ eq:diagmatrix

In general, if A is the matrix of T with respect to some basis (not necessarily a basis of eigenvectors) and if S is a matrix whose columns are n linearly independent eigenvectors then

$$AS = S\mathrm{Diag}\{\lambda_1, \dots, \lambda_n\} \qquad \Rightarrow \qquad S^{-1}AS = \mathrm{Diag}\{\lambda_1, \dots, \lambda_n\} \tag{10.15}$$

As we shall see, not every matrix has a basis of eigenvectors. Depending on the field, a matrix might have no eigenvectors at all. A simple example is that over the field $\kappa = \mathbb{R}$ the matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{10.16}$$

has no eigenvectors at all. Thus, the following fact is very useful:

Theorem 10.3.1. If $A \in M_n(\mathbb{C})$ is any complex matrix then it has at least one nonvanishing eigenvector.

[11] Alternative terminology: characteristic value, characteristic vector.

In order to prove this theorem it is very useful to introduce the characteristic polynomial:

Definition The *characteristic polynomial* of a matrix $A \in M_n(\kappa)$ is

$$p_A(x) =: \det(x\mathbf{1_n} - A) \tag{10.17}$$

eq:charpoly

Proof of Theorem 10.3.1: The characteristic polynomial $p_A(x)$ has at least one root, call it λ, over the complex numbers. Since, $\det(\lambda\mathbf{1_n} - A) = 0$ the matrix $\lambda\mathbf{1_n} - A$ has a nonzero kernel $K \subset \mathbb{C}^n$. If v is a nonzero vector in K then it is an eigenvector. ♠

So - a natural question is -

Given a matrix A, does it have a basis of eigenvectors? Equivalently, can we diagonalize A via $A \to S^{-1}AS$?

NO! YOU CAN'T DIAGONALIZE EVERY MATRIX!

Definition A matrix $x \in M_n(\mathbb{C})$ is said to be *semisimple* if it can be diagonalized.

Remarks:

1. Note well: The eigenvalues of A are zeroes of the characteristic polynomial.
2. We will discuss Hilbert spaces in Section §13 below. When discussing operators T on Hilbert space one must distinguish eigenvalues of T from the elements of the *spectrum* of T. For a (bounded) operator T on a Hilbert space we define the *resolvent* of T to be the set of complex numbers λ so that $\lambda\mathbf{1} - T$ is 1-1 and onto. The complement of the resolvent is defined to be the *spectrum* of T. In infinite dimensions there are different ways in which the condition of being $1-1$ and invertible can go wrong. The *point spectrum* consists of the eigenvalues, that is, the set of λ so that $\ker(\lambda\mathbf{1} - T)$ is a nontrivial subspace of the Hilbert space. In general it is a proper subset of the spectrum of T. See section 18.3 below for much more detail.
2. Theorem 10.3.1 is definitely false if we replace $\mathbb{C}$ by $\mathbb{R}$. It is also false in infinite dimensions. For example, the Hilbert hotel operator has no eigenvector. To define the Hilbert hotel operator choose an ON basis $\phi_1, \phi_2, \ldots$ and let $S : \phi_i \to \phi_{i+1}$, $i = 1, \ldots$. In terms of harmonic oscillators we can represent S as

$$S = \frac{1}{\sqrt{a^\dagger a}} a^\dagger . \tag{10.18}$$

Exercise
If A is 2×2 show that

$$p_A(x) = x^2 - x\mathrm{tr}\,(A) + \det(A) \tag{10.19}$$

We will explore the generalization later.

Exercise
Show that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{10.20}$$

cannot be diagonalized.
(Note that it does have an eigenvector, of eigenvalue 0.)

10.4 Jordan Canonical Form

Although you cannot diagonalize every matrix, there is a canonical form which is "almost" as good: Every matrix $A \in M_n(\mathbb{C})$ can be brought to <u>*Jordan canonical form*</u> by conjugation by $S \in GL(n, \mathbb{C})$:

$$A \to S^{-1}AS \tag{10.21}$$

We will now explain this

Definition: A $k \times k$ matrix of the form:

$$J_\lambda^{(k)} = \begin{pmatrix} \lambda & 1 & & & & & 0 \\ 0 & \lambda & 1 & & & & \cdot \\ 0 & 0 & \lambda & 1 & & & \cdot \\ & & & \cdot & & & \cdot \\ & & & & \cdot & & 0 \\ & & & & & \cdot & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \lambda \end{pmatrix} \tag{10.22}$$

is called an <u>*elementary Jordan block belonging to* λ</u>. We can also write

$$J_\lambda^{(k)} = \lambda 1 + \sum_{i=1}^{k-1} e_{i,i+1} \tag{10.23}$$

Example: The first few elementary Jordan blocks are:

$$J_\lambda^{(1)} = \lambda, \qquad J_\lambda^{(2)} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \qquad J_\lambda^{(3)} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \cdots \tag{10.24}$$

Exercise *Jordan blocks and nilpotent matrices*
Write the Jordon block as

$$J_\lambda^{(k)} = \lambda \mathbf{1_k} + N^{(k)} \tag{10.25}$$

Show that $(N^{(k)})^\ell \neq 0$ for $\ell < k$ but $(N^{(k)})^k = 0$.

The $J_\lambda^{(k)}$ are the atoms in the world of matrices with complex matrix elements. They cannot be broken into more elementary parts by similarity transformation. For $k > 1$ the Jordan blocks cannot be diagonalized. One easy way to see this is to write:

$$J_\lambda^{(k)} - \lambda \mathbf{1_k} = N^{(k)} \tag{10.26}$$

eq:nilpjord

If $J_\lambda^{(k)}$ could be diagonalized, then so could the LHS of (10.26). However $N^{(k)}$ cannot be diagonalized since $(N^{(k)})^k = 0$, and $N^{(k)} \neq 0$. Another proof uses the characteristic polynomial. The characteristic polynomial of a diagonalizable matrix is $p_A(x) = \prod(x - \lambda_i)$. Now note that the characteristic polynomial of the Jordan matrix is:

$$p_{J^{(k)}}(x) := \det[x1 - J_\lambda^{(k)}] = (x - \lambda)^k \tag{10.27}$$

eq:charpol

Hence if J could be diagonalized it would have to equal $SJ_\lambda^{(k)}S^{-1} = Diag\{\lambda\}$. But then we can invert this to $J_\lambda^{(k)} = S^{-1}Diag\{\lambda\}S = Diag\{\lambda\}$, a contradiction for $k > 1$.

Theorem/Definition: *Every* matrix $A \in M_n(\mathbb{C})$ can be conjugated to Jordan canonical form over the complex numbers. A Jordan canonical form is a matrix of the form:

$$A = \begin{pmatrix} A_1 & & & & & 0 \\ & \cdot & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \\ 0 & & & & & A_s \end{pmatrix} \tag{10.28}$$

where we have blocks A_i corresponding to the *distinct* roots $\lambda_1, \dots, \lambda_s$ of the characteristic polynomial $p_A(x)$ and each block A_i has the form:

$$A_i = \begin{pmatrix} J_{\lambda_i}^{(k_1^i)} & & & & & & \\ & J_{\lambda_i}^{(k_2^i)} & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & J_{\lambda_i}^{(k_{\ell_i}^i)} \end{pmatrix} \tag{10.29}$$

where $J_{\lambda_i}^{(k_j^i)}$ is the elementary Jordan blocks belonging to λ_i and

$$n = \sum_{i=1}^{s} \sum_{t=1}^{\ell_i} k_t^i \tag{10.30}$$

Proof: We sketch a proof briefly below. See also Herstein section 6.6.

Note that the *characteristic polynomial* looks like

$$\begin{aligned} p(x) &\equiv \det[x1 - A] = \prod_i (x-\lambda_i)^{\kappa_i} \\ \kappa_i &= \sum_{j=1}^{\ell_i} k_j^i \end{aligned} \tag{10.31}$$ eq:charpolp

Remarks

1. Thus, if the roots of the characteristic polynomial are all distinct then the matrix is diagonalizable. This condition is sufficient, but not necessary, since $\lambda \mathbf{1_n}$ is diagonal, but has multiple characteristic values.

2. The above theorem implies that every matrix A can be put in the form:

$$A = A_{ss} + A_{nilp} \tag{10.32}$$

where A_{ss} ("the semisimple part") is diagonalizable and A_{nilp} is nilpotent. Note that if D is diagonal then

$$\mathrm{tr}\, D(N^{(k)})^{\ell} = 0 \tag{10.33}$$

for $\ell > 0$ and hence

$$\mathrm{tr}\, A^n = \mathrm{tr}\, A_{ss}^n. \tag{10.34}$$

Thus, the traces of powers of a matrix do not characterize A uniquely, unless it is diagonalizable.

Exercise *Jordan canonical form for nilpotent operators*

If $T: V \to V$ is a nilpotent linear transformation show that there are vectors $v_1, \dots, v_\ell$ so that V has a basis of the form:

$$\begin{aligned} \mathcal{B} = \{ & v_1, Tv_1, T^2 v_1, \dots, T^{b_1 - 1} v_1; \\ & v_2, Tv_2, \dots, T^{b_2-1} v_2; \\ & \cdots \\ & v_\ell, Tv_\ell, \dots, T^{b_\ell - 1} v_\ell \} \end{aligned} \tag{10.35}$$ eq:JCF

where $\dim V = b_1 + \cdots + b_\ell$ and

$$T^{b_j} v_j = 0 \qquad j = 1, \dots, \ell \tag{10.36}$$

Exercise *The Cayley-Hamilton theorem*

If $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ is a polynomial in x then we can evaluate it on a matrix $A \in Mat_n(k)$:

$$f(A) := a_0 + a_1 A + \cdots + a_m A^m \in Mat_n(k) \tag{10.37}$$

a.) Show that if $p_A(x)$ is the characteristic polynomial of A then

$$p_A(A) = 0 \tag{10.38}$$

b.) In general, for any matrix A, there is a smallest degree polynomial $m_A(x)$ such that $m_A(x = A) = 0$. This is called *the minimal polynomial of* A. In general $m_A(x)$ divides $p_A(x)$, but might be different from $p_A(x)$. Give an example of a matrix such that $m_A(x) \neq p_A(x)$. [12]

Exercise *A useful identity*

In general, given a power series $f(x) = \sum_{n\geq 0} a_n x^n$ we can evaluate it on a matrix $f(A) = \sum_{n\geq 0} a_n A^n$. In particular, we can define $\exp(A)$ and $\log A$ for a matrix A by the power series expansions of $\exp(x)$ in x and $\log(x)$ in $(1-x)$.

a.) Show that

$$\det\exp A = \exp\mathrm{Tr}A \tag{10.39}$$

Sometimes written in the less accurate form

$$\det M = \exp\mathrm{Tr}\log M \tag{10.40}$$

b.) Suppose M is invertible, and δM is "small" compared to M. Show that

$$\begin{aligned}\sqrt{\det(M+\delta M)} = \sqrt{\det M}\Bigg[&1 + \frac{1}{2}\mathrm{Tr}(M^{-1}\delta M) \\ &+ \frac{1}{8}(\mathrm{Tr}(M^{-1}\delta M))^2 - \frac{1}{4}\mathrm{Tr}(M^{-1}\delta M)^2 + \mathcal{O}((\delta M)^3)\Bigg]\end{aligned} \tag{10.41}$$

---

[12] *Answer*: The simplest example is just $A = \lambda \mathbf{1}_k$ for $k > 1$. For a more nontrivial example consider $A = J_\lambda^{(2)} \oplus J_\lambda^{(2)}$.

Exercise *Jordan canonical form and cohomology*

Recall from Section **** that a chain complex has a degree one map $Q : M \to M$ with $Q^2 = 0$. If M is a complex vector space of dimension $d < \infty$ show that the Jordan form of Q is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes 1_{d_1} \oplus 0_{d_2} \tag{10.42}$$

where $d = 2d_1 + d_2$. Show that the cohomology is isomorphic to $\mathbb{C}^{d_2}$.

Exercise *Nilpotent* 2×2 *matrices*

A matrix such that $A^m = 0$ for some $m > 0$ is called *nilpotent*.

a.) Show that any 2×2 nilpotent matrix must satisfy the equation:

$$A^2 = 0 \tag{10.43}$$

b.) Show that any 2×2 matrix solving $A^2 = 0$ is of the form:

$$A = \begin{pmatrix} x & y \\ z & -x \end{pmatrix} \tag{10.44}$$

where

$$x^2 + yz = 0. \tag{10.45}$$ eq:aone

The solutions to Equation (10.45) form a complex variety known as the A_1 singularity. It is a simple example of a (singular, noncompact) Calabi-Yau variety.

c.) If A is a 2×2 matrix do $\mathrm{tr}\, A, \det A$ determine its conjugacy class?

Exercise *Nilpotent matrices*

A matrix such that $A^m = 0$ for some $m > 0$ is called *nilpotent*.

a.) Characterize the matrices for which $p_A(x) = x^n$.

b.) Characterize the $n \times n$ matrices for which $N^2 = 0$.

c.) Characterize the $n \times n$ matrices for $N^k = 0$ for some integer k.

Exercise *Flat connections on a punctured sphere*

The moduli space of *flat* $GL(2, \mathbb{C})$ *connections on the three punctured sphere* is equivalent to the set of pairs (M_1, M_2) of two matrices in $GL(2, \mathbb{C})$ up to simultaneous conjugation. Give an explicit description of this space.

10.4.1 Proof of the Jordan canonical form theorem

We include the proof here for completeness.

♣Should say exactly where you use the ground field being $\mathbb{C}$. ♣

Part 1: Decompose the space according to the different characteristic roots:

Let $A \in Mat_n(\mathbb{C})$ be a matrix acting on $\mathbb{C}^n$. Let $m_A(x)$ be the minimal polynomial of A. This is the polynomial of least degree such that $m_A(A) = 0$. Suppose $m_A(x) = q_1(x)q_2(x)$, where q_1, q_2 are relatively prime polynomials. Then define

$$\begin{aligned} V_1 &:= \{v : q_1(A)v = 0\} \\ V_2 &:= \{v : q_2(A)v = 0\} \end{aligned} \tag{10.46}$$

eq:jordproof

We claim $\mathbb{C}^n = V_1 \oplus V_2$, and A acts block-diagonally in this decomposition.

To see this, note that there exist polynomials $r_1(x), r_2(x)$ such that

$$q_1(x)r_1(x) + q_2(x)r_2(x) = 1 \tag{10.47}$$

eq:relprime

If $u \in V_1 \cap V_2$ then, applying (10.47) with $x = A$ to u we get $u = 0$. Thus

$$V_1 \cap V_2 = \{0\} \tag{10.48}$$

eq:inter

Now, apply (10.47) to any vector u to get:

$$\begin{aligned} u &= q_1(A)(r_1(A)u) + q_2(A)(r_2(A)u) \\ &= u_1 \oplus u_2 \end{aligned} \tag{10.49}$$

Since $q_2(A)u_1 = m_A(A)(r_1(A)u) = 0$ we learn that $u_1 \in V_2$, and similarly $u_2 \in V_1$. Thus, any vector u is in the sum $V_1 + V_2$. Combined with (10.48) this means

$$\mathbb{C}^n = V_1 \oplus V_2. \tag{10.50}$$

Finally, V_1, V_2 are invariant under A. Thus, A acts block diagonally on $V_1 \oplus V_2$.

Now, factoring $m_A(x) = \prod_i (x - \lambda_i)^{\rho_i}$ into distinct roots we obtain a block decomposition of A on $\mathbb{C}^n = \oplus_i V_i$ where $(x - \lambda_i)^{\rho_i}$ is the minimal polynomial of A on V_i. Consider A restricted to V_i. We can subtract $\lambda_i 1$, which is invariant under all changes of basis to assume that $A_i^{\rho_i} = 0$.

Part 2: Thus, the proof of Jordan decomposition is reduced to the Jordan decomposition of matrices M on $\mathbb{C}^n$ such that the minimal polynomial is $M^\rho = 0$. [13]

We will approach this by showing using induction on dim$V$ that a nilpotent operator $T : V \to V$ has a basis of the form (10.35). The initial step is easily established for dim$V = 1$ (or dim$V = 2$). Now for the inductive step note that if T is nilpotent and nonzero then $T(V) \subset V$ is a *proper subspace*. After all, if $T(V) = V$ then applying T successively we would obtain a contradiction. Then, by the inductive hypothesis there must be a basis for $T(V)$ of the form given in (10.35).

[13] Here we follow a very elegant proof given by M. Wildon at http://www.ma.rhul.ac.uk/ uvah099/Maths/JNFfinal.pdf.

Now, let us consider the kernel of T. This contains the *linearly independent* vectors $T^{b_1}v_1, \ldots, T^{b_\ell}v_\ell$. We can complete this to a basis for $\mathrm{ker}T$ with some vectors $w_1, \ldots, w_m$. Now, since $v_i \in T(V)$ there must be vectors $u_i \in V$ with $T(u_i) = v_i$. Then, we claim,

$$\begin{aligned}
\mathcal{B} = \{&u_1, Tu_1, T^2v_1, \ldots, T^{b_1}u_1; \\
&u_2, Tu_2, \ldots, T^{b_2}u_2; \\
&\cdots \\
&u_\ell, Tu_\ell, \ldots, T^{b_\ell}u_\ell; \\
&w_1, \ldots, w_m\}
\end{aligned} \tag{10.51}$$

eq:JCF-2

is a basis for V. Of course, we have $T^{b_j+1}u_j = 0$ and $Tw_i = 0$. First, these vectors are linearly independent: Suppose there were a relation of the form

$$\begin{aligned}
0 = &\kappa_1^0 u_1 + \kappa_1^1 Tu_1 + \cdots + \kappa_1^{b_1} T^{b_1} u_1 \\
&+ \kappa_2^0 u_2 + \kappa_2^1 Tu_2 + \cdots + \kappa_2^{b_2} T^{b_2} u_2 \\
&+ \cdots \\
&+ \kappa_\ell^0 u_\ell + \kappa_\ell^1 Tu_\ell + \cdots + \kappa_\ell^{b_\ell} T^{b_\ell} u_\ell + \\
&+ \xi_1 w_1 + \cdots + \xi_m w_m
\end{aligned} \tag{10.52}$$

Apply T to this equation and use linear independence of the basis for $T(V)$, then use linear independence of the basis for $\mathrm{ker}T$. Finally, we can see that (10.51) is complete because

$$\begin{aligned}
\mathrm{dim}V &= \mathrm{dimker}T + \mathrm{dimim}T \\
&= (\ell + m) + (b_1 + \cdots + b_\ell) \\
&= \sum_{j=1}^{\ell} (b_j + 1) + m
\end{aligned} \tag{10.53}$$

This completes the argument. ♠

10.5 The stabilizer of a Jordan canonical form

Given a matrix $x \in M_n(\mathbb{C})$ it is often important to understand what matrices will commute with it.

For example, if $x \in GL(n, \mathbb{C})$ we might wish to know the stabilizer of the element under the action of conjugation:

$$Z(x) := \{g : g^{-1}xg = x\} \subset GL(n, \mathbb{C}) \subset M_n(\mathbb{C}) \tag{10.54}$$

In order to find the commutant of x it suffices to consider the commutant of its Jordan canonical form. Then the following theorem becomes useful:

Lemma Suppose k, ℓ are positive integers and A is a $k \times \ell$ matrix so that

$$J_{\lambda_1}^{(k)} A = A J_{\lambda_2}^{(\ell)} \tag{10.55}$$

eq:CommuteJ

Then

1. If $\lambda_1 \neq \lambda_2$ then $A = 0$.
2. If $\lambda_1 = \lambda_2 = \lambda$ and $k = \ell$ then A is of the form

$$A^{(k)}(\alpha) = \alpha_0 \cdot 1 + \alpha_1 J_\lambda^{(k)} + \alpha_2 (J_\lambda^{(k)})^2 + \cdots + \alpha_{k-1} (J_\lambda^{(k)})^{k-1} \tag{10.56}$$ eq:A-alpha

   for some set of complex numbers $\alpha_0, \dots, \alpha_{k-1}$.

3. If $\lambda_1 = \lambda_2 = \lambda$ and $k < \ell$ then A is of the form

$$\begin{pmatrix} 0 & A^{(k)}(\alpha) \end{pmatrix} \tag{10.57}$$

4. If $\lambda_1 = \lambda_2 = \lambda$ and $k > \ell$ then A is of the form

$$\begin{pmatrix} A^{(\ell)}(\alpha) \\ 0 \end{pmatrix} \tag{10.58}$$

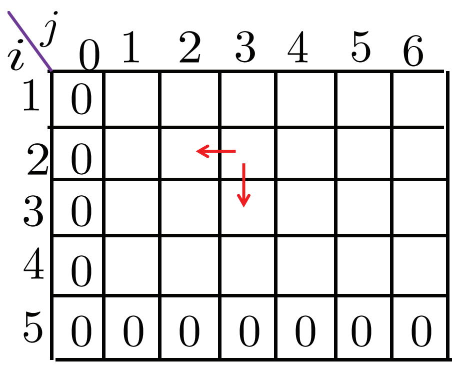

Figure 4: The commutation relation implies that entries in a box are related to those to the left and underneath in this enhanced matrix, as indicated by the arrows. fig:JORDANCOMM

Proof: Write

$$A = \sum_{i=1}^{k} \sum_{j=1}^{\ell} A_{ij} e_{ij} \tag{10.59}$$

Then (10.55) is equivalent to

$$(\lambda_2 - \lambda_1) A = \sum_{i=1}^{k-1} \sum_{j=1}^{\ell} A_{i+1,j} e_{ij} - \sum_{i=1}^{k} \sum_{j=2}^{\ell} A_{i,j-1} e_{ij} \tag{10.60}$$ eq:CommuteJ2

Now enhance the matrix A_{ij} to $\hat{A}_{ij}$ by adding a row $i = k+1$ and a column $j = 0$ so that

$$\hat{A}_{ij} = \begin{cases} A_{ij} & 1 \leq i \leq k, 1 \leq j \leq \ell \\ 0 & i = k+1 \quad \text{or} \quad j = 0 \end{cases} \tag{10.61}$$

so (10.60) becomes

$$(\lambda_2 - \lambda_1)\hat{A}_{ij} = \hat{A}_{i+1,j} - \hat{A}_{i,j-1} \qquad 1 \le i \le k, 1 \le j \le \ell \tag{10.62}$$

Now consider Figure 4. If $\lambda_1 \neq \lambda_2$ then we use the identity in the $i = k, j = 1$ box to conclude that $A_{k,1} = 0$. Then we successively use the identity going up the $j = 1$ column to find that $A_{i,1} = 0$ for all i. Then we start again at the bottom of the $j = 2$ column and work up. In this way we find $A = 0$. If $\lambda_1 = \lambda_2$ the identity just tells us that two entries along a diagonal have to be the same. Thus all diagonals with one of the zeros on the edge must be zero. The other diagonals can be arbitrary. But this is precisely what a matrix of type $A^{(k)}(\alpha)$ looks like. ♠

Remark: Note that the matrix $A^{(k)}(\alpha)$ in (10.56) above can be rewritten in the form

$$A^{(k)} = \beta_0 + \beta_1 N^{(k)} + \beta_2 (N^{(k)})^2 + \cdots + \beta_{k-1} (N^{(k)})^{k-1} \tag{10.63}$$

which has the form, for example for $k = 5$:

$$A^{(5)} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ 0 & \beta_0 & \beta_1 & \beta_2 & \beta_3 \\ 0 & 0 & \beta_0 & \beta_1 & \beta_2 \\ 0 & 0 & 0 & \beta_0 & \beta_1 \\ 0 & 0 & 0 & 0 & \beta_0 \end{pmatrix} \tag{10.64}$$

It is clear from this form that the matrix is invertible iff $\beta_0 \neq 0$.

A consequence of this Lemma is that for any $x \in M_n(\mathbb{C})$ the subgroup $Z(x) \subset GL(n, \mathbb{C})$ must have complex dimension at least n. Some terminology (which is common in the theory of noncompact Lie groups) that one might encounter is useful to introduce here in its simplest manifestation:

Definition x is said to be *regular* if the dimension of $Z(x)$ is precisely n. That is, a regular element has the smallest possible centralizer.

Exercise *Regular and semisimple*

Show that x is regular and semisimple iff *all the roots of the characteristic polynomial are distinct.* We will use this term frequently in following sections.

10.5.1 Simultaneous diagonalization

A second application of the above Lemma is simultaneous diagonalization:

Theorem: Two diagonalizable matrices which commute $[A_1, A_2] = 0$ are simultaneously diagonalizable.

Proof: If we first diagonalize A_1 then we get diagonal blocks corresponding to the distinct eigenvalues λ_i:

$$A_1 = \begin{pmatrix} \lambda_1 \mathbf{1_{r_1\times r_1}} & 0 & \cdots \\ 0 & \lambda_2 \mathbf{1_{r_2\times r_2}} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \tag{10.65}$$

We have now "broken the gauge symmetry" to $GL(r_1,\mathbb{C}) \times GL(r_2,\mathbb{C}) \times \cdots$. Moreover, since the λ_i are distinct, a special case of our lemma above says that A_2 must have a block-diagonal form:

$$A_2 = \begin{pmatrix} A_2^{(1)} & 0 & \cdots \\ 0 & A_2^{(2)} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \tag{10.66}$$

and we can now diagonalize each of the blocks without spoiling the diagonalization of A_1. ♠

In quantum mechanics this theorem has an important physical interpretation: Commuting Hermitian operators are observables whose eigenvalues can be simultaneously measured.

Exercise *Hilbert scheme of points*
Warning: This is a little more advanced than what is generally assumed here.
Suppose two $N \times N$ *complex* matrices X_1, X_2 commute

$$[X_1, X_2] = 0 \tag{10.67}$$

eq:hilscha

What can we say about the pair up to simultaneous conjugation? If they can be simultaneously diagonalized then we may write:

$$S X_i S^{-1} = Diag\{z_i^{(1)}, \dots, z_i^{(N)}\} \qquad i = 1,2 \tag{10.68}$$

eq:hilschb

and hence we can associate to (10.67) N points $(z_1^{(k)}, z_2^{(k)}) \in \mathbb{C}^2$, $k = 1, 2, \dots, N$. In fact, because of conjugation, they are N unordered points. Thus the set of diagonalizable X_i satisfying (10.67) is the symmetric product $\mathrm{Sym}^N(\mathbb{C}^2)$.

In general, we can only put X_1, X_2 into Jordan canonical form. Thus, the "moduli space" of conjugacy classes of pairs (X_1, X_2) of simultaneously diagonalizable matrices is more complicated. This is still not a good space. To get a good space we consider only the conjugacy classes of "stable triples." These are triples (X_1, X_2, v) where $[X_1, X_2] = 0$ and $v \in \mathbb{C}^N$ is a vector such that $X_1^{n_1} X_2^{n_2} v$ span $\mathbb{C}^N$. In this case we get a very interesting smooth variety known as the *Hilbert scheme of N points on $\mathbb{C}^2$*.

a.) Show that $Hilb^N(\mathbb{C}^2)$ can be identified with the set of ideals $I \subset \mathbb{C}[z_1, z_2]$ such that $\mathbb{C}[z_1, z_2]/I$ is a vector space of dimension N.

Hint: Given an ideal of codimension N, observe that multiplication by z_1, z_2 on $\mathbb{C}[z_1, z_2]/I$ define two commuting linear operators. Conversely, given (X_1, X_2, v) define $\phi : \mathbb{C}[z_1, z_2] \to \mathbb{C}^N$ by $\phi(f) := f(X_1, X_2)v$.

b.) Write matrices corresponding to the ideal

$$I = (z_1^N, z_2 - (a_1 z_1 + \cdots + a_{N-1} z_1^{N-1})) \tag{10.69}$$

The point is, by allowing nontrivial Jordan form, the singular space $\mathrm{Sym}^N(\mathbb{C}^2)$ is resolved to the nonsingular space $\mathrm{Hilb}^N(\mathbb{C}^2)$.

Exercise *ADHM equations*
Find the general solution of the 2×2 ADHM equations:

$$\begin{aligned} &[X_1, X_1^\dagger] + [X_2, X_2^\dagger] + ii^\dagger - j^\dagger j = \zeta_R \\ &[X_1, X_2] + ij = 0 \end{aligned} \tag{10.70}$$ eq:adhm

modulo $U(2)$ transformations.

---

11. Sesquilinear forms and (anti)-Hermitian forms

sec:sips

Definition 11.1 A *sesquilinear form* on a complex vector space V is a function

$$h : V \times V \to \mathbb{C} \tag{11.1}$$

which is anti-linear in the first variable and linear in the second. That is: for all $v_i, w_i \in V$ and $\alpha_i \in \mathbb{C}$:

$$\begin{aligned} h(v, \alpha_1 w_1 + \alpha_2 w_2) &= \alpha_1 h(v, w_1) + \alpha_2 h(v, w_2) \\ h(\alpha_1 v_1 + \alpha_2 v_2, w) &= \alpha_1^* h(v_1, w) + \alpha_2^* h(v_2, w) \end{aligned} \tag{11.2}$$

Note that h defines a $\mathbb{C}$-linear map $\bar{V} \times V \to \mathbb{C}$ and hence (by the universal property) factors through a unique $\mathbb{C}$-linear map

$$\tilde{h} : \bar{V} \otimes V \to \mathbb{C} \tag{11.3}$$

Conversely, such a map defines a sesquilinear form. *Thus, the space of all sesquilinear forms, which is itself a complex vector space, is isomorphic to* $(\bar{V} \otimes V)^*$. We write:

$$\mathrm{Sesq}(V) \cong (\bar{V} \otimes V)^* \tag{11.4}$$

Now note that there are canonical maps

$$\mathrm{Sesq}(V) \otimes V \to \bar{V}^* \tag{11.5}$$ eq:sesqmap-1

$$\mathrm{Sesq}(V) \otimes \bar{V} \to V^* \tag{11.6}$$ eq:sesqmap-2

given by the contraction $V^* \otimes V \to \mathbb{C}$ and $\bar{V}^* \otimes \bar{V} \to \mathbb{C}$, respectively. Written out more explicitly, what equation (11.6) means is that if we are given a sesquilinear form h and an element $\bar{w} \in \bar{V}$ we get a corresponding element $\ell_{h,\bar{w}} \in V^*$ given by

$$\ell_{h,\bar{w}}(v) := h(w, v) \tag{11.7}$$

eq:sesquisom-3

and similarly, for (11.5), given an element $w \in V$ and an h we get an element $\tilde{\ell}_{h,w} \in (\bar{V})^*$ given by

$$\tilde{\ell}_{h,w}(\bar{v}) := h(v, w) \tag{11.8}$$

Definition 11.2

1. A sesquilinear form is said to be *nondegenerate* if for all nonvanishing $v \in V$ there is some $w \in V$ such that $h(v, w) \neq 0$ and for all nonvanishing $w \in V$ there is some v with $h(v, w) \neq 0$.

2. An *Hermitian form* on a complex vector space V is a sesquilinear form such that for all $v, w \in V$:

$$h(v, w) = (h(w, v))^* \tag{11.9}$$

3. If $h(v, w) = -(h(w, v))^*$ then h is called *skew-Hermitian* or *anti-Hermitian.*

Remarks

1. If h is nondegenerate then (11.5) and (11.6) define *isomorphisms* $V \cong \bar{V}^*$ and $\bar{V} \cong V^*$, respectively. In general there is a canonical anti-linear isomorphism $V \cong \bar{V}$ and hence also a canonical antilinear isomorphism $V^* \cong (\bar{V})^*$. However, as we have stressed, there is no canonical isomorphism $V \cong V^*$. What the above definitions imply is that such an isomorphism is provided by a nondegenerate sesquilinear form.

2. In particular, an anti-linear isomorphism $V \cong V^*$ is provided by a nondegenerate Hermitian form. This is used in quantum mechanics in the Dirac bra-cket formalism where the anti-linear isomorphism $V \to V^*$ is denoted

$$|\psi\rangle \to \langle\psi| \tag{11.10}$$

So say this more precisely in the above language: If $v \in V$ is denoted $|\psi\rangle$ and $w \in V$ is denoted $|\chi\rangle$ then, *given an underlying nondegenerate sesquilinear form* we can write

$$h(v, w) = \ell_{h,\bar{v}}(w) = \langle\psi|\chi\rangle \tag{11.11}$$

If the underlying sesquilinear form is hermitian then $\langle\psi|\psi\rangle$ will be real. In fact, since probabilities are associated with such expressions we want it to be positive. This leads us to inner product spaces.

Exercise

Show that the most general Hermitian form on a complex vector space, expressed wrt a basis e_i is

$$h(\sum z_i e_i, \sum w_j e_j) = \sum_{i,j} z_i^* h_{ij} w_j \tag{11.12}$$

where $(h_{ij})^* = h_{ji}$. That is, h_{ij} is an Hermitian matrix.

---

12. Inner product spaces, normed linear spaces, and bounded operators

12.1 Inner product spaces

Definition 11.3 An *inner product space* is a vector space V over a field k ($k = \mathbb{R}$ or $k = \mathbb{C}$ here) with a positive Hermitian form. That is we have a k-valued function

$$(\cdot, \cdot) : V \times V \to k \tag{12.1}$$

satisfying the four axioms:

i.) $(x, y + z) = (x, y) + (x, z)$

ii.) $(x, \alpha y) = \alpha(x, y)$ $\alpha \in k$

iii.) $(x, y) = (y, x)^*$

iv.) $\forall x$, the *norm of* x:

$$\| x \|^2 := (x, x) \geq 0 \tag{12.2}$$

and moreover $(x, x) = 0 \leftrightarrow x = 0$.

Axioms (i),(ii),(iii) say we have a symmetric quadratic form for $k = \mathbb{R}$ and an Hermitian form for $k = \mathbb{C}$. The fourth axiom tells us that the form is not only nondegenerate, but positive. In quantum mechanics, we usually deal with such inner products because of the probability interpretation of the values (ψ, ψ). Probabilities should be positive.

Example 1: $\mathbb{C}^n$ with

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} (x^i)^* y^i \tag{12.3}$$

Example 2: $\mathbb{R}^n$, here $k = \mathbb{R}$ and

$$(\vec{x}, \vec{y}) = \vec{x} \cdot \vec{y} = \sum x^i y^i \tag{12.4}$$

Example 3: $\mathcal{C}[a, b]$= the set of complex-valued continuous functions on the interval $[a, b]$.

$$(f, g) := \int_a^b f(x)^* g(x) dx \tag{12.5}$$ eq:ContFun

Definition 11.4: A set of vectors $\{x_i\}$ in an inner product space is called *orthonormal* [14] if $(x_i, x_j) = \delta_{ij}$.

[14] We often abbreviate to ON

Theorem 11.1: If $\{x_i\}_{i=1,\dots,N}$ is an ON set then

$$\parallel x \parallel^2 = \sum_{i=1}^{N} |(x, x_i)|^2 + \parallel x - \sum_{i=1}^{N}(x_i, x)x_i \parallel^2 \tag{12.6}$$

Proof: Note that $\sum(x_i, x)x_i$ and $x - \sum(x_i, x)x_i$ are orthogonal. ♠

Theorem 11.2: (Bessel's inequality)

$$\parallel x \parallel^2 \geq \sum_{i=1}^{N} |(x, x_i)|^2 \tag{12.7}$$

Proof: Immediate from the previous theorem.

Theorem 11.3: (Schwarz inequality)

$$\parallel x \parallel \cdot \parallel y \parallel \geq |(x, y)| \tag{12.8}$$

Proof: It is true for $y = 0$. If $y \neq 0$ then note that $\{y/ \parallel y \parallel\}$ is an ON set. Apply Bessel's inequality ♠

12.2 Normed linear spaces

A closely related notion to an inner product space is

Definition: A *normed linear space* or *normed vector space* is a vector space V (over $k = \mathbb{R}$ or $k = \mathbb{C}$) with a function $\parallel \cdot \parallel : V \to \mathbb{R}$ such that

i.) $\parallel v \parallel \geq 0$, $\forall v \in V$
ii.) $\parallel v \parallel = 0$ iff $v = 0$
iii.) $\parallel \alpha v \parallel = |\alpha| \parallel v \parallel$
iv.) $\parallel v + w \parallel \leq \parallel v \parallel + \parallel w \parallel$

An inner product space is canonically a normed linear space because we can define

$$\parallel v \parallel = +\sqrt{(v, v)} \tag{12.9}$$

and verify all the above properties. However, the converse is not necessarily true. See the exercise below. The canonical example of normed linear spaces which are not inner product spaces are the bounded operators on an infinite-dimensional Hilbert space. See below.

Exercise *Another proof of the Schwarz inequality*
Note that $\parallel x - \lambda y \parallel^2 \geq 0$. Minimize wrt λ.

Exercise *Polarization identity and the parallelogram theorem*

a.) Given an inner product space V, prove that the inner product can be recovered from the norm using the polarization identity:

$$(x,y) = \frac{1}{4}[(\parallel x+y \parallel^2 - \parallel x-y \parallel^2) - i(\parallel x+iy \parallel^2 - \parallel x-iy \parallel^2)] \tag{12.10}$$

which is also sometimes written as

$$4(y,x) = \sum_{k=0}^{4} i^k (x+i^k y, x+i^k y) \tag{12.11}$$

b.) Prove that conversely a normed linear space is an inner product space if and only if the norm satisfies the parallelogram law:

$$\parallel x+y \parallel^2 + \parallel x-y \parallel^2 = 2 \parallel x \parallel^2 + 2 \parallel y \parallel^2 \tag{12.12}$$

Warning: This is hard. Start by proving additive linearity in y.

c.) Give an example of a normed linear space which is *not* an inner product space.

12.3 Bounded linear operators

Definition: A *bounded linear transformation* of a normed linear space $(V_1, \parallel \cdot \parallel_1)$ to $(V_2, \parallel \cdot \parallel_2)$ is a linear map $T : V_1 \to V_2$ such that $\exists C \geq 0$ with

$$\parallel Tv \parallel_2 \leq C \parallel v \parallel_1 \tag{12.13}$$ eq:bdd

for all $v \in V_1$. In this case, we define the norm of the operator

$$\parallel T \parallel = \sup_{\parallel v \parallel_1 = 1} \parallel Tv \parallel_2 = \inf C \tag{12.14}$$ eq:bddi

Theorem: For a linear operator between two normed vector spaces

$$T : (V_1, \parallel \cdot \parallel_1) \to (V_2, \parallel \cdot \parallel_2) \tag{12.15}$$

the following three statements are equivalent:

1. T is continuous at $x = 0$.
2. T is continuous at every $x \in V_1$.
3. T is bounded

Proof: The linearity of T shows that it is continuous at one point iff it is continous everywhere. If T is bounded by C then for any $\epsilon > 0$ we can choose $\delta < \epsilon/C$ and then $\parallel x \parallel < \delta$ implies $\parallel T(x) \parallel < \epsilon$. Conversely, if T is continuous at $x = 0$ then choose any $\epsilon > 0$ you like. We know that there is a δ so that $\parallel x \parallel < \delta$ implies $\parallel T(x) \parallel < \epsilon$. But this means that for any $x \neq 0$ we can write

$$\parallel T(x) \parallel = \frac{\parallel x \parallel}{\delta} \parallel T\left(\delta \frac{x}{\parallel x \parallel}\right) \parallel < \frac{\epsilon}{\delta} \parallel x \parallel \tag{12.16}$$

and hence T is bounded ♠

12.4 Constructions with inner product spaces

A natural question at this point is how the constructions we described for general vector spaces generalize to inner product spaces and normed vector spaces.

1. *Direct sum.* This is straightforward: If V_1 and V_2 are inner product spaces then so is $V_1 \oplus V_2$ where we define

$$(x_1 \oplus y_1, x_1 \oplus y_2)_{V_1 \oplus V_2} := (x_1, x_2)_{V_1} + (y_1, y_2)_{V_2} \tag{12.17}$$

2. *Tensor product* One can extend the tensor product to primitive vectors in the obvious way

$$(x_1 \otimes y_1, x_2 \otimes y_2)_{V_1 \otimes V_2} := (x_1, x_2)_{V_1} (y_1, y_2)_{V_2} \tag{12.18}$$

and then extending by linearity, so that

$$(x_1 \otimes y_1, x_2 \otimes y_2 + x_3 \otimes y_3)_{V_1 \otimes V_2} := (x_1 \otimes y_1, x_2 \otimes y_2)_{V_1 \otimes V_2} + (x_1 \otimes y_1, x_3 \otimes y_3)_{V_1 \otimes V_2} \tag{12.19}$$

3. For $\mathrm{Hom}(V_1, V_2)$ and dual spaces see Section §14.

What about quotient space? If $W \subset V$ is a subset of an inner product space then W of course becomes an inner product space. Can we make V/W an inner product space? Clearly this is problematic since the obvious attempted definition

$$([v_1], [v_2]) \stackrel{?}{=} (v_1, v_2)_V \tag{12.20}$$

would only be well-defined if if

$$(v_1 + w_1, v_2 + w_2) \stackrel{?}{=} (v_1, v_2) \tag{12.21}$$

for all $w_1, w_2 \in W$ and $v_1, v_2 \in V$. This is clearly impossible!

Although we cannot put an inner product on V/W one might ask if there is a canonical complementary space inner product space to W. There is a natural candidate, the *orthogonal complement*, defined by:

$$W^\perp := \{y \in V | \forall x \in W, (x, y) = 0\} \tag{12.22}$$

So the question is, do we have

$$V \stackrel{?}{=} W \oplus W^\perp \tag{12.23}$$

Note that we certainly have $W \cap W^\perp = \{0\}$ by positivity of the inner product. So the question is whether $V = W + W^\perp$. We will see that this is indeed always true, when V is finite dimensional, in Section §15. In infinite-dimensions we must be more careful as the following example shows:

Example: Let $V = \mathcal{C}[0, 1]$ be the inner product space of continuous complex-valued functions on the unit interval with inner product (12.5). Let $W \subset V$ be the subspace of functions which *vanish* on $[\frac{1}{2}, 1]$:

$$W = \{f \in \mathcal{C}[0, 1] | \qquad f(x) = 0 \qquad \frac{1}{2} \leq x \leq 1\} \tag{12.24}$$

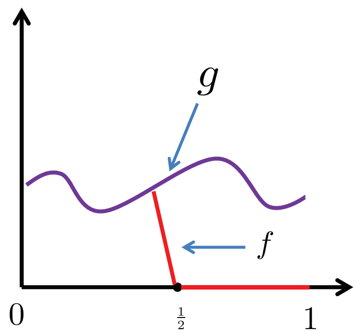

Figure 5: The function $g(x)$ must be orthogonal to all functions $f(x)$ in W. We can use the functions $f(x)$ shown here to see that $g(x)$ must vanish for $0 \leq x \leq \frac{1}{2}$. fig:Continuous

What is $W^\perp$? Any function $g \in W^\perp$ must be orthogonal to the function $f \in W$ which agrees with g for $x \in [0, \frac{1}{2} - \epsilon]$ and then continuously interpolates to $f(\frac{1}{2}) = 0$. See Figure 5. This implies that $g(x) = 0$ for $x < \frac{1}{2}$, but since g is continuous we must have $g(\frac{1}{2}) = 0$. Thus

$$W^\perp = \{g \in \mathcal{C}[0,1] | \qquad g(x) = 0 \qquad 0 \leq x \leq \frac{1}{2}\} \tag{12.25}$$

Now, clearly, $W + W^\perp$ is a *proper* subset of V since it cannot contain the simple function $h(x) = 1$ for all x. We will see that by making the inner product space *complete* we can eliminate such pathology.

13. Hilbert space

:HilbertSpace

In order to do the analysis required for quantum mechanics one often has to take limits. It is quite important that these limits exist. The notion of inner product space is too flexible.

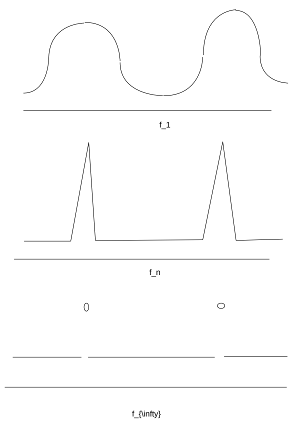

Figure 6: A sequence of continuous functions approaches a normalizable, but discontinuous function. fig:discfunc

For example, $\mathcal{C}[a,b]$ is an inner product space, but it is certainly possible to take a sequence of continuous functions $\{f_n\}$ such that $\| f_n - f_m \| \to 0$ for large n, m but f_n has no limiting continuous function as in 6. That's bad.

Definition: A *complete* inner product space is called a *Hilbert space*

Complete means: every Cauchy sequence converges to a point in the space.

Example 1: $\mathbb{R}^n$ and $\mathbb{C}^n$ are real and complex Hilbert spaces, respectively.

Counter-Example 2: $\mathcal{C}[a,b]$, the continuous complex-valued functions on $[a,b]$ is *not* a Hilbert space.

Example 3: Define

$$L^2[a,b] \equiv \left\{ f : [a,b] \to \mathbb{C} : |f|^2 \text{is measurable} \quad \text{and} \qquad \int_a^b |f(x)|^2 dx < \infty \right\} \tag{13.1}$$

It is not obvious, but it is true, that this defines a Hilbert space. To make it precise we need to introduce "measurable functions." For a discussion of this see Reed and Simon. This is the guise in which Hilbert spaces arise most often in quantum mechanics.

Example 4:

$$\ell^2 := \{\{x_n\}_{n=1}^{\infty} : \sum_{n=1}^{\infty} |x_n|^2 < \infty\} \tag{13.2}$$

♣Should give the example of Bargmann Hilbert space $Hol(\mathbb{C}, e^{-|z|^2} d^2 z)$ since this is a very nice representation for the HO algebra. ♣

Definition Two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are said to be *isomorphic* if there is an inner-product-preserving 1-1 linear transformation *onto* $\mathcal{H}_2$,

$$U : \mathcal{H}_1 \to \mathcal{H}_2 \tag{13.3}$$

such that

$$\forall x, y \in \mathcal{H}_1 : (Ux, Uy)_{\mathcal{H}_2} = (x, y)_{\mathcal{H}_1} \tag{13.4}$$

Such an operator U is also known as a *unitary transformation.*

Remark: In particular, U is norm preserving, that is, it is an *isometry.* By the polarization identity we could simply define U to be an isometry.

What are the invariants of a Hilbert space? By a general set-theoretic argument it is easy to see that every Hilbert space $\mathcal{H}$ has an ON basis. (For a proof see Reed & Simon Theorem II.5). The Bessel and Schwarz inequalities apply. If the basis is countable then $\mathcal{H}$ is called *separable.*

Theorem: Let $\mathcal{H}$ be a separable Hilbert space. Let S be an ON basis. Then

a.) If $|S| = N < \infty$, then $\mathcal{H} \cong \mathbb{C}^N$

b.) If $|S| = \infty$ then $\mathcal{H} \cong \ell^2$

Proof: (Case b): Let $\{y_n\}$ be a complete ON system. Then $U : x \to \{(y_n, x)\}$ is a unitary isomorphism with ℓ_2 ♠

Example 5 Consider $L^2[0, 2\pi]$. Then

$$\{\phi_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}\} \tag{13.5}$$

is a complete ON system. This statement summarizes the Fourier decomposition:

$$f(x) = \sum c_n \phi_n(x) \tag{13.6}$$

Example 6 $L^2(\mathbb{R})$. Note that the standard plane waves $e^{ik\cdot x}$ are *not* officially elements of $L^2(\mathbb{R})$. Nevertheless, there are plenty of ON bases, e.g. take the Hermite functions $H_n(x)e^{-x^2}$ (or any complete set of eigenfunctions of a Schrödinger operator whose potential goes to ∞ at $x \to \pm\infty$) and this shows that $L^2(\mathbb{R})$ is a separable Hilbert space. Indeed, all elementary quantum mechanics courses show that there is a highest weight representation of the Heisenberg algebra so that we have an ON basis $|n\rangle$, $n = 0, 1, \dots$ and

$$a|n\rangle = \sqrt{n}|n-1\rangle \tag{13.7}$$

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \tag{13.8}$$

The mapping of normalized Hermite functions to $|n\rangle$ gives an isometry of $L^2(\mathbb{R})$ with ℓ^2.

Remarks

1. *Dual space* Our various linear algebra operations can be carried over to Hilbert space provided one uses some care. The direct sum is straightforward. In order for the dual space $\mathcal{H}^*$ to be a *Hilbert space* we must use the *continuous* linear functions to define $\mathcal{H}^*$. Therefore, these are the *bounded* operators in the vector space of all linear operators $\mathrm{Hom}(\mathcal{H}, \mathbb{C})$. This space is again a Hilbert space, and the isomorphism $\mathcal{H} \cong \mathcal{H}^*$ is provided by the *Riesz representation theorem*, which says that every bounded linear functional $\ell : \mathcal{H} \to \mathbb{C}$ is of the form

$$\ell(v) = (w, v) \tag{13.9}$$

   for some unique $w \in \mathcal{H}$. This is the Hilbert space analog of the isomorphism (11.7).

2. *Tensor product.* We have indicated how to define the tensor product of inner product spaces. However, for Hilbert spaces completeness becomes an issue. See Reed+Simon Sec. II.4 for details. Of course, the tensor product of Hilbert spaces is very important in forming Fock spaces.

3. *"All Hilbert spaces are the same."* One sometimes encounters this slogan. What this means is that all separable Hilbert spaces are isomorphic as Hilbert spaces. Nevertheless Hilbert spaces of states appear in very different physical situations.

E.g. the Hilbert space of QCD on a lattice is a separable Hilbert space, so is it "the same" as the Hilbert space of a one-dimensional harmonic oscillator? While there is indeed an isomorphism between the two, it is not a physically natural one. The physics is determined by the kinds of operators that we consider acting on the Hilbert space. Very different (algebras) of operators are considered in the harmonic oscillator and in QCD. The isomorphism in question would take a physically natural operator in one context to a wierd one in a different context. So, in the next sections we will study in more detail linear operators on vector spaces.

Exercise *An application of the Schwarz inequality to quantum mechanics*
Consider the quantum mechanics of a particle on a line. Show that

$$(\langle\psi|x^2|\psi\rangle)^2 \leq \langle\psi|x^4|\psi\rangle \tag{13.10}$$

by applying the Schwarz inequality to $\psi(x)$ and $x^2\psi(x)$.

Exercise *The uncertainty principle*
Apply the Schwarz inequality to $\psi_1 = x\psi(x)$ and $\psi_2 = k\hat{\psi}(k)$ (where $\hat{\psi}(k)$ is the Fourier transform) by using the Plancherel theorem. Deduce the *uncertainty principle*:

$$\langle\psi|x^2|\psi\rangle\langle\psi|p^2|\psi\rangle \geq \frac{1}{4} \tag{13.11}$$

with minimal uncertainty (saturating the inequality) for the Gaussian wavepackets:

$$\psi(x) = Ae^{-Bx^2} \tag{13.12}$$

---

14. Banach space

ec:BanachSpace

The analog of a Hilbert space for normed linear spaces is called a Banach space:
Definition: A complete normed linear space is called a *Banach space*.

All Hilbert spaces are Banach spaces, but the converse is not true. A key example is the set of bounded operators on Hilbert space:

Theorem $\mathcal{L}(\mathcal{H})$ is a Banach space with the operator norm.

Sketch of proof: $\mathcal{L}(\mathcal{H})$ is a complex vector space and the operator norm makes it a normed linear space (the axioms are easily checked). If $\{A_n\}_{n=1}^{\infty}$ is a Cauchy sequence of operators in the operator norm then for all x, A_nx is a Cauchy sequence of vectors, and hence we can define $Ax = \lim_{n\to\infty} A_nx$, and it is not difficult to show that A is a bounded linear operator and $A_n \to A$ in the operator norm. See Reed-Simon Theorem III.2 for details ♠

Remarks:

1. In fact, the proof of the above theorem proves more: If V_1, V_2 are normed linear spaces and V_2 is complete, that is, if V_2 is a Banach space, then $\mathcal{L}(V_1, V_2)$ is a Banach space, with the operator norm.

♣So, state the theorem in this generality and let $\mathcal{L}(\mathcal{H})$ be a corollary. ♣

2. There are several different notions of convergence of a sequence of operators A_n between normed linear spaces $V_1 \to V_2$, and consequently several different topologies on the space of linear operators. The operator norm defines one topology. Another is the "strong topology" whereby $A_n \to A$ if for all $x \in V_1$, $\lim_{n\to\infty} \parallel A_n x - Ax \parallel = 0$. This is different from the norm topology because the rate at which $\parallel A_n x - Ax \parallel$ goes to zero might depend in an important way on x. There is also an intermediate "compact-open" topology which can also be useful.

♣Explain more here. ♣

♣Also point out that $C[0,1]$ with the sup norm *is a complete normed linear space.* So it is a Banach space, but not an inner product space. ♣

15. Projection operators and orthogonal decomposition

ctionOperators

As we discussed above, if $W \subset V$ be a subspace of an inner product space then we can define the *orthogonal subspace* by:

$$W^{\perp} = \{y : (x,y) = 0 \quad \forall x \in W\} \tag{15.1}$$

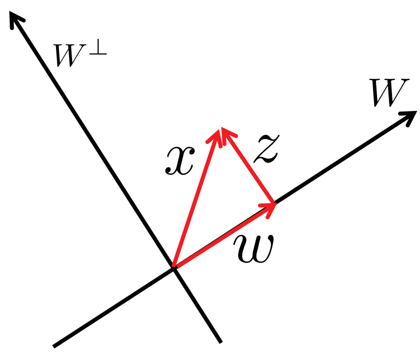

Figure 7: Vectors used in the projection theorem.

fig:Projection

Theorem(*The projection theorem*) Suppose V is a Hilbert space and $W \subset V$ is a *closed* subspace. Then:

$$V \cong W \oplus W^{\perp} \tag{15.2}$$

that is, every vector $x \in V$ can be *uniquely* decomposed as $x = z + w$ with $w \in W, z \in W^{\perp}$.

Proof: We follow the proof in Reed-Simon:

The first point to establish (and the point which fails if we use an inner product space which is not a Hilbert space) is that given any $x \in V$ there is a unique vector $w \in W$ which is closest to x. Let $d := \inf_{w\in W} \parallel x - w \parallel \geq 0$. There must be a sequence $w_n \in W$ with

$\lim_{n\to\infty} \| x - w_n \| = d$. But then we can write:

$$\begin{aligned} \| w_n - w_m \|^2 &= \| (w_n - x) + (x - w_m) \|^2 \\ &= 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - \| 2x - (w_n + w_m) \|^2 \\ &= 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - 4 \| x - \frac{1}{2}(w_n + w_m) \|^2 \\ &\leq 2 \| x - w_n \|^2 + 2 \| x - w_m \|^2 - 4d^2 \end{aligned} \tag{15.3}$$

where in line 2 we used the parallelogram identity. Now we know that the limit of the RHS is zero, therefore, for all $\epsilon > 0$ there is an N so that for $n, m > N$ the RHS is less than ϵ. Therefore $\{w_n\}$ is a Cauchy sequence and, since W is closed, $\lim w_n = w \in W$ exists and minimizes the distance.

Now denote $z := x - w$, where w is the distance minimizing $w \in W$ we have just found. We need to prove that z is in $W^\perp$. Let $d = \| x - w \|$. Then for all $t \in \mathbb{R}, y \in W$:

$$d^2 \leq \| x - (w + ty) \|^2 = \| z - ty \|^2 = d^2 - 2tRe(z, y) + t^2 \| y \|^2 \tag{15.4}$$

This must hold for all t and therefore,

$$Re(z, y) = 0 \tag{15.5}$$

for all y. If we have a real vector space then $(z, y) = 0$ for all $y \in W \Rightarrow z \in W^\perp$, so we are done. If we have a complex vector space we replace $t \to it$ above and prove

$$Im(z, y) = 0 \tag{15.6}$$

Therefore, z is orthogonal to every vector $y \in W$. ♠

Remarks:

1. The theorem definitely *fails* if we drop the positive definite condition on $(\cdot, \cdot)$. Consider $\mathbb{R}^2$ with "inner product" defined on basis vectors e, f by

$$(e, e) = (f, f) = 0 \qquad (e, f) = (f, e) = 1 \tag{15.7}$$

   The product is nondegenerate, but if $W = \mathbb{R}e$ then $W^\perp = W$.

In general, for *any* vector space V, not necessarily an inner product space, we can define a *projection operator* to be an operator $P \in End(V)$ such that $P^2 = P$. It then follows from trivial algebraic manipulation that if we define $Q = 1 - P$ we have

1. $Q^2 = Q$ and $QP = PQ = 0$
2. $1 = P + Q$

Moreover we claim that

$$V = \ker P \oplus \mathrm{im} P \tag{15.8}$$

Proof: First, for any linear transformation $\ker T \cap \mathrm{im} T = \{0\}$, so that applies to $T = P$. Next we can write, for any vector v, $v = Pv + (1 - P)v$, and we note that $(1 - P)v \in \ker P$. ♠

In finite dimensions, P can be brought to Jordan form and the equation $P^2 = P$ then shows that it is diagonal with diagonal eigenvalues 0 and 1.

Now, if we are in a *Hilbert space* V and $W \subset V$ is a closed subspace the projection theorem says that every $x \in V$ has a unique decomposition $x = w + w^\perp$ in $W \oplus W^\perp$. Therefore, if we define $P(x) = w$ we have a projection operator. Clearly, Q is the projector to $W^\perp$. We now note that these projectors satisfy the following relation

$$(x_1, Px_2) = (w_1 + w_1^\perp, w_2) = (w_1, w_2) \tag{15.9}$$

$$(Px_1, x_2) = (w_1, w_2 + w_2^\perp) = (w_1, w_2) \tag{15.10}$$

and so for all $x_1, x_2 \in \mathcal{H}$, we have $(Px_1, x_2) = (x_1, Px_2)$. As we will see below, this means that P (and likewise Q) is *self-adjoint.* In general, a self-adjoint projection operator is also known as an *orthogonal projection operator.*

Conversely, given a self-adjoint projection operator P we have an orthogonal decomposition $V = W \oplus W^\perp$. Therefore, *there is a 1-1 correspondence between closed subspaces of V and self-adjoint projection operators.*

As an application we prove

Theorem [Riesz representation theorem]: If $\ell : \mathcal{H} \to \mathbb{C}$ is a bounded linear functional then there exists a unique $y \in \mathcal{H}$ so that $\ell(x) = (y, x)$. Moreover $\parallel \ell \parallel = \parallel y \parallel$ so that $\mathcal{H}^* \cong \mathcal{H}$.

Proof: If $\ell = 0$ then we take $y = 0$. If $\ell \neq 0$ then there is some vector not in the kernel. Now, because ℓ is continuous $\ker\ell$ is a closed subspace of $\mathcal{H}$ and hence there is an orthogonal projection P to $(\ker\ell)^\perp$. Now, the equation $\ell(x) = 0$ is one (complex) equation and hence the kernel should have complex codimension one. That is, P is a projector to a one-dimensional subspace. Therefore we can choose any nonzero $x_0 \in (\ker\ell)^\perp$ and then

$$P(x) = \frac{(x_0, x)}{(x_0, x_0)} x_0 \tag{15.11}$$

is the projector to the orthogonal complement to $\ker\ell$. If $Q = 1 - P$ then Q is the orthogonal projector to $\ker\ell$ and hence

$$\ell(x) = \ell(P(x) + Q(x)) = \ell(P(x)) = \ell(x_0) \frac{(x_0, x)}{(x_0, x_0)} = (y, x) \tag{15.12}$$

where $y = \frac{\ell(x_0)^*}{(x_0, x_0)} x_0$. Given this representation one easily checks $\parallel \ell \parallel = \parallel y \parallel$ ♠

Remarks Thus, there is a 1-1 correspondence between projection operators and closed linear subspaces. The operator norm defines a topology on the space of bounded operators, and then the *space of projection operators* is a very interesting manifold, known as the Grassmannian. The Grassmannian has several disconnected components, labelled by the rank of P. Each component has very intricate and interesting topology. Grassmannians of projection operators on Hilbert space are very important in the theory of fiber bundles. In physics they arise naturally in the quantization of free fields in curved spacetime and more recently have played a role both in mathematical properties of the S-matrix of N=4 supersymmetric Yang-Mills theory as well as in the classification of topological phases of matter in condensed matter theory.

Exercise
Suppose that Γ is an operator such that $\Gamma^2 = 1$. Show that

$$\Pi_\pm = \frac{1}{2}(1 \pm \Gamma) \tag{15.13}$$

are projection operators, that is, show that: Show that

$$\Pi_+^2 = \Pi_+ \qquad \Pi_-^2 = \Pi_- \qquad \Pi_+\Pi_- = \Pi_-\Pi_+ = 0 \tag{15.14}$$

---

16. Unitary, Hermitian, and normal operators

Let V, W be finite dimensional inner product spaces. It is important in this section that we are working with finite dimensional vector spaces, otherwise the theory of operators is much more involved. See Reed and Simon and below.

Given

$$T : V \to W \tag{16.1}$$

we can *define* the adjoint operator

$$T^\dagger : W \to V \tag{16.2}$$ eq:defadji

by:

$$\forall x \in W, y \in V \qquad (T^\dagger x, y) := (x, Ty) \tag{16.3}$$

Here we are using the property that the inner product is a nondegenerate form: To define the vector $T^\dagger x$ it suffices to know its inner product with all vectors y. (In fact, knowing the inner product on a basis will do.)

If $T : V \to V$ and $\{v_i\}$ is an ordered *orthonormal* basis, then it follows that the matrices wrt this basis satisfy:

$$(T^\dagger)_{ij} = (T_{ji})^* \tag{16.4}$$ eq:daggmat

It also follows immediately from the definition that

$$\ker T^\dagger = (\mathrm{im} T)^\perp \tag{16.5}$$

Thus, we have the basic picture of an operator between inner product spaces given in Figure 8

Definition:

If V is an inner product space and $T : V \to V$ is linear then:
1. $T^\dagger = T$ defines an *Hermitian* or *self-adjoint* operator.

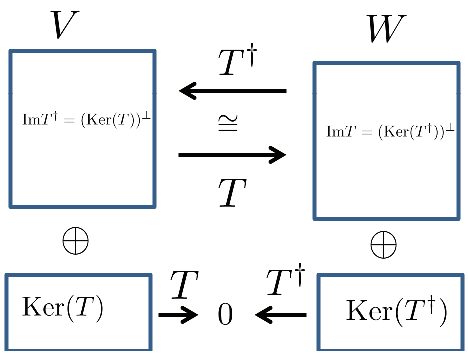

Figure 8: Orthogonal decomposition of domain and range associated to an operator T between inner product spaces. fig:INNPRODSPA

2. $TT^{\dagger} = T^{\dagger}T = 1$ defines a *unitary operator*.
3. $TT^{\dagger} = T^{\dagger}T$ defines a *normal operator*.

Put differently, a unitary operator on an inner product space is a norm preserving linear operator (a.k.a. an *isometry*) i.e.

$$\forall v \in V \qquad \parallel T(v) \parallel^2 = \parallel v \parallel^2 \tag{16.6}$$

It is worthwhile expressing these conditions in terms of the matrices of the operators wrt an ordered ON basis v_i of V. Using (16.4) we see that with respect to an ON basis:

$$\boxed{\text{Hermitian matrices } H_{ij} \text{ satisfy: } H_{ij} = H^*_{ji}.}$$

$$\boxed{\text{Unitary matrices } U_{ij} \text{ satisfy: } \sum_k U_{ik} U^*_{jk} = \delta_{ij}}$$

Remarks

1. Recalling that $\mathrm{cok}T := W/\mathrm{im}T$ we see that

$$\mathrm{cok}T \cong \mathrm{ker}T^{\dagger} \tag{16.7}$$

2. For a finite dimensional Hilbert space $\mathcal{L}(\mathcal{H})$ is an inner product space with natural inner product [15]

$$(A_1, A_2) = \mathrm{Tr} A_1^{\dagger} A_2 \tag{16.8}$$

but this will clearly not work in infinite dimensions since, for example, the unit operator is bounded but has no trace. In finite dimensions this shows that if V_1, V_2 are inner product spaces then so is $\mathrm{Hom}(V_1, V_2)$ in a natural way. In particular, V^* is an inner product space.

[15] Warning, the norm $\parallel A \parallel$ defined by this inner product is *not* the same as the operator norm!

Exercise

a.) If $\mathrm{Tr} T^\dagger T = 0$, show that $T = 0$.

Exercise

Let $\{v_i\}, \{\hat{v}_i\}$ be any two ON bases for a vector space V. Show that the operator defined by $U : v_i \to \hat{v}_i$ is unitary.

Exercise *Unitary vs. Orthogonal*

Show that if $U \in M_n(\mathbb{C})$ is both unitary and real then it is an orthogonal matrix.

Exercise *Eigenvalues of hermitian and unitary operators*

a.) The eigenvalues λ of Hermitian operators are real $\lambda^* = \lambda$. This mathematical fact is compatible with the postulate of quantum mechanics that states that observables are represented by Hermitian operators.

b.) The eigenvalues of unitary operators are phases: $|\lambda| = 1$.

Exercise *The Cayley transform*

The Mobius transformation $z \to w = \frac{1+iz}{1-iz}$ maps the upper half of the complex plane to the unit disk. This transform has an interesting matrix generalization:

a.) Show that if H is Hermitian then

$$U = (1 + iH)(1 - iH)^{-1} \tag{16.9}$$

is a unitary operator.

b.) Show that if U is a unitary operator then $H = i(1 - U)(1 + U)^{-1}$ is Hermitian provided $(1 + U)$ is invertible.

Exercise

a.) Show that if H is Hermitian and α is real then $U = \exp[i\alpha H]$ is unitary.

b.) Show that if H is an Hermitian operator and $\parallel H \parallel \leq 1$ then

$$U_{\pm} = H \pm i\sqrt{1 - H^2} \tag{16.10}$$

are unitary operators.

c.) Show that any matrix can be written as the sum of four unitary matrices.

---

17. The spectral theorem: Finite Dimensions

A nontrivial and central fact in math and physics is:

The (finite dimensional) spectral theorem: Let V be a finite dimensional inner product space over $\mathbb{C}$ and let $H : V \to V$ be an Hermitian operator. Then:

a.) If $\lambda_1, \dots, \lambda_k$ are the *distinct* eigenvalues of H then there are mutually orthogonal Hermitian projection operators $P_{\lambda_1}, \dots, P_{\lambda_k}$ such that:

$$H = \sum_i \lambda_i P_{\lambda_i} = \lambda_1 P_{\lambda_1} \oplus \lambda_2 P_{\lambda_2} \oplus \cdots \lambda_k P_{\lambda_k} \tag{17.1}$$ eq:spctrldecom

$$1 = P_{\lambda_1} + \cdots + P_{\lambda_k} \tag{17.2}$$ eq:sprtrlii

b.) There exists an ON basis of V of eigenvectors of H.

Idea of the proof:

We proceed by induction on $\dim V$. The case $\dim V = 1$ is clear.

Now suppose $\dim V = n > 1$. By theorem 10.3.1 of section 10.3 H has one nonvanishing eigenvector v. Consider $W = L(\{v\})$, the span of v. The orthogonal space $W^{\perp}$ is also an inner product space. Moreover H takes $W^{\perp}$ to $W^{\perp}$: If $x \in W^{\perp}$ then

$$(Hx, v) = (x, Hv) = \lambda(x, v) = 0 \tag{17.3}$$

so $Hx \in W^{\perp}$. Further, the restriction of H to $W^{\perp}$ is Hermitian. Since $\dim W^{\perp} = \dim V - 1$ by the inductive hypothesis there is an ON basis for H on $W^{\perp}$ and hence there is one for V. ♠

In terms of matrices, any Hermitian matrix is unitarily diagonalizable:

$$H^{\dagger} = H \Rightarrow \exists U, \quad UU^{\dagger} = 1 \quad s.t. : UHU^{\dagger} = Diag\{E_1, \dots, E_n\}, E_i \in \mathbb{R} \tag{17.4}$$

Exercise

Show that the orthogonal projection operators P_i for H can be written as polynomials in H. Namely, we can take

$$P_{\lambda_i} = \prod_{j \neq i} \left(\frac{H - \lambda_j}{\lambda_i - \lambda_j} \right) \tag{17.5}$$

Exercise
Let $f(x)$ be a convergent power series and let H be Hermitian. Then

$$f(H) = \sum_i f(\lambda_i) P_{\lambda_i} \tag{17.6}$$

Exercise
Let X_n be the symmetric $n \times n$ matrix whose elements are all zero except for the diagonal above and below the principal diagonal. These matrix elements are 1. That is:

$$X_n = \sum_{i=1}^{n-1} e_{i,i+1} + \sum_{i=2}^{n} e_{i,i-1} \tag{17.7}$$

where $e_{i,j}$ is the matrix with 1 in row i and column j, and zero otherwise.
Find the eigenvalues and eigenvectors of X_n.

Exercise
Let H be a positive definite Hermitian matrix. Evaluate

$$\lim_{N\to\infty} \frac{\log(v, H^N v)}{N} \tag{17.8}$$

Exercise *The Seesaw Mechanism*
Consider the matrix

$$H = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{17.9}$$

eq:mtrxp

a.) Assuming H is Hermitian find the eigenvalues and eigenvectors of H.

An important discovery in particle physics of the past few years is that neutrinos have nonzero masses. One idea in neutrino physics is the seesaw mechanism - there are two "flavors" of neutrinos with a mass matrix

$$\begin{pmatrix} 0 & m \\ m^* & M \end{pmatrix} \tag{17.10}$$

eq:seesaw

where M is real. The absolute values of the eigenvalues give the masses of the two neutrinos.

c.) Find the eigenvalues of (17.10), and give a simplified expression for the eigenvalues in the limit where $|m| \ll |M|$.

d.) Suppose it is known that $|m| \cong 1\text{TeV} = 10^3\text{GeV}$, and a neutrino mass of $1eV$ is measured experimentally. What is the value of the large scale M?

e.) For what values of m, M does the kernel of (17.10) jump in dimension? Verify the constancy of the index.

17.1 Normal and Unitary matrices

Together with simultaneous diagonalization we can extend the set of unitarily diagonalizable matrices. Recall that a complex $n \times n$ matrix A is called *normal* if $AA^\dagger = A^\dagger A$,

Theorem:

1. Every normal matrix is diagonalizable by a *unitary matrix*.

2. Every normal operator T on a finite-dimensional inner product space has a spectral decomposition

$$T = \sum_{i=1}^{k} \mu_i P_{\mu_i} \tag{17.11}$$

where $\mu_i \in \mathbb{C}$ and $\{P_{\mu_i}\}$ are mutually orthogonal projection operators summing to the identity.

Proof: Note that we can decompose any matrix as

$$A = H + K \tag{17.12}$$

with $H^\dagger = H$ and $K^\dagger = -K$, antihermitian. Thus, iK is hermitian. If A is normal then $[H, K] = 0$. Thus we have two commuting hermitian matrices, which can be simultaneously diagonalized. ♠

As an immediate corollary we have

Theorem The eigenvectors of a unitary operator on an inner product space V form a basis for V. Every unitary operator on a finite dimensional inner product space is unitarily diagonalizable.

Proof: $U^\dagger U = 1 = UU^\dagger$ so U is a normal matrix. ♠

17.2 Singular value decomposition and Schmidt decomposition

17.2.1 Bidiagonalization

Theorem Any matrix $A \in M_n(\mathbb{C})$ can be *bidiagonalized* by unitary matrices. That is, there always exist unitary matrices $U, V \in U(n)$ such that

$$UAV^\dagger = \Lambda \tag{17.13}$$

is diagonal with nonnegative entries

Proof: First diagonalize $AA^\dagger$ and $A^\dagger A$ by U, V, so $UAA^\dagger U^\dagger = D_1$ and $VA^\dagger AV^\dagger = D_2$. Then note that $\mathrm{Tr}D_1^\ell = \mathrm{Tr}D_2^\ell$ for all positive integers ℓ. Therefore, up to a permutation, D_1, D_2 are the same diagonal matrices, but a permutation is obtained by conjugation with a unitary matrix, so we may assume $D_1 = D_2$. Then it follows that $UAV^\dagger$ is normal, and hance can be unitarily diagonalized. Since we have separate phase degrees of freedom for U and V we can rotate away the phases on the diagonal to make the diagonal entries nonnegative. ♠.

Remarks:

1. By suitable choice of U and V the diagonal elements of Λ can be arranged in monotonic order. The set of these values are called the *singular values* of A.

2. This theorem is very useful when investigating moduli spaces of vacua in supersymmetric gauge theories.

♣Should give some examples... ♣

Exercise
Find a unitary bidiagonalization of the Jordan block $J_\lambda^{(2)}$.

17.2.2 Application: The Cabbibo-Kobayashi-Maskawa matrix, or, how bidiagonalization can win you the Nobel Prize

This subsubsection assumes some knowledge of relativistic field theory.

One example where bidiagonalization is important occurs in the theory of the "Kobayashi-Maskawa matrix" describing the mixing of different quarks. The $SU(3) \times SU(2) \times U(1)$ standard model has quark fields U_{Ri}, D_{Ri} neutral under $SU(2)$ and

$$\psi_{iL} = \begin{pmatrix} U_{Li} \\ D_{Li} \end{pmatrix} \tag{17.14}$$

forming a doublet under the gauge group $SU(2)$. Here i is a *flavor index* running over the families of quarks. In nature, at observable energies, i runs from 1 to 3. $SU(3)$ color and spin indices are suppressed in this discussion.

In the Standard Model there is also a doublet of scalars fields, the Higgs field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \qquad \tilde{\phi} = \begin{pmatrix} (\phi^0)^* \\ \phi^- \end{pmatrix} \tag{17.15}$$

where $\phi^- = -(\phi^+)^*$ and $\phi^0, \phi^+ \in \mathbb{C}$. Both $\phi, \tilde{\phi}$ transform as $SU(2)$ doublets.

The Yukawa terms coupling the Higgs to the quarks lead to two "quark mass terms":

$$-g_{jk}\bar{D}_{jR}\phi^\dagger\psi_{kL} + hc \tag{17.16}$$

and

$$-\tilde{g}_{jk}\bar{U}_{jR}\tilde{\phi}^{\dagger}\psi_{kL} + hc \tag{17.17}$$

The matrices g_{jk} and $\tilde{g}_{jk}$ are assumed to be generic, and must be fixed by experiment.

For energetic reasons the scalar field ϕ_0 is nonzero (and constant) in nature. ("Develops a vacuum expectation value"). At low energies, $\phi^{\pm} = 0$ and hence these terms in the Lagrangian simplify to give mass terms to the quarks:

$$\sum_{i,j=1}^{3} \bar{U}_{Ri}M^{U}_{ij}U_{Lj} \tag{17.18}$$

and

$$\sum_{i,j=1}^{3} \bar{D}_{Ri}M^{D}_{ij}D_{Lj} \tag{17.19}$$

Here i, j run over the quark flavors, $U, \bar{U}, D, \bar{D}$ are quark wavefunctions. M^U, M^D are arbitrary complex matrices. We would like to go to a "mass eigenbasis" by bi-diagonalizing them with positive entries. The positive entries are identified with quark masses.

By bidiagonalization we know that

$$M^{D}_{ij} = (V_1^{\dagger}m^{D}V_2)_{ij} \tag{17.20}$$

$$M^{U}_{ij} = (V_3^{\dagger}m^{U}V_4)_{ij} \tag{17.21}$$

where m^D, m^U are diagonal matrices with real nonnegative eigenvalues and V_s, $s = 1, 2, 3, 4$ are four 3×3 unitary matrices. It is important that the V_s, are unitary because we want to use them to redefine our quark fields without changing the kinetic energy terms.

How much physical information is in the unitary matrices $V_1, \dots, V_4$? We would like to rotate away the unitary matrices by a field redefinition of the quark fields. The rest of the Lagrangian looks like (again suppressing $SU(3)$ color and hence the gluon fields):

$$\begin{aligned} &\bar{U}_{jR}(\gamma\cdot\partial + \frac{2}{3}\gamma\cdot B)U_{jR} + \bar{D}_{jR}(\gamma\cdot\partial - \frac{1}{3}\gamma\cdot B)D_{jR} \\ &+ \bar{\psi}_{jL}(\gamma\cdot\partial + \frac{1}{6}\gamma\cdot B + \gamma\cdot A)\psi_{jL} \end{aligned} \tag{17.22}$$

where we just sum over the flavor index and the operator $\gamma\cdot\partial + q\gamma\cdot B$ with $q \in \mathbb{R}$ is diagonal for our purposes. Crucially, it is diagonal in the flavor space. The $SU(2)$ gauge field A_μ has off-diagonal components $W^{\pm}_{\mu}$:

$$A_\mu = \begin{pmatrix} A^3_\mu & W^+_\mu \\ W^-_\mu & -A^3_\mu \end{pmatrix} \tag{17.23}$$

leading, among other things, to the charged current interaction

$$W^+_\mu \bar{U}_{Li}\gamma^\mu D_{Li} \tag{17.24}$$ eq:cc-int

Clearly, we can rotate away V_1, V_3 by a field redefinition of D_{jR}, U_{jR} (taking the case of three flavors and giving the quarks their conventional names):

$$\begin{aligned} \begin{pmatrix} \bar{d} \ \bar{s} \ \bar{b} \end{pmatrix}_R &= \bar{D}_R V_1^\dagger \\ \begin{pmatrix} \bar{u} \ \bar{c} \ \bar{t} \end{pmatrix}_R &= \bar{U}_R V_3^\dagger \end{aligned} \tag{17.25}$$

However, we also need to rotate U_{jL} and D_{jL} to a mass eigenbasis by *different* matrices V_2 and V_4:

$$\begin{aligned} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L &= V_2 D_L \\ \begin{pmatrix} u \\ c \\ t \end{pmatrix}_L &= V_4 U_L \end{aligned} \tag{17.26}$$

Therefore the charged current interaction (17.24) when expressed in terms of mass eigenstate fields is not diagonal in "flavor space." Rather, when we rotate to the mass basis the unitary matrix $S = V_4 V_2^\dagger$ enters in the charged current

$$\begin{pmatrix} \bar{u} \ \bar{c} \ \bar{t} \end{pmatrix}_L \gamma^\mu S \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{17.27}$$

where u, c, t, d, s, b are mass eigenstate fields.

The unitary matrix S is called the Kobayashi-Maskawa matrix. It is still not physically meaningful because that by using further diagonal phase redefinitions that S only depends on 4 physical parameters, instead of the 9 parameters in an arbitrary unitary matrix.

Much effort in current research in experimental particle physics is devoted to measuring the matrix elements experimentally.

Reference: H. Georgi, *Weak Interactions and Modern Particle Theory.*

Exercise

Repeat the above discussion for N quark flavors. How many physically meaningful parameters are there in the weak interaction currents?

17.2.3 Singular value decomposition

The singular value decomposition applies to any matrix $A \in M_{m\times n}(\mathbb{C})$ and generalizes the bidiagonalization of square matrices. It has a wide variety of applications.

Theorem Suppose that $A \in M_{m\times n}(\mathbb{C})$, which we can take, WLOG so that $m \le n$. Then there exist unitary matrices $U \in U(m)$ and $V \in U(n)$ so that

$$UAV = \begin{pmatrix} \Lambda_{m\times m} & 0_{m\times(n-m)} \end{pmatrix} \tag{17.28}$$ eq:sing-val-de

where Λ is a diagonal matrix with *nonnegative* entries (known as the *singular values* of A).

Proof: We have already proven the case with $m = n$ so assume that $m < n$. Enhance A to

$$\hat{A} = \begin{pmatrix} A \\ 0_{(n-m)\times n} \end{pmatrix} \tag{17.29}$$

Then bidiagonalization gives $\hat{U}, \hat{V} \in U(n)$ so that $\hat{U}\hat{A}\hat{V}$ is diagonal. Note that

$$\hat{U}\hat{A}\hat{A}^\dagger\hat{U} = \begin{pmatrix} D & 0_{m\times(n-m)} \\ 0_{(n-m)\times m} & 0_{(n-m)\times(n-m)} \end{pmatrix} \tag{17.30}$$

and hence if we break up $\hat{U}$ into blocks

$$\hat{U} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \tag{17.31}$$

then $U = U_{11}$ is in fact a unitary matrix in $U(m)$. But then

$$UA\hat{V} = \begin{pmatrix} \Lambda_{m\times m} & 0_{m\times(n-m)} \end{pmatrix} \tag{17.32}$$

and we can WLOG take Λ to have nonnegative entires ♠

Remarks

1. The eigenvalues of D are known as the *singular values* of A, the the decomposition (17.28) is known as the *singular value decomposition* of A.

2. The singular value decomposition can be rephrased as follows: If $T : V_1 \to V_2$ is a linear map between finite dimensional inner product spaces V_1 and V_2 then there exist ON sets $\{u_n\} \subset V_1$ and $\{w_n\} \subset V_2$ (not necessarily complete) and positive numbers λ_n so that

$$T = \sum_{n=1}^{N} \lambda_n (u_n, \cdot) w_n \tag{17.33}$$ eq:SVP-DEC-2

17.2.4 Schmidt decomposition

Let us consider a tensor product of two finite-dimensional inner product spaces $V_1 \otimes V_2$. We will assume, WLOG that $\dim V_1 \le \dim V_2$. Then a vector $v \in V_1 \otimes V_2$ is called *separable* or *primitive* if it is of the form $v = v_1 \otimes v_2$ where $v_1 \in V_1$ and $v_2 \in V_2$. In quantum information theory vectors which are not separable are called *entangled*. The reader might want to keep in mind that V_1, V_2 are finite-dimensional Hilbert spaces.

The general vector in $V_1 \otimes V_2$ is a linear combination of separable vectors. Schmidt decomposition is a canonical form of this linear combination:

Theorem Given an arbitary vector $v \in V_1 \otimes V_2$ there exist ordered ON bases $\{u_i\}_{i=1}^{\dim V_1}$ for V_1 and $\{w_a\}_{a=1}^{\dim V_2}$ for V_2, so that

$$v = \sum_{i=1}^{\dim V_1} \lambda_i u_i \otimes w_i \tag{17.34}$$

with $\lambda_i \geq 0$.

Proof: Choose arbitrary ON bases $\{\tilde{u}_i\}$ for V_1 and $\{\tilde{w}_a\}$ for V_2. Then we can expand

$$v = \sum_{i=1}^{\dim V_1} \sum_{a=1}^{\dim V_2} A_{ia} \tilde{u}_i \otimes \tilde{w}_a \tag{17.35}$$

where A is a complex $m \times n$ matrix. Now from the singular value decomposition we can write $A = UDV$, where

$$D = \begin{pmatrix} \Lambda & 0 \end{pmatrix} \tag{17.36}$$

and Λ is an $m \times m$ diagonal matrix with nonnegative entries. Now use U, V to change basis. ♠

Remark: The rank of A is known as the *Schmidt number*. If it is larger than one and v is a quantum state in a bipartite system then the state is entangled.

18. Operators on Hilbert space

18.1 Lies my teacher told me

c:LieMyTeacher

18.1.1 Lie 1: The trace is cyclic:

ubsec:FirstLie

But wait! If

$$\mathrm{Tr} AB = \mathrm{Tr} BA \tag{18.1}$$

then

$$\mathrm{Tr}[A, B] = 0 \tag{18.2}$$

and if we consider $q\psi(x) = x\psi(x)$ and $p\psi(x) = -i\frac{d}{dx}\psi(x)$ on $L^2(\mathbb{R})$ then $[p, q] = -i1$ and $\mathrm{Tr}(1) = \infty$, not zero!

18.1.2 Lie 2: Hermitian operators have real eigenvalues

bsec:SecondLie

But wait! Let us consider $A = q^n p + p q^n$ acting on $L^2(\mathbb{R})$ with $n > 1$. Since p, q are Hermitian this is surely Hermitian. But then

$$A\psi = \lambda\psi \tag{18.3}$$

is a simple first order differential equation whose general solution is easily found to be

$$\psi(x) = \kappa x^{-n/2} \exp[-\frac{i\lambda}{2(n-1)} x^{1-n}] \tag{18.4}$$

where κ is a constant. Then,

$$|\psi(x)|^2 = |\kappa|^2 |x|^{-n} \exp[\frac{\mathrm{Im}\lambda}{(n-1)} x^{1-n}] \tag{18.5}$$

This will be integrable for $x \to \pm\infty$ for $n > 1$ and it will be integrable at $x \to 0$ for $\mathrm{Im}\lambda < 0$ and n odd. Thus it would appear that the spectrum of $q^n p + pq^n$ is the entire lower half-plane!

18.1.3 Lie 3: Hermitian operators exponentiate to form one-parameter groups of unitary operators

ubsec:ThirdLie

But wait! Let us consider p on $L^2[0,1]$. Then $\exp[iap] = \exp[a\frac{d}{dx}]$ is the translation operator

$$(\exp[a\frac{d}{dx}]\psi)(x) = \psi(x+a) \tag{18.6}$$

But this can translate a wavefunction with support on the interval right off the interval! How can such an operator be unitary?!

18.2 Hellinger-Toeplitz theorem

sec:HT-Theorem

One theorem which points the way to the resolution of the above problems is the Hellinger-Toeplitz theorem. First, we begin with a definition:

Definition: A *symmetric everywhere-defined* linear operator, T, on $\mathcal{H}$ is an operator such that

$$(x, Ty) = (Tx, y) \qquad \forall x, y \in \mathcal{H} \tag{18.7}$$

We can also call this an everywhere defined self-adjoint operator. Then one might find the Hellinger-Toeplitz theorem a bit surprising:

Theorem: A symmetric everywhere-defined linear operator must be bounded.

In order to prove this one must use another theorem called the "closed graph theorem." This is one of three closely related theorems from the theory of operators between Banach spaces, the other two being the "bounded inverse theorem" and the "open mapping theorem."

Theorem[Bounded Inverse Theorem]: If $T : \mathcal{B}_1 \to \mathcal{B}_2$ is a bounded 1-1 operator from one Banach space onto (i.e. surjectively) another then $T^{-1} : \mathcal{B}_2 \to \mathcal{B}_1$ is bounded (hence continuous).

Proof: See Reed-Simon.

An immediate consequence of this is the

Theorem[Closed Graph Theorem]: If $T : \mathcal{B}_1 \to \mathcal{B}_2$ is a linear operator then the *graph of* T, defined by

$$\Gamma(T) := \{x \oplus Tx | x \in \mathcal{B}_1\} \subset \mathcal{B}_1 \oplus \mathcal{B}_2 \tag{18.8}$$

is a closed subspace of $\mathcal{B}_1 \oplus \mathcal{B}_2$ iff T is bounded.

Proof: This follows immediately from the bounded inverse theorem: Note that $\Gamma(T)$ is a normed linear space. Therefore, if it is closed it is itself a Banach space. Next, the map $\Gamma(T) \to \mathcal{B}_1$ given by $x \oplus Tx \to x$ is 1-1 bounded and continuous. Therefore the inverse $x \to x \oplus Tx$ is bounded which implies T is bounded. Conversely, if T is bounded then it is continuous so $\Gamma(T)$ is closed. ♠

Given the closed graph theorem the proof of the Hellinger-Toeplitz theorem is quite elegant. We need only prove that the graph of an everywhere-defined symmetric operator $T : \mathcal{H} \to \mathcal{H}$ is closed. Suppose then that $x_n \oplus Tx_n$ is a Cauchy sequence in $\Gamma(T) \subset \mathcal{H} \oplus \mathcal{H}$. Since $\Gamma(T) \subset \mathcal{H} \oplus \mathcal{H}$ and since $\mathcal{H} \oplus \mathcal{H}$ is a Hilbert space we know that

$$\lim_{n\to\infty} x_n \oplus Tx_n = x \oplus y \tag{18.9}$$

for some $x \oplus y \in \mathcal{H} \oplus \mathcal{H}$. Then if $y = Tx$ it will follow that $\Gamma(T)$ is closed. Now we note that for all $z \in \mathcal{H}$:

$$\begin{aligned}
(z, y) &= \lim_{n\to\infty} (z, Tx_n) && \text{def. of } y \\
&= \lim_{n\to\infty} (Tz, x_n) && T \quad \text{is symmetric} \\
&= (Tz, x) && \text{def. of } x \\
&= (z, Tx) && T \quad \text{is symmetric}
\end{aligned} \tag{18.10}$$

Therefore $y = Tx$, so $\Gamma(T)$ is closed, so T is bounded ♠

Now, many of the operators of interest in quantum mechanics are clearly unbounded, for example, the multiplication operator q on $L^2(\mathbb{R})$ satisfies

$$\parallel q\psi \parallel^2 = \int_{\mathbb{R}} x^2 |\psi(x)|^2 dx \tag{18.11}$$

Clearly there are wavefunctions with $\parallel \psi \parallel = 1$ but with support at arbitrarily large x, so q is unbounded. On the other hand it is equally obvious that q is symmetric. There is no contradiction with the HT theorem because of course it is not everywhere defined. Indeed, suppose $\psi(x)$ is a smooth square-integrable function decaying like $x^{-1-\epsilon}$ at large $|x|$ for some ϵ. For $0 < \epsilon \leq \frac{1}{2}$ the wavefunction $x\psi(x)$ is not square-integrable. Similar remarks apply to standard operators such as p and Schrödinger operators. These operators are only *partially defined*, that is, they are only defined on a linear subspace $W \subset \mathcal{H}$. We return to this theme in Section §18.6.

Exercise *Puzzle to resolve*

Consider $\mathcal{H} = \ell_2$ and define $T \in \mathcal{L}(\mathcal{H})$ by

$$T(e_n) = t_n e_n \tag{18.12}$$

where $\{t_n\}$ is a sequence of nonzero complex numbers with $\sum |t_n|^2 < \infty$.

a.) Show that T is an injective bounded operator.

b.) It would seem that this diagonal matrix has an obvious inverse

$$T^{-1}(e_n) = \frac{1}{t_n} e_n \tag{18.13}$$

On the other hand, such an operator is obviously unbounded! Why doesn't this contradict the Bounded Inverse Theorem?

---

ctrumResolvent

18.3 Spectrum and resolvent

Given a bounded operator $T \in \mathcal{L}(\mathcal{H})$ we partition the complex plane into two sets:

Definition:

1. The *resolvent set* or *regular set* of T is the subset $\rho(T) \subset \mathbb{C}$ of complex numbers so that $\lambda 1 - T$ is a bijective, i.e. a $1-1$ onto transformation $\mathcal{H} \to \mathcal{H}$.

2. The *spectrum of* T is the complement: $\sigma(T) := \mathbb{C} - \rho(T)$.

Now there are exactly three mutually exclusive ways the condition that $(\lambda 1 - T)$ is 1-1 can go wrong, and this leads to the decomposition of the spectrum:

Definition: The spectrum $\sigma(T)$ can be decomposed into three disjoint sets:

$$\sigma(T) = \sigma_{\text{point}}(T) \cup \sigma_{\text{res}}(T) \cup \sigma_{\text{cont}}(T) \tag{18.14}$$

1. If $\ker(\lambda 1 - T) \neq \{0\}$, that is, there is an eigenvector of T in $\mathcal{H}$ with eigenvalue λ, then λ is in the *point spectrum.*

2. If $\ker(\lambda 1 - T) = \{0\}$ but $\text{Im}(\lambda 1 - T)$ is not dense, then λ is in the *residual spectrum.*

3. If $\ker(\lambda 1 - T) = \{0\}$ and $\text{Im}(\lambda 1 - T)$ is dense but not all of $\mathcal{H}$, then λ is in the *continuous spectrum.*

Note that by the bounded inverse theorem, if $\lambda \in \rho(T)$ then the inverse, known as the *resolvent*

$$R_\lambda := (\lambda 1 - T)^{-1}, \tag{18.15}$$

is bounded. Now, for any bounded operator T we can try to expand

$$R_\lambda := \frac{1}{\lambda 1 - T} = \frac{1}{\lambda}\left(1 + \sum_{n=1}^{\infty}\left(\frac{T}{\lambda}\right)^n\right) \tag{18.16}$$

eq:Res-Exp

This converges in the norm topology for $|\lambda| > \parallel T \parallel$. One can check that then the series is an inverse to $\lambda 1 - T$ and is bounded. It follows that $\sigma(T)$ is inside the closed unit disk of radius $\parallel T \parallel$.

Similarly the condition that λ be in $\rho(T)$ is an *open condition*: If $\lambda \in \rho(T)$ then so is every complex number in some neighborhood of λ. Therefore, $\sigma(T)$ is a closed subset of $\mathbb{C}$. More formally, we can prove:

Theorem: The resolvent set $\rho(T)$ is open (hence $\sigma(T)$ is closed) and in fact

$$\parallel R_\lambda \parallel \leq \frac{1}{\mathrm{dist}(\lambda, \sigma(T))} \tag{18.17}$$

eq:NormDist

Proof: Suppose $\lambda_0 \in \rho(T)$. Consider the formal expansion

$$\begin{aligned} R_\lambda(T) &= \frac{1}{\lambda - T} = \frac{1}{\lambda - \lambda_0 + \lambda_0 - T} \\ &= \frac{1}{\lambda_0 - T} \frac{1}{1 - \frac{\lambda_0 - \lambda}{\lambda_0 - T}} \\ &= R_{\lambda_0}(T) \left[1 + \sum_{n=1}^{\infty} (\lambda_0 - \lambda)^n (R_{\lambda_0}(T))^n \right] \end{aligned} \tag{18.18}$$

Therefore, if $\parallel R_{\lambda_0}(T) \parallel |\lambda - \lambda_0| < 1$ then the series converges in the norm topology. Once we know it converges the formal properties will in fact be true properties and hence the series represents $R_\lambda(T)$, which will be bounded. Hence $\lambda \in \rho(T)$ for those values ♠

Remarks:

1. The proof shows that the map $\lambda \to R_\lambda(T)$ from $\rho(T)$ to $\mathcal{L}(\mathcal{H})$ is holomorphic.

Definition For any everywhere defined bounded operator T we define the *adjoint of* T, denoted $T^\dagger$ exactly as in the finite-dimensional case:

$$(Tx, y) = (x, T^\dagger y) \qquad \forall x, y \in \mathcal{H} \tag{18.19}$$

Remark If $\lambda \in \sigma_{\mathrm{res}}(T)$ then $\lambda^* \in \sigma_{\mathrm{point}}(T^\dagger)$. To see this, suppose that $\mathrm{im}(\lambda 1 - T)$ is not dense. Then there is some nonzero vector y not in the closed subspace $\overline{\mathrm{im}(\lambda 1 - T)}$ and by the projection theorem we can take it to be orthogonal to $\overline{\mathrm{im}(\lambda 1 - T)}$. That means

$$(y, (\lambda 1 - T)x) = ((\lambda^* 1 - T^\dagger)y, x) = 0 \tag{18.20}$$

for all x, which means that y is an eigenvector of $T^\dagger$ with eigenvalue λ^*. Therefore

$$\sigma_{\mathrm{res}}(T)^* \subset \sigma_{\mathrm{point}}(T^\dagger) \tag{18.21}$$

Example Let us return to the shift operator, or Hilbert hotel operator $S \in \mathcal{L}(\mathcal{H})$ for $\mathcal{H} = \ell^2$:

$$S : (x_1, x_2, \dots) \mapsto (0, x_1, x_2, \dots) \tag{18.22}$$

♣Notation clash: x_i are complex numbers, not a sequence of vectors in Hilbert space. ♣

In terms of harmonic oscillators $S = \frac{1}{\sqrt{a^\dagger a}} a^\dagger$. This is bounded and everywhere defined and one easily computes the adjoint, which just shifts to the left:

$$S^\dagger : (x_1, x_2, \dots) \mapsto (x_2, x_3 \dots) \tag{18.23}$$

Or, if one prefers, $S^\dagger = \frac{1}{\sqrt{a^\dagger a + 1}} a$.

Applying the remark of (18.16) to our case, where one easily checks $\parallel S \parallel = \parallel S^\dagger \parallel = 1$, we conclude that both $\sigma(S)$ and $\sigma(S^\dagger)$ are contained in the closed unit disk D.

Now, one easily shows that if $|\lambda| < 1$ then

$$\Psi_\lambda := (1, \lambda, \lambda^2, \dots) \tag{18.24}$$

is in ℓ^2 and

$$S^\dagger \Psi_\lambda = \lambda \Psi_\lambda \tag{18.25}$$

Since the spectrum must be closed it must be that $\sigma(S^\dagger)$ is the entire closed unit disk D.

On the other hand, let us consider the solutions to

$$y = (\lambda 1 - S)x \tag{18.26}$$

In terms of the components this leads to the equations

$$\begin{aligned} y_1 &= \lambda x_1 \\ y_2 &= \lambda x_2 - x_1 \\ \vdots &\quad \vdots \\ y_n &= \lambda x_n - x_{n-1} \\ \vdots &\quad \vdots \end{aligned} \tag{18.27}$$

which is easily solved, at least formally, to give

$$x_n = \lambda^{-1} y_n + \lambda^{-2} y_{n-1} + \cdots + \lambda^{-n} y_1 \tag{18.28}$$

It immediately follows that if $y = 0$ then $x = 0$, that is, the kernel of $(\lambda 1 - S)$ is $\{0\}$ and hence there is no point spectrum for S. Moreover, if $|y_1| = 1$ while $y_2 = y_3 = \cdots = 0$ then $x_n = \lambda^{-n} y_1$ is clearly not normalizable for $|\lambda| \le 1$. Now let ξ_n be a sequence of positive numbers monotonically decreasing to zero so that $\sum \xi_n^2 |\lambda|^{-2n}$ does not converge. Then using the triangle inequality one easily checks that

$$|x_n| > \xi_n |\lambda^{-n} y_1| \tag{18.29}$$

and hence $\{x_n\}$ cannot be normalizable for $|y_n| < (\xi_{n-1} - \xi_n)|\lambda^{1-n} y_1|$. Therefore $\mathrm{im}(\lambda 1 - S)$ does not intersect the open ball defined by $|y_n| < (\xi_{n-1} - \xi_n)|\lambda^{1-n} y_1|$. and hence is not dense for $|\lambda| \leq 1$. Thus $\sigma(S)$ is the closed unit disk, the point spectrum is zero and the residual spectrum is D.

Finally, we need to consider the nature of the spectrum of $S^\dagger$ when $|\lambda| = 1$. Then Ψ_λ is not in the Hilbert space so there is no point spectrum. On the other hand, if $\mathrm{im}(\lambda 1 - S^\dagger)$ were not dense then there would be a y not in its closure and by the projection theorem we can take it to be orthogonal to $\mathrm{im}(\lambda 1 - S^\dagger)$. But this means $(y, (\lambda - S^\dagger)x) = 0$ for all x which means $(\lambda^* 1 - S)y = 0$, but we know that S has no eigenvectors. Thus $|\lambda| = 1$ is in the spectrum of $S^\dagger$ but is *neither* in the point nor the residual spectrum! We conclude that $\mathrm{im}(\lambda 1 - S^\dagger)$ is dense, but not equal to $\mathcal{H}$. To exhibit a vector outside the image we can try to solve $y = (\lambda 1 - S^\dagger)x$. The formal solution is

$$\begin{aligned} x_1 &= \lambda^{-1} y_1 + \lambda^{-2} y_2 + \lambda^{-3} y_3 + \cdots \\ x_2 &= \lambda^{-1} y_2 + \lambda^{-2} y_3 + \lambda^{-3} y_4 + \cdots \\ x_3 &= \lambda^{-1} y_3 + \lambda^{-2} y_4 + \lambda^{-3} y_5 + \cdots \\ \vdots & \quad \vdots \end{aligned} \tag{18.30}$$

So, if we take $y_n = \lambda^n / n$ then $y \in \ell^2$ but is not in the image.

In summary we have the table:

Operator	Spectrum	Point Spectrum	Residual Spectrum	Continuous Spectrum
S	D	$\emptyset$	D	$\emptyset$
$S^\dagger$	D	Interior(D)	$\emptyset$	$\|\lambda\| = 1$

Theorem Suppose $T : \mathcal{H} \to \mathcal{H}$ is an everywhere defined self-adjoint operator. Then

1. The spectrum $\sigma(T) \subset \mathbb{R}$ is a subset of the real numbers.
2. The residual spectrum $\sigma_{\mathrm{res}}(T) = \emptyset$.

Proof: The usual proof from elementary quantum mechanics courses shows that the point spectrum is real: If $Tx = \lambda x$ then

$$\lambda(x, x) = (x, Tx) = (Tx, x) = \lambda^*(x, x) \tag{18.31}$$

so $\lambda \in \mathbb{R}$.

Now we show the residual spectrum is empty. We remarked above that $\sigma_{\rm res}(T)^* \subset \sigma_{\rm point}(T^\dagger)$. But if $T^\dagger = T$ then if $\lambda \in \sigma_{\rm res}(T)$ we must have $\lambda^* \in \sigma_{\rm point}(T)$ and hence λ is real, but this is impossible since the point and residual spectra are disjoint.

Now, let λ and μ be any real numbers and compute

$$\| (T-(\lambda+i\mu))x \|^2 = (x, (T-\lambda+i\mu)(T-\lambda-i\mu)x) = (x, ((T-\lambda)^2+\mu^2)x) = \| (T-\lambda)x \|^2 + \mu^2 \| x \|^2 \tag{18.32}$$

eq:smeng

This shows that if $\mu \neq 0$ then $T-(\lambda+i\mu)$ is invertible. If we let $x = (T-(\lambda+i\mu))^{-1}y$ then (18.32) implies that $\mu^2 \| x \|^2 \leq \| y \|^2$. But this means that

$$\frac{\| (T-(\lambda+i\mu))^{-1}y \|}{\| y \|} = \frac{\| x \|}{\| y \|} \leq \frac{1}{|\mu|} \tag{18.33}$$

and hence $(T-(\lambda+i\mu)$ has a bounded inverse for $\mu \neq 0$. Therefore, $\lambda + i\mu \in \rho(T)$ for $\mu \neq 0$ and hence $\sigma(T) \subset \mathbb{R}$. ♠

A useful criterion for telling when $\lambda \in \sigma(T)$ is the following:

Definition: A *Weyl sequence* is [16] a sequence of vectors $z_n \in D(T)$ such that $\| z_n \| = 1$ and $\| (\lambda - T)z_n \| \to 0$.

Theorem: [Weyl criterion]

a.) If T has a Weyl sequence then $\lambda \in \sigma(T)$.

b.) If λ is on the boundary of $\rho(T)$ then T has a Weyl sequence.

Proof:

a.) If there is a Weyl sequence and $\lambda \in \rho(T)$ then

$$1 = \| z_n \| = \| R_\lambda(T)(\lambda - T)z_n \| \leq \| R_\lambda(T) \| \| (\lambda - T)z_n \| \to 0 \tag{18.34}$$

which is impossible. Therefore $\lambda \in \sigma(T)$.

b.) Suppose $\lambda \in \overline{\rho(T)} - \rho(T)$, then there is a sequence of complex numbers $\{\lambda_n\}$ with $\lambda_n \in \rho(T)$ and $\lambda_n \to \lambda$ such that $\mathrm{dist}(\lambda_n, \sigma(T)) \to 0$. Therefore, by (18.17) we know $\| R_{\lambda_n}(T) \| \nearrow +\infty$ and therefore there are vectors y_n so that

$$\frac{\| R_{\lambda_n}(T)y_n \|}{\| y_n \|} \nearrow +\infty \tag{18.35}$$

Now set $z_n = R_{\lambda_n}(T)y_n$. These will be nonzero (since $R_{\lambda_n}(T)$ is invertible) and hence we can normalize y_n so that $\| z_n \| = 1$. But then

$$\begin{aligned} \| (\lambda - T)z_n \| &= \| (\lambda - \lambda_n)z_n + (\lambda_n - T)z_n \| \\ &= \| (\lambda - \lambda_n)z_n + y_n \| \\ &\leq |\lambda - \lambda_n| + \| y_n \| \to 0 \end{aligned} \tag{18.36}$$

[16] We state it so that it applies to unbounded operators with dense domain $D(T)$. See below. For bounded operators take $D(T) = \mathcal{H}$.

so $\{z_n\}$ is a Weyl sequence. ♠

Example Consider the operator on $L^2[a,b]$ given by the position operator $q\psi(x) = x\psi(x)$. Clearly this is a bounded operator $\parallel q \parallel \leq b$. It is everywhere defined and symmetric, hence, it is self-adjoint. It does not have eigenvectors [17]. For any $x_0 \in (a,b)$ we can take good approximations to the Dirac delta function:

$$\psi_{\epsilon,x_0} = \left(\frac{2}{\pi}\right)^{1/4} \frac{1}{\epsilon^{1/2}} e^{-(x-x_0)^2/\epsilon^2} \tag{18.37}$$

and, on the real line $\parallel (q-x_0)^2 \psi_{\epsilon,x_0} \parallel^2 = \epsilon^2/\sqrt{2}$, so $(q-x_0)^{-1}$ could hardly be a bounded operator. Thus $\sigma(q) = [a,b]$ and the spectrum is entirely continuous spectrum.

Exercise *The C^* identity*
Show that

$$\parallel T^\dagger T \parallel = \parallel T \parallel^2 \tag{18.38}$$

Remark: In general, a Banach algebra [18] which has an anti-linear involution so that $(ab)^* = b^* a^*$ and which satisfies $\parallel a^* a \parallel = \parallel a \parallel^2$ is known as a *C^*-algebra.* There is a rather large literature on the subject. It can be shown that every C^* algebra is a †-closed subalgebra of the algebra of bounded operators on Hilbert space.

Exercise
Show that $S^\dagger S = 1$ but $SS^\dagger$ is not one, but rather is a projection operator. That means that S is an example of a *partial isometry.*

♣Say more? ♣

♣Give the example of the Harper operator $U + U^* + \lambda(V + V^*)$ whose spectrum is a Cantor set. ♣

18.4 Spectral theorem for bounded self-adjoint operators

Now we would like to explain the statement (but not the proof) of the spectral theorem for self-adjoint operators on Hilbert space - a major theorem of von Neumann.

We begin with an everywhere-defined self-adjoint operator $T \in \mathcal{L}(\mathcal{H})$. As we have seen, T is bounded and $\sigma(T) \subset \mathbb{R}$ is a disjoint union of the point and continuous spectrum.

The spectral theorem says - roughly - that in an appropriate basis T is just a multiplication operator, like $\psi(x) \to x\psi(x)$. Roughly, for each $\lambda \in \sigma(T)$ we choose eigenvectors

[17] The "position eigenstates" of elementary quantum mechanics are distributions, and are not vectors in the Hilbert space.

[18] A *Banach algebra* is an algebra which is also a complete normed vector space and which satisfies $\parallel ab \parallel \leq \parallel a \parallel \cdot \parallel b \parallel$, an inequality which is easily verified for $\mathcal{L}(\mathcal{H})$. So $\mathcal{L}(\mathcal{H})$ is an example of a Banach algebra.

$|\psi_{\lambda,i}\rangle$ where i indicates possible degeneracy of the eigenvalue and then we aim to write something like

$$T \sim \int_{\sigma(T)} \lambda \left(\sum_i |\psi_{\lambda,i}\rangle\langle\psi_{\lambda,i}| \right) d\mu_T(\lambda) \tag{18.39}$$

with some measure $\mu_T(\lambda)$ on the spectrum. Clearly, unless we have a discrete point spectrum with finite dimensional eigenspaces this representation is at best heuristic. In that latter case

$$d\mu_T(\lambda) = \sum_n \delta(\lambda - \lambda_n) d\lambda \tag{18.40}$$

where we sum over the distinct eigenvalues λ_n.

In order to give a precise and general formulation of the spectral theorem von Neumann introduced the notion of a *projection-valued measure*, which we will now define. First we need:

Definition: The (Borel) measurable subsets of the real line $\mathbb{R}$ is the smallest collection $\mathcal{B}(\mathbb{R})$ of subsets of the real line such that

1. All intervals $(a, b) \in \mathcal{B}(\mathbb{R})$.
2. $\mathcal{B}(\mathbb{R})$ is closed under complement: If $E \in \mathcal{B}(\mathbb{R})$ then $\mathbb{R} - E \in \mathcal{B}(\mathbb{R})$.
3. $\mathcal{B}(\mathbb{R})$ is closed under countable union.

Remarks:

1. These axioms imply that $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ and $\emptyset \in \mathcal{B}(\mathbb{R})$.
2. The good thing about this collection of subsets of $\mathbb{R}$ is that one can define a "good" notion of "size" or measure $\mu(E)$ of an element $E \in \mathcal{B}(\mathbb{R})$ such that $\mu((a, b)) = b - a$ and μ is additive on disjoint unions. It turns out that trying to define such a measure μ for arbitrary subsets of $\mathbb{R}$ leads to paradoxes and pathologies.
3. We say a property holds "almost everywhere" if the set where it fails to hold is of measure zero.

Definition: A *projection-valued measure* is a map

$$P : \mathcal{B}(\mathbb{R}) \to \mathcal{L}(\mathcal{H}) \tag{18.41}$$

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

$$P(E) = \mathrm{s} - \lim_{n\to\infty} \sum_{i=1}^{n} P(E_i) \tag{18.42}$$

where the convergence is in the strong topology.

Remarks

1. The meaning of convergence in the strong topology is that a sequence of operators $T_n \to T$ if, for all $x \in \mathcal{H}$, $\parallel T_n x - Tx \parallel \to 0$.

2. Given a PVM and a nonzero vector $x \in \mathcal{H}$ there is a corresponding ordinary measure P_x on $\mathbb{R}$. We define it on a measurable set E by

$$P_x(E) = \frac{(x, P(E)x)}{(x,x)} \tag{18.43}$$

♣No need to divide by (x,x). You can just use a positive measure not normalized to one. ♣

This is a measure because, as is easily verified: $P_x(\emptyset) = 0$, $P_x(\mathbb{R}) = 1$, $P_x(E) \geq 0$, and, if $E = \amalg_{i=1}^{\infty}$ is a countable *disjoint* union of sets $E_i \in \mathcal{B}(\mathbb{R})$ then

$$P_x(E) = \sum_{i=1}^{\infty} P_x(E_i) \tag{18.44}$$

3. It will be convenient below to use the notation

$$P_x(\lambda) := P_x((-\infty, \lambda]) \tag{18.45}$$

eq:Plambda

This will be a measureable function and the corresponding measure $dP_x(\lambda)$ has the property that

$$P_x(E) = \int_E dP_x(\lambda) \tag{18.46}$$

Theorem[Spectral Theorem for bounded operators] If T is an everywhere-defined self-adjoint operator on $\mathcal{H}$ then
1. There is a PVM P_T so that for all $x \in \mathcal{H}$, we have

$$\frac{(x, Tx)}{(x,x)} = \int_{\mathbb{R}} \lambda dP_{T,x}(\lambda) \tag{18.47}$$

eq:DiagMtx

where $P_{T,x}(\lambda)$ is the measurable function associated to P_T via (18.45).
2. If f is a (measurable) function on $\mathbb{R}$ then $f(T)$ makes sense and for all $x \in \mathcal{H}$, we have

$$\frac{(x, f(T)x)}{(x,x)} = \int_{\mathbb{R}} f(\lambda) dP_{T,x}(\lambda) \tag{18.48}$$

eq:DiagMtx-f

Rough idea of the proof: The basic idea of the proof is to look at the algebra of operators generated by T. This is a commutative algebra. For example, it contains all the polynomials in T. If we take some kind of closure then it will contain all *continuous functions of* T. This statement is known as the "continuous functional calculus." Now, this continuous algebra is identified with the continuous functions on the compact set $X = \sigma(T)$. Moreover, given any $x \in \mathcal{H}$ we have a linear map $\Lambda_T : C(X) \to \mathbb{R}$ given by

$$f \mapsto \Lambda_T(f) := (x, f(T)x) \tag{18.49}$$

Moreover, the map is positive, meaning that if f is positive then $\Lambda_T(f) \geq 0$. Then a general theorem - known as the Riesz-Markov theorem - says that any positive linear functional $C(X) \to \mathbb{R}$ where X is Hausdorff is of the form $f \mapsto \int_X f d\mu$. Therefore, given T and x there is a corresponding measure $\mu_{T,x}$ and we have

$$(x, f(T)x) = \int_{\sigma(T)} f d\mu_{T,x} \tag{18.50}$$

Now, using this equation one extends the continuous functional calculus to the Borel functional calculus - namely, now we make sense of operators $g(T)$ where g is not necessarily continuous, but at least measurable. In particular, if g is a characteristic function it is discontinuous, but $g(T)$ will be a projection operator. ♠

Remarks:

1. Note that (18.47) is enough to determine all the matrix elements of T. This equation determines (x, Tx) for all x and then we can use the polarization identity:

$$\begin{aligned}(x, Ty) = \frac{1}{4}\Big[&((x+y), T(x+y)) - ((x-y), T(x-y)) \\ &+ i((x-iy), T(x-iy)) - i((x+iy), T(x+iy))\Big]\end{aligned} \tag{18.51}$$

   which can also be written

$$4(y, Tx) = \sum_{k=0}^{3} i^k (x + i^k y, T(x + i^k y)) \tag{18.52}$$

   Note that on a real Hilbert space we cannot multiply by i, but then $(y, Tx) = (Tx, y) = (x, Ty)$ for self-adjoint T so that it suffices to work just with $x \pm y$ in the corresponding polarization identity.

2. Equation (18.48) is meant to capture the idea that in the block-diagonalized basis provided by P_T the operator T is diagonalized.

3. It follows from the definition of a PVM that if $\parallel x \parallel = 1$

$$(x, P(E)x) = \int_{\mathbb{R}} \chi_E(\lambda) dP_x(\lambda) = \int_E dP_x(\lambda) \tag{18.53}$$

   where $\chi_E(\lambda)$ is the characteristic function of the set E.

4. Using the previous remark Using this we can see that, as expected, for a self-adjoint operator T the PVM P_T has support on the spectrum $\sigma(T)$ in the sense that:

   $\lambda \in \sigma(T)$ *iff* $P_T((\lambda-\epsilon,\lambda+\epsilon)) \neq 0$ *for all* $\epsilon > 0$.

   To prove this suppose first that $P_T((\lambda_0-\epsilon,\lambda_0+\epsilon)) \neq 0$ for all $\epsilon > 0$. Then take the sets $E_n = (\lambda_0 - \frac{1}{n}, \lambda_0 + \frac{1}{n})$. Since $P_T(E_n)$ is nonzero we can take a sequence of nonzero vectors z_n in the image of $P_T(E_n)$ and normalize them to $\parallel z_n \parallel = 1$. Then

$$\begin{aligned} \parallel (T-\lambda_0)z_n \parallel^2 &= \parallel (T-\lambda_0)P_T(E_n)z_n \parallel^2 \\ &= \int_{\mathbb{R}} (\lambda-\lambda_0)^2 \chi_{E_n}(\lambda) dP_{T,z_n}(\lambda) \leq \frac{1}{n^2} \end{aligned} \tag{18.54}$$

   so we have a Weyl sequence and hence $\lambda \in \sigma(T)$. Conversely, suppose that $P_T((\lambda_0 - \epsilon, \lambda_0+\epsilon)) = 0$ for some $\epsilon > 0$. For such an ϵ define the function:

$$f_\epsilon(\lambda) := \begin{cases} 0 & |\lambda-\lambda_0| < \epsilon \\ \frac{1}{\lambda_0-\lambda} & |\lambda-\lambda_0| \geq \epsilon \end{cases} \tag{18.55}$$

   Then

$$\begin{aligned} (\lambda_0 - T)f_\epsilon(T) &= (\lambda_0-T)\int_{|\lambda_0-\lambda|\geq\epsilon} \frac{1}{\lambda_0-\lambda} dP_T(\lambda) \\ &= \int_{|\lambda_0-\lambda|\geq\epsilon} \frac{\lambda_0-\lambda}{\lambda_0-\lambda} dP_T(\lambda) \\ &= \int_{|\lambda_0-\lambda|\geq\epsilon} dP_T(\lambda) = 1 \end{aligned} \tag{18.56}$$

   Similarly, $f(T)(\lambda_0 - T) = 1_{D(T)}$. Therefore, $\lambda_0 - T$ is a bijection of $D(T)$ with $\mathcal{H}$ and hence $\lambda \in \rho(T)$. ♠

Example: Suppose T has a finite pure point spectrum $\{\lambda_n\}_{n=1}^N$ with eigenspaces V_{λ_n}. Then define

$$\delta_\lambda(E) = \int_E \delta(\lambda'-\lambda)d\lambda' = \begin{cases} 1 & \lambda \in E \\ 0 & \lambda \notin E \end{cases} \tag{18.57}$$

Then the projection value measure of T is

$$P_T(E) = \sum_{\lambda_n \in E} P_{V_{\lambda_n}} = \sum_{n=1}^N \delta_{\lambda_n}(E) P_{V_{\lambda_n}} \tag{18.58}$$

In particular, if $T = Id_{\mathcal{H}}$ is the unit operator then

$$P_T(E) = \begin{cases} Id_{\mathcal{H}} & 1 \in E \\ 0 & 1 \notin E \end{cases} \tag{18.59}$$

Exercise
Show that if P is a PVM then [19]

$$P(E_1)P(E_2) = P(E_1 \cap E_2) \tag{18.60}$$

18.5 Defining the adjoint of an unbounded operator

Recall from the Hellinger-Toeplitz theorem that an unbounded operator on an infinite-dimensional Hilbert space cannot be everywhere defined and self-adjoint. On the other hand, as we explained, physics requires us to work with unbounded self-adjoint operators.

Therefore, we should consider *partially defined* linear operators. That is, linear operators T from a proper subspace $D(T) \subset \mathcal{H}$ to $\mathcal{H}$. Giving the domain $D(T)$ of the operator is an essential piece of data in defining the operator.

Definition:
a.) The *graph of* T is the subset

$$\Gamma(T) := \{x \oplus Tx | x \in D(T)\} \subset \mathcal{H} \oplus \mathcal{H} \tag{18.61}$$

b.) T is *closed* if $\Gamma(T)$ is a closed subspace of $\mathcal{H} \oplus \mathcal{H}$.
c.) An operator T_2 is an *extension of an operator* T_1 if $\Gamma(T_1) \subset \Gamma(T_2)$. That is, $D(T_1) \subset D(T_2)$ and when T_2 is restricted to $D(T_1)$ it agrees with T_1. This is usually denoted $T_1 \subset T_2$.

Definition: If $D(T) \subset \mathcal{H}$ is dense then we define the subset $D(T^\dagger) \subset \mathcal{H}$ to be the set of $y \in \mathcal{H}$ so that there exists a $z \in \mathcal{H}$ such that for all $x \in D(T)$,

$$(Tx, y) = (x, z). \tag{18.62}$$

If $y \in D(T^\dagger)$ then z is unique (since $D(T)$ is dense) and we denote

$$z = T^\dagger y \tag{18.63}$$

This defines a linear operator $T^\dagger$ with domain $D(T^\dagger)$ called the *adjoint of* T.

Remark: One way of characterizing $D(T^\dagger)$ is that $y \in D(T^\dagger)$ iff $x \mapsto (y, Tx)$ extends to a to a *bounded* linear operator. Then z exists, by the Riesz representation theorem.

Definition: A densely defined operator T is

[19] *Answer*: First show that if $E_1 \cap E_2 = \emptyset$ then $P(E_1)P(E_2) = P(E_2)P(E_1) = 0$. Do that by using the PVM axiom to see that $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. Square this equation to conclude that $\{P(E_1), P(E_2)\} = 0$. But now, multiply this equation on the left and then on the right by $P(E_1)$ to show that $[P(E_1), P(E_2)] = 0$. Next, write $P(E_1) = P(E_1 - E_1 \cap E_2) + P(E_1 \cap E_2)$ and $P(E_2) = P(E_2 - E_1 \cap E_2) + P(E_1 \cap E_2)$ and multiply.

a.) *Symmetric* if $T \subset T^\dagger$.
b.) *Self-adjoint* if $T = T^\dagger$.

Remarks

1. Let us unpack this definition a bit. An operator T is symmetric iff

$$(x, Ty) = (Tx, y) \tag{18.64}$$

for all $x, y \in D(T)$. However, $x \to (Tx, y)$ might be bounded for a larger class of vectors $y \notin D(T)$. When T is self-adjoint this does not happen and $D(T^\dagger) = D(T)$.

2. Unfortunately, different authors use the term "Hermitian operator" in ways which are inequivalent for unbounded operators. Some authors (such as Reed and Simon) use the term to refer to symmetric operators while other authors (such as Takhtadjan) use the term to refer to self-adjoint operators. So we will use only "symmetric" and "self-adjoint" and reserve the term "Hermitian" for the finite-dimensional case, where no confusion can arise.

Example: Let us use the "momentum" $p = -i\frac{d}{dx}$ to define an operator with a dense domain $D(T)$ within $L^2[0,1]$. The derivative is clearly not defined on all L^2 functions. Now, a function $f : [0,1] \to \mathbb{C}$ is "absolutely continuous" if $f'(x)$ exists almost everywhere and $|f'(x)|$ is integrable. In particular, the fundamental theorem of calculus holds

$$f(b) - f(a) = \int_a^b f'(x)dx \tag{18.65}$$

In this case $f'(x)$ is indeed in $L^2[0,1]$.

We begin with an operator T defined by the domain:

$$D(T) = \{f | f \in AC[0,1], f(0) = f(1) = 0\} \tag{18.66}$$

then elementary integration by parts shows that $T = -i\frac{d}{dx}$ is a symmetric operator. An elaborate argument in Reed-Simon VIII.2, p. 257 shows that with this definition of T we have:

$$D(T^\dagger) = \{f | f \in AC[0,1]\} \tag{18.67}$$

and in Reed-Simon vol. 2, Section X.1, p. 141 it is shown that there is a one-parameter family of self-adjoint extensions $T_\alpha = T_\alpha^\dagger$ labeled by a phase α:

$$D(T_\alpha) = \{f | f \in AC[0,1], f(0) = \alpha f(1)\} \tag{18.68}$$

It easy to appreciate this even without all the heavy machinery of defining self-adjoint extensions of symmetric operators. Formally proving that the operator is symmetric requires that the boundary terms in the integration by parts vanishes, that is:

$$(T\psi_1, \psi_2) = (\psi_1, T\psi_2) \tag{18.69}$$

implies

$$\psi_1^*(1)\psi_2(1) - \psi_1^*(0)\psi_2(0) = 0 \tag{18.70}$$

If $\psi_2 \in D(T)$ then this will be satisfied for $\psi_1 \in D(T^\dagger)$ because both terms separately vanish. We can attempt to extend this definition to a larger domain. If we try to let both $\psi_1, \psi_2 \in D(T^\dagger)$ the condition will fail. The intermediate choice is to choose a phase α and require $\psi(1) = \alpha\psi(0)$.

Exercise
Suppose $T_1 \subset T_2$ are densely defined operators. Show that $T_2^\dagger \subset T_1^\dagger$.

---

ec:UnboundedSA

18.6 Spectral Theorem for unbounded self-adjoint operators

Having introduced the notion of projection valued measures we are now in a position to state the spectral theorem for (possibly unbounded) self-adjoint operators on Hilbert space:

The Spectral Theorem: There is a 1-1 correspondence between self-adjoint operators T on a Hilbert space $\mathcal{H}$ and projection valued measures such that:

a.) Given a PVM P a corresponding self-adjoint operator T_P can be defined by the diagonal matrix elements:

$$\frac{(x, T_P x)}{(x,x)} = \int_{\mathbb{R}} \lambda dP_x(\lambda) \tag{18.71}$$

eq:Gen-ST-i

with domain

$$D(T_P) = \{x \in \mathcal{H} | \int_{\mathbb{R}} \lambda^2 dP_x(\lambda) < \infty\} \tag{18.72}$$

eq:Gen-ST-ii

b.) Conversely, given T there is a corresponding PVM P_T such that (18.71) and (18.72) hold.

c.) Moreover, given a self-adjoint operator T if f is any (Borel measurable) function then there is an operator $f(T)$ with domain

$$D(f(T)) = \{x \in \mathcal{H} | \int_{\mathbb{R}} |f(\lambda)|^2 dP_{T,x}(\lambda) < \infty\} \tag{18.73}$$

eq:Gen-ST-iii

such that

$$\frac{(x, f(T)x)}{(x,x)} = \int_{\mathbb{R}} f(\lambda) dP_{T,x}(\lambda) \tag{18.74}$$

eq:Gen-ST-iv

Example 1. If T has pure point spectrum $\{\lambda_n\}$ with closed eigenspaces V_n then

$$T = \sum_n \lambda_n P_{V_n} \tag{18.75}$$

$$P_T(E) = \sum_{\lambda_n \in E} P_{V_n} \tag{18.76}$$

where P_{V_n} are the orthogonal projections to the subspaces V_n.

Example 2. Take $\mathcal{H} = L^2(\mathbb{R})$ and define $T = q$ by $q\psi(x) = x\psi(x)$ with a domain given by those wavefunctions with falloff at least as fast as $|x|^{-\alpha}$ at infinity, with $\alpha > 3/2$. Then

$$(P_T(E)\psi)(x) = \begin{cases} \psi(x) & x \in E \\ 0 & x \notin E \end{cases} \tag{18.77}$$

Example 3. Take $\mathcal{H} = L^2(\mathbb{R})$ and define $T = p$ by $p\psi(x) = -i\frac{d}{dx}\psi(x)$ with a domain given by absolutely continuous wavefunctions with L^2 integrable derivative. [20] Now, if we make a Fourier transform then this is again a multiplication operator, so now

$$(P_T(E)\psi)(x) = \int_{\mathbb{R}} dy \int_E \frac{dk}{2\pi} e^{ik(x-y)}\psi(y) \tag{18.78}$$

Remarks

1. This is a major theorem and proofs can be found in many places. We mention just Chapters VII and VIII of Reed-Simon, and Chapter 3 of G. Teschl, *Mathematical Methods in Quantum Mechanics.*

2. The resolvent set and spectrum of a *closed* but possibly unbounded operator are defined along the same lines as in the bounded case: $\lambda \in \rho(T)$ if $\lambda - T$ is a bijection of $D(T)$ onto $\mathcal{H}$. It follows that the resolvent $R_\lambda = (\lambda - T)^{-1}$ is a bounded operator $\mathcal{H} \to D(T)$. The spectrum is the complement of the resolvent set as before, and, as before, if T is self-adjoint $\sigma(T)$ is a subset of $\mathbb{R}$.

3. One can show that if T is a bounded self-adjoint operator then $\sigma(T)$ is a bounded subset of $\mathbb{R}$.

4. What is going on with our example $q^n p + pq^n$ in Section §18.1.2 above? At least a partial answer is that the most obvious domain on which one can prove the operator is symmetric (and hence has real point spectrum) is the set of wavefunctions so that $q^n p\psi$ is L^2. These must fall off as $\psi(x) \sim |x|^{-\alpha}$ for $\alpha > n - 1/2$ and the putative eigenfunctions exhibited above lie outside that domain.

18.7 Commuting self-adjoint operators

Once we start admitting partially defined operators on Hilbert space we have stepped onto a slippery slope. If T_1 and T_2 are only defined on $D(T_1)$ and $D(T_2)$ respectively then T_1+T_2 is only defined on $D(T_1) \cap D(T_2)$ and $T_1 \circ T_2$ is similarly only defined on

$$D(T_1 \circ T_2) = \{x | x \in D(T_2) \qquad \text{and} \qquad T_2(x) \in D(T_1)\} \tag{18.79}$$

[20] This is an example of something called a *Sobolev space.*

The problem is that these subspaces might be small, or even the zero vector space.

Example 1 Take any $y \notin D(T_1)$ and let $T_2(x) = (z, x)y$ for some z. Then $T_1 \circ T_2$ is only defined on the zero vector.

Example 2: Let $\{x_n\}$ be an ON basis for $\mathcal{H}$ and let $\{y_n\}$ be another ON basis so that each y_n is an infinite linear combination of the x_m's and vice versa. Then let $D(T_1)$ be the set of finite linear combinations of the x_n's and $D(T_2)$ be the set of finite linear combinations of the y_n's. Then $D(T_1)$ and $D(T_2)$ are dense and $D(T_1) \cap D(T_2) = \{0\}$. In order to produce an example of two such ON bases consider $\mathcal{H} = \ell^2(\mathbb{C})$ and take the Cayley transform of $T = \lambda(S + S^\dagger)$ where λ is real and of magnitude $|\lambda| < \frac{1}{2}$. Then

$$U = (1 + iT)(1 - iT)^{-1} \tag{18.80}$$

is a well-defined unitary operator which takes the standard basis $e_n = (0, \dots, 0, 1, 0 \dots)$ of ℓ^2 to a new basis f_n all of which are infinite linear combinations of the e_n.

Definition: Two self-adjoint operators T_1 and T_2 are said to *commute* if their PVM's commute, that is, for all $E_1, E_2 \in \mathcal{B}(\mathbb{R})$

$$P_{T_1}(E_1)P_{T_2}(E_2) = P_{T_2}(E_2)P_{T_1}(E_1) \tag{18.81}$$

When T_1, T_2 are bounded the spectral theorem shows that this reduces to the usual notion that $[T_1, T_2] = 0$.

18.8 Stone's theorem

Part of the proof of the spectral theorem involves showing that if $x \to f(x)$ is a continuous function, or more generally a measurable function, then if T is self-adjoint the operator $f(T)$ is densely defined and makes sense. In particular, if this is applied to the exponential function $f(x) = \exp[ix]$ one obtains an operator with domain all of $\mathcal{H}$ (by (18.73)) which is in fact a bounded operator. All the good formal properties that we expect of this operator are in fact true:

Theorem: If T is a self-adjoint operator then the family of operators $U(t) = \exp[itT]$ satisfies

1. $U(t)U(s) = U(t + s)$
2. $t \to U(t)$ is continuous in the strong operator topology.
3. The limit $\lim_{t\to 0}(U(t)x - x)/t$ exists iff $x \in D(T)$ in which case the limit is equal to $iT(x)$

♣Should explain that $t \to U(t)$ is continuous in the norm topology iff T is bounded. ♣

Stone's theorem, is a converse statement: First, we define a *strongly continuous one parameter group* is a homomorphism from $\mathbb{R}$ to the group of unitary operators on Hilbert space which is continuous in the strong topology, that is:

1. $U(t)U(s) = U(t + s)$

2. For each $x \in H$, $\lim_{t_1 \to t_2} U(t_1)x = U(t_2)x$.

Theorem [Stone's theorem]: If $U(t)$ is a strongly continuous one-parameter group of unitary operators on $\mathcal{H}$ then there is a self-adjoint operator T on $\mathcal{H}$ such that $U(t) = \exp[itT]$.

Remarks

1. If T is bounded then we can simply define

$$U(t) = \sum_{n=0}^{\infty} \frac{(itT)^n}{n!} \tag{18.82}$$

   This converges in the operator norm. In particular $t \to U(t)$ is continuous in the operator norm. However, such a definition will not work if T is an unbounded operator.

2. The proof is in Reed-Simon Theorem VIII.8, p.266.

3. Let us return to our third lie of §18.1.3. We cannot exponentiate $T = -id/dx$ on $L^2[0,1]$ simply because it is unbounded and not self-adjoint with the domains $D(T)$ and $D(T^\dagger)$ given above. Since T is not defined on functions with nonzero values at $x = 0, 1$ it is not surprising that we cannot define the translation of a wavepacket past that point. If we take one of the self-adjoint extensions T_α then we are working with twisted boundary conditions on the circle. Now is is quite sensible to be able to translate by an arbitrary amount around the circle.

4. What about the translation operator on the half-line? Our naive discussion above should make it clear that in this case p is not even *essentially self-adjoint.* [21] So there is no self-adjoint extension of p acting on the Sobelev space with $\psi(x) = 0$ at $x = 0$.

5. Stone's theorem can also be proven for continuity in the compact-open topology.

18.9 Traceclass operators

We would like to define the trace of an operator but, as the first lie in Section 18.1.1 shows, we must use some care.

For simplicity we will restrict attention in this section to *bounded operators.*

Definition: An operator is called *positive* if [22]

$$(x, Tx) \geq 0 \qquad\qquad \forall x \in \mathcal{H} \tag{18.83}$$

Three immediate and easy properties of positive bounded operators are:

Theorem: If T is a positive bounded operator on a complex Hilbert space $\mathcal{H}$ then

[21] If the closure of the graph $\overline{\Gamma(T)}$ is the graph of an operator we call that operator $\bar{T}$: $\overline{\Gamma(T)} = \Gamma(\bar{T})$, and we say that T is closeable. A symmetric operator T is essentially self-adjoint if $\bar{T}$ is self-adjoint.

[22] A more accurate term would be *nonnegative*. But this is standard terminology.

1. T is self-adjoint.

2. $|(x, Ty)|^2 \le (x, Tx)(y, Ty)$ for all $x, y \in \mathcal{H}$.

3. T has a unique positive square-root $S \in \mathcal{L}(\mathcal{H})$, i.e. S is positive and $S^2 = T$.

Proof:

1. Note that $(x, Tx) = \overline{(x, Tx)} = (Tx, x)$. Now, if $\mathcal{H}$ is a complex Hilbert space we can then use the polarization identity to prove $(x, Ty) = (Ty, x)$ and hence T is self-adjoint. The statement fails for real Hilbert spaces.

2. Consider the inequalities $0 \le (x + \lambda e^{i\theta} y, T(x + \lambda e^{i\theta} y))$, where λ is real and we choose a suitable phase $e^{i\theta}$ and require the discriminant of the resulting quadratic polynomial in λ to be nonpositive.

3. Follows immediately from the spectral theorem. ♠

Theorem/Definition: If T is a positive operator on a separable Hilbert space we define the trace of T by

$$\mathrm{Tr}(T) := \sum_{n=1}^{\infty} (u_n, T u_n) \tag{18.84}$$ eq:TraceDef-1

where $\{u_n\}$ is an ON basis for $\mathcal{H}$. This sum, (which might be infinite) does not depend on the ON basis.

Proof: The sum in (18.84) is a sum of nonnegative terms and hence the partial sums are strictly increasing. They either diverge to infinity or have a limit. We use the square-root property, namely $T = S^2$ with S self-adjoint and positive to check independence of basis. Let $\{v_m\}$ be any other ON basis:

$$\begin{aligned}\mathrm{Tr}(T) &= \sum_{n=1}^{\infty}(u_n, Tu_n) = \sum_{n=1}^{\infty} \parallel S u_n \parallel^2 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |(Su_n, v_m)|^2 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |(u_n, Sv_m)|^2 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |(u_n, Sv_m)|^2 = \sum_{m=1}^{\infty} \parallel S v_m \parallel^2 = \sum_{m=1}^{\infty} (v_m, T v_m)\end{aligned} \tag{18.85}$$

The exchange of infinite sums in going to the second line is valid because all terms are nonnegative ♠

We now use the squareroot lemma and the above theorem to define the traceclass ideal $\mathcal{I}_1$:

Definition: The *traceclass operators* $\mathcal{I}_1 \subset \mathcal{L}(\mathcal{H})$ are those operators so that $|T| := \sqrt{T^\dagger T}$ has a finite trace: $\mathrm{Tr}(|T|) < \infty$.

With this definition there is a satisfactory notion of trace:

Theorem: [Properties of traceclass operators]

1. $\mathcal{I}_1 \subset \mathcal{L}(\mathcal{H})$ is a $*$-closed *ideal*: This means: if $T_1, T_2 \in \mathcal{I}_1$ then $T_1 + T_2 \in \mathcal{I}_1$ and if $T_3 \in \mathcal{L}(\mathcal{H})$ then $T_1 T_3 \in \mathcal{I}_1$ and $T_3 T_1 \in \mathcal{I}_1$, and, finally, T is traceclass iff $T^\dagger$ is.

2. If $T \in \mathcal{I}_1$ the trace, defined by

$$\mathrm{Tr}(T) := \sum_{n=1}^{\infty} (u_n, T u_n) \tag{18.86}$$

where $\{u_n\}$ is any ON basis, is independent of the choice of ON basis and defines a linear functional $\mathcal{I}_1 \to \mathbb{C}$.

3. If $T_1 \in \mathcal{I}_1$ and $T_2 \in \mathcal{L}(\mathcal{H})$ then the trace is cyclic:

$$\mathrm{Tr}(T_1 T_2) = \mathrm{Tr}(T_2 T_1) \tag{18.87}$$

Proofs: The proofs are straightforward but longwinded. See Reed-Simon, Section VI.6.

Finally, we mention one more commonly used class of operators intermediate between traceclass and bounded operators:

Definition: The *compact operators* on $\mathcal{H}$, denoted $\mathcal{K}(\mathcal{H})$ is the norm-closure of the operators of finite rank. [23]

Thanks to the singular value decomposition, the canonical form of a compact operator follows immediately: There are ON sets $\{u_n\}$ and $\{w_m\}$ in $\mathcal{H}$ (not necessarily complete) and positive numbers λ_n so that

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) w_n \tag{18.88}$$

where the convergence of the infinite sum is in the operator norm. Hence the only possible accumulation point of the λ_n is zero. For a compact self-adjoint operator there is a complete ON basis $\{u_n\}$ with

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) u_n \tag{18.89}$$

where λ_n are real and $\lim_{n\to\infty} \lambda_n = 0$. This is called the Hilbert-Schmidt theorem.

Next, we have $\mathcal{I}_1 \subset \mathcal{K}(\mathcal{H})$. This follows because if T is traceclass so is $T^\dagger T$, but this implies that for any ON basis $\{u_n\}$ consider the linear span L_N of $\{u_n\}_{n=1}^N$. Then if $y \in L_N^\perp$

[23] This definition is fine for operators on Hilbert space but will not work for operators on Banach space. In that case one must use a different criterion, equivalent to the above for Hilbert spaces. See Reed-Simon Section VI.5.

is nonzero we can normalize it to $\parallel y \parallel = 1$ and since $L_N \cup \{y\}$ can be completed to an ON basis

$$\parallel Ty \parallel^2 + \sum_{n=1}^{N} \parallel Tu_n \parallel^2 \leq \mathrm{Tr} T^\dagger T < \infty \tag{18.90}$$

so

$$\parallel Ty \parallel^2 \leq \mathrm{Tr} T^\dagger T - \sum_{n=1}^{N} \parallel Tu_n \parallel^2 \tag{18.91}$$

Since $\mathrm{Tr} T^\dagger T < \infty$ the RHS goes to zero for $N \to \infty$. This means that

$$T = \sum_{n=1}^{\infty} (u_n, \cdot) T u_n \tag{18.92}$$

with converges in the operator norm because

$$\begin{aligned} \parallel T - \sum_{n=1}^{N} (u_n, \cdot) T u_n \parallel &= \mathrm{sup}_{y \neq 0} \frac{\parallel Ty - \sum_{n=1}^{N} (u_n, y) T u_n \parallel}{\parallel y \parallel} \\ &= \mathrm{sup}_{y \neq 0} \frac{\parallel T \left(y - \sum_{n=1}^{N} (u_n, y) u_n \right) \parallel}{\parallel y \parallel} \\ &= \mathrm{sup}_{y \neq 0} \frac{\parallel T y^\perp \parallel}{\sqrt{\parallel y^\perp \parallel^2 + \parallel y^\parallel \parallel^2}} \end{aligned} \tag{18.93}$$

where $y^\perp$ is the orthogonal projection to $L_N^\perp$.

In particular, for a positive traceclass operator T there is an ON basis with

$$T = \sum_{n=1}^{\infty} \lambda_n (u_n, \cdot) u_n \tag{18.94}$$

eq:Schmidt-Dec

where $\lambda_n \geq 0$, and $\mathrm{Tr}(T) = \sum_{n=1}^{\infty} \lambda_n$. This theorem is important when we discuss physical states and density matrices in quantum mechanics in Section §19 below.

Exercise

Give an example of a positive operator on a real Hilbert space which is not self-adjoint.
[24]

---

19. The Dirac-von Neumann axioms of quantum mechanics

sec:DvN-Axioms

The Dirac-von Neumann axioms attempt to make mathematically precise statements associated to the physical description of quantum systems.

[24] *Answer*: Consider $1 + J^{(2)}$ on $\mathbb{R}^2$.

1. *"Space of states"*: To a "physical system" we assign a complex separable $\mathbb{Z}_2$-graded Hilbert space $\mathcal{H}$ known as the "space of states."

2. *Physical observables*: The set of physical quantities which are "observable" in this system is in 1-1 correspondence with the set $\mathcal{O}$ of self-adjoint operators on $\mathcal{H}$.

3. *Physical states*: The set of physical states of the quantum system is in 1-1 correspondence with the set $\mathcal{S}$ of positive (hence self-adjoint) trace-class operators ρ with $\mathrm{Tr}(\rho) = 1$.

4. *Physical Measurement*: Physical measurements of an observable $T \in \mathcal{O}$ when the system is in a state $\rho \in \mathcal{S}$ are governed by a probability distribution $P_{T,\rho}$ on the real line of possible outcomes. The probability of measuring the value in a set $E \in \mathcal{B}(\mathbb{R})$ is defined by

$$P_{T,\rho}(E) = \mathrm{Tr} P_T(E)\rho \qquad (19.1)$$

eq:Born-VN

where P_T is the projection-valued measure of the self-adjoint operator T.

5. *Symmetries.* To state the postulate we need a definition:

Definition An *automorphism* of a quantum system to be a pair of bijective maps $\beta_1 : \mathcal{O} \to \mathcal{O}$ and $\beta_2 : \mathcal{S} \to \mathcal{S}$ where β_1 is real linear on $\mathcal{O}$ such that (β_1, β_2) preserves probability measures:

$$P_{\beta_1(T),\beta_2(\rho)} = P_{T,\rho} \qquad (19.2)$$

The automorphisms form a group QuantAut.

Now, the symmetry axiom posits that if a physical system has a group G of symmetries then there is a homomorphism $\rho : G \to \mathrm{QuantAut}$.

♣Need to put some continuity properties on ρ. For example, for continuous groups we probably want ρ to be continuous in the compact-open topology. ♣

6. *Time evolution.* If the physical system has a well-defined notion of time, then evolution of the system in time is governed by a strongly continuous groupoid of unitary operators [25] and

$$\rho(t_2) = U(t_1, t_2)^{-1} \rho(t_1) U(t_1, t_2) \qquad (19.3)$$

eq:TimeEvolve

7. *Collapse of the wavefunction* If a measurement of a physical observable corresponding to a self-adjoint operator T is made on a state ρ then the state changes discontinuously according to the result of the measurement. If T has pure point spectrum $\{\lambda_n\}_n$ with eigenspaces V_n then when T is measured the state changes discontinuously to

$$\rho \to \hat{\rho} = \sum_n P_{V_n} \rho P_{V_n} \qquad (19.4)$$

eq:StateCollap

When T has a continuous spectrum an analogous, but more complicated, formula holds which takes into account the resolution of the measuring apparatus measuring a continuous spectrum.

[25] By this we simply mean that $(t_1, t_2) \to U(t_1, t_2)$ is continuous in the strong operator topology and $U(t_1, t_2)U(t_2, t_3) = U(t_1, t_3)$.

8. *Combination of systems.* If two physical systems, represented by Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are "combined" (for example, they are allowed to interact or are otherwise considered to be part of one system) then the combined system is described by the $\mathbb{Z}_2$-graded tensor product $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$.

♣This doesn't have much content. Need to say something about observables. ♣

Remarks

1. First, let us stress that the terms "physical system," "physical observables," and "symmetries of a physical system," are not *a priori* defined mathematical terms, although we do hope that they are meaningful terms to the reader. The point of the first few axioms is indeed to identify concrete mathematical objects to associate with these physical notions.

2. *"Space of states"*: The $\mathbb{Z}_2$-grading is required since we want to incorporate fermions. See below for $\mathbb{Z}_2$-graded linear algebra. Some authors have toyed with the idea of using real or quaternionic Hilbert spaces but despite much effort nobody has given a compelling case for using either. I have not seen any strong arguments advanced for why the Hilbert space should be separable. But the theory of self-adjoint operators is likely to be a good deal more complicated for non-separable Hilbert spaces.

3. *Pure states vs. mixed states.* The "physical states" referred to in axiom 3 are often called *density matrices* in the physics literature. By the Schmidt decomposition of positive traceclass operators (18.94) we can write

$$\rho = \sum \rho_n |n\rangle\langle n| \tag{19.5}$$

where $\rho_n \geq 0$ with $\sum \rho_n = 1$ and $|n\rangle$ is an ON basis for $\mathcal{H}$. Note that the set $\mathcal{S}$ is a convex set: If $\rho_1, \rho_2 \in \mathcal{S}$ then so is $t\rho_1 + (1-t)\rho_2$ for $t \in [0,1]$. For any convex set there is a notion of the *extremal points*. These are the points which are not of the form $t\rho_1 + (1-t)\rho_2$ with $0 < t < 1$ for any value of $\rho_1 \neq \rho_2$. In the convex set of physical states the extremal points correspond to projection operators onto one-dimensional subspaces

$$\rho_\ell = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \qquad\qquad \psi \in \ell \tag{19.6}$$

The space of these extremal points are called the *pure states*. States which are not pure states are called *mixed states*. The pure states are equivalently the one-dimensional projection operators and hence are in 1-1 correspondence with the *space of lines in* $\mathcal{H}$. The space of lines in $\mathcal{H}$ is known as the *projective Hilbert space* $\mathbb{P}\mathcal{H}$. Any nonzero vector $\psi \in \mathcal{H}$ determines a line $\ell = \{z\psi | z \in \mathbb{C}\}$ so $\mathbb{P}\mathcal{H}$ is often thought of as vectors up to scale, and we can identify $\mathbb{P}\mathcal{H} = (\mathcal{H} - \{0\})/\mathbb{C}^*$. Thus, calling $\mathcal{H}$ a "space of states" is a misnomer for two reasons. First, vectors in $\mathcal{H}$ can only be used to define pure states rather than general states. Second, different vectors, namely those in the same line define the same state, so a pure state is an equivalence class of vectors in $\mathcal{H}$.

4. *Born-von Neumann formula.* Equation (19.1) goes back to the Born interpretation of the absolute square of the wavefunction as a probability density. Perhaps we should call it the Born-von Neumann formula. To recover Born's interpretation, if $\rho = |\psi\rangle\langle\psi|$ is a pure state defined by a normalized wavefunction $\psi(x) \in L^2(\mathbb{R})$ of a quantum particle on $\mathbb{R}$ and $T = q$ is the position operator then the probability of finding the particle in a measurable subset $E \in \mathcal{B}(\mathbb{R})$ is

$$P_{q,\rho}(E) = \mathrm{Tr} P_q(E)\rho = \int_E |\psi(x)|^2 dx \tag{19.7}$$

5. *Heisenberg Uncertainty Principle.* Based on its physical and historical importance one might have thought that the Heisenberg uncertainty principle would be a fundamental axiom, but in fact it is a consequence of the above. To be more precise, for a bounded self-adjoint operator $T \in \mathcal{O}$ the average value of T in state ρ is

$$\langle T\rangle_\rho := \mathrm{Tr}(T\rho) \tag{19.8}$$

We then define the *variance* or mean deviation $\sigma_{T,\rho}$ by

$$\sigma^2_{T,\rho} := \mathrm{Tr}(T - \langle T\rangle)^2\rho \tag{19.9}$$

Then if T_1 and T_2 are bounded self-adjoint operators we have, for any real number λ and phase $e^{i\theta}$,

$$0 \leq \mathrm{Tr}\left((T_1 + e^{i\theta}\lambda T_2)^\dagger (T_1 + e^{i\theta}\lambda T_2)\rho\right) \tag{19.10}$$

(provided $T_1 + e^{i\theta}\lambda T_2$ has a dense domain) and since the discriminant of the quadratic polynomial in λ is nonpositive we must have

$$\mathrm{Tr}(T_2^2\rho)\mathrm{Tr}(T_1^2\rho) \geq \frac{1}{4}\left(\langle (e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)\rangle_\rho\right)^2 \tag{19.11}$$

Note that $(e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)$ is (at least formally) self-adjoint and hence the quantity on the RHS is nonnegative. We can replace $T \to T - \langle T\rangle$ in the above and we deduce the general Heisenberg uncertainty relation: For all $e^{i\theta}$ we have the inequality:

$$\sigma^2_{T_1,\rho}\sigma^2_{T_2,\rho} \geq \frac{1}{4}\left(\langle (e^{i\theta}T_1T_2 + e^{-i\theta}T_2T_1)\rangle_\rho - 2\cos\theta\langle T_1\rangle_\rho\langle T_2\rangle_\rho\right)^2 \tag{19.12}$$

If we specialize to $\theta = \pi/2$ we get the Heisenberg uncertainty relation as usually stated:

$$\sigma^2_{T_1,\rho}\sigma^2_{T_2,\rho} \geq \frac{1}{4}\left(\langle i[T_1, T_2]\rangle_\rho\right)^2 \tag{19.13}$$

Actually, this does not quite accurately reflect the real uncertainty in *successive* measurements of noncommutative observables because the first measurement alters the state. For a recent discussion see [26]

6. The data of the first four axioms are completely general and are not specific to any physical system. The next two axioms rely on properties specific to a physical system.

[26] J. Distler and S. Paban, "On Uncertainties of Successive Measurements," arXiv:1211.4169.

7. *Symmetries* The meaning of β_1 being linear on $\mathcal{O}$ is that if $T_1, T_2 \in \mathcal{O}$ and $D(T_1) \cap D(T_2)$ is a dense domain such that $\alpha_1 T_1 + \alpha_2 T_2$, with α_1, α_2 real has a unique self-adjoint extension then $\beta_1(\alpha_1 T_1 + \alpha_2 T_2) = \alpha_1 \beta_1(T_1) + \alpha_2 \beta_1(T_2)$. A consequence of the symmetry axiom is that β_2 is affine linear on states:

$$\beta_2(t\rho_1 + (1-t)\rho_2) = t\beta_2(\rho_1) + (1-t)\beta_2(\rho_2) \tag{19.14}$$

eq:Aff-Lin

The argument for this is that (β_1, β_2) must preserve expectation values $\langle T \rangle_\rho$. However, positive self-adjoint operators of trace one are themselves observables and we have $\langle \rho_1 \rangle_{\rho_2} = \langle \rho_2 \rangle_{\rho_1}$, so the restriction of β_1 to $\mathcal{S}$ must agree with β_2. Now apply linearity of β_1 on the self-adjoint operators. From (19.14) it follows [27] that β must take extreme states to extreme states, and hence β_2 induces a map $\beta : \mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$. Now, define the overlap of two lines $\ell_1, \ell_2 \in \mathbb{P}\mathcal{H}$ by

$$\mathcal{P}(\ell_1, \ell_2) = \frac{|\langle \psi_1 | \psi_2 \rangle|^2}{\parallel \psi_1 \parallel^2 \parallel \psi_2 \parallel^2} \tag{19.15}$$

where $\psi_1 \in \ell_1$ and $\psi_2 \in \ell_2$ are nonzero vectors. Preservation of probabilities implies that

$$\mathcal{P}(\beta(\ell_1), \beta(\ell_2)) = \mathcal{P}(\ell_1, \ell_2) \tag{19.16}$$

eq:PreservePro

So we can think of the group QuantAut as the group of maps $\mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$ which satisfy (19.16). Note that if $T : \mathcal{H} \to \mathcal{H}$ is linear or anti-linear and preserves norms $\parallel T\psi \parallel = \parallel \psi \parallel$ then T descends to a map $\bar{T} := \mathbb{P}\mathcal{H} \to \mathbb{P}\mathcal{H}$ satisfying (19.16). Now *Wigner's theorem*, proved in Chapter *** below asserts that *every* map $\beta \in$ QuantAut is of this form. More precisely, there is an exact sequence:

$$1 \to U(1) \to \mathrm{Aut}_{\mathbb{R}}(\mathcal{H}) \to \mathrm{QuantAut} \to 1 \tag{19.17}$$

eq:WignerTheor

where $\mathrm{Aut}_{\mathbb{R}}(\mathcal{H})$ is the group of unitary *and* anti-unitary transformations of $\mathcal{H}$ and $U(1)$ acts on $\mathcal{H}$ by scalar multiplication. See Chapter *** below for more detail.

♣Need to comment on "Virasoro symmetry" in 2d CFT and spontaneously broken symmetries in QFT. ♣

8. *Schrödinger equation.* Suppose that the system has time-translation invariance. Then we can use the symmetry axiom and the dynamics axiom to conclude that $U(t_1, t_2) = U(t_2 - t_1)$ is a strongly-continuous one parameter group of unitary operators. Then by Stone's theorem there is a self-adjoint generator known as the *Hamiltonian* H and usually normalized by

$$U(t) = \exp\left[-\frac{i}{\hbar} tH\right] \tag{19.18}$$

where $\hbar$ is Planck's constant, so H has units of energy. Then if $\rho(t)$ is a pure state and it can be described by a differentiable family of vectors $|\psi(t)\rangle$ in $\mathcal{H}$ then these vectors should satisfy the Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle \tag{19.19}$$

More generally, if $t \to H(t)$ is a family of self-adjoint operators then $U(t_1, t_2) = P\exp - \frac{i}{\hbar} \int_{t_1}^{t_2} H(t')dt'$.

♣Need to say in what sense it is continuous ♣

[27] For some interesting discussion of related considerations see B. Simon, "Quantum Dynamics: From Automorphism to Hamiltonian."

9. *Collapse of the wavefunction.* We have given the result where a measurement is performed but the value of the measurement is not recorded. We can also speak of conditional probabilities. If the measured value of T is λ_i then

$$\rho \to \rho' = \frac{P_{V_i} \rho P_{V_i}}{\mathrm{Tr} P_{V_i} \rho} \tag{19.20}$$

eq:MeausreE1

This rule should be thought of in terms of conditional probability. When we know something happened we should renormalize our probability measures to account for this. This is related to "Bayesian inference" and "Bayesian updating." The denominator in (19.20) is required so that ρ' has trace one. (Note that the denominator is nonzero because by assumption the value λ_i was measured.)

Of course this axiom is quite notorious and generates a lot of controversy. Briefly, there is a school of thought which denies that there is any such discontinuous change in the physical state. Rather, one should consider the quantum mechanics of the full system of measuring apparatus together with the measured system. All time evolution is smooth unitary evolution (19.3). If the measuring apparatus is macroscopic then the semiclassical limit of quantum mechanics leads to classical probability laws governing the measuring apparatus and one can derive the *appearance* of the collapse of the wavefunction. This viewpoint relies on *phase decoherence* of nearly degenerate states in a large Hilbert space of states describing a fixed value of a classical observable. According to this viewpoint Axiom 7 should *not* be an axiom. Rather, it is an effective description of "what really happens." For references see papers of W. Zurek. [28]

10. *Simultaneous measurement.* If T_1 and T_2 commute then they can be "simultaneously measured." What this means is that if we measure T_1 then the change (19.4) of the physical state does not alter the probability of the subsequent measurement of T_2:

$$\begin{aligned} P_{T_2,\hat{\rho}}(E) &= \mathrm{Tr} P_{T_2}(E) \sum_n P_{V_n} \rho P_{V_n} \\ &= \sum_n \mathrm{Tr} P_{T_2}(E) P_{V_n} \rho P_{V_n} \\ &= \sum_n \mathrm{Tr} P_{V_n} P_{T_2}(E) \rho \\ &= \mathrm{Tr}(\sum_n P_{V_n}) P_{T_2}(E) \rho \\ &= \mathrm{Tr} P_{T_2}(E) \rho = P_{T_2,\rho}(E) \end{aligned} \tag{19.21}$$

Although sometimes stated as an axiom this is really a consequence of what was said above. (And we certainly don't want any notion of simultaneity to be any part of the fundamental axioms of quantum mechanics!)

[28] Short expositions are in T. Banks, "Locality and the classical limit of quantum mechanics," arXiv:0809.3764 [quant-ph]; "The interpretation of quantum mechanics," http://blogs.discovermagazine.com/cosmicvariance/files/2011/11/banks-qmblog.pdf We also recommend S. Coleman's classic colloquium, "Quantum Mechanics: In Your Face." http://media.physics.harvard.edu/video/?id=SidneyColeman QMIYF

11. The fact that we work with $\mathbb{Z}_2$-graded Hilbert spaces and take a $\mathbb{Z}_2$-graded tensor product can have important consequences. An example arises in the currently fashionable topic of Majorana fermions in condensed matter theory.

12. *Relation to classical mechanics.* There is a formulation of classical mechanics which is closely parallel to the above formulation of quantum mechanics. In order to discuss it one should focus on C^*-algebras. In quantum mechanics one can consider the C^* algebra of bounded operators $\mathcal{L}(\mathcal{H})$ or its subalgebra of compact operators $\mathcal{K}(\mathcal{H})$. The positive elements are self-adjoint, and hence observables. If $\mathcal{M}$ is a phase space, i.e. a symplectic manifold, then the analogous C^* algebra is $C_0(\mathcal{M})$, the commutative C^* algebra of complex-valued continuous functions $f : \mathcal{M} \to \mathbb{C}$ vanishing at infinity[29] with $\parallel f \parallel = \sup_{x\in\mathcal{M}}|f(x)|$. The observables are the real-valued functions and the positive functions are necessarily observables. Now, in general, one defines

    Definition A *state* on a C^* algebra $\mathfrak{A}$ is a linear map $\omega : \mathfrak{A} \to \mathbb{C}$ which is positive, i.e., $\omega(A) \geq 0$ if $A \geq 0$, and of norm 1.

    Then there are two relevant theorems:

    Theorem 1: If $\mathfrak{A} = \mathcal{K}(\mathcal{H})$ then the space of states in the sense of C^*-algebra theory is in fact the set of positive traceclass operators of trace 1, i.e., the set of density matrices of quantum mechanics.

    Theorem 2: If X is any Hausdorff space and $\mathfrak{A} = C_0(X)$ then the space of states is the space of probability measures on X.

    Therefore, in the formulation of classical mechanics one defines the observables $\mathcal{O}_{\rm class}$ to be $f \in C_0(\mathcal{M};\mathbb{R})$ and the states $\mathcal{S}_{\rm class}$ to be probability measures $d\mu$ on $\mathcal{M}$. Then, given an observable f and a state $d\mu$ we get a probability measure on $\mathbb{R}$, which, when evaluated on a Borel set $E \in \mathcal{B}(\mathbb{R})$ is

    $$P_{f,d\mu}(E) := \int_{f^{-1}(E)} d\mu \tag{19.22}$$

    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 $d\mu_{\rm Liouville} = \frac{\omega^n}{n!}$ where ω is the symplectic form and given a state $d\mu$ we can define $d\mu(x) = \rho(x)d\mu_{\rm Liouville}$. Then the classical analog of the Schrödinger equation is the Liouville equation

    $$\frac{d\rho(x;t)}{dt} = -\{H,\rho\} \tag{19.23}$$

    This is a good formalism for describing semiclassical limits and coherent states.

13. Of course, our treatment does not begin to do justice to the physics of quantum mechanics. Showing how the above axioms really lead to a description of Nature requires an entire course on quantum mechanics. We are just giving the bare bones axiomatic framework.

[29] This means that for all $\epsilon > 0$ the set $\{x \in \mathcal{M} || f(x)| \geq \epsilon\}$ is compact.

References: There is a large literature on attempts to axiomatize quantum mechanics. The first few chapters of Dirac's book is the first and most important example of such an attempt. Then in his famous 1932 book *The Mathematical Foundations of Quantum Mechanics* J. von Neumann tried to put Dirac's axioms on a solid mathematical footing, introducing major advances in mathematics (such as the theory of self-adjoint operators) along the way. For an interesting literature list and commentary on this topic see the Notes to Section VIII.11 in Reed and Simon. We are generally following here the very nice treatment of L. Takhtadjan, *Quantum Mechanics for Mathematicians*, GTM 95 which in turn is motivated by the approach of G. Mackey, *The Mathematical Foundations of Quantum Mechanics*, although we differ in some important details.

Exercise
Show that (19.1) is in fact a probability measure on the real line.

Exercise *States of the two-state system*

Consider the finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$. Show that the physical states can be parametrized as:

$$\rho = \frac{1}{2}(1 + \vec{x} \cdot \vec{\sigma}) \tag{19.24}$$

where $\vec{x} \in \mathbb{R}^3$ and $\vec{x}^2 \leq 1$. Note that this is a convex set and the set of extremal points is the sphere $S^2 \cong \mathbb{C}P^1 = \mathbb{P}\mathcal{H}$.

Exercise *Von Neumann entropy*

The von Neumann entropy of a state ρ is defined to be $S(\rho) := -\mathrm{Tr}(\rho \log \rho)$.

Suppose that $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is a product system. For a state ρ define $\rho_B \in \mathcal{L}\mathcal{H}_B$ by $\rho_B = \mathrm{Tr}_{\mathcal{H}_A}(\rho)$ and similarly for ρ_A.

Show that if ρ is a pure state then $S(\rho_A) = S(\rho_B)$.

---

20. Canonical Forms of Antisymmetric, Symmetric, and Orthogonal matrices

anFormSymmOrth

20.1 Pairings and bilinear forms

20.1.1 Perfect pairings

Definition. Suppose M_1, M_2, M_3 are R-modules for a ring R.

1.) Then a M_3*-valued pairing* is a bilinear map

$$b : M_1 \times M_2 \to M_3 \tag{20.1}$$

2.) It is said to be *nondegenerate* if the induced maps

$$L_b : M_1 \to \mathrm{Hom}_R(M_2, M_3) \qquad m_1 \mapsto b(m_1, \cdot) \tag{20.2}$$

and

$$R_b : M_2 \to \mathrm{Hom}_R(M_1, M_3) \qquad m_2 \mapsto b(\cdot, m_2) \tag{20.3}$$

are injective.

3.) It is said to be a *perfect pairing* if these maps are injective and surjective, i.e. L_b and R_b define isomorphisms.

Examples

1. $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $b(x, y) = kxy$ is nondegenerate for $k \neq 0$ but only a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups) for $k = \pm 1$.
2. $\mathbb{Z} \times U(1) \to U(1)$ defined by $b(n, e^{i\theta}) = e^{in\theta}$ is a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups) but $b(n, e^{i\theta}) = e^{ikn\theta}$ for k an integer of absolute value > 1 is not a perfect pairing.
3. $\mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ given by $b(r, s) = rs \mathrm{mod} n$ is a perfect pairing of $\mathbb{Z}$-modules (i.e. abelian groups).

20.1.2 Vector spaces

Now we specialize to a pairing of vector spaces over a field κ with $M_3 = \kappa$. Then a pairing is called a bilinear form:

Definition. A bilinear form on a vector space is a map:

$$\langle \cdot, \cdot \rangle : V \times V \to \kappa \tag{20.4}$$

which is *linear* in both variables.

It is called a *symmetric quadratic form* if:

$$\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle \qquad \forall v_1, v_2 \in V \tag{20.5}$$

eq:symmt

and *antisymmetric quadratic form* if:

$$\langle v_1, v_2 \rangle = -\langle v_2, v_1 \rangle \qquad \forall v_1, v_2 \in V \tag{20.6}$$

eq:symmtp

and it is *alternating* if: $\langle v, v \rangle = 0$ for all $v \in V$.

Remark: Note we are using angle brackets for bilinear forms and round brackets for sesquilinear forms. Some authors have the reverse convention!

Exercise

a.) Use the universal property of the tensor product to show that the space of bilinear forms on a vector space over κ is

$$\mathrm{Bil}(V) \cong (V \otimes V)^* = V^* \otimes V^*. \tag{20.7}$$

b.) Thus, a bilinear form induces two maps $V \to V^*$ (by contracting with the first or second factor). Show that if the bilinear form is nondegenerate then this provides two isomorphisms of V with V^*.

Exercise

a.) Show that if the field κ is not of characteristic 2 then a form is alternating iff it is antisymmetric.

b.) Show that alternating and antisymmetric are not equivalent for a vector space over the field $\mathbb{F}_2$. [30]

20.1.3 Choosing a basis

If we choose a an ordered basis $\{v_i\}$ for V then a quadratic form is given by a matrix

$$Q_{ij} = \langle v_i, v_j \rangle \tag{20.8}$$

Under a change of basis

$$w_i = \sum_j S_{ji} v_j \tag{20.9}$$

the matrix changes by

$$\tilde{Q}_{ij} = \langle w_i, w_j \rangle = (S^{tr} Q S)_{ij} \tag{20.10}$$

Note that the symmetry, or anti-symmetry of Q_{ij} is thus preserved by arbitrary change of basis. (This is not true under similarity transformations $Q \to SQS^{-1}$.)

Remark: The above definitions apply to any module M over a ring R. We will use the more general notion when discussing abstract integral lattices.

20.2 Canonical forms for symmetric matrices

Theorem If $Q \in M_n(\kappa)$ is symmetric, (and κ is any field of characteristic $\neq 2$) then there is a nonsingular matrix $S \in GL(n, \kappa)$ such that $S^{tr} Q S$ is diagonal:

$$S^{tr} Q S = Diag\{\lambda_1, \ldots, \lambda_n\} \tag{20.11}$$ `eq:brdg`

[30] *Answer*: Alternating implies antisymmetric, but not vice versa.

Proof:[31]

Suppose we have a quadratic form Q. If $Q = 0$ we are done. If $Q \neq 0$ then there exists a v such that $Q(v, v) \neq 0$, because

$$2Q(v,w) = Q(v+w, v+w) - Q(v,v) - Q(w,w) \tag{20.12}$$

(Note: Here we have used symmetry and the invertibility of 2.)

Now, we proceed inductively. Suppose we can find $v_1, \ldots, v_k$ so that $Q(v_i, v_j) = \lambda_i \delta_{ij}$ with $\lambda_i \neq 0$, and define V_k to be the span of $\{v_1, \ldots, v_k\}$.

Let

$$V_k^{\perp} := \{w | Q(w,v) = 0 \qquad \forall v \in V_k\} \tag{20.13}$$

We claim $V = V_k \oplus V_k^{\perp}$. First note that if $u = \sum a_i v_i \in V_k \cap V_k^{\perp}$ then $Q(u, v_i) = 0$ implies $a_i = 0$ since $\lambda_i \neq 0$. Moreover, for any vector $u \in V$

$$u^{\perp} = u - \sum_i Q(u, v_i) \lambda_i^{-1} v_i \in V_k^{\perp} \tag{20.14}$$

and therefore u is in $V_k + V_k^{\perp}$.

Now consider the restriction of Q to $V_k^{\perp}$. If this restriction is 0 we are done. If the restriction is not zero, then there exists a $v_{k+1} \in V_k^{\perp}$ with $Q(v_{k+1}, v_{k+1}) = \lambda_{k+1} \neq 0$, and we proceed as before. On a finite dimensional space the procedure must terminate $\spadesuit$.

Remark: The above theorem definitely fails for a field of characteristic 2. For example the symmetric quadratic form

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{20.15}$$

on $\mathbb{F}_2 \oplus \mathbb{F}_2$ cannot be diagonalized.

Returning to fields of characteristic $\neq 2$, the diagonal form above still leaves the possibility to make further transformations by which we might simplify the quadratic form. Now by using a further diagonal matrix D we can bring it to the form:

$$(SD)^{tr} Q (SD) = Diag\{\mu_1^2 \lambda_1, \ldots, \mu_n^2 \lambda_n\} \tag{20.16}$$ eq:diagtrmn

Now, at this point, the choice of field κ becomes very important.

Suppose the field $\kappa = \mathbb{C}$. Then note that by a further transformation of the form (20.16) we can always bring A to the form

$$Q = \begin{pmatrix} 1_r & 0 \\ 0 & 0 \end{pmatrix} \tag{20.17}$$

However, over $k = \mathbb{R}$ there are further invariants:

[31]Taken from Jacobsen, Theorem 6.5.

Theorem: [Sylvester's law]. For any real symmetric matrix A there is an invertible real matrix S so that

$$SQS^{\mathrm{tr}} = Diag\{1^p, (-1)^q, 0^n\} \qquad S \in GL(n, \mathbb{R}) \tag{20.18}$$

Proof: Now, λ_i, μ_i in (20.16) must both be real. Using real μ_i we can set $\mu_i^2 \lambda_i = \pm 1, 0$. ♠

Remarks

1. The point of the above theorem is that, since μ_i are real one *cannot* change the sign of the eigenvalue λ_i. The rank of A is $p+q$. The **signature** is (p, q, n) (sometimes people use $p-q$). If $n = 0$ A *is nondegenerate*. If $n = q = 0$ A *is positive definite*

2. If $\kappa = \mathbb{Q}$ there are yet further invariants, since not every positive rational number is the square of a rational number, so the invariants are in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$.

Finally, we note that the transformations $A \to SAS^{-1}$ and $A \to SAS^{tr}$ do not interact very well in the following sense: Suppose you know a complex matrix A is symmetric. Does that give you any useful information about its diagonalizability or its Jordan form under $A \to SAS^{-1}$ with $S \in GL(n, \mathbb{C})$? The answer is *no!*:

Theorem An arbitrary complex matrix is similar to a complex symmetric matrix.

Idea of proof: Since there is an S with SAS^{-1} in Jordan form it suffices to show that $J_\lambda^{(k)}$ is similar to a complex symmetric matrix. Write

$$J_\lambda^{(k)} = \lambda 1 + \frac{1}{2}(N + N^{tr}) + \frac{1}{2}(N - N^{tr}) \tag{20.19}$$

One diagonalizes $(N - N^{tr})$ by a unitary matrix U such that $U(N + N^{tr})U^{-1}$ remains symmetric. ♠

Exercise

a.) Show that the complex symmetric matrix

$$\begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \tag{20.20}$$

has a zero trace and determinant and find its Jordan form.

b.) Show that there is no nonsingular matrix S such that

$$S \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} S^{tr} \tag{20.21}$$

is diagonal.

20.3 Orthogonal matrices: The real spectral theorem

Sometimes we are interested only in making transformations for S an orthogonal matrix.

♣Clarify that now you are adding the data of an inner product space together with a bilinear form! ♣

Theorem [Real Finite Dimensional Spectral Theorem]. If A is a real symmetric matrix then it can be diagonalized by an orthogonal transformation:

$$A^{tr} = A, A \in M_n(\mathbb{R}) \to \exists S \in O(n, \mathbb{R}) : \qquad SAS^{tr} = Diag\{\lambda_1, \dots, \lambda_n\} \tag{20.22}$$

Proof: The proof is similar to that for the finite dimensional spectral theorem over $\mathbb{C}$. If we work over the complex numbers, then we have at least one characteristic vector $Av = \lambda v$. Since λ is real and A hermitian, $Av^* = \lambda v^*$, so $A(v + v^*) = \lambda(v + v^*)$. Thus, A in fact has a real eigenvector. Now take the orthogonal complement to $(v + v^*)$ and use induction. ♠

Example: In mechanics we use this theorem to define moments of inertia

As an application we have the analog of the theorem that unitary matrices can be unitarily diagonalized:

Theorem Every real orthogonal matrix O can be brought to the form:

$$SOS^{tr} = Diag\{+1^r, -1^q, R(\theta_i)\} \qquad S \in O(n, \mathbb{R}) \tag{20.23}$$

by an orthogonal transformation. Here

$$R(\theta_i) = \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix} \qquad \theta_i \neq n\pi \tag{20.24}$$

Proof: Consider $T = O + O^{-1} = O + O^{tr}$ on $\mathbb{R}^n$. This is real symmetric, so by the spectral theorem (over $\mathbb{R}$) there is an orthogonal basis in which $\mathbb{R}^n = \oplus_i V_i$ where T has eigenvalue λ_i on the subspace V_i and λ_i are the distinct eigenvalues of T. They are all real. Note that for all vectors $v \in V_i$ we have

$$(O^2 - \lambda_i O + 1)v = 0 \tag{20.25}$$

eq:Quad01

so O restricted to V_i satisfies $O^2 - \lambda_i O + 1 = 0$. Therefore, if $v \in V_i$ then the vector space $W = Span\{v, Ov\}$ is preserved by O. Moreover, O preserves the decomposition $W \oplus W^\perp$. Therefore by induction we need only analyze the cases where W is 1 and 2-dimensional. If W is one dimensional then $Ov = \mu v$ and we easily see that $\mu^2 = 1$ so $\mu = \pm 1$. Suppose W is two-dimensional and O acting on W satisfies

$$O^2 - \lambda O + 1 = 0 \tag{20.26}$$

eq:Quad0

Now, by complexification we know that O is unitary and hence it is diagonalizable (over $\mathbb{C}$) and its eigenvalues must be phases $\{e^{i\theta_1}, e^{i\theta_2}\}$. On the other hand (20.26) will be true after diagonalization (over $\mathbb{C}$) so $\lambda = e^{i\theta} + e^{-i\theta} = 2\cos\theta$ for both angles $\theta = \theta_1$

and $\theta = \theta_2$. Now, on this two-dimensional space we have $O^{tr} = O^{-1}$ as 2×2 matrices. Therefore det$O$ is ± 1 and moreover

$$O = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow O^{tr} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \qquad \text{and} \qquad O^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{20.27}$$

Thus, $O^{tr} = O^{-1}$ implies $a = d$ and $b = -c$ if det$O = +1$ and it implies $a = -d$ and $b = c$ if det$O = -1$. In the first case we go back to equation (20.26) and solve for a, b to find $O = R(\pm\theta)$. In the second case we find

$$O = R(\theta) P \tag{20.28}$$

with

$$P = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{20.29}$$

Next we observe that $P^2 = 1$ and $PR(\phi)P = R(-\phi)$ so we may then write

$$O = R(\phi) P R(\phi)^{tr} \tag{20.30}$$

with $2\phi = \theta$, and hence in the second case we can transform O to P which is of the canonical type given in the theorem. ♠

20.4 Canonical forms for antisymmetric matrices

Theorem Let κ be any field. If $A \in M_n(\kappa)$ is antisymmetric there exists $S \in GL(n, \kappa)$ that brings A to the canonical form:

$$SAS^{tr} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus 0_{n-r} \tag{20.31}$$

Proof: The proof is very similar to the case of symmetric forms. Let us suppose that A is the matrix with respect to an antisymmetric quadratic form Q on a vector space V. If $Q = 0$ we are done. If $Q \neq 0$ then there must be linearly independent vectors u, v with $Q(u, v) = q \neq 0$. Define $u_1 = u$ and $v_1 = q^{-1} v$. Now Q has the required canonical form with respect to the ordered basis $\{u_1, v_1\}$.

Now we proceed by induction. Suppose we have constructed linearly independent vectors $(u_1, v_1, \cdots, u_k, v_k)$ such that $Q(u_i, v_j) = \delta_{ij}$, and $Q(u_i, u_j) = Q(v_i, v_j) = 0$. Let $V_k = Span\{u_1, v_1, \cdots, u_k, v_k\}$. Then again $V = V_k \oplus V_k^{\perp}$ where again we define

$$V_k^{\perp} := \{w | Q(w, v) = 0 \qquad \forall v \in V_k\} \tag{20.32}$$

It is easy to see that $V_k \cap V_k^{\perp}$ and if $w \in V$ is any vector then

$$w^{\perp} = w + \sum_i (Q(w, u_i) v_i - Q(w, v_i) u_i) \in V_k^{\perp} \tag{20.33}$$

so $V = V_k + V_k^{\perp}$. Restricting Q to $V_k^{\perp}$ we proceed as above. ♠

As usual, if we put a restriction on the change of basis we get a richer classification:

Theorem Every *real antisymmetric* matrix $A^{tr} = -A$ can be skew-diagonalized by $S \in O(n, \mathbb{R})$, that is, A can be brought to the form:

$$SAS^{tr} = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 & \cdots \\ -\lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_2 & \cdots \\ 0 & 0 & -\lambda_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{20.34}$$

The λ_i are called the *skew eigenvalues.* Note that, without a choice of orientation they are only defined up to sign.

Idea of proof: There are two ways to prove this. One way is to use the strategies above by using induction on the dimension. Alternatively, we can view A as an operator on $\mathbb{R}^n$ and observe that iA is an Hermitian operator on $\mathbb{C}^n$ so there is a basis of ON eigenvectors of iA. But

$$iAv = \lambda v \tag{20.35}$$

with λ real implies

$$\begin{aligned} Av &= -i\lambda v \\ Av^* &= i\lambda v^* \end{aligned} \tag{20.36}$$

So if an ON basis on $\mathbb{C}^n$ then we get an orthogonal basis on $\mathbb{R}^n$ consisting of $u_1 = v + v^*$ and $u_2 = i(v - v^*)$ and

$$\begin{aligned} Au_1 &= \lambda u_2 \\ Au_2 &= -\lambda u_1 \end{aligned} \tag{20.37}$$

Letting S be the matrix given by the columns of these vectors we have $S^{tr}AS$ is in the above block-diagonal form ♠

20.5 Automorphism Groups of Bilinear and Sesquilinear Forms

Given a bilinear form, or a sesquilinear form Q on a vector space V, the *automorphism group of the form* is the subgroup of operators $T \in GL(V)$ such that

$$Q(Tv, Tw) = Q(v, w) \tag{20.38}$$ eq:autgroup

for all $v, w \in V$.

In special cases these groups have special names:

1. *Q is a nondegenerate symmetric bilinear form on V*: Then the group of automorphisms is denoted $O(Q)$ and it is called the orthogonal group of the form. If V is a complex vector space of dimension n and we choose a basis with $Q = 1$ then we

define the matrix group $O(n,\mathbb{C})$ as the group of $n\times n$ complex invertible matrices S such that $S^{tr}S=1$. If V is a real vector space and Q has signature $(-1)^p,(+1)^q$ then we can choose a basis in which the matrix form of Q is

$$\eta := \begin{pmatrix} -1_p & 0 \\ 0 & +1_q \end{pmatrix} \tag{20.39}$$

and the resulting matrix group, denoted $O(p,q;\mathbb{R})$, is the group of invertible matrices so that

$$S^{tr}\eta S=\eta \tag{20.40}$$

2. *Q is a nondegenerate anti-symmetric bilinear form on V*: In this case the group of automorphisms is called the *symplectic group $Sp(Q)$*. If V is finite dimensional and we are working over any field κ we can choose a basis for V in which Q has matrix form

$$J=\begin{pmatrix} 0_n & 1_n \\ -1_n & 0_n \end{pmatrix} \tag{20.41}$$

and then the resulting matrix group, which is denoted $Sp(2n;\kappa)$ for field κ, is the set of invertible matrices with matrix elements in κ such that

$$S^{tr}JS=J \tag{20.42}$$

3. *Sesquilinear forms.* It also makes sense to talk about the automorphism group of a sesquilinear form on a complex inner product space. If there is an ON basis $\{e_i\}$ in which the matrix $h(e_i,e_j)$ is of the form

$$h_{ij}=\begin{pmatrix} 1_p & 0 \\ 0 & -1_q \end{pmatrix} \tag{20.43}$$

and then the resulting matrix group, which is denoted $U(p,q)$ or $U(p,q;\mathbb{C})$, is the set of invertible matrices in $GL(n,\mathbb{C})$ so that

$$U^{\dagger}hU=h \tag{20.44}$$

Remarks

1. When working over rings and not fields we might not be able to bring Q to a simple standard form like the above, nevertheless, $Aut(Q)$ remains a well-defined group.

2. We will look at these groups in much more detail in our chapter on a Survey of Matrix Groups

21. Other canonical forms: Upper triangular, polar, reduced echelon

21.1 General upper triangular decomposition

Theorem 13 Any complex matrix can be written as $A = UT$, where U is unitary and T is upper triangular.

This can be proved by successively applying reflections to the matrix A. I.e. we define

$$R_{ij}(v) = \delta_{ij} - 2\frac{v_i v_j}{v \cdot v} \tag{21.1}$$

This is a reflection in the hyperplane $v^{\perp}$ and hence an orthogonal transformation. Consider the vector A_{k1}. If it is zero there is nothing to do. If it is nonzero then this vector, together with e_1 span a 2-dimensional plane. We can reflect in a line in this plane to make A_{k1} parallel to e_1. Now consider everything in $e_1^{\perp}$. Then A_{k2} is a vector which forms a 2-dimensional plane with e_2. We can repeat the process. In this way one can choose vectors $v_1, \dots, v_n$ so that $R(v_n) \cdots R(v_1)A$ is upper triangular.

For this and similar algorithms G.H. Golub and C.F. Van Loan, *Matrix Computations.*

21.2 Gram-Schmidt procedure

c:Gram-Schmidt

In the case where A is nonsingular the above theorem can be sharpened. In this case the upper triangular decomposition is closely related to the Gram-Schmidt procedure. Recall that the Gram-Schmidt procedure is the following:

Let $\{u_i\}$ be a set of linearly independent vectors. The GS procedure assigns to this an ON set of vectors $\{v_i\}$ with the same linear span:

The procedure:

a.) Let $w_1 = u_1$, define $v_1 = w_1 / \parallel w_1 \parallel$.

b.) Let $w_2 = u_2 - (v_1, u_1)v_1$, define $v_2 = w_2 / \parallel w_2 \parallel$.

c.) Let $w_n = u_n - \sum_{k=1}^{n-1}(v_k, u_n)v_k$, define $v_n = w_n / \parallel w_n \parallel$.

Theorem 14 Any nonsingular matrix $A \in GL(n, \mathbb{C})$ can be uniquely written as

$$A = UT \tag{21.2}$$

where U is unitary and T is upper triangular, with positive real diagonal entries. Any nonsingular matrix $A \in GL(n, \mathbb{R})$ can be uniquely written as

$$A = OT \tag{21.3}$$

where O is orthogonal and T is upper triangular with positive real diagonal entries.

Proof: Note that in the Gram-Schmidt procedure the bases are related by

$$v_i = \sum T_{ji} u_j \tag{21.4}$$

eq:gmi

where T is an invertible *upper triangular* matrix. Now, let A_{ij} be any nonsingular matrix. Choose any ON basis $\tilde{v}_i$ for $\mathbb{C}^n$ and define:

$$u_j := \sum_{i=1}^{n} A_{ij}\tilde{v}_i \tag{21.5}$$ eq:gmii

This is another basis for the vector space. Then applying the GS procedure to the system $\{u_j\}$ we get an ON set of vectors v_i satisfying (21.4). Therefore,

$$v_j = \sum_{j,k} A_{kj}T_{ji}\tilde{v}_k \tag{21.6}$$ eq:gmiii

Since v_i and $\tilde{v}_i$ are two ON bases, they are related by a unitary transformation, therefore

$$U_{ki} = \sum_{j} A_{kj}T_{ji} \tag{21.7}$$ eq:gmiv

is unitary. Since T is invertible the theorem follows ♠.

Exercise *Gram-Schmidt at a glance*

Let $u_1, \dots, u_n$ be n linearly independent vectors. Show that the result of the Gram-Schmidt procedure is summarized in the single formula:

$$\begin{aligned}
v_n &= \frac{(-1)^{n-1}}{\sqrt{D_{n-1}D_n}} \det\begin{pmatrix} u_1 & \cdots & u_n \\ (u_1,u_1) & \cdots & (u_n,u_1) \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ (u_1,u_{n-1}) & \cdots & (u_n,u_{n-1}) \end{pmatrix} \\
D_n &= \det\begin{pmatrix} (u_1,u_1) & \cdots & (u_n,u_1) \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ (u_1,u_n) & \cdots & (u_n,u_n) \end{pmatrix}
\end{aligned} \tag{21.8}$$ eq:gssci

21.2.1 Orthogonal polynomials

Let $w(x)$ be a nonnegative function on $[a,b]$ such that

$$\int_a^b x^N w(x)dx < \infty \qquad N \geq 0 \tag{21.9}$$ eq:opi

We can define a Hilbert space by considering the complex-valued functions on $[a,b]$ such that

$$\int_a^b |f(x)|^2 w(x) dx < \infty \tag{21.10}$$ eq:opii

Let us call this space $L^2\big([a,b], d\mu\big)$, the L^2 functions wrt the measure $d\mu(x) = w(x)dx$. Applying the Gram-Schmidt procedure to the system of functions

$$\{1, x, x^2, \dots\} \tag{21.11}$$

leads to a system of *orthogonal polynomials* in terms of which we may expand any smooth function.

Example. Legendre polynomials. Choose $[a,b] = [-1,1]$, $w(x) = 1$. We obtain:

$$\begin{aligned} u_0 = 1 \qquad &\rightarrow \qquad \phi_0 = \frac{1}{\sqrt{2}} \\ u_1 = x \qquad &\rightarrow \qquad \phi_1 = \sqrt{\frac{3}{2}} x \\ u_2 = x^2 \qquad &\rightarrow \qquad \phi_2 = \sqrt{\frac{5}{2}} \frac{3x^2 - 1}{2} \end{aligned} \tag{21.12}$$ eq:legpolys

In general

$$\phi_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x) \tag{21.13}$$

where $P_n(x)$ are the Legendre polynomials. We will meet them (more conceptually) later.

Exercise *Systems of orthogonal poynomials*
Work out the first few for
1. Tchebyshev I: $[-1,1]$, $w(x) = (1-x^2)^{-1/2}$
2. Tchebyshev II: $[-1,1]$, $w(x) = (1-x^2)^{+1/2}$
3. Laguerre: $[0,\infty)$, $w(x) = x^k e^{-x}$
4. Hermite: $(-\infty,\infty)$, $w(x) = e^{-x^2}$
For tables and much information, see Abramowitz-Stegun.

Remarks Orthogonal polynomials have many uses:

1. Special functions, special solutions to differential equations.

2. The general theory of orthogonal polynomials has proven to be of great utility in investigations of large N matrix integrals.

3. See B. Simon, *Orthogonal Polynomials on the Unit Circle*, Parts 1,2 for much more about orthogonal polynomials.

21.3 Polar decomposition

There is an analog for matrices of *polar decompositions*, generalizing the representation of complex numbers by phase and magnitude: $z = re^{i\theta}$. Here is the matrix analog:

♣This section should include the important case of complex ips and comment on the generalization to Hilbert space with partial isometries. ♣

Theorem Any matrix $A \in M_n(\mathbb{C})$ can be written as

$$A = UP \tag{21.14}$$

or

$$A = P'U \tag{21.15}$$

where P, P' are positive semidefinite and U is unitary. Moreover, the decomposition is unique if A is nonsingular.

Proof: The proof is a straightforward application of the singular value decomposition. Recall that we can write

$$A = U\Lambda V \tag{21.16}$$

where Λ is diagonal with nonnegative entries and U and V are unitary. Therefore we can write

$$A = (U\Lambda U^{-1}) \cdot (UV) = (UV) \cdot (V^{-1}\Lambda V) \tag{21.17}$$

Now note that both $(U\Lambda U^{-1})$ and $(V^{-1}\Lambda V)$ are positive semidefinite, and if A is nonsingular, positive definite. ♠

Remarks:

1. Taking the determinant recovers the polar decomposition of the determinant: $\det A = re^{i\theta}$ with $r = \det P$ and $e^{i\theta} = \det U$.
2. Note that $A^\dagger A = P^2$ so we could define P as the positive squareroot $P = \sqrt{A^\dagger A}$. This gives another approach to proving the theorem.

The version of the theorem over the real numbers is:

Theorem. Any invertible real $n \times n$ matrix A has a unique factorization as

$$A = PO \tag{21.18}$$

where P is a positive-definite symmetric matrix and O is orthogonal.

Proof: Consider AA^{tr}. This matrix is symmetric and defines a positive definite symmetric form. Since such forms can be diagonalized we know that there is a squareroot.

Let $P := (AA^{tr})^{1/2}$. There is a unique positive definite square root. Now check that $O := P^{-1}A$ is orthogonal. ♠

A related theorem is

Theorem. Any nonsingular matrix $A \in Mat_n(\mathbb{C})$ can be decomposed as:

$$A = SO \tag{21.19}$$

where S is complex symmetric and O is complex orthogonal.

Proof: Gantmacher, p.7.

Finally, we consider the generalization to operators on Hilbert space. Here there is an important new phenomenon. We can see it by considering the shift operator S on ℓ^2:

$$S : (x_1, x_2, x_3, \dots) \to (0, x_1, x_2, x_3, \dots) \tag{21.20}$$

Recall that $S^\dagger$ is the shift to the left, so $S^\dagger S = 1$, but $SS^\dagger$ is *not* one, rather it is $1 - |0\rangle\langle 0|$, in harmonic oscillator language.

Definition A *partial isometry* $V : \mathcal{H} \to \mathcal{H}$ is an operator so that

$$VV^\dagger V = V \tag{21.21}$$

The shift operator above is a good example of a partial isometry.

In general, note that $V^\dagger V = P_i$ and $VV^\dagger = P_f$ are both projection operators. Now we claim that

$$1 - V^\dagger V \tag{21.22}$$

is the orthogonal projection to $\ker(V)$. It is clear that if $\psi \in \ker(V)$ then $(1 - V^\dagger V)\psi = \psi$ and conversely $V(1 - V^\dagger V)\psi = 0$. Therefore, $V^\dagger V$ is the orthogonal projector to $(\ker V)^\perp$. Similarly, $VV^\dagger$ is the orthogonal projector to $\mathrm{im}(V)$.

$(\ker V)^\perp$ with orthogonal projector $V^\dagger V$ is called the *initial subspace*

$\mathrm{im}(V)$ with orthgonal projector $VV^\dagger$ is called the *final subspace.*

Note that V is an isometry when restricted to $V : (\ker V)^\perp \to \mathrm{im}(V)$, hence the name "partial isometry."

Theorem If T is a bounded operator on Hilbert space there is a partial isometry V so that

$$T = V\sqrt{T^\dagger T} = V|T| \tag{21.23}$$

and V is uniquely determined by $\ker V = \ker T$.

For the proof, see Reed-Simon Theorem VI.10 and for the unbounded operator version Theorem VIII.32. Note that for compact operators it follows from the singular value decomposition, just as in the finite dimensional case.

Remark: In string field theory and noncommutative field theory partial isometries play an important role. In SFT they are used to construct solutions to the string field equations. In noncommutative field theory they are used to construct "noncommutative solitons."

21.4 Reduced Echelon form

Theorem 17. Any matrix $A \in GL(n, \mathbb{C})$ can be factorized as

$$A = N\Pi B \tag{21.24}$$

where N is upper-triangular with $1's$ on the diagonal, Π is a permutation matrix, and B is upper-triangular.

Proof: See Carter, Segal, and MacDonald, *Lectures on Lie Groups and Lie Algebras*, p. 65.

Remarks

1. When we work over nonalgebraically closed fields we sometimes can only put matrices into *rational canonical form.* See Herstein, sec. 6.7 for this.

22. Families of Matrices

In many problems in mathematics and physics one considers continuous, differentiable, or holomorphic families of linear operators. When studying such families one is led to interesting geometrical constructions. In this section we illustrate a few of the phenomena which arise when considering linear algebra in families.

22.1 Families of projection operators: The theory of vector bundles

Let us consider two simple examples of families of projection operators.

Let $\theta \sim \theta + 2\pi$ be a coordinate on the circle. Fix $V = \mathbb{R}^2$, and consider the operator

$$\Gamma(\theta) = \cos\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \sin\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{22.1}$$

Note that $\Gamma^2 = 1$ and accordingly we can define two projection operators

$$P_{\pm}(\theta) = \frac{1}{2}(1 \pm \Gamma(\theta)) \tag{22.2}$$

Let us consider the eigenspaces as a function of θ. For each θ the image of $P_+(\theta)$ is a real line in $\mathbb{R}^2$. So, let us consider the set

$$\mathcal{L}_+ := \{(e^{i\theta}, v) | P_+(\theta)v = v\} \subset S^1 \times \mathbb{R}^2 \tag{22.3}$$

At fixed θ we can certainly choose a basis vector, i.e. an eigenvector, in the line given by the image of $P_+(\theta)$. What happens if we try to make a *continuous* choice of such a basis vector as a function of θ? The most natural choice would be the family of eigenvectors:

$$\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \tag{22.4}$$

eq:Mobius-1

Now, θ is identified with $\theta + 2\pi$, and the projection operator only depends on θ modulo 2π. However, (22.4) is *not globally well-defined*! If we shift $\theta \to \theta + 2\pi$ then the eigenvector changes by a minus sign.

But we stress again that even though there is no globally well-defined continuous choice of eigenvector the real line given by the image of $P_+(\theta)$ is well-defined. For example, we can check:

$$\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \mathbb{R} = - \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} \mathbb{R} \subset \mathbb{R}^2 \tag{22.5}$$

eq:moblinone

The family of real lines over the circle define what is called a *real line bundle*. Another example of a real line bundle is $S^1 \times \mathbb{R}$ which is, topologically, the cylinder. However, our family is clearly different from the cylinder. One can prove that it is impossible to find a *continuous* choice of basis for all values of θ. Indeed, one can picture this real line bundle as the Mobius strip, which makes its topological nontriviality intuitively obvious.

Example 2 In close analogy to the previous example consider

$$\Gamma(\hat{x}) = \hat{x} \cdot \vec{\sigma} \tag{22.6}$$

for $\hat{x} \in S^2$, the unit sphere $\hat{x}^2 = 1$. Once again $(\Gamma(\hat{x}))^2 = 1$. Consider the projection operators

$$P_\pm(\hat{x}) = \frac{1}{2}(1 \pm \hat{x} \cdot \vec{\sigma}) \tag{22.7}$$

On $S^2 \times \mathbb{C}^2$ the eigenspaces of $P_\pm$ define a complex line for each point $\hat{x} \in S^2$. If we let $\mathcal{L}_+$ denote the total space of the line bundle, that is

$$\mathcal{L}_+ = \{(\hat{x}, v) | \hat{x} \cdot \vec{\sigma} v = +v\} \subset S^2 \times \mathbb{C}^2 \tag{22.8}$$

Then you can convince yourself that this is NOT the trivial complex line bundle $\mathcal{L}_+ \neq S^2 \times \mathbb{C}$. Using standard polar coordinates for the sphere, away from the south pole we can take the line to be spanned by

$$e_+ = \begin{pmatrix} \cos\frac{1}{2}\theta \\ e^{i\phi}\sin\frac{1}{2}\theta \end{pmatrix} \tag{22.9}$$

while away from the north pole the eigenline is spanned by

$$e_- = \begin{pmatrix} e^{-i\phi}\cos\frac{1}{2}\theta \\ \sin\frac{1}{2}\theta \end{pmatrix} \tag{22.10}$$

But note that, just as in our previous example, there is no continuous choice of nonzero vector spanning the eigenline for all points on the sphere.

What we are describing here is a nontrivial *line bundle* with transition function $e_+ = e^{i\phi}e_-$ on the sphere minus two points. This particular line bundle is called the Hopf line bundle, and of great importance in mathematical physics.

A generalization of this construction using the quaternions produces a nontrivial rank two complex vector bundle over S^4 known as the instanton bundle.

These two examples have a magnificent generalization to the theory of vector bundles. We will just summarize some facts. A full explanation would take a different Course.

Definition: A (complex or real) *vector bundle* E over a topological space X is the space of points

$$E := \{(x,v) : P(x)v = v\} \subset X \times \mathcal{H} \tag{22.11}$$

where $P(x)$ is a continuous family of orthogonal finite rank projection operators in a (complex or real) separable infinite-dimensional Hilbert space $\mathcal{H}$.

This is not the standard definition of a vector bundle, but it is equivalent to the usual definition. Note that there is an obvious map

$$\pi : E \to X \tag{22.12}$$

given by $\pi(x,v) = x$. The fibers of this map are vector spaces of dimension n:

$$E_x := \pi^{-1}(x) = \{(x,v) | v \in \mathrm{im} P(x)\} \tag{22.13}$$

carries a natural structure of a vector space:

$$\alpha(x,v) + \beta(x,w) := (x, \alpha v + \beta w) \tag{22.14}$$

so we are just doing linear algebra in families. We define a *section* of E to be a continuous map $s : X \to E$ such that $\pi(s(x)) = x$. It is also possible to talk about tensors on X with values in E. In particular $\Omega^1(X; E)$ denotes the sections of the space of 1-forms on X with values in E.

Note that, in our definition, a vector bundle is the same thing as a continuous family of projection operators on Hilbert space. So a vector bundle is the same thing as a *continuous* map

$$P : X \to Gr_n(\mathcal{H}) \tag{22.15}$$

where $Gr_n(\mathcal{H})$ is the Grassmannian of rank n projection operators in the norm topology.

Definition: Two vector bundles E_1, E_2 of rank n are *isomorphic* if there is a homotopy of the corresponding projection operators P_1, P_2. Therefore, the isomorphism classes of vector bundles is the same as the set of homotopy classes $[X, Gr_n(\mathcal{H})]$.

This viewpoint is also useful for defining connections. If $\psi : X \to \mathcal{H}$ is a continuous map into Hilbert space then

$$s(x) := (x, P(x)\psi(x) \tag{22.16}$$

is a section of E. Every section of E can be represented in this way. Then a *projected connection* on E is the map $\nabla^{\text{proj}} : \Gamma(E) \to \Omega^1(X;E)$ given by $P \circ d \circ P$ where d is the exterior differential and we have written $\circ$ to emphasize that we are considering the composition of three operators. With local coordinates x^μ we can write:

$$\nabla^{\text{proj}} : (x, s(x)) \to dx^\mu (x, P(x) \frac{\partial}{\partial x^\mu}(P(x)s(x))) \tag{22.17}$$

For those who know about curvature, the curvature of this connection is easily shown to be:

$$F = PdPdPP \tag{22.18}$$

and representatives of the characteristic classes $\mathrm{ch}_n(E)$ (thought of as DeRham cohomology classes) are then defined by the differential forms

$$\omega_n(E) = \frac{1}{n!(2\pi i)^n} \mathrm{Tr}(PdPdP)^n \tag{22.19}$$

Remarks

1. In physics the projected connection is often referred to as the *Berry connection*, and it is related to the quantum adiabatic theorem. The formula for the curvature is important in, for example, applications to condensed matter physics. In a typical application there is a family of Hamiltonians with a gap in the spectrum and P is the projector to eigenspaces below that gap. For example in the quantum Hall effect with P the projector to the lowest Landau level the first Chern class is given by

$$\omega_1 = \frac{1}{2\pi i} \mathrm{Tr} PdPdP \tag{22.20}$$

   which turns out to be related to the Kubo formula, as first noted by TKNN.

2. A beautiful theorem of Narasimhan-Ramanan shows that *any* connection on a vector bundle can be regarded as the pull-back of a projected connection. See [32]

3.

Exercise *The Bott's projector and Dirac's monopole*
In a subsection below we will make use of the *Bott projector*:

$$P(z, \bar{z}) = \frac{1}{1 + |z|^2} \begin{pmatrix} 1 & \bar{z} \\ z & |z|^2 \end{pmatrix} \tag{22.21}$$

[32] D. Quillen, "Superconnection character forms and the Cayley transform," Topology, **27**, (1988) 211; J. Dupont, R. Hain, and S. Zucker, "Regulators and characteristic classes of flat bundles," arXiv:alg-geom/9202023.

a.) Check that this is a projection operator. [33]

b.) Show that the projector has a well-defined limit for $z \to \infty$.

c.) Show that under stereographic projection $S^2 \to \mathbb{R}^2 \cong \mathbb{C}$

$$z = \frac{x^1 + i x^2}{1 + x^3} \qquad \frac{1 + x^3}{2} = \frac{1}{1 + |z|^2} \tag{22.22}$$

we have

$$P_+(\hat{x}) = P(z, \bar{z}) \tag{22.23}$$

$$P_-(\hat{x}) = Q(z, \bar{z}) = \frac{1}{1 + |z|^2} \begin{pmatrix} |z|^2 & -\bar{z} \\ -z & 1 \end{pmatrix} \tag{22.24}$$

d.) Regarding the projector $P_+(\hat{x})$ as a projection operator on $\mathbb{R}^3 - \{0\}$, show that the projected connection is the same as the famous Dirac magnetic monopole connection by computing

♣Need to check these two equations ♣

$$\nabla^{\text{proj}} e_+ = \frac{1}{2}(1 - \cos\theta) d\phi e_+ \tag{22.25}$$

$$\nabla^{\text{proj}} e_- = -\frac{1}{2}(1 + \cos\theta) d\phi e_- \tag{22.26}$$

e.) Show that for the Bott projector [34]

$$\text{Tr}(PdPdP) = -\text{Tr}(QdQdQ) = -\frac{dz d\bar{z}}{(1 + |z|^2)^2} \tag{22.27}$$

and show that $\int_{\mathbb{C}} \text{ch}_1(\mathcal{L}_\pm) = \pm 1$.

Exercise

Show that any vector bundle E has a complementary bundle $E^\perp$ so that $E \oplus E^\perp = X \times \mathbb{C}^N$, where $\oplus$ is a family version of the direct sum of vector spaces.

22.2 Codimension of the space of coinciding eigenvalues

♣This sub-section is a little out of order because it uses group actions and homogeneous spaces which are only covered in Chapter 3. But it fits very naturally in this Section. ♣

We now consider more general families of linear operators $T(s) : V \to V$ where V is finite-dimensional and s is a set of parameters. Mathematically, we have a space $\mathcal{S}$ and a map $T : \mathcal{S} \to \text{End}(V)$ and we can put various conditions on that map: Continuous, differentiable, analytic,... Physically, $\mathcal{S}$ is often a set of "control parameters," for example coupling constants, masses, or other "external" parameters which can be varied.

We no longer assume the $T(s)$ are projection operators.

[33] This is obvious if you note that the second column is just $\bar{z}$ times the first column. But it is also good to do the matrix multiplication.

[34] *Hint*: In order to avoid a lot of algebra write $P = n\tilde{P}$ with $n = (1 + r^2)^{-1}$, $r^2 = |z|^2$ so that dn is proportional to dr^2.

22.2.1 Families of complex matrices: Codimension of coinciding characteristic values

A very important question that often arises is: What is the subset of points in $\mathcal{S}$ where T changes in some important way. Here is a sharp version of that question:

> What is the set of points $\mathcal{S}_{\text{sing}} \subset \mathcal{S}$ where characteristic values of $T(s)$ coincide?

In equations:

$$\mathcal{S}_{\text{sing}} = \{s \in \mathcal{S} | p_{T(s)}(x) \quad \text{has multiple roots}\} \tag{22.28}$$

where $p_T(x) = \det(x1 - T)$ is the characteristic polynomial.

We can only give useful general rules for *generic* families. We first argue that it suffices to look at the space $\text{End}(V) \cong M_n(\mathbb{C})$ itself. Let $\mathcal{D} \subset \text{End}(V)$ be the sublocus where the characteristic polynomial has multiple eigenvalues:

$$\mathcal{D} := \{T | p_T(x) \quad \text{has multiple roots}\} \tag{22.29}$$

If $s \to T(s)$ is generically $1-1$ and $\mathcal{F} \subset \text{End}(V)$ is the image of the family then

$$\mathcal{S}_{\text{sing}} = T^{-1}(\mathcal{D} \cap \mathcal{F}) \tag{22.30}$$

Now, the codimension of $\mathcal{S}_{\text{sing}}$ in $\mathcal{S}$ is the same as:

$$\text{cod}(\mathcal{D} \cap \mathcal{F}) - \text{cod}(\mathcal{F}) \tag{22.31}$$

See Figure 9.

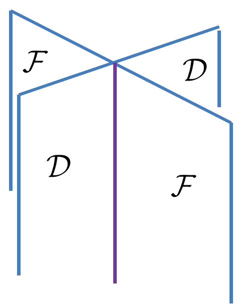

Figure 9: For generic families if there are d transverse directions to $\mathcal{D}$ in $M_n(\mathbb{C})$ and there are f transverse dimensions to $\mathcal{F}$ in $M_n(\mathbb{C})$ then there will be $d + f$ transverse dimensions to $\mathcal{D} \cap \mathcal{F}$ in $M_n(\mathbb{C})$ and d transverse dimensions to $\mathcal{S}_{\text{sing}}$ in $\mathcal{S}$. fig:CODIMENSIO

On the other hand, for generic subspaces, the *codimensions* add for intersections:

$$\text{cod}(\mathcal{D} \cap \mathcal{F}) = \text{cod}(\mathcal{D}) + \text{cod}(\mathcal{F}) \tag{22.32}$$

Therefore, if $s \to T(s)$ is a generic 1-1 family then the codimension of the set where characteristic values coincide in $\mathcal{S}$ is the same as the codimension of $\mathcal{D}$ in $M_n(\mathbb{C})$.

In general $\mathcal{D} \subset \mathrm{End}(V)$ can be exhibited as an algebraic variety. This follows from Exercises **** and **** at the end of chapter 3 where we show that the subspace is defined by the resultant of two polynomials in x:

$$\mathrm{Res}(p_T(x), p'_T(x)) = 0 \tag{22.33}$$

The resultant is a polynomial in the coefficients of $p_T(x)$ which in turn can be expressed in terms of polynomials in the matrix elements of T with respect to an ordered basis.

Example: If $n = 2$ then

$$\begin{aligned} \det(x1 - T) &= x^2 - \mathrm{Tr}(T)x + \det(T) \\ &= x^2 + a_1 x + a_0 \\ a_1 &= -\mathrm{Tr}(T) \\ a_0 &= \frac{1}{2}(\mathrm{Tr}(T))^2 - \mathrm{Tr}(T^2) \end{aligned} \tag{22.34}$$

The subspace in the space of all matrices where two characteristic roots coincide is clearly

$$a_1^2 - 4a_0 = 4\mathrm{Tr}(T^2) = (\mathrm{Tr}(T))^2 = 0 \tag{22.35}$$

which is an algebraic equation on the matrix elements.

The complex codimension of the solutions to one algebraic equation in $M_n(\mathbb{C})$ is one, and therefore we have the general rule:

The general element of a generic family of complex matrices will be diagonalizable, and the sublocus where at least two characteristic roots coincide will be real codimension two.

**** FIGURE OF $\mathcal{D}$ WITH TRANSVERSE SPACE IDENTIFIED WITH $\mathbb{C}$ ****

Note that it follows that, in a generic family the generic operator $T(s)$ will be diagonalizable, and will only fail to be diagonalizable on a complex codimension one subvariety of $\mathcal{S}$.

22.2.2 Orbits

It is interesting to view the above rule in a different way by counting dimensions of orbits. Using techniques discussed in chapters below the space $DIAG^*$ of diagonalizable matrices with distinct eigenvalues is fibered over

$$(\mathbb{C}^n - \Delta)/S_n \tag{22.36}$$

where Δ is the subspace where any two entries coincide. The fibration is the map of T to the unordered set of eigenvalues. The fiber is the set of matrices with a given set of eigenvalues and this is a homogeneous space, hence the fiber is

$$GL(n)/GL(1)^n \tag{22.37}$$

We can therefore compute the dimension of $DIAG^*$. The base is n-dimensional and the fiber is $n^2 - n$ dimensional so the total is n^2-dimensional, in accordance with the idea that $DIAG^*$ is the complement of a positive codimension subvariety.

However, let us now consider $DIAG$, the set of all *diagonalizable* matrices in $M_n(\mathbb{C})$, possibly with coinciding eigenvalues. Then of course $DIAG$ is still of full dimension n^2, but the subspace $\mathcal{D} \cap DIAG$ where two eigenvalues coincide is in fact complex codimension three!

Example: The generic 2×2 complex matrix is diagonalizable. However, the space of diagonalizable 2×2 matrices with coinciding eigenvalues is the space of matrices $\lambda 1$ and is one complex dimensional. Clearly this has codimension three!

More generally, if we consider the space of diagonalizable matrices and look at the subspace where two eigenvalues coincide, but all the others are distinct then we have a fibration over $(\mathbb{C}^{n-1} - \Delta)$ with fiber

$$GL(n)/(GL(2)^n \times GL(1)^{n-2}) \tag{22.38}$$

of dimension $n^2 - (n - 2 + 4) = n^2 - n - 2$. The base is of dimension $(n - 1)$ so the total space is of dimension $n^2 - 3$ and hence complex codimension 3 in $Mat_n(\mathbb{C})$.

The above discussion might seem puzzling since we also argued that the space of matrices where characteristic roots of the $p_A(x)$ coincide is complex codimension one, not three. Of course, the difference is accounted form by considering matrices with nontrivial Jordan form. The orbit of

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \tag{22.39}$$

is two-complex-dimensional. Together with the parameter λ this makes a complex codimension one subvariety of $M_n(\mathbb{C})$.

More generally

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \oplus_i \lambda_i 1 \tag{22.40}$$

with distinct λ_i has a stabilizer of dimension $2 + (n - 2) = n$ so the space of such matrices is $(n^2 - n) + (n - 2) + 1 = n^2 - 1$ dimensional.

22.2.3 Local model near $\mathcal{S}_{\text{sing}}$

It is of some interest to make a good model of how matrices degenerate when approaching a *generic* point in $\mathcal{S}_{\text{sing}}$.

Suppose $s \to s_* \in \mathcal{S}_{\text{sing}}$ and two distinct roots of the characteristic polynomial say, $\lambda_1(s), \lambda_2(s)$, have a common limit $\lambda(s_*)$. We can choose a family of projection operators

$P(s)$ of rank 2, whose range is the two-dimensional subspace spanned by the eigenvectors $v_1(s)$ and $v_2(s)$ so that

$$T(s) = P_s t(s) P_s + Q_s T(s) Q_s \tag{22.41}$$

and such that $Q_s T(s) Q_s$ has a limit with distinct eigenvalues on the image of Q_s. The operator $t(s)$ is an operator on a two-dimensional subspace and we may choose some fixed generic ordered basis and write:

$$t(s) = \lambda(s) 1 + \begin{pmatrix} z(s) & x(s) - iy(s) \\ x(s) + iy(s) & -z(s) \end{pmatrix} \tag{22.42}$$

where λ, x, y, z are all *complex.* $\lambda(s)$ is some smooth function going to $\lambda(s_*)$ while

$$x(s_*)^2 + y(s_*)^2 + z(s_*)^2 = 0 \tag{22.43}$$

is some generic point on the nilpotent cone of 2×2 nilpotent matrices. That generic point will have $z(s_*) \neq 0$ and hence $x(s_*) \pm iy(s_*) \neq 0$.

***** FIGURE OF A DOUBLE-CONE WITH A PATH ENDING ON THE CONE AT A POINT s_* NOT AT THE TIP OF THE CONE ****

Therefore we can consider the smooth matrix

$$S(s) = \begin{pmatrix} z(s) & 1 \\ x(s) + iy(s) & 0 \end{pmatrix} \tag{22.44}$$

which will be invertible in some neighborhood of s_*. Now a small computation shows that

$$S^{-1} t(s) S = \lambda(s) 1 + \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} \tag{22.45}$$

$$w = x^2 + y^2 + z^2 \tag{22.46}$$

Therefore we can take w to be a coordinate in the normal bundle to $\mathcal{D} \subset M_2(\mathbb{C})$ and the generic family of degenerating matrices has (generically) a nonsingular family of bases where the operator can be modeled as

$$T(w) = P \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} P + T^{\perp} \qquad w \in \mathbb{C} \tag{22.47}$$

22.2.4 Families of Hermitian operators

If we impose further conditions then the rule for the codimension can again change.

An important example arises if we consider families of *Hermitian* matrices. In this case, the codimension of the subvariety where two eigenvalues coincide is *real* codimension 3.

Let us prove this for a family of 2×2 matrices. Our family of matrices is

$$\begin{pmatrix} d_1(s) & z(s) \\ \bar{z}(s) & d_2(s) \end{pmatrix} \tag{22.48}$$

where d_1, d_2 are real and z is complex.

The eigenvalues coincide when the discriminant of the characteristic polynomial vanishes. This is the condition

$$b^2 - 4ac = (d_1 + d_2)^2 - 4(d_1 d_2 - |z|^2) = (d_1 - d_2)^2 + 4|z|^2 = 0 \tag{22.49}$$

Thus, $d_1 = d_2$ and $z = 0$ is the subvariety. Moreover, in the neighborhood of this locus the family is modeled on

$$\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} + \vec{x} \cdot \vec{\sigma} \tag{22.50}$$

For the general case the subspace of Hermitian matrices where exactly two eigenvalues coincide and are otherwise distinct is a fibration over

$$(\mathbb{R}^{(n-2)} - \Delta)/S_{n-2} \times \mathbb{R} \tag{22.51}$$

(the fibration being given by the map to the unordered set of eigenvalues) with fiber:

♣Again, this discussion is out of place. We have not discussed quotients yet. ♣

$$U(n)/(U(2) \times U(1)^{n-2}) \tag{22.52}$$

The fiber has real dimension $n^2 - (4 + (n-2)) = n^2 - n - 2$ and the base has real dimension $n - 1$ so the total dimension is $(n^2 - n - 2) + (n - 1) = n^2 - 3$. So the codimension is 3.

**** FIGURE OF $\mathcal{D}$ AS LINE WITH TRANSVERSE SPACE A PLANE, BUT LABELED AS $\mathbb{R}^3$ *****

Near to the level crossing the universal form of the matrix is

$$T(\vec{x}) = P_{\vec{x}} \left((\lambda + \vec{a} \cdot \vec{x}) 1_{2\times 2} + \vec{x} \cdot \vec{\sigma} + \mathcal{O}(x^2) \right) P_{\vec{x}} + T^{\perp}(\vec{x}) \tag{22.53}$$

where $\vec{x}$ is a local parameter normal to the codimension three subspace, $P_{\vec{x}}$ is a smooth family of rank 2 projectors with a smooth limit at $\vec{x} = 0$, $\lambda \in \mathbb{R}$, and $\vec{a} \in \mathbb{R}^3$.

Example In solid state physics the energy levels in bands cross at *points* in the Brillouin zone. (See Chapter 4 below.)

Exercise

Show that the subspace of Hermitian matrices where exactly k eigenvalues coincide and are otherwise distinct is of real codimension $k^2 - 1$ in the space of all Hermitian matrices.

The way in which these subspaces fit together is quite intricate. See

V. Arnold, "Remarks on Eigenvalues and Eigenvectors of Hermitian Matrices. Berry Phase, Adiabatic Connections and Quantum Hall Effect," Selecta Mathematica, Vol. 1, No. 1 (1995)

References: For some further discussion see Ref: von Neumann + Wigner, Avron-Seiler, Ann. Phys. 110(1978)85; B. Simon, "Holonomy, the Quantum Adiabatic Theorem, and Berry's Phase," Phys. Rev. Lett. **51**(1983)2167

22.3 Canonical form of a family in a first order neighborhood

In this section we consider families in a slightly different way: Can we put families into canonical form by conjugation?

Suppose we have a family of complex matrices $T(s)$ varying continuously with some control parameter $s \in S$ and we allow ourselves to make similarity transformations $g(s)$,

$$T(s) \to g(s)T(s)g(s)^{-1} \tag{22.54}$$

where $g(s)$ must vary continuously with s.

We know that if we work at a *fixed* value of s then we can put $T(s)$ into Jordan canonical form. We can ask – can we put an arbitrary family of matrices into some canonical form?

For example, if $T(s_0)$ is diagonal for some s_0, can we choose $g(s)$ for s near s_0 so that $T(s)$ is diagonal in the neighborhood of s_0?

This is a hard question in general. Let us consider what we can say to first order in perturbations around s_0. For simplicity, suppose S is a neighborhood around 0 in the complex plane, so s_0 is zero. Write

$$T(s) = T_0 + s\delta M + \mathcal{O}(s^2) \tag{22.55}$$

and let us assume that T_0 has been put into some canonical form, such as Jordan canonical form. Now write $g(s) = 1 + s\delta g + \mathcal{O}(s^2)$. Then

$$g(s)T(s)g(s)^{-1} = T_0 + s\delta M + s[\delta g, T_0] + \mathcal{O}(s^2) \tag{22.56}$$ eq:firstpert

For any matrix $m \in M_n(\mathbb{C})$ let us introduce the operator

$$Ad(m) : M_n(\mathbb{C}) \to M_n(\mathbb{C}) \tag{22.57}$$

$$Ad(m)(X) := [m, X] \tag{22.58}$$

What we learn from (22.56) is that we can "conjugate away" anything in the image of $Ad(T_0)$.

That is, *the space of nontrivial perturbations is the cokernel of the operator* $Ad(T_0)$.

Let us suppose that $T_0 = Diag\{\lambda_1, \dots, \lambda_n\}$ is diagonal with *distinct* eigenvalues. Then if $\delta g \in M_n(\mathbb{C})$

$$(\mathrm{Ad}(T_0)\delta g)_{ij} = (\lambda_i - \lambda_j)\delta g_{ij} \tag{22.59}$$

Thus, the range is the space of off-diagonal matrices. We can always conjugate away any off-diagonal element of $(\delta M)_{ij}$, but we cannot conjugate away the diagonal elements. The cokernel can be represented by the diagonal matrices.

Thus, near a matrix with distinct eigenvalues we can, at least to first order in perturbation theory, take the perturbed matrix to be diagonal.

If some eigenvalues coincide then we might not be able to conjugate away some off-diagonal elements.

Moreover, as we have seen above, it is perfectly possible to have a family of matrices degenerate from diagonalizable to nontrivial Jordan form.

Exercise

Compute the cokernel of the operator

$$Ad(J_\lambda^{(k)}) : M_k(\mathbb{C}) \to M_k(\mathbb{C}) \tag{22.60}$$

and find a natural subspace complementary to $\mathrm{im}(Ad(J_\lambda^{(k)}))$ in $M_k(\mathbb{C})$.

Answer: As we showed in equation (10.55) above the kernel of $Ad(J_\lambda^{(k)})$ consists of matrices of the form

$$A = a_1\mathbf{1_k} + a_2(e_{1,2}+e_{2,3}+\cdots+e_{k-1,k}) + a_3(e_{1,3}+e_{2,4}+\cdots+e_{k-2,k}) + \cdots + a_k e_{1,k} \tag{22.61}$$

That is, it is a general polynomial in $J_\lambda^{(k)}$. It is therefore a k-dimensional space. By the index, we know that the cokernel is therefore k-dimensional. Therefore, there is a k-dimensional space of nontrivial perturbations.

Example: By direct computation we find for $k = 2$ the general perturbation is equivalent to

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \delta_1 & \delta_2 \end{pmatrix} \tag{22.62}$$

and in general the matrices

$$J_\lambda^{(k)} + \sum_{i=1}^{k} \delta_i e_{ki} \tag{22.63}$$

for $\delta_1, \dots, \delta_k$ free parameters give a set of representatives of the cokernel.

b.) Show that

$$\det\left(J_\lambda^{(k)} + \sum_{i=1}^{k} \delta_i e_{ki} \right) = \lambda^k + \delta_k \lambda^{k-1} - \delta_{k-1}\lambda^{k-2} \pm \cdots + (-1)^{k-1}\delta_1 \tag{22.64}$$

See R. Gilmore, *Catastrophe Theory*, ch. 14 for further details. There are many applications of the above result.

22.4 Families of operators and spectral covers

In this subsection we describe the spectral cover construction which allows us to translate the data of a family of operators into a purely geometrical object.

Thus, suppose we have a generic family $T(s)$ of complex $n \times n$ matrices over $s \in \mathcal{S}$. As we saw above, generically $T(s)$ is regular semisimple so we let

$$\mathcal{S}^* = \mathcal{S} - \mathcal{S}_{\text{sing}} \tag{22.65}$$

and for $s \in \mathcal{S}^*$ we have

$$T(s) = \sum_{i=1}^{n} \lambda_i(s) P_{\lambda_i}(s) \tag{22.66}$$

where $P_{\lambda_i}(s)$ are orthogonal projection operators and the $\lambda_i(s)$ are *distinct*. Note that the sum on the RHS makes sense without any choice of ordering of the λ_i.

Now consider the space

$$\widetilde{\mathcal{S}} := \{(s, \lambda) | p_{T(s)}(\lambda) = 0\} \subset \mathcal{S} \times \mathbb{C} \tag{22.67}$$

That is, the fiber of the map $\widetilde{\mathcal{S}} \to \mathcal{S}$ is the space of λ's which are characteristic values of $T(s)$ at $s \in \mathcal{S}$.

Over $\mathcal{S}^*$ all the eigenvalues are distinct so

$$\widetilde{\mathcal{S}}^* \to \mathcal{S}^* \tag{22.68}$$

is a smooth n-fold cover. In general since $\pi_1(\mathcal{S}^*) \neq 0$ it will be a nontrivial cover, meaning that there is only locally, but not globally a well-defined ordering of the eigenvalues $\lambda_1(s), \dots, \lambda_n(s)$. In general only the unordered set $\{\lambda_1(s), \dots, \lambda_n(s)\}$ is well-defined over $\mathcal{S}^*$. See equation (22.75) et. seq. below for the case $n = 2$.

Now, note that there is a well-defined map to the space of rank one projectors:

$$P : \widetilde{\mathcal{S}}^* \to Gr_1(\mathbb{C}^n) \tag{22.69}$$

Namely, if $(s, \lambda) \in \widetilde{\mathcal{S}}^*$ then λ is an eigenvalue of $T(s)$ and $P(s, \lambda)$ is the projector to the eigenline $L_\lambda \subset V = \mathbb{C}^n$ spanned by the eigenvector with eigenvalue λ.

In general, a vector bundle whose fibers form a dimension one vector space is called a *line bundle*. Moreover, if we look at the fibers over the different sheets of the covering space $\widetilde{\mathcal{S}}^* \to \mathcal{S}^*$ we get an s-dependent decomposition of V into a sum of lines:

$$V = \oplus_{i=1}^n L_{\lambda_i(s)} \tag{22.70}$$

Now the RHS does depend on the ordering of the $\lambda_i(s)$, but up to isomorphism it does not depend on an ordering. The situation for $n = 2$ is sketched schematically in Figure 10

> Therefore, there is a 1-1 correspondence between the family of operators $T(s)$ parametrized by $s \in \mathcal{S}^*$ and complex line bundles over the n-fold covering space $\widetilde{\mathcal{S}}^* \to \mathcal{S}^*$.

Now, let us ask what happens when we try to consider the full family over $\mathcal{S}$. Certainly $\widetilde{\mathcal{S}} \to \mathcal{S}$ still makes sense, but now, over $\mathcal{S}_{\text{sing}}$ some characteristic values will coincide and the covering is an n-fold *branched covering*.

Let us recall the meaning of the term *branched covering*. In general, a branched covering is a map of pairs $\pi : (Y, R) \to (X, B)$ where $R \subset Y$ and $B \subset X$ are of real codimension two. R is called the *ramification locus* and B is called the *branch locus*. The map $\pi : Y - R \to X - B$ is a regular covering and if it is an n-fold covering we say the branched covering is an n-fold branched cover. On the other hand, near any point $b \in B$ there is a neighborhood U of b and local coordinates

$$(x_1, \dots, x_{d-2}; w) \in \mathbb{R}^{d-2} \times \mathbb{C}, \tag{22.71}$$

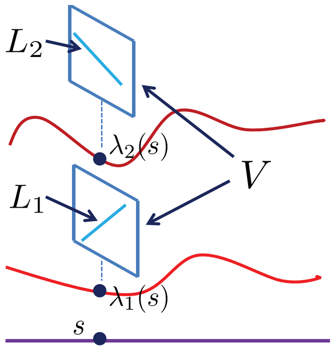

Figure 10: We consider a family of two-by-two matrices $T(s)$ with two distinct eigenvalues $\lambda_1(s)$ and $\lambda_2(s)$. These define a two-fold covering of $\mathcal{S}$, the spectral cover. Moreover, the eigenlines associated to the two sheets give a decomposition of $V = \mathbb{C}^2$ into a sum of lines which varies with s. fig:TWOEIGENLI

where $\dim_{\mathbb{R}} X = d$ and $w = 0$ describes the branch locus $B \cap U$. More importantly, $\pi^{-1}(U) = \amalg_\alpha \widetilde{U}_\alpha$ is a disjoint union of neighborhoods in Y of points $r_\alpha \in R$ with local coordinates

$$(x_1, \dots, x_{d-2}; \xi_\alpha) \in \mathbb{R}^{d-2} \times \mathbb{C} \tag{22.72}$$

so that the map $\pi_\alpha : \widetilde{U}_\alpha \to U$ (where π_α is just the restriction of π) is just given by

$$\pi_\alpha : (x_1, \dots, x_{d-2}; \xi_\alpha) \to (x_1, \dots, x_{d-2}; \xi_\alpha^{e_\alpha}) \tag{22.73}$$

where e_α are positive integers called *ramification indices.*

In plain English: For any $b \in B$ there are several points r_α in the preimage of π above b and near any r_α the map π looks like a mapping of unit disks in the complex plane $\xi \to w = \xi^e$. Note that for an n-fold covering

$$\sum_\alpha e_\alpha = n \tag{22.74}$$

The case where exactly one ramification index is $e = 2$ and all the others are equal to one is called a *simple branch point.* ♣Put in figure of disks covering disks. ♣

Now, we have argued that the interesting part of $T(s)$ near a generic point $s_* \in \mathcal{S}_{\rm sing}$ is of the form

$$T(w) = \begin{pmatrix} 0 & 1 \\ w & 0 \end{pmatrix} \tag{22.75}$$

eq:SimpleDegFa

For this family the characteristic polynomials are clearly

$$p_{T(w)}(x) = x^2 - w \tag{22.76}$$

and the set of characteristic roots is the unordered set $\{+\sqrt{w}, -\sqrt{w}\}$. This is just the unordered set $\{+\xi, -\xi\}$ labeling the two sheets of the 2-fold branched cover. By taking appropriate real slices we may picture the situation as in Figure 11.

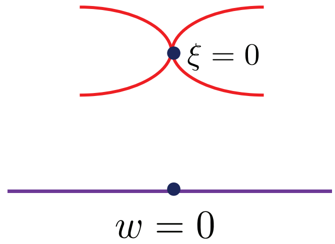

Figure 11: A particular real representation of a complex 2-fold branched cover of a disk over a disk. The horizontal axis in the base is the real line of the complex w plane. The verticle axis corresponds to the real axis of the complex ξ plane on the right and the purely imaginary axis of the complex ξ plane on the left. fig:BRANCHCOVE

The monodromy of the eigenvalues is nicely illustrated by considering the closed path

$$w(t) = w_0 e^{2\pi i t} \qquad\qquad 0 \leq t \leq 1 \tag{22.77}$$ eq:closedcurve

If we choose at $t = 0$ a particular squareroot ξ_0 of w_0 then this closed path lifts to

$$\xi(t) = \xi_0 e^{\pi i t} \qquad\qquad 0 \leq t \leq 1 \tag{22.78}$$ eq:liftedcurve

and the value at $t = 1$ is the *other* squareroot $-\xi_0$ of w_0.

Now let us consider the eigenlines at w. If $w \neq 0$ (i.e. $s \in \mathcal{S}^*$) then there are two distinct eigenlines which are the span of the eigenvectors

$$v_\pm = \begin{pmatrix} 1 \\ \pm\xi \end{pmatrix} \tag{22.79}$$

The eigenlines are just the images of the two Bott projectors:

$$P_\epsilon(\xi) = \frac{1}{1+|\xi|^2} \begin{pmatrix} 1 & \epsilon\xi^* \\ \epsilon\xi & |\xi|^2 \end{pmatrix} \tag{22.80}$$

where $\epsilon = \pm 1$. Along the lifted curve (22.78) P_+ evolves into P_- and so the two eigenlines get exchanged under monodromy.

Note well that $P_\epsilon(\xi)$ has a nice smooth limit as $\xi \to 0$. Therefore the two eigenlines smoothly degenerate to a single line over the single point $\xi = 0$.

In the holomorphic case one can use the fact that holomorphic functions of many variables cannot have complex codimension two singularities (Hartog's theorem) to conclude that:

Therefore, at least for generic holomorphic families over a complex space $\mathcal{S}$ there is a 1-1 correspondence between the families of $n \times n$ matrices $T(s)$ and holomorphic line bundles over n-fold branched covers $\widetilde{\mathcal{S}} \to \mathcal{S}$.

Remarks

1. *Higgs bundles.* The method of spectral covers is very important in certain aspects of Yang-Mills theory. In particular, a version of the Yang-Mills equations on $C \times \mathbb{R}^2$, where C is a Riemann surface lead to differential equations on a connection on C known as the *Hitchin equations*. It turns out that solutions to these are equivalent to holomorphic families of operators on C with the technical difference that the operators $T(s)$, with $s \in C$, (known as *Higgs fields*) are valued in $(1,0)$-forms on C. In this case the spectral cover technique converts a difficult analytic problem such as the solution of Yang-Mills equations into a purely geometrical problem: Describing holomorphic line bundles on n-fold coverings in T^*C.

Exercise
Describe the spectral cover for the family

$$T(z) = \begin{pmatrix} 0 & 1 \\ z^2 - E & 0 \end{pmatrix} \tag{22.81}$$

over the complex plane $z \in \mathbb{C}$.

Exercise
Suppose that $T(z)$ is a two-dimensional holomorphic matrix. Show that the spectral cover is a hyperelliptic curve and write an equation for it in terms of the matrix elements of $T(z)$.

Exercise *Hitchin map*
The characteristic polynomial defines a natural map $h : \mathrm{End}(V) \to \mathbb{C}^n$, known in this context as the Hitchin map:

$$\begin{aligned} h : T \mapsto (a_1(T), \dots, a_n(T)) \\ \det(x\mathbf{1_n} - T) = x^n + a_1 x^{n-1} + \cdots + a_n \end{aligned} \tag{22.82}$$ eq:hitchin

Define the universal spectral cover over $\mathbb{C}^n$ to be

$$\widetilde{\mathbb{C}^n} := \{(a_1, \dots, a_n; t) | t^n + a_1 t^{n-1} + \cdots + a_n = 0\} \tag{22.83}$$ eq:univers

Show that $\bar{S}$ is a fiber product of S with $\widetilde{\mathbb{C}^n}$. [35]

22.5 Families of matrices and differential equations

Another way families of matrices arise is in the theory of differential equations.

As motivation, let us consider the Schrödinger equation:

$$\left(-\hbar^2 \frac{d^2}{dx^2} + V(x)\right)\psi(x) = E\psi(x) \tag{22.84}$$

where we have rescaled V and E by $2m$. We keep $\hbar$ because we are going to discuss the WKB approximation. We could scale it to one and then some other (dimensionless!) physical parameter must play the role of a small parameter.

We can write this as the equation

$$\hbar^2 \frac{d^2}{dx^2}\psi(x) = v(x)\psi(x) \tag{22.85}$$

with $v(x) = V(x) - E$. Moreover, if $v(x)$ is the restriction to the real line of a meromorphic function $v(z)$ on the complex plane we can write an ODE on the complex plane:

$$\hbar^2 \partial_z^2 \psi = v(z)\psi \tag{22.86}$$

This can be converted to a 2×2 matrix equation:

$$\hbar \frac{\partial}{\partial z}\psi = \begin{pmatrix} 0 & 1 \\ v(z) & 0 \end{pmatrix}\psi \tag{22.87}$$

eq:mtrx-schro

where

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{22.88}$$

is a column vector.

Given this reformulation it is natural to consider more generally the matrix differential equation

$$\hbar \frac{\partial}{\partial z}\psi = A(z)\psi \tag{22.89}$$

eq:norderf

where $A(z)$ is some meromorphic matrix function of z.

For example, the Airy differential equation corresponds to

$$A(z) = \begin{pmatrix} 0 & 1 \\ z & 0 \end{pmatrix} \tag{22.90}$$

while the harmonic oscillator corresponds to

$$A(z) = \begin{pmatrix} 0 & 1 \\ z^2 - E & 0 \end{pmatrix} \tag{22.91}$$

eq:HO-MATRIX

[35] *Answer*: T defines a map $p_T : \mathcal{S} \to \mathbb{C}^n$ by taking the coefficients of the characteristic polynomial. You can then pull back the covering $\widetilde{\mathbb{C}^n} \to \mathbb{C}^n$ along T to get $\widetilde{\mathcal{S}}$.

For the theory we are developing it is interesting to generalize further and let $A(z)$ be a meromorphic $n \times n$ matrix. This subsumes the theory of n^{th} order linear ODE's, but the more general setting is important in applications and leads to greater flexibility.

Now if $A(z)$ is nonsingular in a simply connected region $\mathcal{R}$ then there is an n-dimensional space of solutions of (22.89) in $\mathcal{R}$. If $\psi^{(1)}, \dots, \psi^{(n)}$ are n linearly independent solutions they can all be put together to make an $n \times n$ matrix solution

$$\Psi = \begin{pmatrix} \psi^{(1)} & \cdots & \psi^{(n)} \end{pmatrix} = \begin{pmatrix} \psi_1^{(1)} & \cdots & \psi_1^{(n)} \\ \vdots & & \vdots \\ \psi_n^{(1)} & \cdots & \psi_n^{(n)} \end{pmatrix} \tag{22.92}$$

The solutions $\psi^{(i)}$ are linearly independent, iff Ψ is invertible.

The best way to think about solutions of linear ODE's is in fact to look for invertible matrix solutions to

$$\hbar \frac{\partial}{\partial z} \Psi = A(z) \Psi \tag{22.93}$$ eq:bestway

Note that if C is a constant matrix, then $\Psi \to \Psi C$ is another solution. We can think of this freedom as corresponding to the choices of initial conditions specifying the independent solutions $\psi^{(i)}$.

What happens if we multiply from the left by a constant matrix? The equation is not preserved in general since in general C will not commute with $A(z)$. This is not a bug, but a feature, and indeed it is useful to consider more generally making a redefinition

$$\Psi(z) = g(z) \tilde{\Psi}(z) \tag{22.94}$$

where $g(z)$ is an invertible $n \times n$ matrix function of z. Then we change the equation to:

$$\hbar \frac{\partial}{\partial z} \tilde{\Psi} = A^g(z) \tilde{\Psi} \tag{22.95}$$

where

$$A^g = g^{-1} A g - \hbar g^{-1} \partial_z g \tag{22.96}$$

is the "gauge-transform" of A.

Thus we learn, that when working with families of matrices, the proper notion of equivalence might not be that of conjugation, but of gauge transformation. Which is the proper notion depends on the problem under consideration.

From this point of view, solving (22.93) can be interpreted as finding the gauge transformation Ψ that gauges A to 0.

What can we say about the existence of solutions? Let us think of z as valued on the complex plane. Locally, if $A(z)$ does not have singularities then we can always solve the equation with the path ordered exponential. We choose a path $z(t)$ starting from z_{in}

$$\begin{aligned}\Psi(z(t)) &= \text{Pexp}\left[\frac{1}{\hbar}\int_{z_{\text{in}}}^{z(t)} A(w)dw\right]\Psi(z_{\text{in}})\\ &:= \left(1+\sum_{n=1}^{\infty}\hbar^{-n}\int_{z_{\text{in}}}^{z} dz_1\int_{z_{\text{in}}}^{z_1} dz_2\cdots\int_{z_{\text{in}}}^{z_{n-1}} dz_n A(z_1)\cdots A(z_n)\right)\Psi(z_{\text{in}})\\ &:= \left(1+\sum_{n=1}^{\infty}\frac{1}{\hbar^n n!}\int_0^t dt_1\frac{dz(t_1)}{dt_1}\int_0^t dt_2\frac{dz(t_2)}{dt_2}\cdots\int_0^t dt_n\frac{dz(t_n)}{dt_n}P[A(z(t_1))\cdots A(z(t_n))]\right)\Psi(z_{\text{in}})\end{aligned} \tag{22.97}$$

eq:Pexp

where all the integrations are along the path: $\int dz ... = \int dt\frac{dz}{dt}...$ and $P[...]$ is the time-ordered product with later times on the left.

If $A(z)$ is meromorphic then the path ordered exponential exists provided the path $z(t)$ does not go through any singularities. This is clear from the third line of (22.97) since $A(z(t))$ will be bounded on the interval $[0,t]$. The second line makes clear that $\Psi(z)$ is locally holomorphic. However, the solution *does* depend on the homotopy class of the path in $\mathbb{C}$ minus the singularities. We will return to this below.

22.5.1 The WKB expansion

Sometimes the path-ordered exponential is prohibitively difficult to evaluate and we have to make due with approximate solutions. One way to get such approximate solutions is the WKB method. Note that if we could gauge transform $A(z)$ to be *diagonal* then the equation is easily solved, because in the 1×1 case:

$$\hbar\partial_z\psi = a(z)\psi \Rightarrow \psi(z) = \exp\left(\frac{1}{\hbar}\int_{z_0}^{z} a(w)dw\right)\psi(z_0) \tag{22.98}$$

This observation motivates us to find a gauge transformation which implements the diagonalization order by order in $\hbar$. To this end we introduce the ansatz:

$$\Psi_{wkb} = S(z,\hbar)e^{\frac{1}{\hbar}\Delta(z,\hbar)} \tag{22.99}$$

eq:WKB-ansatz

where $\Delta(z,\hbar)$ is *diagonal*, and we assume there are series expansions

$$S(z,\hbar) = S_0(z)+\hbar S_1(z)+\hbar^2 S_2(z)+\cdots \tag{22.100}$$

$$\Delta(z,\hbar) = \Delta_0(z)+\hbar\Delta_1(z)+\hbar^2\Delta_2(z)+\cdots \tag{22.101}$$

Now, to determine these series we substitute (22.99) into equation (22.93) and bring the resulting equation to the form

$$(A(z)-\hbar S' S^{-1})S = S\Delta' \tag{22.102}$$

eq:wkbi

where $\Delta' = \frac{\partial\Delta}{\partial z}$ and $S' = \frac{\partial S}{\partial z}$. Now we look at this equation order by order in $\hbar$. At zeroth order we get

$$A(z)S_0 = S_0\Delta_0' \tag{22.103}$$

Thus, $S_0(z)$ must diagonalize $A(z)$ and Δ'_0 is the diagonal matrix of eigenvalues of $A(z)$. Of course, $A(z)$ might fail to be diagonalizable at certain places. We will return to this.

Now write (22.102) as

$$A(z)S - \hbar S' = S\Delta' \tag{22.104}$$ eq:wkbii

Now substitute $A(z) = S_0\Delta'_0 S_0^{-1}$. Next, make a choice of diagonalizing matrix S_0 and multiply the equation on the left by S_0^{-1} to get

$$\Delta'_0 S_0^{-1} S - \hbar S_0^{-1} S' = S_0^{-1} S \Delta' \tag{22.105}$$ eq:wkbiii

Equation (22.105) is the best form in which to substitute the series in $\hbar$. At order $\hbar^n$, $n > 0$ we get

$$\Delta'_0 S_0^{-1} S_n - S_0^{-1} S'_{n-1} = \sum_{i=0}^{n} S_0^{-1} S_i \Delta'_{n-i} \tag{22.106}$$ eq:wkbiv

separating out the $i = 0$ and $i = n$ terms from the RHS the equation is easily rearranged to give

$$\begin{aligned} &[\Delta'_0, S_0^{-1} S_1] - \Delta'_1 = S_0^{-1} S'_0 \\ &[\Delta'_0, S_0^{-1} S_n] - \Delta'_n = S_0^{-1} S'_{n-1} + \sum_{i=1}^{n-1} S_0^{-1} S_i \Delta'_{n-i} \qquad\qquad n > 1 \end{aligned} \tag{22.107}$$ eq:wkbivp

In an inductive procedure, every term on the RHS is known. On the left-hand side Δ'_n is *diagonal*, so we take it to be the diagonal component of the RHS.

As long as the eigenvalues of Δ'_0 are *distinct* $Ad(\Delta'_0)$ maps ONTO the space of off-diagonal matrices, and hence we can solve for $S_0^{-1} S_n$. In this way we generate the WKB series.

Remarks

1. The WKB procedure will fail when $A(z)$ has nontrivial Jordan form. This happens when the characteristic polynomial has multiple roots. These are the branch points of the Riemann surface defined by the characteristic equation.

2. Indeed, returning to the matrix $A(z)$ corresponding to the Schrödinger equation (22.87). The characteristic equation $\lambda^2 - v(z) = 0$ has branchpoints at zeroes $v(z_0)$. Recalling that $v(x) = V(x) - E$ these are just the turning points of the usual mechanics problem. In the neighborhood of such points one can write an exact solution in terms of Airy functions, and then match to the WKB solution to produce a good approximate solution.

3. Note that - so long as $A(z)$ is diagonalizable with distinct eigenvalues - the above procedure only determines $S_0(z)$ up to right-multiplication by a diagonal matrix $D_0(z)$. However, the choice of diagonal matrix then enters in the equation determining Δ_1 and S_1: $\Delta_1' = -Diag(S_0^{-1}S_0')$. Thus, if we change our choice

$$S_0(z) \to \tilde{S}_0(z) = S_0(z)D_0(z) \tag{22.108}$$

then we have:

$$\tilde{\Delta}_1(z) = -\mathrm{Diag}(\tilde{S}_0^{-1}\tilde{S}_0') = -\mathrm{Diag}(S_0^{-1}S_0') - \mathrm{Diag}(D_0^{-1}D_0') \tag{22.109}$$

and when substituting into (22.99) the change of S_0 is canceled by the change of Δ_1. Similarly, S_n is only determined up to the addition of a matrix of the form S_0D_n where D_n is an arbitrary diagonal matrix function of z. However, at the next stage in the procedure D_n will affect Δ_{n+1} in such a way that the full series $S\exp[\frac{1}{\hbar}\Delta]$ is unchanged.

4. In general, the WKB series is only an *asymptotic series*.

Exercise
Write the general n^{th} order linear ODE in matrix form.

Exercise
Show that for the matrix

$$A(z) = \begin{pmatrix} 0 & 1 \\ v(z) & 0 \end{pmatrix} \tag{22.110}$$

We can must have $\Delta_0'(z) = \sqrt{v(z)}\sigma^3$ and we can choose S_0 to be

$$S_0 = \begin{pmatrix} v^{-1/4} & v^{-1/4} \\ v^{1/4} & -v^{1/4} \end{pmatrix} \tag{22.111}$$

Show that with this choice of $S_0(z)$ we have

$$[\Delta_0', S_0^{-1}S_1] - \Delta_1' = -\frac{1}{4}\sigma^1\frac{d}{dz}\log v(z) \tag{22.112}$$

and therefore

$$\Delta(z,\hbar) = \sigma_3 \int_{z_0}^{z} \sqrt{v(z')}dz' + \mathcal{O}(\hbar^2) \tag{22.113}$$

22.5.2 Monodromy representation and Hilbert's 21st problem

Consider again the matrix equation (22.93).

If $A(z)$ is holomorphic near z_0 then so is the solution $\Psi(z)$. On the other hand, there will be interesting behavior when $A(z)$ has singularities.

Definition A *regular singular point* z_* is a point where $A(z)$ has Laurent expansion of the form

$$A(z) = \frac{A_{-1}}{z - z_*} + \cdots \tag{22.114}$$

with A_{-1} regular semisimple.

We have

Theorem [Fuchs' theorem]: Near a RSP there exist convergent series solutions in a disk around $z = z_*$. and if $A_{-1} = \mathrm{Diag}\{\lambda_1, \dots, \lambda_n\}$ then

$$\Psi(z) = \mathrm{Diag}\{(z-z_*)^{\lambda_1}, \dots, (z-z_*)^{\lambda_n}\} \left(1 + \Psi_1(z-z_*) + \Psi_2(z-z_*)^2 + \cdots\right) \tag{22.115}$$

where $\Psi_1, \Psi_2, \dots$ are constant matrices, is a *convergent series* in some neighborhood of z_*

Note that in general the solution will have monodromy around $z = z_*$: Analytic continuation around a counterclockwise oriented simple closed curve around $z = 0$ gives

$$\Psi(z_* + (z - z_*)e^{2\pi i}) = \mathrm{Diag}\{e^{2\pi i \lambda_1}, \dots, e^{2\pi i \lambda_n}\}\Psi(z) \tag{22.116}$$

If $A(z)$ only has regular singular points at, say, $p_1, \dots, p_s$ then analytic continuation defines a *representation*

$$\rho : \pi_1(\mathbb{C} - \{p_1, \dots, p_s\}, z_0) \to GL(n, \mathbb{C}) \tag{22.117}$$ eq:MonRep

known as the *monodromy representation of the differential equation.*

Remark: Riemann was the first to investigate this problem, completely solving the case of $n = 2$ with three regular singular points. In his famous address to the International Congress of Mathematicians in Paris in 1900 D. Hilbert presented a list of 23 problems for the mathematics of the 20^{th} century. The 21^{st} problem was, roughly, in modern terms:

Given an irreducible n-dimensional representation (22.117)*, find a differential equation for which it is a monodromy representation.*

This problem has a complicated history, with claimed solutions and counterexamples. We note that there is a very physical approach to the problem using free fermion conformal field theory correlation functions which was pursued by the Kyoto school of Miwa, Jimbo, et. al. It is also the first example of what is called the "Riemann-Hilbert correspondence," which plays an important role in algebraic geometry.

♣Explain more clearly if the problem has been solved or not! ♣

22.5.3 Stokes' phenomenon

A subject closely related to the WKB analysis is Stokes' phenomenon. We give a brief account here.

Definition An *irregular singular point* is a singular point of the form

$$A(z) = \frac{A_{-n}}{z^n} + \cdots \tag{22.118}$$

with $n > 1$.

Let us consider the simplest kind of ISP, which we put at $z = 0$:

$$A(z) = \frac{R}{z^2} + \frac{A_{-1}}{z} + \cdots \tag{22.119}$$

with $R = \mathrm{Diag}\{r_1, \dots, r_n\}$. Then the series method will produce a *formal* solution

$$\Psi_f = \left(1 + \Psi_1 z + \Psi_2 z^2 + \cdots\right) e^{-R/z} \tag{22.120}$$

The big difference will now be that the series is only *asymptotic* for $z \to 0$.

♣HERE GOES DEFINITION OF ASYMPTOTIC SERIES. STRESS DEPENDENCE ON CHOICE OF RAY. ♣

Example: Consider

$$\frac{d}{dz}\Psi = \left(\frac{r\sigma^3}{z^2} + \frac{s\sigma^1}{z}\right)\Psi \tag{22.121}$$

Substituting

$$\Psi_f = U e^{-r\sigma^3/z} \tag{22.122}$$

with $U = 1 + zU_1 + z^2 U_2 + \cdots$ we find

$$\frac{dU}{dz} = \mathrm{Ad}(\frac{r\sigma^3}{z^2})U + \frac{s\sigma^1}{z}U \tag{22.123}$$

Writing this out we get the equations

$$\begin{aligned} [r\sigma^3, U_1] &= -s\sigma^1 \\ [\sigma^3, U_{n+1}] &= \frac{(n - s\sigma^1)}{r}U_n \end{aligned} \tag{22.124}$$

and the factor of n on the RHS in the second line shows that coefficients in U_n are going to grow like $n!$ and hence the series will only be asymptotic.

Indeed, in this case the formal series is easily shown to be

$$U_n = -\left(\frac{-1}{2r}\right)^n \frac{\prod_{j=1}^{n-1}(j^2 - s^2)}{n!}(s + n\sigma^1)(\sigma^3)^n \tag{22.125}$$ eq:FormSerEx

so the prefactor grows like $n!$.

Definition The rays $(r_i - r_j)\mathbb{R}_+$ starting at $z = 0$ are known as *Stokes rays* and the open regions between these rays are *Stokes sectors*. See Figure 12.

Now one can prove

Theorem: Let ρ be a ray which is *not* a Stokes ray, and let $\mathbb{H}_\rho$ be the half-plane containing ρ as in Figure 13. Then there is a unique solution Φ_ρ which is asymptotic to the formal solution as $z \to 0$ along any ray in $\mathbb{H}_\rho$:

$$\Phi_\rho e^{R/z} \to 1 \tag{22.126}$$

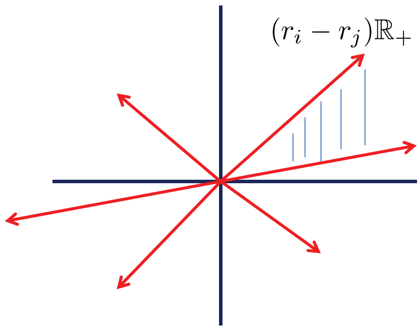

Figure 12: Stokes sectors fig:STOKESECTO

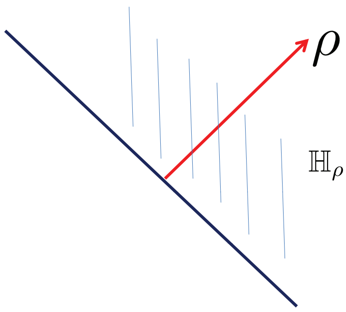

Figure 13: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$. fig:HALFPLANE

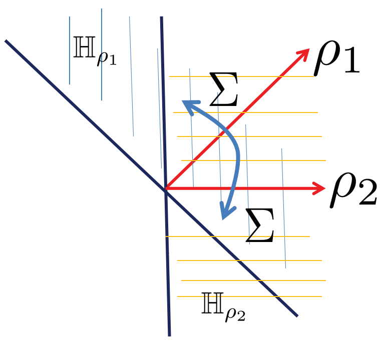

Figure 14: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$. fig:HPOVERLAP

where $z \to 0$ along any ray in $\mathbb{H}_\rho$, and hence

$$\lim_{z\to 0} z^{-n} \left(\Phi_\rho(z) e^{R/z} - (1 + z\Psi_1 + \cdots + z^n \Psi_n) \right) = 0 \qquad (22.127)$$

It is very important here that the limit is taken along a ray in $\mathbb{H}_\rho$, otherwise the statement will be false. Indeed, in general one *cannot* find a formal solution in a *larger* domain which is asymptotic to the formal solution! This is one version of *Stokes phenomenon.*

Now, consider two rays ρ_1, ρ_2, neither of which is a Stokes ray. The half-planes overlap as in Figure 14. Then, by uniqueness of the solution to the differential equation we know that

$$\Phi_{\rho_1} = \Phi_{\rho_2} S_\Sigma \qquad \text{on} \qquad \mathbb{H}_{\rho_1} \cap \mathbb{H}_{\rho_2} \tag{22.128}$$

where S_Σ is a matrix which is constant as a function of z. (It might well depend on other parameters in the differential equation.) Moreover, $S_\Sigma = 1$ if there is no Stokes' ray in Σ, but $S_\Sigma \neq 1$ if there are Stokes' rays in Σ. If there is precisely one Stokes ray ℓ in Σ then we set $S_\Sigma = S_\ell$ and call S_ℓ the *Stokes factor* for ℓ. ♣That sentence is really a theorem. ♣

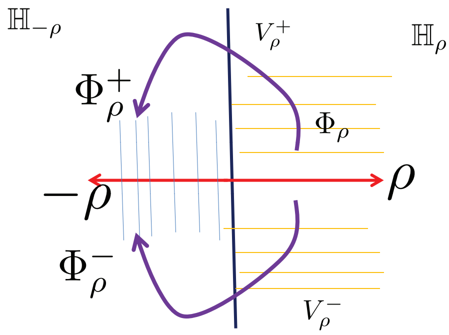

Figure 15: There is a true solution to the differential equation asymptotic to the formal solution in the half-plane $\mathbb{H}_\rho$. fig:TWOAC

Now we can describe an analog of monodromy for ISP's: Choose rays $\pm\rho$ which are not Stokes rays. Starting with Φ_ρ in $\mathbb{H}_\rho$ there are two analytic continuations to $\mathbb{H}_{-\rho}$, as shown in Figure 15. Call these two analytic continuations $\Phi_\rho^\pm$ then we have:

Theorem:

$$\Phi_\rho^+ = \Phi_{-\rho} S_+ \qquad \text{in} \qquad \mathbb{H}_{-\rho} \tag{22.129}$$

$$\Phi_\rho^- = \Phi_{-\rho} S_- \qquad \text{in} \qquad \mathbb{H}_{-\rho} \tag{22.130}$$

are given by

$$S_+ =: \prod_{\ell \in V_+(\rho)}^{ccw} S_\ell : \tag{22.131}$$

$$S_- =: \prod_{\ell \in V_-(\rho)}^{cw} S_\ell : \tag{22.132}$$

where the products are ordered so successive rays are counterclockwise or clockwise. These are called Stokes matrices, and serve as the analogs of monodromy matrices in the irregular singular point case.

Remarks

1. There is a generalization of this story to higher order poles. If

$$A(z) = \frac{R}{z^{\ell+1}} + \cdots \tag{22.133}$$

with R regular semisimple then the formal series solution has the form

$$\Psi_f = U(z)e^{Q(z)} \tag{22.134}$$

$$U(z) = 1 + zU_1 + z^2U_2 + \cdots \tag{22.135}$$

$$Q(z) = \frac{Q_\ell}{z^\ell} + \cdots + \frac{Q_1}{z} + Q_0 \log z \tag{22.136}$$

Moreover, a true solution asymptotic to the formal solution will generally only exist in angular sectors of angular width $|\Delta\theta| < \frac{\pi}{\ell}$. For more details about this see Coddington and Levinson, or Hille's book on ODEs.

2. There is a great deal more to be said about the kind of groups the Stokes matrices live in and their use in parametrizing flat connections and their applications to Yang-Mills theory.

Exercise

a.) Derive the formal series (22.125).

b.) Show that the equation can be reduced to a second order ODE with one irregular singular point and one regular singular point.

Answer:

a.) Write

$$U_n = w_n 1 + x_n\sigma^1 + y_n\sigma^2 + z_n\sigma^3 \tag{22.137}$$

so that

$$\begin{aligned} 2ix_{n+1}\sigma^2 - 2iy_{n+1}\sigma^1 &= \frac{n}{r}(x_n\sigma^1 + y_n\sigma^2 + z_n\sigma^3) \\ &\quad - \frac{s}{r}(x_n + iy_n\sigma^3 - iz_n\sigma^2) \end{aligned} \tag{22.138}$$

Deduce that $nw_n = sx_n$ and $nz_n = isy_n$ and

$$x_{n+1} = \frac{1}{2ir}\frac{n^2 - s^2}{n}y_n \qquad y_{n+1} = -\frac{1}{2ir}\frac{n^2 - s^2}{n}x_n \tag{22.139}$$

The induction starts with

$$U_1 = \frac{s}{2ir}\sigma^2 + \frac{s^2}{2r}\sigma^3 \tag{22.140}$$

so $x_{2n+1} = y_{2n} = 0$. The rest is simple induction.

b.) Write out the differential equation on a two-component column vector and eliminate ψ_2 to get:

$$\psi'' + \frac{2ir - sz}{z(ir - sz)}\psi' - \frac{r^2 + s^2z^2}{z^4}\psi = 0 \tag{22.141}$$

♣Transform this to the confluent hg equation. ♣

Exercise *Harmonic Oscillator*

Consider the matrix (22.91) corresponding to the harmonic oscillator differential equation. Using the known asymptotics of the parabolic cylinder functions work out the Stokes sectors and Stokes factors for this equation.

Exercise *Three singular points*

Suppose that a second order differential equation on the extended complex plane has three singular points with monodromy given by regular semisimple elements conjugate to

$$\begin{pmatrix} \mu_i & 0 \\ 0 & \mu_i^{-1} \end{pmatrix} \qquad i = 1, 2, 3 \tag{22.142}$$

Show that for generic $\mu_i \in \mathbb{C}$ the equation $M_1 M_2 M_3 = 1$ can be solved, up to simultaneous conjugation of the M_i, in terms of the μ_i, and write M_2, M_3 in a basis where M_1 is diagonal.

In fancy mathematical terms: The moduli space of flat $SL(2, \mathbb{C})$ connections with fixed conjugacy class of regular semisimple monodromy around three points on $\mathbb{C}P^1 - \{p_1, p_2, p_3\}$ has no moduli.

♣YOU SHOULD GIVE AN EXAMPLE OF HOW THIS GIVES REFLECTION/TRANSMISSION COEFFICIENTS IN SOME SCATTERING PROBLEMS. ♣

---

23. $\mathbb{Z}_2$-graded, or super-, linear algebra

rLinearAlgebra

In this section "super" is merely a synonym for "$\mathbb{Z}_2$-graded." Super linear algebra is extremely useful in studying supersymmetry and supersymmetric quantum theories, but its applications are much broader than that and the name is thus a little unfortunate.

Superlinear algebra is very similar to linear algebra, but there are some crucial differences, which we highlight in this section. *It's all about signs.*

We are going to be a little bit pendantic and long-winded in this section because the subject is apt to cause confusion.

23.1 Super vector spaces

It is often useful to add the structure of a $\mathbb{Z}_2$*-grading* to a vector space. A $\mathbb{Z}_2$-graded vector space over a field κ is a vector space over κ which, moreover, is written as a direct sum

$$V = V^0 \oplus V^1. \tag{23.1}$$

eq:zeet

The vector spaces V^0, V^1 are called the even and the odd subspaces, respectively. We may think of these as eigenspaces of a "parity operator" P_V which satisfies $P_V^2 = 1$ and is $+1$ on V^0 and -1 on V^1. If V^0 and V^1 are finite dimensional, of dimensions m, n respectively we say the super-vector space has graded-dimension or superdimension $(m|n)$.

A vector $v \in V$ is called *homogeneous* if it is an eigenvector of P_V. If $v \in V^0$ it is called *even* and if $v \in V^1$ it is called *odd*. We may define a *degree* or *parity* of homogeneous vectors by setting $\deg(v) = \bar{0}$ if v is even and $\deg(v) = \bar{1}$ if v is odd. Here we regard $\bar{0}, \bar{1}$ in the additive abelian group $\mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\}$. Note that if v, v' are homogeneous vectors of the same degree then

$$\deg(\alpha v + \beta v') = \deg(v) = \deg(v') \tag{23.2}$$

eq:dnge

for all $\alpha, \beta \in \kappa$. We can also say that $P_V v = (-1)^{\deg(v)} v$ acting on homogeneous vectors. For brevity we will also use the notation $|v| := \deg(v)$. Note that $\deg(v)$ is *not defined* for general vectors in V.

Mathematicians define the *category of super vector spaces* so that a *morphism* from $V \to W$ is a linear transformation which *preserves grading*. We will denote the space of morphisms from V to W by $\underline{\mathrm{Hom}}(V, W)$. These are just the ungraded linear transformations of ungraded vector spaces, $T : V \to W$, which commute with the parity operator $TP_V = P_W T$.

So far, there is no big difference from, say, a $\mathbb{Z}$-graded vector space. However, important differences arise when we consider *tensor products*.

So far we defined a category of supervector spaces, and now we will make it into a *tensor category*. (See definition below.)

The tensor product of two $\mathbb{Z}_2$ graded spaces V and W is $V \otimes W$ as vector spaces over κ, but the $\mathbb{Z}_2$-grading is defined by the rule:

$$\begin{aligned} (V \otimes W)^0 &:= V^0 \otimes W^0 \oplus V^1 \otimes W^1 \\ (V \otimes W)^1 &:= V^1 \otimes W^0 \oplus V^0 \otimes W^1 \end{aligned} \tag{23.3}$$

eq:tsnp

Thus, under tensor product the degree is additive on homogeneous vectors:

$$\deg(v \otimes w) = \deg(v) + \deg(w) \tag{23.4}$$

eq:tnesvct

If κ is any field we let $\kappa^{p|q}$ denote the supervector space:

$$\kappa^{p|q} = \underbrace{\kappa^p}_{\text{even}} \oplus \underbrace{\kappa^q}_{\text{odd}} \tag{23.5}$$

Thus, for examples:

$$\mathbb{R}^{n_e|n_o} \otimes \mathbb{R}^{n'_e|n'_o} \cong \mathbb{R}^{n_e n'_e + n_o n'_o | n_e n'_o + n_o n'_e} \tag{23.6}$$

and in particular:

$$\mathbb{R}^{1|1} \otimes \mathbb{R}^{1|1} = \mathbb{R}^{2|2} \tag{23.7}$$

$$\mathbb{R}^{2|2} \otimes \mathbb{R}^{2|2} = \mathbb{R}^{8|8} \tag{23.8}$$

$$\mathbb{R}^{8|8} \otimes \mathbb{R}^{8|8} = \mathbb{R}^{128|128} \tag{23.9}$$

Now, in fact we have a *braided tensor category*:

In ordinary linear algebra there is an isomorphism of tensor products

$$c_{V,W} : V \otimes W \to W \otimes V \tag{23.10}$$

eq:BrdIso

given by $c_{V,W} : v \otimes w \mapsto w \otimes v$. In the super-commutative world there is also an isomorphism (23.10) defined by taking

$$c_{V,W} : v \otimes w \to (-1)^{|v| \cdot |w|} w \otimes v \tag{23.11}$$

eq:SuperBraid

on homogeneous objects, and extending by linearity.

Let us pause to make two remarks:

1. Note that in (23.11) we are now viewing $\mathbb{Z}/2\mathbb{Z}$ as a ring, not just as an abelian group. Do not confuse $\deg v + \deg w$ with $\deg v \deg w$! In computer science language $\deg v + \deg w$ corresponds to *XOR*, while $\deg v \deg w$ corresponds to *AND*.

2. It is useful to make a general rule: In equations where the degree appears it is understood that all quantities are homogeneous. Then we extend the formula to general elements by linearity. Equation (23.11) is our first example of a general rule: In the supercommutative world, commuting any object of homogeneous degree A with an object of homogeneous degree B results in an "extra" sign $(-1)^{AB}$. This is sometimes called the "*Koszul sign rule.*"

With this rule the tensor product of a collection $\{V_i\}_{i \in I}$ of supervectorspaces

$$V_{i_1} \otimes V_{i_2} \otimes \cdots \otimes V_{i_n} \tag{23.12}$$

eq:TensSupVect

of supervector spaces is well-defined and *independent of the ordering of the factors*. This is a slightly nontrivial fact. See the remarks below.

We define the $\mathbb{Z}_2$-graded-symmetric and $\mathbb{Z}_2$-graded-antisymmetric products to be the images of the projection operators

$$P = \frac{1}{2}\left(1 \pm c_{V,V}\right) \tag{23.13}$$

Therefore the $\mathbb{Z}_2$-graded-symmetric product of a supervector space is the $\mathbb{Z}_2$-graded vector space with components:

$$\begin{aligned} S^2(V)^0 &\cong S^2(V^0) \oplus \Lambda^2(V^1) \\ S^2(V)^1 &\cong V^0 \otimes V^1 \end{aligned} \tag{23.14}$$

and the $\mathbb{Z}_2$-graded-antisymmetric product is

$$\begin{aligned} \Lambda^2(V)^0 &\cong \Lambda^2(V^0) \oplus S^2(V^1) \\ \Lambda^2(V)^1 &\cong V^0 \otimes V^1 \end{aligned} \tag{23.15}$$

Remarks

1. In this section we are stressing the differences between superlinear algebra and ordinary linear algebra. These differences are due to important signs. If the characteristic of the field κ is 2 then ± 1 are the same. Therefore, in the remainder of this section we assume κ is a field of characteristic different from 2.

2. Since the transformation $c_{V,W}$ is nontrivial in the $\mathbb{Z}_2$-graded case the fact that (23.12) is well-defined is actually slightly nontrivial. To see the issue consider the tensor product $V_1 \otimes V_2 \otimes V_3$ of three super vector spaces. Recall the relation $(12)(23)(12) = (23)(12)(23)$ of the symmetric group. Therefore, we should have "coherent" isomorphisms:

$$(c_{V_2,V_3} \otimes 1)(1 \otimes c_{V_1,V_3})(c_{V_1,V_2} \otimes 1) = (1 \otimes c_{V_1,V_2})(c_{V_1,V_3} \otimes 1)(1 \otimes c_{V_2,V_3}) \tag{23.16}$$

eq:z2yb

   and this is easily checked.

   In general a *tensor category* is a category with a bifunctor $\mathcal{C}\times\mathcal{C} \to \mathcal{C}$ denoted $(X, Y) \to X\otimes Y$ with an associativity isomorphism $F_{X,Y,Z} : (X\otimes Y)\otimes Z \cong X\otimes(Y\otimes Z)$ satisfying the pentagon coherence relation. A braiding is an isomorphism $c_{X,Y} : X\otimes Y \to Y\otimes X$. The associativity and braiding isomorphisms must satisfy "coherence equations." The category of supervector spaces is perhaps the simplest example of a braided tensor category going beyond the category of vector spaces.

3. Note well that $S^2(V)$ as a supervector space does *not* even have the same dimension as $S^2(V)$ in the ungraded sense! Moreover, if V has a nonzero odd-dimensional summand then $\Lambda^n(V)$ does not vanish no matter how large n is.

4. With this notion of symmetric product we can nicely unify the bosonic and fermionic Fock spaces. If we have a system with bosonic and fermionic oscillators then there is a natural supervector space V spanned by the bosonic and fermionic creation operators, where the bosonic oscillators are even and the fermionic oscillators are odd. Then the $\mathbb{Z}_2$-graded, or super-Fock space $S^\bullet(V)$ naturally gives the full Fock space of the free boson-fermion system. That is, we have the isomorphism of ungraded vector spaces:

$$\underbrace{S^\bullet V}_{\text{graded symmetrization}} = \underbrace{S^\bullet V^0 \otimes \Lambda^\bullet V^1}_{\text{ungraded tensor product of vector spaces}} \tag{23.17}$$

Exercise
a.) Show that $c_{V,W}c_{W,V} = 1$.
b.) Check (23.16).

Exercise *Reversal of parity*

a.) Introduce an operation which switches the parity of a supervector space: $(\Pi V)^0 = V^1$ and $(\Pi V)^1 = V^0$. Show that Π defines a functor of the category of supervector spaces to itself which squares to one.

b.) In the category of finite-dimensional supervector spaces when are V and ΠV isomorphic? [36]

c.) Show that one can identify ΠV as the functor defined by tensoring V with the canonical odd one-dimensional vector space $\kappa^{0|1}$.

23.2 Linear transformations between supervector spaces

If the ground field κ is taken to have degree 0 then the dual space $V^\vee$ in the category of supervector spaces consists of the morphisms $V \to \kappa^{1|0}$. Note that $V^\vee$ inherits a natural $\mathbb{Z}_2$ grading:

$$\begin{aligned}(V^\vee)^0 &:= (V^0)^\vee \\ (V^\vee)^1 &:= (V^1)^\vee\end{aligned} \tag{23.18}$$

eq:duals

Thus, we can say that $(V^\vee)^\epsilon$ are the linear functionals $V \to \kappa$ which vanish on $V^{1+\epsilon}$.

Taking our cue from the natural isomorphism in the ungraded theory:

$$\mathrm{Hom}(V, W) \cong V^\vee \otimes W \tag{23.19}$$

we use the same definition so that the space of linear transformations between two $\mathbb{Z}_2$-graded spaces becomes $\mathbb{Z}_2$ graded. We also write $\mathrm{End}(V) = \mathrm{Hom}(V, V)$.

In particular, a linear transformation is an even linear transformation between two $\mathbb{Z}_2$-graded spaces iff $T : V^0 \to W^0$ and $V^1 \to W^1$, and it is odd iff $T : V^0 \to W^1$ and $V^1 \to W^0$. Put differently:

$$\begin{aligned}\mathrm{Hom}(V, W)^0 &\cong \mathrm{Hom}(V^0, W^0) \oplus \mathrm{Hom}(V^1, W^1) \\ \mathrm{Hom}(V, W)^1 &\cong \mathrm{Hom}(V^0, W^1) \oplus \mathrm{Hom}(V^1, W^0)\end{aligned} \tag{23.20}$$

The general linear transformation is neither even nor odd.

If we choose a basis for V made of vectors of homogeneous degree and order it so that the even degree vectors come first then with respect to such a basis even transformations have block diagonal form

$$T = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \tag{23.21}$$

eq:matev

while odd transformations have block diagonal form

[36] *Answer*: An isomorphism is a degree-preserving isomorphism of vector spaces. Therefore if V has graded dimension $(m|n)$ then ΠV has graded dimension $(n|m)$ so they are isomorphic in the category of supervector spaces iff $n = m$.

$$T = \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix} \tag{23.22}$$ eq:matevp

Remarks

1. Note well! There is a difference between $\mathrm{Hom}(V, W)$ and $\underline{\mathrm{Hom}}(V, W)$. The latter is the space of morphisms from V to W in the category of supervector spaces. They consist of just the even linear transformations: [37]

$$\underline{\mathrm{Hom}}(V, W) = \mathrm{Hom}(V, W)^0 \tag{23.23}$$

2. If $T : V \to W$ and $T' : V' \to W'$ are linear operators on super-vector-spaces then we can define the $\mathbb{Z}_2$ graded tensor product $T \otimes T'$. Note that $\deg(T \otimes T') = \deg(T) + \deg(T')$, and on homogeneous vectors we have

$$(T \otimes T')(v \otimes v') = (-1)^{\deg(T')\deg(v)} T(v) \otimes T'(v') \tag{23.24}$$ eq:tensortmsn

As in the ungraded case, $\mathrm{End}(V)$ is a ring, but now it is a $\mathbb{Z}_2$-graded ring under composition: $T_1 T_2 := T_1 \circ T_2$. That is if $T_1, T_2 \in \mathrm{End}(V)$ are homogeneous then $\deg(T_1 T_2) = \deg(T_1) + \deg(T_2)$, as one can easily check using the above block matrices. These operators are said to *graded-commute*, or *supercommute* if

$$T_1 T_2 = (-1)^{\deg T_1 \deg T_2} T_2 T_1 \tag{23.25}$$ eq:KRule

Remark: Now what shall we take for the definition of $GL(\kappa^{p|q})$? This should be the group of automorphisms of the object $\kappa^{p|q}$ in the category of super-vector-spaces. These must be even invertible maps and so

$$GL(\kappa^{p|q}) \cong GL(p; \kappa) \times GL(q; \kappa). \tag{23.26}$$ eq:kpq-iso

Some readers will be puzzled by equation (23.26). The Lie algebra of this group (see Chapter 8) is

$$\mathfrak{gl}(p; \kappa) \oplus \mathfrak{gl}(q; \kappa) \tag{23.27}$$

and is *not* the standard super Lie algebra $\mathfrak{gl}(p|q; \kappa)$. (See Chapter 12).

Another indication that there is something funny going on is that the naive definition of $GL(\kappa^{p|q})$, namely that it is the subset of $\mathrm{End}(\kappa^{p|q})$ of invertible linear transformations, will run into problems with (23.25). For example consider $\kappa^{1|1}$. Then, choosing a basis (say 1) for κ we get a basis of homogeneous vectors on $\kappa^{1|1}$. Then the operator

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{23.28}$$

[37] Warning! Some authors use the opposite convention for distinguishing hom in the category of super-vector spaces from "internal hom."

is an odd element with $T^2 = 1$. On the other hand, we might have expected T to super-commute with itself, but then the sign rule (23.25) implies [38] that if it super-commutes with itself then $T^2 = 0$, but this is not the case.

We will define a more general group with the correct super Lie algebra, but to do so we need to discuss the notion of supermodules over a superalgebra.

Exercise

Show that if $T : V \to W$ is a linear transformation between two super-vector spaces then

a.) T is even iff $TP_V = P_W T$

b.) T is odd iff $TP_V = -P_W T$.

23.3 Superalgebras

The set of linear transformations $\mathrm{End}(V)$ of a supervector space is an example of a superalgebra. In general we have:

Definition

a.) A *superalgebra* $\mathcal{A}$ is a supervector space over a field κ together with a morphism

$$\mathcal{A} \otimes \mathcal{A} \to \mathcal{A} \tag{23.29}$$

of supervector spaces. We denote the product as $a \otimes a' \mapsto aa'$. Note this implies that

$$\deg(aa') = \deg(a) + \deg(a'). \tag{23.30}$$

We assume our superalgebras to be unital so there is a $1_{\mathcal{A}}$ with $1_{\mathcal{A}} a = a 1_{\mathcal{A}} = a$. Henceforth we simply write 1 for $1_{\mathcal{A}}$.

b.) The superalgebra is *associative* if $(aa')a'' = a(a'a'')$.

c.) Two elements a, a' in a superalgebra are said to *graded-commute*, or *super-commute* provided

$$aa' = (-1)^{|a||a'|} a'a \tag{23.31}$$

If every pair of elements a, a' in a superalgebra graded-commmute then the superalgebra is called *graded-commutative* or *supercommutative.*

d.) The *supercenter*, or $\mathbb{Z}_2$*-graded center* of an algebra, denoted $Z_s(\mathcal{A})$, is the subsuperalgebra of $\mathcal{A}$ such that all homogeneous elements $a \in Z_s(\mathcal{A})$ satisfy

$$ab = (-1)^{|a||b|} ba \tag{23.32}$$

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

Example 1: *Matrix superalgebras.* If V is a supervector space then $\mathrm{End}(V)$ as described above is a matrix superalgebra. One can show that the supercenter is isomorphic to κ, consisting of the transformations $v \to \alpha v$, for $\alpha \in \kappa$.

[38] so long as the characteristic of κ is not equal to two

Example 2: *Grassmann algebras.* The Grassmann algebra of an ordinary vector space W is just the exterior algebra of W considered as a $\mathbb{Z}_2$-graded algebra. We will denote it as $\mathrm{Grass}[W]$.

In plain English, we take vectors in W to be odd and use them to generate a superalgebra with the rule that

$$w_1 w_2 + w_2 w_1 = 0 \tag{23.33}$$

for all w_1, w_2. In particular (provided the characteristic of κ is not two) we have $w^2 = 0$ for all w.

Thus, if we choose basis vectors $\theta^1, \dots, \theta^n$ for W then we can view $\mathrm{Grass}(W)$ as the quotient of the supercommutative polynomial superalgebra $\kappa[\theta^1, \dots, \theta^n]/I$ where the relations in I are:

$$\theta^i\theta^j + \theta^j\theta^i = 0 \qquad\qquad (\theta^i)^2 = 0 \tag{23.34}$$

The typical element then is

$$a = x + x_i\theta^i + \frac{1}{2!}x_{ij}\theta^i\theta^j + \cdots + \frac{1}{n!}x_{i_1,\dots,i_n}\theta^{i_1}\cdots\theta^{i_n} \tag{23.35}$$

The coefficients $x_{i_1,\dots,i_m}$ are m^{th}-rank totally antisymmetric tensors in $\kappa^{\otimes m}$.

We will sometimes also use the notation $\mathrm{Grass}[\theta^1, \dots, \theta^n]$.

Definition Let $\mathcal{A}$ and $\mathcal{B}$ be two superalgebras. The *graded tensor product* $\mathcal{A}\widehat{\otimes}\mathcal{B}$ is the superalgebra which is the graded tensor product as a vector space and the multiplication of homogeneous elements satisfies

$$(a_1 \otimes b_1) \cdot (a_2 \otimes b_2) = (-1)^{|b_1||a_2|}(a_1 a_2) \otimes (b_1 b_2) \tag{23.36}$$ eq:GradedTens

Remarks

1. Every $\mathbb{Z}_2$-graded algebra is also an ungraded algebra: We just forget the grading. However this can lead to some confusions:

2. An algebra can be $\mathbb{Z}_2$-graded-commutative and not ungraded-commutative: The Grassmann algebras are an example of that. We can also have algebras which are ungraded commutative but not $\mathbb{Z}_2$-graded commutative. The Clifford algebras $C\ell_{\pm 1}$ described below provide examples of that.

3. The $\mathbb{Z}_2$-graded-center of an algebra can be different from the center of an algebra as an ungraded algebra. Again, the Clifford algebras $C\ell_{\pm 1}$ described below provide examples.

4. One implication of (23.36) is that when writing matrix representations of graded algebras we do *not* get a matrix representation of the graded tensor product just by taking the tensor product of the matrix representations.

Example 3: *The real Clifford algebras $C\ell_{r+,s-}$.* Clifford algebras are defined for a general quadratic form Q on a vector space V over κ. We will study the Clifford algebras extensively in Chapter 10(??). Nevertheless, a few comments here nicely illustrate some important general points. If we take the case of a real vector space $\mathbb{R}^d$ with quadratic form

$$Q = \begin{pmatrix} +1_r & 0 \\ 0 & -1_s \end{pmatrix} \tag{23.37}$$

Then we get the real Clifford algebras $C\ell_{r+,s-}$. They can also be defined as the $\mathbb{Z}_2$ graded algebra over $\mathbb{R}$ generated by *odd* elements e_i with relations

$$\{e_i, e_j\} = 2Q_{ij} \tag{23.38}$$

Note that since $e_i^2 = \pm 1$ the algebra only admits a $\mathbb{Z}_2$ grading and moreover it is *not supercommutative*, because an odd element squares to zero in a supercommutative algebra.

It is instructive to look at some small values of r, s. Consider $C\ell_{-1}$. This has a single generator e with relation $e^2 = -1$. Therefore

$$C\ell_{-1} = \mathbb{R} \oplus \mathbb{R}e \tag{23.39}$$

as a vector space. The multiplication is

$$(a \oplus be)(c \oplus de) = (ac - bd) \oplus (bc + ad)e \tag{23.40}$$

so $C\ell_{-1}$ is isomorphic to the complex numbers $\mathbb{C}$ as an *ungraded algebra*, although not as a graded algebra. Similarly, $C\ell_{+1}$ is

$$C\ell_{+1} = \mathbb{R} \oplus \mathbb{R}e \tag{23.41}$$

as a vector space with multiplication:

$$(a \oplus be)(c \oplus de) = (ac + bd) \oplus (bc + ad)e. \tag{23.42}$$

As an ungraded algebra this is sometimes known as the "double numbers."

Note that both $C\ell_{-1}$ and $C\ell_{+1}$ are *commutative* as ungraded algebras but *noncommutative* as superalgebras. Thus the centers of these as ungraded algebras are $C\ell_{\pm 1}$ but the supercenter of $C\ell_{\pm 1}$ as graded algebras are $Z_s(C\ell_{\pm 1}) \cong \mathbb{R}$. In fact, for Q nondegenerate it can be shown that

$$Z_s(C\ell(Q)) \cong \mathbb{R} \tag{23.43}$$

We can also look at graded tensor products. First, note that for $n > 0$:

$$C\ell_n \cong \underbrace{C\ell_1 \widehat{\otimes} \cdots \widehat{\otimes} C\ell_1}_{\text{n times}} \tag{23.44}$$

$$C\ell_{-n} \cong \underbrace{C\ell_{-1} \widehat{\otimes} \cdots \widehat{\otimes} C\ell_{-1}}_{\text{n times}} \tag{23.45}$$

More generally we have

$$C\ell_{r+,s-} = \underbrace{C\ell_1 \widehat{\otimes} \cdots \widehat{\otimes} C\ell_1}_{\text{r times}} \widehat{\otimes} \underbrace{C\ell_{-1} \widehat{\otimes} \cdots \widehat{\otimes} C\ell_{-1}}_{\text{s times}} \tag{23.46}$$

We can similarly discuss the complex Clifford algebras $\mathbb{C}\ell_n$. Note that over the complex numbers if $e^2 = +1$ then $(ie)^2 = -1$ so we do not need to account for the signature, and WLOG we can just consider $\mathbb{C}\ell_n$ for $n \geq 0$. In particular, let $D \cong \mathbb{C}\ell_1$. Note that D is *not* a matrix superalgebra since it's dimension as an ordinary complex vector space, namely 2, is not a perfect square.

Definition A super-algebra over κ is *central simple* if, after extension of scalars to an algebraic closure $\bar{\kappa}$ it is isomorphic to a matrix super algebra $\mathrm{End}(V)$ or to $\mathrm{End}(V)\widehat{\otimes}D$.

This is the definition one finds in Section 3.3 of Deligne's *Notes on Spinors.* In particular, it is shown in Chapter 10, with this definition, that the Clifford algebras over $\mathbb{R}$ and $\mathbb{C}$ are central simple.

Exercise *The opposite algebra*

a.) For any ungraded algebra A we can define the *opposite algebra* A^{opp} by the rule

$$a \cdot^{\mathrm{opp}} b := ba \tag{23.47}$$

Show that A^{opp} is still an algebra.

b.) Show that $A \otimes A^{\mathrm{opp}} \cong \mathrm{End}(A)$.

c.) For any superalgebra A we can define the *opposite superalgebra* A^{opp} by the rule

$$a \cdot^{\mathrm{opp}} b := (-1)^{|a||b|} ba \tag{23.48}$$

Show that A^{opp} is still an superalgebra.

d.) Show that A is supercommutative iff $A = A^{\mathrm{opp}}$.

e.) Show that $A\widehat{\otimes}A^{\mathrm{opp}} \cong \mathrm{End}(A)$ as superalgebras.

f.) Show that if $\mathcal{A} = C\ell_{r+,s-}$ then $\mathcal{A}^{\mathrm{opp}} = C\ell_{s+,r-}$.

Exercise *Super Ideals*

An ideal I in a superalgebra is an ideal in the usual sense: For all $a \in \mathcal{A}$ and $b \in I$ we have $ab \in I$ (left ideal) or $ba \in I$ (right ideal), or both (two-sided ideal). The ideal is *homogeneous* if I is the direct sum of $I^0 = I \cap \mathcal{A}^0$ and $I^1 = I \cap \mathcal{A}^1$. (Explain why this is a nontrivial condition!)

a.) Show that the ideal $\mathcal{I}^{\text{odd}}$ generated by all odd elements in $\mathcal{A}$ is homogeneous and given by

$$\mathcal{I}^{\text{odd}} = (\mathcal{A}^1)^2 \oplus \mathcal{A}^1 \tag{23.49}$$

b.) Show that

$$\mathcal{A}/\mathcal{I}^{\text{odd}} \cong \mathcal{A}^0/((\mathcal{A}^1)^2) \tag{23.50}$$

c.) Another definition of *central simple* is that there are no nontrivial homogeneous two-sided ideals. Show that this is equivalent to the definition above.

d.) Describe an explicit basis for the ideal generated by all odd elements in the Grassmann algebra $\kappa[\theta_1, \dots, \theta_n]$.

e.) Give an example of a supercommutative algebra which is *not* a Grassmann algebra. [39]

Exercise *Invertibility lemma*

Let $\mathcal{A}$ be a *supercommutative* superalgebra and let $\mathcal{I}^{\text{odd}} = (\mathcal{A}^1)$ be the ideal generated by odd elements. Let π be the projection

$$\pi : \mathcal{A} \to \mathcal{A}_{\text{red}} = \mathcal{A}/\mathcal{I}^{\text{odd}} \cong \mathcal{A}^0/((\mathcal{A}^1)^2) \tag{23.51}$$

a.) Show that a is invertible iff $\pi(a)$ is invertible. [40]

b.) Show that in a Grassmann algebra the map π is the same as reduction modulo nilpotents, or, more concretely, just putting the θ^i's to zero.

Exercise *Supercommutators and super Lie algebras*

The *graded commutator* or *supercommutator* of even elements in a superalgebra is

$$[a, b] := ab - (-1)^{|a||b|} ba \tag{23.52}$$

Since the expression $ab - ba$ still makes sense this notation can cause confusion so one must exercise caution when reading.

Show that the graded commutator satisfies:

1. $[\cdot, \cdot]$ is linear in both entries.

[39] *Answer*: Hint: Consider the possiblity that there are even nilpotent elements in $\mathcal{A}^0$ which are not the square of odd elements. Or consider functions on an algebraic supermanifold.

[40] *Answer*: One direction is trivial. If $\pi(a)$ is invertible then, since π is onto, there is an element $b \in \mathcal{A}$ with $1 = \pi(a)\pi(b) = \pi(ab)$. Therefore it suffices to show that if $\pi(a) = 1$ then a is invertible. But if $\pi(a) = 1$ then there is a finite set of odd elements ξ_i and elements $c_i \in \mathcal{A}$ so that $a = 1 - \nu$ with $\nu = \sum_{i=1}^n c_i \xi_i$. Note that $\nu^{n+1} = 0$ (by supercommutativity and the pigeonhole principle) so that $a^{-1} = 1 + \nu + \cdots + \nu^n$.

2. $[b,a] = (-1)^{1+|a||b|}[a,b]$

3. The *super Jacobi identity*:

$$(-1)^{x_1 x_3}[X_1,[X_2,X_3]] + (-1)^{x_2 x_1}[X_2,[X_3,X_1]] + (-1)^{x_3 x_2}[X_3,[X_1,X_2]] = 0 \quad (23.53)$$

where $x_i = \deg(X_i)$.

These two conditions are abstracted from the properties of super-commutators to define super Lie algebras in Chapter 12 below. Briefly: We define a super vector space $\mathfrak{g}$ to be a *super Lie algebra* if there is an (abstract) map $[\cdot,\cdot] : \mathfrak{g}\times\mathfrak{g} \to \kappa$ which satisfies the conditions 1,2,3 above.

Exercise *Super-Derivations*

Definition: A *derivation* of a superalgebra is a homogeneous linear map $D : A \to A$ such that

$$D(ab) = D(a)b + (-1)^{|D||a|}aD(b) \quad (23.54)$$

a.) Show that the supercommutator of two superderivations is a superderivation.

b.) Show that the odd derivations of the Grassmann algebra are of the form

$$\sum_i f^i \frac{\partial}{\partial\theta^i} \quad (23.55)$$

where f^i are even.

Exercise *Multiplying Clifford algebras*

a.) Show that the real Clifford algebras are of dimension $\dim_{\mathbb{R}} C\ell_n = 2^{|n|}$, for any $n \in \mathbb{Z}$.

b.) Show that if n,m are integers with the same sign then $C\ell_n \widehat{\otimes} C\ell_m \cong C\ell_{n+m}$. Show that if n,m are *any* integers, then

$$C\ell_n \widehat{\otimes} C\ell_m \cong C\ell_{n+m} \widehat{\otimes} M \quad (23.56)$$

where M is a matrix superalgebra.

23.4 Modules over superalgebras

Definition A *super-module M over a super-algebra $\mathcal{A}$* (where $\mathcal{A}$ is itself a superalgebra over a field κ) is a supervector space M over κ together with a κ-linear map $\mathcal{A} \times M \to M$ defining a left-action or a right-action. That is, it is a left-module if, denoting the map by $L : \mathcal{A} \times M \to M$ we have

$$L(a, L(b, m)) = L(ab, m) \tag{23.57}$$

and it is a right-module if, denoting the map by $R : \mathcal{A} \times M \to M$ we have

$$R(a, R(b, m)) = R(ba, m) \tag{23.58}$$

In either case:

$$\deg(R(a, m)) = \deg(L(a, m)) = \deg(a) + \deg(m) \tag{23.59}$$

The notations $L(a, m)$ and $R(a, m)$ are somewhat cumbersome and instead we write $L(a, m) = am$ and $R(a, m) = ma$ so that $(ab)m = a(bm)$ and $m(ab) = (ma)b$. We also sometimes refer to a super-module over a super-algebra $\mathcal{A}$ just as a *representation of $\mathcal{A}$*.

Definition A *linear transformation* between two super-modules M, N over $\mathcal{A}$ is a κ-linear transformation of supervector spaces such that if T is homogeneous and M is a left $\mathcal{A}$-module then $T(am) = (-1)^{|T||a|} aT(m)$ while if M is a right $\mathcal{A}$-module then $T(ma) = T(m)a$. We denote the space of such linear transformations by $\mathrm{Hom}_{\mathcal{A}}(M, N)$. If N is a left $\mathcal{A}$-module then $\mathrm{Hom}_{\mathcal{A}}(M, N)$ is a left $\mathcal{A}$-module with $(a \cdot T)(m) := a \cdot (T(m))$. If N is a right $\mathcal{A}$-module then $\mathrm{Hom}_{\mathcal{A}}(M, N)$ is a right $\mathcal{A}$-module with $(T \cdot a)(m) := (-1)^{|a||m|} T(m)a$. When $M = N$ we denote the module of linear transformations by $\mathrm{End}_{\mathcal{A}}(M)$.

Example 1- continued *Matrix superalgebras.* In the ungraded world a matrix algebra $\mathrm{End}(V)$ for a finite dimensional vector space, say, over $\mathbb{C}$, has a unique irreducible representation, up to isomorphism. This is just the space V itself. A rather tricky point is that if V is a supervector space $V = \mathbb{C}^{p|q}$ then V and ΠV are *inequivalent* representations of $\mathrm{End}(V)$. One way to see this is that if η is a generator of $\Pi = \mathbb{C}^{0|1}$ then $T(\eta v) = (-1)^{|T|} \eta T(v)$ is a priori a different module. In terms of matrices

$$\begin{pmatrix} D & -C \\ -B & A \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{23.60}$$

So the LHS gives a representation of the matrix superalgebra, but it is not related by an automorphism $GL(\mathbb{C}^{p|q})$. The even subalgebra $\mathrm{End}(\mathbb{C}^p) \oplus \mathrm{End}(\mathbb{C}^q)$ has a unique faithful representation $\mathbb{C}^p \oplus \mathbb{C}^q$ and hence the matrix superalgebra $\mathrm{End}(\mathbb{C}^{p|q})$ has exactly two irreducible modules.

Example 3- continued *Clifford Modules.* A good example of supermodules over a superalgebra are the $\mathbb{Z}_2$-graded modules for the $\mathbb{Z}_2$-graded Clifford algebras.

Already for $C\ell_0 \cong \mathbb{R}$ there is a difference between graded and ungraded modules. There is a unique irreducible ungraded module, namely $\mathbb{R}$ acting on itself. But there are two inequivalent graded modules, $\mathbb{R}^{1|0}$ and $\mathbb{R}^{0|1}$.

Let us also discuss the representations of $C\ell_{\pm 1}$. As an ungraded algebra $C\ell_{+1} \cong \mathbb{R} \oplus \mathbb{R}$ because we can introduce projection operators $P_\pm = \frac{1}{2}(1 \pm e)$, so

$$C\ell_{+1} \cong \mathbb{R}P_+ \oplus \mathbb{R}P_- \qquad\qquad \text{ungraded!} \tag{23.61}$$

Therefore, there are *two* inequivalent ungraded irreducible representations with carrier space $\mathbb{R}$ and $\rho(e) = \pm 1$. However, as a graded algebra there is a unique irreducible representation, $\mathbb{R}^{1|1}$ with

$$\rho(e) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{23.62}$$

since e is odd and squares to 1.

Similarly, $C\ell_{-1}$ as an *ungraded* algebra is isomorphic to $\mathbb{C}$ and has a unique ungraded irreducible representation: $\mathbb{C}$ acts on itself. (As representations of a real algebra $\rho(e) = \pm i$ are equivalent.) However, as a graded algebra there a unique irreducible representation, $\mathbb{R}^{1|1}$ with

$$\rho(e) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{23.63}$$

Now, $C\ell_{1,-1}$ has two irreducible graded representations $\mathbb{R}^{1|1}_\pm$ with

$$\rho(e_1) = \pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad\qquad \rho(e_2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} := \epsilon \tag{23.64}$$

Note that these are both odd, they anticommute, and they square to ± 1, respectively. Moreover, they generate all linear transformations on $\mathbb{R}^{1|1}$. Therefore, $C\ell_{1,-1}$ is a supermatrix algebra:

$$C\ell_{1,-1} \cong \mathrm{End}(\mathbb{R}^{1|1}) \tag{23.65}$$

It is interesting to compare this with $C\ell_{+2}$. Now, as an *ungraded algebra* we have a representation

$$\rho(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad\qquad \rho(e_2) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{23.66}$$ eq:epsdef

since these matrices anticommute and both square to $+1$. These generate the full matrix algebra $M_2(\mathbb{R})$ as an ungraded algebra. However, if we try to use these operators on $\mathbb{R}^{1|1}$ this is *not* a representation of $C\ell_{+2}$ as a graded algebra because $\rho(e_2)$ is not odd.

In fact, $C\ell_{+2}$ is *not* equivalent to a matrix superalgebra. In Chapter 10(??) we prove the beautiful periodicity theorem (closely related to Bott periodicity):

Theorem $C\ell_{r+,s-}$ is equivalent to a supermatrix algebra iff $(r-s) = 0\mathrm{mod}8$.

There is a unique irreducible representation of $C\ell_{+2}$ as a superalgebra. The carrier space is the (2|2)-dimensional space $\mathbb{R}^{2|2}$ and is given - up to similarity - by

$$\rho(e_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \tag{23.67}$$

It is true that $\mathbb{R}^{2|2} = \mathbb{R}^{1|1}\widehat{\otimes}\mathbb{R}^{1|1}$. But the tensor product of matrix representations does *not* give a matrix representation of the graded tensor product.

If we work with complex Clifford algebras the story is slightly different. $\mathbb{C}\ell_1$ as an ungraded algebra is $\mathbb{C}\oplus\mathbb{C}$ and has two inequivalent ungraded representations. As a graded algebra it has a unique irreducible graded representation $\mathbb{C}^{1|1}$; we could take, for example $\rho(e) = \sigma^1$. Then $\mathbb{C}\ell_2$ as an ungraded algebra is the matrix algebra $M_2(\mathbb{C})$ and as a graded algebra is a matrix superalgebra $\mathrm{End}(\mathbb{C}^{1|1})$. As a matrix superalgebra it actually has *two* inequivalent graded representations, both of which have carrier space $\mathbb{C}^{1|1}$. We could take, for example, $\rho(e_1) = \sigma^1$ and $\rho(e_2) = \pm\sigma^2$. One way to see these are inequivalent is to note that the volume form $\rho(e_1 e_2)$ restricted to the even subspace is a different scalar in the two cases.

We will discuss much more about Clifford modules in Chapter 10, for now, we summarize the discussion here in the following table:

Clifford Algebra	Ungraded algebra	Graded algebra	Ungraded irreps	Graded irreps
$C\ell_{-1}$	$\mathbb{C}$	$\mathbb{R}[e], e^2 = -1$	$\mathbb{C}$	$\mathbb{R}^{1\vert 1}, \rho(e) = \epsilon$
$C\ell_0$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}^{1\vert 0}, \mathbb{R}^{0\vert 1}$
$C\ell_{+1}$	$\mathbb{R}\oplus\mathbb{R}$	$\mathbb{R}[e], e^2 = 1$	$\mathbb{R}_\pm, \rho(e) = \pm 1$	$\mathbb{R}^{1\vert 1}, \rho(e) = \sigma^1$
$C\ell_{+2}$	$M_2(\mathbb{R})$	$C\ell_{+2}$	$\mathbb{R}^2$	$\mathbb{R}^{2\vert 2}$
$C\ell_{+1,-1}$	$M_2(\mathbb{R})$	$\mathrm{End}(\mathbb{R}^{1\vert 1})$	$\mathbb{R}^2$	$\mathbb{R}^{1\vert 1}_\pm$
$\mathbb{C}\ell_{+1}$	$\mathbb{C}\oplus\mathbb{C}$	$\mathbb{C}\ell_{+1}$	$\mathbb{C}_\pm, \rho(e) = \pm 1$	$\mathbb{C}^{1\vert 1}$
$\mathbb{C}\ell_{+2}$	$M_2(\mathbb{C})$	$\mathrm{End}(\mathbb{C}^{1\vert 1})$	$\mathbb{C}^2$	$\mathbb{C}^{1\vert 1}_\pm$

Remark: In condensed matter physics a *Majorana fermion* is a *real* operator γ which squares to one. If there are several γ_i they anticommute. Therefore, the Majorana fermions generate a *real Clifford algebra* within the set of observables of a physical system admitting Majorana fermions. If we have two sets of Majorana fermions then we expect their combined system to be a tensor product. Here we see that only the graded tensor product will produce the expected rule for the physical observables. This is one reason why it is important to take a graded tensor product in the amalgamation axiom in the Dirac-von Neuman axioms.

What about the Hilbert spaces representing states of a Majorana fermion? If we view these as representations of an ungraded algebra then we encounter a famous paradox. (For the moment, take the Hilbert spaces to be real.) $C\ell_{+2}$ as an ungraded algebra has irreducible representation $\mathbb{R}^2$. On the other hand, this is a system of *two* Majorana fermions γ_1 and γ_2 so we expect that each Majorana fermion has a Hilbert space $\mathcal{H}_1$ and $\mathcal{H}_2$ and moreover these are isomorphic, so $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ implies that $\dim_{\mathbb{R}}\mathcal{H}_1 = \dim_{\mathbb{R}}\mathcal{H}_2 = \sqrt{2}$. This is nonsense! If we view the Hilbert space representations as complexifications of real representations then the paradox evaporates with the use of the graded tensor products: As irreducible representations we have:

$$\mathbb{R}^{2|2} = \mathbb{R}^{1|1}\widehat{\otimes}\mathbb{R}^{1|1} \tag{23.68}$$

The situation is a bit more tricky if we use complex graded representations. The paradox returns if we insist on using an irreducible representation of $\mathbb{C}\ell_2$, both in the graded and ungraded cases. However, in the graded case we can say that the 7th DvN axiom is satisfied if the physical (graded) representation is

$$\mathbb{C}^{1|1}\widehat{\otimes}\mathbb{C}^{1|1} \cong \mathbb{C}_+^{1|1} \oplus \mathbb{C}_-^{1|1} \tag{23.69}$$

Exercise *Tensor product of modules*

Let $\mathcal{A}$ and $\mathcal{B}$ be superalgebras with modules M and N, respectively. Show that the rule

$$(a \otimes b) \cdot (m \otimes n) := (-1)^{|b||m|}(am) \otimes (bn) \tag{23.70}$$

does indeed define $M \otimes N$ as an $\mathcal{A}\widehat{\otimes}\mathcal{B}$ module. Be careful with the signs!

Exercise *Left modules vs. right modules*

Suppose $\mathcal{A}$ is supercommutative.

a.) Show that if $(a, m) \to L(a, m)$ is a left-module then the new product $R : \mathcal{A} \times M \to M$ defined by

$$R(a, m) := (-1)^{|a||m|}L(a, m) \tag{23.71}$$

defines M as a *right-module*, that is,

$$R(a_1, R(a_2, m)) = R(a_2 a_1, m) \tag{23.72}$$

b.) Similarly, show that if M is a right-module then it can be canonically considered also to be a left-module.

Because of this we will sometimes write the module multiplication on the left or the right, depending on which order is more convenient to keep the signs down.

Exercise *Representations of Clifford algebras*
Show that

$$\rho(e_1) = \begin{pmatrix} 0 & \sigma^1 \\ \sigma^1 & 0 \end{pmatrix} \qquad \rho(e_2) = \begin{pmatrix} 0 & \sigma^3 \\ \sigma^3 & 0 \end{pmatrix} \tag{23.73}$$

is a graded representation of $C\ell_{+2}$ on $\mathbb{R}^{2|2}$. Show that it is equivalent to the one given above.

23.5 Free modules and the super-General Linear Group

Now let $\mathcal{A}$ be supercommutative. Then we can define a *free right $\mathcal{A}$-module*, $\mathcal{A}^{p|q}$ to be

$$\mathcal{A}^{p|q} = \mathcal{A}^{\oplus p} \oplus (\Pi\mathcal{A})^{\oplus q} \tag{23.74}$$

as a supervector space with the obvious right $\mathcal{A}$-module action.

Since it is a free module we can choose a basis. Set $n = p + q$ and choose a basis e_i, $1 \le i \le n$ with e_i even for $1 \le i \le p$ and odd for $p + 1 \le i \le p + q = n$. Then we can identify

$$\mathcal{A}^{p|q} \cong e_1\mathcal{A} \oplus \cdots \oplus e_n\mathcal{A} \tag{23.75}$$

We define the degree of $e_i a$ to be $\deg(e_i) + \deg(a)$ so that the even part of $\mathcal{A}^{p|q}$ is

$$(\mathcal{A}^{p|q})^0 = e_1\mathcal{A}^0 \oplus \cdots \oplus e_p\mathcal{A}^0 \oplus e_{p+1}\mathcal{A}^1 \oplus \cdots \oplus e_n\mathcal{A}^1 \tag{23.76}$$

Definition: If $\mathcal{A}$ is supercommutative we define $GL(\mathcal{A}^{p|q})$ to be the group of *automorphisms* of $\mathcal{A}^{p|q}$. Recalling that morphisms in the category of supervector spaces are parity-preserving this may be identified with the group of invertible even elements in $\mathrm{End}_{\mathcal{A}}(\mathcal{A}^{p|q})$.

We stress that even though $GL(\mathcal{A}^{p|q})$ is called a supergroup it is actually an honest group. However, it is not an honest manifold, but actually a supermanifold.

♣Mathematicians take a more categorical approach to defining supergroups. Make sure this is compatible. ♣

It is useful to give a matrix description of these groups. We represent the general element m of $\mathcal{A}^{p|q}$ by $m = e_i x^i$, with $x^i \in \mathcal{A}$. Then the general module map $T : \mathcal{A}^{p|q} \to \mathcal{A}^{r|s}$ is determined by its action on basis vectors:

$$T(e_j) = \tilde{e}_\alpha X^\alpha{}_j \qquad X^\alpha{}_j \in \mathcal{A} \tag{23.77}$$

where $\tilde{e}_\alpha$, $\alpha = 1, \ldots, r + s$, are the generators of $\mathcal{A}^{r|s}$.

We say the matrix X with matrix elements $X^{\alpha}{}_{j}$ (which are elements of $\mathcal{A}$) where rows and columns have a parity assigned is an $(r|s) \times (p|q)$ *supermatrix*.

If we choose a basis for $\mathcal{A}^{p|q}$ then we may represent an element $m \in \mathcal{A}^{p|q}$ by a column vector

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \tag{23.78}$$

then the (active) transformation T is given by a matrix multiplication from the left with block form:

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.79}$$

eq:blockform

The supermatrix representing the composition of transformations $T_1 \circ T_2$ is the ordinary matrix product of X_1 and X_2.

When T is an even transformation

$$A \in M_{r\times p}(\mathcal{A}^0) \qquad B \in M_{r\times q}(\mathcal{A}^1) \tag{23.80}$$

$$C \in M_{s\times p}(\mathcal{A}^1) \qquad D \in M_{s\times q}(\mathcal{A}^0) \tag{23.81}$$

or, more informally, X is of the form:

$$\begin{pmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{pmatrix} \tag{23.82}$$

eq:even-blockf

When T is an odd transformation

$$A \in M_{r\times p}(\mathcal{A}^1) \qquad B \in M_{r\times q}(\mathcal{A}^0) \tag{23.83}$$

$$C \in M_{s\times p}(\mathcal{A}^0) \qquad D \in M_{s\times q}(\mathcal{A}^1) \tag{23.84}$$

or, more informally, X is of the form:

$$\begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix} \tag{23.85}$$

Example 1: $\mathcal{A} = \kappa$. Then there are no odd elements in $\mathcal{A}$ and we have invertible morphisms. This group of automorphisms of $\kappa^{p|q}$ is isomorphic to $GL(p;\kappa) \times GL(q;\kappa)$ and in a homogeneous basis will have block diagonal form

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \tag{23.86}$$

with A, D invertible.

Example 2: $\mathcal{A} = \kappa[\theta^1, \ldots, \theta^r]$ is a Grassmann algebra, then $GL(\mathcal{A}^{p|q})$ consists of matrices (23.79) which are even, i.e. of the form (23.82), with A, D invertible, which is the same as A, D being invertible modulo θ^i. (See Section §23.7 below.) For example

$$\begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \tag{23.87}$$

$$\begin{pmatrix} 1+\theta_1\theta_2 & \theta_1 \\ \theta_2 & 1-\theta_1\theta_2 \end{pmatrix} \tag{23.88}$$

are examples of such general linear transformations. In fact they are both of the form $\exp(Y)$ for an even supermatrix Y. (Find it!)

Remarks

1. Note a tricky point: If $T : \mathcal{A}^{p|q} \to \mathcal{A}^{r|s}$ is a linear transformation and we have chosen bases as above so that T is represented by a supermatrix X then the supermatrix representing aT is *not* $aX^{\alpha}{}_{j}$, rather it is the supermatrix:

$$\begin{pmatrix} a1_{r\times r} & 0 \\ 0 & (-1)^{|a|}a1_{s\times s} \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.89}$$

Similarly, the matrix representing Ta is *not* $X^{\alpha}{}_{j}a$, rather it is the supermatrix:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a1_{p\times p} & 0 \\ 0 & (-1)^{|a|}a1_{q\times q} \end{pmatrix} \tag{23.90}$$

2. In Chapter 8 (?) we describe the relation between Lie groups and Lie algebras. Informally this is just given by the exponential map and Lie algebra elements exponentiate to form one-parameter subgroups $g(t) = \exp(tA)$ of G. The same reasoning applies to $GL(\mathcal{A}^{p|q})$ and the super Lie algebra $gl(\mathcal{A}^{p|q})$ is - as a supervector space - the same as $\mathrm{End}(\mathcal{A}^{p|q})$.

23.6 The Supertrace

There are analogs of the trace and determinant for elements of $\mathrm{End}(\mathcal{A}^{p|q})$ with $\mathcal{A}$ supercommutative.

For $X \in \mathrm{End}(\mathcal{A}^{p|q})$ we define the supertrace on homogeneous elements by

$$\mathrm{STr}(X) = \mathrm{STr}(\begin{pmatrix} A & B \\ C & D \end{pmatrix}) := \begin{cases} \mathrm{tr}\,(A) - \mathrm{tr}\,(D) & X \text{ even} \\ \mathrm{tr}\,(A) + \mathrm{tr}\,(D) & X \text{ odd} \end{cases} \tag{23.91}$$

that is

$$\mathrm{STr}(X) = \mathrm{tr}\,(A) - (-1)^{|X|}\mathrm{tr}\,(D) \in \mathcal{A} \tag{23.92}$$

The supertrace satisfies $\mathrm{STr}(X+Y) = \mathrm{STr}(X) + \mathrm{STr}(Y)$ so we can extend it to all of $\mathrm{End}(\mathcal{A}^{p|q})$ by linearity.

Now one can easily check (do the exercise!!) that the supertrace satisfies the properties:

1. $\mathrm{STr}(XY) = (-1)^{|X||Y|}\mathrm{STr}(YX)$ and therefore the supertrace of a graded commutator vanishes. Note that the signs in the definition of the supertrace are crucial for this to be true.

2. $\mathrm{STr}(aX) = a\mathrm{STr}(X)$.

3. If g is even and invertible then $\mathrm{STr}(g^{-1}Xg) = \mathrm{STr}(X)$. This follows from the cyclicity property we just stated. Therefore, the supertrace is basis independent for a free module and hence is an intrinsic property of the linear transformation.

Remark: In the case where $\mathcal{A} = \kappa$ and we have a linear transformation on a supervector space we can say the supertrace of $T \in \mathrm{End}V$ is

$$\mathrm{STr}T := \mathrm{Tr}(P_V T) \tag{23.93}$$ eq:strce

In supersymmetric field theories P_V is often denoted $(-1)^F$, where F is a fermion number and the supertrace becomes $\mathrm{Tr}(-1)^F T$. These traces are very important in obtaining exact results in supersymmetric field theories.

Exercise
a.) Show that in general

$$\mathrm{STr}(T_1T_2) \neq \mathrm{STr}(T_2T_1) \tag{23.94}$$

b.) Check that if T_1, T_2 are homogeneous then

$$\mathrm{STr}(T_1T_2) = (-1)^{\mathrm{deg}T_1 \cdot \mathrm{deg}T_2}\mathrm{STr}(T_2T_1) \tag{23.95}$$

---

sec:Berezinian

23.7 The Berezinian of a linear transformation

Let $\mathcal{A}$ be supercommutative and consider the free $\mathcal{A}$ module $\mathcal{A}^{p|q}$.

While the determinant of a matrix can be defined for any matrix, the the *superdeterminant* or *Berezinian* can only be defined for elements of $GL(\mathcal{A}^{p|q})$. If $X \in \mathrm{End}(\mathcal{A}^{p|q})$ is even and invertible then the value of $\mathrm{Ber}(X)$ lies in the subalgebra of invertible elements of $\mathcal{A}^0$, which we can consider to be $GL(\mathcal{A}^{1|0})$.

The conditions which characterize the Berezinian are :

1. When X can be written as an exponential of a matrix $X = \exp Y$, with $Y \in \mathrm{End}(\mathcal{A}^{p|q})$ we must have

$$\mathrm{Ber}(X) = \mathrm{Ber}(\exp Y) := \exp(\mathrm{STr}\ Y) \tag{23.96}$$ eq:sdettr

2. The Berezinian is multiplicative:

$$\mathrm{Ber}(X_1X_2) = \mathrm{Ber}(X_1)\mathrm{Ber}(X_2) \tag{23.97}$$ eq:MultBer

Note that the two properties (23.96) and (23.97) are compatible thanks to the Baker-Campbell-Hausdorff formula. (See Chapter 8 below.)

♣Actually, need to show the supertrace condition is well-defined. ♣

We can use these properties to give a formula for the Berezinian of a matrix once we know the key result

Lemma Let $\mathcal{A}$ be supercommutative and $\pi : \mathcal{A} \to \mathcal{A}_{\mathrm{red}} = \mathcal{A}/\mathcal{I}^{\mathrm{odd}}$. This defines a map

$$\pi : \mathrm{End}(\mathcal{A}^{p|q}) \to \mathrm{End}(\mathcal{A}^{p|q}_{\mathrm{red}}) \tag{23.98}$$

by applying π to the matrix elements. Then: the supermatrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{23.99}$$

1. Is invertible iff $\pi(A)$ and $\pi(D)$ are invertible.
2. Is in the image of the exponential map iff $\pi(A)$ and $\pi(D)$ are.

Proof: The proof follows closely that of the invertibility lemma above. Note that since X is even then $\pi(X)$ is block diagonal so $\pi(X)$ is invertible iff $\pi(A)$ and $\pi(D)$ are invertible. If $\pi(X)$ is invertible then, since π is onto there is a $Y \in \mathrm{End}(\mathcal{A}^{p|q})$ so that $1 = \pi(X)\pi(Y) = \pi(XY)$. Then $XY = 1 - Z$ for some Z such that $\pi(Z) = 0$. All the matrix elements of Z are nilpotent so there is an N so that $Z^{N+1} = 0$. Then $Y(1 + Z + \cdots + Z^N)$ is the inverse of X.

By the same token if $\pi(X) = \exp(\alpha)$ then we can lift α to $\tilde{\alpha} \in \mathrm{End}(\mathcal{A}^{p|q})$ and $\pi(X\exp[-\tilde{\alpha}]) = 1$ so $X\exp[-\tilde{\alpha}] = 1 - Z$ where Z is nilpotent. Therefore $1 - Z = \exp[z]$ is well-defined because the series for $\log(1 - Z)$ terminates. ♠.

Now – assuming that a Berezinian function actually exists – we can give a formula for what it must be. From the first condition we know that when $\pi(A), \pi(D)$ are in the image of the exponential map then

$$\mathrm{Ber}\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} = \frac{\det A}{\det D} \tag{23.100}$$

Note that the entries of A and D are all even so in writing out the usual definition of determinant there is no issue of ordering. Together with multiplicativity and the fact that the exponential map is onto for $GL(n, \kappa)$ this determines the formula for all block diagonal matrices.

Moreover, upper triangular matrices are in the image of the exponential once again because all the matrix elements of B are nilpotent so that

$$\log\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} = -\sum_{k=1}^{\infty} \frac{(-1)^k}{k} \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}^k \tag{23.101}$$

terminates and is a well-defined series. Moreover it is clear that it has supertrace = 0, and therefore

$$\mathrm{Ber}(\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix}) = 1 \tag{23.102}$$

$$\mathrm{Ber}(\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}) = 1 \tag{23.103}$$

Now for general invertible X we can write

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & BD^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ D^{-1}C & 1 \end{pmatrix} \tag{23.104}$$

and hence multiplicativity implies

$$\mathrm{Ber}(X) = \frac{\det(A - BD^{-1}C)}{\det D} = \frac{\det A}{\det D}\det(1 - A^{-1}BD^{-1}C) \tag{23.105}$$

eq:SDETFORM

There is one more point to settle here. We have shown that the two properties (23.96) and (23.97) uniquely determine the Berezinian of a supermatrix and even determine a formula for it. But, strictly speaking, we have not yet shown that the Berezinian actually exists, because we have not shown that the formula (23.105) is indeed multiplicative. A brute force approach to verifying this would be very complicated.

A better way to proceed is the following. We want to prove that $\mathrm{Ber}(gh) = \mathrm{Ber}(g)\mathrm{Ber}(h)$ for any two group elements g, h. Let us consider the subgroups G^+, G^0, G^- of upper triangular, block diagonal, and lower triangular matrices:

$$G^+ = \{X : X = \begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix}\} \tag{23.106}$$

$$G^0 = \{X : X = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}\} \tag{23.107}$$

$$G^- = \{X : X = \begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}\} \tag{23.108}$$

Any group element can be written as $g = g^+g^0g^-$ where $g^{\pm,0} \in G^{\pm,0}$. So now we need to consider $gh = g^+g^0g^-h^+h^0h^-$. It would be very complicated to rewrite this again as a Gauss decomposition. On the other hand, it is completely straightforward to check multiplicativity of the formula for products of the form g^+k, g^0k, kg^0, and kg^- for any k and $g^{\pm,0} \in G^{\pm,0}$. For example, to check multiplicativity for g^+k we write

$$\begin{pmatrix} 1 & B' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A + B'C & B + B'D \\ C & D \end{pmatrix} \tag{23.109}$$

and now we simply note that

$$\det((A + B'C) - (B + B'D)D^{-1}C) = \det(A - BD^{-1}C) \tag{23.110}$$

The other cases are similarly straightforward. (Check them!!) Therefore, to check multiplicativity we need only check that multiplicativity for products of the form $g^- h^+$. Therefore we need only show

$$\mathrm{Ber}\begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix} = 1 \tag{23.111}$$

eq:check-ber-m

because

$$\begin{pmatrix} 1 & 0 \\ C & 1 \end{pmatrix}\begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix} \tag{23.112}$$

This is not completely obvious from (23.105). Nevertheless, it is easily shown: Note that the matrix in (23.111) is in the image of the exponential map. But it is trivial from the relation to the supertrace that if $g = \exp(Y)$ then $\mathrm{Ber}(g^{-1}) = (\mathrm{Ber}(g))^{-1}$. On the other hand,

$$\begin{pmatrix} 1 & B \\ C & 1+CB \end{pmatrix}^{-1} = \begin{pmatrix} 1+BC & -B \\ -C & 1 \end{pmatrix} \tag{23.113}$$

and applying the formula (23.105) to the RHS trivially gives one.

Finally, we remark that from the multiplicativity property it follows that $\mathrm{Ber}(g^{-1}Xg) = \mathrm{Ber}(X)$ and hence the Berezinian is invariant under change of basis. Therefore, it is intrinsically defined for an even invertible map $T \in \mathrm{End}(\mathcal{A}^{p|q})$.

Exercise
Let $\alpha, \beta \in \mathbb{C}^*$. Evaluate

$$\mathrm{Ber}\begin{pmatrix} \alpha & \theta_1 \\ \theta_2 & \beta \end{pmatrix} \tag{23.114}$$

Using both of the expressions above.

Exercise
Show that two alternative formulae for the Berezinian are

$$\mathrm{Ber}(X) = \frac{\det A}{\det(D - CA^{-1}B)} = \frac{\det A}{\det D}(\det(1 - D^{-1}CA^{-1}B))^{-1} \tag{23.115}$$

eq:SDET-ALT

Note that the equality of (23.105) and (23.115) follows because

$$\det(1 - A^{-1}BD^{-1}C) = (\det(1 - D^{-1}CA^{-1}B))^{-1} \tag{23.116}$$

and this in turn is easily established because both matrices are $1 + Nilpotent$ and hence in the image of the exponential map and since $A^{-1}B$ and $D^{-1}C$ are both odd we have

$$\mathrm{STr}(A^{-1}BD^{-1}C)^k = -\mathrm{STr}(D^{-1}CA^{-1}B)^k \tag{23.117}$$

This in turn gives another easy proof of multiplicativity, once one has reduced it to (23.111), which follows immediately from (23.115).

♣This is a better proof. Put this in the text and make the other an exercise. ♣

Exercise

Let $\alpha, \beta \in \mathbb{C}^*$. Evaluate

$$\mathrm{Ber}\begin{pmatrix} \alpha & \theta_1 \\ \theta_2 & \beta \end{pmatrix} \tag{23.118}$$

Using both of the expressions above.

23.8 Bilinear forms

Bilinear forms on super vector spaces $\mathfrak{b} : V \otimes V \to \kappa$ are defined as in the ungraded case: $\mathfrak{b}$ is a bilinear morphism of supervector spaces.

It follows that $\mathfrak{b}(x, y) = 0$ if x and y in V are homogeneous and of opposite parity.

We can identify the set of bilinear forms with $V^\vee \otimes V^\vee$. We can then apply super-symmetrization and super-antisymmetriziation.

Thus, symmetric bilinear forms have a very important extra sign.

$$\mathfrak{b}(x, y) = (-1)^{|x||y|}\mathfrak{b}(y, x) \tag{23.119}$$

This means $\mathfrak{b}$ is symmetric when restricted to $V^0 \times V^0$ and antisymmetric when restricted to $V^1 \times V^1$.

Similarly, antisymmetric bilinear forms have the reverse situation:

$$\mathfrak{b}(x, y) = (-1)^{1+|x||y|}\mathfrak{b}(y, x) \tag{23.120}$$

This means $\mathfrak{b}$ is anti-symmetric when restricted to $V^0 \times V^0$ and symmetric when restricted to $V^1 \times V^1$.

The definition of a *nondegenerate* form is the same as before. A form is nondegenerate iff its restrictions to $V^0 \times V^0$ and $V^1 \times V^1$ are nondegenerate. Therefore, applying the canonical forms of symmetric and antisymmetric matrices we discussed in Section §20 above we know that if $\kappa = \mathbb{R}$ and $\mathfrak{b}$ is a nondegenerate $\mathbb{Z}_2$-graded symmetric form then there is a basis where its matrix looks like

$$Q = \begin{pmatrix} 1_r & 0 & 0 & 0 \\ 0 & -1_s & 0 & 0 \\ 0 & 0 & 0 & -1_m \\ 0 & 0 & 1_m & 0 \end{pmatrix} \tag{23.121}$$ eq:OSPQ

The automorphisms of the bilinear form are the even invertible morphisms $g : V \to V$ such that

$$\mathfrak{b}(gv, gw) = \mathfrak{b}(v, w) \tag{23.122}$$

for all $v, w \in V$. This is just the group $O(r,s) \times Sp(2m;\mathbb{R})$.

As with the general linear group, to define more interesting automorphism groups of a bilinear form we need to consider bilinear forms on the free modules $\mathcal{A}^{p|q}$ over a supercommutative superalgebra $\mathcal{A}$.

Definition: Let $\mathcal{A}$ be a superalgebra. A bilinear form on a (left) $\mathcal{A}$-module M is a morphism of supervector spaces

$$M \otimes M \to \mathcal{A} \tag{23.123}$$

such that

$$\mathfrak{b}(am, m') = a\mathfrak{b}(m, m') \qquad \mathfrak{b}(m, am') = (-1)^{|a||m|}a\mathfrak{b}(m, m') \tag{23.124}$$

Now, if we apply this to the free module $\mathcal{A}^{p|q}$ over a supercommutative algebra $\mathcal{A}$ (which can be considered to be either a left or right $\mathcal{A}$-module then we have simply

$$a\mathfrak{b}(m, m') = \mathfrak{b}(am, m') \qquad \mathfrak{b}(ma, m') = \mathfrak{b}(m, am') \qquad \mathfrak{b}(m, m'a) = \mathfrak{b}(m, m')a \tag{23.125}$$

The automorphism group of $\mathfrak{b}$ is the group of $g \in \mathrm{End}(\mathcal{A}^{p|q})$ which are even and invertible and for which

$$\mathfrak{b}(gm, gm') = \mathfrak{b}(m, m') \tag{23.126}$$

for all $m, m' \in \mathcal{A}^{p|q}$. In the case where $\mathfrak{b}$ is a nondegenerate $\mathbb{Z}_2$-graded symmetric on $\mathcal{A}^{p|q}$ we define an interesting generalization of both the orthogonal and symplectic groups which plays an important role in physics. We could denote it $OSp_{\mathcal{A}}(\mathcal{A}^{p|q})$.

Using ideas from Chapter 8, discussed for the super-case in Chapter 12, we can use this discussion to derive the superLie algebra in the case where we specialize to a Grassmann algebra $\mathcal{A} = \mathbb{R}[\theta^1, \dots, \theta^{q'}]/I$ so that $\mathcal{A}_{\mathrm{red}} = \mathbb{R}$. If $\mathfrak{b}$ is nondegenerate then on the reduced module $\mathbb{R}^{p|q}$ (where q and q' are not related) it can be brought to the form (23.121), so $p = r + s$ and $q = 2m$. Writing $g(t) = e^{tA}$ and differentiating wrt t at $t = 0$ we derive the Lie algebra of the supergroup. It is the subset of $\mathrm{End}(\mathcal{A}^{p|q})$ such that

$$\mathfrak{b}(Am, m') + \mathfrak{b}(m, Am') = 0 \tag{23.127}$$

If we finally reduce this equation mod nilpotents we obtain an equation on $\mathrm{End}(\mathbb{R}^{p|q})$. ♣notation conflict m ♣
That defines a Lie algebra over $\mathbb{R}$ which is usually denoted $\mathrm{osp}(r, s|2m;\mathbb{R})$.

23.9 Star-structures and super-Hilbert spaces

There are at least *three* notions of a real structure on a complex superalgebra which one will encounter in the literature:

1. It is a $\mathbb{C}$-antilinear involutive *automorphism* $a \mapsto a^{\bigstar}$. Hence $\deg(a^{\bigstar}) = \deg(a)$ and $(ab)^{\bigstar} = a^{\bigstar}b^{\bigstar}$.

2. It is a $\mathbb{C}$-antilinear involutive *anti-automorphism*. Thus $\deg(a^*) = \deg(a)$ but

$$(ab)^* = (-1)^{|a||b|}b^*a^* \tag{23.128}$$

3. It is a $\mathbb{C}$-antilinear involutive *anti-automorphism*. Thus $\deg(a^\star) = \deg(a)$ but

$$(ab)^\star = b^\star a^\star \tag{23.129}$$

If $\mathcal{A}$ is a supercommutative complex superalgebra then structures 1 and 2 coincide: $a \to a^\bigstar$ is the same as $a \to a^*$. See remarks below for the relation of 2 and 3.

Definition A sesquilinear form h on a complex supervector space $\mathcal{H}$ is a map $h : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ such that

1. It is even, so that $h(v, w) = 0$ if v and w have opposite parity
2. It is $\mathbb{C}$-linear in the second variable and $\mathbb{C}$-antilinear in the first variable
3. An *Hermitian form* on a supervector space is a sesquilinear form which moreover satisfies the symmetry property:

$$(h(v, w))^* = (-1)^{|v||w|} h(w, v) \tag{23.130}$$

4. If in addition for all nonzero $v \in \mathcal{H}^0$

$$h(v, v) > 0 \tag{23.131}$$

while for all nonzero $v \in \mathcal{H}^1$

$$i^{-1} h(v, v) > 0, \tag{23.132}$$

then $\mathcal{H}$ endowed with the form h is a *super-Hilbert space.*

For bounded operators we define the adjoint of a homogeneous linear operator $T : \mathcal{H} \to \mathcal{H}$ by

$$h(T^* v, w) = (-1)^{|T||v|} h(v, Tw) \tag{23.133}$$

The spectral theorem is essentially the same as in the ungraded case with one strange modification. For even Hermitian operators the spectrum is real. However, for odd Hermitian operators the point spectrum sits in a real subspace of the complex plane which is *not* the real line! If T is odd then an eigenvector v such that $Tv = \lambda v$ must have even and odd parts $v = v_e + v_o$. Then the eigenvalue equation becomes

$$\begin{aligned} Tv_e &= \lambda v_o \\ Tv_o &= \lambda v_e \end{aligned} \tag{23.134}$$

Now the usual proof that the point spectrum is real is modified to:

$$\begin{aligned} \lambda^* h(v_o, v_o) &= h(\lambda v_o, v_o) = h(Tv_e, v_o) = h(v_e, Tv_o) = \lambda h(v_e, v_e) \\ \lambda^* h(v_e, v_e) &= h(\lambda v_e, v_e) = h(Tv_o, v_e) = -h(v_o, Tv_e) = -\lambda h(v_o, v_o) \end{aligned} \tag{23.135}$$

These two equations have the same content: Since $v \neq 0$ and we are in a superHilbert space it must be that

$$h(v_e, v_e) = i^{-1} h(v_o, v_o) > 0 \tag{23.136}$$

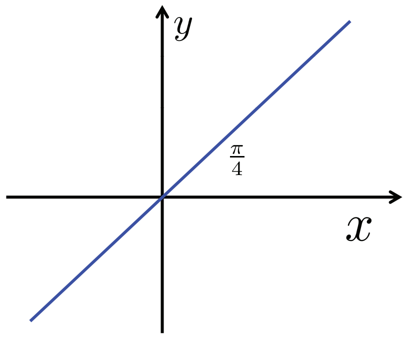

Figure 16: When the Koszul rule is consistently implemented odd super-Hermitian operators have a spectrum which lies along the line through the origin which runs through $1+i$. fig:SUPERHERMI

and therefore the phase of λ is determined. It lies on the line passing through $e^{i\pi/4} = (1+i)/\sqrt{2}$ in the complex plane, as shown in Figure 16

Example: An example of a natural super-Hilbert space is the Hilbert space of L^2-spinors on an even-dimensional manifold with $(-1)^F$ given by the chirality operator. An odd self-adjoint operator which will have nonhomogeneous eigenvectors is the Dirac operator on an even-dimensional manifold. One usually thinks of the eigenvalues as real for this operator and that is indeed the case if we use the star-structure $\star$, number 3 above. See the exercise below.

Remarks

1. In general star-structures 2 and 3 above are actually closely related. Indeed, given a structure $a \to a^*$ of type 2 we can define a structure of type 3 by defining either

$$a^\star = \begin{cases} a^* & |a| = 0 \\ ia^* & |a| = 1 \end{cases} \tag{23.137}$$

or

$$a^\star = \begin{cases} a^* & |a| = 0 \\ -ia^* & |a| = 1 \end{cases} \tag{23.138}$$ eq:relstar

It is very unfortunate that in most of the physics literature the definition of a star structure is that used in item 3 above. For example a typical formula used in manipulations in superspace is

$$\overline{\theta_1\theta_2} = \bar\theta_2\bar\theta_1 \tag{23.139}$$

and the fermion kinetic energy

$$\int dt i\bar\psi\frac{d}{dt}\psi \tag{23.140}$$

is only "real" with the third convention. The rationale for this convention, especially for fermionic fields, is that they will eventually be quantized as operators on a Hilbert space. Physicists find it much more natural to have a standard Hilbert space structure, even if it is $\mathbb{Z}_2$-graded. On the other hand, item 2 implements the Koszul rule consistently and makes the analogy to classical physics as close as possible. So, for example, the fermionic kinetic term is

$$\int dt \bar{\psi} \frac{d}{dt} \psi \tag{23.141}$$

and is "manifestly real."

Fortunately, as we have just noted one convention can be converted to the other, but the difference will, for example, show up as factors of i in comparing supersymmetric Lagrangians in the different conventions, as the above examples show.

Exercise

a.) Show that a super-Hermitian form h on a super-Hilbert space can be used to define an ordinary Hilbert space structure on $\mathcal{H}$ by taking $\mathcal{H}^0 \perp \mathcal{H}^1$ and taking

$$\begin{aligned} (v,w) &:= h(v,w) \qquad & v,w \in \mathcal{H}^0 \\ (v,w) &:= i^{-1} h(v,w) \qquad & v,w \in \mathcal{H}^1 \end{aligned} \tag{23.142}$$

eq:UnbradedHer

b.) Show that if T is an operator on a super-Hilbert-space then the super-adjoint T^* and the ordinary adjoint $T^\dagger$, the latter defined with respect to (23.142), are related by

$$T^* = \begin{cases} T^\dagger & |T| = 0 \\ iT^\dagger & |T| = 1 \end{cases} \tag{23.143}$$

c.) Show that $T \to T^\dagger$ is a star-structure on the superalgebra of operators on superspace which is of type 3 above.

d.) Show that if T is an odd self-adjoint operator with respect to $*$ then $e^{-i\pi/4} T$ is an odd self-adjoint operator with respect to $\dagger$. In particular $e^{-i\pi/4} T$ has a point spectrum in the real line.

e.) More generally, show that if a is odd and real with respect to $*$ then $e^{-i\pi/4} a$ is real with respect to $\star$ defined by (23.138).

23.9.1 SuperUnitary Group

Let us return to a general finite-dimensional Hermitian form on a complex supervector-space. Restricted to V^0 it can be brought to the form $\mathrm{Diag}\{+1_r, -1_s\}$ while restricted to the odd subspace it can be brought to the form $\mathrm{Diag}\{+1_t, -1_u\}$. The automorphism group of (V,h) is therefore $U(r,s) \times U(t,u)$. If we consider instead a free module $\mathcal{A}^{n_e,n_o}$

over a supercommutative algebra $\mathcal{A}$ (where $\mathcal{A}$ is a vector space over $\kappa = \mathbb{C}$) we can still define an Hermitian form $h : \mathcal{A}^{n_e,n_o} \times \mathcal{A}^{n_e,n_o} \to \mathcal{A}$. If $\mathcal{A}_{\rm red} = \mathbb{C}$ and h is of the above type with $n_e = r + s$ and $n_o = t + u$ then the automorphism group of h is $U_{\mathcal{A}}(r, s|p, q)$. If we derive the Lie algebra and reduce modulo nilpotents we then obtain the super Lie algebra $\mathrm{u}(r, s|p, q; \mathbb{C})$ which is the subset of $\mathrm{End}(\mathbb{C}^{n_e|n_o})$

$$h(Av, v') + (-1)^{|A||v|}h(v, Av') = 0 \tag{23.144}$$

i.e $\mathrm{u}(r, s|p, q)$ is the real super Lie algebra of super-anti-unitary operators. We will say much more about this in Chapter 12.

Exercise *Fixed points*
Let $\eta_{r,s}$, and $\eta_{t,u}$ be diagonal matrices...
Show that $\mathrm{u}(r, s|p, q)$ the the set of fixed points of the antilinear involution ...
♣ FILL IN ♣

23.10 Functions on superspace and supermanifolds

23.10.1 Philosophical background

Sometimes one can approach the subjects of topology and geometry through algebra and analysis. Two famous examples of this are

1. Algebraic geometry: The geometry of vanishing loci of systems of polynomials can be translated into purely algebraic questions about commutative algebra.
2. Gelfand's Theorem on commutative C^*-algebras

We now explain a little bit about Gelfand's theorem:

There is a 1-1 correspondence between Hausdorff topological spaces and commutative C^*-algebras.

If X is a Hausdorff topological space then we can form $C_0(X)$, which is the space of all *continuous* complex valued functions $f : X \to \mathbb{C}$ which "vanish at infinity." What this means is that for all $\epsilon > 0$ the set of $x \in X$ so that $|f(x)| \geq \epsilon$ is a compact set. This is a C^*-algebra with involution $f \mapsto f^*$ where $f^*(x) := (f(x))^*$ and the norm is

$$\parallel f \parallel := \sup_{x\in X}|f(x)| \tag{23.145}$$

Then there is a 1-1 correspondence between isomorphism classes of topological spaces and isomorphism classes of commutative C^*-algebras.

The way one goes from a commutative C^* algebra $\mathcal{A}$ to a topological space is that one defines $\Delta(\mathcal{A})$ to be the set of - any of

a.) The C^*-algebra morphisms $\chi : \mathcal{A} \to \mathbb{C}$.

b.) The maximal ideals

c.) The irreducible representations.

For a commutative C^* algebra the three notions are equivalent. The space $\Delta(\mathcal{A})$ carries a natural topology since there is a norm on linear maps $\mathcal{A} \to \mathbb{C}$ of Banach spaces. It turns out that $\Delta(\mathcal{A})$ is a Hausdorff space. Gelfand's theorem then says that $C_0(\Delta(\mathcal{A}))$ is in fact isomorphic as a C^* algebra to $\mathcal{A}$, while $\Delta(C_0(X))$ is homeomorphic as a topological space to X.

The correspondence is very natural if we interpret a,b,c in terms of $\mathcal{A} = C_0(X)$. then, given a point $x \in X$ we have

a.) The morphism $\chi_x : f \mapsto f(x)$

b.) The maximal ideal $\mathfrak{m}_x = \ker(\chi_x) = \{f \in C_0(X) | f(x) = 0\}$

c.) The representations $\rho_x(f) = f(x)$.

Remarks:

1. As a simple and very important example of how geometry is transformed into algebra, a continuous map of topological spaces $f : X \to Y$ is in 1-1 correspondence with a C^*-algebra homomorphism $\varphi_f : C_0(Y) \to C_0(X)$, given by the "pullback": $\varphi_f(g) := g \circ f$. We are going to exploit this idea over and over in the following pages.

2. Notice that if X is just a finite disjoint union of n points then $C_0(X) \cong \mathbb{C} \oplus \cdots \oplus \mathbb{C}$ is finite-dimensional, and if X has positive dimension then $C_0(X)$ is infinite-dimensional.

3. Now, on a vector space like $\mathbb{R}^n$ the symmetric algebra $S^\bullet(\mathbb{R}^n)$ can be interpreted as the algebra of polynomial functions on $\mathbb{R}^n$. These are dense (Stone-Weierstrass theorem) in the algebra of continuous functions $C_0(\mathbb{R}^n)$.

Algebraic geometry enhances the scope of geometry by considering more general commutative rings. Perhaps the simplest example is the "thickened point." (The technical term is "connected zero dimensional nonreduced scheme of length 2.") The "thickened point" is defined by saying that its algebra of functions is the commutative algebra $D = \mathbb{C}[\eta]/(\eta^2)$. As a vector space it is $\mathbb{C} \oplus \mathbb{C}\eta$ and the algebra structure is defined by $\eta^2 = 0$. This is an example of an algebra of functions on a "thickened point." How do we study the "thickened point" ? Let us look at maps of this "point" into affine spaces such at $\mathbb{C}^n$. Using the philosophy motivated by the mathematics mentioned above a "map from the thickened point into $\mathbb{C}^n$" is the same thing as an algebra homomorphism $\varphi : \mathbb{C}[t^1, \ldots, t^n] \to D$ where we recall that $\mathbb{C}[t^1, \ldots, t^n]$ is just the algebra of polynomials. Such a homomorphism must be of the form

$$P \mapsto \varphi(P) = \varphi_1(P) + \varphi_2(P)\eta \tag{23.146}$$

Since this is an algebra homomorphism φ_1, φ_2 are linear functionals on the algebra of polynomials and moreover $\varphi(PQ) = \varphi(P)\varphi(Q)$ implies that

$$\begin{aligned}\varphi_1(PQ) &= \varphi_1(P)\varphi_1(Q)\\ \varphi_2(PQ) &= \varphi_1(P)\varphi_2(Q) + \varphi_1(Q)\varphi_2(P)\end{aligned} \tag{23.147}$$

The first equation tells us that φ_1 is just evaluation of the polynomial at a point $\vec{t_0} = (t_0^1, \dots, t_0^n)$. The second is then precisely the algebraic way to define a vector field at that point! Thus

$$\varphi_2(P) := \sum_{i=1}^{n} v^i \frac{\partial}{\partial t^i} P|_{\vec{t_0}} \tag{23.148}$$

Therefore, *for every "map from the thickened point to $\mathbb{C}^n$" we associate the data of a point $\vec{t_0} \in \mathbb{C}^n$ and a vector field at that point.* This amply justifies the term "thickened point." An obvious generalization is to consider instead the commutative algebra $D_N = \mathbb{C}[\eta]/(\eta^N)$. These give different "thickened points." Technically, this is the ring of functions on a "connected zero dimensional nonreduced scheme of length N" which we will just call a "thickened point of order $N-1$." In this case a map into $\mathbb{C}^n$ is characterized by a suitable linear functional on the set of Taylor expansion coefficients of f around some point $\vec{t_0}$. ♣?? check this last sentence. ♣

Noncommutative "geometry" develops this idea by starting with *any* (C^*-) algebra $\mathcal{A}$, not necessarily commutative, and interpreting $\mathcal{A}$ as the "algebra of functions" on some mythical "noncommutative space" and proceeding to study geometrical questions translated into algebraic questions about $\mathcal{A}$. So, for example, if $\mathcal{A} = M_n(\mathbb{C})$ is the algebra of $n \times n$ matrices then there is only one maximal ideal, and the only algebra homomorphism to $M_n(\mathbb{C}) \to \mathbb{C}$ is $\phi(M) = 0$, so $M_n(\mathbb{C})$ is the set of functions on a "space" which is a kind of "nonabelian thickened point."

23.10.2 The model superspace $\mathcal{R}^{p|q}$

Supergeometry is a generalization of algebraic geometry and a specialization of general noncommutative geometry where the algebras we use are supercommutative.

A superpoint has a real or complex algebra of functions given by a Grassmann algebra $\mathrm{Grass}[\theta^1, \dots, \theta^q]$, depending on whether κ is $\mathbb{R}$ or $\mathbb{C}$, respectively.

Note that this algebra is just $S^\bullet(\mathbb{R}^{0|q})$ where we use the $\mathbb{Z}_2$-graded symmetric algebra. We can say there are q odd coordinates and we are considering polynomial functions of these coordinates.

This motivates the definition of the superspace $\mathcal{R}^{p|q}$ as the "space" whose super-algebra of polynomial functions is

$$S^\bullet(\mathbb{R}^{p|q}) \tag{23.149}$$

where we take the $\mathbb{Z}_2$-graded symmetric algebra. As a $\mathbb{Z}_2$-graded vector space this algebra is just

$$S^\bullet(\mathbb{R}^{p|0}) \widehat{\otimes} \Lambda^\bullet(\mathbb{R}^q) \tag{23.150}$$

(Here we view $\Lambda^\bullet(\mathbb{R}^q) = \Lambda^{\mathrm{ev}}(\mathbb{R}^q) \oplus \Lambda^{\mathrm{odd}}(\mathbb{R}^q)$ as a $\mathbb{Z}_2$-graded vector space.)

Given a choice of basis $\{\theta^1, \ldots, \theta^q\}$ of $\mathbb{R}^{0|q}$ a general super-polynomial on $\mathcal{R}^{p|q}$ can be written as

$$\Phi = \phi_0 + \phi_i \theta^i + \frac{1}{2!}\phi_{i_1 i_2}\theta^{i_1}\theta^{i_2} + \cdots + \frac{1}{n!}\phi_{i_1\cdots i_q}\theta^{i_1}\cdots\theta^{i_q} \tag{23.151}$$

where the $\phi_{i_1,\ldots,i_m}$ are *even*, totally antisymmetric in $i_1, \ldots, i_m$, and for fixed $i_1, \ldots, i_m$ are polynomials on $\mathbb{R}^{p|0}$.

Given an *ordered* basis $\{\theta^1, \ldots, \theta^q\}$ of $\mathbb{R}^{0|q}$ we can furthermore introduce a *multi-index* $I = (i_1 < i_2 < \cdots < i_k)$ where we say I has length k, and we write $|I| = k$. We denote $I = 0$ for the empty multi-index. Then we can write

$$\Phi = \sum_I \phi_I \theta^I \tag{23.152}$$

where the ϕ_I are ordinary even polynomials on $\mathbb{R}^p$.

Similarly, we can extend these expressions by allowing the ϕ_I to be *smooth* (not just polynomial) functions on $\mathcal{R}^p$ and then we define the algebra of smooth functions on $\mathcal{R}^{p|q}$ to be the commutative superalgebra

$$\begin{aligned} \mathcal{C}^\infty(\mathcal{R}^{p|q}) &:= \mathcal{C}^\infty(\mathbb{R}^p)\widehat{\otimes} S^\bullet(\mathbb{R}^{0|q}) \\ &= \mathcal{C}^\infty(\mathbb{R}^p)[\theta^1, \ldots, \theta^q]/(\theta^i\theta^j + \theta^j\theta^i = 0) \end{aligned} \tag{23.153}$$

eq:DefineRpq

An element of $\mathcal{C}^\infty(\mathcal{R}^{p|q})$ was called by Wess and Zumino a "superfield." The idea is that we have a "function" of (x, θ) where $x = (x^1, \ldots, x^p)$ and "Taylor expansion" in the odd coordinates must terminate so

♣Actually, parities can get a little confusing if we consider things like $\mathcal{W}_\alpha$ in SYM. ♣

$$\Phi(x, \theta) = \sum_I \phi_I(x)\theta^I \tag{23.154}$$

where $\phi_I(x)$ are smooth functions of x. Trying to take this too literally can lead to confusing questions. *What is a "point" in a superspace? Can we localize a function at the coordinate* $\frac{1}{2}\theta$ *instead of* θ*? What is the "value" of a function at a point on superspace?* One way of answering such questions is explained in the remark below about the "functor of points," but often physicists just proceed with well-defined rules and get well-defined results at the end, and leave the philosophy to the mathematicians.

23.10.3 Superdomains

The official mathematical definition of a supermanifold, given below, makes use of the idea of sheaves. To motivate that we first define a *superdomain* $\mathcal{U}^{p|q}$ to be a "space" whose superalgebra of functions is analogous to (23.155):

$$\begin{aligned} \mathcal{C}^\infty(\mathcal{U}^{p|q}) &:= \mathcal{C}^\infty(U)\widehat{\otimes} S^\bullet(\mathbb{R}^{0|q}) \\ &= \mathcal{C}^\infty(U)[\theta^1, \ldots, \theta^q]/(\theta^i\theta^j + \theta^j\theta^i = 0) \end{aligned} \tag{23.155}$$

eq:DefineRpq

where $U \subset \mathbb{R}^p$ is any open set. Denote $\mathcal{O}^{p|q}(U) := \mathcal{C}^\infty(\mathcal{U}^{p|q})$. When $V \subset U$ there is a well-defined morphism of superalgebras

$$r_{U\to V} : \mathcal{O}^{p|q}(U) \to \mathcal{O}^{p|q}(V), \tag{23.156}$$

given simply by restricting from U to V the smooth functions ϕ_I on U. These morphisms are called, naturally enough, the *restriction morphisms*. Note that they are actually morphsims of superalgebras. It is often useful to denote

$$r_{U\to V}(\Phi) := \Phi|_V. \tag{23.157}$$

The restriction morphisms satisfy the following list of fairly evident properties:

1. $r_{U\to U} = Identity$.
2. $(\Phi|_V)|_W = \Phi|_W$ when $W \subset V \subset U$.
3. Suppose $U = \cup_\alpha U_\alpha$ is a union of open sets and $\Phi_1, \Phi_2 \in \mathcal{O}^{p|q}(U)$. Then if $(\Phi_1)|_{U_\alpha} = (\Phi_2)|_{U_\alpha}$ for all α we can conclude that $\Phi_1 = \Phi_2$.
4. Suppose $U = \cup_\alpha U_\alpha$ is a union of open sets and Φ_α is a collection of elements $\Phi_\alpha \in \mathcal{O}^{p|q}(U_\alpha)$. Then if, for all α, β,

$$(\Phi_\alpha)|_{U_\alpha\cap U_\beta} = (\Phi_\beta)|_{U_\alpha\cap U_\beta} \tag{23.158}$$

   then we can conclude that there exists a $\Phi \in \mathcal{O}^{p|q}(U)$ such that $(\Phi)|_{U_\alpha} = \Phi_\alpha$.

23.10.4 A few words about sheaves

The properties we listed above for functions on superdomains are actually a special case of a defining list of axioms for a more general notion of a *sheaf*. Since this has been appearing in recent years in physics we briefly describe the more general concept.

Definition

a.) A *presheaf* $\mathcal{F}$ on a topological space X is an association of a set $\mathcal{F}(U)$ to every open set [41] $U \subset X$ such that there is a coherent system of restriction maps. That is, whenever $V \subset U$ there is a map $r_{U\to V}$ so that

$$r_{U,U} = \text{Identity} \qquad r_{V\to W}\circ r_{U\to V} = r_{U\to W} \qquad W\subset V\subset U \tag{23.159}$$

b.) Elements $f \in \mathcal{F}(U)$ are called *sections over* U. If $V \subset U$ we denote $r_{U\to V}(f) := f|_V$.

c.) A *sheaf* $\mathcal{F}$ on a topological space is a presheaf which moreover satisfies the two additional properties when $U = \cup_\alpha U_\alpha$ is a union of open sets:

1. If $f, g \in \mathcal{F}(U)$ and for all α, $f|_{U_\alpha} = g|_{U_\alpha}$, then $f = g$.

2. If for all α we are given $f_\alpha \in \mathcal{F}(U_\alpha)$ such that for all α, β we have $(f_\alpha)|_{U_\alpha\cap U_\beta} = (f_\beta)|_{U_\alpha\cap U_\beta}$ then there exists an $f \in \mathcal{F}(U)$ so that $f|_{U_\alpha} = f_\alpha$.

A good example is the sheaf of $\mathbb{C}^\infty$ functions on a smooth manifold. Another good example is the sheaf of holomorphic functions (The extension axiom is analytic continuation.)

[41] Technically $\emptyset$ is an open set. We should define $\mathcal{F}(\emptyset)$ to be the set with one element. If we have a sheaf of groups, then it should be the trivial group. etc.

In many common examples the sets $\mathcal{F}(U)$ in a sheaf carry some algebraic structure. Thus, we assume there is some "target" category $\mathcal{C}$ so that $\mathcal{F}(U)$ are objects in that category. So, if $\mathcal{C} = \mathbf{GROUP}$ then $\mathcal{F}(U)$ is a group for every open set, and we have a sheaf of groups; if $\mathcal{C} = \mathbf{ALGEBRA}$ is the category of algebras over κ then we have a sheaf of algebras, etc. ♣More examples, and some pictures ♣

If $\mathcal{F}$ and $\mathcal{G}$ are two sheaves on a topological spaces X then a *morphism of sheaves* is the data of a morphism (in the category $\mathcal{C}$ where the sheaf is valued) $\phi(U) : \mathcal{F}(U) \to \mathcal{G}(U)$ for every open set $U \subset X$. Note this must be a morphism in whatever target category $\mathcal{C}$ we are using. Thus, if we have a sheaf of groups, then for each U, $\phi(U)$ is a group homomorphism, and if we have a sheaf of algebras $\phi(U)$ is an algebra homomorphism, etc. Moreover, the morphisms must be compatible with restriction maps:

$$\phi(V) \circ r^{\mathcal{F}}_{U\to V} = r^{\mathcal{G}}_{U\to V} \circ \phi(U) \tag{23.160}$$

♣Write out as a commutative diagram. ♣

We can also speak of morphisms between sheaves on different topological spaces X and Y. To do this, we first define the *direct image sheaf.* Given a continuous map $\varphi : X \to Y$ and a sheaf $\mathcal{F}$ on X we can define a new sheaf $\varphi_*(\mathcal{F})$ on Y. By definition if U is an open set of Y then

$$\varphi_*(\mathcal{F})(U) := \mathcal{F}(\varphi^{-1}(U)) \tag{23.161}$$

Now a morphism of sheaves $(X, \mathcal{F}) \to (Y, \mathcal{G})$ can be defined to be a continuous map $\varphi : X \to Y$ together with a morphism of sheaves over Y, $\phi : \mathcal{G} \to \varphi_*(\mathcal{F})$. ♣Explain contravariance here. ♣

Finally, we will need the notion of the *stalk* of a sheaf at a point $\wp$. If you are familiar with directed limits then we can just write

$$\mathcal{F}(\wp) := \lim_{U:p\in U} \mathcal{F}(U) \tag{23.162}$$

What this means is that we look at sections in infinitesimal neighborhoods of $\wp$ and identify these sections if they agree. To be precise, we consider $\amalg_{U:p\in U}\mathcal{F}(U)$ and identify $f_1 \in \mathcal{F}(U_1)$ with $f_2 \in \mathcal{F}(U_2)$ if there is an open set $p \in W \subset U_1 \cap U_2$ such that $f_1|_W = f_2|_W$. So, with this equivalence relation

$$\mathcal{F}(\wp) = \amalg_{U:p\in U}\mathcal{F}(U)/\sim \tag{23.163}$$

Example: Consider the sheaf of holomorphic functions on $\mathbb{C}$. Then the stalk at z_0 can be identified with the set of formal power series expansions at z_0. For the sheaf of $\mathcal{C}^\infty$ functions on a manifold the stalk at $\wp$ is just $\mathbb{R}$.

Finally, with these definitions we can say that the superdomains $\mathcal{U}^{p|q}$ defined above describe a *sheaf of Grassmann algebras* $\mathcal{O}^{p|q}$ with value on an open set $U \subset \mathbb{R}^p$ given by the Grassmann algebra

$$\mathcal{O}^{p|q}(U) = \mathcal{C}^\infty(\mathcal{U}^{p|q}) = \mathcal{C}^\infty(U)\widehat{\otimes}S^\bullet(\mathbb{R}^{0|q}) \tag{23.164}$$

A *super-change of coordinates* is an invertible morphism of the sheaf $\mathcal{O}^{p|q}$ with itself. Concretely it will be given by an expression like

$$\begin{aligned} \tilde{t}^a &= f^a(t^1, \dots, t^p|\theta^1, \dots, \theta^q) & a &= 1, \dots, p \\ \tilde{\theta}^i &= \psi^i(t^1, \dots, t^p|\theta^1, \dots, \theta^q) & i &= 1, \dots, q \end{aligned} \tag{23.165}$$

where f^a are even elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ and ψ^i are odd elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$, respectively.

Exercise *Alternative definition of a presheaf*

Given a topological space X define a natural category whose objects are open sets $U \subset X$ and whose morphisms are inclusions of open sets.

Given any category $\mathcal{C}$ show that a presheaf with values in $\mathcal{C}$ can be defined as a contravariant functor from the category of open sets in X to $\mathcal{C}$.

Exercise

Show that the stalk of $\mathcal{O}^{p|q}$ at a point $\wp \in \mathbb{R}^p$ is the finite-dimensional Grassmann algebra $\Lambda^*(\mathbb{R}^q)$ over $\mathbb{R}$.

23.10.5 Definition of supermanifolds

One definition of a supermanifold is the following:

Definition A supermanifold M of dimension $(p|q)$ is an ordinary manifold M_{red} with a sheaf $\mathcal{F}$ of Grassmann algebras which is locally equivalent to the supermanifold $\mathcal{R}^{p|q}$. That is, near any $p \in M_{\text{red}}$ there is a neighborhood $p \in U$ so that the restriction of the sheaf to U is equivalent (isomorphism of sheaves) to a superdomain $\mathcal{U}^{p|q}$.

In this definition M_{red} is the called the "reduced space" or the "body" of the supermanifold. The sheaf $\mathcal{F}$ of Grassmann algebras has a subsheaf $\mathcal{I}^{\text{odd}}$ generated by the odd elements and the quotient sheaf $\mathcal{F}/\mathcal{I}^{\text{odd}}$ is the sheaf of $\mathcal{C}^\infty$ functions of the reduced space M_{red}.

There is a second (equivalent) definition of supermanifolds which strives to make a close parallel to the definition of manifolds in terms of atlases of charts.

We choose a manifold M of dimension p and define a *super-chart* to be a pair $(\mathcal{U}^{p|q}, c)$ where $c : U \to M$ is a homeomorphism. Then a supermanifold will be a collection of supercharts $(\mathcal{U}_\alpha^{p|q}, c_\alpha)$ so that if $c_\alpha(U_\alpha) \cap c_\beta(U_\beta) = \hat{U}_{\alpha\beta}$ is nonempty then there is a change of coordinates between coordinates $(t^1_\alpha, \dots, t^p_\alpha | \theta^1_\alpha, \dots, \theta^q_\alpha)$ on $\mathcal{O}^{p|q}(c_\alpha^{-1}(\hat{U}_{\alpha\beta}))$ and $(t^1_\beta, \dots, t^p_\beta | \theta^1_\beta, \dots, \theta^q_\beta)$ on $\mathcal{O}^{p|q}(c_\beta^{-1}(\hat{U}_{\alpha\beta}))$ given by a collection of functions:

$$\begin{aligned} t^a_\alpha &= f^a_{\alpha\beta}(t^1_\beta, \dots, t^p_\beta | \theta^1_\beta, \dots, \theta^q_\beta) & \qquad a = 1, \dots, p \\ \theta^i_\alpha &= \psi^i_{\alpha\beta}(t^1_\beta, \dots, t^p_\beta | \theta^1_\beta, \dots, \theta^q_\beta) & \qquad i = 1, \dots, q \end{aligned} \tag{23.166}$$

where $f^a_{\alpha\beta}$ are even elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ and $\psi^i_{\alpha\beta}$ are odd elements of $\mathcal{C}^\infty(\mathbb{R}^{p|q})$, respectively. These maps need to be invertible, in an appropriate sense, and they need to satisfy a version of the cocycle identity when there are nonempty triple overlaps $c_\alpha(U_\alpha) \cap c_\beta(U_\beta) \cap c_\gamma(U_\gamma)$. For a more careful discussion see Chapter III of Leites.

Example A good example of a nontrivial supermanifold is super-complex projective space $\mathbb{C}P^{m|n}$. The reduced manifold is just $\mathbb{C}P^m$. Recall that $\mathbb{C}P^m$ is the space of complex lines in $\mathbb{C}^{m+1}$ and can be thought of as the set of nonzero points $(X^0, \dots, X^m) \in \mathbb{C}^{m+1}$ modulo the scaling action $X^A \to \lambda X^A$. We denote the equivalence class by $[X^0 : \cdots : X^m]$. Informally, we can define $\mathbb{C}P^{m|n}$ as the "set of points" $(X^0, \dots, X^m | \theta^1, \dots, \theta^n) \in \mathbb{C}^{m+1|n}$ with $(X^0, \dots, X^m) \neq 0$ again with identification by scaling

$$(X^0, \dots, X^m | \theta^1, \dots, \theta^n) \sim \lambda (X^0, \dots, X^m | \theta^1, \dots, \theta^n) \tag{23.167}$$

To make proper sense of this we could define a standard superatlas by choosing the usual atlas on $\mathbb{C}P^m$ defined by the nonvanishing of one of the homogeneous coordinates

$$U_\alpha := \{[X^0 : X^1 : \cdots : X^m] | X^\alpha \neq 0\} \qquad \alpha = 0, \dots, p \tag{23.168}$$

so local coordinates are given by $t^A_\alpha := X^A / X^\alpha$. (Note that $t^\alpha_\alpha = 1$ is not a coordinate, so coordinates are given by $A = 1, \dots, m$ omitting α.) Then the supermanifold has

$$\mathcal{F}(U_\alpha) = \mathcal{C}^\infty(U_\alpha)[\theta^1_\alpha, \dots, \theta^n_\alpha]/(\theta^i_\alpha \theta^j_\alpha + \theta^j_\alpha \theta^i_\alpha = 0) \tag{23.169}$$

and on $U_{\alpha\beta}$ we have the change of coordinates

$$\begin{aligned} t^A_\alpha &= \frac{X^\beta}{X^\alpha} t^A_\beta & \qquad A = 0, \dots, m \\ \theta^i_\alpha &= \frac{X^\beta}{X^\alpha} \theta^i_\beta & \qquad i = 1, \dots, n \end{aligned} \tag{23.170}$$

(You can put $A = \alpha$ to learn that $\frac{X^\beta}{X^\alpha} = 1/t^\alpha_\beta$ to get the honest formula.)

Now we can go on to produce more nontrivial examples of supermanifolds by choosing homogeneous even polynomials $P(X^A | \theta^i)$ and dividing the sheaf by the ideal generated by these.

For example, see [42] for an interesting discussion of the sub-supermanifold of $\mathbb{C}P^{2|2}$ defined by

$$X_1^2 + X_2^2 + X_3^2 + \theta_1 \theta_2 = 0 \tag{23.171}$$

Remarks

[42] E. Witten, "Notes on Supermanifolds and Integration," arXiv:1209.2199, Section 2.3.1

1. In general there is no reality condition put on the odd generators θ^i. Therefore, it is natural to consider a supermanifold $\mathbb{R}^{p|*q}$ where the ring of functions can be expanded as above but only ϕ_0 is real, and all the other ϕ_I with $|I| > 0$ are complex polynomials. Gluing these together gives a *cs-supermanifold.*

2. *More philosophy: The functor of points.* There is a way of speaking about "points of a supermanifold" which is a generalization of a standard concept in algebraic geometry. We first give the background in algebraic geometry. For simplicity we just work with a ground field $\kappa = \mathbb{C}$. There is a generalization of algebraic varieties known as "schemes." Again we characterize them locally by their algebras of "polynomial functions," but now we are allowed to introduce nilpotents as in the "thickened point" example discussed above. To characterize the "points" on a scheme X we probe it by taking an arbitrary scheme S and consider the set of all morphisms of schemes $\mathrm{Hom}(S, X)$. The set $\mathrm{Hom}(S, X)$ is, roughly speaking, just the homomorphisms from the algebra of functions on X to the algebra of functions on S. In this context the set $\mathrm{Hom}(S, X)$ is called *the set of S-points of X*. Now, the map $X \mapsto \mathrm{Hom}(S, X)$ is (contravariantly) "functorial in S." This means that if $f : S \to S'$ is a morphism of schemes then there is a natural morphism of sets $F_X(f) : \mathrm{Hom}(S', X) \to \mathrm{Hom}(S, X)$. Therefore, given a scheme X, there is a functor F_X from the category of all schemes [43] to the category of sets, $F_X : \mathbf{SCHEME}^{\mathrm{opp}} \to \mathbf{SET}$ defined on objects by

$$F_X : S \mapsto \mathrm{Hom}(S, X) \tag{23.172}$$

This functor is called the *functor of points.* If we let the "probe scheme" S be a point then its algebra of functions is just $\mathbb{C}$ and $F_X(S) = \mathrm{Hom}(S, X)$ is the set of algebra homomorphisms from functions on X to $\mathbb{C}$. That is, indeed the set of points of the underlying topological space X_{red} of X. More generally, if S is an ordinary algebraic manifold then we should regard $F_X(S)$ as a set of points in X parametrized by S. Of course, we could probe the scheme structure of X more deeply by using a more refined probe, such as a nonreduced point of order N described above. In fact, if we use too few probe schemes S we might miss structure of X, therefore mathematicians use the functor from *all* schemes S to sets. Now, a key theorem justifying this approach (known as the *Yoneda theorem*) states that:

Two schemes X and X' are isomorphic as schemes iff there is a natural transformation between the functors F_X and $F_{X'}$.

Now, we can apply all these ideas to supermanifolds with little change. If M is a supermanifold, and $S = \mathbb{R}^{0|0}$ then the set of S-points of M is precisely the set of points of the underlying manifold M_{red}.

We will not go very deeply into supermanifold theory but we do need a notion of vector fields:

[43] actually, the opposite category $\mathbf{SCHEME}^{\mathrm{opp}}$

23.10.6 Supervector fields and super-differential forms

In the ordinary theory of manifolds the space of vector fields on the manifold is in 1-1 correspondence with the derivations of the algebra of functions. The latter concept makes sense for supermanifolds, provided we take $\mathbb{Z}_2$-graded derivations, and is taken to define the super-vector-fields on a supermanifold.

For $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ the space of derivations is a left supermodule for $\mathcal{C}^\infty(\mathbb{R}^{p|q})$ generated by

$$\frac{\partial}{\partial t^1}, \dots, \frac{\partial}{\partial t^p}, \frac{\partial}{\partial \theta^1}, \dots, \frac{\partial}{\partial \theta^q} \tag{23.173}$$

where we have to say whether the odd derivatives act from the left or the right. We will take them to act from the left so, for example, if $q = 2$ then

$$\begin{aligned} \frac{\partial}{\partial \theta^1}\Phi &= \phi_1 + \phi_{12}\theta^2 \\ \frac{\partial}{\partial \theta^2}\Phi &= \phi_2 - \phi_{12}\theta^1 \end{aligned} \tag{23.174}$$

In general, for any $i \in \{1, \dots, q\}$ we can write always expand Φ in the form $\Phi = \sum_{I:i\notin I}(\phi_I\theta^I + \phi_{i,I}\theta^a\theta^I)$ and then

$$\frac{\partial}{\partial \theta^i}\Phi = \sum_{I:i\notin I} \phi_{i,I}\theta^I \tag{23.175}$$

A simple, but important lemma says that the $\mathcal{O}^{p|q}$ module of derivations is free and of dimension $(p|q)$.

Defining differential forms turns out to be surprisingly subtle. The problems are related to how one defines a grading of expressions like $d\theta^i$ and, related to this, how one defines the exterior derivative. There are two (different) ways to do this.

One way to proceed is to consider the "stalk" of the tangent sheaf $T\mathcal{R}^{p|q}$ at a point $\wp$. This is a module for the real Grassmann algebra Grass$[\theta^1, \dots, \theta^q]$. (That is, the coefficients of $\frac{\partial}{\partial t^a}$ and $\frac{\partial}{\partial \theta^i}$ are functions of θ^i but not of the t^a, because we restricted to a point $\wp$.) The dual module is denoted $\Omega^1\mathcal{R}^{p|q}(\wp)$. There is an even pairing

$$T\mathcal{R}^{p|q}(\wp) \otimes \Omega^1\mathcal{R}^{p|q}(\wp) \to \mathcal{O}^{p|q}(\wp) \tag{23.176}$$

The pairing is denoted $\langle v, \omega\rangle$ and if Φ_1, Φ_2 are superfunctions then

$$\langle \Phi_1 v, \Phi_2\omega\rangle = (-1)^{|v||\Phi_2|}\Phi_1\Phi_2\langle v, \omega\rangle \tag{23.177}$$

If we have a system of coordinates $(t^a|\theta^i)$ then $\Omega^1\mathcal{R}^{p|q}(\wp)$ is a free module of rank $(p|q)$ generated by symbols dt^a and $d\theta^i$. Thus,

$$\langle \frac{\partial}{\partial t^a}, dt^b\rangle = \delta_a^b \qquad \langle \frac{\partial}{\partial \theta^i}, d\theta^j\rangle = \delta_j^i \tag{23.178}$$

and so forth.

Now, to define the differential forms at the point $\wp$ we take the exterior algebra of $\Omega^1\mathcal{R}^{p|q}(\wp)$ to define

$$\Omega^\bullet\mathcal{R}^{p|q}(\wp) := \Lambda^\bullet(\Omega^1\mathcal{R}^{p|q}(\wp)) \tag{23.179}$$

where - very importantly - we are using the $\mathbb{Z}_2$-graded antisymmetrization to define $\Lambda^\bullet$. Thus, the generators dt^a are *anti-commuting* (as usual) while the generators $d\theta^i$ are *commuting*. The stalks can be used to define a sheaf $\Omega^\bullet\mathcal{R}^{p|q}$ and the general section in $\Omega^\bullet\mathcal{R}^{p|q}(U)$ is an expression:

$$\sum_{k=0}^{p}\sum_{\ell=0}^{\infty}\omega_{a_1,\dots,a_k;i_1,\dots,i_\ell}dt^{a_1}\cdots dt^{a_k}d\theta^{i_1}\cdots d\theta^{i_\ell} \tag{23.180}$$

eq:GenSuperFor

where $\omega_{a_1,\dots,a_k;i_1,\dots,i_\ell}$ are elements of $\mathcal{O}^{p|q}(U)$ which are totally antisymmetric in the $a_1,\dots,a_k$ and totally symmetric in the $i_1,\dots,i_\ell$.

If we consider dt^a to be odd and $d\theta^i$ to be even then expressions such as (23.180) can be multiplied. Then, finally, we can define an exterior derivative by saying that $d:\mathcal{O}^{p|q}\to\Omega^1\mathcal{R}^{p|q}$ takes $d: t^a\mapsto dt^a$ and $d:\theta^i\mapsto d\theta^i$ and then we impose the super-Leibniz rule

$$d(\omega_1\omega_2) = d\omega_1\omega_2 + (-1)^{|\omega_1|}\omega_1 d\omega_2 \tag{23.181}$$

It is still true that $d^2=0$ and we have the Super-Poincaré lemma: If $d\omega=0$ in $\Omega^\bullet\mathcal{R}^{p|q}$ then $\omega = d\eta$.

Remarks

1. We are following the conventions of Witten's paper cited below. For a nice interpretation of the differential forms on a supermanifold in terms of Clifford and Heisenberg modules see Section 3.2. Note that with the above conventions

$$d(\theta^1\theta^2) = \theta^2 d\theta^1 - \theta^1 d\theta^2 \tag{23.182}$$

2. However, there is another, equally valid discussion which is the one taken in Deligne-Morgan. The superderivations define a sheaf of super-modules for the sheaf $\mathcal{O}^{p|q}$ and it is denoted by $T\mathcal{R}^{p|q}$. Then the cotangent sheaf, denoted $\Omega^1\mathcal{R}^{p,q}$ is the dual module with an even pairing: ♣?♣

$$T\mathcal{R}^{p|q}\otimes\Omega^1\mathcal{R}^{p|q}\to\mathcal{O}^{p|q} \tag{23.183}$$

The pairing is denoted $\langle v,\omega\rangle$ and if Φ_1,Φ_2 are superfunctions then

$$\langle\Phi_1 v,\Phi_2\omega\rangle = (-1)^{|v||\Phi_2|}\Phi_1\Phi_2\langle v,\omega\rangle \tag{23.184}$$

It we have a system of coordinates $(t|\theta)$ then Ω^1 freely generated as an $\mathcal{O}^{p|q}$-module by dt^a and $d\theta^i$.

Now we define a differential $d:\mathcal{O}^{p|q}\to\Omega^1\mathcal{R}^{p|q}$ to be an *even* morphism of sheaves of super-vector spaces by

$$\langle v, df\rangle := v(f) \tag{23.185}$$

In particular this implies

$$d(\theta^1\theta^2) = -\theta^2 d\theta^1 + \theta^1 d\theta^2 \tag{23.186}$$

The issue here is that $\Omega^\bullet\mathcal{R}^{p|q}$ is really *bigraded* by the group $\mathbb{Z}\oplus\mathbb{Z}_2$. It has "cohomological degree" in $\mathbb{Z}$ coming from the degree of the differential form in addition to "parity." In general, given vector spaces which are $\mathbb{Z}$-graded *and also* $\mathbb{Z}_2$-graded, so that $V = V^0 \oplus V^1$ as a super-vector-space, and V^0 and V^1 are also $\mathbb{Z}$-graded, then there are *two* conventions for defining the commutativity morphism:

$$c_{V,W} : V \otimes W \to W \otimes V \tag{23.187}$$

1. We have the modified Koszul rule: $c_{V,W} : v\otimes w \mapsto (-1)^{(|v|+\deg(v))(|w|+\deg(w))} w\otimes v$, where $\deg(v), \deg(w)$ refer to the integer grading.
2. We have the modified Koszul rule: $c_{V,W} : v\otimes w \mapsto (-1)^{|v||w|+\deg(v)\deg(w)} w\otimes v$.

In convention 1 we have simply taken a homomorphism of the $\mathbb{Z}\oplus\mathbb{Z}_2$ grading to $\mathbb{Z}_2$. In our notes we have adopted the first convention in making d odd. This makes $d\theta^i$ even because we sum the degree of d (which is one) with the degree of θ^i (which is one modulo two) to get zero, modulo two. In the second convention it would still be true that the $d\theta^i$ commute with the $d\theta^j$, but for a different reason. Convention 2 is adopted in Deligne-Morgan, for reasons explained on p.62. ♣Should discuss this more. What are the $\mathbb{Z}_2$-valued bilinear forms on $\mathbb{Z}\oplus\mathbb{Z}_2$? ♣

Exercise

Suppose that M is a manifold and TM is its tangent bundle. Let ΠTM be the supermanifold where the Grassmann algebra is the Grassmann algebra of the sections of TM.

Show that the $\mathbb{C}^\infty$ functions on TM can be identified with the DeRham complex of the bosonic manifold $\Omega^\bullet$ and, under this correspondence, write d as a supervector field on $\mathbb{C}^\infty(\Pi TM)$.

This observation is often used in applications of supersymmetric field theory to topological invariants.

Exercise

a.) Show that the graded commutator of super-derivations is a superderivation.

b.) Consider the odd vector fields $D = \frac{\partial}{\partial\theta} + \theta\frac{\partial}{\partial t}$ and $Q = \frac{\partial}{\partial\theta} - \theta\frac{\partial}{\partial t}$ on $\mathcal{R}^{1|1}$. Compute $[D,D]$, $[Q,Q]$, and $[Q,D]$.

ANOTHER EXERCISE WITH MORE THETAS.

23.11 Integration over a superdomain

In this section we will say something about how to define an integral of superfunctions on the supermanifold $\mathcal{R}^{p|q}$.

As motivation we again take inspiration from the theory of commutative C^*-algebras. A beautiful theorem - the Riesz-Markov theorem - says that if $\mathcal{A}$ is a commutative C^*-algebra and $\Lambda : \mathcal{A} \to \mathbb{C}$ is a linear functional then there is a (complex-valued) measure $d\mu$ on $X = \Delta(\mathcal{A})$ so that this linear functional is just

$$\Lambda(f) = \int_X f d\mu \tag{23.188}$$

(Recall that $f \in \mathcal{A}$ is canonically a function on X, so the expression on the RHS makes sense.)

So, we will view an integral over $\mathcal{R}^{p|q}$ as a linear functional

$$\Lambda : \mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R} \tag{23.189}$$

To guide us in our discussion there are three criteria we want from our integral:

1. We want integration by parts (Stokes' theorem) to be valid.
2. We want the Fubini theorem to be valid.
3. We want the definition to reduce to the usual Riemannian integration when $q = 0$.

Let us begin with $p = 0$, the fermionic point. For brevity denote $\mathcal{O}^q := \mathcal{O}^{0|q}(pt) = S^\bullet(\mathbb{R}^{0|q})$. The space of linear functionals

$$\mathcal{D}^q := \mathrm{Hom}(\mathcal{O}^q, \mathbb{R}) \tag{23.190}$$

is a real supervector space of dimension $(2^{q-1}|2^{q-1})$. Indeed, given an *ordered* basis $\theta^1, \dots, \theta^q$ for $\mathbb{R}^{0|q}$ there is a canonical dual basis δ_I for $\mathcal{D}^q$ defined by $\delta_I(\theta^J) = \delta_I{}^J$ where I, J are multi-indices.

On the other hand, $\mathcal{D}^q$ is also a right $\mathcal{O}^q$-module since if Λ is a linear functional and $g \in \mathcal{O}^q$ we can define

$$(\Lambda \cdot g)(f) := \Lambda(gf) \tag{23.191}$$

It is important to distinguish $\mathcal{D}^q$ as a vector space over $\mathbb{R}$ from $\mathcal{D}^q$ as a module over the supercommutative superalgebra $\mathcal{O}^q$. In the latter case, $\mathcal{D}^q$ is free and of dimension $(1|0)$ or $(0|1)$.

For example, suppose $q = 2$ and we choose an ordered basis $\{\theta^1, \theta^2\}$ for $\mathbb{R}^{0|2}$. Then let $\delta = \delta_{12}$. Then

$$\begin{aligned} \delta &= \delta_{12} \\ \delta \cdot \theta^1 &= \delta_2 \\ \delta \cdot \theta^2 &= -\delta_1 \\ \delta \cdot \theta^1\theta^2 &= \delta_0 \end{aligned} \tag{23.192}$$

In general, given an ordered basis, $\delta = \delta_I$, where I is the multi-index $I = 12\dots q$, is a basis vector for $\mathcal{D}^q$ as an $\mathcal{O}^q$-module: Indeed, right-multiplication by elements of $\mathcal{O}^q$ gives a vector space basis over $\mathbb{R}$ as follows:

$$\begin{aligned} \delta &= \delta_{1\dots q} \\ (\delta \cdot \theta^i) &= (-1)^{i-1}\delta_{1\cdots\hat{\imath}\cdots q} \\ (\delta \cdot \theta^i\theta^j) &= \pm\delta_{1\cdots\hat{\imath}\cdots\hat{\jmath}\cdots q} \\ \vdots \qquad & \qquad \vdots \end{aligned} \tag{23.193}$$

Moreover, since the scalar 1 is even δ has degree $q$mod2 and hence, as a right $\mathcal{O}^q$-module, $\mathcal{D}^q$ has parity $q$mod2, so it is a free module of type $(\mathcal{O}^q)^{1|0}$ or $(\mathcal{O}^q)^{0|1}$, depending on whether q is even or odd, respectively. Of course, if N_q is a nonzero real number then $N_q\delta$ is also a perfectly good generator of $\mathcal{D}^q$ as a $\mathcal{O}^q$-module.

Now we claim that, given an ordered basis $\{\theta^1, \dots, \theta^q\}$ for $\mathbb{R}^{0|q}$ there is a canonical generator for $\mathcal{D}^q$ which we will denote by

$$\Lambda_q = \int [d\theta^1 \cdots d\theta^q] \tag{23.194}$$ eq:CanonicalMe

The notation is apt because this functional certainly satisfies the integration-by-parts property:

$$\int [d\theta^1 \cdots d\theta^q] \frac{\partial f}{\partial \theta^i} = 0 \tag{23.195}$$

for any i and and f. Thus, criterion 1 above is automatic in our approach.

However, the integration-by-parts property is satisfied by *any* generator of $\mathcal{D}^q$ as an $\mathcal{O}^q$-module, that is, it is satisfied by any nonzero multiple of Λ_q. How should we normalize Λ_q? We can answer this question by appealing to criterion 2. That is, we require an analog of the Fubini theorem. There is a canonical isomorphism $\mathcal{R}^{0|q_1} \times \mathcal{R}^{0|q_2} \cong \mathcal{R}^{0|q_1+q_2}$, that is there are canonical isomorphism $\mathcal{O}^{q_1}\widehat{\otimes}\mathcal{O}^{q_2} \cong \mathcal{O}^{q_1+q_2}$ (simply given by multiplying the polynomials) and hence canonical isomorphisms

$$\mathcal{D}^{q_1}\widehat{\otimes}\mathcal{D}^{q_2} \cong \mathcal{D}^{q_1+q_2} \tag{23.196}$$

given by

$$(\ell_1\widehat{\otimes}\ell_2)(f_1\widehat{\otimes}f_2) = (-1)^{|\ell_2||f_1|}\ell_1(f_1)\ell_2(f_2) \tag{23.197}$$

Now we require that our canonical integrals Λ_q satisfy

$$\begin{tikzcd} \mathcal{D}^{q_1} \otimes \mathcal{D}^{q_2} \arrow[rr, "\cong"] \arrow[dr, "\Lambda_{q_1}\otimes\Lambda_{q_2}"'] & & \mathcal{D}^{q_1+q_2} \arrow[dl, "\Lambda_{q_1+q_2}"] \\ & \mathbb{R} & \end{tikzcd} \tag{23.198}$$ eq:Compat-Lamb

Let $\Lambda_q(\theta^1 \cdots \theta^q) := N_q$. Then (23.198) implies that

$$\begin{aligned} N_{q_1+q_2} &= \Lambda_{q_1+q_2}(\theta^1 \cdots \theta^{q_1+q_2}) \\ &= (\Lambda_{q_1} \otimes \Lambda_{q_2})(\theta^1 \cdots \theta^{q_1} \otimes \theta^{q_1+1} \cdots \theta^{q_1+q_2}) \\ &= (-1)^{q_1q_2} N_{q_1} N_{q_2} \end{aligned} \tag{23.199}$$

The general solution to the equation $N_{q_1+q_2} = (-1)^{q_1 q_2} N_{q_1} N_{q_2}$ is

$$N_q = (-1)^{\frac{1}{2}q(q-1)}(N_1)^q \tag{23.200}$$

So this reduces the question to $q = 1$.

It is customary and natural to normalize the integral so that

$$\int [d\theta]\theta = 1 \tag{23.201}$$

That is, $N_1 = 1$. With this normalization, the Berezin integral on $\mathcal{R}^{0|1}$ is the functional:

$$\int [d\theta](a + b\theta) = b \tag{23.202}$$

Now, for $q > 1$, noticing that $\theta^1 \cdots \theta^q = (-1)^{\frac{1}{2}q(q-1)}\theta^q \cdots \theta^1$ we have shown that demanding that the integral satisfy the "Fubini theorem" (as interpreted above) normalizes the canonical measure so that

$$\boxed{\int [d\theta^1 \cdots d\theta^q]\theta^q \cdots \theta^1 = +1} \tag{23.203}$$

Now let us consider $p > 0$. Criterion 3 above tells us that we don't want our integrals to be literally all linear functionals $\mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R}$. For example, that would include distributions in the bosonic variables. So we have the official definition

Definition: A *density* on the superspace $\mathcal{R}^{p|q}$ is a linear functional $\mathcal{O}^{p|q}(\mathbb{R}^p) \to \mathbb{R}$ of the form

$$\Phi = \sum_I \phi_I \theta^I \mapsto \sum_I \int_{\mathbb{R}^p} [dt^1 \cdots dt^p] d_I(t)\phi_I(t) \tag{23.204}$$

where $[dt^1 \cdots dt^p]$ is the standard Riemann measure associated with a coordinate system $(t^1, \dots, t^p)$ for $\mathbb{R}^p$ and $d_I(t)$ are some collection of smooth functions. We denote the space of densities on $\mathcal{R}^{p|q}$ by $\mathcal{D}^{p|q}$.

When $p = 0$ this reduces to our previous description, and $\mathcal{D}^{0|q} = \mathcal{D}^q$. Now, analogous to the previous discussion, $\mathcal{D}^{p|q}$ is once again a $\mathcal{O}^{p|q}(\mathbb{R}^p)$-module of rank $(1|0)$ or $(0|1)$, depending on whether $q = 0 \mathrm{mod} 2$ or $q = 1 \mathrm{mod} 2$, respectively. Once again, given an ordered coordinate system $(t^1, \dots, t^p|\theta^1, \dots, \theta^q)$ for $\mathcal{R}^{p|q}$ we have a canonically normalized density which we denote

$$\int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q] \tag{23.205}$$

defined by

$$\int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q]\Phi := \int_{\mathbb{R}^p} [dt^1 \cdots dt^p] (\int [d\theta^1 \cdots d\theta^q]\Phi) \tag{23.206}$$

where $[dt^1 \cdots dt^p]$ is the Riemannian measure. Thus, we first integrate over the odd coordinates and then over the reduced bosonic coordinates.

♣The conventions on indices t^i, θ^a are opposite to those of the previous section. Use consistent conventions. ♣

Finally, let us give the change of variables formula. Suppose $\mu : \mathcal{R}^{p|q} \to \mathcal{R}^{p|q}$ is an invertible morphism. Then we can define new "coordinates:

$$\begin{aligned} \tilde{t}^i &= \mu^*(t^i) \qquad\qquad i = 1, \dots, p \\ \tilde{\theta}^a &= \mu^*(\theta^a) \qquad\qquad a = 1, \dots, q \end{aligned} \tag{23.207}$$

Then, again because $\mathcal{D}^{p|q}$ is one-dimensional as an $\mathcal{O}^{p|q}(\mathbb{R}^p)$-module, we know that there is an even invertible element of $\mathcal{O}^{p|q}(\mathbb{R}^p)$ so that

$$\int [d\tilde{t}^1 \cdots d\tilde{t}^p | d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \Phi(\tilde{t}|\tilde{\theta}) = \int [dt^1 \cdots dt^p | d\theta^1 \cdots d\theta^q] j(\mu) \mu^* \Phi \tag{23.208}$$

where $\mu^*\Phi$ is a function of $(t|\theta)$ given by $\mu^*\Phi = \Phi(\tilde{t}(t|\theta)|\tilde{\theta}(t|\theta))$.

Some special cases will make the general formula clear:

1. If $\mu : \mathbb{R}^p \to \mathbb{R}^p$ is an ordinary diffeomorphism then it can be lifted to a superdiffeomorphism just by setting $\mu^*(\theta^a) = \theta^a$ and $\mu^*(t^i) = \tilde{t}^i$. Then the standard change-of-variables result says that

$$\begin{aligned} [d\tilde{t}^1 \cdots d\tilde{t}^p] &= [dt^1 \cdots dt^p] \cdot \left|\det \frac{\partial \tilde{t}^i}{\partial t^j}\right| \\ &= [dt^1 \cdots dt^p] \cdot \mathrm{or}(\mu) \cdot \det \frac{\partial \tilde{t}^i}{\partial t^j} \end{aligned} \tag{23.209}$$

where $\mathrm{or}(\mu) = +1$ if μ is orientation preserving and $\mathrm{or}(\mu) = -1$ if it is orientation reversing.

2. On the other hand, if $\mu^*(t^i) = t^i$ and $\mu^*(\theta^a) = D^a{}_b \theta^b$ then

$$\begin{aligned} \tilde{\theta}^q \cdots \tilde{\theta}^1 &= D^q{}_{b_1} \cdots D^1{}_{b_q} \theta^{b_1} \cdots \theta^{b_q} \\ &= \left(\sum_{\sigma \in S_q} \epsilon(\sigma) D^q{}_{\sigma(q)} \cdots D^1{}_{\sigma(1)} \right) \theta^q \cdots \theta^1 \\ &= \det(D^a{}_b) \theta^q \cdots \theta^1 \end{aligned} \tag{23.210}$$

and therefore

$$[d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] = [d\theta^1 \cdots d\theta^q] (\det(D^a{}_b))^{-1} \tag{23.211}$$

For the general formula we consider the Jacobian

$$\mathrm{Jac}(\mu) = \begin{pmatrix} \frac{\partial \tilde{t}^i}{\partial t^j} & \frac{\partial \tilde{t}^i}{\partial \theta^b} \\ \frac{\partial \tilde{\theta}^a}{\partial t^j} & \frac{\partial \tilde{\theta}^a}{\partial \theta^b} \end{pmatrix} \tag{23.212}$$

which we regard as an element of $\mathrm{End}(\Omega^1 \mathcal{R}^{p|q})$. (Recall that $\Omega^1 \mathcal{R}^{p|q}(U)$ is a free module of rank $(p|q)$ over $\mathcal{O}^{p|q}(U)$.) The formula is

$$j(\mu) = \mathrm{or}(\mu_{\mathrm{red}}) \mathrm{Ber}(\mathrm{Jac}(\mu)) \tag{23.213}$$

♣Needs a full proof.
♣

Example: Consider $\mathcal{R}^{1|2}$. Let $\Phi(\tilde{t}|\tilde{\theta}) = h(\tilde{t})$. Change variables by $\tilde{t} = t + \theta^1\theta^2$ and $\tilde{\theta}^i = \theta^i$. Then

$$\begin{aligned} 0 &= \int [d\tilde{t}|d\tilde{\theta}] h(\tilde{t}) \\ &= \int [dt|d\theta] \left(h(t) + h'(t)\theta^1\theta^2\right) \\ &= -\int dt \frac{\partial h}{\partial t} \end{aligned} \tag{23.214}$$

eq:SuperIntExp

Note that this identity relies on the validity of integration by parts.

Remarks

1. Note well that $d\theta^a$ are commutative objects but $[d\theta^a]$ are anti-commutative objects in the sense that $[d\theta^1 d\theta^2] = -[d\theta^2 d\theta^1]$, and so on.
2. The possible failure of boundary terms to vanish in examples like 23.214 leads to important subtleties in string perturbation theory. On a supermanifold it might not be possible to say, globally, which even variables are "purely bosonic," that is, "free of nilpotents." This is related to the issue of whether the supermanifold is "split" or not. For recent discussions of these problems see Witten, arXiv:1304.2832, 1209.5461.

23.12 Gaussian Integrals

23.12.1 Reminder on bosonic Gaussian integrals

Let Q_{ij} be a symmetric quadratic form with positive definite real part on $\mathbb{R}^p$. Then the Gaussian integral over $\mathcal{R}^{p|0}$ is

$$(2\pi)^{-p/2} \int [dt^1 \cdots dt^p] \exp[-\frac{1}{2} t^i Q_{ij} t^j] = \frac{1}{(\det Q)^{1/2}} \tag{23.215}$$

where we choose the sign of the square root so that $(\det Q)^{1/2}$ is in the positive half-plane, i.e., we choose the principal branch of the logarithm.

One could analytically continue in Q from this result.

23.12.2 Gaussian integral on a fermionic point: Pfaffians

Let us now consider the Gaussian integral over a fermionic point $\mathcal{R}^{0|q}$.

Let A_{ij} be a $q \times q$ antisymmetric matrix. Consider the Gaussian integral:

$$\int [d\theta^1 \cdots d\theta^q] \exp[\frac{1}{2} \theta^a A_{ab} \theta^b] \tag{23.216}$$

Our first observation is that if q is odd then this integral must vanish! To see this, we recall that we can always skew-diagonalize A:

$$SAS^{tr} = \begin{pmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \lambda_2 \\ -\lambda_2 & 0 \end{pmatrix} \oplus \cdots \tag{23.217}$$

By the change-of-variable formula if we change coordinates $\tilde{\theta}^a = S^a{}_b \theta^b$ then the integral is

$$\begin{aligned}\int [d\theta^1 \cdots d\theta^q] \exp[\frac{1}{2}\theta^a A_{ab}\theta^b] &= \det S \int [d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \exp[\frac{1}{2}\tilde{\theta}^a \tilde{A}_{ab}\tilde{\theta}^b] \\ &= \det S \int [d\tilde{\theta}^1 \cdots d\tilde{\theta}^q] \exp[\lambda_1 \tilde{\theta}^1 \tilde{\theta}^2 + \lambda_2 \tilde{\theta}^3 \tilde{\theta}^4 + \cdots]\end{aligned} \tag{23.218}$$

Now, if q is odd, then in this expression $\tilde{\theta}^q$ does not appear in the exponential. Therefore the integral has a factor of $\int [d\tilde{\theta}^q] = 0$. This is a very simple example of how an "*unpaired fermion zeromode*" leads to the zero of a fermionic Gaussian integral. See remarks below.

Suppose instead that $q = 2m$ is even. Then the integral can be evaluated in terms of the skew eigenvalues as

$$\det S \prod_{i=1}^{m} (-\lambda_i) \tag{23.219}$$

Recall that an antisymmetric matrix can be skew-diagonalized by an orthogonal matrix S. We didn't quite fix which one, because we didn't specify the signs of the λ_i. Therefore, up to sign, the Gaussian integral is just the product of skew eigenvalues.

On the other hand, the integral can also be evaluated as a polynomial in the matrix elements of A. Indeed the *Pfaffian* of the antisymmetric matrix can be defined as:

$$\mathrm{pfaff}(A) := \int [d\theta^1 \cdots d\theta^{2m}] \exp[\frac{1}{2}\theta^a A_{ab}\theta^b]. \tag{23.220}$$

With a little thought one shows that expanding this out leads to

$$\begin{aligned}\mathrm{pfaff} A &= \frac{1}{m! 2^m} \sum_{\sigma \in S_{2m}} \epsilon(\sigma) A_{\sigma(1)\sigma(2)} \cdots A_{\sigma(2m-1)\sigma(2m)} \\ &= A_{12} A_{34} \cdots A_{2m-1,2m} + \cdots\end{aligned} \tag{23.221}$$

`eq:pfaffian`

This definition of the Pfaffian resembles that of the determinant of a matrix, but note that it is slightly different. Since A is a bilinear form it transforms as $A \to S^{tr} A S$ under change of basis. Therefore, the Pfaffian is slightly basis-dependent:

$$\mathrm{pfaff}(S^{tr} A S) := \det S \ \cdot \ \mathrm{pfaff}(A) \tag{23.222}$$

`eq:pfaff-tmn`

We can easily prove this using the change-of-variables formula for the Berezin integral. (Do that!)

Now a beautiful property of the Pfaffian is that it is a canonical square-root of the determinant of an antisymmetric matrix.

$$(\mathrm{pfaff} A)^2 = \det A. \tag{23.223}$$

`eq:DetSqrt`

(In particular, the determinant of an antisymmetric matrix - a complicated polynomial in the matrix elements - has a canonical polynomial square root.)

Using the Berezin integral we will now give a simple proof of (23.223). First, note that if M is *any* $n \times n$ matrix and we have two sets of generators $\theta_{\pm}^{a}$, $a = 1, \ldots, n$ of our Grassmann algebra then

$$\boxed{\int [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \exp[\theta_+^i M_{ia} \theta_-^a] = \det M} \tag{23.224}$$

eq:DET-M-GRASS

An easy way to prove this is to make a transformation $M \to S^{-1} M S$ to Jordan canonical form. The change of variables of θ_+^a, θ_-^a by $S^{tr,-1}$ and S, respectively, cancel each other out. [44] Assuming that the matrix is diagonalizable we have

$$\int [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \exp[\theta_+^1 \theta_-^1 \lambda_1 + \theta_+^2 \theta_-^2 \lambda_2 + \cdots] = \prod_{i=1}^{n} \lambda_i = \det M \tag{23.225}$$

To check the sign we observe that the following moves always involve moving an even number of θ's past each other. For example,

$$\begin{aligned} \theta_+^1 \theta_-^1 \theta_+^2 \theta_-^2 \theta_+^3 \theta_-^3 \theta_+^4 \theta_-^4 &= \theta_+^1 \theta_+^2 \theta_-^2 \theta_-^1 \theta_+^3 \theta_-^3 \theta_+^4 \theta_-^4 \\ &= \theta_+^1 \theta_+^2 \theta_+^3 \theta_-^3 \theta_-^2 \theta_-^1 \theta_+^4 \theta_-^4 \\ &= \theta_+^1 \theta_+^2 \theta_+^3 \theta_+^4 \theta_-^4 \theta_-^3 \theta_-^2 \theta_-^1 \end{aligned} \tag{23.226}$$

We leave the case when M has nontrivial Jordan form as a (good) exercise.

Now apply this to $M_{ia} \to A_{ij}$ with $n = 2m$ and consider

$$\det A = \int [d\theta_-^1 \cdots d\theta_-^{2m} d\theta_+^{2m} \cdots d\theta_+^1] \exp[\theta_+^i A_{ij} \theta_-^j] \tag{23.227}$$

Change variables to

$$\theta_{\pm}^i := \frac{1}{\sqrt{2}} (\psi^i \pm \chi^i) \qquad i = 1, \ldots, 2m \tag{23.228}$$

and note that

$$\theta_+^i A_{ij} \theta_-^j = \frac{1}{2} \psi^i A_{ij} \psi^j - \frac{1}{2} \chi^i A_{ij} \chi^j \tag{23.229}$$

To compute the superdeterminant of the change of variables perhaps the simplest way to

[44] The reader might worry about a sign at this point. To allay this fear note that we could rewrite the measure as $\epsilon[d\theta_-^1 \cdots d\theta_-^n d\theta_+^1 \cdots d\theta_+^n]$. With the latter measure the two factors in the Berezinian clearly cancel each other. But then we encounter the same sign ϵ going back to the desired ordering with $[d\tilde{\theta}_+^1 \cdots d\tilde{\theta}_+^n] = \epsilon[d\tilde{\theta}_+^n \cdots d\tilde{\theta}_+^1]$.

proceed is to compute

$$
\begin{aligned}
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \theta_+^1 \cdots \theta_+^{2m} \theta_-^{2m} \cdots \theta_-^1 = \\
\frac{1}{2^{2m}} \int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\psi^1 + \chi^1) \cdots (\psi^{2m} + \chi^{2m})(\psi^{2m} - \chi^{2m}) \cdots (\psi^1 - \chi^1) = \\
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\chi^{2m}\psi^{2m})(\chi^{2m-1}\psi^{2m-1}) \cdots (\chi^1\psi^1) = \\
\int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] (\chi^{2m} \cdots \chi^1 \psi^1 \cdots \psi^{2m}) = \\
\int [d\psi^1 \cdots d\psi^{2m}] (\psi^1 \cdots \psi^{2m}) = \\
(-1)^{\frac{1}{2}(2m)(2m-1)} = (-1)^m
\end{aligned}
\tag{23.230}
$$

from which we conclude that

$$
[d\theta_-^1 \cdots d\theta_-^{2m} d\theta_+^{2m} \cdots d\theta_+^1] = (-1)^m [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \tag{23.231}
$$

So our change of variables gives

$$
\begin{aligned}
\det A &= (-1)^m \int [d\psi^1 \cdots d\psi^{2m} d\chi^1 \cdots d\chi^{2m}] \exp\left[\frac{1}{2}\psi^i A_{ij}\psi^j - \frac{1}{2}\chi^i A_{ij}\chi^j\right] \\
&= (\mathrm{pfaff} A)^2
\end{aligned}
\tag{23.232}
$$

Which concludes the proof of (23.223) ♠

Remarks

1. ♣ Remarks on localization from integral over a fermion zeromode
2. ♣ Remarks on use of Pfaffian in the general definition of Euler characteristic.
3. Why is the transformation (23.222) compatible with (23.223) and the invariance of the determinant under $A \to S^{-1}AS$? The reason is that for $S^{tr} = S^{-1}$ we have S is orthogonal so that $\det S = \pm 1$ and hence $(\det S)^2 = 1$.
4. *Pfaffians in families.* Sometimes the Pfaffian is defined as a squareroot of the determinant $\det A$ of an antisymmetric matrix. This has the disadvantage that the sign of the Pfaffian is not well-defined. In our definition, for a finite-dimensional matrix, there is a canonical Pfaffian. On the other hand, in some problems it is important to make sense of the Pfaffian of an anti-symmetric form on an infinite-dimensional Hilbert space. So, one needs another definition. Since determinants of infinite-dimensional operators can be defined by zeta-function regularization of the product of their eigenvalues one proceeds by defining the Pfaffian from the square-root of the determinant. So, we try to define the Pfaffian as:

$$
\mathrm{pfaff} A \quad \stackrel{?}{=} \quad \prod_{\lambda > 0} \lambda \tag{23.233}
$$

eq:postv

where the product runs over the skew eigenvalues. But this is not a good definition for many purposes because in families skew eigenvalues can smoothly change sign. Consider, for example the family

$$A(\alpha) = \begin{pmatrix} 0 & \cos\alpha \\ -\cos\alpha & 0 \end{pmatrix} \qquad 0 \le \alpha \le 2\pi \tag{23.234}$$ eq:smple

Then the Pfaffian according to the above definition (23.233) would not be differentiable at $\alpha = \pi/2$ and $3\pi/2$. Really, the pfaffian is a section of a line bundle, as we explain in Section §24 below.

One approach to pinning down the sign of the Pfaffian is simply to choose a sign at one point of the family, and then follow the skew eigenvalues continuously. With this definition

$$\text{pfaff}(A(\alpha)) = \cos\alpha \tag{23.235}$$

(in agreement with our definition for finite-dimensional forms). This is a perfectly reasonable definition. However, in some problems involving gauge invariance one meets quadratic forms A which should be identified up to gauge transformation. Suppose we identify A up to orthogonal transformations. Then the equivalence class $[A(\alpha)]$ is a closed family of operators for $0 \le \alpha \le \pi$. If we take a smooth definition of the Pfaffian of $A(\alpha)$ then we find that it changes sign under $\alpha \to \alpha + \pi$, so in fact, it behaves more like the section of the Mobius band over the circle. We return to this in Section §24.7 below.

Exercise

a.) Show that for $q = 2$ $\det A = (a_{12})^2$

b.) Show that for $q = 4$

$$\text{Pfaff}(A) = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23} \tag{23.236}$$

Check by direct compuation that indeed

$$\det A = (a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23})^2 \tag{23.237}$$

Exercise

a.) Prove equation (23.221).

b.) Explain why we divide the sum by the order of the group of centrally-symmetric shuffles (See Chapter 1, Sections 4 and 5) WB_m.

Exercise

Prove (23.224) by completing the argument for nontrivial Jordan form.

b.) Prove (23.224) by a direct evaluation of the integral without changing variables to obtain the standard expression

$$\det M = \sum_{\sigma\in S_n} \epsilon(\sigma) M_{1\sigma(1)} M_{2\sigma(2)} \cdots M_{n\sigma(n)} \tag{23.238}$$

Exercise

Let $z_1, \dots, z_{2N}$ be points in the complex plane. Show that

$$\left(\text{Pfaff}\frac{1}{z_i - z_j}\right)^k \prod_{i<j} (z_i - z_j)^\ell \tag{23.239}$$

is a polynomial in z_i of degree $N(2N-1)\ell - kN$, so long as $k \le \ell$, which transforms under S_{2N} with the sign $\epsilon(\sigma)^{k+\ell}$.

Expressions like this have proven useful in the theory of the fractional quantum Hall effect.

23.12.3 Gaussian integral on $\mathcal{R}^{p|q}$

Now we put these results together and consider the general Gaussian integral on $\mathcal{R}^{p|q}$:

$$(2\pi)^{-p/2} \int_{\mathcal{R}^{p|q}} [dt|d\theta] \exp\left[-\frac{1}{2} t^a Q_{ab} t^b - t^a B_{ai} \theta^i + \frac{1}{2}\theta^i A_{ij} \theta^j\right] \tag{23.240}$$

We can consider the quadratic form to have matrix elements in a general supercommutative ring (but they are constant in t^a, θ^i) so we allow odd off-diagonal terms like B_{ai}.

We can complete the square with the change of variables:

$$\begin{aligned} \tilde{t}^a &= t^a \\ \tilde{\theta}^i &= \theta^i + (A^{-1})_{ij} t^a B_{aj} \end{aligned} \tag{23.241}$$

The change of variables formula gives $[d\tilde{t}|d\tilde{\theta}] = [dt|d\theta]$ and hence we evaluate the integral to get

$$\frac{\text{Pfaff}(A)}{\sqrt{\det(Q - BA^{-1}B^{tr})}} \tag{23.242}$$

This can be written as $(\text{Ber}(\mathcal{Q}))^{-1/2}$ where $\mathcal{Q}$ is the super-quadratic form

$$\mathcal{Q} = \begin{pmatrix} Q & B \\ B^{tr} & A \end{pmatrix} \tag{23.243}$$

but the latter expression is slightly ambiguous since there are two squareroots of $\det A$.

23.12.4 Supersymmetric Cancelations

Suppose a super-quadratic form on $\mathcal{R}^{n|2n}$ is of the special form

$$\mathcal{Q} = \begin{pmatrix} M^{tr}M & 0 & 0 \\ 0 & 0 & M \\ 0 & -M^{tr} & 0 \end{pmatrix} \tag{23.244}$$ eq:TFT-1

where M is nonsingular and $\mathrm{Re}(M^{tr}M) > 0$. Then the Gaussian integral is just

$$\mathrm{sign}(\mathrm{det}M) \tag{23.245}$$ eq:TFT-2

Note that M (reduced modulo nilpotents) might be a complex matrix, and the integral is still sensible so long as $\mathrm{Re}(M^{tr}M) > 0$. Therefore we define

$$\mathrm{sign}(\mathrm{det}M) := \begin{cases} +1 & |\arg(\mathrm{det}M)| < \pi/4 \\ -1 & |\arg(\mathrm{det}M) - \pi| < \pi/4 \end{cases} \tag{23.246}$$

Thus, the result of the Gaussian integral (23.245) is "almost" independent of the details of M. There is a nice "theoretical" explanation of this fact which is a paradigm for arguments in supersymmetric field theory and topological field theory.

So, let us denote, for brevity

$$[d\theta_- d\theta_+] := [d\theta_-^1 \cdots d\theta_-^n d\theta_+^n \cdots d\theta_+^1] \tag{23.247}$$

and we consider the integral

$$I[M] := (2\pi)^{-n/2} \int_{\mathcal{R}^{n|2n}} [dt|d\theta_- d\theta_+] \mathrm{exp}\left[-\frac{1}{2} t^i (M^{tr}M)_{ik} t^k + \theta_+^i M_{ij} \theta_-^j\right] = \mathrm{sign}(\mathrm{det}M) \tag{23.248}$$ eq:TFT-3

It is useful to introduce n additional bosonic coordinates H^i and instead write this as an integral over $\mathcal{R}^{2n|2n}$:

$$I[M] = (2\pi)^{-n} \int_{\mathcal{R}^{2n|2n}} [dtdH|d\theta_- d\theta_+] \mathrm{exp}\left[-\frac{1}{2} H^i H^i + \sqrt{-1} H^i M_{ij} t^j + \theta_+^i M_{ij} \theta_-^j\right] \tag{23.249}$$ eq:TFT-1

Now, introduce the odd vector field

$$Q := \theta_-^k \frac{\partial}{\partial t^k} - \sqrt{-1} H^k \frac{\partial}{\partial \theta_+^k} \tag{23.250}$$

Note that, on the one hand, the "action" can be written as

$$Q(\Psi) = -\frac{1}{2} H^i H^i + \sqrt{-1} H^i M_{ij} t^j + \theta_+^i M_{ij} \theta_-^j \tag{23.251}$$

where

$$\Psi = -\frac{i}{2} \theta_+^k H^k - \theta_+^i M_{ij} t^j \tag{23.252}$$

and on the other hand,

$$Q^2 = \frac{1}{2}[Q,Q]_+ = 0 \tag{23.253}$$

So, now suppose we perturb $M \to M + \delta M$. Then the change in the Gaussian integral can be written as

$$\delta I[M] = \int [dtdH|d\theta_- d\theta_+] Q(\delta\Psi) e^{Q(\Psi)} = \int [dtdH|d\theta_- d\theta_+] Q(\delta\Psi e^{Q(\Psi)}) = 0 \tag{23.254}$$

where the last inequality follows from integration by parts.

The reader might be bothered by this. The answer (23.248) *does* depend a little bit on M. Moreover, why can't we just use the argument to put M to zero? But then, of course, the integral would seem to be $\infty \times 0$ for $M = 0$. If a perturbation makes M singular then we get a factor of $\infty \times 0$ where the ∞ comes from the integral dt and the 0 comes from the fermionic integral. Recall, however, that in the definition of the integral we do the fermionic integral first and therefore $\int [dtdH|d\theta_- d\theta_+]1 = 0$. Therefore, we could replace the integrand by

$$e^{Q(\Psi)} - 1 = Q\left(\Psi + \frac{1}{2!}\Psi Q(\Psi) + \cdots\right) \tag{23.255}$$

There will be a term which survives the fermionic integral but it is a total derivative in $\frac{\partial}{\partial t^i}$ which does not vanish at infinity in field space. Thus, *singular perturbations can change the value of the integral.*

23.13 References

In preparing this section we have used the following references:

1. P. Deligne and J.W. Morgan, *Notes on Supersymmetry (following Joseph Bernstein* in *Quantum Fields and Strings: A Course for Mathematicians*, Vol. 1, pp.41-96 AMS

2. D. Leites, "Introduction to supermanifolds," 1980 Russ. Math. Surv. 35 1.

3. E. Witten, "Notes on Supermanifolds and Integration," arXiv:1209.2199.

4. www.math.ucla.edu/ vsv/papers/ch3.pdf

5. J. Groeger, Differential Geometry of Supermanifolds, http://www.mathematik.hu-berlin.de/ groegerj

   Possibly useful, but I haven't seen them yet:

6. V. Varadarajan, *Supersymmetry for Mathematicians: An Introduction*

7. *Super Linear Algebra* by Kandasamy and Smarandache

   For an extremely accessible discussion of the theory of schemes See

8. D. Eisenbud and J. Harris, *The Geometry of Schemes*, Springer GTM

24. Determinant Lines, Pfaffian Lines, Berezinian Lines, and anomalies

sec:DETERMINANTS

24.1 The determinant and determinant line of a linear operator in finite dimensions

Recall that a one-dimensional vector space over κ is called a *line*. If L is a line then a linear transformation $T : L \to L$ can canonically be identified with an element of κ. Indeed, choose any basis vector v for L then $T(v) = tv$, with $t \in \kappa$, and the number t does not depend on v. On the other hand, suppose we have two lines L_1, L_2. They are, of course, isomorphic, but not naturally so. In this case if we have a linear transformation $T : L_1 \to L_2$ then there is not canonical way of identifying T with an element of κ because there is no *a priori* way of identifying a choice of basis for L_1 with a choice of basis for L_2. Put differently, $\mathrm{Hom}(L, L) \cong L^\vee \otimes L \cong \kappa$ is a natural isomorphism, but $\mathrm{Hom}(L_1, L_2) \cong L_1^\vee \otimes L_2$ is just another line.

Now suppose that

$$T : V \to W \tag{24.1}$$

is a linear transformation between different vector spaces over κ where the dimension is possibly larger than one. Then there is no canonical notion of the determinant as a number. Choose ordered bases $\{v_i\}$ for V and $\{w_a\}$ for W, and define M_{ai} to be the matrix of T with respect to that basis. Then under this isomorphism $T \in \mathrm{Hom}(V, W) \cong V^* \otimes W$ may be written as

$$T = \sum_{i,a} M_{ai} \hat{v}_i \otimes w_a \tag{24.2}$$

and if $\mathrm{dim}V = \mathrm{dim}W$ one can of course define the number $\mathrm{det}M_{ai}$. However, a change of basis of V by S_1 and of W by S_2 changes the matrix $M \to S_2^{-1} M S_1$ and hence leads to

$$\mathrm{det}M' = \mathrm{det}S_2^{-1} \mathrm{det}M \mathrm{det}S_1 \tag{24.3}$$

in the new basis. *If V, W are not naturally isomorphic, then we cannot naturally assign a number we would want to call* $\mathrm{det}T$.

Nevertheless, there *is* a good mathematical construction of $\mathrm{det}T$ which is natural with respect to the domain and range of T. What we do is consider the 1-dimensional vector spaces $\Lambda^d V$ and $\Lambda^{d'} W$ where $d = \mathrm{dim}V, d' = \mathrm{dim}W$. Then there is a canonically defined linear transformation

$$\mathrm{det}T : \Lambda^d V \to \Lambda^{d'} W \tag{24.4}$$

For $d \neq d'$ it is zero, and for $d = d'$ we can write it by choosing bases v_i, w_a. Denote the dual basis $\hat{v}_i$ so that $T = \sum_{i,j} M_{ai} \hat{v}_i \otimes w_a$. Then

$$\mathrm{det}T := \frac{1}{(d!)^2} \sum_{a_s, i_s} M_{a_1 i_1} \cdots M_{a_d i_d} \hat{v}_{i_1} \wedge \cdots \wedge \hat{v}_{i_d} \otimes w_{a_1} \wedge \cdots \wedge w_{a_d} \tag{24.5}$$ eq:detelement

The important thing about this formula is that, as opposed to the determinant defined as a polynomial in matrix elements, the object (24.5) is *natural* with respect to both V and W. That is, it is independent of basis (even though we chose a basis to write it out, if we change basis we get the same object. This is not true of the determinant defined as a polynomial in matrix elements.)

While (24.5) is natural it requires interpretation. It is not a number, it is an element of a one-dimensional vector space, i.e. a *line*. This line is called the *determinant line of* T

$$\mathrm{DET}(T) := \Lambda^{\dim V}(V^*) \otimes \Lambda^{\dim W}(W) \tag{24.6}$$ eq:detline

Thus, we have a one-dimensional vector space $\mathrm{DET}(T)$ and an element of that vector space $\det(T) \in \mathrm{DET}(T)$.

This is a nontrivial concept because:

1. Linear operators often come in families T_s. Then $\mathrm{DET}(T)$ becomes a nontrivial line bundle over parameter space.

2. The theory extends to infinite dimensional operators such as the Dirac operator. Indeed, in finite dimensions $\mathrm{DET}(T)$ does not depend on the choice of operator T except through its domain and target. This is no longer true in infinite dimensions.

Remarks

1. When $\dim V \neq \dim W$ the line $\mathrm{DET}(T)$ defined in (24.6) still makes sense, but we must define the element in that line, $\det T \in \mathrm{DET}(T)$ to be $\det T = 0$.
2. When $W = V$, that is, when they *are* canonically isomorphic then $\det M_{ij}$ is basis independent. Indeed, in this case there is a canonical isomorphism

$$\Lambda^d(V) \otimes \Lambda^d(V^\vee) \cong \kappa \tag{24.7}$$

(κ can be $\mathbb{R}$ or $\mathbb{C}$, or any field.)

Example Above we considered an interesting family of one-dimensional vector spaces

$$\mathcal{L}_+ = \{(\hat{x}, v) | \hat{x} \cdot \vec{\sigma} v = +v\} \subset S^2 \times \mathbb{C}^2 \tag{24.8}$$

For each point $\hat{x} \in S^2$ we have a one-dimensional subspace $\mathcal{L}_{+,\hat{x}} \subset \mathbb{C}^2$. Let us let

$$L = S^2 \times \mathbb{C} \tag{24.9}$$

be another family of one-dimensional vector spaces. We can define an operator

$$T_{\hat{x}} : \mathcal{L}_{+,\hat{x}} \to L_{\hat{x}} \tag{24.10}$$

defined by just projecting to the first component of the vector $v \in \mathcal{L}_{+,\hat{x}}$. To find a matrix of $T_{\hat{x}}$ we need to choose a basis. As we discussed we might choose a basis vector

$$e_+ = \begin{pmatrix} \cos\frac{1}{2}\theta \\ e^{i\phi}\sin\frac{1}{2}\theta \end{pmatrix} \tag{24.11}$$

away from the south pole, while away from the north pole we might choose:

$$e_- = \begin{pmatrix} e^{-i\phi}\cos\frac{1}{2}\theta \\ \sin\frac{1}{2}\theta \end{pmatrix} \tag{24.12}$$

We can also simply choose 1 as a basis for $\mathbb{C}$ for the target of $T_{\hat{x}}$.

With respect to this basis we would have a determinant function

$$\det M_{ai} = \cos\frac{1}{2}\theta \qquad 0 \le \theta < \pi \tag{24.13}$$

$$\det M_{ai} = e^{-i\phi}\cos\frac{1}{2}\theta \qquad 0 < \theta \le \pi \tag{24.14}$$

Note that the second expression does make sense at the south pole because $\cos\frac{\pi}{2} = 0$. Clearly this does not define a function on S^2. Rather, it defines a section of a line bundle. Moreover, it has exactly one zero.

Exercise

Consider a finite dimensional vector space V of dimension d.

a.) Show that there is a canonical isomorphism $\Lambda^d(V^\vee) \cong (\Lambda^d V)^\vee$

b.) Show that there is a canonical isomorphism

$$\Lambda^d V^\vee \otimes \Lambda^d V \to \kappa \tag{24.15}$$

24.2 Determinant line of a vector space and of a complex

If V is a finite-dimensional vector space of dimension d then the line $\Lambda^d V$ is often called the *determinant line of* V and denoted

$$\mathrm{DET}(V) := \Lambda^{\mathrm{dim}V} V \tag{24.16}$$

Because there is a natural isomorphism

$$\mathrm{DET}(V)^\vee \otimes \mathrm{DET}(V) \cong \kappa \tag{24.17}$$

and since the one-dimensional space κ acts like a multiplicative identity under $\otimes$, we also denote

$$\mathrm{DET}(V)^{-1} := \mathrm{DET}(V)^\vee \tag{24.18}$$

Therefore we will also denote

$$(v_1 \wedge \cdots \wedge v_d)^{-1} := (v_1 \wedge \cdots \wedge v_d)^\vee \tag{24.19}$$

for any choice of basis $\{v_i\}$.

Again, when we consider families of vector spaces, we can get interesting line bundles from this construction:

Example Consider the Grassmannian $Gr(k,n)$ of k-dimensional subspaces of $\mathbb{C}^n$. For each subspace $W \in Gr(k,n)$ we may associate the line $\Lambda^k W$. This gives another nontrivial family of lines which is naturally a determinant line bundle. A notable use of this Determinant line in physics is that it can be interpreted as the vacuum line in quantization of n free fermions. Then the possible sets of creation operators form the Grassmannian $Gr(n,2n)$. ♣Say more ♣

Now consider a short exact sequence of vector spaces:

$$0 \to V_1 \to V_2 \xrightarrow{\pi} V_3 \to 0 \tag{24.20}$$

Then we claim there is a canonical isomorphism

$$\mathrm{DET}(V_1) \otimes \mathrm{DET}(V_3) \cong \mathrm{DET}(V_2) \tag{24.21}$$

eq:DET-MULT

To see this, consider any ordered basis $\{v_1, \dots, v_{d_1}\}$ for V_1 and $\{w_1, \dots, w_{d_3}\}$ for V_3. Then lift the w_a to vectors $\tilde{w}_a$ in V_2 so that $\pi(\tilde{w}_a) = w_a$. Then

$$\{v_1, \dots, v_{d_1}, \tilde{w}_1, \dots, \tilde{w}_{d_3}\} \tag{24.22}$$

is an ordered basis for V_2. Our canonical isomorphism (24.21) is defined by

$$(v_1 \wedge \cdots \wedge v_{d_1}) \otimes (w_1 \wedge \cdots \wedge w_{d_3}) \mapsto v_1 \wedge \cdots \wedge v_{d_1} \wedge \tilde{w}_1 \wedge \cdots \wedge \tilde{w}_{d_3} \tag{24.23}$$

The main point to check here is that the choice of lifts $\tilde{w}_a$ do not matter on the RHS. This is clear since for a different choice $\tilde{w}'_a - \tilde{w}_a \in V_1$ and the difference is annihilated by the first part of the product. Next, under changes of bases both sides transform in the same way, so the isomorphism is basis-independent, and hence natural. Put differently, the element

$$(v_1 \wedge \cdots \wedge v_{d_1})^{-1} \otimes (v_1 \wedge \cdots \wedge v_{d_1} \wedge \tilde{w}_1 \wedge \cdots \wedge \tilde{w}_{d_3}) \otimes (w_1 \wedge \cdots \wedge w_{d_3})^{-1} \tag{24.24}$$

of the line given in the LHS of (24.25) is actually *independent* of the choice of bases $\{v_1, \dots, v_{d_1}\}$, $\{w_1, \dots, w_{d_3}\}$, and the lifts $\tilde{w}_a$. So even though we chose bases to exhibit this element it is basis-independent, and hence natural. It gives therefore a natural isomorphism

$$\mathrm{DET}(V_1)^{-1} \otimes \mathrm{DET}(V_2) \otimes \mathrm{DET}(V_3)^{-1} \cong \kappa \tag{24.25}$$

eq:3-Term-Cplx

The same kind of reasoning as we used to prove (24.21) can be used to prove that if

$$0 \to V_1 \to \cdots \to V_i \xrightarrow{T_i} V_{i+1} \to \cdots \to V_n \to 0 \tag{24.26}$$

is an exact sequence, so $\mathrm{im}T_i = \mathrm{ker}T_{i+1}$, then there is a canonical isomorphism

$$\bigotimes_{i \text{ odd}} \mathrm{DET}(V_i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(V_i) \cong \kappa \tag{24.27}$$

eq:ExactSequen

Exercise *Determinant of a complex*
A multiplicative version of the Euler-Poincaré principal is that if

$$0 \to V_1 \to \cdots \to V_i \xrightarrow{d_i} V_{i+1} \to \cdots \to V_n \to 0 \tag{24.28}$$

is a complex $d_{i+1}d_i = 0$ (i.e. not necessarily exact) then there is a natural isomorphism:

$$\bigotimes_{i \text{ odd}} \mathrm{DET}(V_i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(V_i) \cong \bigotimes_{i \text{ odd}} \mathrm{DET}(H^i)^{-1} \otimes \bigotimes_{i \text{ even}} \mathrm{DET}(H^i) \tag{24.29}$$

eq:DetComplex

24.3 Abstract defining properties of determinants

Because we want to speak of determinants and determinant lines in infinite dimensions it can be useful to take a more abstract approach to the determinant and determinant line of T. These can be abstractly characterized by the following three properties:

1. $\det(T) \neq 0$ iff T is invertible.
2. $\mathrm{DET}(T_2 \circ T_1) \cong \mathrm{DET}(T_2) \otimes \mathrm{DET}(T_1)$
3. If T_1, T_2, T_3 map between two short exact sequences

$$\begin{array}{ccccccccc} 0 & \to & E_1 & \to & E_2 & \overset{\pi}{\to} & E_3 & \to & 0 \\ & & T_1 \downarrow & & T_2 \downarrow & & T_3 \downarrow & & \\ 0 & \to & F_1 & \to & F_2 & \overset{\pi}{\to} & F_3 & \to & 0 \end{array} \tag{24.30}$$

then

$$\mathrm{DET}(T_2) \cong \mathrm{DET}(T_1) \otimes \mathrm{DET}(T_3) \tag{24.31}$$

canonically, and under this isomorphism

$$\det(T_2) \to \det T_1 \otimes \det T_3 \tag{24.32}$$

Exercise

Show that property (3) means, essentially, that T_2 has block upper-triangular form. Show this by considering

$$\begin{array}{ccccccccc} 0 & \to & \mathbb{C}^n & \overset{\iota}{\to} & \mathbb{C}^{n+m} & \overset{\pi}{\to} & \mathbb{C}^m & \to & 0 \\ & & T_1 \downarrow & & T_2 \downarrow & & T_3 \downarrow & & \\ 0 & \to & \mathbb{C}^n & \overset{\iota}{\to} & \mathbb{C}^{n+m} & \overset{\pi}{\to} & \mathbb{C}^m & \to & 0 \end{array} \tag{24.33}$$

where

$$\begin{aligned} \iota : (x_1, \dots, x_n) &\to (x_1, \dots, x_n, 0, \dots, 0) \\ \pi : (y_1, \dots, y_{n+m}) &\to (y_{n+1}, \dots, y_{n+m}) \end{aligned} \tag{24.34}$$

Show that

$$T_2 = \begin{pmatrix} T_1 & T_{12} \\ 0 & T_2 \end{pmatrix} \tag{24.35}$$

where T_{12} is in general nonzero.

---

sec:PfaffLine

24.4 Pfaffian Line

Just as for determinants, Pfaffians are properly regarded as sections of line (bundles).

Recall that for an antisymmetric matrix A, $\mathrm{Pfaff}(S^{tr} A S) = (\det S)\mathrm{Pfaff}(A)$. So the Pfaffian is basis-dependent, yet, once again, there is a basis independent notion of a Pfaffian.

Let T be an antisymmetric bilinear form on a finite-dimensional vector space V.

Recall that a bilinear form T on a vector space V can be regarded as an element of

$$T : V \to V^{\vee} \tag{24.36}$$

Then, from our definition above

$$\mathrm{DET}(T) = (\Lambda^d V^{\vee})^{\otimes 2} \tag{24.37}$$

For an antisymmetric bilinear form we define the Pfaffian line of T to be the "squareroot":

$$\mathrm{PFAFF}(T) := \Lambda^d V^{\vee} \tag{24.38}$$

On the other hand, a bilinear form defines a map $V \otimes V \to \kappa$ and hence is an element of

$$T \in V^{\vee} \otimes V^{\vee} \tag{24.39}$$

Moreover, if T is antisymmetric then it defines a 2-form

$$\omega_T \in \Lambda^2 V^{\vee} \tag{24.40}$$

If $d = 2m$ we define the Pfaffian element to be:

$$\mathrm{pfaff}T := \frac{\omega_T^m}{m!} \in \Lambda^n V^{\vee} \tag{24.41}$$

Exercise *Comparing Liouville and Riemann volume forms*

a.) Suppose V is a real vector space with a a positive definite symmetric form g. Show that if A_{ij} is the anti-symmetric matrix of T with respect to an orthonormal basis with respect to g, then

$$\frac{\omega_T^m}{m!} = \mathrm{pfaff}(A)\mathrm{vol}\,(g) \tag{24.42}$$

where $\mathrm{vol}\,(g)$ is the volume form of the metric, i.e. $\mathrm{vol}\,(g) = e^1 \wedge \cdots \wedge e^n$ where $\{e^1, \ldots, e^n\}$ is an ordered ON basis.

Note that if we change ordered ON bases then both $\mathrm{pfaff}(A)$ and $\mathrm{vol}\,(g)$ change by $\det S$, where $S \in O(2m)$, so the product is well-defined.

b.) Let M be a symplectic manifold with symplectic form ω. Choose any Riemannian metric g on M. Then let $A(g)$ be the antisymmetric matrix of ω with respect to an ON frame for g. Note that $\det(A(g))$ does not depend on the choice of ON frame, and is hence a globally well-defined function. Show that M is orientable iff $\det(A(g))$ admits a globally-defined square root on M.

24.5 Determinants and determinant lines in infinite dimensions

24.5.1 Determinants

Now let us consider a linear operator $T : \mathcal{H} \to \mathcal{H}$. We will take $\mathcal{H}$ to be an infinite-dimensional separable Hilbert space. [45]

Can we speak meaningfully of det$T$ in this case? Suppose that $T = 1 + \Delta$, where Δ is traceclass. Then it has a set of eigenvalues $\{\delta_k\}_{k=1}^{\infty}$ with

$$\sum_{k=1}^{\infty} |\delta_k| < \infty \tag{24.43}$$

In this case $\sum_{k=1}^{\infty} \log\delta_k$ is a well-defined absolutely convergent series and

$$\mathrm{det}T := \lim_{N\to\infty} \prod_{k=1}^{N} (1 + \delta_k) \tag{24.44}$$

is well-defined, and can be taken to be the determinant of T.

This is a good start, but the class of operators $1 + \textit{traceclass}$ is far too small for use in physics and mathematics. Another definition, known as the ζ-function determinant, introduced by Ray and Singer [46] can be defined as follows:

Let $T : \mathcal{H} \to \mathcal{H}$ be a self-adjoint operator with discrete point spectrum $\{\lambda_k\}_{k=1}^{\infty}$ assume that

1. No $\lambda_k = 0$ (otherwise det$T = 0$).

2. $|\lambda_k| \to \infty$ for $k \to \infty$.

3. If we form the series

$$\zeta_T(s) := \sum_k \lambda_k^{-s} := \sum_{\lambda_k > 0} \lambda_k^{-s} + \sum_{\lambda_k < 0} e^{-i\pi s} |\lambda_k|^{-s} \tag{24.45}$$

eq:zeta-functi

then the spectrum goes to infinity rapidly enough so that $\zeta_T(s)$ converges to an analytic function on the half-plane Re$(s) > R_0$, for some R_0, and admits an analytic continuation to a holomorphic function of s near $s = 0$.

When these conditions are satisfied we may define

$$\mathrm{det}_\zeta(T) := \exp[-\zeta_T'(0)] \tag{24.46}$$

Remark: A typical example of an operator for which these conditions apply is an elliptic operator on a compact manifold. For example, the Laplacian acting on tensors on a smooth compact manifold, or the Dirac operator on a smooth compact spin manifold are common examples. See Example **** below.

[45] This is not strictly necessary for some definitions and constructions below.
[46] GIVE REFERENCE.

The next natural question is to consider determinant lines for operators $T : \mathcal{H}^1 \to \mathcal{H}^2$ between two "different" Hilbert spaces. Of course, we proved in Section §13 that any two separable Hilbert spaces are isomorphic. However, there is no natural isomorphism, so the question where an expression like det$T$ should be valued is just as valid as in finite dimensions. A good example is the *chiral* Dirac operator on an even dimensional spin manifold. First, we must identify a suitable class of operators where such determinant lines can make sense.

24.5.2 Fredholom Operators

Definition

a.) An operator $T : \mathcal{H}^1 \to \mathcal{H}^2$ between two separable Hilbert spaces is said to be *Fredholm* if ker$T$ and cok$T$ are finite-dimensional.

b.) The *index* of a Fredholm operator is defined to be

$$\text{Ind}(T) := \dim \ker T - \dim \text{cok} T \tag{24.47}$$

Comments and Facts

1. We are generally interested in *unbounded* operators. Then, as we have discussed the domain $D(T)$ is an important part of the definition of T. The above definition is a little sloppy for unbounded operators. For some purposes, such as index theory one can replace T by $T/(1 + T^\dagger T)$ and work with a bounded operator.

2. One often sees Fredholm operators defined with the extra requirement that im$(T) \subset \mathcal{H}^2$ is a *closed* subspace. In fact, it can be shown (using the closed graph theorem) that if T is bounded and ker$T$ and cok$T$ are finite dimensional then this requirement is satisfied. ♣Give argument. ♣

3. Another definition one finds is that a bounded operator T is Fredholm iff there is an inverse up to compact operators. That is, T is Fredholm iff there is a bounded operator $S : \mathcal{H}^2 \to \mathcal{H}^1$ so that $TS - 1$ and $ST - 1$ are compact operators. (Recall that compact operators are finite-rank operators, or limits in the norm topology of finite-rank operators.) The equivalence of these definitions is known as *Atkinson's theorem.*

4. The space of all bounded Fredholm operators $\mathcal{F}(\mathcal{H}^1, \mathcal{H}^2)$ inherits a topology from the operator norm.

5. If T is Fredholm then there is an $\epsilon > 0$ (depending on T) so that if $\parallel K \parallel < \epsilon$ then $T + K$ is Fredholm and $\text{Ind}(T) = \text{Ind}(T + K)$. Therefore, the index is a continuous map:
$$\text{Ind} : \mathcal{F}(\mathcal{H}) \to \mathbb{Z} \tag{24.48}$$
and is hence constant on connected components.

6. In fact, the space of Fredholm operators has infinitely many connected components, in the norm topology, and these are in 1-1 correspondence with the integers and can be labeled by the index.

7. Warning: In the compact-open and strong operator topologies $\mathcal{F}(\mathcal{H})$ is *contractible.* [47]

♣♣: EXPLAIN: Even for unbounded operators (with dense domain of definition) the definition of Fredholm is that the kernel and cokernel are finite dimensional. There is no need to say that the range is closed. But only when the range is closed is there an isomorphism of the kernel of $T^\dagger$ with the cokernel. A good example is d/dx on $L^2([1,\infty)$ (or do we need the half-line?) The kernel is zero because 1 is not normalizable, so the kernel of $T^\dagger$ is also zero. But we can construct a lot of states which are not in the cokernel. For example $1/x^n$ is in the image of d/dx if $n > 3/2$ but not if $n \leq 3/2$ since the preimage would not be L^2-normalizable. So the range is not closed and the cokernel is not isomorphic to the kernel of the adjoint.

24.5.3 The determinant line for a family of Fredholm operators

There are two descriptions of the determinant line:

Construction 1: We define a line bundle DET first over the index zero component $\mathcal{F}(\mathcal{H}^1, \mathcal{H}^2)_0$ whose fiber at $T \in \mathcal{F}(\mathcal{H}^1, \mathcal{H}^2)$ is

$$\mathrm{DET}|_T := \{(S, \lambda) | S : \mathcal{H}^2 \to \mathcal{H}^1, S^{-1}T \in \mathcal{I}_1\}/\sim \tag{24.49}$$

where the equivalence relation is

$$(S_1, \lambda_1) \sim (S_2, \lambda_2) \qquad \leftrightarrow \qquad \lambda_2 = \lambda_1 \mathrm{det}(S_2^{-1}S_1) \tag{24.50}$$

where $S_2^{-1}S_1 = 1 + traceclass$ and we use the standard definition for this.

To check this one has to check that $\sim$ really is an equivalence relation. Next, note that $\mathrm{DET}|_T$ is indeed a one-dimensional vector space with vector space structure $z \cdot [(S, \lambda)] := [(S, z\lambda)]$, so any two vectors are proportional: $[(S_1, \lambda_1)] = \xi[(S_2, \lambda_2)]$ with $\xi = \lambda_1\lambda_2^{-1}\mathrm{det}(S_2^{-1}S_1)$.

The next thing to check is that, in the norm topology, the lines are indeed a continuous family of lines.

There is a canonical section of this line: $\det(T) := [(T, 1)]$ (in the index zero component).

[47] Here is an argument provided by Graeme Segal: Identify $\mathcal{H}$ with $L^2[0,1]$ and then write $\mathcal{H} \cong \mathcal{H}_t \oplus \mathcal{H}_{1-t}$ where $\mathcal{H}_t$ and $\mathcal{H}_{1-t}$ are the Hilbert spaces on the intervals $[0,t]$ and $[t,1]$, respectively. Then $\mathcal{H}$ is also isomorphic to $\mathcal{H}_t$. For an operator A on $\mathcal{H}$ let A_t be its image under this isomorphism. Then, one can check that $t \mapsto A_t \oplus 1_{\mathcal{H}_{1-t}}$ is continuous in the norm topology, deforms A to 1, and stays in the set of Fredholm operators if A is Fredholm.

One can show that for any T there is a canonical isomorphism

$$\mathrm{DET}|_T \cong \mathrm{DET}(\mathrm{ker}T)^{-1} \otimes \mathrm{DET}(\mathrm{cok}T) \tag{24.51}$$

The reason we can't just use the RHS as a definition of the determinant line is that in families $T(s)$, we can have the spaces $\mathrm{ker}T(s)$ and $\mathrm{cok}T(s)$ jump discontinuously in dimension.

This leads us to consider the second construction:

Construction 2: Let $\mathcal{S}$ be a family of Fredholm operators $T(s)$. For any positive real number a define

$$\mathcal{U}_a = \{T | a \notin \sigma(T^\dagger T)\} \tag{24.52}$$

If $T \in \mathcal{U}_a$ then we can use the spectral decomposition of the Hermitian operator $T^\dagger T$ to split the Hilbert space into the "low energy" and "high energy" modes:

$$\mathcal{H} = \mathcal{H}_{<a} \oplus \mathcal{H}_{>a} \tag{24.53}$$

(i.e. we use the spectral projection operators). Moreover, since T is Fredholm, the "low energy" space $\mathcal{H}_{<a}$ is in fact *finite-dimensional.*

Now notice that we have an exact sequence:

$$0 \to \mathrm{ker}T \to \mathcal{H}^1_{<a} \stackrel{T}{\to} \mathcal{H}^2_{<a} \to \mathrm{cok}T \to 0 \tag{24.54}$$

Now, using the property of determinant lines in exact sequences we conclude that there is a *canonical* isomorphism

$$\mathrm{DET}(\mathrm{ker}T)^{-1} \otimes \mathrm{DET}(\mathrm{cok}T) \cong \mathrm{DET}(\mathcal{H}^1_{<a})^{-1} \otimes \mathrm{DET}(\mathcal{H}^2_{<a}) \tag{24.55}$$

The advantage of using the RHS of this equation is that now we can consider what happens on overlaps $\mathcal{U}_{ab} = \mathcal{U}_a \cap \mathcal{U}_b$, where we can assume WLOG that $a < b$. Then

$$\mathcal{H}_b = \mathcal{H}_a \oplus \mathcal{H}_{a,b} \tag{24.56}$$

where $\mathcal{H}_{a,b}$ is the sum of eigenspaces with eigenvalues between a, b. Note that there is an isomorphism $T_{a,b} : \mathcal{H}^1_{a,b} \to \mathcal{H}^2_{a,b}$ (which is just the restriction of T) and hence $\mathrm{det}T_{a,b}$ is nonzero and a canonical trivialization of

$$\mathrm{DET}(\mathcal{H}^1_{a,b})^{-1} \otimes \mathrm{DET}(\mathcal{H}^2_{a,b}) \tag{24.57}$$

Using these trivializations the determinant line bundles patch together to give a smooth determinant line bundle over the whole family. ♣More detail. Warn about gerbes. ♣

24.5.4 The Quillen norm

In physical applications we generally want our path integrals and correlation functions to be numbers, rather than sections of line bundles. (Sometimes, we just have to live with the latter situation.)

$$\parallel (v^1 \wedge \cdots \wedge v^n)^{-1} \otimes (w^1 \wedge \cdots \wedge w^m) \parallel^2 = \frac{\det(w^a, w^b)}{\det(v^i, v^j)} \mathrm{det}'_\zeta(T^\dagger T) \tag{24.58}$$

♣ SHOW IT PATCHES NICELY ♣

24.5.5 References

1. G. Segal, Stanford Lecture 2, http://www.cgtp.duke.edu/ITP99/segal/
2. D. Freed, On Determinant Line Bundles
3. D. Freed, "Determinants, Torsion, and Strings," Commun. Math. Phys. **107**(1986)483
4. D. Freed and G. Moore, "Setting the Quantum Integrand of M Theory," hep-th/0409135.

24.6 Berezinian of a free module

There is an analog of the determinant line of a vector space also in $\mathbb{Z}_2$-graded linear algebra.

Let $\mathcal{A}$ be a supercommutative superalgebra and consider a free module $M \cong \mathcal{A}^{p|q}$. Then we can define a free module $\mathrm{Ber}(M)$ of rank $(1|0)$ or $(0|1)$ over $\mathcal{A}$ by assigning for every isomorphism

$$M \cong e_1\mathcal{A} \oplus \cdots \oplus e_{p+q}\mathcal{A} \tag{24.59}$$

a basis vector in $\mathrm{Ber}(M)$ denoted by $[e_1, \dots, e_p|e_{p+1} \dots, e_{p_q}]$ so that if T is an automorphism of M then $T(e_i) = e_k X^k{}_j$, with the matrix elements $X^k{}_j \in \mathcal{A}$ and we take

$$[Te_1, \dots, Te_p|Te_{p+1} \dots, Te_{p_q}] = \mathrm{Ber}(X)[e_1, \dots, e_p|e_{p+1} \dots, e_{p_q}] \tag{24.60}$$

To complete the definition we take the parity of $\mathrm{Ber}(M)$ to be

$$\mathrm{Ber}(M) \cong \begin{cases} \mathcal{A}^{1|0} & q = 0 \mathrm{mod} 2 \\ \mathcal{A}^{0|1} & q = 1 \mathrm{mod} 2 \end{cases} \tag{24.61}$$

Now, given an exact sequence

$$0 \to M' \to M \to M'' \to 0 \tag{24.62}$$

with $M' \cong \mathcal{A}^{p'|q'}$, $M'' \cong \mathcal{A}^{p''|q''}$, and $M \cong \mathcal{A}^{p|q}$ then there is a *natural* isomorphism

$$\mathrm{Ber}(M') \otimes \mathrm{Ber}(M'') \to \mathrm{Ber}(M) \tag{24.63}$$ eq:BerLineMult

This isomorphism is defined by choosing a basis $\{e'_1, \dots, e'_{p'+q'}\}$ for M' and a complementary basis $\{\tilde{e}''_1, \dots, \tilde{e}''_{p''+q''}\}$ for M which projects to a basis $\{e''_1, \dots, e''_{p''+q''}\}$ for M''. Then the multiplicativity isomorphism (24.63) is defined by

$$\begin{aligned} &[e'_1, \dots, e'_{p'}|e'_{p'+1}, \dots e'_{p'+q'}] \otimes [e''_1, \dots, e''_{p''}|e''_{p''+1}, \dots, e''_{p''+q''}] \\ &\qquad \to [e'_1, \dots, e'_{p'}, \tilde{e}''_1, \dots, \tilde{e}''_{p''}|e'_{p'+1}, \dots e'_{p'+q'}, \tilde{e}''_{p''+1}, \dots, \tilde{e}''_{p''+q''}] \end{aligned} \tag{24.64}$$

Although we have chosen bases to define the isomorphism one can check that under changes of bases the isomorphism remains of the same form, so in this sense it is "natural."

Now in our discussion of integration over a superdomain the densities $\mathcal{D}^{p|q}$ on $\mathcal{R}^{p|q}$ can be recognized to be simply $\mathrm{Ber}(\Omega^1\mathcal{R}^{p|q})$. This behaves well under coordinate transformations and so defines a sheaf on $\mathcal{M}_{\mathrm{red}}$ on a supermanifold, and so:

The analog of a density for a manifold is a global section of $\mathrm{Ber}(\Omega^1\mathcal{M})$ on a supermanifold $\mathcal{M}$.

In supersymmetric field theories, this is the kind of quantity we should be integrating. ♣Should complete the parallel discussion by Berezinian line for linear transformation between modules etc. ♣

omalyComments

24.7 Brief Comments on fermionic path integrals and anomalies

24.7.1 General Considerations

♣☡Warning: This section assumes a knowledge of many things we have yet covered above. It is really out of place here. ♣

Determinant lines are of great importance in quantum field theory, especially in the theory of anomalies.

Typically one has a fermionic field $\psi(x)$ and a Dirac-like-operator and the path integral involves an expression like

$$\int [d\psi d\bar{\psi}]\exp[\int i\bar{\psi}D\psi] \tag{24.65}$$

where $\int[d\psi d\bar{\psi}]$ is formally an infinite-dimensional version of the Berezin integral. At least formally this should be the determinant of D.

However, it is often the case that D is an operator between two different spaces, e.g. on an even-dimensional spin manifold M the chiral Dirac operator is an operator

$$D: L^2(M, S^+\otimes E)\to L^2(M, S^-\otimes E) \tag{24.66}$$

where $S^{\pm}$ are the chiral spin bundles on M and E is a bundle with connection (in some representation of the gauge group). There is no canonical way of relating bases for these two Hilbert spaces, so $\mathrm{det}D$ must be an element of the determinant line $\mathrm{DET}(D)$.

If we have families of Dirac operators parametrized, say, by gauge fields, then we have a determinant line bundle $\mathrm{DET}(D)$ over that family, and $\mathrm{det}D$ is a section of that line bundle. If D is Fredholm then we can still define

$$\mathrm{DET}(D) = \Lambda^{mx}(\mathrm{ker}D)^\vee\otimes\Lambda^{mx}(\mathrm{cok}D) \tag{24.67}$$

♣Some of above paragraph now redundant ♣

The above remarks have implications for the theory of anomalies. In particular the *geometrical theory of anomalies* due to Atiyah and Singer.

In a general Lagrangian quantum field theory the path integral might look like

$$Z\sim\int_{\mathcal{B}}[d\phi]\int[d\psi d\bar{\psi}]\exp[S_{\mathrm{bosonic}}(\phi)+\int\bar{\psi}D_\phi\psi+S_{\mathrm{interaction}}] \tag{24.68}$$

where $\mathcal{B}$ is some space of bosonic fields. For example it might consist of the set of maps from a worldvolume $\mathcal{W}$ to a target manifold M, in the case of a nonlinear sigma model, or it might be the set of gauge equivalence classes of gauge fields, or some combination of these ingredients. Note that the Dirac-like operator D_ϕ typically depends on ϕ and $S_{\mathrm{interaction}}$ here indicates interactions of higher order in the fermions. Let us suppose that the interactions between bosons and fermions can be handled perturbatively. If we first integrate over the fermions then we obtain an expression like

$$Z \sim \int_{\mathcal{B}} [d\phi] \exp[S_{\text{bosonic}}(\phi)] \det D_\phi \tag{24.69}$$

where $\det D_\phi$ is a section of a line bundle $DET(D_\phi)$ over $\mathcal{B}$, rather than a $\mathbb{C}^*$-valued function on $\mathcal{B}$.

This expression is meaningless, even at the most formal level, unless the line bundle has been trivialized with a trivial flat connection.

Remarks

1. In physical examples there is a natural connection on $\mathrm{DET}(D_\phi)$ and it turns out that the vanishing of the curvature is the vanishing of the "perturbative anomaly."
2. If the perturbative anomaly is nonzero then the path integral does not even make formal sense. Of course, there are many other demonstrations using techniques of local field theory that the theory is ill-defined. But this is one elegant way to understand that fact.
3. There can be anomaly-canceling mechanisms. One of the most interesting is the *Green-Schwarz mechanism*. In this method one introduces an action $\exp[S_{GS}[\phi]]$ which is actually not a well-defined function from $\mathcal{B} \to \mathbb{C}^*$, but rather is a section of a line bundle over $\mathcal{B}$, where $\mathcal{B}$ is the space of bosonic fields. If the line bundle with connection is dual to that of the fermions so that $\mathcal{L}_{GS} \otimes \mathrm{DET}(D_\phi)$ has a flat trivialization then $e^{S_{GS}[\phi]} \det D_\phi$ can be given a meaning as a well-defined function on $\mathcal{B}$. Then it can - at least formally - be integrated in the functional integral.
4. Even if the perturbative anomalies cancel, i.e. the connection is flat, if the space of bosonic fields is not simply connected then the flat connection on the determinant line can have nontrivial holonomy. This is the "global anomaly."

♣Actually, the general mechanism was known before GS, but GS is a very nice example of the general idea. ♣

In Section §24.7.3 below we will give perhaps the simplest illustration of a global anomaly.

24.7.2 Determinant of the one-dimensional Dirac operator

As a warmup, let us consider the odd-dimensional Dirac operator on the circle. This maps $L^2(S^1, S) \to L^2(S^1, S)$ where S is the spin bundle on the circle. There are two spin structures. The tangent bundle is trivial and hence we can simply think of spinors as complex functions on the circle with periodic or anti-periodic boundary conditions. So, concretely, we identify the Dirac operator coupled to a real $U(1)$ gauge field as

$$D_a = \frac{d}{dt} + ia(t) \tag{24.70}$$

where $t \sim t+1$ is a coordinate on S^1, $a(t) \in \mathbb{R}$ is periodic and identified via gauge transformations, and D acts on complex-valued functions which are periodic or antiperiodic.

The operator D_a is Fredholm and maps $\mathcal{H} \to \mathcal{H}$, $\det D_a$ as a complex number. The first simplification we can make is that we can gauge $a(t)$ to be constant. But we cannot remove the constant by a well-defined gauge transformation since $\oint a(t)dt$ is gauge invariant. Once this is done the eigenfunctions are $e^{2\pi i(n+\epsilon)t}$ where $n \in \mathbb{Z}$ and $\epsilon = \frac{1}{2}$ for AP and $\epsilon = 0$ for P boundary conditions. The eigenvalue is then $2\pi i(n+\epsilon) + ia$. Note that we can account for both boundary conditions by shifting $a \to a \pm 2\pi\epsilon$ so we will temporarily set $\epsilon = 0$ to simplify the formulae.

We could proceed by evaluating the ζ function. However, it is actually easier to proceed formally as follows:

$$\det D_a \stackrel{?}{=} \prod_{n\in\mathbb{Z}} (2\pi i n + ia) \tag{24.71}$$

It is good to rearrange this formal expression by putting together the positive and negative integers and separating out the $n = 0$ term:

$$\begin{aligned} \prod_{n\in\mathbb{Z}} (2\pi i n + ia) &= ia \prod_{n=1}^{\infty} (2\pi i n + ia)(-2\pi i n + ia) \\ &= (ia) \prod_{n=1}^{\infty} (2\pi n)^2 (1 - \frac{a^2}{(2\pi n)^2}) \\ &= \left(\prod_{n=1}^{\infty} (2\pi n)^2 \right) \left(ia \prod_{n=1}^{\infty} (1 - \frac{a^2}{(2\pi n)^2}) \right) \end{aligned} \tag{24.72}$$

Note that we can separate out the infinite factor $\prod_{n=1}^{\infty}(2\pi n)^2$, and what remains contains all the a-dependence and is in fact a well-defined product so we have:

$$\det(D_a) = (ia) \prod_{n=1}^{\infty} (1 - \frac{a^2}{(2\pi n)^2}) = 2i \sin(a/2) \tag{24.73}$$ eq:DetCF

where we have used the famous formula

$$\prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) = \frac{\sin \pi z}{\pi z} \tag{24.74}$$

This is the result for P-boundary conditions. The result for AP-boundary conditions is obtained by shifting $a \to a + \pi$

Remark: Using ζ-function regularization the "constant" $\left(\prod_{n=1}^{\infty}(2\pi n)^2\right)$ can be argued to be in fact $= 1$ as follows:

$$\begin{aligned}\prod_{n=1}^{\infty}(2\pi n)^2 &= \exp[\sum_{n=1}^{\infty}\log(2\pi n)^2] \\ &= \exp[-\frac{d}{ds}|_{s=0}\sum_{n=1}(2\pi n)^{-2s}] \\ &= \exp[-\frac{d}{ds}|_{s=0}(2\pi)^{-2s}\zeta(2s)] \\ &= \exp[2\log(2\pi)\zeta(0) - 2\zeta'(0)] \\ &= 1\end{aligned} \tag{24.75}$$

where in the last line we used the expansion of the Riemann zeta function around $s = 0$:

$$\zeta(s) = -\frac{1}{2} - s\log\sqrt{2\pi} + \mathcal{O}(s^2) \tag{24.76}$$

The first equality above is formal. The second is a definition. The remaining ones are straightforward (and rigorous) manipulations.

24.7.3 A supersymmetric quantum mechanics

-SpinStructure

Now, let us consider a supersymmetric quantum mechanics with target space given by a Riemannian manifold $(M, g_{\mu\nu})$. If x^μ, $\mu = 1, \dots, n = \dim M$, are local coordinates in a patch of M then there are maps from the one-dimensional worldline $(x^\mu(t), \psi^\mu(t))$. We can think of $(x^\mu(t), \psi^\mu(t))$ as functions on the supermanifold ΠTM and we have a map from $\mathbb{R}^{1|0} \to \Pi TM$. The action is:

$$S = \int dt\{g_{\mu\nu}(x(t))\dot{x}^\mu\dot{x}^\nu + i\psi^a[\frac{d}{dt}\delta^{ab} + \dot{x}^\mu(t)\omega_\mu^{ab}(x(t))]\psi^b\} \tag{24.77}$$

eq:SQM

Here $g_{\mu\nu}$ is a Riemannian metric on M and $\omega_\mu^{ab}dx^\mu$ is a spin connection. The $a, b = 1, \dots, n$ refer to tangent space indices.

Let us consider just the theory of the fermions in a fixed bosonic background.

Consider briefly the Hamiltonian quantization of the system. Classically the bosonic field is just a point $x_0 \in M$. The canonical quantization relations on the fermions ψ^a gives a real Clifford algebra:

$$\{\psi^a, \psi^b\} = 1 \tag{24.78}$$

If $n = 2m$ then the irreducible representations are the chiral spin representations of M. Therefore, the wavefunctions of the theory are sections of a spinor bundle over M. Therefore, the Hilbert space is $L^2(M; S)$, the L^2 sections of the spin bundle over M. Therefore, if the theory is sensible, M should be spin. It is interesting to see how that constraint arises just by considering the path integral on the circle.

Let us consider the path integral on the circle, so $t \sim t + 1$. Again, we focus on the fermionic part of the path integral, so fix a loop $x : S^1 \to M$.

The fermionic path integral gives, formally pfaff(D_A) where

$$D_A = i\left(\frac{d}{dt}\delta^{ab} + A^{ab}(t)\right) \tag{24.79}$$

eq:DeeA

where $A^{ab}(t)$ is the real so$(2m)$ gauge field

$$A^{ab}(t) = \dot{x}^{\mu}(t)\omega_{\mu}^{ab}(x(t)). \tag{24.80}$$ eq:LoopGF

It is very useful to generalize the problem and consider the operator D_A in (24.79) for an *arbitrary* $SO(2m)$ gauge field $A^{ab}(t)$, with $a, b = 1, \dots, 2m$. So we consider the Berezin integral

$$Z = \int [d\psi^a(t)] e^{\int_0^1 i\psi^a(t)\left(\frac{d}{dt}\delta^{ab} + A^{ab}(t)\right)\psi^b(t)dt} \tag{24.81}$$

Formally, this is just pfaff(D_A). In this infinite-dimensional setting we have two approaches to defining it:

1. We can consider the formal product of skew eigenvalues and regularize the product, say, using ζ-function regularization. Then we must choose a sign for each skew eigenvalue.

2. We can evaluate the determinant and attempt to take a squareroot.

We will explain (2), leaving (1) as an interesting exercise.

24.7.4 Real Fermions in one dimension coupled to an orthogonal gauge field

So, we want to evaluate $\det(D_A)$, where $D_A : L^2(S^1, \mathbb{R}^{2m}) \to L^2(S^1, \mathbb{R}^{2m})$. (Note we have complexified our fermions by doubling the degrees of freedom to compute the determinant.)

All the gauge invariant information is in $P\exp \oint A(t)dt \in SO(2m)$. By a constant orthogonal transformation the path ordered exponent can be put in a form

$$P\exp \oint A(t)dt = R(\alpha_1) \oplus R(\alpha_2) \oplus \cdots \oplus R(\alpha_m) \tag{24.82}$$ eq:Cartan

and by a single-valued gauge transformation $A^{ab}(t)$ can be gauged to a form which is t-independent. Recall that

$$R(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \tag{24.83}$$

so that the gauge invariant information only depends on $\alpha_i \sim \alpha_i + 2\pi$.

Therefore using gauge transformations we can reduce the problem of evaluating $\det(D_A)$ to the evaluation of the determinant of

$$D_\alpha = \frac{d}{dt} + \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix} \tag{24.84}$$ eq:RealD

We can diagonalize the matrix and hence we get the Dirac operator

$$D_\alpha = \frac{d}{dt} + \begin{pmatrix} i\alpha & 0 \\ 0 & -i\alpha \end{pmatrix} \tag{24.85}$$

Now, using the result (24.73) above we learn that

$$(\text{Pfaff}\, D_\alpha)^2 = 4\sin^2(\alpha/2) = \det(1 - R(\alpha)) \tag{24.86}$$ eq:twotwo

In general, for the antisymmetric operator D_A (24.79) coupled to any $SO(2m)$ gauge field on the circle we have

$$(\mathrm{Pfaff} D_A)^2 = \det(1 - \mathrm{hol}(A)) \tag{24.87}$$

Now we would like to take a square root of the determinant to define the Pfaffian. Let us consider a family of operators parametrized by $g \in SO(2m)$ with $P\exp \oint A dt = g$. Then $\det(1-g)$ is a function on $SO(2m)$ which is conjugation invariant, and hence we can restrict to the Cartan torus (24.82). It is clear that this does not have a well-defined square-root. If we try to take $\prod_{i=1}^m 2\sin(\alpha_i/2)$ then the expression has an ill-defined sign because we identify $\alpha_i \sim \alpha_i + 2\pi$. The expression *does* have a good meaning as a section of a principal $\mathbb{Z}_2$ bundle over $SO(2m)$. Put differently, if we pull back the function to $Spin(2m)$:

$$1 \to \mathbb{Z}_2 \to Spin(2m) \to SO(2m) \to 1 \tag{24.88}$$

then the function $\det(1-\tilde{g})$ (where we take the determinant in the $2m$-dimensional representation) does have a well-defined square-root. To see this it suffices to work again with the Cartan torus since the functions are conjugation invariant and therefore we need really only consider

$$1 \to \mathbb{Z}_2 \to Spin(2) \to SO(2) \to 1 \tag{24.89}$$

The group $Spin(2)$ is the group of even invertible elements

$$r(\beta) := \exp[\beta e_1 e_2/2] = \cos(\beta/2) + \sin(\beta/2) e_1 e_2 \tag{24.90}$$

in the Clifford algebra $C\ell_{+2}$. Here $\beta \sim \beta + 4\pi$. The projection to $SO(2)$ is given by $r(\beta) \mapsto R(\alpha)$ with $\alpha = \beta$, but a full period in α only lifts to a half-period in β. So on $Spin(2m)$ we have

$$\sqrt{\det(1-\tilde{g})} = \prod_{i=1}^m (2\sin\beta_i/2) \tag{24.91}$$

and this expression is well-defined.

Remark: In fact, this expression has a nice interpretation in terms of the characters in the chiral spin representations:

$$\prod_{i=1}^m (2\sin\beta_i/2) = i^{-m}\left(\mathrm{ch}_{S^+}(\tilde{g}) - \mathrm{ch}_{S^-}(\tilde{g})\right) \tag{24.92}$$

24.7.5 The global anomaly when M is not spin

Let us now return to our supersymmetric quantum mechanics above. We have learned that after integrating out the fermions the path integral on the circle is

$$Z(S^1) = \int [dx^\mu(t)] e^{-\int_0^1 dt g_{\mu\nu}(x(t))\dot{x}^\mu \dot{x}^\nu} \sqrt{\det(1 - \mathrm{Hol}(x^*\omega))} \tag{24.93}$$

where now for a given loop in the manifold $x: S^1 \to M$ we are using the holonomy of the orthogonal gauge field (24.80).

The question is: *Can we consistently define the sign of the square root over all of loop space $LM = Map(S^1 \to M)$?*

Let us fix a point $x_0 \in M$ and choose a basis for the tangent space at x_0. Then consider based loops $x(t)$ that begin and end at x_0. Then $\mathrm{Hol}(x^*\omega)$ defines a map from based loops $\Omega_{x_0}(M) \to SO(2m)$. If M is a spin manifold then there is a well-defined lift of this map to $Spin(2m)$. Then a well-defined square-root exists.

♣Put in commutative diagram ♣

On the other hand, it can be shown using topology that if M is not spin then there will be a family of based loops: $x^\mu(t;s) = x^\mu(t;s+1)$ so that at $s = 0$ we have a constant map to a point $x^\mu(t;0) = x_0^\mu \in M$ and $x^\mu(t;s)$ loops around a nontrivial 2-sphere in M such that

♣Picture of the lasso of the 2-sphere ♣

$$\sqrt{\det(1 - \mathrm{Hol}(x^*\omega))}|_{s=0} = -\sqrt{\det(1 - \mathrm{Hol}(x^*\omega))}|_{s=1} \qquad (24.94)$$

Thus, if M is not spin, the fermionic path integral in the SQM theory (24.77) cannot be consistently defined for all closed paths $x^\mu(t)$ and therefore the path integral does not makes sense. This is an example of a global anomaly.

Remark: Recall the exercise from §24.4. Apply this to the 2-form on LM defined by:

$$\omega = \oint_0^1 dt \delta x^a(t) \left(\frac{d}{dt}\delta^{ab} + \dot{x}^\mu(t)\omega_\mu^{ab}(x(t)) \right) \delta x^b(t) \qquad (24.95)$$

One can argue that ω is closed and nondegenerate, hence it is a symplectic form. Then we can interpret the above remarks as the claim that *A manifold M is spin iff LM is orientable.*

24.7.6 References

1. M.F. Atiyah, "Circular Symmetry and the Stationary Phase Approximation," Asterisque

2. Atiyah and Singer, PNAS

3. O. Alvarez, I.M. Singer, and B. Zumino

4. G. Moore and P. Nelson, "Aetiology of sigma model anomalies,"

5. D. Freed, "Determinants, Torsion, and Strings," Commun. Math. Phys. **107**(1986)483

6. D. Freed and G. Moore, "Setting the Quantum Integrand of M Theory," hep-th/0409135.

25. Quadratic forms and lattices

Lattices show up in many ways in physics. The study of lattices in 2 and 3 dimensions is quite useful in solid state physics, in part because the types of atoms in a crystal are (by definition of a crystal) invariant under translation by a lattice. Higher dimensional lattices have also played a role in solid state physics, in the context of "quasicrystals" and also in the classification of quantum Hall states.

In this course we will encounter many very symmetric higher dimensional lattices as root lattices and weight lattices of Lie algebras. These encode many important aspects of the representation theory of compact Lie groups.

It turns out that many special lattices, such as even unimodular lattices play a distinguished role in string theory and in conformal field theory through vertex operator constructions. Lattices also play an important role in the study of compactifications of string theory.

Lattices of charges are again of importance in studying duality symmetries in supersymmetric quantum field theory and string theory.

In math lattices are studied for their own sake, as very beautiful objects, but they also have far flung connections to other branches of math such as number theory and error-correcting codes. See Conway & Sloane, *Sphere Packings, Lattices, and Groups* for a comprehensive survey. Finally, they are also important in topology (as intersection forms), especially in the topology of four-manifolds.

25.1 Definition

The word "lattice" means different things to different people. For some people it is a finitely generated free abelian group. In this case, up to isomorphism there is just one invariant: the rank, and any "lattice" in this sense is just $\mathbb{Z}^r$, up to isomorphism. To some people it is a regular array of points in some space. In these notes we take a "lattice" to be a finitely generated abelian group with the extra data of a nondegenerate symmetric bilinear form: [48]

Definition A *lattice* Λ is a free abelian group equipped with a nondegenerate, symmetric bilinear quadratic form:

$$\langle \cdot, \cdot \rangle : \Lambda \times \Lambda \to R \tag{25.1}$$

where R is a $\mathbb{Z}$-module.

Thus:

1. $\langle v_1, v_2 \rangle = -\langle v_2, v_1 \rangle, \qquad \forall v_1, v_2 \in \Lambda.$
2. $\langle n v_1 + m v_2, v_3 \rangle = n \langle v_1, v_3 \rangle + m \langle v_2, v_3 \rangle, \qquad \forall v_1, v_2, v_3 \in \Lambda$, and $n, m \in \mathbb{Z}$.
3. $\langle v, w \rangle = 0$ for all $w \in \Lambda$ implies $v = 0$.

When $R = \mathbb{Z}$ we say we have an *integral lattice*. We will also consider the cases $R = \mathbb{Q}, \mathbb{R}$.

We say that two lattice $(\Lambda_1, \langle \cdot, \cdot \rangle_1)$ and $(\Lambda_2, \langle \cdot, \cdot \rangle_2)$ are *equivalent* if there is a group isomorphism $\phi : \Lambda_1 \to \Lambda_2$ so that $\phi^*(\langle \cdot, \cdot \rangle_2) = \langle \cdot, \cdot \rangle_1$. The *automorphism group* of the lattice is the group of ϕ's which are isomorphisms of the lattice with itself. These can be a finite or infinite discrete groups and can be very interesting.

[48] Hence, sometimes the term "quadratic module" is used.

There is a simple way of thinking about lattices in terms of matrices of integers. A (finitely generated) free abelian group of rank n is isomorphic, as a group, to $\mathbb{Z}^n$. Therefore, we can choose an ordered integral basis $\{e_i\}_{i=1}^n$ for the lattice (that is, a set of generators for the abelian group) and then define the $n \times n$ *Gram-Matrix*

$$G_{ij} := \langle e_i, e_j \rangle. \tag{25.2}$$

Of course, the basis is not unique, another one is defined by

$$\tilde{e}_i := \sum_j S_{ji} e_j \tag{25.3}$$

eq:chgbss

where, now, the matrix S must be both *invertible* and *integral valued*, that is

$$S \in GL(n, \mathbb{Z}) \tag{25.4}$$

eq:intgrlequi

Under the change of bases (25.3) the Gram matrix changes to

$$G \to \tilde{G} = S^{tr} G S \tag{25.5}$$

eq:intgrlequi

So lattices can be thought of as symmetric nondegenerate matrices of integers with an equivalence relation given by (25.5).

Remark. As we will soon begin to see, the classification of lattices is somewhat nontrivial. In fact, it is an extremely subtle and beautiful problem, only partially solved.

By contrast, the classification of integral antisymmetric forms is fairly straightforward. Any such form can be brought by an integral transformation to the shape: [49]

$$\begin{pmatrix} 0 & d_1 \\ -d_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & d_2 \\ -d_2 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & d_3 \\ -d_3 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & d_k \\ -d_k & 0 \end{pmatrix} \tag{25.6}$$

eq:antisymm

and this form is unique if we require $d_i > 0$ and $d_1 | d_2 | \cdots | d_k$. This is important in the quantization of certain mechanical systems with compact coordinates and momenta. It classifies the integral symplectic forms on a torus, for example.

Exercise

a.) Show that the group of integer invertible matrices whose inverses are also integer matrices form a group. This is the group $GL(n, \mathbb{Z})$.

Note that it is *not* the same as the set of integer matrices which are invertible. For example

$$\begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix} \in GL(2, \mathbb{Z}) \tag{25.7}$$

[49] For a proof see Lang, Algebra, p. 380.

but

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \notin GL(2,\mathbb{Z}) \tag{25.8}$$

b.) Show that for $S \in GL(n,\mathbb{Z})$, we necessarily have $|\mathrm{det} S| = 1$.

c.) $SL(n,\mathbb{Z})$ is the subgroup of matrices of determinant 1. What is the center of $SL(n,\mathbb{Z})$?

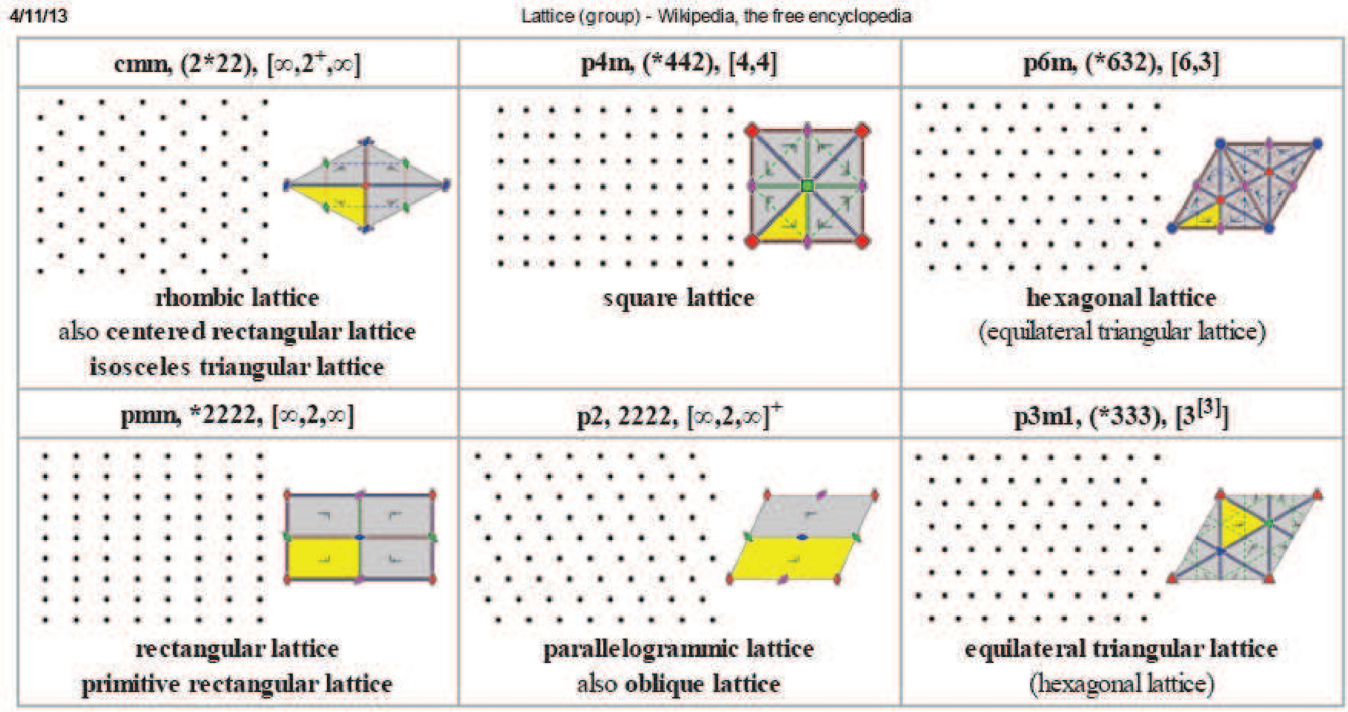

Figure 17: A picture of some important two-dimensional lattices embedded into Euclidean $\mathbb{R}^2$. From Wikipedia. fig:latticeii

25.2 Embedded Lattices

Quite often we do not think of lattices in the above abstract way but rather as a discrete subgroup of $\mathbb{R}^m$. [50] See for example Figure 17, above for some rank 2 lattices in $\mathbb{R}^2$ and Figures 18 and 19 for some embedded lattices in $\mathbb{R}^3$.

To describe an embedded lattice we can consider the generators to be linearly independent vectors $\vec{e}_1, \cdots, \vec{e}_n \in \mathbb{R}^m$. (Necessarily, $m \geq n$). Define

$$\Lambda \equiv \{\sum_{i=1}^{n} \ell_i \vec{e}_i | \ell_i \in \mathbb{Z}\}. \tag{25.9}$$ eq:lattice

[50] By a *discrete subgroup* we mean, heuristically, that there are no accumulation points. Technically, the action on G should be properly discontinuous.

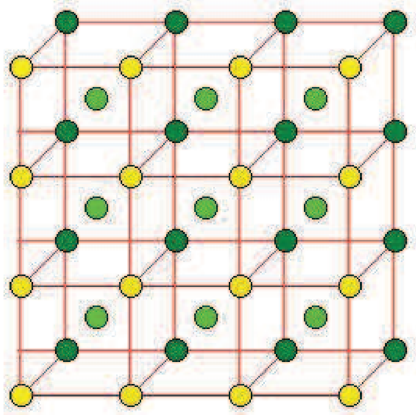

Figure 18: A three-dimensional lattice, known as the body centered cubic lattice. fig:BodyCenter

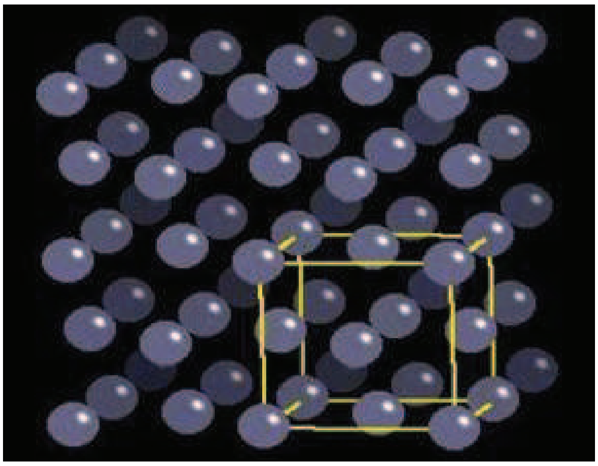

Figure 19: A three-dimensional lattice, known as the face centered cubic lattice. fig:FCC-Lattic

Under vector addition Λ is isomorphic to $\mathbb{Z}^n$. Moreover, if $\mathbb{R}^m$ is equipped with a symmetric quadratic form (e.g. the Euclidean metric) then the lattice inherits one:

$$\langle\cdot,\cdot\rangle : \Lambda \times \Lambda \to \mathbb{R} \tag{25.10}$$

Abstractly, when embedding a lattice into $\mathbb{R}^n$ we are using the simple tensor product

$$\Lambda \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathbb{R}^n \tag{25.11}$$

If the coordinates of the vectors are $\vec{e}_i = (e_{i1}, \ldots, e_{im})$ then we can form an $n \times m$ *generating matrix*

$$M = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ \vdots & \vdots & & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix} \tag{25.12}$$

The lattice is the span of vectors ξM where $\xi \in \mathbb{Z}^n$. If we use the Euclidean metric on $\mathbb{R}^m$ to induce the bilinear form on Λ then the Gram-Matrix is $G = MM^{tr}$. Different generating matrices are related by $M \mapsto S^{tr}M$, for $S \in GL(n, \mathbb{Z})$.

Example 1: The most obvious example is $\Lambda = \mathbb{Z}^n \subset \mathbb{R}^n$. For $n = 2$ this is a square lattice, for $n = 3$ it is the simple cubic lattice. For general n we will refer to it as a "hypercubic lattice." The automorphisms must be given by integer matrices which are simultaneously in $O(n)$. Since the rows and columns must square to 1 and be orthogonal these are signed permutation matrices. Therefore

$$\mathrm{Aut}(\mathbb{Z}^n) = \mathbb{Z}_2^n \rtimes S_n \tag{25.13}$$ eq:AutZn

where S_n acts by permuting the coordinates $(x_1, \dots, x_n)$ and $\mathbb{Z}_2^n$ acts by changing signs $x_i \to \epsilon_i x_i$, $\epsilon_i \in \{\pm 1\}$.

Example 2 As a good example of the utility of allowing $m > n$ let us define:

$$A_n := \{(x_0, x_1, \dots, x_n) \in \mathbb{Z}^{n+1} | \sum_{i=0}^{n} x_i = 0\} \subset \mathbb{R}^{n+1} \tag{25.14}$$ eq:AnDef

A group of automorphisms of the lattice A_n is is rather obvious from (25.14), namely the symmetric group S_{n+1} acts by permutation of the coordinates. Another obvious symmetry is $(x_0, \dots, x_n) \mapsto (-x_0, \dots, -x_n)$. These generate the full automorphism group

$$\mathrm{Aut}(A_n) = \mathbb{Z}_2 \times S_{n+1} \qquad n > 1 \tag{25.15}$$

for $n = 1$, $A_1 \cong \sqrt{2}\mathbb{Z} \subset \mathbb{R}$ and the automorphism group is just $\mathbb{Z}_2$.

A nice basis is given by

$$\alpha_i = \vec{e}_i - \vec{e}_{i-1} \qquad\qquad i = 1, \dots, n \tag{25.16}$$ eq:Anbasis

where $\vec{e}_i$, $i = 0, \dots, n$, are the standard basis vectors in $\mathbb{R}^{n+1}$. The Gram matrix is then the famous Cartan matrix for A_n:

$$G_{ij} = C_{ij} = 2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1} \qquad\qquad i, j = 1, \dots, n \tag{25.17}$$

The corresponding matrix is tridiagonal:

$$C(A_n) = \begin{pmatrix} 2 & -1 & 0 & \cdots & \cdots & 0 \\ -1 & 2 & -1 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 2 & -1 & 0 \\ 0 & \cdots & \cdots & -1 & 2 & -1 \\ 0 & \cdots & \cdots & 0 & -1 & 2 \end{pmatrix} \tag{25.18}$$

Note that we could project two basis vectors of A_2 into a plane $\mathbb{R}^2$ to get

$$\begin{aligned} \alpha_1 &= \sqrt{2}(1, 0) \\ \alpha_2 &= \sqrt{2}(-\frac{1}{2}, \frac{\sqrt{3}}{2}) \end{aligned} \tag{25.19}$$ eq:atwo

and as a check we can use these to compute

$$C(A_2) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{25.20}$$

♣Give a figure and say how the projection is done. ♣

Using these vectors we generate a beautifully symmetric hexagonal lattice in the plane.

♣ FIGURE OF HEXAGONAL LATTICE HERE ♣

Similarly we have

$$C(A_3) = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \tag{25.21}$$

and we could also realize the lattice as the span of vectors in $\mathbb{R}^2$

$$\begin{aligned} \alpha_1 &= \\ \alpha_2 &= \\ \alpha_3 &= \end{aligned} \tag{25.22}$$

eq:atwo

Example 3 Consider the set of points $D_n \subset \mathbb{Z}^n$ defined by

$$D_n := \{(x_1, \ldots, x_n) \in \mathbb{Z}^n | x_1 + \cdots + x_n = 0 \mod 2\} \tag{25.23}$$

If we embed into $\mathbb{R}^n$ in the obvious way then we can use the Euclidean metric to induce an integer-valued quadratic form. To get some intuition, let us consider D_3. This is known as the "face-centered-cubic" or fcc lattice. To justify the name note that there is clearly a lattice proportional to the cubic lattice and spanned by $(2,0,0)$, $(0,2,0)$, and $(0,0,2)$. However in each xy, xz, and yz plane the midpoint of each 2×2 square is also a lattice vector. Choosing such midpoint lattice vectors in each plane gives a generating matrix:

$$M = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & -1 & 0 \end{pmatrix} \tag{25.24}$$

and one then computes

$$MM^{tr} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \tag{25.25}$$

so in fact $D_3 \cong A_3$. (This reflects a special isomorphism of simple Lie algebras.)

Example 4 The n-dimensional *bcc lattice*, BCC_n, where *bcc* stands for "body-centered cubic" is the sublattice of $\mathbb{Z}^n$ consisting of $(x_1, \ldots, x_n)$ so that the x_i are either all even or all odd. Note that if all the x_i are odd then adding $\vec{e}$ produces a vector with all x_i even, where $\vec{e} = (1, 1, \ldots, 1) = \vec{e}_1 + \cdots + \vec{e}_n$. Therefore, we can write:

$$BCC_n = 2\mathbb{Z}^n \cup (2\mathbb{Z}^n + \vec{e}) \tag{25.26}$$

Clearly $2\mathbb{Z}^n$ is proportional to the "cubic" lattice. Adding in the orbit of $\vec{e}$ produces one extra lattice vector inside each n-cube of side length 2, hence the name bcc.

Example 5: Now take $\mathbb{R}^2$ as a vector space but we do not use the Euclidean metric on $\mathbb{R}^2$ to induce the bilinear form on Λ. Rather we use the Minkowski signature metric

$$\mathbb{R}^{1,1} = \{(t,x) | \langle (t,x),(t,x)\rangle = -t^2 + x^2\} \tag{25.27}$$ eq:2DMink

Let $R > 0$, and consider the lattice $\Lambda(R)$ generated by

$$\begin{aligned} e_1 &= \frac{1}{\sqrt{2}}(\frac{1}{R}, \frac{1}{R}) \\ e_2 &= \frac{1}{\sqrt{2}}(-R, R) \end{aligned} \tag{25.28}$$ eq:nafi

Note that for any R we have simply:

$$G_{ij} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{25.29}$$

So, manifestly, as lattices these are all isomorphic, although as embedded lattices they depend on R.

♣ FIGURE OF e_1, e_2 IN THE PLANE ♣

Remarks: Solids comprised of a single element will form simple three-dimensional lattices in nature, at least in the limit that they are infinitely pure and large. Those on the LHS of the periodic table tend to be bcc e.g. the alkali metals of column one and Ba, Ra, while those towards the right tend to be fcc (e.g. Cu, Ni, Ag, Au, Pt, Ir, Al, Pb) or the column of noble gases (except He).

Exercise

If Λ is a lattice, let 2Λ be the lattice of elements divisible by 2, i.e., vectors $\vec{v}$ such that $\frac{1}{2}\vec{v} \in \Lambda$. Show that 2Λ is a subgroup of Λ. Suppose $\vec{v}$ is not divisible by 2. Is $2\Lambda + \vec{v}$ a subgroup?

Exercise *Automorphisms of $\mathbb{Z}^n$*

Check that the group of signed permutations is isomorphic to the semidirect product (25.13). Write explicitly $\alpha : S_n \to \mathrm{Aut}(\mathbb{Z}_2^n)$.

25.3 Some Invariants of Lattices

What can we say about the classification of lattices?

If we were allowed to take $S \in GL(n, \mathbb{R})$ then Sylvester's theorem guarantees that we can change basis to put the Gram matrix into diagonal form so that

$$\tilde{G}_{ij} = Diag\{-1^t, +1^s\} \tag{25.30}$$

This provides us with two important invariants of the lattice, the signature and the rank. The rank is

$$r = t + s \tag{25.31}$$

and we will define the signature to be

$$\sigma = s - t \tag{25.32}$$

When we work of $R = \mathbb{Z}$ there are going to be more invariants:

Example: Consider two lattices:

A. $\Lambda_A = \mathbf{e_1}\mathbb{Z} \oplus \mathbf{e_2}\mathbb{Z}$ with form:

$$G_A = \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \tag{25.33}$$

B. $\Lambda_B = \mathbf{e_1}\mathbb{Z} \oplus \mathbf{e_2}\mathbb{Z}$ with form:

$$G_B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{25.34}$$

We ask:

Can these be transformed into each other by a change of basis with $S \in GL(2, \mathbb{Z})$?

The answer is clearly "yes" over $\mathbb{R}$ because they both have Lorentzian signature.

The answer is clearly "no" over $\mathbb{Z}$ because the norm-square of any vector in Λ_B is even $(n_1 e_1 + n_2 e_2)^2 = 2n_1 n_2$, while this is not true of Λ_A.

The lattice Λ_B is an important lattice, it is denoted by $II^{1,1}$ or by $H(1)$. [51]

Definition. A lattice Λ is called an *even lattice* if, for all $\mathbf{x} \in \Lambda$

$$\langle \mathbf{x}, \mathbf{x} \rangle \in 2\mathbb{Z} \tag{25.35}$$

(Note: This does not preclude $\langle x, y \rangle$ being odd for $x \neq y$.) A lattice is called *odd* if it is not even.

Note that under $G \to S^{tr} G S$

$$(S^{tr} G S)_{ii} = \sum_k (S_{ki})^2 G_{kk} + 2 \sum_{k<j} G_{kj} S_{ki} S_{ji} \tag{25.36}$$

so if the diagonal elements of G are even in one basis then they are even in all bases.

[51] Some authors use the notation $U(1)$. We do not use this notation since it can cause confusion.

In order to describe our next invariant of lattices we need to introduce the dual lattice. Given a lattice Λ we can define the *dual lattice* [52]

$$\Lambda^* := \mathrm{Hom}_{\mathbb{Z}}(\Lambda, \mathbb{Z}) \tag{25.37}$$

eq:dualf

where Hom means a $\mathbb{Z}$-linear mapping.

As we have now seen several times, given the data of a bilinear form there is a $\mathbb{Z}$-linear map $\ell : \Lambda \to \Lambda^*$ defined by

$$\ell(v)(v') := \langle v, v' \rangle. \tag{25.38}$$

This has no kernel for a nondegenerate form and hence we can consider $\Lambda \subset \Lambda^*$ and so we may form:

$$D(\Lambda) := \Lambda^*/\Lambda \tag{25.39}$$

This abelian group is known as the *discriminant group*, or *glue group*.

Next we make Λ^* into a lattice by declaring ℓ to be an isometry onto its image:

$$\langle v, w \rangle_\Lambda = \langle \ell(v), \ell(w) \rangle_{\Lambda^*} \tag{25.40}$$

eq:isometry

We then extend to the rest of Λ^* to make it a lattice,

To make this more concrete suppose e_i is a basis for Λ and let $\hat{e}^i$ be the dual basis for Λ^* so that

$$\hat{e}^i(e_j) = \delta^i{}_j \tag{25.41}$$

It follows that

$$\ell(e_i) = \sum_j G_{ij} \hat{e}^j \tag{25.42}$$

Now, using (25.40) it follows that Λ^* has the Gram matrix

$$\langle \hat{e}^i, \hat{e}^j \rangle = G^{ij} \tag{25.43}$$

where $G^{ij} G_{jk} = \delta^i{}_k$.

Note that in general Λ^* is not an integral lattice since G^{ij} will be a rational matrix if G_{ij} is an integral matrix. Let us denote

$$g := \mathrm{det} G_{ij} \tag{25.44}$$

$$G^{ij} = \frac{1}{g} \widehat{G}^{ij} \tag{25.45}$$

where $\widehat{G}^{ij}$ is a matrix of integers.

Lemma The discriminant group is a *finite* abelian group of order g

[52]The lattice Λ^* is closely related to the *reciprocal lattice* of solid state physics. However, there are some differences. Conceptually, the reciprocal lattice is an embedded lattice, embedded in momentum space. Moreover, there are some normalization differences by factors of 2π. More importantly, the reciprocal lattice depends on things like lattice spacings.

Proof: To see that it is finite we note that

$$g\hat{e}^j = \ell(\widehat{G}^{ij} e_j) \tag{25.46}$$

and hence $[g\hat{e}^j] = 0$ in the discriminant group. Therefore, every element is torsion and hence the group is finite. By the classification of finite abelian groups we see that the order $|D(\Lambda)|$ divides g.

In fact $|D(\Lambda)| = g$ as the following argument shows:

If $\Lambda \subset \mathbb{R}^n$ is an embedded lattice of maximal rank, and both the dual pairing and the Gram-matrix are inherited from the standard Euclidean bilinear form on $\mathbb{R}^n$ and we may write:

$$\begin{aligned} \hat{e}^i &= \sum_j G^{ij} e_j \\ e_i &= \sum_j G_{ij} \hat{e}^j \end{aligned} \tag{25.47}$$

eq:dualbs

Now use the notion of fundamental domain defined in the next chapter. By comparing the volume of a unit cell of Λ^* to that of Λ we find:

$$|D(\Lambda)| = \frac{\sqrt{\det G_{ij}}}{\sqrt{\det G^{ij}}} = \det G_{ij} \tag{25.48}$$

concluding the proof ♠

♣Would be better to give an argument that does not use the fundamental domain. ♣

Moreover, $D(\Lambda)$ inherits a bilinear form valued in $\mathbb{Q}/\mathbb{Z}$. (Recall that an abelian group is a $\mathbb{Z}$-module and one can define bilinear forms modules over a ring.) Specifically, we define

$$b\left([v_1], [v_2]\right) = \langle v_1, v_2 \rangle \mathrm{mod}\mathbb{Z} \tag{25.49}$$

The finite group $D(\Lambda)$ together with its bilinear form to $\mathbb{Q}/\mathbb{Z}$ is an invariant of the lattice.

Example 1: Consider $\Lambda = \nu\mathbb{Z} \subset \mathbb{R}$. We use the standard Euclidean metric on $\mathbb{R}$ so that ν^2 must be an integer n. Then $\Lambda^* = \frac{1}{\nu}\mathbb{Z}$. Note that $\Lambda \subset \Lambda^*$, indeed, $D(\Lambda) \cong \mathbb{Z}/n\mathbb{Z}$ so

$$[\Lambda^* : \Lambda] = n \tag{25.50}$$

There are only two choices of basis for Λ, namely $\mathbf{e_1} = \pm\nu$. The Gram matrix is $G_{11} = \nu^2 = n$. The bilinear form on the discriminant group is

$$b\left(\frac{r}{\nu} + \nu\mathbb{Z}, \frac{s}{\nu} + \nu\mathbb{Z}\right) = \frac{rs}{n}\mathrm{mod}\mathbb{Z} \tag{25.51}$$

Example 2: A_1^*. The Gram matrix is just the 1×1 matrix 2 so if $A_1 = \mathbb{Z}\alpha$ then $A_1^* = \mathbb{Z}\lambda$, with $\lambda = \frac{1}{2}\alpha$. The discriminant group is clearly $\mathbb{Z}_2$.

Example 3: A_2^*: Consider the Cartan matrix

$$C(A_2) = G_{ij} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{25.52}$$ eq:gmtx

and $\det G_{ij} = 3$.

We easily compute

$$G^{ij} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{25.53}$$ eq:gmtxop

and hence the dual basis is given by

$$\begin{aligned} \lambda^1 &= \frac{1}{3}(2\alpha_1 + \alpha_2) \\ \lambda^2 &= \frac{1}{3}(\alpha_1 + 2\alpha_2) \end{aligned} \tag{25.54}$$ eq:dualbas

The group $\Lambda^*/\Lambda \cong \mathbb{Z}_3$. Since $\alpha_1 = 2\lambda^1 - \lambda^2, \alpha_2 = -\lambda^1 + 2\lambda^2$, $2\lambda^1 = \lambda^2 \mathrm{mod}\Lambda$. So one set of representatives is given by $0\mathrm{mod}\Lambda, \lambda^1\mathrm{mod}\Lambda, 2\lambda^1\mathrm{mod}\Lambda$. Alternatively, we could take $\lambda^2\mathrm{mod}\Lambda$ as the generator.

If we take the embedding *** above then

$$\begin{aligned} \lambda^1 &= (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{6}}) \\ \lambda^2 &= (0, \sqrt{\frac{2}{3}}) \end{aligned} \tag{25.55}$$ eq:dualbasp

generate a triangular lattice.

♣ FIGURE ♣

As we shall see, the hexagonal lattice Λ is the root lattice of $SU(3)$, while Λ^* is the weight lattice.

Example 4: A_n^*. One could just invert the Cartan matrix and proceed as above. (See exercise below.) However, an alternative route is to view A_n as embedded in $\mathbb{Z}^{n+1} \otimes \mathbb{R}$ as above. Then relative to the basis α_i (25.16) above we will find a dual basis:

$$\lambda^i \cdot \alpha_j = \delta^i{}_j. \tag{25.56}$$

Writing out the equation in components one easily finds (and even more easily checks):

$$\lambda^1 = \left(-\frac{n}{n+1}, \underbrace{\frac{1}{n+1}, \dots, \frac{1}{n+1}}_{n \text{ times}} \right) \tag{25.57}$$

Now that we have λ^1 it is also easy to solve for the λ^i, $i > 1$ in terms of λ^1 from:

$$\begin{aligned} \alpha_1 &= 2\lambda^1 - \lambda^2 \\ \alpha_2 &= -\lambda^1 + 2\lambda^2 - \lambda^3 \\ \vdots \quad & \quad \vdots \\ \alpha_n &= -\lambda^{n-1} + 2\lambda^n \end{aligned} \tag{25.58}$$

to get

$$\begin{aligned} \lambda^2 &= 2\lambda^1 - \alpha_1 \\ \lambda^3 &= 3\lambda^1 - 2\alpha_1 - \alpha_2 \\ \lambda^4 &= 4\lambda^1 - 3\alpha_1 - 2\alpha_2 - \alpha_1 \\ \vdots \quad & \quad \vdots \\ \lambda^{n+1} &= (n+1)\lambda^1 - n\alpha_1 - (n-1)\alpha_2 - \cdots - \alpha_n \end{aligned} \tag{25.59}$$

and explicit substitute of the vectors shows that $\lambda^{n+1} = 0$.

Thus

$$\lambda^i = \left(\underbrace{-\frac{i}{n+1}, \dots, -\frac{i}{n+1}}_{j \text{ times}}, \underbrace{\frac{j}{n+1}, \dots, \frac{j}{n+1}}_{i \text{ times}} \right) \tag{25.60}$$

where $i + j = n + 1$. ♣CHECK THIS!! ♣

Thus, the discriminant group is cyclic,

$$D(A_n) \cong \mathbb{Z}/(n+1)\mathbb{Z} \tag{25.61}$$

and is generated, for example, by $[\lambda^1]$. Therefore, to compute the quadratic form it suffices to compute

$$b([\lambda^1], [\lambda^1]) = -\frac{1}{n+1} \mathrm{mod} \mathbb{Z} \tag{25.62}$$

The inverse Cartan matrix is given in the exercise below.

Example 5: D_n^*: We claim the dual lattice of the "n-dimensional fcc lattice" is one half of the "n-dimensional bcc lattice":

$$D_n^* = \frac{1}{2} BCC_n \tag{25.63}$$

eq:fcc-dual

To see this note first that if $x \in BCC_n$ and $y \in D_n$ then $x \cdot y$ is even. This is obvious if all the x_i are even, and if they are all odd then $\sum_i x_i y_i = \sum_i (2n_i + 1) y_i = \sum_i y_i = 0 \mathrm{mod} 2$, by the definition of D_n. Therefore $\frac{1}{2} BCC_n \subset D_n^*$. Conversely, if $v \in D_n^*$ then $2v_i$ must be integer, since $2e_i \in D_n$, and moreover, looking at the products with $(1, -1, 0, \dots, 0)$, $(0, 1, -1, 0, \dots, 0)$ and so forth gives

$$\begin{aligned} v_1 - v_2 &= k_1 \\ v_2 - v_3 &= k_2 \\ &\vdots\vdots \\ v_{n-1} - v_n &= k_{n-1} \end{aligned} \tag{25.64}$$

Multiply these equations by 2. Then $2v_i$ are integers, and on the RHS we have even integers. Therefore the $2v_i$ are all even or all odd. Therefore $D_n^* \subset \frac{1}{2}BCC_n$, and this establishes (25.63).

♣So give the discriminant group! ♣

Remark: Combining Example 5 with the observation above on the periodic table we see that the periodic table has a (very approximate) self-duality! It works best for exchanging the first and last column (excluding H, He). A conceptual reason for this is the following. [53] The noble gases have a filled electron shell and to a good approximation act as hard spheres. So their crystal structure should be a minimal sphere packing in three-dimensional space. This means they should be hcp or fcc, and many-body effects break the degeneracy to fcc. On the other hand, in the first column we have a filled shell with a single electron. Therefore, to a good approximation the metals can be treated in the free one-electron picture. Then their Fermi surface is a sphere in *momentum space*. Now, this Fermi surface is a good approximation to the boundary of the Wigner-Seitz cell. Therefore the crystal structure is given by solving a sphere-packing problem in momentum space! Fcc in momentum space implies bcc in real space.

Exercise *Inverse of a generalized Cartan matrix*

Let a_α, $\alpha = 1, \dots, r$ be a set of positive integers and consider the generalized Cartan matrix

$$G_{\alpha\beta} = a_\alpha \delta_{\alpha,\beta} - \delta_{\alpha+1,\beta} - \delta_{\alpha-1,\beta} \tag{25.65}$$

a.) Show that the inverse matrix is

$$G^{\alpha\beta} = \begin{cases} \frac{1}{n} q_\alpha p_\beta & 1 \le \alpha \le \beta \le r \\ \frac{1}{n} p_\alpha q_\beta & 1 \le \alpha \le \beta \le r \end{cases} \tag{25.66}$$

Where $n = \det G_{\alpha\beta}$ and the integers p_α, q_α and n are defined as follows.

Define $[x, y] := x - \frac{1}{y}$, then $[x, y, z] = [x, [y, z]]$, then $[x, y, z, w] = [x, [y, z, w]]$, etc. That is, these are continued fractions with signs. Now in terms of these we define p_α, q_α from

$$\begin{aligned} \frac{p_{j-1}}{p_j} &= [a_j, a_{j+1}, \dots, a_r] \\ \frac{q_{j+1}}{q_j} &= [a_j, a_{j-1}, \dots, a_1] \qquad\qquad 1 \le j \le r \end{aligned} \tag{25.67}$$

with boundary conditions $q_1 = 1, p_r = 1$.

b.) Show that $n = p_0$.

c.) Show that

$$\underbrace{[2, 2, \dots, 2]}_{r \text{ times}} = \frac{r+1}{r} \tag{25.68}$$

♣So give the inverse Cartan matrix in matrix form ♣

---

[53] This argument was worked out with K. Rabe.

25.3.1 The characteristic vector

The next useful invariant of a lattice is based on a

Definition A *characteristic vector* on an integral lattice is a vector $w \in \Lambda$ such that

$$\langle v, v\rangle = \langle w, v\rangle \text{mod} 2 \tag{25.69}$$ eq:chrctvctr

for every $v \in \Lambda$.

Lemma A characteristic vector always exists.

Proof: Consider the lattice $\Lambda/2\Lambda$. Denote elements in the quotient by $\bar{v}$. Note that the quadratic form $Q(v) = \langle v, v\rangle$ descends to a $\mathbb{Z}_2$-valued form $q(\bar{v}) = \langle \bar{v}, \bar{v}\rangle \text{mod} 2$. Moreover, over the field $\kappa = \mathbb{Z}_2$ note that q is *linear*:

$$q(\bar{v}_1 + \bar{v}_2) = q(\bar{v}_1) + q(\bar{v}_2) + 2\langle \bar{v}_1, \bar{v}_2\rangle = q(\bar{v}_1) + q(\bar{v}_2) \tag{25.70}$$

But any linear function must be of the form $q(\bar{v}) = \langle \bar{v}, \bar{w}\rangle$. Now let $w \in \Lambda$ be any lift of $\bar{w}$. This will do. ♠

Note that characteristic vectors are far from unique. Indeed, if w is a characteristic vector and v is any other vector then $w + 2v$ is characteristic. Moreover, any characteristic vector is of this form if the form is nondegenerate when reduced mod two. Therefore the quantity

$$\mu(\Lambda) := \langle w, w\rangle \text{mod} 8 \tag{25.71}$$

does not depend on the choice of w and is an invariant of the lattice Λ.

Remark There is a great deal of magic associated with the number 8 in lattice theory. ♣Say more? ♣

25.3.2 The Gauss-Milgram relation

The invariants we have just described are all related by a beautiful formula sometimes called the Gauss-Milgram sum formula:

Let Λ be an integral lattice. Choose a characteristic vector $w \in \Lambda$ and define the quadratic function

$$Q : \Lambda \otimes \mathbb{R} \to \mathbb{R} \tag{25.72}$$

by

$$Q(v) = \frac{1}{2}\langle v, v - w\rangle \tag{25.73}$$

Note that Q takes integral values on Λ and rational values on Λ^*. Moreover, if $x \in \Lambda^*$ note that

$$Q(x + v) = Q(x) + \langle x, v\rangle + \frac{1}{2}\langle v, v - w\rangle \tag{25.74}$$ eq:Qshift

and the second and third terms are in fact integral, so that if we may define

$$q : D(\Lambda) \to \mathbb{Q}/\mathbb{Z} \tag{25.75}$$

by

$$q(\bar{x}) := Q(x)\text{mod}\mathbb{Z} \tag{25.76}$$

where x is any lift of $\bar{x}$ to Λ. Thanks to (25.74), $q(\bar{x})$ is well-defined.

This is an example of a quadratic function on a finite group. It satisfies

$$q(\bar{x}+\bar{y}) - q(\bar{x}) - q(\bar{y}) + q(\bar{0}) = b(\bar{x},\bar{y}) \tag{25.77}$$

Now, the Gauss-Milgram sum formula states that if $\mathcal{D} = D(\Lambda)$ then

$$\boxed{\sum_{\bar{x}\in\mathcal{D}} e^{2\pi i q(\bar{x})} = \sqrt{|\mathcal{D}|}e^{2\pi i\left(\frac{\sigma(\Lambda)-\mu(\Lambda)}{8}\right)}} \tag{25.78}$$

Proof:

Let us begin with the one-dimensional case $\Lambda = \nu\mathbb{Z}$ with $\nu^2 = n$ is a positive integer. Then, as we have seen $\mathcal{D} = \mathbb{Z}/n\mathbb{Z}$, and

$$b(\bar{x},\bar{y}) = \frac{\bar{x}\bar{y}}{n}\text{mod } 1 \tag{25.79}$$

Moreover, we can take

♣Fix normalization? ♣

$$w = \begin{cases} 0 & n \quad \text{even} \\ 1 & n \quad \text{odd} \end{cases} \tag{25.80}$$

eq:charctr

So we have $Q(x) = \frac{x(x-w)}{2n}$, Let $q(x) = Q(x)\text{mod}1$.

Now we would like to evaluate:

$$S_n = \sum_{\mathcal{D}} e^{2\pi i q(x)} \tag{25.81}$$

Evaluation: Let $g(t) := \sum_{x=0}^{n-1} e^{2\pi i Q(x+t)}$. Note that $g(t+1) = g(t)$ so that

$$g(t) = \sum_{-\infty}^{+\infty} c_k e^{-2\pi i k t} \tag{25.82}$$

We want $g(0) = \sum c_k$. Write

$$c_k = \int_0^1 g(t)e^{2\pi i k t}dt = \int_0^n e^{2\pi i(Q(t)+kt)}dt \tag{25.83}$$

So now write

$$g(0) = \sum_{k\in\mathbb{Z}} c_k = \sum_{k=-\infty}^{\infty}\int_0^n e^{2\pi i Q(t)}e^{2\pi i k t}dt \tag{25.84}$$

But

$$Q(t+kn) = Q(t) + Q(kn) + tk \tag{25.85}$$

and $Q(kn)$ is an integer. Therefore,

$$\sum_{k=-\infty}^{+\infty} \int_0^n e^{2\pi i Q(t+kn)} dt = \int_{-\infty}^{+\infty} e^{2\pi i Q(t)} dt = \sqrt{\frac{\pi}{-i\pi/n}} e^{2\pi i \frac{w^2}{8n}} \tag{25.86}$$

so

$$\sum_{\mathcal{D}} e^{2\pi i q(x)} = \sqrt{n} = \sqrt{n} \exp[2\pi i \left(\frac{1}{8} - \frac{\langle w, w\rangle}{8}\right)] \tag{25.87}$$

Now, for the opposite signature we just take the complex conjugate.

Finally, to go to the general case note that we could have run a very similar argument by considering

$$g(t) = \sum_{\bar{x} \in \Lambda^*/\Lambda} e^{2\pi i Q(x+t)} \tag{25.88}$$

The Fourier analysis is very similar. Once we get to the Gaussian integral we can diagonalize it over $\mathbb{R}$ and then factorize the result into the one-dimensional case. ♠

♣Spell out the details some more here! ♣

Remarks

1. Note that it follows that for self-dual lattices $\mu(\Lambda) = \sigma(\Lambda)\text{mod}8$.

2. *Quadratic functions on finite abelian groups and quadratic refinements*

   ♣ Explain the general problem of finding a quadratic refinement of a bilinear form on a finite abelian group. ♣

3. *Gauss sums in general*

25.4 Self-dual lattices

Definition: An integral lattice is *self-dual*, or *unimodular* if $\Lambda = \Lambda^*$. Equivalently, Λ is unimodular if the determinant of the integral Gram matrix is $\det G_{ij} = \pm 1$.

Example 1: The Narain lattices generated by (25.28) above satisfy $\Lambda(R)^* = \Lambda(R)$ and are unimodular for all R.

Example 2: Of course, if Λ_1 and Λ_2 are unimodular then so is $\Lambda_1 \oplus \Lambda_2$. So $H(1) \oplus \cdots H(1)$ with d factors is an even unimodular lattice of signature (d, d). Similarly, $I^{t,s} \cong \mathbb{Z}^d$ with quadratic form:

$$Diag\{(-1)^t, (+1)^s\} \tag{25.89}$$

on $\mathbb{Z}^d$, $d = t + s$, is an odd unimodular lattice.

Example 3: *Positive definite even unimodular.* There is a class of very interesting positive definite even unimodular lattices which are of rank $r = 8k$ for k a positive integer. Introduce the vector

$$s = (\frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{Q}^{8k} \tag{25.90}$$

$$
\begin{aligned}
\Gamma_{8k} := & \Big\{(x_1, \cdots x_{8k}) \ \in \mathbb{Z}^{8k} | \qquad \sum x_i = 0(2) \Big\} \\
& \cup \Big\{(x_1, \cdots , x_8) \ \in \mathbb{Z}^{8k} + s | \qquad \sum x_i = 0(2) \Big\}
\end{aligned}
\tag{25.91}
$$

Let us check this lattice is even unimodular:

a.) Integral: The only nonobvious part is whether the product of two vectors from $\mathbb{Z}^{8k}+s$ is integral. Write these as $x_i = n_i + \frac{1}{2}, y_i = m_i + \frac{1}{2}$ where $n_i, m_i \in \mathbb{Z}$ and $\sum n_i = 0(2)$ and $\sum m_i = 0(2)$. Then

$$
\sum (n_i + \frac{1}{2})(m_i + \frac{1}{2}) = \sum n_i m_i + \frac{1}{2}(\sum n_i + m_i) + 2k \in \mathbb{Z} \tag{25.92}
$$

b.) Even: Use $n_i^2 = n_i(2)$ for n_i integral.

c.) Self-dual: Suppose $(v_1, \ldots, v_{8k}) \in \Gamma_{8k}^*$. Then, $v_i \pm v_j \in \mathbb{Z}$ and therefore $2v_i \in \mathbb{Z}$ and moreover, the v_i are either all integral or all half-integral. Now,

$$
s \cdot v = \frac{1}{2} \sum_i v_i \in \mathbb{Z} \tag{25.93}
$$

implies $\sum v_i = 0(2)$, hence $v \in \Gamma_{8k}$, and hence $\Gamma_{8k}^* \subset \Gamma_{8k}$ implies it is unimodular.

The case $k = 1$ defines what is known as the E_8-lattice, which is of particular interest in group theory and some areas of physics. Here is a particular lattice basis:

$$
\begin{aligned}
\alpha_1 &= \frac{1}{2}(e_1 + e_8) - \frac{1}{2}(e_2 + e_3 + e_4 + e_5 + e_6 + e_7) \\
\alpha_2 &= e_1 + e_2 \\
\alpha_3 &= e_2 - e_1 \\
\alpha_4 &= e_3 - e_2 \\
\alpha_5 &= e_4 - e_3 \\
\alpha_6 &= e_5 - e_4 \\
\alpha_7 &= e_6 - e_5 \\
\alpha_8 &= e_7 - e_6
\end{aligned}
\tag{25.94}
$$

The form $\alpha_i \cdot \alpha_j$ is the famous E_8 matrix:

$$
\begin{bmatrix}
2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\
0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 2
\end{bmatrix}
\tag{25.95}
$$

One can check that this is in fact of determinant 1. This data is often encoded in a *Dynkin diagram* shown in Figure 20. A dot corresponds to a basis vector. Two dots are connected by a single line if $\alpha_i \cdot \alpha_j = -1$ (i.e. if the angle between them is $2\pi/3$). One gets

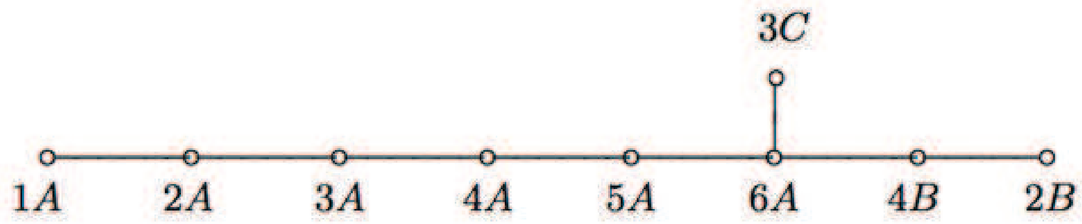

Figure 20: Dynkin diagram of the E_8 lattice. The numbers attached to the nodes have some interesting magical problems which will be discussed later. fig:E8-Dynkin-

Remarks

1. The automorphism group of the E_8 lattice is an extremely intricate object. It is known as the Weyl group of E_8 and it is generated by the reflections in the hyperplanes orthogonal to the simple roots α_i listed above. There is an obvious subgroup isomorphic to

$$(\mathbb{Z}_2)^7 \ltimes S_8 \tag{25.96}$$

where the S_8 acts by permuting the coordinates and $(\mathbb{Z}_2)^7 \cong (\mathbb{Z}_2)^8/\mathbb{Z}_2$ is the group of sign-flips $x_i \to \epsilon_i x_i$ where an *even* number of signs are flipped. What is not obvious

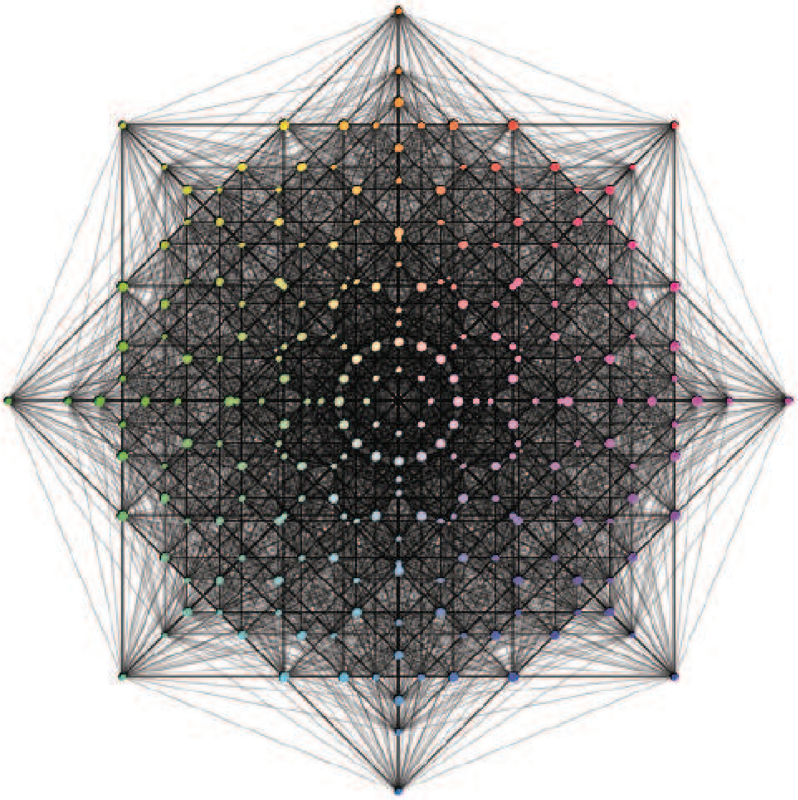

Figure 21: A projection of the 240 roots of the E8 root lattice in a two-dimensional plane. Copied from http://www.madore.org/ david/math/e8w.html.

fig:ProjectedE

is that this group is only a subgroup and in fact the full Weyl group has order

$$\begin{aligned} |W(E_8)| &= 2^7 8! \times 135 \\ &= 8! \times (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 4 \cdot 2 \cdot 3) \\ &= 2^{14} \times 3^5 \times 5^2 \times 7 \\ &= 696729600 \end{aligned} \tag{25.97}$$

2. ♣ SAY SOMETHING ABOUT VECTORS OF SQUARELENGTH TWO ♣

25.4.1 Some classification results

There are some interesting results on the classification of unimodular lattices. We now briefly review some of the most important ones.

The nature of the classification of lattices depends very strongly on the signature and rank of the form. For example, the classification of definite integral forms is a very difficult problem which remains unsolved in general.

By contrast, the classification is much simpler for indefinite signature: (i.e. $t > 0, s > 0$).

♣ EXPLAIN THE PROOF IN SERRE'S BOOK. THIS IS A BEAUTIFUL AND SIMPLE APPLICATION OF GENERAL IDEAS OF K-THEORY ♣

1. Odd unimodular lattices are unique. By change of basis we always get the lattice:

$$\Gamma \approx I^{t,s}. \tag{25.98}$$

2. Even unimodular lattices only exist for $(t-s) = 0 mod 8$ and are again unique for s, t both nonzero. They are denoted:

$$\Gamma \approx II^{t,s} \tag{25.99}$$

An explicit consturction of $II^{t,s}$ may be given by taking the lattice of d-tuples of points $(x_1, \dots, x_d) \in \mathbb{R}^{t,s}$, with $d = t + s$, where the x_i are either all integral, or all half-integral, and in either case $\sum x_i = 0 \mathrm{mod} 2$.

Although the indefinite even unimodular lattices are unique, their *embedding* into $\mathbb{R}^{p,q}$ is highly nonunique. We have already seen this in example 3 above.

Note that $\Lambda(R) = \Lambda(1/R)$. The inequivalent embeddings of $II^{1,1}$ into $\mathbb{R}^{1,1}$ are parametrized by $R \geq 1$.

There are some partial results on positive definite even unimodular lattices.

1. In fact, they only exist for

$$\mathrm{dim}\Lambda = 0 \mathrm{mod} 8 \tag{25.100}$$

2. In any dimension there is a *finite* number of inequivalent lattices. In fact, we can count them! That is, we can count them using the Smith-Minkowski-Siegel "mass formula" which gives a formula for

$$\sum_{[\Lambda]} \frac{1}{|\mathrm{Aut}(\Lambda)|} = \frac{|B_{n/2}|}{n} \prod_{1 \leq j < n/2} \frac{|B_{2j}|}{4j} \tag{25.101}$$

The sum on the left is over inequivalent even unimodular positive definite lattice of dimension n. The B_{2j} are the Bernoulli numbers defined by

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{720} \pm \cdots \tag{25.102}$$

The growth of the Bernoulli number is given by Euler's result:

$$B_{2n} = (-1)^{n+1} \frac{2(2n)!}{(2\pi)^{2n}} \zeta(2n) \tag{25.103}$$

and hence the product on the RHS grows very fast. (Note that $\zeta(2n)$ is exponentially close to 1.

♣Say how RHS grows with the rank r. ♣

♣ Reference: A. Eskin, Z. Rudnik, and P. Sarnak, "A Proof of Siegel's Weight Formula," (There should be a simple topological field theory proof.) ♣

In higher dimensions there can be many inequivalent even integral unimodular lattices. The number $n(d)$ of such lattices is known to be:

$$
\begin{aligned}
n(8) &= 1 \\
n(16) &= 2 \\
n(24) &= 24 \\
n(32) &> 80 \times 10^6
\end{aligned}
\tag{25.104}
$$

eq:inequiv

Indeed, if we compute the RHS of the SMS formula for $n = 8$ then we get

$$
\frac{B_4}{4} \times \frac{B_2}{4} \times \frac{B_4}{8} \times \frac{B_6}{12} = \left(\frac{1}{120}\right) \times \left(\frac{1}{24}\right) \times \left(\frac{1}{240}\right) \times \left(\frac{1}{504}\right) = \frac{1}{696729600} \tag{25.105}
$$

This is exactly one over the order of the automorphism group of E_8, confirming $n(d) = 1$.

♣Surely there is a simpler proof of uniqueness... ♣

Of the 24 even unimodular lattices in dimension 24 one stands out, it is the *Leech lattice*, which is the unique lattice whose minimal length-square is 4.

There are many constructions of the Leech lattice, but one curious one is that we consider the light-like vector:

$$
w = (70; 24, 23, \dots, 3, 2, 1, 0) \tag{25.106}
$$

in $II^{1,25} \subset \mathbb{R}^{1,25}$ and consider the lattice $w^{\perp}/w\mathbb{Z}$. This is a positive definite even integral lattice of rank 24. Note that the vectors of length-squared two are not orthogonal to w.

♣Need to explain why this construction gives a self-dual lattice. ♣

Remarks

1. Topology of 4-manifolds.
2. Abelian Chern-Simons theory.

Exercise
Compute the number of vectors of square-length $= 2$ in $\Lambda_R(E_8)$.

Exercise For any integer k construct an even unimodular lattice whose minimum length vector is 2^k.

Exercise *Narain lattices*
Let V be a vector space and $V^{\vee}$ be the dual space.

Using only the dual pairing one defines a natural signature (d,d) nondegenerate metric on $V \oplus V^{\vee}$:

$$\langle (x,\ell),(x',\ell')\rangle := \ell'(x) - \ell(x') \tag{25.107}$$

a.) Show that if $\Lambda \subset V$ is a lattice then

$$\Lambda_N := \{(p+w, p-w) | p \in \Lambda, w \in \Lambda^*\} \tag{25.108}$$

is an even unimodular lattice.

b.) Show that the space of inequivalent embedded lattices isomorphic to $II^{d,d}$ is $O(d,d;\mathbb{Z})\backslash O(d,d;\mathbb{R})$

♣This is out of place. We haven't done quotients yet. ♣

Exercise

Show that the lattice of $(x_1,\dots,x_d) \in \mathbb{R}^{t,s}$, with x_i all integral or all half-integral and $\sum x_i = 0(2)$ is an even self-dual lattice.

Hint: Use the same procedure as in the E_8 case above.

Exercise

Show that if $s > t$ then $II^{s,t}$ must be isomorphic as a lattice to a lattice of the form

$$II^{1,1} \oplus \cdots \oplus II^{1,1} \oplus E_8 \oplus \cdots \oplus E_8 \tag{25.109}$$

while if $t > s$ it is of the form

$$II^{1,1} \oplus \cdots \oplus II^{1,1} \oplus E_8(-1) \oplus \cdots \oplus E_8(-1) \tag{25.110}$$

of signature $((+1)^l, (-1)^{l+8m})$.

Exercise *Lattice Theta functions*

If Λ is a positive definite lattice one can associate to it the *Theta function*

$$\Theta_\Lambda := \sum_{v\in\Lambda} q^{\frac{1}{2}(v,v)} \tag{25.111}$$

This is a series in $q^{1/2}$ and it converges absolutely for $|q^{1/2}| < 1$. This function counts the number of lattice vectors of a given length.

For τ a complex number of positive imaginary part define $q := e^{2\pi i\tau}$. Using the Poisson summation formula show that the theta functions for Λ and Λ^* are related by:

$$\Theta_\Lambda(-1/\tau) = (-i\tau)^{\dim\Lambda/2}\frac{1}{\sqrt{|D(\Lambda)|}}\Theta_{\Lambda^*}(\tau) \tag{25.112}$$

♣♣ EXPLAIN RELATION OF FINITE HEISENBERG GROUPS AND THETA FUNCTIONS ♣♣

25.5 Embeddings of lattices: The Nikulin theorem

♣ Explain the terminology "glue group" ♣

25.6 References

Reference: For much more about lattices, see

J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices, and Groups.*

A beautiful and concise treatment of some of the material above can be found in:

J.-P. Serre, *A Course on Arithmetic*

26. Positive definite Quadratic forms

Criteria for A to be positive definite:

$A > 0$ iff the determinants of all minors is positive.

If A is a positive definite matrix with integer entries then it satisfies some remarkable properties:

1. Kronecker's theorem: $\| A \| \geq 2$ or $\| A \| = 2\cos(\pi/q)$, $q \geq 3$.

2. Perron-Frobenius theorem: A has a maximal positive eigenvalue and the eigenvector can be taken to have all positive entries. (Actually, the PF theorem is far more general.)

See V. Jones, et. al. *Coxeter graphs...* and Gantmacher, for a discussion.

Nice Application: GOOGLE search algorithm PageRank

27. Quivers and their representations

Nice application of linear algebra.