Math diary

Ethan Lake

Preface:

This is a diary containing worked-out math problems. These problems are either elaborations on calculations in papers which I wanted to work out in detail, explanations of well-known facts that I wanted to remember, or problems which arose when doing research. There are doubtless many typos, and I have not been very diligent about adding citations. Some entries were written near the beginning of grad school, and I take no responsibility for any misguided beliefs that my younger self decided to write down.

Below is an index of problems arranged by subject matter; some diary entries appear under multiple sections as appropriate.

Algebra and topology

Geometry

Groups and representation theory

Statistics and information theory

Miscellaneous

Why exactly do Poisson brackets map to commutators during (pre)quantization?

Today we'll motivate why Poisson brackets become commutators under (pre)quantization, which I've only ever seen presented in a rather axiomatic way.

◇ ◇ ◇

First, a disclaimer: it is *not* true that Poisson brackets map to commutators under quantization. Indeed, it can be proven that there does not exist any quantization scheme for which this is true. The "best" quantization scheme, viz. Weyl quantization (which is the scheme where products of x and p are symmetrized when quantized) is such that the PB of two functions f, g maps to the commutator if at least one of f, g is at most quadratic in the x and p, but this is essentially the best we can do (see e.g. Hall's "Quantum theory for mathematicians"). However, it is true that we can get PBs to map to commutators within the context of pre-quantization, where we're allowed to work with a larger Hilbert space before committing to a choice of polarization. This latter context is the one we will discuss in the present diary entry.

To start, we will prove a math fact that will be useful later.

Proposition 1. *For any two vector fields X, Y, we have*

$$i_{[X,Y]}\Lambda = (\mathcal{L}_X i_Y - i_Y \mathcal{L}_X)\Lambda, \tag{1}$$

where $\Lambda \in \Omega^n(M)$ is any differential form on M, i_X is the interior product, and where $\mathcal{L}_X$ is the Lie derivative.

My proof is inelegant, but oh well. First, we calculate

$$\mathcal{L}_X i_Y \Lambda = ([i_{\nabla_X Y}\Lambda]_{\boldsymbol{\lambda}} + [i_Y(\nabla_X \Lambda)]_{\boldsymbol{\lambda}} + Y^\mu \Lambda_{\mu\alpha\lambda_3\ldots\lambda_n}\partial_{\lambda_2}X^\alpha + \ldots)dx^{\boldsymbol{\lambda}}, \tag{2}$$

where $\boldsymbol{\lambda}$ is a multi-index $\boldsymbol{\lambda} = \lambda_2, \lambda_3, \ldots, \lambda_n$, so that $dx^{\boldsymbol{\lambda}} = dx^{\lambda_2} \wedge \cdots \wedge dx^{\lambda_n}$. Here the $+\ldots$ refers to other terms in the Lie derivative of Λ, viz. other indices that get contracted with the derivative of X. I know the notation is opaque, but thinking of something better seemed to not be worth the trouble—if in doubt, just re-write it out. Anyway, the other relevant term is

$$\begin{aligned} i_Y \mathcal{L}_X \Lambda &= i_Y(\nabla_X \Lambda + \partial_{\lambda_1} X^\mu \Lambda_{\mu\boldsymbol{\lambda}} dx^1 \wedge dx^{\boldsymbol{\lambda}}) \\ &= i_Y(\nabla_X \Lambda) + i_{\nabla_Y X}\Lambda - \partial_{\lambda_2} X^\mu Y^\alpha \Lambda_{\mu\alpha\lambda_3\ldots\lambda_n} dx^{\boldsymbol{\lambda}}. \end{aligned} \tag{3}$$

We can flip the sign on the last term by exchanging the μ and α indices in Λ, and when this is subtracted from (2), it will cancel with the last term there. Similarly, the first term in the above equation will cancel with the second in (2). This leaves only two terms, and we find that indeed,

$$(\mathcal{L}_X i_Y - i_Y \mathcal{L}_X)\Lambda = (i_{\nabla_X Y} - i_{\nabla_Y X})\Lambda = i_{[X,Y]}\Lambda. \tag{4}$$

Now we can talk about Poisson brackets. Let X_f and Y_g be two Hamiltonian vector fields on phase space associated to Hamiltonians f and g, respectively, and let ω be the symplectic form. This means that

$$i_{X_f}\omega = -df \implies \omega_{\mu\nu} X_f^\mu = \partial_\mu f, \qquad X_f^\mu = \omega^{\mu\nu}\partial_\nu f, \tag{5}$$

and likewise for Y_g. Keeping track of the signs here is supremely important, unfortunately. We have made a choice that means commutators will become negative Poisson brackets, but this was only realized retrospectively—oh well, can't be helped now.

Our goal will be to related the vector field $[X_f, Y_g]$ to the vector field $V^\mu_{\{f,g\}} = \omega^{\mu\nu}\partial_\nu\{f,g\}$ (if f, g are Hamiltonians, then their Poisson bracket also generates a Hamiltonian vector field). In the following, we will drop the subscripts on X and Y for clarity.

First, we calculate, using the result derived above,

$$i_{[X,Y]}\omega = (\mathcal{L}_X i_Y - i_Y \mathcal{L}_X)\omega = \mathcal{L}_X i_Y \omega = \mathcal{L}_X(Y^\sigma \omega_{\sigma\nu})dx^\nu = -\mathcal{L}_X dg, \tag{6}$$

where our sign convention is such that $Y^\sigma\omega_{\sigma\nu} = -\partial_\nu g$, and where we used the stationarity of ω under the flow of X:

$$\mathcal{L}_X\omega = (di_X + i_X d)\omega = d(i_X\omega) = d(-df) = 0. \tag{7}$$

Expanding this out, we have

$$\begin{aligned} -[\mathcal{L}_X dg]_\nu &= -X^\mu\partial_\mu\partial_\nu g - \partial_\nu X^\mu \partial_\mu g = -\omega^{\mu\lambda}\partial_\lambda f \partial_\mu\partial_\nu g - \partial_\nu(\omega^{\mu\lambda}\partial_\lambda f)\partial_\mu g \\ &= \partial_\nu(\omega^{\lambda\mu}\partial_\lambda f \partial_\mu g) = \partial_\nu\{f,g\}. \end{aligned} \tag{8}$$

Now the vector field generated by the Poisson bracket is

$$V^\mu_{\{f,g\}} = \partial_\lambda(\omega^{\sigma\nu}\partial_\sigma f \partial_\nu g)\omega^{\mu\lambda}, \tag{9}$$

so that

$$i_{V_{\{f,g\}}}\omega = -\partial_\nu(\omega^{\sigma\lambda}\partial_\sigma f \partial_\lambda g)dx^\nu = -\partial_\nu\{f,g\}dx^\nu, \tag{10}$$

which when compared with (8) and (6) gives

$$i_{[X,Y]}\omega = -i_{V_{\{f,g\}}}\omega. \tag{11}$$

Since ω is invertible, the two vector fields in the interior products must be negatives of each other, and hence (restoring subscripts on X and Y for clarity)

$$[X_f, Y_g] = -V_{\{f,g\}}. \tag{12}$$

Graahh, that minus sign is annoying. Oh well.

This tells us that the commutator of Hamtilonian vector fields, when represented as differential operators, generates the Poisson bracket algebra. In geometric quantization we send functions on phase space (f and g) to differential operators (first order ones in e.g. Kahler quantization) that perform the flow along their respective vector fields. Therefore if we think of (pre)quantization as a mapping of functions on phase space to differential operators via $f \mapsto X_f$, the above equation tells us that the (pre)quantization of the Poisson bracket $\{f,g\}$ is the commutator of the (pre)quantizations of f and g.

Torsion-free first cohomology

We will prove that $H^1(X;\mathbb{Z})$ is torsion-free for any X (this holds in both singular cohomology and in group cohomology). This can be done in just a paragraph.

$\diamond \diamond \diamond$

This is just a quick application of the universal coefficient theorem, which tells us that for any Abelian coefficient group G,

$$0 \to \mathrm{Ext}(H_0(X;\mathbb{Z}), G) \to H^1(X;G) \to \mathrm{Hom}(H_1(X;\mathbb{Z}), G) \to 0. \tag{13}$$

Since $H_0(X;\mathbb{Z})$ is always free, the Ext vanishes and we have

$$H^1(X;G) \cong \mathrm{Hom}(H_1(X;\mathbb{Z}), G). \tag{14}$$

Thus in particular if $G = \mathbb{Z}$, none of the elements in $\mathrm{Hom}(H_1(X;\mathbb{Z}), G)$ can have torsion, since if $kh(c) = 0$ for some $k \in \mathbb{Z}, c \in H_1(X;\mathbb{Z}), h \in \mathrm{Hom}(H_1(X;\mathbb{Z}), \mathbb{Z})$, then $h(c) = 0$ (more generally, $\mathrm{Hom}(H(X), \mathbb{Z})$ is isomorphic to the free part of $H(X)$, provided $H(X)$ is finitely generated). If we were to take $p \geq 1$ cohomology groups, things would change, since the torsion part of $H^p(X;\mathbb{Z})$ is sensitive to the torsion part of $H_{p-1}(X;\mathbb{Z})$, which can be non-zero. Done! Note that unlike $H^1(Xl\mathbb{Z})$, the homology $H_1(X;\mathbb{Z})$ is not torsion-free in general (e.g. $SO(3)$, with $H_1(SO(3);\mathbb{Z}) = \mathbb{Z}_2$).

Orientability and the fundamental group

We will show that if M is a connected manifold such that $\pi_1(M)$ has no subgroups of index 2, then M is orientable.

$\diamond \diamond \diamond$

We will prove the contrapositive, namely that if M is non-orientable then $\pi_1(M)$ must have an index 2 subgroup.

Intuitively, the reason for this is as follows. Suppose M is non-orientable, and consider curves in $\pi_1(M)$. Since M is non-orientable it should have a cycle $c \in C_1(M;\mathbb{Z})$ around which the orientation cannot be consistently defined, since the orientation bundle of M must be nontrivial. Consider going around c twice. Then $c \times c$ in $\pi_1(M)$ looks like two paths next to eachother, both traveling in the same direction. However since the orientation cannot be defined consistently along c, swapping the direction that one of the paths travels in does not change what $c \times c$ is in $\pi_1(M)$. But then we have two copies of the same path traveling in opposite directions, and so we can "annihilate" these paths, producing the trivial element in $\pi_1(M)$. Thus we expect $\pi_1(M)$ to have some sort of $\mathbb{Z}_2$ character coming from loops of this sort.

Now for the more precise argument. If M is non-orientable then it has a two-sheeted double cover $\mathcal{M}$, which is a covering space $p : \mathcal{M} \to M$ for M. Let the $\mathbb{Z}_2$ fiber at a given point $x \in M$ be denoted by F, and consider the monodromy action of the fundamental group based at x on the fiber;

$$\pi_1(M, x) \times F \to F. \tag{15}$$

The action is defined by lifting up elements in $\pi_1(M, x)$ into $\mathcal{M}$, and taking $h \cdot x = h(1)$ for h any element of $\pi_1(M, x)$ lifted up into $\mathcal{M}$. Thus the monodromy action permutes the (two) elements of F. This action is always transitive, since if x, x' are two distinct points in F, the path in $\mathcal{M}$ connecting them projects down into $\pi_1(M, x)$, and gives us a lift with which to have an action that permutes x and x'.

So, we have a homomorphism

$$f : \pi_1(M) \to \mathbb{Z}_2 \tag{16}$$

defined by the monodromy action on the fibers. Now $\ker(f)$ is a subgroup of $\pi_1(M)$, and so since $\mathrm{im}(f) = \mathbb{Z}_2$, we have

$$[\pi_1(M) \ : \ \ker(f)] = \dim \mathrm{im}(f) = 2. \tag{17}$$

Thus $\ker(h)$ provides us with the desired index-two subgroup.

Fundamental groups and homotopies into S^1

Suppose X is path connected, with $\pi_1(X)$ finite. We will show that all maps from X into S^1 are nullhomotopic.

◇ ◇ ◇

We will show this by using the covering spaces $\widetilde{X}$ and $\widetilde{S}^1 = \mathbb{R}$, so first some preliminaries about these. For a general map $f : X \to Y$, with $\widetilde{X}$ the universal cover of X and $\widetilde{Y}$ any cover of Y, consider the diagram

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{\widetilde{f}} & \widetilde{Y} \\ \big\downarrow{\scriptstyle \pi_X} & & \big\downarrow{\scriptstyle \pi_Y} \\ X & \xrightarrow{f} & Y \end{array} \tag{18}$$

We want to first show the existence of the lift $\widetilde{f}$ that maps between the covering spaces. Consider the map $f \circ \pi_X : \widetilde{X} \to Y$. This will lift to the desired map $\widetilde{f} : \widetilde{X} \to \widetilde{Y}$ if the lifting criterion is satisfied, namely if (sorry for the profusion of π's)

$$(f \circ \pi_X)(\pi_1(\widetilde{X})) \subset \pi_Y(\pi_1(Y)). \tag{19}$$

This is true for us since $\widetilde{X}$ is the universal cover of X, and hence has trivial fundamental group. Thus $\widetilde{f}$ exists. Note that we are *not* assuming the existence of a lift of f to a map $g : X \to \widetilde{Y}$, since this imposes too strong a constraint on $\pi_1(X)$.

Now we can start proving the claim. Throughout, we will be cavalier about keeping explicit track of basepoints and related issues.

Suppose that the map f were not nullhomotopic. Then we must have $f(X) = [k] \in \pi_1(S^1) = \mathbb{Z}$ such that $k \neq 0$, where $[k]$ stands for a representative of the element in $\pi_1(S^1)$ which wraps the circle k times. Now consider pulling back $[k]$ into X. The inverse image of a given point on S^1 will of course not usually be unique, but we can choose from among the points in the inverse image so that as we move around continuously in S^1, we move around continuously in X. The point of this procedure is just to pullback $[k]$ to a curve in X. The pullback of $[k]$ cannot be an open curve in X, since then $f(X)$ would be contractible and $f(X)$ would be nullhomotopic. An example of this is the map

$$g : S^2 \to S^1, \qquad (\theta, \phi) \mapsto 2\theta. \tag{20}$$

This covers the S^1, but it does so in a contractible way, and the curves in the inverse image of [1] are all contractble. Similarly, the pullback of $[k]$ cannot be contractible, since then $f(X)$ would have been contractible. So, we conclude that the pullback of $[k]$ must be non-contractible in X, and hence if $f(X)$ is not nullhomotopic, then it must map some nontrivial element in $\pi_1(X)$ to $[k]$.

We will now show that this is a contradiction. Consider the (left) action of deck transformations

$$\text{Deck} : \pi(X, x_0) \to \text{aut}(\pi_X^{-1}(x_0)), \qquad \gamma \mapsto (\gamma : x_0 \mapsto [\gamma] \cdot x_0), \tag{21}$$

where $\gamma \cdot x_0$ is the endpoint in $\widetilde{X}$ of the curve γ based at x_0, when that curve is lifted up to $\widetilde{X}$. The lifted map $\widetilde{f}$ is such that it allows us to either lift to $\widetilde{X}$ and then map, or map and then lift to $\widetilde{Y}$ — the results are the same. This means that the action of deck transformations commutes with the mapping between the two spaces:

$$[f(\gamma)] \cdot f(x_0) = \widetilde{f}([\gamma] \cdot x_0). \tag{22}$$

Now since we are assuming $f(X)$ has degree k around the S^1, we can always choose γ to be such that $[f(\gamma)] = [k]$ in $\pi_1(S^2) = \mathbb{Z}$. We know how deck transformations on the covering space of the circle work: they shift us up by one coil on the helix over S^1. Thus if the S^1 has radius 1, the deck transformation must act as (after choosing an appropriate lift of $\theta = f(x_0)$ into $\mathbb{R}$)

$$[f(\gamma)] \cdot \theta = \theta + 2\pi k. \tag{23}$$

In particular, we see that

$$\widetilde{f}([\gamma^n] \cdot x_0) = [f(\gamma)^n] \cdot \theta = \theta + 2\pi n k, \tag{24}$$

for any $n \in \mathbb{Z}$.

Now we get a contradiction: we have assumed that $\pi_1(X)$ is finite, and so in particular, all elements in $\pi_1(X)$ must have finite order. This is in contradiction to the action of the deck transformations we derived above, which implies that all of the elements in $\pi_1(X)$ have infinite order.

As a quick application of this result, we can derive that $\pi_n(S^1) = 0$ for all $n > 1$ (not obvious), which follows from the fact that $\pi_1(S^n) = 0$ for $n > 1$ (obvious).

Relations between different group cohomologies and different Chech cohomologies with $U(1)$ and $\mathbb{Z}$ coefficients

Today is a quick recollection of some algebraic topology basics. Remind yourself why for finite groups G, we have the isomorphism in group cohomology

$$H^k(G; \mathbb{R}/2\pi\mathbb{Z}) \cong H^{k+1}(G; \mathbb{Z}) \tag{25}$$

for $k > 0$.

◇ ◇ ◇

This is basically just a covariant analogue of the contravariant statement about exact sequences of chain complexes inducing long exact sequences in cohomology. First, we have the exact sequence

$$1 \to \mathbb{Z} \to \mathbb{R} \to U(1) \to 1. \tag{26}$$

Let C be a free chain complex. $\mathrm{Hom}(C, \cdot)$ is a *covariant* functor (as opposed to $\mathrm{Hom}(\cdot, M)$ which is contravariant), and so the above short exact sequence induces the short exact sequence

$$1 \to \mathrm{Hom}(C, \mathbb{Z})) \to \mathrm{Hom}(C, \mathbb{R}) \to \mathrm{Hom}(C, U(1)) \to 1. \tag{27}$$

Thus by writing down the three chain complexes side-by-side with the rows of the resulting diagram being exact, we get the associated long exact sequence in cohomology ($H^*(G; \cdot)$ is also a covariant functor)

$$\ldots \to H^k(G; \mathbb{Z}) \to H^k(G; \mathbb{R}) \to H^k(G; U(1)) \to H^{k+1}(G; \mathbb{Z}) \to \ldots \tag{28}$$

The maps from guys with coefficients in $\mathbb{Z}$ to those with coefficients in $\mathbb{R}$ and the subsequent maps to guys with coefficients in $U(1)$ are the maps induced by the maps in $\mathbb{Z} \to \mathbb{R} \to U(1)$. The maps between degrees are induced by the coboundary operator,

$$\delta : H^k(G; U(1)) \to H^{k+1}(G; \mathbb{Z}). \tag{29}$$

Hitting an element in $H^k(G; U(1))$ with δ produces something nontrivial here since the coefficient group is changing: if $\delta\omega^k = 0$ for some $U(1)$-valued cochain ω^k, then we really mean that $\delta\omega^k = 2\pi\omega^{k+1}$, where ω^{k+1} is $\mathbb{Z}$-valued.

Now we use the fact that for finite groups G, the group cohomology with coefficients in $\mathbb{R}$ vanishes

$$H^k(G; \mathbb{R}) = 0 \tag{30}$$

for $k > 0$. This can be seen in several ways; one is by writing the UCT for group cohomology for $k > 0$ as

$$H^k(G; M) \cong \mathrm{Hom}(H_k(G; \mathbb{Z}), M) \oplus \mathrm{Ext}(H_{k-1}(G; \mathbb{Z}), M). \tag{31}$$

When G is a finite group the homology $H_k(G; \mathbb{Z})$ will be finite when $k \neq 0$, and so $\mathrm{Hom}(H_k(G; \mathbb{Z}), \mathbb{R}) = 0$ when $k > 0$. If $k = 1$ then we have $H_0(G; \mathbb{Z}) = \mathbb{Z}$, but $\mathrm{Ext}(\mathbb{Z}, \mathbb{R}) =$

0. When $k > 1$ $H_{k-1}(G;\mathbb{Z})$ will be finite, and then $\mathrm{Ext}(H_{k-1}(G;\mathbb{Z}),\mathbb{R})$ will vanish (e.g. $\mathrm{Ext}(\mathbb{Z}_n; M) = M/nM$). One might also say that $\mathbb{R}$ is an injective $\mathbb{Z}G$ module, so that group cohomology with $\mathbb{R}$ coefficients vanishes (see e.g. Brown's book on group cohomology). Anyway, since the terms with $\mathbb{R}$ coefficients in the long exact sequence die, we get

$$H^k(G;\mathbb{R}/2\pi\mathbb{Z}) \cong H^{k+1}(G;\mathbb{Z}). \tag{32}$$

As mega-simple examples, we can use the power of homological algebra to calculate (for us $0 \notin \mathbb{N}$)

$$H^k(\mathbb{Z}_n;\mathbb{Z}) = \begin{cases} \mathbb{Z}_n & k \in 2\mathbb{N} \\ 0 & k \in 2\mathbb{N}+1 \\ \mathbb{Z} & k = 0 \end{cases}. \tag{33}$$

This can be easily checked against the groups with $U(1)$ coefficients in low dimensions: $H^1(\mathbb{Z}_n;U(1)) = \mathbb{Z}_n$ since with the action being trivial the first cohomology just measures $\mathrm{Hom}(\mathbb{Z}_n, U(1))$, while e.g. $H^2(\mathbb{Z}_n;U(1)) = 0$ as can by checked by writing out the cocycle relations explicitly.

Formula (32) is only true for group cohomology with finite groups; the more general statement would be that (I think)

$$H^k(G;\mathbb{R}/\mathbb{Z}) \cong \mathrm{Tor}[H^{k+1}(G;\mathbb{Z})], \tag{34}$$

which reduces to the less general statement in the case of G finite since if G finite the whole cohomology group must be pure torsion.

We can also apply this formula to the case of Cech cohomology where the coefficients are valued in some sheaf, usually that of the smooth functions on a manifold X valued in a group G (G needn't be Abelian, although if it isn't then the corresponding Cech cohomologies are just sets, and not groups). In this case an n-cochain is a smooth function from intersections $U_1 \cap \cdots \cap U_{n+1}$ into G. From the same SES, we get

$$\ldots \to H^k(X;\mathcal{C}^\infty(\mathbb{Z})) \to H^k(X;\mathcal{C}^\infty(\mathbb{R})) \to H^k(X;\mathcal{C}^\infty(\mathbb{C}^\times)) \to H^{k+1}(X;\mathcal{C}^\infty(\mathbb{Z})) \to \ldots \tag{35}$$

Now of course when the group is $\mathbb{Z}$, we can just replace the Cech cohomology above with regular cohomology valued in $\mathbb{Z}$. The Cech cohomology groups that involve $\mathcal{C}^\infty(\mathbb{R})$ are all zero (in degree greater than zero)[1]. We can prove this as follows. Consider the case $k = 1$. Let u_α be a partition of unity, where each u_α has compact support in U_α and where $\sum_\alpha u_\alpha = \mathbf{1}$. Given a 1-cochain $\lambda_{\alpha\beta}$, consider the 0-cochain

$$\omega_\alpha = \sum_\beta \lambda_{\alpha\beta} u_\beta. \tag{36}$$

Then

$$(\delta\omega)_{\alpha\beta} = \sum_\gamma u_\gamma(\lambda_{\alpha\gamma} - \lambda_{\beta\gamma}) = \sum_\gamma u_\gamma \lambda_{\alpha\beta} = \lambda_{\alpha\beta}, \tag{37}$$

[1] This is what allows one to prove the equivalence between Cech cohomology with coefficients in the constant sheaf $\mathbb{R}$ and de Rham cohomology: construct the Cech de-Rham bicomplex and use the triviality of all the cohomology groups in the "interior" of the complex to snake from the first column of the complex to the last row.

where we used the cocycle condition for λ and the fact that $\lambda_{\alpha\beta} = -\lambda_{\beta\alpha}$. Therefore $\lambda = \delta\omega$ is exact! A similar argument goes through for all higher cohomology groups. This argument fails in the case of e.g. $\mathcal{C}^\infty(U(1))$ coefficients though, since then we cannot conclude that $\lambda_{\alpha\beta} + \lambda_{\beta\gamma} + \lambda_{\gamma\alpha} = 0$, only that it is equal to $2\pi n_{\alpha\beta\gamma}$, where $n_{\alpha\beta\gamma} \in \mathbb{Z}$ (here we have lifted a $U(1)$ cochain to an $\mathbb{R}$ cochain $\lambda_{\alpha\beta}$).

Returning to the SES and using the above result, we get the isomorphism

$$H^k(X; \mathcal{C}^\infty(U(1))) \cong H^{k+1}(X; \mathbb{Z}). \tag{38}$$

Note in particular the case when $k = 1$. The RHS gives the Chern class of a $U(1)$ bundle over X, and so this tells us that

$$\begin{aligned} H^1(X; \mathcal{C}^\infty(U(1))) &\cong \{\text{Isomorphism classes of principal } U(1) \text{ bundles over } X\} \\ &\cong H^2(X; \mathbb{Z}), \end{aligned} \tag{39}$$

where the first line comes from the fact that the cohomology group on the left classifies all possible transition functions for $U(1)$ bundles over X, up to isomorphism. This is kind of cool since the bundles are all characterized by a second cohomology group, independent of $\dim X$. This is also essentially the same as saying that for $U(1)$ bundles, only the first Chern class is non-vanishing. The first statement about bundles is true more generally:

$$H^1(X; \mathcal{C}^\infty(G)) \cong \{\text{Isomorphism classes of principal } G \text{ bundles over } X\}. \tag{40}$$

Here we think of these bundles as having a distinguished "identity" element $X \times G \to X$ given by the trivial bundle. For example, taking G to be some k-dimensional subgroup of $GL(\dim X; \mathbb{R})$ gives us the set of all k-plane bundles over X. Note that if G is not Abelian, the LHS will not be a group in general; it will just be a set with a distinguished "identity" element. In the case where it is a group and G is abelian, like when $G = U(1)$, the group operation corresponds to taking $\otimes$s of bundles. The analogue of $H^2(X; \mathbb{Z})$ for G-bundles is $H^2(X; \mathcal{C}^\infty(\pi_1(G))) \cong H^2(X; \pi_1(G))$, with the isomorphism holding for compact Lie groups G, since then $\pi_1(G)$ is discrete and abelian. That $\pi_1(G)$ is relevant here comes from the SES $1 \to \pi_1(G) \to \widetilde{G} \to G \to 1$ for $\widetilde{G}$ the universal cover, and from then applying the same construction described above to this SES. In particular, taking $G = SO(n)$ with $n > 2$ gives

$$H^1(X; \mathcal{C}^\infty(SO(n))) \cong H^2(X; \mathbb{Z}_2), \tag{41}$$

since $\pi_1(SO(n)) = \mathbb{Z}_2$ for all $n > 2$. The classes on the RHS are the second Stiefel-Whitney classes.

Writhe and linking numbers

Write down an integral formula for the writhe $w(L)$ of a link in S^3. If the link has components K_i, the appropriate formula to reproduce is

$$w(L) = \sum_{i,j} \mathcal{L}(K_i, K_j) + \sum_i w(K_i), \tag{42}$$

where $\mathcal{L}$ is the linking number and $w(K_i)$ is the writhe of the K_i link component.

◇ ◇ ◇

The linking number $\mathcal{L}(K_i, K_j)$ can be conveniently defined as the signed intersection number of K_i with the Seifert surface of K_j (or vice versa). Recall that the Seifert surface of a knot is a (possibly non-unique) *orientable* manifold $M_j \subset S^3$ that bounds K_j, so that $\partial M_j = K_j$. Since the intersection product of manifolds is Poincare dual to the wedge product, we guess that the correct formula will come from taking wedge products between the Poincare duals of the knots K_i and the Seifert surfaces.

To this end, let S_i be the Seifert surface of K_i. Its Poincare dual in S^3 is a 1-form $A_i = \widehat{S}_i$. To deal with the write of each link component properly, we need to fatten the link components into solid tori of radius r_i (we do this so that after fattening, none of the fattened K_i intersect one another). Let $B_i = \widehat{K}_i$ be the Poincare dual of the solid tori K_i. In the usual way (Thom classes and stuff) we can choose B_i to be a bump function oriented orthogonally to K_i, with total integral 1 in any plane intersecting K_i transversely. If $\eta_i : S^3 \to K_i$ is the restriction, then

$$K_i = \eta_i^*(f_i(r) r dr \wedge d\theta), \tag{43}$$

where r, θ are the coordinates on a given cross section of the solid torus, and where $\int_0^{r_i} f_i(r) r dr \wedge d\theta = 1$.

Now, we claim that the answer for the writhe of the link L is the Chern-Simons-like result[2]

$$w(L) = \int_{S^3} \sum_{i,j} A_i \wedge B_j. \tag{47}$$

To see this, let us first look at the $i \neq j$ terms. These give

$$\int_{S^3} A_i \wedge B_j = \int_{S^3} A_i \wedge \eta_j^*(f_j(r) r dr \wedge d\theta). \tag{48}$$

The integrand will be non-zero when the Seifert surface S_i meets the support of $\eta_j^*(f_j(r) r dr \wedge d\theta)$. If z is the coordinate along the non-bounding cycle of the solid torus K_j, then we can write this as

$$\int_{K_j} A_{i,z} f_j(r) r dr \wedge d\theta \wedge dz = \sum_{p \in S_i \cap K_j} \int \pm f_j(r) r dr \wedge d\theta = \sum_{p \in S_i \cap K_j} \mathrm{sgn}(p). \tag{49}$$

[2]That Chern-Simons theory is counting the linking number is not surprising. Partly because it's topological so what else could it be doing, but also because the 2-point functions for the gauge field A reproduce the Gauss formula for the linking number. To get the propagator we need to invert d, i.e. we need to solve

$$\epsilon^{\mu\nu\lambda} \partial_\lambda \Delta_{\mu\nu}(x-y) = \delta(x-y), \tag{44}$$

where the derivative acts on x. We can do this with

$$\Delta_{\mu\nu}(x-y) = \epsilon_{\mu\nu\sigma} \frac{(x-y)^\sigma}{4\pi |x-y|^3}, \tag{45}$$

which one can check using $\nabla \cdot \frac{r}{|r|^3} = 4\pi\delta(r)$. Thus we have

$$\left\langle \int_C A_\mu dx^\mu \int_{C'} A_\nu dy^\nu \right\rangle = \frac{1}{4\pi} \int_C \int_{C'} dx^\mu \wedge dy^\nu \epsilon_{\mu\nu\lambda} \frac{(x-y)^\lambda}{|x-y|^3} = \mathcal{L}(C, C') \tag{46}$$

where we've used Gauss' formula for the linking number.

Hopefully the notation is clear: the $\pm$ signs are determined by whether the orientation of A_i agrees or disagrees with the normal vector determined by $\star\eta^*(dr \wedge d\theta)$ in S^3. Thus for $i \neq j$ we have

$$\int_{S^3} A_i \wedge B_j = \mathcal{L}(K_i, K_j). \tag{50}$$

Now for the $i = j$ case. This is slightly trickier, since $S_i \cap K_i$ isn't well-defined. To handle this, we split apart a given solid torus K into a bunch of smaller solid tori K_λ, so that K looks like a coaxial cable or one of those cables that are used in e.g. suspension bridges. The writhe of K is then given by the linking number of any two distinct K_λ, K_γ (the decomposition is such that as the framing of K twists around in S^3, all of the small component tori twist around rigidly, maintaining a constant cross-section). Each K_λ has a Poincare dual 2-form that we write similarly as

$$K_\lambda = \eta_\lambda^*(f_\lambda(r) r dr \wedge d\theta), \tag{51}$$

where now $f_\lambda(r)$ does not integrate to $1/2\pi$, but rather $\sum_\lambda \int f_\lambda(r) r dr \wedge d\theta = 1$, so that the f_λ are a bump-function resolution of the identity for f. Now let A_λ be the Poincare dual to the Seifert surface of K_λ, except modified so that it is a delta function supported on S_λ with weight $\int f_\lambda(r) r dr \wedge d\theta \neq 1$. Then we have, for $\lambda \neq \gamma$,

$$\begin{aligned} \int_{S^3} A_\lambda \wedge B_\gamma &= \sum_{p \in S_\lambda \cap K_\gamma} \pm \left(\int f_\lambda(r) r dr \wedge d\theta \right) \left(\int f_\gamma(r) r dr \wedge d\theta \right) \\ &= w(K) \left(\int f_\lambda(r) r dr \wedge d\theta \right) \left(\int f_\gamma(r) r dr \wedge d\theta \right), \end{aligned} \tag{52}$$

where we used the fact that the linking number of any two distinct smaller component tori is equal to the write of the link K that we are focusing on.

The $\lambda = \gamma$ terms are still problematic, but if we decompose K into a larger and larger number N of constituent smaller solid tori, then since the individual values of $\int f_\lambda(r) r dr \wedge d\theta$ go to zero, and since there are N^2 $\lambda \neq \gamma$ terms but only N $\lambda = \gamma$ terms, in the $N \to \infty$ limit, we can ignore the $\gamma = \lambda$ terms. Thus

$$\sum_{\gamma,\lambda} \int A_\gamma \wedge B_\lambda = w(K) \left(\sum_\gamma \int f_\gamma(r) r dr \wedge d\theta \right) \left(\sum_\lambda \int f_\lambda(r) r dr \wedge d\theta \right) = w(K). \tag{53}$$

Since the $\{f_\lambda\}$ are a partition of unity for f, and $\int f(r) r dr \wedge d\theta = 1$.

So the $i \neq j$ terms give us the linking numbers of the various link components, while the $i = j$ terms give us the write of each component, and we get

$$w(L) = \sum_{i,j} \mathcal{L}(K_i, K_j) + \sum_i w(K_i), \tag{54}$$

which is what we wanted to show.

The Hopf map

This is from Nakahara. Consider a map $f : S^{2n-1} \to S^n$. Do a few things:

- Write down the generator v for $H^n(S^n;\mathbb{R})$. Show that $f^*(v)$ is closed and exact, with
$$f^*(v) = d\omega. \tag{55}$$
- Define the Hopf invariant by the self-linking number of the worldlines defined by the Poincare dual of $d\omega$:
$$H(f) = \int_{S^{2n-1}} \omega \wedge d\omega. \tag{56}$$
Show that while ω is not uniquely determined, this ambiguity does not affect $H(f)$. Also show that if $g : S^{2n-1} \to S^n$ is homotopic to f, then $H(f) = H(g)$.
- Compute $H(f)$ in the case $n = 2$, where f is the Hopf map which relates coordinates on $S^2 \subset \mathbb{R}^3 = \langle x, y, z\rangle$ and $S^3 \subset \mathbb{R}^4 = \langle a, b, c, d\rangle$ by
$$x = 2(ad + cb), \qquad y = 2(bc - ad), \qquad z = a^2 + b^2 - c^2 - d^2. \tag{57}$$

◇ ◇ ◇

There are a few ways to write down the volume form. One which is kind of cool is
$$v = \frac{1}{r}\sum_i (-1)^{i+1} x_i dx^1 \wedge \cdots \wedge \widehat{dx}^i \wedge \cdots \wedge dx^{n+1}, \tag{58}$$
where the hat means that we omit that term. To see that this is the volume form, we wedge with the 1-form $dr = x_i dx^i$:
$$dr \wedge v = dx^1 \wedge \cdots \wedge dx^{n+1}, \tag{59}$$
which is the volume form on $\mathbb{R}^{n+1}$. Since we know that this splits as dr times the sphere part, v must indeed be a generator for the top cohomology of S^n.

Obviously v is closed, and so $f^*(v)$ is closed by virtue of the fact that exterior differential commutes with pullbacks. This can be proved by writing down the definitions. For example, if $h : X \to Y$ is a smooth map (so that locally we can invert the Jacobian) and γ is a 1-form on Y (just for simplicity), then one can check that both $df^*(v)$ and $f^*(dv)$ are given by
$$df^*(v) = f^*(dv) = \frac{\partial}{\partial y^\alpha}\alpha_\mu(h(y))\frac{\partial x^\mu}{\partial y^\beta}\, dy^\alpha \wedge dy^\beta. \tag{60}$$

Now $f^*(v)$ is a closed n-form on S^{2n-1}. Since S^{2n-1} has no nontrivial n-cycles to integrate this over, it must be exact. Thus we can write
$$f^*(v) = d\omega, \tag{61}$$
for some $n-1$ form ω.

Now for the second bullet point. ω is only defined up to an exact form. This doesn't affect $H(f)$ though, since shifting ω by an exact form changes the integrand by a total derivative.

Now, suppose $f \sim g$. We claim that since f and g are homotopic, their Hopf invariants must be equal. Indeed this is the case: to see this, set up the mapping cylinder $Cy = S^{2n-1} \times I$, where on $S^{2n-1} \times \{0\}$ we use the map f and on $S^{2n-1} \times \{1\}$ we use g. Then

$$H(g) - H(f) = \int_{S^{2n-1}\times\{1\}} \omega_g \wedge g^*(v) - \int_{S^{2n-1}\times\{0\}} \omega \wedge f^*(v). \tag{62}$$

Defining f_t in the natural way,

$$H(g) - H(f) = \int_{Cy} d\omega_t \wedge d\omega_t, \tag{63}$$

where $f_t^*(v) = d\omega_t$. Since pullbacks distribute over wedge products, we have

$$\begin{aligned} H(g) - H(f) &= \int_{Cy} f_t^*(v) \wedge f_t^*(v) \\ &= \int_{Cy} f_t^*(v \wedge v) = 0, \end{aligned} \tag{64}$$

since $v \wedge v = 0$ as the wedge product is being carried out in S^n and v is an n-form.

Finally for the third bullet point. This is kind of heinous and we won't write everything out. We will first deploy the form of v given above, so that

$$v = \frac{1}{4\pi r}(x\ dy \wedge dz - y\ dx \wedge dz + z\ dx \wedge dy). \tag{65}$$

Here we've normalized v by dividing by 4π so that it has integral periods.

Life can be made slightly easier by noting that $d(r^2) = x\ dx + y\ dy + z\ dz = 0$, so that the above becomes

$$v = -\frac{1}{4\pi}\frac{dx \wedge dy}{z}. \tag{66}$$

We know x and y in terms of the 4-sphere coordinates a, b, c, d, and so we can just substitute them in the above equation and simplify. This is done with the help of $a\ da + b\ db + c\ dc + d\ dd = 0$, but we won't write out all the algebra. We end up getting (sorry for the awful notation dd by the way)

$$f^*(v) = \frac{1}{\pi}(da \wedge db + dc \wedge dd). \tag{67}$$

This is a total derivative, and so we can identify ω with

$$\omega = \frac{1}{\pi}(a \wedge db + c \wedge dd). \tag{68}$$

Thus the integral is

$$H(f) = \frac{1}{\pi^2}\int_{S^3}(a \wedge db \wedge dc \wedge dd + c \wedge dd \wedge da \wedge db). \tag{69}$$

The two contributions are equal since the a, b, c, d are all on equal footing, and so we only need to worry about the first one.

We can do the integral by remembering that the $n+1$ sphere is a suspension of the n sphere. This means that we can build a coordinate system on the 3 sphere as an iterative process by attaching 1-sphere coordinates to $n < 3$ sphere coordinates that are already in place. Concretely, we start with the 1-sphere $(\cos\phi, \sin\phi)$. Then we add on a sphere to one of the components, so that we get the 2-sphere with coordinates $(\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi)$. Finally we add on another sphere to get the coordinates for the 3-sphere, namely

$$(a, b, c, d) = (\cos\phi\cos\theta\cos\psi,\ \cos\phi\cos\theta\sin\psi,\ \cos\phi\sin\theta,\ \sin\phi). \tag{70}$$

We then have

$$db \wedge dc \wedge dd = \cos^3\phi\cos\theta^2\cos\psi\ d\psi d\theta d\phi, \tag{71}$$

so that the integral is

$$H(f) = \frac{2}{\pi^2}\int \cos^4\phi\cos^3\theta\cos^2\psi\ d\psi d\theta d\phi, \tag{72}$$

where the integral over ϕ runs from 0 to 2π and the others go from $-\pi/2$ to $\pi/2$. The integral is $\pi^2/2$, and so

$$H(f) = 1. \tag{73}$$

Reminder about Stiefel-Whitney classes

In this entry we will remind ourselves what the first and second Stiefel-Whitney classes are, and what they obstruct. In dimensions $d \leq 4$, we will show that the second Stiefel-Whitney class, when capped with the fundamental class of an orientable 2-submanifold S, computes the mod 2 self-intersection number of S. This will be done with the help of the Whitney sum formula:

$$w(E \oplus F) = w(E) \cup w(F), \tag{74}$$

where $w = 1 + w_1 + w_2 + \dots$ is the total Stiefel-Whitney class. Finally, we will convince ourselves that all orientable 3-manifolds are spin (we can do so without doing computations with Wu classes). "The Wild World of Four Manifolds" is good inspirational reading for this diary entry.

$\diamond\ \diamond\ \diamond$

Suppose we're given a manifold M of dimension m. Consider the tangent bundle of M. This is a vector bundle with structure group $O(m)$ (reduced from $GL(m;\mathbb{R})$ with the help of a metric), where $O(m)$ acts on the orthonormal frames e^i_α on each patch U_α. Let $t_{\alpha\beta} \in O(m)$ be the transition function between frames on different patches, so that $e^i_\alpha = [t_{\alpha\beta}]^i_j e^j_\beta$. We will throughout assume we are working with a cover which is refined sufficiently so that all the patches and their intersections are simply connected. We can get such a cover by starting from a triangulation and taking each U_α to be the maximal star-shaped region containing only the 0-handle labeled α.

First Stiefel-Whitney class: Define w_1 as the cohomology class of the determinant Cech 1-cochain

$$w_1 = [\det t_{\alpha\beta}] \in H^1(M; \mathbb{Z}_2), \tag{75}$$

where the cohomology is Cech cohomology. From the cocycle condition on the transition functions and multiplicative property of the determinant, we see that $\delta w_1 = 0$. Suppose we perform a rotation on the framing at each patch by $e^i_\alpha \mapsto [h_\alpha]^i_j e^j_\alpha$. We see that this acts on $g_{\alpha\beta}$ with the adjoint action, and so

$$w_1 \mapsto [\det h_\alpha \det^{-1} h_\beta \det t_{\alpha\beta}] = [\det t_{\alpha\beta}], \tag{76}$$

since $\det h_\alpha \det^{-1} h_\beta$ is a $\mathbb{Z}_2$ coboundary. Thus w_1 is independent of the exact choice of local frame. We should think of this as a $\mathbb{Z}_2$ gauge field: with the 0-skeleton of M living at the center of patches, $\det t_{\alpha\beta}$ naturally lives on the 1-skeleton, and so w_1 behaves like a $\mathbb{Z}_2$ gauge field, with local frame rotations playing the role of gauge transformations.

Now we mention what this has to do with orientability. If M is orientable then we can further reduce the structure group to $SO(m)$, thereby trivializing the $\mathbb{Z}_2$ bundle over the 1-skeleton and setting $w_1 = 0$. Conversely, if w_1 is a trivial cohomology class, then $\det g_{\alpha\beta} = \det h_\alpha \det^{-1} h_\beta$ for some local rotations h_α, h_β, and so M is seen to be orientable after performing a rotation of the framing. Thus, the first Stiefel-Whitney class w_1 measures an obstruction to orientability. This in turn represents an obstruction to finding m linearly independent sections of TM. The intuitive picture for this is that as we go around a path over which w_1 has nontrivial holonomy, two of the basis vectors e^i in any orthonormal frame must change places, so that the orientation flips. This can only be done if two of the e^i are linearly dependent at some point along the path. Equivalently, we can think of w_1 as measuring the obstruction to extending a trivialization of the frame bundle from the 0-skeleton (which can always be trivialized) to a trivialization of the 1-skeleton. If $\det g_{\alpha\beta} = -1$ then the frames at U_α and U_β cannot be smoothly connected, and so the trivialization cannot be extended across the 1-handle which links U_α and U_β.

Since w_1 is a 1-cocycle, it is Poincare dual to a codimension-1 submanifold $\widehat{w}_1$ of M. According the definition of w_1, the orientation flips when passing through $\widehat{w}_1$. Now if w_1 is nontrivial in cohomology, then $\widehat{w}_1$ is not a boundary. This means we can travel along a curve in M that intersects the $\widetilde{w}_1$ surface only once, and thus defining a global orientation is impossible. In this way, the parts of the 1-skeleton of M that intersect $\widehat{w}_1$ transversely are the parts of the 1-skeleton across which a trivialization on the 0-skeleton cannot be extended. In terms of the basis vectors for the framing, they must become degenerate when passing through $\widehat{w}_1$, so that $\widehat{w}_1$ determines the location where a collection of m basis vectors for the framing must fail to be linearly independent. Said another way, if N is any closed one-dimensional submanifold of M and $[N]$ is its fundamental class, then

$$w_1 \frown [N] \in \mathbb{Z}_2 \tag{77}$$

is equal to 1 if the orientation is preserved around N and -1 if the orientation is reversed.

Since w_1, a first cohomology class, measures the obstruction to having an orientation, we expect that different orientations are classified by a zeroth cohomology class. Of course this is true — one picks an orientation for each connected component of M — but let us see why this happens in a more systematic way. We start from the exact sequence

$$1 \to SO(m) \to O(m) \to \mathbb{Z}_2 \to 1, \tag{78}$$

where the second to last map is the determinant map. This gives rise to a long exact sequence in cohomology. Since the groups involved are not Abelian, we have to clarify exactly what cohomology we are dealing with. We will be thinking in a Cech frame of mind, and considering cochains valued in smooth functions (and not constant functions like in regular Cech cohomology, where we assign n-fold intersections of patches with constant group elements) which take values in some group (I'll try to not use sheafy language). Thus the relevant part of the long exact sequence we get is

$$\ldots \to H^0(M;\mathbb{Z}_2) \to H^1(M;\mathcal{C}^\infty(SO(m))) \to H^1(M;\mathcal{C}^\infty(O(m))) \to H^1(M;\mathbb{Z}_2) \to \ldots, \tag{79}$$

since cohomology with coefficients in $\mathcal{C}^\infty$ is the same thing as regular cohomology. Note that the Cech cohomology $H^1(M;\mathcal{C}^\infty(G))$ is only a set, and not a group, if G is not Abelian. This set comes with a distinguished "identity" element, and the exactness of the sequence means that the image of one map is the preimage of this distinguished element under the subsequent map.

We see that if the image of the map into $H^1(M;\mathbb{Z}_2)$ is non-zero, then by exactness the map $H^1(M;\mathcal{C}^\infty(SO(m))) \to H^1(M;\mathcal{C}^\infty(O(m)))$ cannot be surjective. This means that we have frame bundles that can be trivialized when using $O(m)$ as the structure group, but not $SO(m)$, meaning that the image of $H^1(M;\mathcal{C}^\infty(O(m)))$ determines the non-orientability of the manifold. Furthermore, suppose that M is orientable. Then the map $H^1(M;\mathcal{C}^\infty(SO(m))) \to H^1(M;\mathcal{C}^\infty(O(m)))$ is surjective, and so we can always reduce the structure group to $SO(m)$, since every $O(m)$ bundle comes from an $SO(m)$ bundle. Furthermore, from the exact sequence we see that different orientations are classified by $H^0(M;\mathbb{Z}_2)$, as promised.

Second Stiefel-Whitney class and beyond: Now for w_2. Assume now that M is orientable. As we will see, w_2 measures the obstruction to having a spin structure over M. Let $\widetilde{t}$ be a lift of the t transition functions (which since $w_1 = 0$ we can take to be in $SO(m)$) to $\mathrm{Spin}(m)$. Now the t's must satisfy the cocycle condition, but in general the $\widetilde{t}$'s do not: all that we require is that

$$(\delta\widetilde{t})_{\alpha\beta\gamma} = f_{\alpha\beta\gamma}\mathbf{1}, \qquad f_{\alpha\beta\gamma} \in \mathbb{Z}_2, \tag{80}$$

since the projection $\phi : \mathrm{Spin}(m) \to SO(m)$ has kernel $\ker\phi = \mathbb{Z}_2$, so if $f_{\alpha\beta\gamma} = -1$ then we still project down to a well-defined oriented frame bundle. There is an ambiguity in the choice of $\widetilde{t}$, since we can lift t to either $\widetilde{t}$ or $-\widetilde{t}$. Let us specify this choice of lift by

$$t_{\alpha\beta} \mapsto h_{\alpha\beta}\widetilde{t}_{\alpha\beta}, \tag{81}$$

where h is a Cech 1-cochain valued in $\mathbb{Z}_2$. We see that doing this changes $f_{\alpha\beta\gamma}$ by a 2-coboundary. The cohomology class of the $\mathbb{Z}_2$-valued 2-cochain $f_{\alpha\beta\gamma}$ is precisely w_2, and by construction it measures the obstruction to lifting an $SO(m)$ frame bundle to a Spin bundle: thus w_2 is the obstruction to having a spin structure over M.

How is this the natural degree-2 version of w_1? It is the natural analogue since while w_1 measured the inability to extend a trivialization of the frame bundle over the 1-skeleton of M, w_2 measures the obstruction to extend the trivialization over the 2-skeleton. We think about this as follows: since we have assumed $w_1 = 0$, we can trivialize the bundle over the 1-skeleton. Consider a 2-cell c, and look at what the framing does as we travel around ∂c. Since $\pi_1(SO(m)) = \mathbb{Z}_2$, we have two options. If the trivialization on ∂c traces out a trivial path, then we can extend the framing into c. However if it traces out the nontrivial loop in $SO(m)$, then we cannot extend the framing into c (the induced spin structure on ∂c is non-bounding). That this inability to extend the framing is captured by w_2 is seen by realizing that -1 in $\mathrm{Spin}(m)$ corresponds to the nontrivial element of $\pi_1(SO(m))$ (think about Deck transformations), and so if $(\widetilde{\delta t})_{\alpha\beta\gamma} = -1$ then the framing on the boundary of the section of the 2-skeleton determined by $U_\alpha \cap U_\beta \cap U_\gamma$ twists by 2π, and hence $w_2 = (\widetilde{\delta t})$ indeed measures the obstruction to extending the framing over the 2-skeleton. Also, analogously to the w_1 case, one can show that w_2 defines an obstruction to having $m-1$ linearly independent sections.

By Poincare duality, w_2 defines a codimension-2 submanifold $\widehat{w}_2$. This submanifold should be thought of as the region in M on which fermions cannot be defined. By the above arguments, we know that the framing must rotate by 2π along curves which link $\widehat{w}_2$, and so we can think about $\widehat{w}_2$ as representing the location of vortices around which the fermions pick up an extra minus sign: they essentially behave like 2π vortices in a type II superconductor (and w_2 would be nontrivial if we had an odd number of such vortices). More formally, we can say that $\widehat{w}_2$ is such that the 2-handles pierced by it are precisely those 2-handles over which the trivialization of the 1-skeleton of M does not extend.

We can see this things from a more algebraic perspective by looking at the exact sequence

$$1 \to \mathbb{Z}_2 \to \mathrm{Spin}(m) \to SO(m) \to 1, \tag{82}$$

where the second to last map is the projection. This gives rise to a long exact sequence in cohomology, the relevant part of which is

$$\ldots \to H^2(M;\mathbb{Z}_2) \to H^1(M;\mathcal{C}^\infty(\mathrm{Spin}(m))) \to H^1(M;\mathcal{C}^\infty(SO(m))) \to H^2(M;\mathbb{Z}_2) \to \ldots, \tag{83}$$

again since cohomology with coefficients in $\mathcal{C}^\infty(\mathbb{Z})$ is the same thing as regular cohomology with coefficients in $\mathbb{Z}$.

We see that if the image of the map into $H^2(M;\mathbb{Z}_2)$ is non-zero, then by exactness the map $H^1(M;\mathcal{C}^\infty(\mathrm{Spin}(m))) \to H^1(M;\mathcal{C}^\infty(SO(m)))$ cannot be surjective. This means that we have $SO(m)$ frame bundles that do not come from projecting a Spin bundle, meaning that the image of $H^1(M;\mathcal{C}^\infty(SO(m)))$ in $H^2(M;\mathbb{Z}_2)$ determines the inability of the manifold to possess a spin structure. Thus the image of $H^1(M;\mathcal{C}^\infty(SO(m)))$ in $H^2(M;\mathbb{Z}_2)$ is w_2. Like in the w_1 case, if we assume that M admits a spin structure, then by the above exact sequence we see that such spin structures are classified (non-canonically) by $H^1(M;\mathbb{Z}_2)$.

Now we want to mention another more geometric way of thinking about w_2, which works particularly well in four dimensions. Let $S \subset M$ be an oriented and closed 2-submanifold. Consider capping w_2 with the fundamental class of S. We get

$$w_2 \frown [S] = w_2(TM|_S), \tag{84}$$

where on the RHS we are indicating the Stiefel-Whitney class of the restriction of the tangent space to S. Now we use the fact that $TM|_S = TS \oplus TN$ where N is the part of the tangent space normal to the surface S. Now we use the Whitney product formula to conclude that

$$w_2 \frown [S] = w_2(TS) + w_2(TN) + w_1(S) \cup w_1(N), \tag{85}$$

since the terms on the RHS are the only possible things in the product of degree 2. Since S was assumed orientable, the last term vanishes. Since S is 2-dimensional, $w_2(TS)$ is actually the mod 2 reduction of the Euler class $e(S)$ (to be discussed in a little bit), and so will be zero for any compact orientable surface,[3] and so $w_2(TS) = 0$. Thus

$$\widehat{w}_2 \cap S = \widehat{w}_2(TN). \tag{86}$$

Now TN is also 2-dimensional, so that $w_2(TN)$ is the mod 2 reduction of $e(TN)$, which is a Poincare dual to a collection of points. In fact, $\widehat{w}_2(TN)$ precisely measures the self-intersection number of S! One can see this by taking another copy of S and displacing it slightly along directions determined locally by the normal bundle TN, so that this copy of S is a section of TN. Each time that this copy intersects the original surface will correspond to a zero of the section, which is measured by $e(TN)$. This is precisely what we do to define the self-intersection number.[4] Recapitulating, we have

$$\widehat{w}_2 \cap S = S \cap S \qquad \text{mod } 2, \tag{87}$$

where on both sides we really mean the signed intersection number. This is essentially where the splitting of w_2 into Wu classes comes from.

We now want to convince ourselves that all orientable 3-manifolds have $w_2 = 0$, and consequently all admit a spin structure. The half-hazard way of doing this is the following: consider as before an arbitrary closed oriented submanifold S of M. Again using the product formula, we have

$$w_2 \frown [S] = w_2(TS) + w_1(S) \cup w_1(N), \tag{88}$$

but again by the fact that $w_2(TS)$ is the mod 2 reduction of $e(S)$ and S was assumed orientable, this vanishes since $e(S) \in 2\mathbb{Z}$. We conclude that the integral of w_2 over any closed, oriented surface is zero (I think this implies that w_2 is zero, but perhaps it is feasible for w_2 to have a nontrivial cap product with an unorientable N? Should come back to this).

The third Stiefel-Whitney class w_3 is usually pretty boring, because it is zero if both w_1 and w_2 are zero. This is because

$$\pi_2(SO(m)) = 0 \tag{89}$$

for all m. This implies that, assuming we have been able to extend the trivialization onto the 2-skeleton, $w_3 = 0$. Indeed, there will be an obstruction to extending the trivialization into a given 3-cell c if the framing on ∂c winds around in a noncontractible way. But this never happens, since $\pi_2(SO(m)) = 0\,\forall m$. Thus if we have found a trivialization over the 0,1, and 2 cells, then we can always extend this to a trivialization over the 3-cells, and we don't need to worry about w_3 being nonzero.

[3] Just because $\chi(M) = 2 - 2g \in 2\mathbb{Z}$ vanishes mod 2.

[4] Recall that we do this by thinking of fermions as ribbons and defining the linking number as the intersection number of one edge of the ribbon with the Seifert surface determined by the other edge.

As a corollary, the Euler characteristic of any orientable three-manifold is even.[5] Now in the case of a 3-manifold M, since all the obstructions $\pi_i(SO(m))$ to extending the trivialization are $\mathbb{Z}_2$-valued, the vanishing of the SW class w_3 actually means that $e(M) = 0$, not just $e(M) \in 2\mathbb{Z}$, but in other cases where the relevant $\pi_i(SO(m))$ are not just $\mathbb{Z}_2$s, the vanishing of the SW classes only provide necessary, but not sufficient, conditions for extending the trivialization. As another way of re-stating the same thing, any orientable spin 3-manifold is parallelizable.

In general, the nonvanishing of the nth Stiefel-Whitney class indicates an obstruction to having $m - n + 1$ linearly independent nonvanishing sections of TM, or equivalently, it measures the obstruction to trivializing the frame bundle up to the n-skeleton of M (note that the nontriviality of w_n is a sufficient, but not necessary, condition for there not to exist $m - n + 1$ linearly independent non-zero sections). As we have mentioned, the top class w_m is the mod 2 Euler characteristic, by virtue of the fact that it is the $\mathbb{Z}_2$ reduction of the Euler class (representing the obstruction to fully trivializing the frame bundle throughout all of M). That is,

$$w_m \frown [M] = \chi(M) \mod 2. \tag{90}$$

Geometrically, the dual $\widehat{w}_m$ is a collection of points at which there is an obstruction to defining a single vector field, i.e. $\widehat{w}_m$ tells us where vector fields must vanish (so that e.g. $\widehat{w}_m$ is the location of the cowlicks on a hairy ball). Since w_m is a mod 2 index, it doesn't distinguish between positive and negative index critical points. In particular since $\chi(M) \in 2\mathbb{Z}$ for any closed, orientable 2-manifold M, the associated w_2 vanishes, and so fermions can be defined on any of these manifolds (naively one might think that a framing would be hard to define on S^2 since S^2 is not parallelizable, but this is not the case since $\chi(S^2) = 2$ [and of course fermions can exist on a 2-sphere]).

The statement made at the beginning of the last paragraph can be proved by using the Whitney sum formula. Indeed, suppose we have a bundle E which admits n linearly independent nonzero sections. Then we can decompose E as

$$E = E_\perp \oplus \mathcal{O}_n, \tag{91}$$

where $\mathcal{O}_n$ is a trivial rank-n bundle. Using the Whitney sum formula and the fact that $w_k(F) = 0$ if $k > \mathrm{Rank}(F)$ for any bundle F, we have

$$w(E) = w(E_\perp \oplus \mathcal{O}_n) = w(E_\perp) = 1 + w_1(E_\perp) + \cdots + w_{m-n}(E_\perp). \tag{92}$$

Therefore we conclude that if E has n linearly independent nonzero sections, $w_k(E) = 0$ for $k > m - n$. In particular, the non-vanishing of $w_{m-n+1}(E)$ implies that there cannot exist n such sections. Changing the names of the variables, we can equivalently say that $w_n(E) \neq 0$ implies that E does not admit $m - n + 1$ linearly independent nonzero sections.

[5] Of course, Poincare duality does better and tells us that all odd-dimensional manifolds actually have zero Euler characteristic.

Level repulsion for kindergarteners

Today's problem is super basic but was new to me so it was still worth a diary entry. Consider a 2-by-2 Hamiltonian

$$H = \begin{pmatrix} H_1 & \Delta \\ \Delta^* & H_2 \end{pmatrix}. \tag{93}$$

Assume that the probability distribution for the components of H is a Gaussian:

$$P(H_1, H_2, \Delta) \propto \exp\left(-\frac{1}{2}\mathrm{Tr}H^2\right). \tag{94}$$

Let E_+, E_- be the two eigenvalues of H (with $E_+ > E_-$), and define the variable $\xi = E_+ - E_-$. Find the probability distribution for ξ, first for real Δ and then for complex Δ.

◇ ◇ ◇

First let $\Delta \in \mathbb{R}$. It will be helpful to change to variables

$$H_\pm = \frac{H_1 \pm H_2}{2}, \tag{95}$$

in terms of which the eigenvalues are

$$E_\pm = H_+ \pm \sqrt{\Delta^2 + H_-^2}. \tag{96}$$

First we need the probability distribution for getting the two values E_+, E_-. This is

$$P(E_+, E_-) \propto \int dH_+ dH_- d\Delta\ \delta(E_+ - H_+ - \sqrt{\Delta^2 + H_-^2})\delta(E_+ - H_+ - \sqrt{\Delta^2 + H_-^2})e^{-(E_+^2+E_-^2)/2}. \tag{97}$$

Now do the integral over H_+:

$$P(E_+, E_-) \propto \int dH_- d\Delta\ \delta(E_+ - E_- - 2\sqrt{\Delta^2 + H_-^2})e^{-(E_+^2+E_-^2)/2}. \tag{98}$$

This is just an integral over $\mathbb{R}^2$, with $r = \sqrt{\Delta^2 + H_-^2}$ being the radial coordinate. Thus we get a factor of r in the integration measure and

$$P(E_+, E_-) \propto (E_+ - E_-)e^{-(E_+^2+E_-^2)/2}. \tag{99}$$

Thus we have

$$P(\xi) \propto \int dE_+ dE_-\ \delta(E_+ - E_- - \xi)(E_+ - E_-)e^{-(E_+^2+E_-^2)/2}, \tag{100}$$

so that

$$P(\xi) \propto \int dE_+\ \xi e^{-(E_+-\xi/2)^2-\xi^2/4} \implies P(\xi) \propto \xi e^{-\xi^2/4}. \tag{101}$$

Note that this goes to zero as $\xi \to 0$, which is level repulsion.

For the more general case of $\Delta \in \mathbb{C}$, the only thing that changes is we get a probability distribution like

$$P(E_+, E_-) \propto \int dH_- d\Delta_1 d\Delta_2 \; \delta(E_+ - E_- - 2\sqrt{\Delta_1^2 + \Delta_2^2 + H_-^2}), \tag{102}$$

and so now we're doing an integral over $\mathbb{R}^3$ instead and get an r^2 in the measure. Thus for $\Delta \in \mathbb{C}$ the level repulsion is even stronger, going as ξ^2:

$$P(\xi) \propto \xi^2 e^{-\xi^2/4}. \tag{103}$$

Another perspective on level repulsion comes from the Schor-Horn theorem, which says the following: if $\{d_1, \dots, d_n\}$ and $\{\lambda_1, \dots, \lambda_n\}$ are sets of real numbers, then there exists a Hermitian matrix H with diagonal entries $\{d_1, \dots, d_n\}$ and eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ iff the vector $(d_1, \dots, d_n) \in \mathbb{R}^n$ lies within the convex hull spanned by the vectors $(\lambda_{\sigma(1)}, \dots, \lambda_{\sigma(n)})$, for all permutations $\sigma \in S_n$. If we view $H = H_0 + V$ where H_0 is diagonal and V is a perturbation, then the fact that the diagonal entries of H are constrained to lie in the convex hull spanned by the vectors containing the eigenvalues of H means that the eigenvalues of H_0 must be "squeezed in" closer together than the eigenvalues of H; hence level repulsion.

We can see this explicitly from the baby 2×2 example. The fact that the trace and determinant are preserved when diagonalizing means that

$$H_1 H_2 = E_+ E_- + |\Delta|^2, \qquad H_1 + H_2 = E_+ + E_-, \tag{104}$$

which of course is obvious given that $E_\pm$ are easily solved for. The first equation tells us that the geometric mean of H_1 and H_2 is greater than that of E_1 and E_2. Now if we fix the sum $\sum_i x_i = C$ of a bunch of variables x_i, the geometric mean $(\prod_i x_i)^{1/N}$ is maximized for $x_i = C/N$. This means that H_1, H_2 must be closer to being equal than E_+, E_-. Since the sums of these two sets are equal, this then tells us that

$$\{H_1, H_2\} \in [E_-, E_+]. \tag{105}$$

This is in accordance with the Schor-Horn theorem, since the convex hull of the two vectors $(E_+, E_-), (E_-, E_+) \in \mathbb{R}^2$ is just the straight line connecting them; drawing the picture then makes it clear that (105) is satisfied.

Verlinde formula for finite groups

This is problem 10.8 in the big yellow book, chapter 10. In what follows, it is helpful to recall the orthogonality relations for characters of a finite group G, namely

$$\begin{aligned} \sum_{j \in \mathrm{Rep}(G)} \chi_j(a) \chi_j^*(b) &= \frac{|G|}{|C_a|} \delta_{a,b} \\ \frac{1}{|G|} \sum_{g \in G} \chi_j(g) \chi_k^*(g) &= \delta_{j,k} \end{aligned} \tag{106}$$

where a, b label representative elements of conjugacy classes C_a, C_b and $\text{Rep}(G)$ is the irreps of G. The second one is the counterpart of $\int dx e^{ix(k-k')} = \delta(k-k')$ and the first one is the dual.

Define the group S matrix as

$$S_j(a) = \sqrt{\frac{|C_a|}{|G|}}\chi_j(a). \tag{107}$$

The matrix elements here are $[S]_{ja} = S_j(a)$, with the mixing of conjugacy class and irreps allowed since for finite groups the number of conjugacy classes is the same as the number of irreps.

Use the orthogonality relations for the character to prove the Verlinde formula for the group G, relating the fusion coefficients of the group to the S matrix elements. Also prove a dual formula giving the structure constants of the class algebra in terms of the S matrix elements. Getting to these results requires several intermediate steps, which are sub-problems that are listed in the big yellow book—see there for details (although I'm not sure if the formula in part h is correct?).

◇ ◇ ◇

Let's first see if we can derive the orthogonality relations, just for fun. We'll get the second one first. We need Schur's lemma. Let $\psi : V_i \to V_j$ be a map between the two vector spaces that the irreps ρ_i, ρ_j act on, and let it be a map which commutes with the action of the representation, in that

$$\psi\rho_i(g) = \rho_j(g)\psi, \quad \forall g \in G. \tag{108}$$

Then since elements in $\ker\psi$ are thus killed by $\psi\rho_i(g)$, $\ker\psi$ is $\rho_i(g)$-invariant. Since ρ_i is an irrep, we must either have $\ker\psi = 0$ or $\ker\psi = V_i$. If it's the latter, then $\psi = 0$, so wolog we can take it to be the former. Similarly, since $\text{im }\psi$ is $\rho_j(g)$-invariant, then if $\text{im }\psi \neq 0$, $\text{im }\psi = V_j$. This means that either ψ is 0, or it is an isomorphism $\psi : V_i \cong V_j$. From $\psi\rho_i(g) = \rho_j(g)\psi$, we see that it is invariant under conjugation by $\rho_i(g)$ for all $g \in G$, and thus must be central in $GL(V)$. Thus it must be of the form $\psi = \gamma\mathbf{1}$ for some constant γ.

We can make maps that satisfy the condition above by averaging over the group. For any map ϕ and representations ρ_i, ρ_j, define the average by

$$\widetilde{\phi} = \frac{1}{|G|}\sum_{g\in G}\rho_i^{-1}(g)\phi\rho_j(g). \tag{109}$$

By using the linearity of the representations and shifting the sum, one checks that $\widetilde{\phi}\rho_i(g) = \rho_j(g)\widetilde{\phi}$. Thus $\widetilde{\phi} = 0$ unless $i = j$. Assume this is the case, so that $\widetilde{\phi} = \gamma\mathbf{1}$, and take the trace. Then using the cyclicity of the trace,

$$\gamma\dim_i = \frac{1}{|G|}\sum_{g\in G}\text{Tr}[\phi] \implies \widetilde{\phi} = \frac{\text{Tr}[\phi]}{\dim_i}\mathbf{1}, \tag{110}$$

where $\dim_i$ is the dimension of ρ_i.

To get the character orthogonality relation, pick some representation ρ_i let $\phi = E_{\alpha\beta}$ be one of the basis elements of $GL(\dim_i)$. Then $\widetilde{\phi} = \mathbf{1}\delta_{\alpha\beta}/\dim_i$. So then since the RHS is zero if the two representation were different, we can write

$$\delta_{ij}\frac{\delta_{\gamma\sigma}\delta_{\alpha\beta}}{\dim_i} = \frac{1}{|G|}\sum_{g\in G}[\rho_i(g^{-1})]_{\gamma\alpha}[\rho_j(g)]_{\beta\sigma}. \tag{111}$$

Now take $\gamma = \alpha$ and $\beta = \sigma$, and sum over α, β. On the RHS, this gives a sum over $g \in G$ of the product $\chi_i(g^{-1})\chi_j(g)$. On the LHS, this gives a sum over α, β of $\delta_{\alpha\beta}/\dim_i = 1$. So then

$$|G|\delta_{ij} = \sum_{g\in G}\chi_i(g^{-1})\chi_j(g). \tag{112}$$

Now we need the relation

$$\chi_j(g^{-1}) = \chi_j^*(g). \tag{113}$$

Proof: since G is finite, $g^n = 1$ for some $n \in \mathbb{Z}$, and thus $\chi_j(g^{-1}) = \chi_j(g^{n-1})$. Working in a basis where g is diagonal, since $g^{n-1}g = 1$, the matrix elements on the diagonals of g, g^{n-1} must be roots of unity, and since the product of each pair of diagonal entries is 1, they must be conjugates. Thus $\rho_j(g^{-1}) = \rho_j^*(g)$ and so $\chi_j(g^{-1}) = \chi_j^*(g)$.

Finally, we use that $\chi_i(g)$ is a class function to change the sum over $g \in G$ to a sum over conjugacy classes. Doing so gives

$$\sum_a |C_a|\chi_j(a)\chi_k^*(a) = |G|\delta_{j,k} \tag{114}$$

which is the orthogonality relation we wanted.

The other orthogonality relation is in some sense the dual of the first, where duality exchanges the conjugacy class index a with the representation index i. Conjugacy classes and irreps are dual to one another by way of the characters:

$$\chi : \mathrm{Rep}(G) \otimes \mathrm{Conj}(G) \to \mathbb{C}, \tag{115}$$

so that the intuition is $\mathrm{Rep}(G) \cong (\mathrm{Conj}(G))^*$. Thus we can think of the i index in $\chi_i(a)$ as a contravariant index and a as a covariant index. When G is Abelian $\mathrm{Conj}(G) = G$, and so we get the familiar $\mathrm{Rep}(G) \cong G^* \cong G$ for finite G, with the duality between group elements and representations being the Fourier transform.

To get the other orthogonality relation, define the character table matrix by $X_{ij} = \chi_i(g_j)$, where i is an irrep and g_j is some (representative of) a conjugacy class. Again, we're allowed to write it like this since there are just as many conjugacy classes as irreps[6] . Also defining the matrix $D_{ij} = \delta_{ij}|C_{g_j}|/|G|$, the orthogonality relation we already derived reads

$$\mathbf{1} = XDX^\dagger. \tag{117}$$

[6]Proof: the conjugacy classes $C_a = \sum_{g\in C_a} g$ form a basis of the center $Z(\mathbb{C}[G])$, so that the number of conjugacy classes is

$$N_C = \dim(Z(\mathbb{C}[G])). \tag{116}$$

On the other hand, $\mathbb{C}[G]$ is isomorphic to a product $\prod_i \mathrm{Mat}_{\dim_i}(\mathbb{C})$, where the product runs over irreps. Since the centers of the matrix algebras are all one-dimensional regardless of their size, the dimension of the center of $\mathbb{C}[G]$ is equal to the number of terms in the product, proving the claim.

Now we multiply by X^{-1} on the left and X on the right to get $\mathbf{1} = DX^\dagger X$. This means that

$$\delta_{ij} = \sum_k \frac{|C_{g_i}|}{|G|} \chi_k^*(g_i)\chi_k(g_j). \tag{118}$$

Switching back to denoting the conjugacy classes by a and rearranging gives

$$\sum_{j \in \mathrm{Rep}(G)} \chi_j(a)\chi_j^*(b) = \frac{|G|}{|C_a|}\delta_{a,b}, \tag{119}$$

which is what we wanted.

Now for the Verlinde formula. Consider the tensor product

$$\rho_i \otimes \rho_j = \bigoplus_k N_{ij}^k \rho_k. \tag{120}$$

Taking the trace of both sides, writing $\mathrm{Tr}[\rho_i \otimes \rho_j] = \mathrm{Tr}[\rho_i]\mathrm{Tr}[\rho_j]$, and then evaluating the representations on a particular (representative of) a conjugacy class C_a, we have

$$\chi_i(a)\chi_j(a) = \sum_k N_{ij}^k \chi_k(a). \tag{121}$$

Of course, the χ_j are class functions, and so the exact choice of representative of the conjugacy class is not important. Now multiply both sides by $|C_a|\chi_l^*(a)$ and sum over all conjugacy classes a, so that orthogonality lets us write (re-labeling $l \to k$)

$$N_{ij}^k = \frac{1}{|G|}\sum_a |C_a|\chi_i(a)\chi_j(a)\chi_k^*(a). \tag{122}$$

In terms of the S matrix,

$$N_{ij}^k = \sum_a \sqrt{\frac{|G|}{|C_a|}} S_i(a)S_j(a)S_k^*(a). \tag{123}$$

Since $\chi_0(a) = 1$ for all a, $S_0(a) = (|C_a|/|G|)^{1/2}$, and so we have the Verlinde formula

$$N_{ij}^k = \sum_a \frac{S_i(a)S_j(a)S_k^*(a)}{S_0(a)}. \tag{124}$$

A few comments on the S matrix. First, it is unitary. Indeed,

$$[SS^\dagger]_{ja} = \sum_b S_j(b)S_b^\dagger(a) = \sum_b S_j(b)S_a^*(b) = \sum_b \frac{|C_b|}{|G|}\chi_j(b)\chi_a^*(b) = \delta_{j,a}. \tag{125}$$

The S matrix is symmetric for Abelian groups, because for finite Abelian groups Rep and Vec are the same. If G is Abelian then every irrep is one dimensional, and it acts by multiplication by some $g \in G$, putting the irreps and the group elements (which since G is Abelian are the same thing as conjugacy classes) in bijection. In this case $S_j(a) = \chi_j(a)/\sqrt{|G|}$, and $\chi_j(a) = \chi_a(j)$ since the representations just act by group multiplication and G is Abelian. However, S is not symmetric as constructed for non-Abelian groups, since the conjugacy classes can have different orders which mess things up. For example in the group S_3,

$$S_{(12)}((123)) = \sqrt{\frac{2}{6}}\chi_{(12)}((123)) = -1/\sqrt{3}, \tag{126}$$

while

$$S_{(123)}((12)) = \sqrt{\frac{3}{6}}\chi_{(123)}((12)) = -1/\sqrt{2}. \tag{127}$$

Now we will get a dual version of the Verlinde formula, obtained with the $\mathrm{Rep}(G) \cong (\mathrm{Conj}(G))^*$ duality. Instead of summing over conjugacy classes to get the fusion coefficients, we will sum over irreps. To this end we introduce the class algebra. In what follows, when we write a conjugacy class C_a, we mean the sum $\mathbb{Z}[G] \ni C_a = \sum_{g\in C_a} g$. We can form an algebra with the classes through the following product operation:

$$\star : \mathbb{Z}[G] \times \mathbb{Z}[G] \to \mathbb{Z}[G], \qquad C_a \star C_b = \sum_c \mathcal{N}_{ab}^c C_c. \tag{128}$$

Just to be explicit, this is nothing more than

$$C_a \star C_b = \sum_{g\in C_a}\sum_{h\in C_b} gh. \tag{129}$$

Here the coefficients $\mathcal{N}_{ab}^c$ are integers and we can restrict to $\mathbb{Z}[G]$ from $\mathbb{C}[G]$. This is because if $C_c \ni f = gh$ appears in the product $C_a \star C_b$ for $g \in C_a, h \in C_b$ then so does $g'fg'^{-1} = (g'gg'^{-1})(g'hg'^{-1})$, so if one element in a conjugacy class appears on the RHS, all elements appear. This product is the conjugacy class dual of the $\otimes$ of irreps. As an example, for S_3 we have

$$C_{(12)} \star C_{(12)} = 3C_{\mathbf{1}} + 3C_{(123)}, \quad C_{(123)} \star C_{(123)} = C_{(123)} + 2C_{\mathbf{1}}, \quad C_{(123)} \star C_{(12)} = 2C_{(12)}. \tag{130}$$

Since we associate the sign representation of S_3 with the class of (123) and the two-dimensional representation with the class of $C_{(12)}$, there is no (is there? maybe we can normalize by the sizes of conjugacy classes or something?) obvious relation between the $\mathcal{N}$ coefficients and the N coefficients.

Now take both sides of the $\star$ product decomposition formula and act on them with some representation ρ_j, and then take the trace. This gives

$$\chi_j(C_a \star C_b) = \sum_c \mathcal{N}_{ab}^c \chi_j(C_c). \tag{131}$$

Now multiply by $\chi_j^*(C_d)$ and sum over all irreps $j \in \mathrm{Rep}(G)$. We can simplify the resulting expression by using the orthogonality of characters in their "covariant" representation index, which for the classes reads

$$\sum_{j\in\mathrm{Rep}(G)} \chi_j(C_c)\chi_j^*(C_d) = |C_c||C_d| \sum_{j\in\mathrm{Rep}(G)} \chi_j(c)\chi_j^*(d) = \delta_{c,d}|C_c||C_d|\frac{|G|}{|C_c|}. \tag{132}$$

since the characters are class functions. Using this and re-naming $d \to c$,

$$\mathcal{N}_{ab}^c = \frac{1}{|G||C_c|} \sum_j \chi_j(C_a \star C_b)\chi_j^*(C_c). \tag{133}$$

Now since $gC_a g^{-1} = C_a \quad \forall g \in G$, we have

$$C_a \in Z(\mathbb{C}[G]) \implies \chi_j(C_a) = \lambda_a \dim_j, \tag{134}$$

for some constant λ. Thus we have

$$\chi_j(C_a \star C_b) = \lambda_a \lambda_b \dim_j = \frac{1}{\dim_j}\chi_j(C_a)\chi_j(C_b) = \frac{|C_a||C_b|}{\dim j}\chi_j(a)\chi_j(b). \tag{135}$$

This means that

$$\mathcal{N}_{ab}^c = \frac{|C_a||C_b|}{|G|} \sum_{j \in \mathrm{Rep}(G)} \frac{1}{\dim_j}\chi_j(a)\chi_j(b)\chi_j^*(c). \tag{136}$$

Since $\chi_j(1) = \dim_j$, we have

$$\mathcal{N}_{ab}^c = \frac{|C_a||C_b|}{|G|} \sum_{j \in \mathrm{Rep}(G)} \frac{\chi_j(a)\chi_j(b)\chi_j^*(c)}{\chi_j(1)}. \tag{137}$$

Finally, we just need to write this in terms of the S matrix. This gives the dual Verlinde formula

$$\mathcal{N}_{ab}^c = \sqrt{\frac{|C_a||C_b|}{|C_c|}} \sum_{j \in \mathrm{Rep}(G)} \frac{S_j(a)S_j(b)S_j^*(c)}{S_j(1)}. \tag{138}$$

Note that because of these square roots, unless the group G is Abelian, the coefficients

$$\mathcal{M}_{ab}^c = \sum_{j \in \mathrm{Rep}(G)} \frac{S_j(a)S_j(b)S_j^*(c)}{S_j(1)} \tag{139}$$

will *not* generally be integers. From our expression for N_{ij}^k we saw that the "dual version" of the $\mathcal{M}_{ab}^c$'s (where we replace conjugacy classes with irreps) are integers, so unless G is Abelian there is an asymmetry between the conjugacy class approach and the irrep approach.

WZ consistency conditions and descent equations

In the BRST formalism, show that the BRST operator is exact when acting on both the gauge field and on the ghosts. Use these facts to derive the descent equations linking the different types of anomalies in different dimensions together. This is pretty standard stuff, but I got rather confused about signs / commutativity issues reading the presentation in e.g. Nakahara, and decided that for posterity's sake it'd be good to write out the details.

◇ ◇ ◇

We will let S be the BRST operator which implements gauge transformations along the fiber of the bundle $\mathcal{A}/\mathcal{G}$, where $\mathcal{A}$ is the space of all gauge field configurations and $\mathcal{G}$ is the group of local gauge transformations. The anomaly has a chance to be nontrivial when the cohomology of S is nontrivial, i.e. when the topology of $\mathcal{A}/\mathcal{G}$ is nontrivial.

As usual, we will work with a bicomplex of forms that have a degree under d (the degree of the differential form) and a degree under S (the ghost number), with d and S anticommuting:

$$dS + Sd = 0. \tag{140}$$

Both d and S have the following graded product rule:

$$d(a \wedge b) = da \wedge b + (-1)^{|a|} a \wedge db, \qquad S(a \wedge b) = Sa \wedge b + (-1)^{|a|} a \wedge Sb, \tag{141}$$

where $|a|$ is the total degree (differential form degree + ghost number) of a (it needs to be the total degree since d and S anticommute), and the wedge product is meant to take place in both complexes.

We need to figure out how S acts on the ghosts and the gauge fields. To do this, it is helpful to define a basepoint $A_\bullet$ from which we can reach other points in $\mathcal{A}$. We can reach a point A by (using math conventions for is and stuff)

$$A = g^{-1}(A_\bullet + d)g, \tag{142}$$

with $F = gF_\bullet g^{-1}$. Note that since $A, A_\bullet$ are differential 1-forms, they both anticommute with S. The ghost ω is defined by varying g along the fiber, and then transporting the variation back to the origin with g^{-1}:

$$\omega = g^{-1}Sg = g^{-1}\delta g|_{\text{fiber}}. \tag{143}$$

The action of S on the ghost is easy:

$$S\omega = -g^{-1}Sgg^{-1}Sg = -\omega^2, \tag{144}$$

so that

$$S^2\omega = -S\omega^2 = -(S\omega)\omega + \omega(S\omega) = \omega^3 - \omega^3 = 0, \tag{145}$$

since ω has total degree 1.

The action of S on the gauge field is (using $Sd = -dS$ and $A_\bullet = g(A-d)g^{-1}$)

$$\begin{aligned} SA &= S[g^{-1}(A_\bullet + d)g] = -g^{-1}SgA - g^{-1}A_\bullet Sg - g^{-1}d(Sg) \\ &= -\omega A - (A-d)g^{-1}Sg - g^{-1}d(gg^{-1}Sg) \\ &= -\omega A - A\omega - dw \\ &= -D_A\omega. \end{aligned} \tag{146}$$

S is also nilpotent on A, since

$$\begin{aligned} SSA &= -SD_A\omega \\ &= \omega^2 A + \omega(SA) - (SA)\omega - A\omega^2 - d(\omega^2) \\ &= \omega^2 A - \omega^2 A - \omega A\omega - \omega d\omega + \omega A\omega + A\omega^2 + (d\omega)\omega - A\omega^2 - (d\omega)\omega + \omega(d\omega) \\ &= 0. \end{aligned} \tag{147}$$

Let α_n be the anomaly, which is a top-dimensional form closed under S (so that $S\alpha_n = d\alpha_n = 0$). Here the subscript and superscript indicate that it is a differential form of degree n, with ghost number 0. Now $dS\alpha_n = 0$, so locally we have

$$\alpha_n = d\alpha^0_{n-1}. \tag{148}$$

Then we also have

$$dS\alpha^0_{n-1} = -S\alpha_n = 0, \tag{149}$$

so that locally

$$S\alpha^0_{n-1} = d\alpha^1_{n-2}. \tag{150}$$

The pattern continues, and we derive the descent equations

$$S\alpha^{k-1}_{n-k} = d\alpha^k_{n-k-1}, \tag{151}$$

for $0 \le k < n$.

The usual example of an anomaly which satisfies the descent equations is the $(n/2)$th Chern character

$$\alpha_n = \mathrm{ch}_{n/2}(F) = \frac{1}{(n/2)!}\mathrm{Tr}\left(\frac{iF}{2\pi}\right)^{n/2}. \tag{152}$$

This is the total derivative of a Chern-Simons term, whose gauge variation is the total derivative of ω wedged with the $(n/2-1)$th Chern character, and so on. For a $U(1)$ gauge field, the basic relations are

$$\alpha_4 = \frac{1}{8\pi^2}F\wedge F = \frac{1}{8\pi^2}d(A\wedge F) = d\alpha^0_3, \tag{153}$$

and

$$S\alpha_3 = \frac{1}{8\pi^2}d\lambda\wedge F = \frac{1}{8\pi^2}d(\lambda\wedge F) = d\alpha^1_2, \tag{154}$$

where we've written $d\lambda$ for the gauge transformation instead of ω for clarity.

Vandermonde determinant and matrix integrals

What is the integration measure on the space of N by N Hermitian matrices? Write it down in a form involving an integration over the eigenvalues.

◇ ◇ ◇

Let M be a Hermitian matrix. We can write it as $M = U\Lambda U^\dagger$, where U is unitary and Λ is diagonal. The distance on the space of Hermitian matrices is provided by the trace

$$ds^2 = \mathrm{Tr}(dMdM). \tag{155}$$

Now

$$dM = U(d\Lambda)U^\dagger + UU^\dagger(dU)\Lambda U^\dagger + U\Lambda(dU^\dagger)UU^\dagger = U(d\Lambda + U^\dagger dU\Lambda - \Lambda U^\dagger dU)U^\dagger, \tag{156}$$

since $(dU^\dagger)U = -U^\dagger dU$. Now define $\omega \equiv U^\dagger dU$. The trace distance is then

$$\mathrm{Tr}(dMdM) = \mathrm{Tr}[(d\Lambda + [\omega, \Lambda])^2]. \tag{157}$$

The commutator is

$$[\omega, \Lambda]_{ij} = \omega_{ij}\lambda_j - \omega_{ij}\lambda_i, \tag{158}$$

where $\Lambda_{ij} = \lambda_i\delta_{ij}$. Thus the trace of the commutator squared gives

$$\mathrm{Tr}([\omega, \Lambda]^2) = \sum_{ij} \omega_{ij}(\lambda_j - \lambda_i)\omega_{ji}(\lambda_i - \lambda_j) = -\sum_{ij} \omega_{ij}\omega_{ji}(\lambda_i - \lambda_j)^2. \tag{159}$$

Now define $dN \equiv -i\omega$. Now $\omega^\dagger = -\omega$, so that dN is Hermitian. Furthermore, we may choose dN to have zeros on the diagonal. Indeed, notice that the relation $M = U\Lambda U^\dagger$ is preserved under the map

$$U \mapsto UD, \qquad D \in U(1)^N. \tag{160}$$

The notation here means that D is a diagonal unitary matrix. Making this shift changes

$$\omega \mapsto D^\dagger U^\dagger(dU)D + D^\dagger dD. \tag{161}$$

We claim that we can choose D such that the diagonal part of this is zero. Doing so means that for all i,

$$D_i^\dagger dD_i = -D_i^\dagger \omega_{ii} D_i. \tag{162}$$

Now write $D = e^{iH^D}$, where H^D is Hermitian and diagonal. Also write $U = e^{iH^U}$, where H^U is Hermitian. Then $\omega = idH^U$, and the above equation is solved provided that we take

$$H_i^D = -H_{ii}^U. \tag{163}$$

Thus, taking $U_{ij} \mapsto U_{ij}e^{-i\delta_{ij}H_{ii}^U}$ kills of the diagonal parts of dN. Thus we may write

$$dN_{ij} = \frac{da_{ij} + idb_{ij}}{\sqrt{2}}, \qquad dN_{ii} = 0, \tag{164}$$

where a, b are both real, so that a is symmetric and b is antisymmetric. With this notation we have

$$\mathrm{Tr}([\omega, \Lambda]^2) = \frac{1}{2}\sum_{ij} dN_{ij}dN_{ji}(\lambda_i - \lambda_j)^2 = \sum_{i<j} |dN_{ij}|^2(\lambda_i - \lambda_j)^2 = \sum_{i<j}(da_{ij}^2 + db_{ij}^2)(\lambda_i - \lambda_j)^2. \tag{165}$$

The quadratic in $d\Lambda$ part of $\mathrm{Tr}(dMdM)$ is just $\sum_i d\lambda_i^2$, while the remaining term is

$$\mathrm{Tr}[\{d\Lambda, [\omega, \Lambda]\}] = 0, \tag{166}$$

since Λ and $d\Lambda$ commute. Thus the metric is

$$\mathrm{Tr}(dMdM) = \sum_i d\lambda_i^2 + \sum_{i<j}(da_{ij}^2 + db_{ij}^2)(\lambda_i - \lambda_j)^2. \tag{167}$$

The integration measure is obtained from $\sqrt{\det g}$. In our case, $\det g$ is just

$$\det g = \prod_{i<j}(\lambda_i - \lambda_j)^4 \implies \sqrt{\det g} = \prod_{i<j}(\lambda_i - \lambda_j)^2 \equiv \Delta^2(M). \tag{168}$$

The fact that the difference in eigenvalues is raised to the fourth power on the LHS is since we get a factor of $(\lambda_i - \lambda_j)^2$ from both the da^2 and db^2 parts of the metric. Thus the integration measure is

$$\mathcal{D}M = \prod_i d\lambda_i \prod_{j<k} da_{jk} db_{jk} \Delta^2(M). \tag{169}$$

Thus the integration over a, b is like integrating over angles in spherical coordinates, the integration over λ is like integrating over the radial direction, and the Jacobian $\Delta^2(M)$ is like r or $r^2 \sin\theta$. If the function we're integrating is only a function of the eigenvalues of M, then the angular integration over a, b just results in a constant.

Compendium of facts about Lie algebras and weights

Today's problem is a little different — instead of solving a particular problem we will just learn about some math facts. I'm just going to go learn about weights and roots and all that, and write down what I learn. Hopefully the compendeum of facts will be useful later on.

◇ ◇ ◇

Let $\mathfrak{g}$ be a semisimple Lie algebra coming from a group G. Semisimple means $\mathfrak{g}$ is a $\oplus$ of simple Lie algebras, where a simple Lie algebra is one with no nontrivial ideals other than $\mathfrak{g}$ itself, i.e. no nontrivial subalgebras $\mathfrak{i}$ with $[\mathfrak{g}, \mathfrak{i}] \subseteq \mathfrak{i}$ other than $\mathfrak{i} = \mathfrak{g}$. An example of a simple Lie algebra is $\mathfrak{su}(2)$, and an example of a semi-simple one is $\mathfrak{g}_1 \oplus \mathfrak{g}_2$, where the latter has the nontrivial ideals $(\mathfrak{g}_1, 0)$ and $(0, \mathfrak{g}_2)$. Note that since $\mathbb{R}$ is Abelian, it has many nontrivial ideals (e.g. $\mathfrak{i} = \{0, x\}$ for $x \in \mathbb{R}$), and so $\mathbb{R}$ is not simple (all simple Lie algebras must be non-Abelian). Equivalently, one can say that a semisimple Lie algebra is one with no nontrivial Abelian ideals. Basically, semisimplicity is ruined by $U(1)$ factors in the Lie group. For example, $\mathfrak{u}(n)$ is not simple, since the diagonal elements (the scalar trace part) constitute a nontrivial ideal.

Let A be an element of $\mathfrak{g}$ such that the zero eigenvalue of the adjoint action Ad_A has greatest multiplicity. I.e. let A be such that $[A, H] = 0$ is satisfied for the greatest number of independent $H \in \mathfrak{g}$. Let H_i be a basis for the space $\ker(\mathrm{Ad}_A)$. The dimesion of $\ker(\mathrm{Ad}_A)$ is called the rank of $\mathfrak{g}$, which we will write as r. We will see momentarily that the H_i generate the Cartan subalgebra of $\mathfrak{g}$ (i.e. they generate a maximal torus in $\mathfrak{g}$).

Roots

Now let R_α denote the eigenvectors of the Ad_A action with eigenvalues $\alpha \neq 0$, so that $[A, R_\alpha] = \alpha R_\alpha$. In fact, all the non-zero eigenvalues α of the adjoint are non-degenerate, so that the eigenspace for α is one-dimensional (we won't proove this). This means we can write $\mathfrak{g} = \ker \oplus \mathrm{im}$ by doing (not bothering to distinguish the complexification $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C} = \mathfrak{g} \oplus i\mathfrak{g}$ from $\mathfrak{g}$ as is the physics tradition)

$$\mathfrak{g} = \mathfrak{t} \oplus \bigoplus_\alpha V_\alpha, \tag{170}$$

where V_α are the one-dimensional vector spaces generated by the R_α, and $\mathfrak{t}$ is the subalgebra generated by the H_i (the zero eigenspace of the adjoint action). Accordingly, we can decompose A as

$$A = c^i H_i + c^\alpha R_\alpha. \tag{171}$$

Now consider the action of Ad_A on $[H_i, R_\alpha]$. By expanding the commutator one gets

$$[A, [H_i, R_\alpha]] = \alpha [H_i, R_\alpha], \tag{172}$$

so that $\mathrm{Ad}_A([H_i, R_\alpha]) \in V_\alpha$. But $\dim V_\alpha = 1$, so we must have

$$[H_i, R_\alpha] = \alpha_i R_\alpha \equiv \alpha(H_i) R_\alpha \tag{173}$$

for some constant α_i which is the evaluation of α on the generator H_i. Now since $H_i \in \ker(\mathrm{Ad}_A)$, we have

$$0 = [A, H_i] = c^j [H_j, H_i] + c^\alpha \alpha_i R_\alpha. \tag{174}$$

Thus by linear independence of the R_α, we need $c^\alpha = 0$ for all α, and so $A = c^i H_i \in \mathfrak{t}$. Also, since we picked H_i arbitrarily from the generators of the kernel of Ad_A, we also need $[H_i, H_j] = 0$ for all i, j. Thus $\mathfrak{t}$ is the Cartan subalgebra of $\mathfrak{g}$. That is, the H_i constintute a maximal set of simultaneously diagonalizable generators of $\mathfrak{g}$. The A such that Ad_A has the biggest kernel is a then a linear combination of these diagonalized generators. The exact choice of linear combination is essentially just a choice of basis.

The α (or sometimes the α_i, or sometimes the R_α) are called the roots of the Lie algebra. Each α can be viewed as a vector in $\mathbb{R}^r$, with components α_i. The best way to think about things is to define α to live in the dual Lie algebra $\mathfrak{t}^*$ via

$$\alpha : \mathfrak{t} \to \mathbb{C}, \quad H \mapsto (\text{eigenvalue of } R_\alpha \text{ under } \mathrm{Ad}_H). \tag{175}$$

A math fact is that all of the H_i can be chosen to be Hermitian (wrt the Killing form, which is built out of the structure constants as $g_{ij} = f^l_{ik} f^k_{lj}$), and thus the α can be chosen to be real. This means that if α is a root then so too is $-\alpha$, with the associated eigenvector being $R_{-\alpha} = R^\dagger_\alpha$. This just follows from applying $\dagger$ to $[H_i, R_\alpha] = \alpha_i R_\alpha$. Another math fact is that if α is a root, then $n\alpha$ is only a root if $n = -1, 0, 1$. Finally, note from the $\oplus$ decomposition of $\mathfrak{g}$ that the number of roots for a given representation is $d - r$, where d is the dimension of the representaiton.

R_α naturally has the interpretation as a raising operator, while $R_{-\alpha}$ is naturally interpreted as a lowering operator. Indeed, if $|W_\mu\rangle$ is an eigenstate of H_i with eigenvalue μ_i (here

$\mu \in \mathbb{R}^r$ — recall that we can make the H_i simultaneously diagonal so i can be chosen freely from $\mathbb{Z}_r$), then

$$H_i R_{\pm\alpha}|W_\mu\rangle = ([H_i, R_{\pm\alpha}] + R_{\pm\alpha}\mu_i)|W_\mu\rangle = (\mu_i \pm \alpha_i)|W_\mu\rangle. \tag{176}$$

Also, by using the Jacobi identity for commutators with $A, R_\alpha, R_{-\alpha}$, one finds that $[A, [R_\alpha, R_{-\alpha}]] = 0$, so that $[R_\alpha, R_{-\alpha}]$ belongs to $\mathfrak{t}$. In fact, the precise linear combination is $[R_\alpha, R_{-\alpha}] = \alpha^i H_i$. Thus the commutator of the raising and lowering operators is consistent with being some sort of σ^z thing.

Given R_α and $R_{-\alpha}$, we can form an $\mathfrak{su}(2)$ Lie algebra by defining

$$H_\alpha = 2\frac{\alpha^i}{\alpha^2} H_i \equiv (\alpha^\vee)^i H_i. \tag{177}$$

This definition makes use of a metric (the Killing form $g_{ij} = f^k_{il} f^l_{jk}$), which is slightly annoying since metrics can be re-scaled. A perhaps better way to define the H_α is just by equating them to the commutator $[R_\alpha, R_{-\alpha}]$. Anyway, the generators of this $\mathfrak{su}(2)$ are then[7]

$$\mathfrak{su}(2)_\alpha = \langle R_\alpha + R_{-\alpha}, -i(R_\alpha - R_{-\alpha}), H_\alpha \rangle\,. \tag{179}$$

Think of these as X, Y, Z. Some people would undoubetdly prefer all of these generators to be divided by 2, but too late[8]. Checking that the $\mathfrak{su}(2)$ commutation relations (really, the $\mathfrak{sl}(2,\mathbb{C}) = \mathfrak{su}(2) \otimes \mathbb{C}$) are satisfied is straightforward. The H_α are known as the co-roots of the Lie algebra, and appropriately lie in $(\mathfrak{t}^*)^* = \mathfrak{t}$ through the map $H_\beta : \alpha \mapsto \alpha(H_\beta)$ (we will also equivalently refer to the coefficients $\alpha_i^\vee = 2\alpha_i/\alpha^2$ as the co-roots). The co-roots span a lattice in $\mathfrak{t}$ called $\Lambda_r^\vee(G)$.

The co-roots obey a quantization condition that will be very helpful to have when we think about monopoles. Namely, we have the following quantization condition:

$$\alpha_i^\vee \beta^i \in \mathbb{Z}, \qquad \forall\, \beta \in \Lambda_r(G),\ \alpha_i^\vee \in \Lambda_r^\vee(G), \tag{181}$$

which we can also write as $\beta(H_\alpha) \in \mathbb{Z}$ for all $\beta, \alpha \in \Lambda_r(G)$. The integers that occur in this quantization condition are known as the Cartan integers. In particular, from $[H_\alpha, R_\alpha] = 2R_\alpha$, we have $\alpha(H_\alpha) = 2$. These facts just follow from angular momentum quantization in $\mathfrak{su}(2)_\alpha$,

[7] Technically this is $\mathfrak{sl}(2,\mathbb{C})$ rather than $\mathfrak{su}(2)$. This is because we are really working with the complexification

$$\mathfrak{su}(2) \otimes \mathbb{C} = \mathfrak{sl}(2,\mathbb{C}). \tag{178}$$

Remember that the $\otimes$ here just means we are allowed to take $\mathbb{C}$-linear combinations of the generators of $\mathfrak{su}(2)$, rather than $\mathbb{R}$-linear combinations (it has nothing to do with the field we use for entries in the matrices of $\mathfrak{su}(2)$). When we allow for linear combinations like $aX + ibZ$, we no longer can satisfy Hermiticity, but tracelessness is preserved. Thus the complexification becomes $\mathfrak{sl}(2,\mathbb{C})$.

[8] An alternate normalization scheme along these lines would be e.g.

$$\mathfrak{su}(2)_\alpha = \left\langle \frac{R_\alpha + R_{-\alpha}}{\sqrt{2\alpha^2}}, \frac{R_\alpha - R_{-\alpha}}{i\sqrt{2\alpha^2}}, H_\alpha \right\rangle, \tag{180}$$

with $H_\alpha = \frac{1}{2}[R_\alpha, R_{-\alpha}] = \alpha^i H_i/\alpha^2$.

and are just saying that the eigenvalues of the diagonal generator of $\mathfrak{su}(2)$ when acting adjointly are integers. Thus we can think of $H_\alpha \in \Lambda_r^*(G)$, since it is a function

$$H_\alpha : \Lambda_r(G) \to \mathbb{Z}, \quad \beta \mapsto \beta(H_\alpha). \tag{182}$$

Thus we have the inclusion $\Lambda_r^\vee(G) \subseteq \Lambda_r^*(G)$. Since the co-roots always have integral eigenvalues in any representation, we can identify $\Lambda_r^\vee(G)$ with the kernal of the exponential map from $\mathfrak{g}$ to $\widetilde{G}$.

One important fact is that for all the simply-laced Lie algebras (like $\mathfrak{su}(N)$), which are the ones for which all the roots are the same length, we can normalize the metric so that $\alpha^2 = 2\,\forall\alpha \in \Lambda_r(G)$. This means that the co-roots and the roots are actually the same for the simply-laced case.

Weights

Weights are the eigenvalues of the H_i in general representations. Note that this means the roots are just the weights when the representation is taken to be the adjoint (but really one should think of them as living in $\mathfrak{t}^*$). Usually when people just say "weight", they mean a weight taken in a fundamental representation, but it seems like they often also mean a weight taken in any representation. Such a weight μ is determined by

$$AW_\mu = \mu W_\mu, \qquad H_i W_\mu = \mu_i W_\mu, \tag{183}$$

where W_μ is a vector transforming in some representation. The number of weights for a given representation is of course the dimension of that representation. We will write $\mu \in \Lambda_w(G)$ to denote a weight in the lattice formed by the weights of $\mathfrak{g}$, in all representations of G (unless specified otherwise). Note the weights really do depend on G, and not just on $\mathfrak{g}$. If we have a representation of $\mathfrak{g}$ that does not lift to one of G, then it is not included in $\Lambda_w(G)$. Since the roots are the weights in the adjoint, and since the adjoint representation of $\mathfrak{g}$ always lifts to one on G regardless of the choice of G, we have

$$\Lambda_r(G) \subseteq \Lambda_w(G). \tag{184}$$

Later we will identify the quotient $\Lambda_w(G)/\Lambda_r(G)$ with something involving a magnetic group.

One potential source of confusion is that several of the physics papers I've been reading do not distinguish between the fundamental weights and the weights. The fundamental weights w_i are defined by the orthongonality condition

$$w^i \alpha_j^\vee = \delta_j^i, \quad \forall \alpha_i^\vee \in \Lambda_r^\vee(G). \tag{185}$$

Thus the fundamental weights form a lattice which we can identify with $(\Lambda_r^\vee(G))^*$, since $\mathbb{Z}$-valued linear combinations of the w_i allow us to generate all $\mathbb{Z}$-valued functions on $\Lambda_r^\vee(\widetilde{G})$. The fundamental weights are the weights for the covering group $\widetilde{G}$, so that we can write the lattice they generate as

$$\Lambda_w(\widetilde{G}) = (\Lambda_r^\vee(\widetilde{G}))^*. \tag{186}$$

This means that the lattice formed by the fundamental weights encompasses the lattices formed by the weights for any other choice of Lie group G with Lie algebra $\mathfrak{g}$:

$$\Lambda_w(G) \subseteq \Lambda_w(\widetilde{G}), \tag{187}$$

since $\widetilde{G}$ is the "biggest" Lie group with Lie algebra $\mathfrak{g}$ (again, since the roots only care about the Lie algebra, $\Lambda_r^\vee(G) = \Lambda_r^\vee(\widetilde{G})$).

Since $\alpha(H_\beta) \in \mathbb{Z}$ for all roots α, β, we have $\Lambda_r(\widetilde{G}) \subseteq (\Lambda_r^\vee(\widetilde{G}))^* = \Lambda_w(\widetilde{G})$. We will show later on that

$$\Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G}) \cong Z(\widetilde{G}). \tag{188}$$

Again, note that we have been writing the various lattices to be a function of G or $\widetilde{G}$, rather than $\mathfrak{g}$. This is a little misleading for the root lattice, since it only depends on the Lie algebra. Indeed, since the roots come from the weights in the adjoint representation which does not see global things like $Z(\widetilde{G})$, it is sensitive only to $\mathfrak{g}$ and we may write $\Lambda_r(\mathfrak{g})$ or $\Lambda_r(\widetilde{G})$ instead of $\Lambda_r(G)$. However, the weights do know about the global stucture of the group through their normalization. The fundamental weights, which can be constructed of the dual of the co-roots, only depend on the Lie algebra. Thus, given a Lie algebra, we can always construct the lattice of fundamental weights. Depending on our choice of Lie group, the representations of the Lie group will be found by selecting out appropriate sublattices of $\Lambda_w(\widetilde{G})$.

To elaborate on this, one way to distinguish regular weights vs fundamental weights is by the following method. Let $H_I(G) \subset \mathfrak{t}$ be the set of all elements of the Cartan algebra that exponentiate to the identity, i.e.

$$\exp(H) = \mathbf{1} \in G \quad \forall\, H \in H_I(G), \tag{189}$$

where we have used the particular choice of exponential map $\exp : \mathfrak{g} \to G$ appropriate for the chosen Lie group. For example, if $G = \widetilde{G}$ is simply connected, then the lattice of such $H_I(G) \subset \mathfrak{t}$ is the co-root lattice. If $\mu \in \mathfrak{t}^*$ is such that $\mu(H) \in \mathbb{Z}$[9] for all $H \in H_I(G)$, then μ is a weight in $\Lambda_w(G)$[10].

For example, consider the exponentiation of Z in $\mathfrak{su}(2)$ (the only generator of the Cartan subalgebra). Suppose in our conventions for the exponential map that $\exp(Z)$ generates a 4π rotation about $\hat{z}$. Then $Z \in H_I(SU(2))$, since a 4π rotation is trivial in $SU(2)$. The associated weight $\mu_\pm(Z) = \pm 1$ is thus a weight vector in $\Lambda_w(SU(2))$, and so $\Lambda_w(SU(2)) = \mathbb{Z}$. However, if $G = SO(3)$ then $(Z/2)$ also exponentiates trivially, so that $Z/2 \in H_I(SO(3))$. But $\mu_\pm$ no longer assings integers to all the elements of $H_I(G)$, since $\mu_\pm(Z/2) = \pm 1/2$. Thus $\mu_\pm \notin \Lambda_w(SO(3))$, and the weight lattice of $SO(3)$ is smaller. It is in fact generated by the weight $\mu'(Z) = \pm 2$, which assigns Z to twice one of its eigenvalues. Then $\mu'(H) \in \mathbb{Z}$ for all $H \in H_I(SO(3))$, and so the weight lattice of $SO(3)$ is $\Lambda_w(SO(3)) = 2\mathbb{Z}$.

Finally, just to build intuition: for a defining representation of dimension d, the fundamental weights are a collection of d vectors in $\mathbb{R}^r$. The vectors connecting these vectors are the roots. This is because the roots are the raising and lowering operators for the $\mathfrak{su}(2)_\alpha$s,

[9]Recall that μ acts on H by assigning H to one of its eigenvalues.

[10]We are tacitly including a factor of $2\pi i$ in the exponential map if we are thinking about matrix algebras, so that $\exp(H) = e^{2\pi i H}$.

which connect states of different H_α eigenvalues (which are precisely the weights). The fact that the roots always come in pairs just comes from the fact that if a root vector connects two weights, then its negative also connects them, just in the opposite direction.

Weyl group

The Weyl group is the group that acts by rotating about the X (or Y) axes of the $\mathfrak{su}(2)_\alpha$'s. Stated another way, it is the symmetry group of the maximal torus $\mathfrak{t}$ (the Weyl groups for different choices of the maximal torus are of course isomoprhic via conjugation in $\mathfrak{g}$). Letting

$$\mathcal{W}_\alpha = \exp\left(i\pi\frac{R_\alpha + R_{-\alpha}}{2}\right), \tag{190}$$

the Weyl group acts on a general matrix $B \in \mathfrak{g}$ as

$$\text{Weyl} : B \mapsto \mathcal{W}_\alpha B \mathcal{W}_\alpha^\dagger. \tag{191}$$

The image of this is in $\mathfrak{g}$ since we translate "out" of the Lie algebra with the first exponential, but then translate back to the tangent space of the identity with the second exponential. From the definition of $\mathcal{W}_\alpha$, we see that it performs a π rotation about an axis perpendicular to the "quantization axis" of the $\mathfrak{su}(2)_\alpha$ (the dumb 1/2 factor is there because of the normalization conventions on the $\mathfrak{su}(2)_\alpha$ generators we chose). Equivalently, it reflects vectors through the hyperplane normal to the quantization axis of $\mathfrak{su}(2)_\alpha$ (namely, the vector $\alpha \in \mathbb{R}^r$). So this just subtracts off twice the projection of a weight μ onto the unit vector α from μ. That is,

$$\text{Weyl} : \mu^i \mapsto \mu^i - 2\frac{(\mu_j \alpha^j)\alpha^i}{\alpha^2} = \mu^i - \alpha^i \mu(H_\alpha). \tag{192}$$

From the angular momentum quantization in $\mathfrak{su}(2)_\alpha$, we see that μ is shifted by $n\alpha$, for some $n \in \mathbb{Z}$.

As an example, the Weyl groups of $SU(n)$ and that of $U(n)$ are both S_n, the symmetric group on n objects, which acts by permuting the entries of the (diagonal) matrices in $\mathfrak{t}$ ($SU(2)$ and $U(1)$ have the same roots since their Lie algebras only differ by a $U(1)$ factor which doesn't show up in the roots since the roots are defined by the adjoint action, which is blind to Abelian subalgebras). A math fact is that the action of the Weyl group maps roots to roots. In particular, if we are interested in the adjoint representation we can take $\beta = \alpha$ to identify β with $-\beta$.

Let's go through all this for the simplest possible example: $\mathfrak{g} = \mathfrak{su}(2)$ (again, this is really $\mathfrak{g} = \mathfrak{su}(2) \otimes \mathbb{C} = \mathfrak{sl}(2, \mathbb{C})$). Part of this we have already done above.

The Cartan subalgebra is one dimensional, with $\mathfrak{t}$ in our normalization being generated just by Z. The weights are the eigenvalues of Z, namely ± 1. The roots are the eigenvalues of the adjoint action of Z. The eigenvectors are $\sigma^\pm$, since $[Z, \sigma^\pm] = \pm 2\sigma^\pm$, and so the roots are ± 2. Thus the fundamental weight lattice and the root lattice are

$$\Lambda_w(\widetilde{G}) = \mathbb{Z}, \qquad \Lambda_r(G) = 2\mathbb{Z}. \tag{193}$$

The roots are ± 2, and are precisey the vectors which connect the fundamental weights ± 1 to eachother. If we were to take $G = SO(3)$, the root lattice would remain unchanged but the weight lattice would be changed, since the fundamental weights ± 1 now no longer lift to representations of the Lie group (since the map $\exp : \mathfrak{g} \to G$ is different for the two choices of G). The weight lattice is instead

$$\Lambda_w(SO(3)) = 2\mathbb{Z} = \Lambda_r(\mathfrak{su}(2)). \tag{194}$$

Thus in this case the weight and root lattices are equal. This is because all the representations of $SO(3)$ are obtained from the adjoint representation of $SU(2)$. Note that we again have that $\Lambda_w(G)/\Lambda_r(G) = 1$ is the center of $G = SO(3)$.

The Weyl group acts by sending $\mu \mapsto -\mu$, and so identifies the weights $+1$ and -1. Thus the weight lattice modulo the action of the Weyl group is

$$\Lambda_w(G)/\text{Weyl} = \mathbb{Z}_2. \tag{195}$$

This is the center of $SU(2)$, which is no accident. Similarly, we have

$$\Lambda_w(G)/\Lambda_r(G) = \mathbb{Z}_2, \tag{196}$$

which also is no accident. The global information about the group is tied up in various lattices in interesting ways, which we turn to now.

Global aspects and covering spaces

Let $\widetilde{G}$ be the simply connected universal cover of G, and let $\widetilde{T} \subset \widetilde{G}$ be a maximal torus of $\widetilde{G}$ that is the image of $\mathfrak{t}$ under the exponential map, which we define to act on $\mathfrak{t}$ by $E : H \mapsto \exp(2\pi i H)$ (note that E is a homonorphism acting on $\mathfrak{t}$ since we can freely combine exponentials). Now in general, given a Lie algebra $\mathfrak{g}$, there may be multiple exponential maps that one can define, each of which "deliniearizes" $\mathfrak{g}$ into a different Lie group. However, a math fact is that there is always a unique simply connected Lie group $\widetilde{G}$ which can be obtained from exponentiating $\mathfrak{g}$, such that all other Lie groups G obtained from $\mathfrak{g}$ are quotients

$$G = \widetilde{G}/\Gamma, \quad \Gamma \subseteq Z(\widetilde{G}). \tag{197}$$

The exponential map which takes us to $\widetilde{G}$ is the only one of the various exp maps that is injective, and until further notice, this map is the one we mean when we talk about the exponential map.

We write

$$\widetilde{T} \cong \mathfrak{t}/\Lambda_I(\widetilde{G}), \tag{198}$$

where the kernel of E being quotiented out by is the integer lattice, which consists of those elements $H \in \mathfrak{t}$ such that $\mu(H) \in \mathbb{Z}$ for all weights $\mu \in \Lambda_w(\widetilde{G})$. We write this as

$$\Lambda_I(\widetilde{G}) = \Lambda_w^*(\widetilde{G}), \tag{199}$$

where the duality involves the group $\mathbb{Z}$, so that $f \in \Lambda_w^*(\widetilde{G})$ means that f is a $\mathbb{Z}$-valued function on the weight lattice of $\widetilde{G}$. Finally, define the central lattice by $\Lambda_Z(\widetilde{G})$, which consists of all those elements in $\mathfrak{t}$ that get exponentiated to something in $Z(\widetilde{G})$. From the definition,

$$Z(\widetilde{G}) = \Lambda_Z(\widetilde{G})/\Lambda_I(\widetilde{G}). \tag{200}$$

Now we need the following: for any lattices Λ_a, Λ_b, we have

$$\Lambda_a \subseteq \Lambda_b \implies \Lambda_b^* \subseteq \Lambda_a. \tag{201}$$

Proof: let Λ_b be generated by the vectors e_i^b, so that any $l \in \Lambda_b$ can be written as $l = m^i e_i^b$, $m^i \in \mathbb{Z}$. Then any $f \in \Lambda_b$ becomes a function on Λ_a through linearity, since we can write any $l' \in \Lambda_b$ as $l' = n^i e_i^b = n^i(k_i e_i^a)$ for some $n^i, k_i \in \mathbb{Z}$. So indeed, $\Lambda_b^* \subseteq \Lambda_a^*$. We will also need

$$\Lambda_b/\Lambda_a \cong \Lambda_a^*/\Lambda_b^*. \tag{202}$$

Proof: by linearity, we can focus on where functions send a single generator of the lattices, so it is enough to consider the case where the lattices are one-dimensional, generated by the numbers e^a, e^b. Since $\Lambda_a \subseteq \Lambda_b$, we can take $e^a = k, e^b = 1$ for some $k \in \mathbb{Z}$ wolog, so that the quotient on the LHS is $\mathbb{Z}_k$. Now on the RHS, a function in Λ_b^* can assign the point k anything in $k\mathbb{Z}$ by linearity, while a function in Λ_a^* can assign the point k anything in $\mathbb{Z}$. Thus there are a $\mathbb{Z}_k$'s worth of functions on Λ_a that do not lift to functions on Λ_b, and so the quotient on the RHS is also $\mathbb{Z}_k$. Applying this argument for each generator of the lattice gives us the result.

One more fact we will need is that

$$\Lambda_Z(\widetilde{G}) \cong \Lambda_r^*(\widetilde{G}). \tag{203}$$

Note that $\Lambda_r^*(\widetilde{G}) = \Lambda_r^*(G)$ since the roots are defined wrt the adjoint action, which only cares about the structure of $\mathfrak{g}$.

Proof: let $H \in \mathfrak{t}$. Then $H \in \Lambda_Z(\widetilde{G})$ iff $\exp(2\pi i H)$ commutes with $\exp(iR_\alpha)$ for all $\alpha \in \Lambda_r(\widetilde{G})$. This is because the R_α generate the part of $\mathfrak{g}$ orthogonal to the maximal torus $\mathfrak{t}$, and so their exponentials generate all the stuff in the Lie group which does not commute with the exponential of H. Thus $H \in \Lambda_Z(\widetilde{G})$ iff we have $e^{-iR_\alpha}e^{2\pi i H}e^{iR_\alpha} = e^{2\pi i H}$. The BCH formula is simplified since $[H, R_\alpha] = \alpha(H)R_\alpha$:

$$e^X e^Y e^{-X} = e^{e^s Y}, \qquad [X, Y] = sY. \tag{204}$$

Making use of this gives $s = 2\pi i\alpha(H)$, and so we see that H is in the central lattice iff $\alpha(H) \in \mathbb{Z}$ for all roots α, which proves the claim.

Now we can use the results of the last few paragraphs to show that

$$Z(\widetilde{G}) \cong \Lambda_Z(\widetilde{G})/\Lambda_I(\widetilde{G}) \cong \Lambda_r^*(\widetilde{G})/\Lambda_w^*(\widetilde{G}) \cong \Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G}). \tag{205}$$

Phew! As a trivial check of this, for $G = SO(3)$ we get $Z(SU(2)) = \mathbb{Z}/2\mathbb{Z}$, as required.

As a less trivial check, we can look at $SU(3)$, which also lets us illustrate some other properites of the various lattices we've been talking about. $SU(3)$ has rank 2, and has three generating vectors for the fundamental weight lattice $\Lambda_w(SU(3))$. The roots are computed by forming the six different differences of pairs of the generating vectors for the fundamental

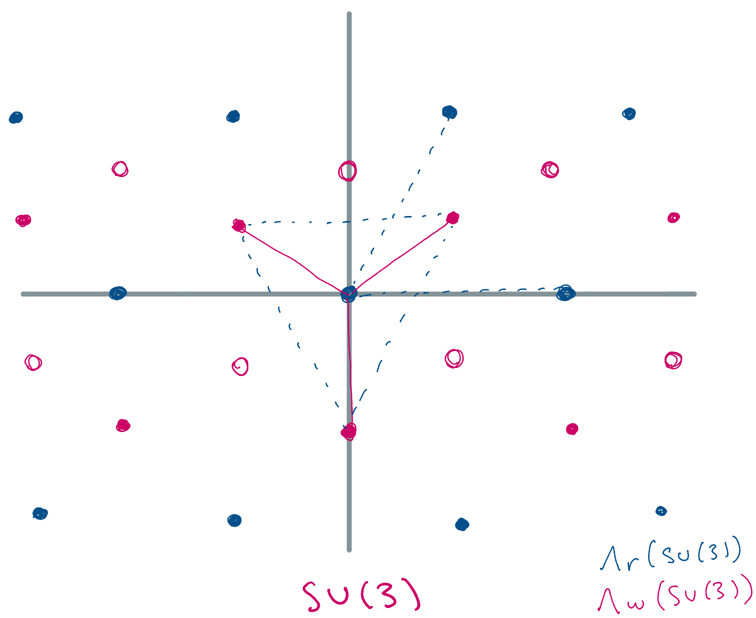


Figure 1: The fundamental weight lattice for $SU(3)$. The blue points indicate the sublattice formed by the root vectors.

weight lattice. This is shown in figure 1. The red solid dots indicate the basis vectors for the fundamental weight lattice, while the blue solid dots indicate the root lattice. The red open dots indicate the sublattice formed by adding any two of the fundamental weight lattice generators together, while adding any three fundamental weight lattice generators together gives something in the root lattice. Thus we have a three-fold periodicity: starting on a given sublattice and adding any three fundamental weight lattice generators will always given one a point in the original sublattice. Any point in a given sublattice can be connected to any other point in the same sublattice through a root vector, and so upon taking the quotient $\Lambda_w(SU(3))/\Lambda_r(SU(3))$, the whole lattice collapses to just three points. This is consistent with $Z(SU(3)) = \Lambda_w(SU(3))/\Lambda_r(SU(3)) = \mathbb{Z}_3$. Also note that the same sort of quotient can be taken by modding out by the action of the Weyl group, which performs reflections in the planes orthogonal to the roots. By taking a look at the figure, one can see that these reflections take the various sublattices to themselves, and that modding out by all three types of reflections collapses each lattice to a single point.

Now we see what can be said when the group in question is not the simply-connected $\widetilde{G}$, but something smaller. We will see that G is always the quotient of $\widetilde{G}$ by some finite subgroup. Now every representation of G lifts to a representation of $\widetilde{G}$ (one in which the deck transformations are represented trivially), and so $\Lambda_w(G) \subseteq \Lambda_w(\widetilde{G})$ (the weights form a sublattice of the lattice generated by the fundamental weights, which are not weights of G unless $G = \widetilde{G}$). Also, the root lattice $\Lambda_r(G) = \Lambda_r(\widetilde{G})$ is contained in the weight lattice, as we saw earlier (just since roots are particular choices of weights). Thus

$$\Lambda_r(G) \subseteq \Lambda_w(G) \subseteq \Lambda_w(\widetilde{G}). \tag{206}$$

Then we can use what we've learned in the last little section to take the dual (maps into $\mathbb{Z}$)

and produce

$$\Lambda_w^*(\widetilde{G}) \subseteq \Lambda_w^*(G) \subseteq \Lambda_r^*(G). \tag{207}$$

Now as we just saw the rightmost lattice is $\Lambda_Z(\widetilde{G})$, while the rightmost lattice is $\Lambda_I(\widetilde{G})$ since if an element in $\mathfrak{t}$ has integral eignevalues under the fundamental representation of $\widetilde{G}$ its exponentiation will be trivial (but elements in $\Lambda_w^*(G) \subset \mathfrak{t}$ are not neccessarily trivial when exponentiated!). So we get

$$\Lambda_I(\widetilde{G}) \subseteq \Lambda_w^*(G) \subseteq \Lambda_Z(\widetilde{G}). \tag{208}$$

Now we exponentate these sublattices. The rightmost one of course gives us $Z(\widetilde{G})$. Define

$$\exp(2\pi i \Lambda_w^*(G)) \equiv \Gamma_G. \tag{209}$$

Then we have

$$\Gamma_G \cong \Lambda_w^*(G)/\Lambda_I(\widetilde{G}), \tag{210}$$

since exp is a homomorphism when acting on $\mathfrak{t}$. So then

$$\Gamma_G \cong \Lambda_w^*(G)/\Lambda_w^*(\widetilde{G}) \cong \Lambda_w(\widetilde{G})/\Lambda_w(G) \tag{211}$$

measures the difference between the lattice generated by the fundamental weights and the actual weights of G. $\Gamma_G \subset Z(\widetilde{G})$ is the subgroup that determines how we obtain G from $\widetilde{G}$:

$$G = \widetilde{G}/\Gamma_G. \tag{212}$$

This in turn implies[11]

$$\pi_1(G) = \Gamma_G. \tag{216}$$

So to summarize, if we are given a Lie group G, it can always be obtained from its universal cover by quotienting by some subgroup Γ_G of $Z(\widetilde{G})$. This is one way to prove that if G is a topological group, $\pi_1(G)$ is always Abelian (since it is a subgroup of $Z(\widetilde{G})$.

Conversely, given any subgroup Γ_G of $Z(\widetilde{G})$ for $\widetilde{G}$ simply connected, one can always constuct an associated Lie algebra G satisfying $\widetilde{G}/\Gamma = G$ (by building it up from its weight lattice).

[11]Proof: intuitively, $\widetilde{G}$ is the universal cover of G, with fiber Γ_G. Thus upon taking the quotient, a path in $\widetilde{G}$ that moves between different sheets of the cover becomes a closed path in the quotient, and contributes to the fundamental group. More formally, consider the fibration

$$\Gamma_G \hookrightarrow \widetilde{G} \to \widetilde{G}/\Gamma_G. \tag{213}$$

This induces a long exact sequence in homotopy groups. Since $\pi_1(\widetilde{G}) = 0$ by assumption, the relevant part of the sequence is

$$0 \to \pi_1(\widetilde{G}/\Gamma_G) \to \pi_0(\Gamma_G) \to \pi_0(\widetilde{G}) \to \pi_0(\widetilde{G}/\Gamma_G) \to 0. \tag{214}$$

Now since Γ_G is discrete we have $\pi_0(\Gamma_G) = \Gamma_G$. Since this is finite, the next homomorphism from this into $\mathbb{Z}$ must be the zero map, since all elements in $\mathbb{Z}$ have infinite order. Thus using exactness of the sequence we obtain

$$\pi_1(\widetilde{G}/\Gamma_G) \cong \Gamma_G, \tag{215}$$

as claimed.

Dual lattices and dual groups

The last topic in our mini survey of Lie algebra stuff are GNO-dual or Langlands-dual Lie algebras, which are important when discussing magnetic charge quantization in gauge theories.

Given a Lie algebra $\mathfrak{g}$ and Lie group G, we can consider the root system $\mathfrak{g}^\vee$ formed by the coroots of $\mathfrak{g}$. This is a simple root system, and so we can always exponentiate it to form a simply-connected Lie group $\widetilde{G}^\vee$. The dual Lie group we are interested in will be a quotient of $\widetilde{G}^\vee$ by some subgroup of $Z(\widetilde{G}^\vee)$. The subgroup is selected by the requirement that the weights of G and $G^\vee$ be dual:

$$\Lambda_w(G^\vee) \equiv (\Lambda_w(G))^*. \tag{217}$$

The fundamental weights for the dual Lie algebra are by definition the dual of the co-root lattice of $\mathfrak{g}^\vee$. But the roots of $\mathfrak{g}^\vee$ are the co-roots of $\mathfrak{g}$, and so the co-roots of $\mathfrak{g}^\vee$ are the roots of $\mathfrak{g}$, since for any root α, we have $(\alpha^\vee)^\vee = \alpha$. Thus the fundamental weights for $\mathfrak{g}^\vee$ are the dual of the root lattice of $\mathfrak{g}$:

$$\Lambda_w(\widetilde{G}^\vee) = \Lambda_r(\widetilde{G})^*. \tag{218}$$

Using the formula for the center of $\widetilde{G}$ that we proved earlier on $\widetilde{G}^\vee$, we see that the universal covers of the group and its dual have the same center:

$$Z(\widetilde{G}^\vee) = \Lambda_w(\widetilde{G}^\vee)/\Lambda_r(\widetilde{G}^\vee) = \Lambda_r^*(\widetilde{G})/\Lambda_r^\vee(\widetilde{G}) = \Lambda_r^*(\widetilde{G})/\Lambda_w^*(\widetilde{G}) = \Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G}) = Z(\widetilde{G}), \tag{219}$$

where we used that the roots of $\widetilde{G}$ are the co-roots of $\widetilde{G}^\vee$ and vice versa.

Now as before, let Γ_G be such that $\widetilde{G}/\Gamma_G = G$. Recall that this group was computed by taking the quotient of the two weight lattices: $\Gamma_G = \Lambda_w(\widetilde{G})/\Lambda_w(G)$. This same formula holds for the dual group $\Gamma_G^\vee$, and so

$$\Gamma_G^\vee = \Lambda_w(\widetilde{G}^\vee)/\Lambda_w(G^\vee) = \Lambda_w^*(G^\vee)/\Lambda_w^*(\widetilde{G}^\vee) = \Lambda_w(G)/\Lambda_r(\widetilde{G}). \tag{220}$$

Now let's see how the two subgroups $\Gamma_G, \Gamma_G^\vee$ are related. We compute

$$\Gamma_G = \frac{\Lambda_w(\widetilde{G})/\Lambda_r(\widetilde{G})}{\Lambda_w(G)/\Lambda_r(\widetilde{G})} = Z(\widetilde{G})/\Gamma_G^\vee. \tag{221}$$

Thus the two subgroups which we use to obtain G and $G^\vee$ are complementary to one another inside of the cetner $Z(\widetilde{G})$. For example, suppose we are working with the group $\widetilde{G} = SU(ab)$, $a, b \in \mathbb{Z}$, and suppose that we obtain G by quotienting by $\Gamma_G = \mathbb{Z}_a$. Then from the above we see that $\Gamma_G^\vee = \mathbb{Z}_b$, and so

$$G = SU(ab)/\mathbb{Z}_a, \qquad G^\vee = SU(ab)/\mathbb{Z}_b. \tag{222}$$

all about instantons goes here!

Characteristic class manipulations for Pontryagin classes

Today we review what pontryagin classes are, and prove some results about their reductions mod 2 and mod 4. These results are helpful to have when dealing with topological terms generated by integrating out massive fermions.

◇ ◇ ◇

First some preliminaries on the Pontryagin classes.

In what follows we will frequently need to complexify real bundles, and realify complex ones. The sequence to keep in mind for complexifying and realifying is

$$U(n) \to SO(2n) \to U(2n). \tag{223}$$

The first map is used to turn a complex $n \times n$ matrix into a real $2n \times 2n$ one, while the second map is used to complexify a real $2n \times 2n$ matrix. The second map in the sequence comes from the inclusion $\mathbb{R} \to \mathbb{C}$, while the first map comes from

$$U(n) \ni A + iB \mapsto \mathbf{1} \otimes A + J \otimes B \in SO(2n), \qquad J = -iY, \tag{224}$$

with A, B real. Here J is how we represent i in $SO(2n)$. Why is the image of $A + iB$ in $SO(2n)$? For $A + iB$ to be unitary, we need

$$(A^T - iB^T)(A + iB) = \mathbf{1} \implies A^TA + B^TB = \mathbf{1}, \quad A^TB - B^TA = 0. \tag{225}$$

Now consider $\mathbf{1} \otimes A + J \otimes B$. Then since $J^T = -J$,

$$(\mathbf{1} \otimes A^T - J \otimes B^T)(\mathbf{1} \otimes A + J \otimes B) = \mathbf{1} \otimes (A^TA + B^TB) + J \otimes (A^TB - B^TA) = \mathbf{1} \otimes \mathbf{1}, \tag{226}$$

and so $A + JB$ is indeed orthogonal.

Anyway, on to Pontryagin classes. For a vector bundle E (usually a real vector bundle), they are defined by

$$p_j(E) = (-1)^j c_{2j}(E \otimes \mathbb{C}), \tag{227}$$

where $E \otimes \mathbb{C}$ is the complexification of E. Pontryagin classes almost obey the same sum formula as the Chern classes. Indeed (writing $\otimes$ for $\otimes_{\mathbb{R}}$),

$$p_j(E \oplus F) = (-1)^j c_{2j}(E \otimes \mathbb{C} \oplus F \otimes \mathbb{C}) = (-1)^j [c(E \otimes \mathbb{C}) \wedge c(F \otimes \mathbb{C})]_{2j} = [p(E) \wedge p(F)]_j + \dots, \tag{228}$$

where $\dots$ are terms that involve odd Chern classes. For example,

$$p_1(E \oplus F) = p_1(E) + p_1(F) - c_1(E \otimes \mathbb{C}) \wedge c_1(F \otimes \mathbb{C}). \tag{229}$$

Now the odd Chern classes of the complexification of a real bundle are 2-torsion[12], so that the Whitney sum formula holds for Pontryagin classes only up to 2-torsion elements. Another way to say this is to realize that if L is a real line bundle, then $L \otimes L$ is trivial, since $L^* \cong L$ by the reality of L means $L \otimes L \cong L \otimes L^* \cong \mathrm{Hom}(L, L)$, which always has a global section given by the identity map. Therefore any cohomology elements that classify real line bundles must be 2-torsion, and so the appropriate cohomology for describing real line bundles is $H^1(M; \mathbb{Z}_2)$. This means that when we map $H^1(M; \mathbb{Z}_2)$ into $H^2(M; \mathbb{Z})$, which classifies complex line bundles, we should get something that's 2-torsion. As we saw above, a similar statement holds for higher degrees.

The Pontryagin classes for a complex vector bundle E, defined by the Chern classes of $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ with $E_{\mathbb{R}}$ the realification of E, can easily be computed in terms of the Chern classes of E. If E is a complex vector bundle, then the isomorphism

$$E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C} \cong E \oplus \bar{E} \tag{232}$$

[12]Proof: Let $\mathcal{L}$ be a complex line bundle. Then we claim that $c_1(\mathcal{L}) = -c_1(\bar{\mathcal{L}})$. Indeed, the first Chern class of a complex line bundle is the same as the Euler class of the underlying real bundle, which locally is expressible in terms of the logarithms of the transition functions. Since $\bar{\mathcal{L}}$ has transition functions which are conjugate to those of $\mathcal{L}$, the Euler class associated to $\mathcal{L}$ is the negative of the one associated to $\bar{\mathcal{L}}$.

Now we use the splitting principle: assume E splits as a direct sum of line bundles, so that

$$E = \bigoplus_j \mathcal{L}_j \implies c(E) = \prod_j (1 + c_1(\mathcal{L}_j)) \implies c(\bar{E}) = \prod_j (1 - c_1(\mathcal{L}_j)). \tag{230}$$

Thus

$$c_k(E) = (-1)^k c_k(\bar{E}). \tag{231}$$

Now suppose that $E = F \otimes \mathbb{C}$, for F a real line bundle. Then $E = F \otimes \mathbb{C} = F \oplus iF$ is isomorphic to $\bar{E} = F \oplus (-iF)$ (this isomorphism does *not* hold if E is a generic complex vector bundle). Therefore by the splitting principle we can conclude that $c_k(F \otimes \mathbb{C}) = -c_k(F \otimes \mathbb{C})$ for k odd, meaning that the odd Chern classes of the complexification of a real bundle are all 2-torsion.

tells us that[13]

$$c(E \otimes_{\mathbb{R}} \mathbb{C}) = c(E \oplus \bar{E}) = (1 + c_1(E) + c_2(E) + \dots)(1 - c_1(E) + c_2(E) - \dots). \quad (238)$$

This relation shows that $c_{2k+1}(E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}) = 0$ since all the odd-degree parts cancel in pairs, and so the Pontryagin classes which are not of degree a multiple of 4 vanish for the realification of a complex vector bundle.

This whole song and dance of defining the Pointryagin classes in terms of the Chern classes of a complexified bundle is mainly just so that we can show that the p_k are only nonzero for $k \in 4\mathbb{Z}$, and that we can show the whitney sum formula for the p_i's. A simpler way to define them would be to use Chern-Weil and just write down the p's explicitly, but then we'd have to do invariant polynomials and stuff to see which ones could be non-zero. This is often the better way to go in terms of computing things, since the complexification is pretty trivial: we just take our real curvature form F_A, and allow ourselves to e.g. diagonalize it using complex numbers. But this approach has the disadvantage that we'd miss torsion phenomena: for example, using the expansion of $\det(\mathbf{1} + F/2\pi)$ it's easy to see that the p_i's obey a Whitney sum formula modulo torsion, but to see the torsion effects we need to work with the complexification.

One such expression is as follows. The general claim is that for an (oriented?) vector

[13] Why is this true? We just have to decipher the proof in Milnor. For any $z = x+iy \in E$, the corresponding element in $E_{\mathbb{R}}$ is obtained just by erasing the i and writing z as a tuple $(x, y) \in E_{\mathbb{R}}$. We create $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ (the fact that it's a $\otimes$ over $\mathbb{R}$ is important! We can't $\otimes$ a real bundle with something unless the tensor unit is $\mathbb{R}$) by forming the sum $E_{\mathbb{R}} \oplus E_{\mathbb{R}}$, and adding in the complex structure with the map $J : (z, w) \mapsto (w, -z)$, where z, w are both regarded as tuples of their real and imaginary parts. Now define the following two maps:

$$\mathcal{I}, \mathcal{I}^* : E \to E_{\mathbb{R}} \oplus E_{\mathbb{R}}, \quad \mathcal{I}(z) = (z, -iz), \quad \mathcal{I}^*(z) = (z, iz). \quad (233)$$

Just to be clear, if $z = x + iy$, we have $\mathcal{I}(z) = ((x, y), (y, -x)) \in E_{\mathbb{R}} \oplus E_{\mathbb{R}}$. Anyway, note that

$$J(\mathcal{I}(z)) = J(z, -iz) = (iz, -z) = \mathcal{I}(iz), \quad (234)$$

so that $\mathcal{I}$ is complex linear with respect to the complex structure on $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ (again, just to be clear, $(iz, -z) = ((-y, x), (-x, -y))$ as an element of $E_{\mathbb{R}} \oplus E_{\mathbb{R}}$). Similarly,

$$J(\mathcal{I}^*(z)) = J(z, iz) = (-iz, z) = -(iz, -z) = -\mathcal{I}^*(iz), \quad (235)$$

so that $\mathcal{I}^*$ is complex anti-linear with respect to the complex structure on $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$. Note however that any $(z, w) \in E_{\mathbb{R}} \oplus E_{\mathbb{R}} \cong E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ can be written uniquely as

$$(z, w) = \frac{1}{2} \left[\mathcal{I}(z + iw) + \mathcal{I}^*(z - iw) \right]. \quad (236)$$

This means that we can take any element in $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ and identify it with one element in the pre-image of $\mathcal{I}$, and one element in the pre-image of $\mathcal{I}^*$. The former is just E, while the latter is $\bar{E}$, since the complex structure on the pre-image of $\mathcal{I}^*$ is opposite to that of the complex structure on $E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$. So finally we can conclude that

$$E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C} \cong E \oplus \bar{E} \quad (237)$$

as claimed.

bundle, we have [4]

$$P(w_{2k}(E)) = p_i(E) + 2\sum_{j=0}^{k-1} w_{2j}(E) \cup w_{4k-2j}(E) \mod 4. \tag{239}$$

Here, the pontryagin square is a map into into $H^*(E;\mathbb{Z}_4)$; hence the mod 4 on the RHS. In particular,

$$P(w_2(E)) = p_1(E) + 2w_4(E) \mod 4. \tag{240}$$

Additionally, from the above general formula, we can conclude that

$$p_k(E) = P(w_{2k}) \mod 2. \tag{241}$$

Thus the mod 2 reduction of the Pointryagin class p_k is *not* given by w_{4k}, but rather by the square of w_{2k}.

This is easy to prove if we are dealing with the realification of a complex vector bundle. In that case, we can follow our earlier manipulations and write

$$p_k(E_{\mathbb{R}}) = c_{2k}(E_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}) = c_{2k}(E \oplus \bar{E}) = [c(E) \wedge c(\bar{E})]_{2k}, \tag{242}$$

where the brackets instruct us to take the degree $2k$ part. Expanding out the RHS,

$$p_k(E_{\mathbb{R}}) = 2\sum_{j=1}^{k-1} c_{2k-2j}(E) \wedge c_{2j}(E) + c_k(E) \wedge c_k(E), \tag{243}$$

where all the terms involving odd Chern classes have canceled. Working mod 4, and using that the mod 2 reduction of the Chern classes for a complex vector bundle E are

$$c_k(E) = w_{2k}(E_{\mathbb{R}}) \mod 2, \tag{244}$$

we obtain (239) (we also need to use the Pontryagin square as the appropriate cohomology operation on the mod 2 reduction of $c_k(E)$).

This proof relied on using the Whitney product formula for the Chern classes of a direct sum of *complex* bundles. Thus we generically won't have $E \otimes \mathbb{C} \cong F \oplus \bar{F}$ for some complex F, unless E happens to be the realification of a complex bundle (viz. $E = F_{\mathbb{R}}$). Since in this case we can't apply the Whitney product theorem, the proof is trickier[14]. The mod 2 version of (239), however, *is* easy to prove when E is a real vector bundle. Indeed, letting E be real, we have

$$p_k(E) = c_{2k}(E \otimes_{\mathbb{R}} \mathbb{C}) = c_{2k}(E \oplus iE). \tag{245}$$

As we mentioned earlier, we can't apply the Whitney product formula on this, since E and iE are *real* vector bundles. Anyway, writing $\rho_n(\cdot)$ for the reduction of $\cdot$ mod n, we have

$$\rho_2[p_k(E)] = \rho_2[c_{2k}(E \oplus iE)] = w_{4k}((E \oplus iE)_{\mathbb{R}}) = w_{4k}(E \oplus E) = P(w_{2k}(E)) \mod 2, \tag{246}$$

[14] Actually from the Chern-Weil point of view, such a product formula kind of makes sense, even though $c(E)$ is only defined for E complex. However since we are interested in torsion phenomena, thinking about expressions like $\mathrm{Tr}[F \wedge F]$ isn't really the road we want to take.

where in the last step we have used the Whitney product formula mod 2 on $w_{4k}(E \oplus E)$. If we change to working mod 4 it really seems like the natural thing that appears should be the rest of the terms in (239), but as of now I don't have a full proof.

For example, consider $\mathbb{CP}^n$, which has nontrivial cohomology in even degrees[15] The Chern classes of the tangent bundle are determined by the Whitney sum formula by taking the product of $n+1$ $\mathbb{C}$ line bundles:

$$c(T\mathbb{CP}^n) = (1+z)^{\wedge(n+1)}, \tag{248}$$

where z is the generator for $H^2(\mathbb{CP}^n;\mathbb{Z})$ and we have to remember to set $c_i = 0$ if $i > n$ (i.e. if $i = n+1$). The SW classes are then obtained by taking the mod-2 reduction of this (the only nonzero SW classes are even, since the odd SW classes of the realification of a complex bundle vanish). For example, take $n = 2$. Then we see that

$$w_2(T\mathbb{CP}^2) = z, \qquad w_4(T\mathbb{CP}^2) = z \cup z, \tag{249}$$

where we are implicitly working mod 2. For $n = 3$ all of the coefficients in the binomial expansion $(1,4,6,4,1)$ are even except the first and the last, and so the total SW class is (there is no contribution from the last 1 since z^4 is an 8-form, which is too big to live on $\mathbb{CP}^3$)

$$w(\mathbb{CP}^3) = 1. \tag{250}$$

Finally for $n = 4$, the binomial expansion is $(1,5,10,10,5,1)$ and so

$$w(T\mathbb{CP}^4) = 1 + z + z^{\cup 4}. \tag{251}$$

Now we can use our characteristic classes formula to find out what the pontryagin square of w_2 is in each of these cases, using our knowledge of p_1. The Pointryagin classes are determined from the Chern classes (remember that we have to complexify the bundle first!)

$$c(T\mathbb{CP}^n \otimes_{\mathbb{R}} \mathbb{C}) = c(T\mathbb{CP}^n \oplus (T\mathbb{CP}^n)^*) = c(T\mathbb{CP}^n) \wedge c((T\mathbb{CP}^n)^*). \tag{252}$$

Since $c((T\mathbb{CP}^n)^*)$ is the same as $c(T\mathbb{CP}^n)$ but with the signs of all the odd degree terms flipped, $c((T\mathbb{CP}^n)^*) = (1-z)^{n+1}$. So then since $p_k = (-1)^k c_{2k}$, we have

$$1 - p_1 + p_2 - \cdots = (1+z)^{n+1}(1-z)^{n+1} = (1-z^2)^{n+1}, \tag{253}$$

where we remember to set $p_k = 0$ if $k \geq \lceil (n+1)/2 \rceil$ by dimensionality reasons. We see that the minus signs in the p_k with k odd on the LHS will always match a minus sign on the RHS for z^{2k}. Thus the non-zero p_k are

$$p_k(T\mathbb{CP}^n_{\mathbb{R}}) = \binom{n+1}{k} z^2, \qquad k < \lceil (n+1)/2 \rceil. \tag{254}$$

[15] This comes from the cell decomposition of $\mathbb{CP}^n$, which we can motivate in the following way. First of all, $\mathbb{CP}^n$ consists of nonzero $n+1$ tuples $(z_0, \dots, z_n)$ modulo scaling by elements in $\mathbb{C}$. When $z_0 \neq 0$, we can normalize by it and get $(1, \widetilde{z}_1, \dots, \widetilde{z}_n)$, with $\widetilde{z}_i = z_i/z_0$. This space is $\mathbb{C}^n$, and it covers all of $\mathbb{CP}^n$ except at "infinity", where $z_0 = 0$. Thus to cover $\mathbb{CP}^n$, we need to attached the space of all $(0, z_1, \dots, z_n)$ to the $\mathbb{C}^n$ space of nonzero z_0. But we still have a re-scaling freedom on the $(0, z_1, \dots, z_n)$ that we place at infinity, so $\mathbb{CP}^n$ is realized by taking $\mathbb{C}^n$ and gluing it up with a copy of $\mathbb{CP}^{n-1}$ at infinity. Iterating this process, which stops at $\mathbb{CP}^0 = \mathbb{C}^0$, we see that

$$\mathbb{CP}^n = \mathbb{C}^n \cup \mathbb{C}^{n-1} \cup \cdots \cup \mathbb{C}^0, \tag{247}$$

which gives us the cell decomposition we're familiar with.

So for example,

$$p(T\mathbb{CP}^2_{\mathbb{R}}) = 1 + 3z^2, \quad p(T\mathbb{CP}^3_{\mathbb{R}}) = 1 + 4z^2, \quad p(T\mathbb{CP}^4_{\mathbb{R}}) = 1 + 5z^2 + 10z^4. \tag{255}$$

Now we can finally check our formula for p_1 mod 4 and the Pontryagin square. Our formula $P(w_2) = p_1 - 2w_4 \mod 4$ tells us that

$$P(w_2(T\mathbb{CP}^2_{\mathbb{R}})) = 3z^2 - 2z^2 \mod 4, \quad P(w_2(T\mathbb{CP}^3_{\mathbb{R}})) = 4z^2 - 2 \cdot 0 \mod 4, \tag{256}$$

and

$$P(w_2(T\mathbb{CP}^4_{\mathbb{R}})) = 5z^2 - 2 \cdot 0 \mod 4. \tag{257}$$

Thus for $\mathbb{CP}^2$ and $\mathbb{CP}^4$, since $w_2 = z$, we get $P(w_2) = w_2^2$, the usual cup square. For $\mathbb{CP}^3$ we get $P(w_2) = 0$, which we needed to get since $\mathbb{CP}^3$ is spin and has $w_2 = 0$. So, everything checks out!

Properties of Clifford algebras, their representations, and a compendium of results vis-a-vis the action of spacetime reflections on fermions

Today we're going to talk about how spacetime symmetries, in particular spacetime reflections, act on fermions. We will try to be as general as possible, covering both even and odd spacetime dimensions, real and imaginary time, and different choices of signature. However, we will stay within the context of free fermions, and won't discuss the transformation properties of any other fields they may couple to, or any other internal global symmetries that they might have. The point of this diary is to build up a fermion cheat-sheet that I can refer to later—the number of different conventions one can choose when dealing with fermions is huge, and so having a reference where the conventions are fixed is very helpful. This is likely all in the literature somewhere, but I wasn't able to find it all in one place (Witten's "fermions and path integrals" is definitely a good place to start, as is chapter 8 of Zinn-Justin).

◇ ◇ ◇

A word on notation: in this diary entry we will as usual be setting $J = -iY$, and will let $\mathcal{K}$ denote complex conjugation. $d = s + t$ will denote the dimension of spacetime, with s the number of spacelike-signature indices (positive signs in $\eta_{\mu\nu}$) and t the number of timelike-signature indices (negative signs in $\eta_{\mu\nu}$). We will assume throughout that there is only one time coordinate, so that the signature is either $(+, -, -, \dots)$, $(-, +, +, \dots)$, $(+, +, +, \dots)$, or $(-, -, -, \dots)$.

The business of working out the action of reflections on fermions is a mess because knowing just the representation theory of the spin groups is not sufficient — the spin groups of course only contain orientation-preserving group elements. The appropriate extensions of the spin groups needed to discuss reflections are the pin groups, and so statements like "the eigenvalue of a spinor under parity / time reversal" are totally meaningless: to talk about orientation-reversing symmetries, one must talk about their action on pinors, not spinors. The representation theory of the pin groups is rather gross, and the aim for this diary entry is to carefully go through and systematically list the different ways in which fermions can transform under reflections. Since charge conjugation often gets mixed into this business, we include a discussion of it as well.

Before we can talk about the pin groups and objects (pinors) which transform under them, we need to review a bit of representation theory.

Representation theory prelude

Suppose the representation $R : G \to \mathrm{Aut}(V)$ that the fermions transform in is isomorphic to its complex conjugate $\bar{R}$ (here we will assume that V is a complex vector space; when we discuss real fermions we will do so explicitly). Then there exists some map $\mathcal{J}$ that takes R to $\bar{R}$; because of the conjugation, $\mathcal{J}$ must be anti-linear. Since $\mathcal{J}^2$ is a linear map taking R to itself, by Schur's lemma we must have $\mathcal{J}^2 \propto \mathbf{1}$, and wolog we can re-scale $\mathcal{J}$ so that the constant of proportionality is ± 1 (but the sign cannot be eliminated with a rescaling by i due to the anti-linear nature of $\mathcal{J}$).

Suppose that $\mathcal{J}^2 = +\mathbf{1}$. In this case, we say that R is a *real* representation. Although we have been representing the group action as automorphisms on a complex vector space V, if the representation is real, we can also restrict the representation to a map $G \to \mathrm{Aut}(V_{\mathbb{R}})$, with $V_{\mathbb{R}}$ a real vector space (another diary entry on Pointryagin classes explains the connection between this and the fact that $\mathcal{J}^2 = +\mathbf{1}$).

Now suppose that $\mathcal{J}^2 = -\mathbf{1}$. In this case, we say that R is a *pseudoreal* or *quaternionic* representation; the reason for the former moniker is because the relations $\mathcal{J}^2 = (i\mathbf{1})^2 = -\mathbf{1}$ and $\mathcal{J}(i\mathbf{1}) = -(i\mathbf{1})\mathcal{J}$ give us a quaternionic structure formed from $i = i\mathbf{1}, j = \mathcal{J}, k = i\mathcal{J}$. In the quaternionic case, we cannot restrict the representation to an action on $\mathrm{Aut}(V_{\mathbb{R}})$.

Note that if R is a real representation and $\widetilde{R}$ is a pseudoreal representation, $R \otimes R$ and $\widetilde{R} \otimes \widetilde{R}$ are both real, while $R \otimes \widetilde{R}$ is pseudoreal (just by looking at the multiplication rule for the $\otimes$ of the $\mathcal{J}$s in question).

If R is real, we have a chance to define a theory with a single Majorana fermion, since then we can form a singlet under R by using a bilinear of a single field (if R is ps$\mathbb{R}$ it cannot be represented by purely real matrices, and therefore the reality condition of Majoranas cannot be preserved by the group action).

Finally, if $R \not\cong \bar{R}$, the representation is complex. In this case there is no invariant bilinear form that pairs two fields transforming in R, and so to construct an action we need to use two fields: one transforming in $\bar{R}$ and one transforming in R (usually called ψ and $\bar{\psi}$). We cannot define single Majorana fermions since we can't make a singlet using a bilinear of a single field (we can write actions in terms of Majorana fermions, but they need to mix flavors between a Majoranna in R and a different Majorana in $\bar{R}$).

One annoying thing is is that the character of the representations of $\mathrm{Spin}(d, 0) \equiv \mathrm{Spin}(d)$ and $\mathrm{Spin}(d-1, 1)$ are not the same (although thankfully $\mathrm{Spin}(d-1, 1)$ and $\mathrm{Spin}(1, d-1)$ have isomorphic representations). For posterity's sake, fixing the spacetime dimension as d, the character of the spin representations as represented over $\mathbb{C}$ are as follows: (in even dimensions, we list the character of the chiral reducible representation $S_\pm$)

d	Euclidean Time	Real Time
1	$\mathbb{R}$	$\mathbb{R}$
2	$\mathbb{C}$	$\mathbb{R}$
3	$\mathbb{H}$	$\mathbb{R}$
4	$\mathbb{H}$	$\mathbb{C}$
5	$\mathbb{H}$	$\mathbb{H}$

(258)

These are easy enough to check: in real time, the first three entries follow from using the matrix $J = -iY$ for the negative-signature coordinate and X, Z for the others, while e.g. the complexity of the $d = 4$ can be verified by computations we will do later and the $d = 5$ entry is from $\text{Spin}(4,1) = Sp(1,1)$. The Euclidean time entries follow from $\text{Spin}(1) = \mathbb{Z}_2, \text{Spin}(2) = U(1), \text{Spin}(3) = SU(2), \text{Spin}(4) = SU(2) \times SU(2)$, and $\text{Spin}(5) = Sp(2)$. Note how $\text{Spin}(d)$ has the same type of spinor representation as $\text{Spin}([d+2]-1, 1)$. The pattern can be continued up to higher d by using Bott periodicity (so that only d mod 8 is relevant).

Now from the above, we see that the representation theory of the spin group strongly depends on our choice of signature. Thus if G is the full symmetry group, the type of representation that G acts on the fermions with will depend on the choice of signature. This does not mean that the symmetries in the Lorentzian and Euclidean theories are different, it just means that the way in which the symmetries act is dependent on the choice of signature: this is true for the action of the spin group and for the action of the pin group (for example, a $\mathbb{R}$ time theory with T symmetry will continue to an $i\mathbb{R}$ time theory with T symmetry, but T will act differently in the $i\mathbb{R}$ time theory). For physics, we should only draw conclusions based on representation theory when working in real time.

Even when we restrict to $\mathbb{R}$ time, the signature matters for reflections. While $\text{Spin}(1, d-1) \cong \text{Spin}(d-1, 1)$, and $O(1, d-1) \cong O(d-1, 1)$, unfortunately $\text{Pin}(1, d-1) \ncong \text{Pin}(1, d-1)$, and $\text{Pin}(0, d) \ncong \text{Pin}(d, 0)$. The fact that pin groups in different signatures aren't isomorphic holds in even the simplest case of 0+1 dimensions, where $\text{Pin}(1,0) \cong \mathbb{Z}_2^2$, while $\text{Pin}(0,1) \cong \mathbb{Z}_4$ (in this case both Spin groups are trivial, and both orthogonal groups are $\mathbb{Z}_2$). Therefore the choice of signature actually is physical in a way—a choice needs to be made in order to determine the algebra obeyed by spacetime reflections, and not all choices of algebra and signature are mutually consistent.

Generalities on representations of Clifford algebras

Let $\mathcal{C}(s,t)$ denote the Clifford algebra generated by the γ matrices γ_μ, $\mu \in \mathbb{Z}_{s+t}$, with

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}, \tag{259}$$

where η has s positive diagonal entries and t negative ones. The γ^μ can always be built out of tensor products of Pauli matrices (with possible factors of i), and so for us will always be unitary.

We will let $\mathcal{C}_\pm(d)$ represent the elements in $\mathcal{C}(s,t)$ that are even / odd under $\gamma_\mu \mapsto -\gamma_\mu \,\forall\, \mu$, respectively. $\mathcal{C}(s,t)$ splits as

$$\mathcal{C}(s,t) \cong \mathcal{C}_+(s,t) \oplus \mathcal{C}_-(s,t). \tag{260}$$

$\mathcal{C}_+(s,t)$ forms a subalgebra of $\mathcal{C}(s,t)$, which we will see is related to $\mathrm{Spin}(s,t)$. The representation theory of $\mathcal{C}(s,t)$ depends on the parity of d, and is sorted out by using the matrix

$$\bar{\gamma} \equiv i^{(d^2-d)/2+t} \prod_{\mu} \gamma_\mu. \tag{261}$$

The dumb factor out front is to ensure that $\bar{\gamma}^2 = \mathbf{1}$ and $\bar{\gamma}^\dagger = \bar{\gamma}$ in every dimension and signature, so that (in even dimensions) it can be employed in the projectors $(\mathbf{1} \pm \bar{\gamma})/2$.[16]

If $d \in 2\mathbb{Z}+1$, $\bar{\gamma}$ commutes with all the γ_μ, and so it will be $\pm\mathbf{1}$. The map $\gamma_\mu \mapsto -\gamma_\mu \,\forall\, \mu$ changes the sign of $\bar{\gamma}$ and preserves the $\mathcal{C}(s,t)$ anticommutaion relations: thus we get two distinct representations, differing my the signs of the γ matrices. When $d \in 2\mathbb{Z}$ we only have one irreducible representation, since the map which changes the sign of all the γ_μ can be obtained by conjugating with $\bar{\gamma}$: $\bar{\gamma}^\dagger \gamma_\mu \bar{\gamma} = -\gamma_\mu$.

$\mathrm{Spin}(s,t)$ is defined as the elements in $\mathcal{C}_+(s,t)$ of unit norm. Since $\mathrm{Spin}(s,t) \subset \mathcal{C}_+(s,t)$, the two distinct representations of $\mathcal{C}(s,t)$ when $d \in 2\mathbb{Z}+1$ are indistinguishable in $\mathrm{Spin}(s,t)$, and in fact the spinor representation of $\mathrm{Spin}(s,t)$ is irreducible. When $d \in 2\mathbb{Z}$, we can form chiral projectors with $\bar{\gamma} \in \mathcal{C}_+(d)$, which commutes with everything in $\mathrm{Spin}(s,t)$. This means we can decompose the representation matrices of $\mathrm{Spin}(s,t)$ in a form which is block-diagonal in the ± 1 eigenspaces of $\bar{\gamma}$, meaning that the spinor representation of $\mathrm{Spin}(s,t)$ is reducible, with the spinor bundle splitting as $S_+ \oplus S_-$. Including spacetime reflections (elements in $\mathcal{C}_-(s,t)$) mixes sections of S_+ with those of S_- since the reflections all anticommute with $\bar{\gamma}$, and leaves us with only one irreducible representation.

The group $\mathrm{Pin}(s,t)$ is defined as the elements of $\mathcal{C}(s,t)$ of unit norm. We will be interested in a representation (the pinor representation) of $\mathrm{Pin}(s,t)$ on $\mathcal{C}(s,t)$, since this representation is what will allow us to determine how spacetime symmetries act on the fields (pinors) in the Lagrangian. This action is determined by the homomorphism

$$\Omega : \mathrm{Pin}(s,t) \to O(s,t) \tag{262}$$

defined for every $\Lambda \in \mathrm{Pin}(s,t)$ by

$$\Lambda^{-1} \gamma_\mu \Lambda = R_{\mu\nu} \gamma^\nu, \tag{263}$$

where $R_{\mu\nu} \in O(s,t)$. This transformation law is what allows $\bar{\psi} \partial\!\!\!/ \psi$, with $\bar{\psi}, \psi$ two pinor fields, to be invariant under the action of Lorentz transformations[17], since it means that $\bar{\psi}\gamma^\mu\psi$ transforms as a vector provided that $\bar{\psi}$ transforms inversely to ψ. The fact that $R_{\mu\nu} \in O(s,t)$ is required can be seen from requiring the anticommutation relations of the Clifford generators to be invariant under the action of $\mathrm{Pin}(s,t)$. From applying the action of $\mathrm{Pin}(s,t)$ on $\eta^{\mu\nu}$, we find

$$2\eta_{\mu\nu} = \{\gamma_\mu, \gamma_\nu\} \mapsto \Lambda^{-1}\{\gamma_\mu, \gamma_\nu\}\Lambda = R_{\mu\lambda} R_{\nu\sigma} \{\gamma^\lambda, \gamma^\sigma\} = 2R_{\mu\lambda}\eta^{\lambda\sigma}[R^T]_{\sigma\nu} \implies [R^T R]_{\mu\nu} = \eta_{\mu\nu}. \tag{264}$$

The homomorphism Ω is obviously not injective, since both Λ and $-\Lambda$ are associated with the same matrix R (this is why the Pin groups are $\mathbb{Z}_2$ extensions of the orthogonal

[16] Note that in other diary entries $\bar{\gamma}$ doesn't have the prefactor.

[17] here $\bar{\psi}$ is a copinor so that $\psi \mapsto \Lambda\psi \implies \bar{\psi} \mapsto \bar{\psi}\Lambda^{-1}$. More on this later.

groups). Whether or not Ω is surjective actually depends on whether d is even or odd. Indeed, consider the action on $\bar{\gamma}$. Then

$$\Lambda^{-1}\bar{\gamma}\Lambda = i^{(d^2-d)/2+t}R_{1\mu_1}R_{2\mu_2}\cdots R_{d\mu_d}\gamma^{\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_d} = R_{1\mu_1}\cdots R_{d\mu_d}\epsilon^{\mu_1\ldots\mu_d}\bar{\gamma} = (\det R)\bar{\gamma}. \tag{265}$$

Here we have used that if any $\mu_i = \mu_j$ but $i \neq j$, then we get a product $R_{i\mu_j}R_{k\mu_j}$ for $i \neq k$, which vanishes by the orthogonality of R. The transformation rule

$$\Lambda^{-1}\bar{\gamma}\Lambda = (\det R)\bar{\gamma} \tag{266}$$

is why $\bar{\gamma}$ is a pseudoscalar.

Now in odd dimensions, $\bar{\gamma}$ commutes with all of $\mathcal{C}(s,t)$, and so in odd dimensions we have $\bar{\gamma} = (\det R)\bar{\gamma} \implies \det R = 1$. Thus in odd dimensions we can only generate an action of $SO(s,t)$. In even dimensions $\bar{\gamma}$ anticommutes with $\mathcal{C}_-(s,t)$, and so if we take Λ to be generated by something in $\mathcal{C}_-(s,t)$, we can pick up matrices with $\det R = -1$, and we get the full $O(s,t)$ algebra. This is basically coming from the fact that unlike in even d, in odd d the matrix $-\mathbf{1}$ is the generator of the $\det R = -1$ part of $O(d)$, which is central and so $O(d) = SO(d) \times \mathbb{Z}_2$: the action of $\mathrm{Pin}(s,t)$ is unable to generate the decoupled $\mathbb{Z}_2$ factor. We can "fix" this and make Ω surjective by defining it instead through

$$(-1)^{s(\Lambda)}\Lambda^{-1}\gamma_\mu\Lambda = R_{\mu\nu}\gamma^\nu, \quad d \in 2\mathbb{Z}+1 \tag{267}$$

where $s(\Lambda) = 0$ if $\Lambda \in \mathcal{C}_+(s,t)$ and $s(\Lambda) = 1$ if $\Lambda \in \mathcal{C}_-(s,t)$.

Now let's look at how the various reflections in $O(s,t)$ are realized. First consider $P = 1 \oplus (-\mathbf{1}_{d-1})$, which acts as parity. We see that this is generated by $\Lambda_P = \gamma_0$. Now for $R_0 = -1 \oplus \mathbf{1}_{d-1}$, which reverses time: this is accomplished with $\Lambda_T = \prod_j \gamma_j$, where the product is over spatial indices. In general, $\Lambda = \gamma_\mu$ reflects all the axes of spacetime except for μ, and so a reflection about the axis μ can be performed with $\Lambda_{R_\mu} = \prod_{\nu\neq\mu}\gamma_\nu$. In even dimensions, we will find it slightly more convenient to write

$$\Lambda_{R_\mu} \equiv \gamma_\mu\bar{\gamma}, \tag{268}$$

which differs from $\prod_{\nu\neq\mu}\gamma_\nu$ by at most a c-number (from the potential power of i in $\bar{\gamma}$), and is easier to work with. In odd dimensions, the $\bar{\gamma}$ will be omitted—the reason for this will be explained in a sec.

The different choices of signature affect what reflections square to, because the choice of signature affects the types of pin group representations that exist. For example, for $d \in 2\mathbb{Z}$,

$$\Lambda^2_{R_\mu} = \gamma_\mu\bar{\gamma}\gamma_\mu\bar{\gamma} = -\gamma_\mu^2 = -\eta_{\mu\mu} \tag{269}$$

holds for all reflections about a single axis. When we talk about parity — reflection of all the spatial indices — we have to make a choice between simplicity and consistency of notation. We will opt for the former and define

$$\Lambda_P \equiv \gamma_0 \implies \Lambda_P^2 = \eta_{00}. \tag{270}$$

This is particularly nice since it means that spatial reflections and parity square to the same thing in real time, and lets the reversal of the time coordinate always square to the negative of Λ_P^2.[18]

Action of $\mathrm{Pin}(s,t)$ on fermions

We now elaborate on the general procedure for determining how C, P, and T act on fermion fields. One technical comment first: none of C, P, or T are connected to the identity in the Lorentz group (and C isn't part of it at all). Thus their actions are only really defined up to arbitrary phases which cancel out in Lorentz-invariant quantities. There are certain canonical choices to make and in what follows we will try to make them, but it is important to keep this ambiguity in mind.

In this section, we will be somewhat abstract and field-theory-centric, and will make use of the technology introduced in the previous section. The free fermion action, in real time, is

$$S = \int d^{d-1}x\, dt\, \bar{\psi} i \not{\partial} \psi. \tag{274}$$

In this expression, ψ is a pinor, and $\bar{\psi}$ is a copinor. This means that the action of $\mathrm{Pin}(s,t)$ is represented on the pinors via

$$\mathrm{Pin}(s,t) \ni g : \psi \mapsto \Lambda_g \psi, \quad \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_g. \tag{275}$$

We will *almost* be setting the barred representation to be the inverse of Λ_g representation, so that $\bar{\Lambda}_g$ is almost Λ_g^{-1}. There are some subtleties involved with reflections in odd dimensions, though, which means this relation won't always hold.

Now in general, Λ_g and $\bar{\Lambda}_g$ will be distinct representations of $\mathrm{Pin}(s,t)$, since the pinor representation may be complex. This means that $\psi, \bar{\psi}$ are generally *distinct* fields in the path integral. If the representation of $\mathrm{Pin}(s,t)$ the pinor ψ transforms under is isomorphic to its dual through some isomorphism $\mathcal{J}$, then we can use $\mathcal{J}$ to relate ψ and $\bar{\psi}$: this is the case where ψ is a Majorana, and because there is only one variable being integrated over in the path integral, the resulting partition function is a Pfaffian, not a determinant. In what follows though, we will simply define $\bar{\psi}$ as a copinor which transforms in the way defined above. When we are trying to be most general we will avoid writing the adjoint as $\bar{\psi} = \psi^\dagger \gamma_0$,

[18] The reason why this is a less consistent choice is that ideally, we would have defined

$$\Lambda_P = \prod_j \Lambda_{R_j}, \tag{271}$$

which after some combinatorial shenanigans is

$$\Lambda_P = i^{d^2/2 - 2d - d + 1 + d/2 - t} \eta_{00} \gamma_0, \tag{272}$$

which is disgusting. It squares to, when d is even,

$$\Lambda_P^2 = (-1)^{1-t+d/2} \eta_{00}, \tag{273}$$

which is really not very pretty. Hence we have used the simpler $\Lambda_P = \gamma_0$ in the main text.

since this will not be true if the fermions are Majorana and just adds more clutter. If the fermion is not Majorana then ψ and $\psi^\dagger$ are independent anyway, but $\bar{\psi}, \psi$ is a conceptually nicer set of independent fields to work with than $\psi^\dagger, \psi$.

Another general point worth mentioning is that in even dimensions, the notion of chirality only makes sense for spinors, not pinors (there are Weyl spinors, but not Weyl pinors). Mathematically, this is because $\mathrm{Pin}(s,t)$ has only a single irrep when $s+t$ is even, in contrast to the spin group $\mathrm{Spin}(s,t)$ which splits as $S_\pm$. Physically, this is just because reflections have determinant -1 and mix chiralities, so something that has a definite transformation rule under reflections cannot have a definite chirality.

Reflections

First for reflections / parity. First consider a reflection R_μ about the coordinate x^μ (if $\mu = 0$ we are reversing the flow of time, but not doing anything antilinear—the full antilinear time reversal will be discussed later). When $d \in 2\mathbb{Z}$, the appropriate action is to take $\psi \mapsto \Lambda_{R_\mu}\psi = \gamma_\mu \bar{\gamma}\psi$. Indeed, the invariance of the kinetic term is demonstrated by

$$\bar{\gamma}^\dagger \gamma_\mu^\dagger (-\partial_\mu \gamma^\mu + \sum_{\nu \neq \mu} \partial_\nu \gamma^\nu) \bar{\gamma}\gamma_\mu = \not{\partial}. \tag{276}$$

This works because when $d \in 2\mathbb{Z}$, $\bar{\gamma}$ anticommutes with all of the γ_μ.

When $d \in 2\mathbb{Z}+1$, we need something different: as we saw above, the homomorphism $\Omega : \mathrm{Pin}(s,t) \to O(s,t)$ is not surjective, and we cannot generate things with odd determinant. The solution to this is to twist the action of $\mathrm{Pin}(s,t)$ on $\bar{\psi}$ by a minus sign. So, we should do something like $\psi \mapsto \Lambda_{R_\mu}\psi$ and $\bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_{R_\mu} = -\bar{\psi}\Lambda_{R_\mu}^{-1}$. Said differently, recall that when d is odd, there are two distinct $\mathrm{Pin}(s,t)$ representations, which differ by the map $\gamma_\mu \mapsto -\gamma_\mu$. That is, if one representation is represented by the matrices $\Lambda_g, g \in \mathrm{Pin}(s,t)$, then the other is represented by the matrices $(\det \Lambda_g)\Lambda_g$, where $\det \Lambda_g$ is the function which splits apart $\mathcal{C}_+(s,t)$ and $\mathcal{C}_-(s,t)$.

If the pinor ψ transforms in one representation, the copinor $\bar{\psi}$ is taken to transform in the other representation, so that it picks up an extra minus sign when acted on by orientation-reversing elements on $\mathrm{Pin}(s,t)$. When d is odd we will thus take $\Lambda_{R_\mu} = \gamma_\mu$ and $\bar{\Lambda}_{R_\mu} = -\gamma_\mu$.[19]

From a slightly different point of view, the difference between odd and even dimensions can be understood in the following way (this paragraph will be in Euclidean signature, for simplicity). When reflections are included, we can consider pinors that are ± 1 eigenpinors under reflection (we can choose ± 1 wolog since $\gamma_\mu^2 = \mathbf{1}$ for all μ). These pinors will be sections of two pinor bundles, that we will denote as $\mathcal{P}$ ($+1$ eigenvalue) and $\mathcal{P}'$ (-1 eigenvalue). These two pinor bundles are related via $\mathcal{P} = \mathcal{P}' \otimes \varepsilon$, where ε is the orientation bundle; basically this is because sections of pinor bundles are glued with reflections along orientation-reversing transition functions, and so changing the signs of these transition functions by tensoring with ε is equivalent to sending $\gamma_\mu \mapsto -\gamma_\mu$ (see e.g. Witten's path integrals and fermions paper). Anyway, in even dimensions, the action of a reflection, $\gamma_\mu\bar{\gamma}, x^\mu \mapsto -x^\mu$, commutes with

[19] The reason for the absence of the $\bar{\gamma}$ here is just because we want the $\bar{\Lambda}_{R_\mu}$ matrix to be obtained from the $\Lambda_{R_\mu}^\dagger$ matrix through the substitution $\gamma_\mu \mapsto -\gamma_\mu$, which interchanges the representations, and since $\bar{\gamma} \mapsto -\bar{\gamma}$ under this map, we would not get the right minus sign if we included the $\bar{\gamma}$.

the Dirac operator $i\not{\partial}$. This means that $i\not{\partial} : \mathcal{P} \to \mathcal{P}$, $\mathcal{P}' \to \mathcal{P}'$, and the Dirac operator is self-adjoint, giving us a determinant.[20] In odd dimensions though, reflections *anticommute* with $i\not{\partial}$, and there is no way to ameliorate this with a factor of $\bar{\gamma}$. Thus in odd dimensions $i\not{\partial} : \mathcal{P} \to \mathcal{P}'$, $\mathcal{P}' \to \mathcal{P}$, and sections of $\mathcal{P}$ must get paired with sections of $\mathcal{P}'$ in the action: the Dirac operator is not self-adjoint, and the parition function is a Pfaffian rather than a determinant. This is just another way of saying that in odd dimensions, the invariant pairing is constructed between pinors transforming in the two distinct representations of $\mathrm{Pin}(s,t)$.

Summarizing, a reflection about the x^μ axis acts as

$$\begin{aligned} &R_\mu : \psi \mapsto \Lambda_{R_\mu}\psi = \gamma_\mu\bar{\gamma}\psi, \qquad && \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_{R_\mu} = \bar{\psi}(\gamma_\mu\bar{\gamma})^{-1} \qquad && d \in 2\mathbb{Z}, \\ &R_\mu : \psi \mapsto \Lambda_{R_\mu}\psi = \gamma_\mu\psi, && \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_{R_\mu} = -\bar{\psi}\gamma_\mu^{-1} && d \in 2\mathbb{Z}+1. \end{aligned} \tag{277}$$

Note that if we take the copinor to be given by the Dirac adjoint $\bar{\psi} = \psi^\dagger\gamma^0$ and the reflection to be about a spatial axis, the $(-1)^d$ factor for the $\bar{\Lambda}_{R_j}$ transformation is picked up when moving the $\gamma_j\bar{\gamma}$ through the γ^0.

An important consequence of the minus sign in the copinor transformation in odd dimensions is that for $d \in 2\mathbb{Z}+1$, the Dirac mass is always odd under spacetime reflections:

$$d \in 2\mathbb{Z}+1 \implies R_\mu : \bar{\psi}\psi \mapsto -\bar{\psi}\psi. \tag{278}$$

The fact that there is no reflection-invariant mass in odd dimensions is what allows the "parity anomaly" (bad name) to be a thing (although this does not tell us that the Dirac mass will be odd under the full antiunitary time-reversal when d is odd).

Reflection P of all the spatial axes is easy: (we are calling it P even though when d is odd it has determinant 1; sorry) in the conventions defined in the previous section,

$$P : \psi \mapsto \Lambda_P\psi = \gamma_0\psi, \qquad \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_P = \bar{\psi}\gamma_0^{-1}. \tag{279}$$

There are no complications in the odd d case since if d is odd P is an element of $\mathrm{Spin}(s,t)$ and is represented identically by both pinor representations of $\mathrm{Pin}(s,t)$.

Now one thing to note here is that we can always work with different reflection / parity operators defined as $\widetilde{\Lambda}_{R_\mu} = \Lambda_{R_\mu}\bar{\gamma}$—this actually seems to be the slightly more popular choice among the few physics papers where these issues are discussed. If d is odd then $\bar{\gamma}$ is central and this modification obviously does nothing, provided we modify the transformation of the copinor in the same way. Thus, if d is odd, this doesn't give us anything new. If d is even though, we can then take the copinor to transform under the matrix $-(\widetilde{\Lambda}_{R_\mu})^{-1}$; this minus sign is naturally generated if the copinor is defined though $\psi^\dagger\gamma_0$. Thus this gives us an alternative way to represent reflections in even dimensions:

$$R_\mu : \psi \mapsto \widetilde{\Lambda}_{R_\mu}\psi = \gamma_\mu\psi, \qquad \bar{\psi} \mapsto -\bar{\psi}\gamma_\mu^{-1} \qquad d \in 2\mathbb{Z}. \tag{280}$$

One then checks that the minus sign in the copinor transformation ensures that e.g. $\bar{\psi}\gamma_\mu\psi$ transforms appropriately as a vector under reflections.

[20]In Witten's paper he does something different: the $\bar{\gamma}$ is not included in his reflection action, but he modifies his Dirac operator by a factor of $\bar{\gamma}$ in a compensating way.

The difference between this alternate reflection and the previous one is simply that in this representation, the reflection squares to something different:

$$(\widetilde{\Lambda}_{R_\mu})^2 = \eta_{\mu\mu} = -\Lambda_{R_\mu}^2. \tag{281}$$

In keeping with this change, we can also define an alternate parity operator $\widetilde{\Lambda}_P = \gamma_0\bar{\gamma}$, so that

$$(\widetilde{\Lambda}_P)^2 = -\eta_{00}. \tag{282}$$

Thus in the tilde representation, parity and reflection square to the same thing in real time.

Time reversal

Now we come to time reversal. We will define the time reversal operation in the conventional way, which complex conjugates scalars but which does not conjugate dynamical fields—that is, we will define time reversal to act as the thing which historically has been called T, and not CT (so that e.g. magnetic fields are odd under our T). There are some reasons to do it the other way (they are motivated by issues that come up when we depart from the context of free fermions; there's a separate diary entry on this), but we will stick with the traditional definition for now.

The full antiunitary time reversal operation is much messier than the spatial reflections: its antiunitary nature means that while it involves the reversal of the time coordinate (which is a transformation in $\mathrm{Pin}(s,t)$), it itself does not act on the (co)pinors through a linear representation of $\mathrm{Pin}(s,t)$. Furthermore because it is antiunitary, it cares about the inner product structure on the Hilbert space in question, and therefore details of how $\bar{\psi}$ is defined will affect the transformation law.

Anyway, first consider the element R_0 of $\mathrm{Pin}(s,t)$ which reverses time, but which does not act with complex conjugation. By the same argument we used for spatial reflections, this acts as

$$\begin{aligned} &R_0 : \psi \mapsto \Lambda_{R_0}\psi = \gamma_0\bar{\gamma}\psi, \quad \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_{R_0} = \bar{\psi}(\gamma_0\bar{\gamma})^{-1} \qquad && d \in 2\mathbb{Z} \\ &R_0 : \psi \mapsto \Lambda_{R_0}\psi = \gamma_0\psi, \quad \bar{\psi} \mapsto \bar{\psi}\bar{\Lambda}_{R_0} = -\bar{\psi}\gamma_0^{-1} \qquad && d \in 2\mathbb{Z}+1. \end{aligned} \tag{283}$$

Similarly to the above discussion of spatial refections, in even dimensions we could also define the transformation through $\widetilde{\Lambda}_{R_0} = \bar{\gamma}\Lambda_{R_0}$, provided that we also took the copinor to transform as $-[\widetilde{\Lambda}_{R_0}]^{-1}$.

Now suppose we let time reversal act on pinors by

$$T : \psi \mapsto \mathcal{K}\Lambda_{R_0}\mathcal{U}\psi, \qquad \bar{\psi} \mapsto \bar{\psi}\bar{\mathcal{U}}\bar{\Lambda}_{R_0}\mathcal{K}, \tag{284}$$

where $\mathcal{U},\bar{\mathcal{U}}$ are unitaries to be determined. $\mathcal{U}$ and $\bar{\mathcal{U}}$ are independent unitaries, in keeping with the fact that $\bar{\psi}$ and ψ are independent fields. That said, $\bar{\mathcal{U}}$ will differ from $\mathcal{U}^\dagger$ only by potential minus signs relating to the signature and choice of γ_μ matrices. In the tilde convention for reflections in even dimensions, the time reversal action would instead be

$$\widetilde{T} : \psi \mapsto \mathcal{K}\gamma_0\mathcal{U}\psi, \qquad \bar{\psi} \mapsto -\bar{\psi}\bar{\mathcal{U}}\gamma_0^{-1}\mathcal{K} \quad (d \in 2\mathbb{Z}). \tag{285}$$

Requiring that $\bar{\psi} i\partial\!\!\!/ \psi$ be T-invariant means that

$$-\partial\!\!\!/ = \bar{\mathcal{U}}\bar{\Lambda}_{R_0}(-\partial^0\gamma_0^* + \partial^j\gamma_j^*)\Lambda_{R_0}\mathcal{U} = \bar{\mathcal{U}}\partial\!\!\!/^*\mathcal{U}, \tag{286}$$

which tells us that

$$\bar{\mathcal{U}}\gamma_\mu^*\mathcal{U} = -\gamma_\mu, \quad \forall\, \mu. \tag{287}$$

If all the γ matrices were Hermitian with then we could take $\bar{\mathcal{U}} = \mathcal{U}^\dagger$ with $\mathcal{U}$ the charge conjugation matrix (to be discussed in a sec); unfortunately in $\mathbb{R}$ time this is never the case. The exact transformation properties of various fermion bilinears depends annoyingly on details like the signature and choice of γ matrices, and so we defer such calculations to the examples section. One thing to note however is that if the pinor representation is real then one possible choice of $\bar{\mathcal{U}}$,$\mathcal{U}$ is always $\bar{\mathcal{U}} = -\mathbf{1}$,$\mathcal{U} = +\mathbf{1}$.

Euclidean time

We now briefly comment on what happens in Euclidean time. In Euclidean time the action is *not* Hermitian, and so we will always avoid writing the copinor field as something dependent on $\psi^\dagger$, since this is needlessly confusing—the spin representation in the Euclidean case is already unitary so the γ_0 isn't needed to get the right transformation rule.

We write the Lagrangian as[21] [22]

$$\mathcal{L}_E = \chi_E^\dagger(\partial\!\!\!/ + m)\psi_E, \tag{291}$$

where $\chi_E^\dagger$ is an independent field (not related to ψ_E by Hermitian conjugation). In even

[21]Some comments on Euclidean time Lagrangians: consider first complex fermions, where a Lorentz invariant bilinear is formed (in $\mathbb{R}$ time) by adding on γ^0 to $\psi^\dagger$ in $\psi^\dagger\psi$ to get $\bar{\psi}\psi$. Now in $\mathbb{R}$ time the operator $\gamma^0 i D\!\!\!\!/\,_A$ is Hermitian, regardless of signature, and so the action is Hermitian as required (we treat $\psi^\dagger$ and ψ as the (independent) integration variables). Now when we go to Euclidean time, we might be tempted to just naively do (independent of signature)

$$\gamma^0 i D\!\!\!\!/\,_A \xrightarrow{t\mapsto it} \gamma^0\partial\!\!\!/, \tag{288}$$

and write the action as $\int d\tau\, d^{D-1}x\, \bar{\psi}\partial\!\!\!/\psi$. However, in Euclidean time, while $\partial\!\!\!/\psi$ transforms as a spinor under $\mathrm{Spin}(D)$, $\bar{\psi}\psi$ is not invariant under $\mathrm{Spin}(D)$, because now the γ^0 in $\bar{\psi}$ screws stuff up. So, we should really write the Euclidean action as

$$S_E = \int d\tau\, d^{D-1}x\, \chi^\dagger\partial\!\!\!/\psi, \tag{289}$$

where if ψ transforms in a representation R of $\mathrm{Spin}(D)$, then $\chi^\dagger$ transforms in the representation $\bar{R}$. This does not actually lead to a doubling of the number of fields, since while $\chi^\dagger$ and ψ are independent, so too were $\psi^\dagger$ and ψ in the real time picture (notice that there is no $\psi^\dagger$ or χ field appearing in S_E). One thing to note however is that in Euclidean time, the action is no longer Hermitian (this is okay; recall e.g. theta angles).

[22]Another choice would be to write

$$\mathcal{L}_E = \chi_E^\dagger\bar{\gamma}(\partial\!\!\!/ + m)\psi_E, \tag{290}$$

but then the reflections act as $\Lambda_{R_j} = \gamma_j$ (without the $\bar{\gamma}$), which is different from the $\mathbb{R}$ time case, so we avoid this.

dimensions, the $\mathrm{Pin}(d)$[23] action $\psi \mapsto \Lambda_g \psi$ acts on χ as $\chi^\dagger \mapsto \chi^\dagger \Lambda_g^\dagger$. With this one checks that $\mathcal{L}_E$ is invariant under $\mathrm{Pin}(d)$, since the γ_E^μ matrices are all Hermitian. If d is odd, we have two different $\mathrm{Pin}(d)$ representations, differing by $\gamma_\mu \mapsto -\gamma_\mu$. In this case, we take χ_E and ψ_E to transform in opposite $\mathrm{Pin}(d)$ representations, so that $\chi_E^\dagger \mapsto \chi_E^\dagger \Lambda_g^\dagger \det \Lambda_g$, allowing the Lagrangian to be reflection-invariant.

Action of charge conjugation

Charge conjugation is not a spacetime symmetry, and therefore is a little out of place here. Furthermore, generic QFTs do not even always come equipped with some invariant notion of a $\mathbb{Z}_2$ charge conjugation transformation (for example, it may be subsumed by another global symmetry group), meaning that C really should not be elevated to the same status as spacetime reflections, as it often is.

However, in the free fermion context we're discussing, a notion of charge conjugation always exists. Charge conjugation, in this context, is a unitary operator that relates ψ and $\bar{\psi}$. There are some circumstances where we will take it to related ψ and $\psi^\dagger$, but I think the relation to $\bar{\psi}$ is more natural from the point of view of representation theory. We take C to act as

$$C : \psi \mapsto \bar{\psi} C^\dagger, \qquad \bar{\psi} \mapsto -C\psi. \tag{292}$$

The reason for the minus sign is to ensure that the Dirac mass is always C-even:

$$\bar{\psi}\psi \mapsto -C_{\alpha\beta}\psi_\beta \bar{\psi}_\lambda [C^\dagger]_{\lambda\alpha} = +\bar{\psi} C^\dagger C \psi = \bar{\psi}\psi. \tag{293}$$

The above conventions mean that a general bilinear $\bar{\psi} M \psi$ then transforms nicely as

$$\bar{\psi} M \psi \mapsto \bar{\psi}(C^\dagger M^T C)\psi. \tag{294}$$

In order for the free term to be invariant, we need

$$i\partial\!\!\!/ = (C^T i\partial\!\!\!/ [C^\dagger]^T)^T = C^\dagger [i\partial\!\!\!/]^T C, \implies C^\dagger \gamma_\mu^T C = -\gamma_\mu, \tag{295}$$

since $\partial_\mu^T = -\partial_\mu$. Since $-\gamma_\mu^T$ and γ_μ obey the same algebra they must be unitarily equivalent, and hence such a matrix C always exists. Exact expressions for C depend on the choice of signature and dimension, though.

As we saw, the Diarc mass is always C invariant. When $d \in 2\mathbb{Z}$ we also have a chiral mass, which transforms as

$$\bar{\psi}\bar{\gamma}\psi \mapsto \bar{\psi}(C^\dagger \bar{\gamma}^T C)\psi. \tag{296}$$

Now

$$C^\dagger \bar{\gamma}^T C = \prod_{\mu=d,\dots,1} C^\dagger \gamma_\mu^T C = \prod_{\mu=d,\dots,1} \gamma_\mu = (-1)^{((d-1)^2+(d-1))/2}\bar{\gamma}. \tag{297}$$

[23] Here by $\mathrm{Pin}(d)$ we really mean $\mathrm{Pin}(d,0)$. We take reflections to act as $\Lambda_{R_j} = \gamma_j \bar{\gamma}$ and $\bar{\Lambda}_{R_j} = (\gamma_j\bar{\gamma})^{-1}$ if d is even and as $\Lambda_{R_j} = \gamma_j$, $\bar{\Lambda}_{R_j} = -\gamma_j$ if d is odd. Just as in the indefinite signature case, were the squares of the reflections were fixed but could be changed by taking $\mathrm{Pin}(s,t) \to \mathrm{Pin}(t,s)$, here we can get different squares for reflections by taking $\mathrm{Pin}(d,0) \to \mathrm{Pin}(0,d)$. This amounts to taking $\gamma_\mu \to i\gamma_\mu$, and so switches the signs in the squares of the reflection matrices, both for d even and d odd.

So, we see that for $d \in 4\mathbb{Z}$ charge conjugation respects chirality and the chiral mass is C-even, while for $d \in 4\mathbb{Z}+2$ charge conjugation exchanges chiral components and the chiral mass is C-odd.

The vector current, since it lacks the ∂_μ of the free term which gives a minus sign when transposed, is (as expected) odd under C:

$$C : \bar{\psi}\gamma^\mu\psi = j_V^\mu \mapsto -j_V^\mu. \tag{298}$$

On the other hand, from the above charge-conjugate of $\bar{\gamma}$, we calculate

$$C : \bar{\psi}\gamma^\mu\bar{\gamma}\psi = j_A^\mu \mapsto \begin{cases} +j_A^\mu & d \in 4\mathbb{Z} \\ -j_A^\mu & d \in 4\mathbb{Z}+2 \end{cases} \tag{299}$$

Majorana fermions

Majorana fermions χ are self-adjoint fermions; this means the co(s)pinor $\bar{\chi}$ is *not* independent from χ. To define a Lorentz-invariant kinetic term with Majorana spinors, we then need to find a matrix that intertwines the spinor representation of $\mathrm{Spin}(s,t)$ with its dual. That is, if Λ_g is a representation matrix of $\mathrm{Spin}(s,t)$ (so that $\Lambda_g^{-1}\gamma^\mu\Lambda_g = R^{\mu\nu}\gamma_\nu$ for $R \in SO(s,t)$), we need to find a matrix $\mathcal{C}$ such that

$$\Lambda_g^T\mathcal{C} = \mathcal{C}\Lambda_g^{-1}. \tag{300}$$

If such a matrix exists, then given a spinor χ we can define a cospinor $\bar{\chi}$ which is linearly dependent on χ, with the two related via $\bar{\chi} \propto \chi^T\mathcal{C}$. The properties of $\mathcal{C}$ then guarantee that $\chi^T\mathcal{C}\chi$ is invariant under $\mathrm{Spin}(s,t)$ transformations.

Recall that the representation of any $g \in \mathrm{Spin}(s,t)$ can be written as

$$\Lambda_g = \exp\left(\frac{1}{8}[\gamma_\mu,\gamma_\nu]\theta_g^{\mu\nu}\right) \implies \Lambda_g^T = \exp\left(-\frac{1}{8}[\gamma_\mu^T,\gamma_\nu^T]\theta_g^{\mu\nu}\right), \tag{301}$$

by the antisymmetry of $\theta_g^{\mu\nu}$. We see that $\mathcal{C}$ will do the job provided that

$$\gamma_\mu^T\mathcal{C} = -\mathcal{C}\gamma_\mu. \tag{302}$$

But this is exactly what charge conjugation does, since C is unitary and satisfies $C^\dagger\gamma_\mu^TC = -\gamma_\mu$. Thus we can take $\mathcal{C} = C$.

Now in order for the theory to be nontrivial, we need $\bar{\chi}i\partial\!\!\!/\chi \neq 0$, i.e. we need $C\partial\!\!\!/$ to be an antisymmetric matrix. Now its transpose is

$$(C\partial\!\!\!/)^T = -\gamma_\mu^T\partial^\mu C^T. \tag{303}$$

Transposing $\gamma_\mu^TC = -C\gamma_\mu$ tells us that $\gamma_\mu^TC^T = -C^T\gamma_\mu$, so

$$(C\partial\!\!\!/)^T = +C^T\partial\!\!\!/, \tag{304}$$

and so if the action is to be nontrivial we need a charge conjugation which is antisymmetric: $C^T = -C$ (the antisymmetry also implies the existence of a nontrivial mass term for the Majoranas). Fortunately, this property is satisfied if $C^2 = \mathbf{1}$ when acting on fermions, since

$$C^2 : \psi_\alpha \mapsto [C(\bar{\psi}_\gamma[C^\dagger]_{\gamma\lambda})]_\alpha = -\psi_\rho C_{\gamma\rho}[C^\dagger]_{\gamma\lambda} = -C^T C^\dagger \psi, \tag{305}$$

and so $-C^T C^\dagger = \mathbf{1} \implies C = -C^T$.

Finally, in $\mathbb{R}$ time (and in $\mathbb{R}$ time only!), we need the action to be Hermitian. So far we have only used $\bar{\chi} \propto \chi^T C$. Now fix the proportionality constant as

$$\bar{\chi} = \lambda_\chi \chi^T C. \tag{306}$$

Now since $C\partial\!\!\!/$ is antisymmetric, in order to have a Hermitian action, we need $\lambda_\chi i\partial\!\!\!/$ to be purely imaginary. Evidently this is only possible if the γ matrices are either all real (in which case we take $\lambda_\chi = 1$) or all imaginary (in which case we take $\lambda_\chi = i$).[24] Having imaginary γ matrices is okay from a representation-theory point of view since the matrices representing $\mathrm{Spin}(s,t)$ are still real. The existence of such all-real or all-imaginary γ matrices holds in real time in all the dimensions we're usually interested in (2, 3, and 4). Again, remember that we are imposing the reality condition on Majorana spinors in $\mathbb{R}$ time (when we Wick-rotate to $i\mathbb{R}$ time, there is no Hermiticity requirement, and no reality condition for Majoranas).

The above shows how to define Majorana spinors. What about Majorana pinors? Evidently we need to find an intertwiner between the pinor representation of $\mathrm{Pin}(s,t)$ and its dual. This is more restrictive, since it relies on having a choice of signature and dimension such that the pinor representation of $\mathrm{Pin}(s,t)$ is real.[25] Said differently, the fact that we have only one field appearing in the action, unlike the two fields appearing in the complex case, means that we have less freedom in choosing transformation properties to make the free kinetic term symmetric. As an example of a constraint coming from the Majorana condition, consider the reflection of all spatial indices. This is usually represented in $\mathrm{Pin}(s,t)$ by $\chi \mapsto \gamma_0\chi$. But we see that $\chi^T C\partial\!\!\!/\chi$ is only invariant under this action if $\gamma_0^2 = -\mathbf{1}$, since invariance implies

$$P : \chi^T C\partial\!\!\!/\chi \mapsto \chi^T \gamma_0^T C\gamma_0\partial\!\!\!/\chi \implies \gamma_0^T C\gamma_0 = C \implies C^\dagger\gamma_0^T C = \gamma_0^{-1} \implies -\gamma_0 = \gamma_0^{-1}. \tag{307}$$

So, whether or not we can define a Majorana pinor depends on the signature (or better, on the Pin structure) that we choose! For example, the constraint that $\gamma_0^2 = -\mathbf{1}$ means we can only have Majorana pinors with this definition if they are "Kramers doublets" under parity.

Comment on CPT

The discussion of CPT is often slightly more confusing than it needs to be, since the CPT theorem is actually saying something very simple. In fact, instead of the CPT theorem,

[24] Another way of saying why Majoranas work when all the γ_μ are imaginary is that the equation $(iD\!\!\!\!/ - m)\psi = 0$ implies $(iD\!\!\!\!/ - m)\psi^* = 0$ if all the γ matrices are imaginary.

[25] Actually, one might argue that a purely imaginary representation of the γ matrices would again work, since the overall factor of i could just be treated as a "coupling constant" in the action. I think I remember finding some path paper where this was referred to as a "pseudo-Majorana" representation, or similar.

it should really be called the CRT theorem, where R is a reflection about any one of the coordinate axes—this is the appropriate operation to replace P with in odd dimensions, and for conceptual clarity it's helpful to use it in even dimensions as well.

In fact, even the moniker CRT is a bit confusing, since not all systems come equipped with a natural notion of charge conjugation. The C is present because in the usual (historical) conventions, T does not $\dagger$ the fields that it acts on, and so the action of C is needed to provide this. However, in systems where the global symmetry group doesn't contain a semi-direct product with a $\mathbb{Z}_2$ factor (that acts as an outer automorphism on the group it is $\rtimes$'d with), there is no real invariant notion of charge conjugation, and it is more natural to just define T as an anitlinear operation that Hermitian conjugates fields—in this definition, the CPT theorem is really just the RT theorem, since this definition has a T that acts as CT, in the cases where C is defined.[26]

Why then should we have an RT theorem? In $\mathbb{R}$ time thinking, the reason is elaborated on in a separate diary entry. The answer in imaginary time is simpler: in Euclidean signature both R and T are ordinary reflections, and so RT is just a π rotation. In particular, RT is in the component of the Lorentz group which is connected to $\mathbf{1}$, and so RT must be a symmetry of any relativistic theory. Note that $(RT)^2$ is a 2π rotation, so that as an operator acting on states of spin $j \in \frac{1}{2}\mathbb{Z}$, we have

$$(RT)^2 = (-1)^{2j}. \tag{308}$$

Examples

We now look at examples in low dimensions. We first look at $i\mathbb{R}$ time, and then at $\mathbb{R}$ time, in both mostly-positive and mostly-negative signatures. We will find the character $(\mathbb{R}, \mathbb{H}, \mathbb{C})$ of the spinor and pinor representations for each signature choice (the representation theory of $\mathrm{Spin}(s,t)$ is the same as that of $\mathrm{Spin}(t,s)$, but this is not true for $\mathrm{Pin}(s,t)$). For the $\mathbb{R}$ time examples, we find the action of time reversal and spatial reflections. We will also consider mass terms: in odd dimensions the only Lorentz scalar is $\bar\psi\psi$ ($\bar\gamma \propto \mathbf{1}$), and so there is only one type of mass term. In even dimensions $\bar\gamma$ is nontrivial, and we also have the chiral mass $\bar\psi\bar\gamma\psi$.

Two dimensions

Euclidean time: In Euclidean signature, we can take our γ matrices to be X and Z. This means that $\mathrm{Spin}(2)$ consists of unit-norm linear combinations of $\mathbf{1}$ and $XZ = J$, so that $\mathrm{Spin}(2) \cong U(1)$. The splitting $S_+ \oplus S_-$ in this case corresponds to spinors that transform as $e^{i\theta}$ or $e^{-i\theta}$. Another basis we commonly use is to take $\gamma^0 = Y, \gamma^1 = X$. Then $\mathrm{Spin}(2)$ has generators $\mathbf{1}, iZ$, the diagonality of which make it clear that the representation is reducible, with the splitting $S_+ \oplus S_-$ being a splitting into left- and right-moving spinors. Each $S_\pm$ spinor representation is complex, and the full spinor representation is the sum $-1 \oplus 1$, which is real. Because of the complexity of $S_\pm$, we can't define chiral Majorana spinors (alias Majorana-Weyl spinors): with the X, Z choice of γ matrices the charge conjugation

[26] In these conventions, the T that appears in CPT sends e.g. $E \mapsto E, B \mapsto -B$ in EM, while the T that appears in RT sends $E \mapsto -E, B \mapsto B$.

matrix is $C = Y$, and the resulting term $\chi^T C \not{\partial} \chi$ mixes S_+ and S_- components; with the X, Y choice the generators of Spin(2) are not real, which is incompatible with the Majorana reality condition. However, while we can't define chiral Majorana spinors, we can define Majorana pinors, since the representation $-1 \oplus 1$, which becomes irreducible when we lift to Pin(2), is real.

Real time: In real time with signature $(-,+)$,we may take

$$\gamma^0 = J, \qquad \gamma^1 = X, \qquad \bar{\gamma} = Z. \tag{309}$$

Therefore Spin(1,1) is generated by diagonal elements $\mathbf{1}$ and Z, which makes the $S_+ \oplus S_-$ decomposition manifest. This time though, the representation is real, since both generators of Spin(1,1) square to $\mathbf{1}$. Therefore chiral majorana spinors may be defined: the charge conjugation matrix is still $C = Y$, but now the free term $\chi^T C \not{\partial} \chi$ does not mix the $S_\pm$ components. Of course, Majorana pinors also exist. The fact that chiral Majorana spinors can be defined only in real time is not something to be bothered about—the reality conditions of the Majoranas are fixed in real time, not in imaginary time, because the reality condition is essentially the statement about $\{\chi_i, \chi_j\}$, and the commutator is defined in real time within the Hamiltonian framework. Therefore, to determine whether a given theory admits Majoranas, one needs to do the analysis in $\mathbb{R}$ time, and not worry about what happens when one continues to $i\mathbb{R}$ time.

Now because the charge conjugation matrix is proportional to γ_0, we may write the action of C more transparently as $C : \psi \to \psi^\dagger$ in the $\mathbb{C}$ case, or $C : \chi \to \chi$ in the Majorana case. The point of writing it like this is to demonstrate that C does not act on the Loretnz indices of the theory. In the $\mathbb{C}$ fermion case ψ transforms with charge 1 under a $U(1)$ symmetry and hence is in a complex irrep of the full symmetry group; in this case C acts nontrivially. However in the Majorana case C really does act trivially, as required by the reality of the representation the Majoranas transform in.

Now for time reversal. The equation $\bar{\mathcal{U}} \gamma_\mu^* \mathcal{U} = -\gamma_\mu$ means that we can set $\bar{\mathcal{U}} = \mathcal{U} = Z$ or $\bar{\mathcal{U}} = -\mathbf{1}, \mathcal{U} = \mathbf{1}$. We will choose the former option; thus in the regular (un-tilded) convention, we have

$$T : \psi \mapsto J\psi, \qquad \bar{\psi} \mapsto \bar{\psi} J^{-1}. \tag{310}$$

This gives $T^2 = (-1)^F$. We may also use the tilded definition of T including $\widetilde{\Lambda}_{R_0}$; this gives

$$\widetilde{T} : \psi \mapsto X\psi, \qquad \bar{\psi} \mapsto -\bar{\psi} X, \tag{311}$$

and so in this case we get a different time-reversal algebra, viz. $\widetilde{T}^2 = \mathbf{1}$. Note that we can also swap T with $\widetilde{T}$ by taking the alternate choice of $\bar{\mathcal{U}} = -\mathbf{1}, \mathcal{U} = \mathbf{1}$. Likewise, for parity, we have either

$$P : \psi \mapsto J\psi, \qquad \bar{\psi} \mapsto \bar{\psi} J^{-1} \tag{312}$$

with $P^2 = (-1)^F$, or

$$\widetilde{P} : \psi \mapsto X\psi, \qquad \bar{\psi} \mapsto -\bar{\psi} X, \tag{313}$$

with $P^2 = \mathbf{1}$. One can also compute,[27] using $CT : \psi \mapsto i\bar{\psi}$, $C\widetilde{T} : \psi \mapsto -i\bar{\psi} Z$,

$$(CT)^2 = \mathbf{1}, \qquad (C\widetilde{T})^2 = (-1)^F, \tag{315}$$

[27]Useful identities for this are

$$C : U\psi \mapsto \bar{\psi} C^\dagger U^T, \qquad \bar{\psi} U \mapsto -\psi C^T U, \tag{314}$$

so that the algebra obeyed by T and $\widetilde{T}$ is swapped when one tacks on charge conjugation.

The two Hermitian mass terms we can add are the Dirac mass $im\bar{\psi}\psi$ or the chiral mass $m\bar{\psi}Z\psi$. Both mass terms are odd under T, but they are both even under $\widetilde{T}$:

$$\begin{aligned} T &: i\bar{\psi}\psi \mapsto -i\bar{\psi}\psi, \quad \bar{\psi}\bar{\gamma}\psi \mapsto -\bar{\psi}\bar{\gamma}\psi, \\ \widetilde{T} &: i\bar{\psi}\psi \mapsto i\bar{\psi}\psi, \quad \bar{\psi}\bar{\gamma}\psi \mapsto \bar{\psi}\bar{\gamma}\psi. \end{aligned} \tag{316}$$

For parity, we see that the Dirac mass is even under P but odd under $\widetilde{P}$. On the other hand, the chiral mass is odd under P, but even under $\widetilde{P}$. The Dirac mass is even under C as it should be, while the chiral mass is odd under C, consistent with the fact that C does not respect chirality when $d \in 4\mathbb{Z}+2$. Therefore

$$\begin{aligned} CT &: i\bar{\psi}\psi \mapsto -i\bar{\psi}\psi, \quad \bar{\psi}\bar{\gamma}\psi \mapsto \bar{\psi}\bar{\gamma}\psi \\ C\widetilde{T} &: i\bar{\psi}\psi \mapsto i\bar{\psi}\psi, \quad \bar{\psi}\bar{\gamma}\psi \mapsto -\bar{\psi}\bar{\gamma}\psi \end{aligned} \tag{317}$$

If we took the signature to be $(+,-)$, then we could use the same γ matrices, just with their indices exchanged, and so the full structure is the same, just with the unitary matrices associated with time reversal and parity switched: in both cases, we are looking at the representation theory of $\mathrm{Pin}(1,1)$.

Three dimensions

Euclidean time: In Euclidean signature, we can take our γ matrices to just be the Pauli matrices; this is just because $\mathrm{Spin}(3) = SU(2)$. Consider then the antilinear map $\mathcal{J} = J\mathcal{K}$: this anticommutes with every γ matrix, and so it commutes with all products of an even number of γ matrices—thus, the spinor representation is pseudoreal.

If we pass to $\mathrm{Pin}(3,0)$, $\mathcal{J}$ is no longer an invariant form, since it anticommutes with each individual γ matrix. Thus the pinor representations of $\mathrm{Pin}^+(3)$ are complex. We say representations because there are two, since we are in an odd dimension (recall that the representation matrices Λ_g of the two representations differ by a factor of $\det \Lambda_g$: $\Lambda'_g = (\det \Lambda_g)\,\Lambda_g$). From our general discussion, we know that these two representations should give an invariant pairing. Indeed they do: if ψ is a pinor transforming under Λ_g and ψ' is a pinor transforming under Λ', then the antisymmetry of the fermions means that

$$\mathrm{Pin}(3,0) \ni g : \psi'_\alpha J^{\alpha\beta}\psi_\beta \mapsto [\Lambda'_g]_{\beta\alpha}\psi'_\alpha J^{\beta\gamma}[\Lambda_g]_{\gamma\lambda}\psi_\lambda = (\det \Lambda_g) J_{\alpha\beta}\Lambda_g^{\alpha\lambda}\Lambda_g^{\beta\rho}\psi'_\lambda\psi_\rho = (\det g)^2 J^{\alpha\beta}\psi'_\alpha\psi_\beta, \tag{318}$$

which is invariant since $\det(\Lambda_g)^2 = 1$. Now since $\mathrm{Pin}(3,0)$ is compact, this invariant pairing gives us an isomorphism between the Λ' representation and the complex conjugate of the Λ representation. To find this isomorphism, we look for a unitary U such that

$$U^\dagger \Lambda_g^* U = (\det \Lambda_g)\Lambda_g = \Lambda'_g. \tag{319}$$

Indeed, such an isomorphism is provided by taking $U = Y$.

for any unitary U.

Real time: In $\mathbb{R}$ time with signature $(-,+,+)$, we will take the γ matrices to be

$$\gamma_0 = J, \qquad \gamma_1 = X, \qquad \gamma_2 = Z. \tag{320}$$

They are all real, and so evidently the spinor representation of $\mathrm{Spin}(2,1)$ and the pinor representations of $\mathrm{Pin}(2,1)$ are all real—both spinor and pinor Majoranas can be defined. Since d is odd, $\mathrm{Pin}(2,1)$ has two irreducible representations. Just as in the $\mathrm{Pin}(3,0)$ case, a fermion ψ transforming in one and a fermion ψ' transforming in the other have an invariant antisymmetric pairing provided by J. However, since $\mathrm{Pin}(2,1)$ is *not compact*, this invariant pairing doesn't imply that the Λ_g representation is related to the complex conjugate of the Λ'_g representation through a unitary transformation. Indeed, now that J is included among the γ matrices there is no unitary U such that $U^\dagger \gamma_\mu^* U = -\gamma_\mu$, so we can not relate the two representations by (319). The non-existence of such a unitary means that we will not be able to choose $\bar{\mathcal{U}} = \mathcal{U}^\dagger$ when figuring out the time-reversal transformation.

Anyway, now for C, R, T. The charge conjugation matrix is $C = Y$, as in two dimensions. Also as in two dimensions we may write $C : \psi \to \psi^\dagger, \chi \to \chi$, which doesn't involve an action on the Lorentz indices; again this is made possible by the reality of the pinor representation of $\mathrm{Pin}(2,1)$.

For T, we need $\bar{\mathcal{U}}\gamma_\mu\mathcal{U} = -\gamma_\mu$. The only solution to this is to take $\bar{\mathcal{U}} = -\mathbf{1}, \mathcal{U} = \mathbf{1}$. Therefore up to a sign, the action of T is fixed to be (the minus signs from the definition of the barred copinor representation and the the $\bar{\mathcal{U}} = -\mathbf{1}$ cancel)

$$T : \psi \mapsto J\psi, \qquad \bar{\psi} \mapsto \bar{\psi} J^{-1}, \tag{321}$$

with $T^2 = (-1)^F$. Note that unlike in two dimensions, we *cannot* choose something that gives an algebra with $T^2 = \mathbf{1}$. From the choice of signature, we see that spatial reflections satisfy $R^2 = +\mathbf{1}$. CT satisfies $CT : \psi \mapsto i\bar{\psi}$; one then finds $(CT)^2 = +\mathbf{1}$. The Dirac mass is $i\bar{\psi}\psi$, which we see is odd under T, CT, and spatial reflections.

In $\mathbb{R}$ time with $(+,-,-)$ signature, we may choose the γ matrices to be purely imaginary:

$$\gamma_0 = Y, \quad \gamma_1 = iX, \qquad \gamma_2 = iZ. \tag{322}$$

This means that Majorana spinors can be defined, but strictly speaking, that Majorana pinors cannot be (although see previous comment on this). Charge conjugation is unchanged. For time reversal, we now need $\bar{\mathcal{U}}\gamma_\mu\mathcal{U} = \gamma_\mu$, which works only if $\bar{\mathcal{U}} = \mathcal{U} = \mathbf{1}$ (up to a sign). Therefore

$$T : \psi \mapsto Y\psi, \qquad \bar{\psi} \mapsto -\bar{\psi} Y. \tag{323}$$

As with mostly negative signature, we have $T^2 = (-1)^F$. Reflections now satisfy $R^2 = (-1)^F$. We now have $CT : \psi \mapsto -\bar{\psi}$, and we find $(CT)^2 = +\mathbf{1}$, which is different from the mostly positive signature. The Dirac mass is $\bar{\psi}\psi$ (no i!), which is still odd under T, CT, and spatial reflections. Therefore in three dimensions, no matter what signature we choose, the Dirac mass is always odd under T and under reflections; this enables the "parity anomaly" (bad terminology) to occur.

Four dimensions

Euclidean time: In Euclidean signature, we can take

$$\gamma^0 = X \otimes \mathbf{1}, \quad \gamma^1 = Y \otimes \mathbf{1}, \quad \gamma^2 = Z \otimes X, \quad \gamma^3 = Z \otimes Z. \tag{324}$$

Consider the antilinear map

$$\mathcal{J} = (X \otimes J)\mathcal{K}. \tag{325}$$

We see that $\mathcal{J}^2 = -\mathbf{1}$ and $\mathcal{J}$ commutes with all products of an even number of γ matrices, and so the spinor representation of Spin(4) is quaternionic, just like for Spin(3) (this is to be expected since $\text{Spin}(4) = SU(2)^2$). In fact, $\mathcal{J}$ actually commutes with every individual γ matrix, implying that the pinor representation of $\text{Pin}^+(4)$ is also pseudoreal.

Real time: In $\mathbb{R}$ time with $(-,+,+,+)$ signature, we may take

$$\gamma^0 = iX \otimes \mathbf{1}, \quad \gamma^1 = Y \otimes X, \quad \gamma^2 = Y \otimes Y, \quad \gamma^3 = Y \otimes Z, \qquad \bar{\gamma} = Z \otimes \mathbf{1}. \tag{326}$$

Note that the product of any two of these is block diagonal—this shows us explicitly how $\text{Spin}(3,1)$ acts reducibly on $S_+ \oplus S_-$. Note that the antilinear map $\mathcal{J} = \mathcal{K}(J \otimes J)$ commutes with all the γ matrices and satisfies $\mathcal{J}^2 = +\mathbf{1}$; therefore the pinor representation must actually be real, in spite of the fact that the above matrices contain i. The full $S_+ \oplus S_-$ spinor representation is of course real since $S_+ \cong S_-^*$, but by looking at a few products of two γ matrices, we can check that the chiral spinor reps acting on $S_\pm$ are in fact complex.[28] Therefore we may have Majorana pinors, but not chiral Majorana spinors.

A more convenient choice for doing calculations is to take

$$\gamma^0 = J \otimes X, \quad \gamma^1 = J \otimes J, \quad \gamma^2 = Z \otimes \mathbf{1}, \quad \gamma^3 = X \otimes \mathbf{1}, \qquad \bar{\gamma} = iJ \otimes Z. \tag{327}$$

Note that all of the γ^μ are real—the reality of the pinor representation is then manifest, but the complexity of $S_\pm$ is harder to see, since $\bar{\gamma}$ is no longer diagonal.

Charge conjugation, in the way presented above, can be chosen to act with $C = i\gamma^0$; the i is so that $C^2 = \mathbf{1}$. This can be written more simply (for either the Dirac adjoint or in the Majorana case) as $C : \psi \mapsto \psi^\dagger$, so that C doesn't act on the Lorentz index of ψ (as before, this is possible due to the reality of the pinor representation).

For time reversal, we need $\bar{\mathcal{U}}\gamma_\mu\mathcal{U} = -\gamma_\mu$; hence as in two dimensions we may choose either $\bar{\mathcal{U}} = \mathcal{U} = \bar{\gamma}$, or $\bar{\mathcal{U}} = -\mathbf{1}, \mathcal{U} = \mathbf{1}$. Opting for the former choice, we have

$$T : \psi \mapsto (J \otimes X)\psi, \quad \bar{\psi} \mapsto \bar{\psi}(J \otimes X)^{-1}, \tag{328}$$

with $T^2 = (-1)^F$. Selecting the other choice for $\mathcal{U}$, for which $\psi \mapsto \gamma_0\bar{\gamma}\psi$ and $\bar{\psi} \mapsto -\bar{\psi}\bar{\gamma}\gamma_0^{-1}$, also gives $T^2 = (-1)^F$, by the the imaginarity of $\bar{\gamma}$. If we choose the tilde transformation rule then we would still have the same two choices—as in two dimensions, changing our choices for $\mathcal{U}$,$\bar{\mathcal{U}}$ is equivalent to changing our choice of the regular or tilded representation. Therefore there are really only two choices for the action of T; the other one we define to be $\widetilde{T}$:

$$\widetilde{T} : \psi \mapsto -(\mathbf{1} \otimes Y)\psi, \qquad \bar{\psi} \mapsto \bar{\psi}(\mathbf{1} \otimes Y), \tag{329}$$

[28] The reality of the pinor representation doesn't imply the reality of the $S_\pm$, since $S_+^* \cong S_-$, which means $S_\pm$ can be complex while keeping $S_+ \oplus S_-$ real, as we have seen.

which as we mentioned has $\widetilde{T}^2 = (-1)^F$.

Because of our signature choice, spatial reflections square to $\gamma_j\bar{\gamma}\gamma_j\bar{\gamma} = (-1)^F$ in the regular representation, or $\gamma_j^2 = +\mathbf{1}$ in the tilde representation. Similarly, we can define parity to act as either γ^0 or $\gamma^0\bar{\gamma}$; in the first case $P^2 = (-1)^F$ while in the second case $P^2 = \mathbf{1}$. Thus we see that time reversal and spatial reflections are fundamentally different for this dimension and signature: the former always squares to $(-1)^F$, while the latter can square to either $(-1)^F$ or $+\mathbf{1}$. When we add in charge conjugation, we find

$$CT : \psi \mapsto i\bar{\psi}, \qquad C\widetilde{T} : \psi \mapsto -\bar{\psi}(J \otimes Z), \tag{330}$$

which tells us that

$$(CT)^2 = \mathbf{1}, \qquad (C\widetilde{T})^2 = (-1)^F. \tag{331}$$

The two mass terms are the Dirac mass $im\bar{\psi}\psi$ and the chiral mass $m\bar{\psi}\bar{\gamma}\psi$. Under the T transformation,

$$\begin{aligned} T &: i\bar{\psi}\psi \mapsto -i\bar{\psi}(J \otimes X)^{-1}(J \otimes X)\psi = -i\bar{\psi}\psi, \\ T &: \bar{\psi}\bar{\gamma}\psi \mapsto \bar{\psi}(J \otimes X)^{-1}(-\bar{\gamma})(J \otimes X)\psi = \bar{\psi}\bar{\gamma}\psi. \end{aligned} \tag{332}$$

Therefore under T, the Dirac mass is odd, while the chiral mass is even. On the other hand, $\widetilde{T}$ satisfies

$$\begin{aligned} \widetilde{T} &: i\bar{\psi}\psi \mapsto -i\bar{\psi}(-\mathbf{1} \otimes Y)(\mathbf{1} \otimes Y)\psi = +i\bar{\psi}\psi, \\ \widetilde{T} &: \bar{\psi}\bar{\gamma}\psi \mapsto -\bar{\psi}(\mathbf{1} \otimes Y)\bar{\gamma}^*(\mathbf{1} \otimes Y)\psi = -\bar{\psi}\bar{\gamma}\psi, \end{aligned} \tag{333}$$

so that the Dirac mass is T-even while the chiral mass is T-odd. Therefore even though the two time reversals square to $(-1)^F$, they are distinguished by their actions on fermion bilinears. Since C preserves chirality when $d \in 4\mathbb{Z}$, the transformation properties of the masses under CT and $C\widetilde{T}$ are the same as under T and $\widetilde{T}$, respectively.

In $\mathbb{R}$ time with $(+,-,-,-)$ signature, we can choose e.g.

$$\gamma^0 = X \otimes \mathbf{1}, \quad \gamma^1 = J \otimes X, \quad \gamma^2 = J \otimes Y, \quad \gamma^3 = J \otimes Z, \qquad \bar{\gamma} = Z \otimes \mathbf{1}, \tag{334}$$

which makes the $S_+ \oplus S_-$ splitting manifest. We see that $\mathcal{J} = (X \otimes J)\mathcal{K}$ commutes with each γ matrix and squares to $-\mathbf{1}$; therefore the pinor representation is pseudoreal. Since the representation theory of $\mathrm{Spin}(s,t)$ is the same as that of $\mathrm{Spin}(t,s)$, we know from the previous bit that the chiral spinor representations $S_\pm$ are still complex. Therefore in this signature we can't define Majorana pinors or chiral Majorana spinors.

We will instead prefer the representation

$$\gamma^0 = X \otimes Y, \quad \gamma^1 = i\mathbf{1} \otimes Z, \quad \gamma^2 = J \otimes Y, \quad \gamma^3 = i\mathbf{1} \otimes X, \qquad \bar{\gamma} = Z \otimes Y. \tag{335}$$

This gives us a purely imaginary representation, with γ^0 antisymmetric and the others symmetric. The antilinear map establishing pseudoreality of the pinor representation is now $\mathcal{J} = (Z \otimes J)\mathcal{K}$. All products of two γ matrices are real, and so again, we can have non-chiral Majorana spinors, but not chiral ones, nor Majorana pinors.

The charge conjugation matrix is again $C = X \otimes Y$. For time reversal, we need $\bar{\mathcal{U}}\gamma_\mu\mathcal{U} = \gamma_\mu$, and so we can either set $\bar{\mathcal{U}} = \mathcal{U} = \mathbf{1}$ or $\bar{\mathcal{U}} = -\bar{\gamma}, \mathcal{U} = \bar{\gamma}$. As in the other signature, the choice between these options is equivalent to a choice between the regular and tilded reflection representations. If we choose the latter for defining T, then we have

$$T : \psi \mapsto (X \otimes Y)\psi, \quad \bar{\psi} \mapsto -\bar{\psi}(X \otimes Y). \tag{336}$$

On the other hand, if we choose $\bar{\mathcal{U}} = \mathcal{U} = \mathbf{1}$ for defining $\widetilde{T}$, then we have

$$\widetilde{T} : \psi \mapsto (J \otimes \mathbf{1})\psi, \quad \bar{\psi} \mapsto -\bar{\psi}(J \otimes \mathbf{1}). \tag{337}$$

Reflections again square to either $(-1)^F$ or as $\mathbf{1}$, depending on the choice of regular or tilde representation. It is straightforward to check that the algebra obeyed by $T, \widetilde{T}, CT$, and $C\widetilde{T}$, as well as the transformation properties of the Dirac and chiral masses under these symmetries, is the same as for the case of mostly positive signature, provided that we stick with the above definitions of T and $\widetilde{T}$.

Fermions, bundles, and Spin$_G$ structures

The goal today is just to write down the right words explaining what bundles are relevant when dealing with fermions coupled to a gauge field. This is because I (like most physicists, probably) knew intuitively what equations to write down for fermions and spinc structures and all of that, but I didn't really know in a precise sense what mathematical structures were being used.

◇ ◇ ◇

In today's diary entry, we will work in Euclidean signature. In Euclidean signature there are no reality conditions on fermions, and so our spinors will always live in complex vector spaces.

Preliminaries

First we review what kind of bundles regular fermions (not coupled to any particular gauge field) are associated with. Let X be spacetime, $n = \dim X$, and let TX denote the tangent bundle of X. The fibers of the tangent space are acted on by $GL(n;\mathbb{R})$, which we can (with the metric) reduce to an action of $SO(n)$. Let LX, the frame bundle, be the principal $SO(n)$ bundle (we are assuming X is oriented) associated[29] to the tangent bundle (i.e. the principal bundle obtained from TX's transition functions).

[29] The word "associated" used to describe a bundle can mean a few different things. First, suppose we are given a principal G bundle $\pi : P_G \to X$, and a vector space V which carries an action of G via some representation $R : G \to \mathrm{Aut}(X)$. Then the vector bundle E associated to P_G is denoted

$$E = P_G \times_R V, \tag{338}$$

and consists of pairs $(u, v) \in P_G \times V$ modulo the equivalence relation

$$(u, v) \sim (gu, R_g^{-1} v), \tag{339}$$

Spinors are constructed with the help of the "square root" of the frame bundle, namely a principal Spin(n) bundle that we will write as SLX. Forming SLX is done, if possible, by lifting the transition functions on the frame bundle from $SO(n)$-valued functions to Spin(n)-valued functions; this can be done provided the cocycle conditions for the transition functions in LX hold in a particular way (more on this in a sec).

Suppose that SLX is well-defined. Then we can then form the spinor bundle S, which is the associated vector bundle made from the principal bundle SLX, the vector space $\mathbb{C}^{2^{\lfloor n/2 \rfloor}}$, and the spinor action of Spin(n) on $\mathbb{C}^{2^{\lfloor n/2 \rfloor}}$.[30][31] To save on writing, we will define the vector space that spinor fields live in as

$$\Delta_n \equiv \mathbb{C}^{2^{\lfloor n/2 \rfloor}}. \tag{341}$$

In our notation,

$$S = SLX \times_{1/2} \Delta_n, \tag{342}$$

where 1/2 is the spinor representation. Fermion fields are sections of S.[32]

To address when the spin frame bundle SLX exists, we need to ask the following question: when is it possible to take the square root of a principal $SO(n)$-bundle? Roughly speaking, the square root of $G/\mathbb{Z}_2$ is G, since $\mathbf{1} \in G/\mathbb{Z}_2$ is the image of $\sqrt{\mathbf{1}} = \{\mathbf{1}, -\mathbf{1}\} \in G$. This means that we can form the bundle "$SLX = \sqrt{LX}$" when the structure group in LX can be lifted to Spin(n) without causing the cocycle condition to fail by $\pm\mathbf{1} \in$ Spin(n) at the

where gu is the action of g on the fiber. This is basically like a $\otimes$ of P_G and V. Here the inverse of the representation is chosen so that $(gu, v) \sim (u, R_g v)$. Anyway, we can use this to see that as we go from patch to patch, the vector space V gets acted on by $R(t_{\alpha\beta})$, where $t_{\alpha\beta}$ are the transition functions. Thus the transition functions for E are given by $R(t_{\alpha\beta})$.

Another notion of an associated bundle is the principal bundle associated to a vector bundle. If the fibers of the vector bundle $\pi : F \to X$ are $\mathbb{K}^n$ then we have transition functions valued in $G \subset GL(n; \mathbb{K})$ (the subgroup G depends on how much extra structure our manifold has, like orientability or a metric, etc), and these transition functions can be used to construct a principal G bundle over X, associated to the vector bundle F.

[30] Here $\lfloor n/2 \rfloor$ is the dimension of the spinor representation, which is what it is for the following reason. The spin group is generated by all elements in the Clifford algebra

$$\text{Spin}(n) \ni \gamma_0^{j_0} \cdots \gamma_n^{j_n}, \qquad \sum_i j_i \in 2\mathbb{Z}. \tag{340}$$

There are 2^{n-1} possible products, and so the dimension of the spin group is 2^{n-1}. This means that it must be represented by matrices of dimension $d \geq 2^{(n-1)/2}$. When $n \in 2\mathbb{Z}$ we cannot have an equality. However when $n \in 2\mathbb{Z}$, the spinor representation is reducible, since we can project onto ± 1 eigenstates of $\bar{\gamma}$. In this case, the reducibility of the representation into two $d/2 \times d/2$ blocks means that $2(d/2)^2 = 2^{n-1}$, so that when d is even, we have $d^2 = 2^n$, and so $d = n/2$. Combining the odd and even case with the floor function, we get the stated result.

[31] if the representation of Spin(n) is real, then we can take the vector space to be real instead; we will ignore this possibility for simplicity

[32] In even dimensions, when the spinor representation is reducible, we can also form the bundles

$$S_\pm = SLX \times_{(1/2)_\pm} \Delta_n^\pm, \qquad \Delta_n^\pm = \mathbb{C}^{2^{\lfloor n/2 \rfloor - 1}}, \tag{343}$$

where $(1/2)_\pm$ is the representation of Spin(n) on the positive / negative chirality reducible component. Sections of these bundles are Weyl fermions.

triple overlaps. The second SW class makes an appearance here through

$$(\delta g^{\mathrm{Spin}(n)})_{\alpha\beta\gamma} = (-1)^{w_2(LX)_{\alpha\beta\gamma}}\mathbf{1}, \tag{344}$$

where the $g^{\mathrm{Spin}(n)}_{\alpha\beta}$ are spin lifts of the transition functions in LX. Here the notation $w_2(LX)$ is a bit sloppy, since the SW classes are defined only for vector bundles. What we really mean is $w_2(V)$, where V is the vector bundle associated to LX via $V = LX \times_1 \mathbb{R}^n$, with 1 the fundamental (vector) representation. V is isomorphic to the tangent bundle TX,[33] and so $w_2(LX)$ is just another way of writing $w_2(TX)$. From the Cech 3-cochain one can construct a class in $H^2(X;\mathbb{Z}_2)$; more on this in the next section.

So roughly speaking, $SLX = \sqrt{LX}$. What about the associated vector bundles? We can use our knowledge of the representations of the spin group to conclude, sgain roughly speaking, that S is the "square root" of the associated bundle $V = LX \times_1 \mathbb{C}^n$ (with the 1 the vector representation; note to self: since our spinors are living in $\mathbb{C}$ I wrote $\mathbb{C}^n$, but should this be $\mathbb{R}^n$?), in the sense that

$$S \otimes S \supset V, \tag{346}$$

with the tensor product of associated bundles being performed by tensoring both the vector spaces and the representations. In the special case of $n = 2$ we have $\mathrm{Spin}(2) = U(1)$ and this is easy to understand, since we are just tensoring two line bundles together.[34]

The Chern class and w_2

The above discussion showed how $w_2(TX)$ appeared from a Cechian point of view. Later on we will need to consider $w_2(E)$ for other types of vector bundles E. In the case where E is a complex vector bundle (which will be relevant to us when discussing fermions coupled to gauge fields), $w_2(E)$ has a very simple relation with $c_1(E)$, which we now describe.

In the following, fix G to be some compact Lie group, containing $\mathbb{Z}_2$ in its center. Further let E be the complex vector bundle associated to a principal $G/\mathbb{Z}_2$ bundle over X, with a given representation of $G/\mathbb{Z}_2$, and with transition functions

$$g_{\alpha\beta} = e^{i2\pi\Lambda_{\alpha\beta}} \in GL(n;\mathbb{C}). \tag{347}$$

Since E is a legit $G/\mathbb{Z}_2$ associated bundle, we have

$$\delta\Lambda \in H^2(X;\mathbb{Z}), \tag{348}$$

where we have identified the Cech 3-cocycle $\delta\Lambda$ with a simplicial 2-cohomology class (more on this in a sec; $\delta\Lambda$ is not exact in $H^2(X;\mathbb{Z})$ since Λ isn't a $\mathbb{Z}$-valued cochain).

---

[33] The isomorphism is given by

$$V = LX \times_1 \mathbb{R}^n \ni (u, v) = ((p, g), v) \mapsto (p, [R_1(g)](v)) \in TX, \tag{345}$$

where $u \in LX$, p is a basepoint in X, $g \in SO(n)$, and $v \in \mathbb{R}^n$. Note that as required, both (u, v) and $(hu, [R_1^{-1}(h)]v)$ get mapped to the same point in TX.

[34] In this case the spinor bundle S can be thought of as containing an action in the "charge 1/2 representation" of $U(1)$ on $\mathbb{R}^2$ (or $\mathbb{C}^2$?).

Taking the square root $E^{1/2}$ of E by passing to an associated G bundle means taking the square root of all of E's transition functions. Suppose first that $\delta\Lambda \in 2H^2(X;\mathbb{Z})$. If this is true, then the transition functions of $E^{1/2}$ will fail the cocycle condition by

$$(\delta g^{E^{1/2}})_{\alpha\beta\gamma} \in \mathbf{1}e^{2\pi i\mathbb{Z}}, \tag{349}$$

which is still acceptable. If however the $\delta\Lambda$ is just a class in $H^2(X;\mathbb{Z})$, then the cocycle condition can fail in $E^{1/2}$ by $\pm\mathbf{1}$ on each patch; this is not acceptable, and the bundle $E^{1/2}$ does not exist.

This means that the square-root-ibility of E is determined by how the transition functions fail the cocycle condition. But we know that if $(\delta\Lambda)_{\alpha\beta\gamma} = n_{\alpha\beta\gamma}$, then the first Chern class of E is just the class in $H^2(X;\mathbb{Z})$ which counts the $n_{\alpha\beta\gamma}$. Now from our above discussion, we see that in this case, $w_2(E)_{\alpha\beta\gamma}$ is precisely the class which counts $n_{\alpha\beta\gamma}$ mod 2. Thus $w_2(E)$ is the mod-2 reduction of c_1,[35] and so the nontriviality of $w_2(E) = c_1(E) \mod 2$ means an obstruction to consistently defining a bundle with square root transition functions.

Relating this to trivializing the 2-skeleton

How does this cocycle-centric definition of w_2 relate to other definitions of it as a SW class? For example, we know that $w_2(TX) \neq 0$ means that there is an obstruction to a extending a trivialization of a $SO(n)$ principal bundle from the 1-skeleton of X to the 2-skeleton[36]: how is this connected to the failure of the square roots of the transition functions to be closed in Cech cohomology? Now one way (I'm not sure if this is a really general statement) to go between a skeleton and a Cechian patch-covering of a manifold is to associate patches to each node of the 0-skeleton, such that the patches are chosen to be the convex regions that extend slightly beyond the halfway point of each of the 1-cells emanating from the 0-cell they are centered on. Choosing the patches this way means that we can associate to each 1-cell a 2-fold overlap of patches; this is nice because we can thinking about moving between 0-cells on the 0-skeleton as moving between patches, applying a transition function when we move across each 1-cell. Anyway, this also means that each 2-cell is associated to a triple overlap of patches, and so on: in this construction, each k-cell in the k-skeleton is associated to a $(k+1)$-fold overlap of patches.

Most importantly for us, each 2-cell corresponds to a triple overlap of patches. Now, what prevents us from extending a trivialization of a principal $SO(n)$ bundle from the 1-skeleton into the 2-skeleton? Such an extension will not be possible at a given 2-cell if the framing winds by the nontrivial element in $\pi_1(SO(d))$ around the boundary of that 2-cell. But this exactly translates to the condition that the square root of the transition functions on the triple patch overlap at the center of that 2-cell fail the cocycle condition by $-\mathbf{1}$: the framing winds by an odd multiple of 2π around the 2-cell, which corresponds to $-\mathbf{1}$ in Spin(n). Thus the Cechian way of thinking about $w_2(TX)$ and the skeleton way of thinking about $w_2(TX)$ are the same: the nontriviality of either indicates that we won't be able to take the square root of our principal $SO(n)$ bundle. The same applies to other groups G and $G/\mathbb{Z}_2$, where the nontriviality of $\pi_1(G/\mathbb{Z}_2)$ prevents the trivialization of a G bundle to be extended into the 2-skeleton.

[35] this is only true for complex vector bundles!

[36] Here by k-skeleton, we mean the k-th dimensional part of a cell complex.

Fermions with a gauge field

Now let the fermions be coupled to a gauge field A, with gauge group G. In general, if we have a field coupled to a background field for $G \times H$, then we will have a principal $G \times H$ bundle, and the vector space used to construct the associated bundle of which the field is a section will transform in a representation of $G \times H$. What representation we choose is up to us, but it will always be expressible as a direct sum of $\otimes$'s between an irrep of G and one of H, since the irreps of $G \times H$ are constructed as the $\otimes$ of irreps of G with irreps of H.[37] An example where we have a principal $G \times H$ bundle but use a reducible representation for the action on the vector space is when we are considering fermions in e.g. four dimensions, and making use of the decomposition $\mathrm{Spin}(4) = SU(2) \times SU(2)$. Here we do not want to take the $\otimes$ of two spin 1/2 irreps; rather, we want to take the representation

$$(1/2)_L \otimes \mathbf{1}_R \oplus \mathbf{1}_L \otimes (1/2)_R, \tag{352}$$

which is reducible. In this case, the associated bundle we get is a direct sum of two associated $SU(2)$ bundles, rather than a tensor product:

$$S = (SLX \times_{(1/2)_+} \mathbb{C}^{2^2}) \oplus (SLX \times_{(1/2)_-} \mathbb{C}^{2^2}). \tag{353}$$

Anyway, back to fermions coupled to a gauge field for an internal symmetry. We start out with a principal $G \times SO(n)$ bundle. If $w_2(TX) = 0$ then we can lift the $SO(n)$ factor to $\mathrm{Spin}(n)$, and we get the spinor bundle which we will write as

$$S_G = P_{G\times \mathrm{Spin}(n)} \times_{R_G \otimes 1/2} \Delta_n^G, \qquad \Delta_n^G = \mathbb{C}^{(\dim R_G) 2^{\lfloor n/2 \rfloor - 1}}. \tag{354}$$

The spinors are sections of this bundle. We will also write this as

$$S_G = (P_G \times_{R_G} \Delta^G) \otimes (P_{\mathrm{Spin}(n)} \times_{1/2} \Delta_n), \qquad \Delta^G = \mathbb{C}^{\dim R_G}. \tag{355}$$

Spin_G structures

There is another option which we use to construct spinors, that can be employed even when $w_2(TX) \neq 0$ in cohomology. This option is available to us when $G = \widetilde{G}/\mathbb{Z}_2$ for some Lie

[37] That is, every rep of $G \times H$ can be written as

$$(G \times H \to \mathrm{Aut}(V)) \ni R = \bigoplus_i \rho_i^g \otimes \rho_i^h, \tag{350}$$

where ρ_i^g, ρ_i^h are irreps of G and H, respectively. Furthermore, each factor in the direct sum is irreducible (as an irrep of $G \times H$). To prove this, we use Peter-Weyl: irreps of a compact group are the same as L^2 functions on that group (which exist because of the assumed compactness condition). A basis for these L^2 functions are precisely the characters. Now $L^2(G \times H) = L^2(G) \otimes L^2(H)$, which since the bases for the L^2 functions are provided by the characters, is the same thing as saying

$$\chi_{\rho^g}(g)\chi_{\rho^h}(h) = \chi_{\rho^g \otimes \rho^h}(g \times h), \tag{351}$$

which is true because $\mathrm{Tr}[A \otimes B] = \mathrm{Tr}[A]\mathrm{Tr}[B]$. Since the L^2 functions of the product group is the $\otimes$ of the individual L^2's, using Peter-Weyl proves the claim.

group $\widetilde{G}$, such that Spin(n) and $\widetilde{G}$ share a common central $\mathbb{Z}_2$ factor in the way they act on the fermions: this is possible if they act on the fermions in a tensor product representation of the spinor representation and a representation of $\widetilde{G}$ which includes this $\mathbb{Z}_2$ factor. For example if $G = SO(3)$, we might take the fermions to transform in the $(1/2)_{\text{Spin}(n)} \otimes (1/2)_{SU(2)}$ representation, since the fundamental of $SU(2)$ has a $-\mathbf{1}$ factor. In this case, the group that couples to the fermions is really $(\widetilde{G} \times \text{Spin}(n))/\mathbb{Z}_2$, and the transition functions in the full bundle are blind to the quotiented $\mathbb{Z}_2$ factor.

Anyway, while we might not be able to construct a bundle $SLX \times_{1/2} \Delta_n$ because the cocycle condition fails as $\delta g^{\text{Spin}(n)} = -\mathbf{1}$ at some points, we still may be able to form the bundle

$$E = P_{(\widetilde{G} \times \text{Spin}(n))/\mathbb{Z}_2} \times_{R_{\widetilde{G}} \otimes 1/2} \Delta_n^{\widetilde{G}}. \tag{356}$$

Here the bundle does *not* split as a tensor product of vector bundles:

$$E \neq (P_{\widetilde{G}} \times_{R_{\widetilde{G}}} \Delta^{\widetilde{G}}) \otimes (SLX \times_{1/2} \Delta_n), \tag{357}$$

since the latter factor does not exist. If E is to be well-defined, since the latter factor in (357) is not well-defined, the principal bundle $P_{\widetilde{G}}$ must also not be well-defined, in a compensating way: the transition functions in G must fail to lift to $\widetilde{G}$-valued transition functions in a way that cancels the ill-defined-ness of the latter tensor factor.

Now the transition functions of E are given by the matrices

$$g^E_{\alpha\beta} = R_{\widetilde{G}}(g^{\widetilde{G}}_{\alpha\beta}) \otimes R_{1/2}(g^{\text{Spin}(n)}_{\alpha\beta}). \tag{358}$$

This means that if we choose our bundles such that cocycle conditions in each of the factors in (357) fail in the same way, the transition functions above will satisfy the cocycle condition. From what we saw earlier, the condition for the transition functions in the SLX and $P_{\widetilde{G}}$ factors to not be closed in Cech cohomology in the same way is given by

$$w_2(TX) = w_2(E_G), \tag{359}$$

where E_G is the vector bundle associated to the principal bundle P_G in the representation R_G.[38] Such an E is called (idk if this is standard?) a Spin$_G$ bundle. Remember that for this to

[38]Since E_G is a complex vector bundle, let's be pedantic and elaborate on exactly what we mean by this. Since we have been working with complex spinors, E_G is formed by

$$E_G = P_G \times_{R_G} \mathbb{C}^{\dim R_G}. \tag{360}$$

Then we really mean

$$w_2(E_G) = w_2([E_G]_\mathbb{R}) = w_2([P_G \times_{R_G} \mathbb{C}^{\dim R_G}]_\mathbb{R}), \tag{361}$$

where the subscript denotes the realification, which is accomplished by taking $\mathbb{C}^{\dim R_G} \to \mathbb{R}^{2\dim R_G}$ and realifying the G action by

$$R_G \to [R_G]_\mathbb{R} : G \to GL(2\dim(R_G); \mathbb{R}) \tag{362}$$

via the inclusion $GL(n;\mathbb{C}) \to GL(2n;\mathbb{R})$.

Now as we have seen, the SW classes of the realification of a complex vector bundle E are the mod-2 reduction of that vector bundle's Chern classes:

$$c_j(E) \xrightarrow{\text{reduction mod 2}} w_{2j}(E_\mathbb{R}), \tag{363}$$

work, we also need to impose the condition that the representation $R_G \otimes 1/2$ that the fermions transform under is such that the representation R_G includes the $\mathbb{Z}_2$ factor corresponding to $-\mathbf{1}$ in G (i.e., if $G = SU(2)$, we must choose a half-odd-integer spin representation).

The simplest case is when $G = U(1)$. Suppose that $w_2(TX) \neq 0$, so that SLX does not exist. Suppose also that the line bundle $L = P_{U(1)}$ is such that

$$w_2(L) = [c_2(L)]_2 = w_2(TX). \tag{365}$$

Then a fermion field of charge q will be a section of the spinc bundle

$$E = P_{[U(1)\times \text{Spin}(n)]/\mathbb{Z}_2} \times_{q/2\otimes 1/2} (U(1) \otimes \Delta_n) \text{ “ = ” } (\sqrt{L} \times_{q/2} U(1)) \otimes (SLX \times_{1/2} \Delta_n), \tag{366}$$

where strictly speaking the second way of writing things is schematic, since neither tensor factor makes sense. Here in order for the representation q of $U(1)$ to include the central $\mathbb{Z}_2$, we also need to take $q \in (2\mathbb{Z}+1)$, so that the element -1 in $U(1)$ is represented nontrivially in the transition functions.

When $G = U(1)$, the classification of such Spin_G structures works just in the same way as for Spin structures, but up a dimension: $\text{Spin}_{U(1)}$ structures, alias $\text{Spin}_{\mathbb{C}}$ structures, are obstructed by $w_3(TX)$, and they are in (non-canonical) bijection with elements in $H^2(X;\mathbb{Z})$, which correspond to different large instantons that can be inserted into the $U(1)$ factor of $\text{Spin}(n) \times U(1)$.

Digression on $\text{Spin}_{\mathbb{C}}$ and CS

The biggest use of $\text{Spin}_{\mathbb{C}}$ connections comes when defining $U(1)$ CS theories on non-spin 3-manifolds. Such connections always exist because

$$w_3(TX) = 0, \qquad \dim X = 3,\ X \text{ compact} \tag{367}$$

We can prove this just by recalling that $w_3(TX) = [e(TX)]_2$, where $e(TX)$ is the Euler class, which is true since $\dim X = 3$. But the Euler characteristic for all *compact* 3-manifolds (orientable or not) vanishes:[39]

$$[\chi(X)]_2 = \sum_{i=0}^{3} \dim(-1)^i H^i(X;\mathbb{Z}_2) = 0, \tag{368}$$

where the sum vanishes by Poincare duality.

To see why this is useful, let a be a $U(1)$ gauge field, and $k = 2l+1$, $l \in \mathbb{Z}$. Then while the CS term at level k is not well-defined unless X is spin, for arbitrary X we may write

$$\mathcal{L} = \frac{k}{4\pi} a \wedge da + \frac{1}{2\pi} a \wedge dB, \tag{369}$$

and so we can equivalently write

$$w_2(E_G) = c_1(E_G) \mod 2. \tag{364}$$

[39] $\chi(\mathbb{R}^3) = 1$ shows that we need to stipulate compactness.

where B is a $\mathrm{Spin}_{\mathbb{C}}$ connection, so that

$$\frac{1}{2}\int_M w_2(TX) = \int_M \frac{F_B}{2\pi} \mod 2, \tag{370}$$

for any closed $M \subset X$. To see that $\mathcal{L}$ makes sense, we consider the following integral over a closed 4-manifold X_4:

$$I = \int_{X_4} \left(\frac{2l+1}{4\pi} F_a \wedge F_a + \frac{1}{2\pi} F_a \wedge F_B \right). \tag{371}$$

Working modulo things in $2\pi\mathbb{Z}$ and letting $\bar{F} = F/2\pi$,

$$\begin{aligned} I &= 2\pi \int_{X_4} \left(\frac{1}{2}\bar{F}_a \wedge \bar{F}_a + \bar{F}_a \wedge \bar{F}_B \right) \mod 2\pi \\ &= 2\pi \int_{X_4} \left(\frac{1}{2}\bar{F}_a \wedge w_2 + \frac{1}{2}\bar{F}_a \wedge (2F_B) \right) \mod 2\pi \\ &= 2\pi \int_{X_4} \bar{F}_a \wedge w_2 = 0 \mod 2\pi. \end{aligned} \tag{372}$$

Thus $\mathcal{L}$ is indeed well-defined.

Stereographic projection and the Hopf fibration; the incredible 3-sphere

---

Today we will try to build intuition for S^3 in a few different ways.

First, we will go over stereographic projection, showing that it is a conformal map, and going into detail for the S^2 case. We will show how geodesics on the sphere are projected onto either lines through the origin or circles, and that either way, the projected geodesics intersect the equator of the sphere at two antipodal points.

We will then move on to the Hopf fibration $S^3 \to S^2$. For an S^3 parametrized by $X^2+Y^2+Z^2+T^2 = 1$ with projected coordinates x, z, y, we will describe how the generators $L_{XY} = X\partial_Y - Y\partial_X$ and $L_{TZ} = T\partial_Z - Z\partial_T$ act on the $\mathbb{R}^3$ of x, y, z. As a biproduct we will show that S^3 is parallelizable by explicitly constructing a Killing field on S^3 with no fixed points.

⋄ ⋄ ⋄

Stereographic projection stuff

In what follows we will use uppercase letters to denote coordinates on S^n, and lowercase letters to denote coordinates of a stereographic projection. We will let X^0 be the “vertical” coordinate in the stereographic projection (the one which controls radial distance in the projected coordinate system), so that the coordinates are related as[40]

$$x^i = \frac{X^i}{1+X_0}. \tag{373}$$

Letting $r^2 = x_i x^i$, the fact that $X_\mu X^\mu = 1$ tells us that, after a bit of algebra,

$$X_0 = \frac{1-r^2}{1+r^2} \implies X^i = \frac{2x^i}{1+r^2}. \tag{374}$$

We claim that in the projected coordinates x^i, the metric is

$$ds^2 = \frac{4}{(1+r^2)^2} dx^2, \tag{375}$$

which is conformally flat. To show this, note that

$$r^2 = \frac{X_i X^i}{(1+X_0)^2} \implies \frac{4}{(1+r^2)^2} = (1+X_0)^2. \tag{376}$$

Thus the above metric is

$$\begin{aligned} ds^2 &= (1+X_0)^2 \left(\frac{dX^i}{1+X^0} - \frac{X_i}{(1+X_0)^2} dX_0 \right)^2 \\ &= dX^i dX_i + \frac{1}{(1+X_0)^2} X^i X_i dX_0^2 - \frac{2}{1+X_0} X^i dX_i dX_0 \\ &= dX_i dX^i + dX_0 dX^0 \left(\frac{1-X_0^2}{(1+X_0)^2} + \frac{2X_0}{1+X_0} \right) \\ &= dX_\mu dX^\mu, \end{aligned} \tag{377}$$

which is indeed the metric on the sphere.

Now specialize to the case of S^2. Then if x, y are the stereographic coordinates in $\mathbb{R}^2$, we can define $z = x+iy$, so that $r^2 = |z|^2$, and $dz d\bar{z} = dx^2 + dy^2$, $i dz \wedge d\bar{z} = 2 dx \wedge dy$. With

[40] This is a projection from the south pole of the sphere, and hence sends the south pole (where $X_0 = -1$) to ∞. To switch to a projection from the north pole, just replace $X_0 \to -X_0$ in the following formulae.

these coordinates, one can show that S^2 is a complex manifold.[41] The metric is then

$$ds^2 = \frac{4}{(1+|z|^2)^2} dz d\bar{z}, \tag{380}$$

which is Hermitian, meaning that we can construct the symplectic form

$$\Omega = \frac{2i}{(1+|z|^2)^2} dz \wedge d\bar{z}. \tag{381}$$

Since $d\Omega = 0$, S^2 is Kahler. Since Ω is closed under both complex differentials, it can be locally written as $\Omega = \partial\bar{\partial}V(z,\bar{z})$. Indeed,

$$\Omega = 2i\partial\bar{\partial}\ln(1+|z|^2). \tag{382}$$

Using $dz \wedge d\bar{z} = -2idx \wedge dy$ we have $\Omega = \sin\theta d\theta \wedge d\phi$ back in spherical coordinates, and so with this normalization the symplectic form is exactly equal to the volume form.

Now we'll go back to a general sphere S^n. Geodesics on the sphere are circles, and as such can be written as $\xi^\mu(\tau) = N^\mu \cos\tau + M^\mu \sin\tau$ for N and M two orthogonal unit vectors. The equator is the S^{n-1} where $X_0 = 0$, and so the points where ξ^μ meets the equator will be at time τ such that $N^0\cos\tau + M^0\sin\tau = 0$. Now if τ solves this equation then so does $\tau+\pi$, so each geodesic intersects the equator at two points. Since $\xi^\mu(\tau+\pi) = -\xi^\mu(\tau)$, this means that the two points where the geodesic intersects the equator are antipodal. Note that this will be true even after stereographic projection: the projected geodesics will meet the unit sphere in the projected plane (the image of the equator) at antipodal points.

The stereographic projection of *any* circle on S^n is either a straight line (if the circle in question passes through the south pole of S^n, from where we are projecting), or a circle. This is kind of crazy, since intuitively one might guess that circles that are not either the equator or geodesics passing through the poles would get projected to ellipses or something.

For an example of how this works, consider the stereographic projection of S^2. Circles are formed by intersecting a plane in $\mathbb{R}^3$ with the S^2. The plane can be parametrized by $aX + bY + cZ + d = 0$. Going to stereographic coordinates, this reads

$$\frac{2}{1+r^2}(ax+by) + c\frac{1-r^2}{1+r^2} + d = 0 \implies r^2(d-c) = -2(ax+by) - c - d. \tag{383}$$

[41] A complex manifold has holomorphic transition functions. The assignment of coordinates above via stereographic projection from the south pole is not good enough to cover the whole sphere, since it is singular there. So, define the coordinate patch

$$\widetilde{x} = \frac{X^1}{1-X_0}, \quad \widetilde{y} = -\frac{X^2}{1-X_0}, \tag{378}$$

which covers the south pole, but not the north pole (the reason for the $-$ sign in the $\widetilde{y}$ definition will be come clear shortly). Away from the poles where both coordinate systems are defined, we see that

$$\widetilde{z} = \frac{X^1 - iX^2}{1-X_0} = 2\frac{x-iy}{1-(1-r^2)/(1+r^2)} = \frac{x-iy}{r^2} = \frac{\bar{z}}{|z|^2} = \frac{1}{z}. \tag{379}$$

So indeed, the tilde'd and un-tilde'd coordinate systems are related holomorphically and hence S^2 is (duh) a $\mathbb{C}$ manifold (if we had not used the minus sign in the definition of $\widetilde{y}$ we'd have gotten $\widetilde{z} = 1/z^*$, which would have made the transition function involve an orientation reversal).

If $d = c$ then the point $(0, 0, -1)$ is on the circle, and the above tells us that we get a straight line in the xy plane. Therefore circles containing the south pole become lines (if the circle also contains the north pole then we need $c = -d$ and so $c = d = 0$, and the projected line passes through the origin). Else, after dividing by $d - c$, we get an equation $x^2 + y^2 = \dots$, where $\dots$ is linear in x, y. Therefore we can complete the square and get $(x - x_0)^2 + (y - y_0)^2 = r_0$, where x_0, y_0, r_0 are messy constants; this gives us a circle in the xy plane.

Hopf stuff

Now we specialize to the case $S^3 \to S^2$, letting the stereographic projection occur from the point $(X, Y, Z, T) = (0, 0, 0, 1)$, and letting it project onto the $T = 0$ subspace, whose coordinates we write as x, y, z.

Consider the rotation generators L_{XY} and L_{TZ}, which rotate the XY and TZ planes, respectively. Now

$$L_{XY} = X\partial_Y - Y\partial_X = \frac{2x}{1+r^2}\frac{1}{1+T}\partial_y - \frac{2y}{1+r^2}\frac{1}{1+T}\partial_x = x\partial_y - y\partial_x. \tag{384}$$

Thus $L_{XY} = L_{xy}$, which just acts to rotate the xy plane. The z axis, which is a geodesic on S^3, is fixed under this action. The only other geodesic that is mapped to itself under L_{xy} is the circle $x^2 + y^2 = 1$ (the equator of the equator). The flow lines of this killing field are circles wrapping the z axis.

Now we can take a look at L_{TZ}. We write

$$L_{TZ} = T\partial_Z - Z\partial_T = \frac{1-r^2}{2}\partial_z - \frac{2z}{1+r^2}\frac{-X^i}{(1+T)^2}\partial_i = \frac{1-r^2}{2}\partial_z + zx^i\partial_i. \tag{385}$$

The first term causes vectors inside the unit sphere in $\mathbb{R}^3$ (the equator of the S^3) to flow up along the z axis, and those outside the equator to flow down along the z axis. The second term causes vectors to flow radially outward / inward with a strength proportional to their z coordinate. The submanifold left invariant under L_{TZ} is evidently the one where $r^2 = 1$ and $z = 0$, i.e. the circle $x^2 + y^2 = 1$ (the equator of the equator). The flow lines under L_{TZ} are circles that wrap the $x^2 + y^2 = 1$ circle.

Now consider the Killing field $\Xi = L_{XY} + L_{ZT}$. Evidently, there are no fixed points under Ξ! This is why S^3 is parallelizable. The flow lines under Ξ are helix-shaped; twisted around both the z axis and the $x^2 + y^2 = 1$ circle. Each flow line defines a torus which it wraps around once, and the family of all such tori defined in this way fill the whole space (see Figure 2).

The cool thing about Ξ is that *all* of its flow lines are geodesics! Indeed, let us switch to complex notation where

$$z_1 = X + iY, \quad z_2 = Z + iT. \tag{386}$$

With this notation, we can project the S^3 onto the S^2 via the map $(z_1, z_2) \mapsto z_1/z_2$, which makes it clear that the fiber is a $U(1)$ coming from the common phase of z_1 and z_2 (the image of this map is S^2 realized as a copy of $\mathbb{R}^2$ compactified at infinity). With this notation, it is also clear that $S^2 = \mathbb{CP}^1$.

Now we can write any geodesic on the S^3 as

$$\xi^\mu(\tau) = A^\mu \cos\tau + B^\mu \sin\tau, \qquad A_\mu A^\mu = B_\mu B^\mu = 1, \quad A_\mu B^\mu = 0. \tag{387}$$

In complex notation, this is

$$\xi^\alpha(\tau) = A^\alpha \cos\tau + B^\alpha \sin\tau, \qquad A_\alpha \bar{A}^\alpha = B_\alpha \bar{B}^\alpha = 1, \quad A_\alpha \bar{B}^\alpha + B_\alpha \bar{A}^\alpha = 0, \tag{388}$$

where now $\alpha = 1, 2$.

Consider the family of geodesics

$$\xi^\alpha(\tau) = e^{i\tau} A^\alpha \tag{389}$$

for some choice of A^α. The tangent lines to this family are (using conventions where e.g. $\partial = \frac{1}{2}(\partial_x - i\partial_y)$)

$$\mathcal{T}_\xi = \partial_\tau \xi^\alpha \partial_\alpha + \partial_\tau \bar{\xi}^\alpha \bar{\partial}_\alpha = i(\xi^\alpha \partial_\alpha - \bar{\xi}^\alpha \bar{\partial}_\alpha). \tag{390}$$

Writing this out,

$$\mathcal{T}_\xi = \xi^X \partial_Y - \xi^Y \partial_X + \xi^Z \partial_T - \xi^T \partial_Z \tag{391}$$

Thus if we choose the unit vector $\xi^\mu = (X, Y, Z, T)$, we see that

$$\mathcal{T}_\xi = L_{XY} + L_{ZT} = \Xi. \tag{392}$$

Thus this family of geodesics are precisely the flow lines of the Killing field Ξ!

Because these geodesics are also the flows of a Killing field, they never intersect. Pedantically, consider two geodesics $\xi^\alpha(\tau) = e^{i\tau}A^\alpha$, $\eta^\alpha(\tau) = e^{i\tau + i\tau_0}B^\alpha$. Here we parametrize the geodesics so that we move around both of them at the same speed (they have the same τ dependence), but we allow for a constant offset τ_0 that allows us to start the parametrization at different offsets between the two geodesics. Anyway, the dot product

$$\bar{\xi} \cdot \eta + \xi \cdot \bar{\eta} = e^{i\tau_0} \bar{A}_\alpha B^\alpha + e^{-i\tau_0} A_\alpha \bar{B}^\alpha \tag{393}$$

is independent of τ, and so the geodesics are a constant distance apart. In particular, any two distinct such geodesics never intersect. This is kind of remarkable, if we think about trying to do the same on the 2-sphere: there all geodesics have to intersect, since the positive curvature of S^2 means that two initially parallel geodesics must intersect eventually. S^3 is also positively curved, but it turns out that we can get non-intersecting geodesics by twisting them in such a way that the twist exactly compensates the tendency for the positive curvature to "pull" them together. We will refer to the circular orbits under the above Killing field as Hopf circles.

Now we talk about why this is a fibration over S^2. First, we ask what the space of Hopf circles is. We can label them with two coordinates, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Here θ is a polar coordinate measured *up from the xy plane to the z axis*, with the value of θ defining the radii of the concentric tori that foliate the space. ϕ is an angle that measures the distance along each torus normal to the Hopf circles, and each pair θ, ϕ has an S^1 growing out of it (a Hopf circle), which wraps once around both cycles around the torus at the angle θ. Note that when $\theta = 0$ (the associated torus degenerates into the unit circle in the xy plane) or

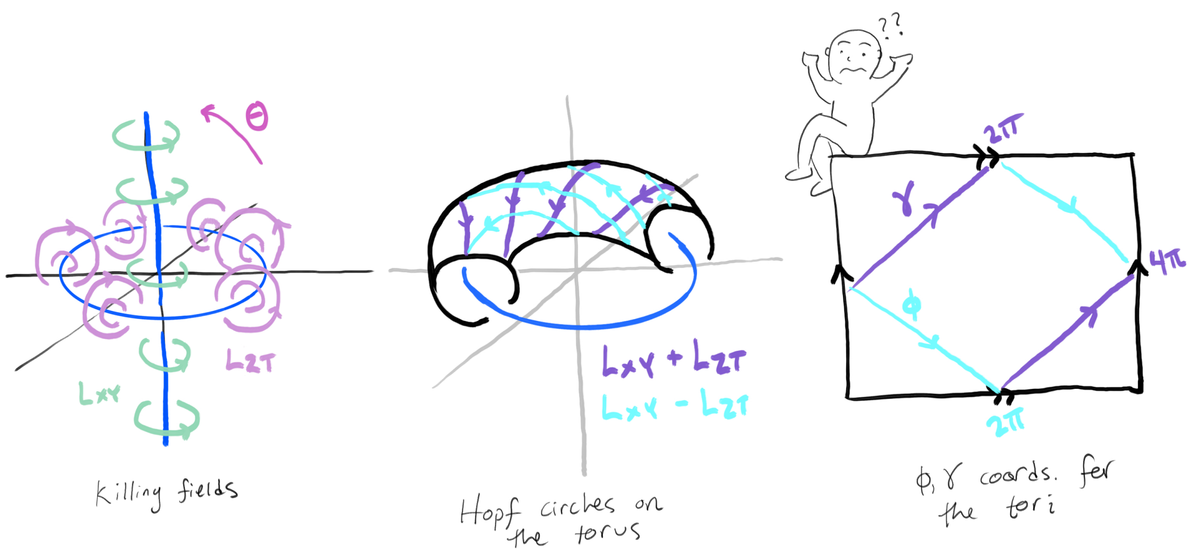


Figure 2: Left: the flow lines of the Killing fields L_{XY} and L_{ZT} in stereographic projection. Center: one torus at a fixed value of θ, with geodesic lines (flows of $L_{XY} \pm L_{ZT}$) drawn on top. Right: the same torus, split open. When we parametrize the torus with γ, ϕ, we only need ϕ to be 2π periodic.

when $\theta = \pi$ (the associated torus degenerates into a circle formed by the z axis and the point at ∞), all of the ϕ values become degenerate, and so the θ, ϕ coordinates give us an S^2: we have an S^2's worth of Hopf circles. We will let the coordinate along the circles be denoted by $\gamma \in [0, 4\pi)$. Thus we have an S^1 bundle over S^2, and the coordinate along the fibers is γ.

This bundle is nontrivial, as there is no global choice of coordinates such that $\gamma = \gamma(\phi, \theta)$ is a well-defined function. Indeed, suppose we tried to smoothly pick out a point on each fiber by marking the points where each Hopf circle pierces the xy plane. This works fine for every Hopf circle except for the one at $\theta = 0$, which lies inside the xy plane. Similarly we could try to mark the points where the Hopf circles pierce the zx plane, but this fails because the Hopf circle at $\theta = \pi$ (viz. the z axis) lies inside this plane. No matter what choice we make, we always miss one (or more) fibers.

Before moving on, we comment on a concise way of understanding the actions of the various Killing fields. First, we package the coordinates z_i as a matrix in the following way:

$$S^3 \ni \begin{pmatrix} z_1 & z_2 \\ -\bar{z}_2 & \bar{z}_1 \end{pmatrix}. \tag{394}$$

Consider the left action of $e^{i\theta\sigma^z}$. This sends $z_i \mapsto e^{i\theta} z_i$, and hence rotates the XY and ZT planes. That is, we can identify

$$i\sigma^z = L_{XY} + L_{ZT}. \tag{395}$$

So, the action of $i\sigma^z$ is precisely that of the flow of the Killing field we identified earlier, which flows in a helix wrapped around the $x^2+y^2 = 1$ circle. Of course, the other generators in $SU(2)$ perform similar flows: acting on the left with $i\sigma^x$ mixes up z_1 and iz_2, and so it mixes X with T and Y with Z. Likewise, $i\sigma^y$ mixes z_1 and z_2, and hence X with Z and Y with T. This means that (I'm not keeping very careful track of signs, sorry!)

$$i\sigma^x = L_{TX} + L_{YZ}, \qquad i\sigma^y = L_{XZ} + L_{YT}. \tag{396}$$

The right action of e.g. $e^{i\theta\sigma^z}$ is similar to the left action, except z_1 and z_2 get opposite phases, so that the XY and ZT planes are rotated in opposite directions. Thus to identify the Killing fields corresponding to the right action, we just change the relative signs between the two terms: $i\sigma^z$ acting on the right gives the field $L_{XY} - L_{ZT}$, and so on. The left and right actions commute, and together provide six linearly independent Killing fields. This identifications are also familiar from the fact that $SU(3) = \text{Spin}(3)$.

Of course, all of these differential operators can be nicely written in terms of quaternion multiplication. We see that multiplication by each quaternion defines a Killing field on S^3, which algebraically is clear because multiplication any of i, j, k is free of fixed points. The decomposition of each quaternion into a pair of differential operators which each rotate a given plane is also clear algebraically: for example, left multiplication by i rotates the $(1, i)$ plane into itself, as well as the (j, k) plane into itself. In accordance with the right action flipping the sign between the two differential operators, we also see algebraically that right multiplication by i rotates the $(1, i)$ plane with the same handedness, while due to the anticommutativity rotates the (j, k) plane with the opposite handedness. Therefore the adjoint action of a given quaternion q, which sends $p \to qpq^{-1}$, allows us to perform a rotation only in the $(1, q)$ plane, while keeping the orthogonal plane fixed. This is of course how one can realize 3d rotations as quaternion conjugation (and is yet another way of seeing why the adjoint representation of $SU(2)$ is the vector representation of $SO(3)$).

Now we return to looking at the structure of the $S^3 \to S^2$ bundle. For this it helps to make another coordinate redefinition, via

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} e^{i(\gamma+\phi)/2} \cos(\theta/2) \\ e^{i(\gamma-\phi)/2} \sin(\theta/2) \end{pmatrix} \tag{397}$$

Here $\theta \in [0, \pi]$, $\gamma \in [0, 4\pi)$, and $\phi \in [0, 2\pi)$. The γ coordinate parametrizes distance along the Hopf circles; i.e., distance along the fibers. As before, the θ coordinate measures the distance away from the $x^2 + y^2 = 1$ circle (the fixed point locus of the L_{ZT} action), and the subspace of constant θ picks out a torus foliated by Hopf circles. Finally, ϕ parametrizes distance along Hopf circles which wind in the *opposite* sense around each of the tori (see Figure 2). The γ, ϕ coordinates thus label points on the tori, while θ parametrizes a direction normal to the tori. As a sanity check on these identifications, consider $\theta = 0$, which is the unit circle in the xy plane. Then $z_2 = 0$, and $z_1 = e^{i\varphi}$, where φ is the polar angle in the xy plane. Now at $\theta = 0$ the two ways of winding around the torus are degenerate, and so $\gamma = \phi = \varphi$, meaning that $(z_1, z_2) = (e^{i\varphi}, 0) = (e^{i(\gamma+\phi)/2}, 0)$, which agrees with (397). A similar sanity check works for $\theta = \pi$: here we are in the ZT plane so that $z_1 = 0$ and $z_2 = e^{i\varphi'}$, with φ' the angle in the ZT plane. We also have $\phi = -\gamma$ (the torus at $\theta \to \pi$ degenerates in the opposite sense as the torus at $\theta \to 0$), and so $(z_1, z_2) = (0, e^{i(\gamma-\phi)/2})$, matching (397).

Why is γ 4π periodic, while ϕ is only defined in the range $[0, 2\pi)$? This is simply because in addition to the γ coordinate, only the first half of each opposite-winding Hopf circle is needed to parametrize the full torus. This is sketched in the right panel of Figure 2: the $\hat{\gamma}$ basis vector points diagonally up; the $\hat{\phi}$ basis vector points diagonally down. Working in units where the torus has sides of length $4\pi/\sqrt{2}$, every point on the torus can written as a linear combination $\alpha\hat{\gamma} + \beta\hat{\phi}$, where $\alpha \in [0, 4\pi), \beta \in [0, 2\pi)$. This is the reason for the different periodicity of the two variables.

In these coordinates, the projection $(z_1, z_2) \mapsto z_1/z_2$ is particularly transparent. Indeed, we have

$$(z_1, z_2) \mapsto e^{i\phi} \cot(\theta/2). \tag{398}$$

The γ coordinate no longer appears when after dividing out (since γ is the coordinate along the fibers), and we see that this mapping gives us a presentation of the S^2 where the north pole is located at ∞ and the south pole at the origin.

To get from these coordinates to a vector $n \in S^2$, we use

$$n^i = \vec{z}^\dagger \sigma^i \vec{z}. \tag{399}$$

Indeed, one checks that e.g. $n^z = \cos^2\theta/2 - \sin^2\theta/2 = \cos\theta$, and similarly for n^x and n^y.

The line element in complex coordinates is of course $ds^2 = dz_1 d\bar{z}_1 + dz_2 d\bar{z}_2$. In terms of the angular coordinates,

$$|dz_1|^2 = \frac{(d\gamma + d\phi)^2}{4} \cos^2\theta/2 + \frac{d\theta^2}{4} \sin^2\theta/2, \tag{400}$$

and

$$|dz_2|^2 = \frac{(d\gamma - d\phi)^2}{4}\sin^2\theta/2 + \frac{d\theta^2}{4}\cos^2\theta/2, \tag{401}$$

so that $(\cos^2\theta/2 - \sin^2\theta/2 = \cos\theta)$

$$ds^2 = \frac{d\gamma^2 + d\theta^2 + d\phi^2}{4} + \frac{1}{2}\cos\theta\, d\gamma\, d\phi. \tag{402}$$

The tori that foliate the S^3 are the fixed θ subspaces. We can confirm our earlier statement about γ being the coordinate along the Hopf circles and ϕ being the coordinate along the opposite-winding Hopf circles by looking at the metric restricted to a given torus at $\theta = \theta_0$. The metric on this torus $T^2_{\theta_0}$ is

$$ds^2|_{T^2_{\theta_0}} = \frac{1}{4}(d\gamma^2 + d\phi^2) + \frac{1}{2}\cos\theta_0\, d\gamma\, d\phi. \tag{403}$$

Now define the variables σ, ρ through $\gamma = (\sigma+\rho)/\sqrt{2}$, $\phi = (\sigma-\rho)/\sqrt{2}$. If our identification of γ, ϕ was correct, then σ should be (proportional to) the coordinate on the torus that travels along the flow of L_{XY}, while ρ should be (proportional to) the coordinate that travels along the flow of L_{ZT} (see Figure 2). Indeed, the metric in these coordinates is

$$ds^2|_{T^2_{\theta_0}} = \frac{1}{4}\left((1+\cos\theta_0)d\sigma^2 + (1-\cos\theta_0)d\rho^2\right). \tag{404}$$

We see that the σ coordinate disappears when $\theta_0 = \pi$, while the ρ coordinate disappears at $\theta_0 = 0$. Since the $\theta_0 = 0$ torus is just the unit circle in the xy plane, the σ coordinate evidently labels the angular direction in the xy plane, which is just as we said it would be based on our identification of γ, ϕ. Likewise, ρ labels the direction on the torus that runs along the flow lines of L_{ZT}, since the torus degenerates to a circle along one of these flow lines at $\theta_0 = \pi$.

We can determine the connection by requiring that it partition the fiber bundle into vertical and horizontal subspaces which are orthogonal with respect to the above metric, via a splitting $TM = TM_h \oplus TM_v$. We will write the connection as the 1-form $\omega = \omega_\mu dx^\mu$. Its goal in life is to give us a definition of the vertical subspace at each point in the bundle, and is chosen such that the contraction of ω with any vector field in the horizontal subspace vanishes: $\omega(X_\tau) = \omega_\mu \frac{dx^\mu}{d\tau} = 0$, for any horizontal vector field $X_\tau = \partial_\tau$. This equation yields a DE relating movement in the vertical subspace to movement in the horizontal subspace, which when solved gives us the holonomies, curvature, etc.

In our application, the vertical subspace is defined by the vector field $X_\gamma = \partial_\gamma$. Consider a curve $(\theta,\phi)(\tau)$ in the base S^2, parametrized by τ. We want to lift this curve into the full bundle in such a way that it is lifted into the horizontal subspace. Now the tangent vector to the lifted curve is of course

$$X_\tau = \dot\gamma\partial_\gamma + \dot\theta\partial_\theta + \dot\phi\partial_\phi, \tag{405}$$

with the overdot denoting differentiation with respect to τ. If this is to be lifted to the horizontal subspace, then we need

$$(X_\tau)_\mu (X_\gamma)^\mu = (X_\tau)_\mu (X_\gamma)_\nu g^{\mu\nu} = 0. \tag{406}$$

Using the form for the metric and $(X_\gamma)_\nu = (1,0,0)$ in the basis (γ, θ, ϕ), we have

$$g_{\gamma\gamma}\dot{\gamma} + g_{\gamma\theta}\dot{\theta} + g_{\gamma\phi}\dot{\phi} = \dot{\gamma} + \cos\theta\dot{\phi} = 0. \tag{407}$$

On the other hand, the above is equal to $\omega(X_\tau) = \omega_\mu dx^\mu/d\tau$. So evidently the connection is

$$\omega = d\gamma + \cos\theta d\phi. \tag{408}$$

Integrating the differential equation $\omega(X_\tau) = 0$ around a closed loop C in the base S^2 tells us that the holonomy is $\Delta\gamma = -\oint_C \cos\theta d\phi$.

The typical physicist usually uses "connection" to mean a 1-form on the base manifold. We can get a 1-form on the base S^2 by choosing a section, i.e. choosing a paremetrization $\gamma(\theta, \phi)$. Of course, since the bundle is nontrivial, we won't be able to choose a globally well-defined such function $\gamma(\theta, \phi)$. Indeed, one choice we might make is $\gamma = \phi$, in which case the 1-form A on S^2 would be

$$A_S = (1 + \cos\theta)d\phi. \tag{409}$$

As the subscript indicates, this is well-defined everywhere except the north pole of the S^2, i.e. the $x^2 + y^2 = 1$ circle in the stereographic $\mathbb{R}^3$. Likewise we might choose $\gamma = -\phi$, which would give $A_N = (-1 + \cos\theta)d\phi$, which works as a section everywhere except the south pole.

Note that regardless of how we project ω onto the S^2, the curvature is

$$dA = \sin\theta \, d\phi \wedge d\theta, \tag{410}$$

which is proportional to the volume form on the S^2. This is why the Hopf fibration is the $U(1)$ bundle on an S^2 enclosing a monopole of unit strength.

Real and quaterionic Hopf fibrations

What we have talked about above is the complex Hopf fibration, since the base space was $S^2 = \mathbb{CP}^1$, and we had the sequence $S^1 \to S^3 \to S^2$, or

$$U(1) \to S^3 \to \mathbb{CP}^1, \tag{411}$$

which is the unit complex numbers, the unit sphere in $\mathbb{C}^2$, and the first complex projective space.

We can also get Hopf fibrations for $\mathbb{RP}^1$ and $\mathbb{HP}^1$. First for the real Hopf fibration, which is rather trivial. The relevant sequence here is $S^0 \to S^1 \to S^1$, or

$$U(1;\mathbb{R}) \to S^1 \to \mathbb{RP}^1. \tag{412}$$

In complete analogy with the complex Hopf fibration, we have the unit real numbers (of course $U(1;\mathbb{R})$, meaning the $\mathbb{R}$ numbers of unit norm, is just a suggestive way of writing $\mathbb{Z}_2$), the unit sphere in $\mathbb{R}^2$, and the first real projective space. Now $\mathbb{RP}^1 = S^1$ (do antipodal identification on S^1 to produce an S^1 of half the circumference—nothing unorientable about this, unlike the case of $\mathbb{RP}^n$ for even n), so we can just as well write $\mathbb{Z}_2 \to \text{Spin}(2) \to S^1$, where Spin(2) is the double cover of S^1 (the boundary of a mobius band). Just like the

$\mathbb{C}$ Hopf fibration, the projection from $\mathrm{Spin}(2) \to S^1$ is realized (on the patches where it is defined) by $(x, y) \to x/y$, which clearly has kernel $S^0 = \mathbb{Z}_2$.

Now for the quaterionic case. By analogy with the previous two, we look for the fibration

$$U(1; \mathbb{H}) \to S^7 \to \mathbb{HP}^1, \tag{413}$$

since S^7 is the unit sphere in $\mathbb{H}^2$. Such a fibration does exist, and can be written $S^3 \to S^7 \to S^4$. Here $U(1; \mathbb{H}) = Sp(1) = SU(2) = S^3$, while $\mathbb{HP}^1 = S^4$. That $\mathbb{HP}^1 = S^7/S^3$ can be seen by realizing that the space of quaternionic lines containing the origin in $\mathbb{H}^2$ can be visualized in the same way as the set of four-dimensional hyperplanes (containing the origin) in $\mathbb{R}^8$. Each hyperplane can be defined by the way in which it intersects the unit sphere S^7 in $\mathbb{R}^8$. But each hyperplane intersects the S^7 in a four-dimensional subspace of that S^7, which cuts out an S^3 inside of the S^7, and so $\mathbb{HP}^1 = S^7/S^3$. Finally, that $\mathbb{HP}^1 = S^4$ can be seen by writing the projection from $\mathbb{H}^2$ with two patches as either as $(q, r) \to q/r$ for $r \neq 0$ or $(q, r) \to r/q$ for $q \neq 0$, with $q, r \in \mathbb{H}$. When $r \neq 0$ we can use the first patch to get a full $\mathbb{R}^4$'s worth of points. When $r = 0$ we use the other patch, and we see that the $r = 0$ subspace gets mapped onto a single point. Putting the two patches together, we see that the quotient space is $\mathbb{R}^4$, but with the points at infinity identified: so, the quotient is indeed S^4.

Details of the $\mathfrak{spin}(n) \cong \mathfrak{so}(n)$ isomorphism

Today we will work out the explicit mapping between the $\mathfrak{spin}(n)$ and $\mathfrak{so}(n)$ Lie algebras. Understanding exactly how this works is a prerequisite for figuring out the correct factors of 1/2 to write in the spin connection, which we will need for the next diary entry.

Let γ_j be Clifford generators, and fix the signature to be Euclidean so that $\{\gamma_i, \gamma_j\} = 2\delta_{ij}$. We can write a generic element in the odd part of the real Clifford algebra $Cl_n = Cl_n^+ \oplus Cl_n^-$ as $v^j\gamma_j$, where $v \in \mathbb{R}^n$. The group $\mathrm{Pin}(n)$ consists of all elements of the form $\prod_{a=1}^k v_a$, where each v_a is a unit vector: $v_{a,i}v_a^i = 1$, and where $k \in \mathbb{N}$ is any integer. The $\mathrm{Spin}(n)$ subgroup is the even subgroup of $\mathrm{Pin}(n)$, i.e. those elements with $k \in 2\mathbb{N}$.

As discussed in the diary entry on Clifford algebras, the map $P : \mathrm{Spin}(n) \to SO(n)$ is realized in the following way: for any $X = x^i\gamma_i \in Cl_n^-$ and any $Y \in \mathrm{Spin}(n)$, we have

$$P : x_i \mapsto R_{ij}x_j, \tag{414}$$

where R is the orthogonal matrix defined by the adjoint action of Y:

$$P : X \mapsto Y^{-1}XY = R_{ij}x^j\gamma^i. \tag{415}$$

That R is orthogonal can be seen by applying the transformation to both $X = \gamma_j, X' = \gamma_k$ and considering the transformation of $\{X, X'\}$.[42]

[42]We can find R_{ij} explicitly by doing

$$R_{ij} = \frac{1}{N}\mathrm{Tr}[Y^{-1}\gamma_i Y\gamma_j], \tag{416}$$

The map P induces a map $P_* : \mathfrak{spin}(n) \to \mathfrak{so}(n)$ via the differential. To find P_*, we first need to write elements in $\mathrm{Spin}(n)$ as points on parametrized curves passing through the identity at $\theta = 0$. To this end, consider the curves

$$\begin{aligned} \mathrm{Spin}(n) \ni Y_{ab}(\theta) &= (-\gamma_b \cos(\theta/2) + \gamma_a \sin(\theta/2))(\gamma_a \sin(\theta/2) + \gamma_b \cos(\theta/2)) \\ &= [\gamma_a, \gamma_b] \cos(\theta/2) \sin(\theta/2) + \mathbf{1}(\sin^2(\theta/2) - \cos^2(\theta/2)). \end{aligned} \tag{419}$$

Using $\cos(\theta/2)\sin(\theta/2) = \sin(\theta)/2$, this is

$$Y_{ab}(\theta) = \gamma_a \gamma_b \sin\theta + \mathbf{1} \cos\theta. \tag{420}$$

The important thing here is that $Y_{ab}(0) = \mathbf{1}$ and $Y'_{ab}(0) = \gamma_a\gamma_b$. Thus $\mathfrak{spin}(n)$ includes all the matrices of the form $\gamma_a\gamma_b$, with $a < b$ (taking products of $Y_{ab}(\theta)$'s for different choices of a, b allows us to pick up all such matrices). Furthermore, the $\gamma_a\gamma_b$s generate all of $\mathfrak{spin}(n)$, since there are $(n^2 - n)/2 = \dim \mathfrak{spin}(n)$ such matrices.

To find P_*, we just need to find how P acts on the $Y_{ab}(\theta)$, and then take the differential at $\theta = 0$. So for any $X = x^i \gamma_i$,

$$P : X \mapsto Y_{ab}^{-1}(\theta) X Y_{ab}(\theta) \implies P_* : X \to \left(\frac{d}{d\theta} [Y_{ab}^{-1}(\theta) X Y_{ab}(\theta)] \right) |_{\theta=0}. \tag{421}$$

The RHS is

$$\begin{aligned} \left(\frac{d}{d\theta} [Y_{ab}^{-1}(\theta) X Y_{ab}(\theta)] \right) |_{\theta=0} &= -Y_{ab}(0)^{-1} Y'_{ab}(0) Y_{ab}(0) X Y_{ab}(0) + Y_{ab}^{-1}(0) X Y'_{ab}(0) \\ &= [X, \gamma_a\gamma_b] \end{aligned} \tag{422}$$

The commutator is

$$\begin{aligned} [X, \gamma_a\gamma_b] = \sum_i x^i [\gamma_i, \gamma_a\gamma_b] \quad &= \sum_i x^i [(-\gamma_a\gamma_i + 2\delta_{ai})\gamma_b - \gamma_a\gamma_b\gamma_i] = 2 \sum_i x^i (\delta_{ai}\gamma_b - \delta_{bi}\gamma_a) \\ = \sum_{ij} (2 A^{ab}_{ij} x^j) \gamma^i, & \end{aligned} \tag{423}$$

where N is a normalization factor for the trace. One might also like to write this as $R_{ij} = \langle \gamma_i | Y \gamma_j Y^{-1} \rangle$, where the γ_is are viewed as basis vectors and the inner product is $\langle X|Y \rangle = \mathrm{Tr}[X^\dagger Y]$; this is then saying that R_{ij} can be found by taking the ith component of the image of the jth basis vector.

As an example of how this works, consider $n = 3$. Using the Pauli matrices for the γ matrices, the claim is that the map $P : \mathrm{Spin}(3) = SU(2) \to SO(3)$ is realized through

$$P : SU(2) \ni U \to R \in SO(3), \qquad R_{ij} = \frac{1}{2} \mathrm{Tr}[U^\dagger \sigma_i U \sigma_j]. \tag{417}$$

As a check, consider the image of $U = e^{i\theta Z/2}$. Then $U = \cos(\theta/2) + iZ \sin(\theta/2)$, and so

$$[P(U)]_{ij} = R_{ij} = \frac{1}{2} \mathrm{Tr} \left[(\cos(\theta/2) - iZ \sin(\theta/2)) \sigma_i (\cos(\theta/2) + iZ \sin(\theta/2)) \sigma_j \right]. \tag{418}$$

If $i = z$ then the RHS is $\mathrm{Tr}[Z\sigma_j]/2 = \delta_{zj}$. If $i = j \neq z$, then we get $\mathrm{Tr}[e^{i\theta Z}]/2 = \cos\theta$, while if $i \neq j \neq z$ then we get $\mathrm{Tr}[e^{i\theta Z} i \epsilon_{jiz} Z]/2 = \epsilon_{ijz} \sin\theta$. Putting these all together gives the 3×3 rotation matrix that we expect.

where A^{ab} is the antisymmetric matrix with a 1 in the bth row and ath column and a -1 in the ath row and bth column: $A^{ab}_{ij} = \delta_{bi}\delta_{aj} - \delta_{bj}\delta_{ai}$ (these matrices form a basis form $\mathfrak{so}(n)$). This means that on the generators of $\mathfrak{spin}(n)$, P_* acts as

$$P_* : \gamma_a\gamma_b \mapsto 2A^{ab} \in \mathfrak{so}(n). \tag{424}$$

This means that the mapping back from $\mathfrak{so}(n)$ to $\mathfrak{spin}(n)$ is, for a generic element $O = \sum_{a<b} \alpha_{ab}A^{ab} \in \mathfrak{so}(n)$ with $\alpha_{ab} = -\alpha_{ba}$,

$$P_*^{-1} : O \mapsto \frac{1}{2}\sum_{a<b} \alpha_{ab}\gamma_a\gamma_b = \frac{1}{4}\sum_{a,b} \alpha_{ab}\gamma_a\gamma_b = \frac{1}{8}\sum_{a,b} \alpha_{ab}[\gamma_a, \gamma_b]. \tag{425}$$

This equation tells us where the factors of 1/8 in the spin connection and the expression for elements in $\text{Spin}(n)$ as exponentials of gamma matrix commutators come from.

The square of the Dirac operator and zero mode solutions; some stuff about veilbeins

Today's diary entry is simple: we want to compute $(i\not{D}_A)^2$ for an arbitrary gauge connection A, on an arbitrary geometry. This will tell us cool stuff about when zero modes of $i\not{D}_A$ can exist.

◇ ◇ ◇

Preliminary veilbein things

Since we are dealing with fermions, we will be using tetrad methods, with $e^a = e^a_\mu dx^\mu$ denoting a 1-form (or using the metric, a vector field) component of the orthonormal basis at each point in spacetime. The roman indices can be raised and lowered with δ_{ab}, while the Greek indices must be raised and lowered using $g_{\mu\nu}$. When acting on e.g a $(1,1)$ tensor V^a_b in the orthonormal basis, the covariant derivative along a vector field X acts as using the spin connection ω_μ as

$$[\nabla_X V]^a_b = X^\mu(\delta^a_c\delta^d_b\partial_\mu + {\omega_\mu}^a{}_c\delta^d_b - {\omega_\mu}^d{}_b\delta^a_c)V^c_d. \tag{426}$$

As usual, covariant indices get minus signs: this is required in order to have contractions like V_aV^a get differentiated appropriately as $\nabla_\mu(V_aV^a) = \partial_\mu(V_aV^a)$. A consequence of this is that the spin connection ${\omega_\mu}^a$ is antisymmetric in a, b, which is a requirement of $\nabla_\mu(\delta^{ab}) = 0$.

One result we will need in the following is that $[\nabla_X e_b]^\mu = 0$: the veilbeins are covariantly constant along the flow of any vector field X. Now the covariant derivative of a veilbein is

$$[\nabla_X e_b]^\nu = X^\mu(\partial_\mu e^\nu_b - {\omega_\mu}^a{}_b e^\nu_a + \Gamma^\nu_{\mu\lambda}e^\lambda_b). \tag{427}$$

This turns out to identically vanish; to see this we need an expression for the spin connection in terms of the veilbeins and the Christoffel symbols. We get this by expressing $\nabla_X V$ in both the spacetime and orthonormal basis as $\nabla_X V = X^\mu \nabla_\mu V^\nu \partial_\nu = X^\mu [\nabla_\mu V]^a e_a$. In the spacetime basis, we of course have $[\nabla_X V]^\nu = X^\mu(\partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda)$. In the orthonormal frame basis, we have $[\nabla_X V]^a = X^\mu(\partial_\mu V^a + \omega_\mu{}^a{}_b V^b)$. So then setting $[\nabla_X V]^a e^\nu_a = [\nabla_X V]^\nu$, we have

$$\begin{aligned} \omega_\mu{}^a{}_b V^b e^\nu_a &= \partial_\mu(V^c e^\nu_c) - e^\nu_a \partial_\mu V^a + \Gamma^\nu_{\mu\lambda} V^b e^\lambda_b \\ &= (\partial_\mu e^\nu_b + \Gamma^\nu_{\mu\lambda} e^\lambda_b) V^b. \end{aligned} \tag{428}$$

Contracting both sides with e^c_ν and then renaming some indices,

$$\omega_\mu{}^a{}_b = e^a_\nu \partial_\mu e^\nu_b + e^a_\nu \Gamma^\nu_{\mu\lambda} e^\lambda_b. \tag{429}$$

so that the spin connection can be thought of as the 1-form

$$\omega^a{}_b = [\nabla^G_{e^a} e_b]_\mu dx^\mu \tag{430}$$

where ∇^G is the "Greek covariant derviative", i.e. the convariant derivative that acts only on Greek (spacetime) indices. Anyway, from (429), we see that the full covariant derivative of the vector field e_b vanishes (and hence so too does the covariant derivative of e^b).

In the following, we will be using the expression for the Riemann curvature tensor in terms of a commutator of two covariant derivatives along the two vector fields e^a, e^b. In terms of the spin connection, we have, for any vector field V^e in the orthonormal frame basis,

$$\begin{aligned} [\nabla_{e^a}, \nabla_{e^b}] V^d &= e^\mu_a e^\nu_b \left((\delta^d_c \partial_\mu + \omega_\mu{}^d{}_c)(\delta^c_e \partial_\nu + \omega_\nu{}^c{}_e) - (\mu \leftrightarrow \nu) \right) V^e \\ &= e^\mu_a e^\nu_b (\delta^d_c \partial_{[\mu} \omega_{\nu]}{}^c{}_e + \omega_{[\mu|}{}^d{}_c \omega_{\nu]}{}^c{}_e) V^e. \end{aligned} \tag{431}$$

Thus we can write the components of the Riemann curvature tensor in the orthonormal frame basis as

$$[R_{ab}]_{de} = [\nabla_{e^a}, \nabla_{e^b}]_{de} = e^\mu_a e^\nu_b [d\omega^d{}_e + \omega^d{}_c \wedge \omega^c{}_e]_{\mu\nu}. \tag{432}$$

Thinking of the curvature tensor as a matrix-valued 2-form on spacetime, this is written more succinctly as $R_{ab} = [d\omega + \omega \wedge \omega]_{ab}$.

Square of the Dirac operator

First we need to set the notation and establish a preliminary result. We will be interested in the Dirac operator $i\not{D}_A : \Gamma(S \otimes E) \to \Gamma(S \otimes E)$, where $\Gamma(S \otimes E)$ denotes C^∞ sections of the bundle $S \otimes E$ with connection A. Here $S = SXL \times_\rho \mathbb{C}^{2^{\lfloor n/2 \rfloor}}$ is the spinor bundle, with ρ the spinor representation and SLX the spin lift of the frame bundle. E is some gauge bundle, and we take the fermion fields (sections of $S \otimes E$) to transform under a $\otimes$ representation of the spinor representation and whatever representation they carry under the gauge group. The covariant derivative on $S \otimes E$ will be written as ∇, and we will write $\nabla^\dagger$ for the adjoint of ∇.

The Dirac operator is $i\not{D}_A = \gamma^a \nabla_{e^a}$. Its square is

$$(i\not{D}_A)^2 = -\gamma^a \nabla_{e^a} \gamma^b \nabla_{e^b} = -\gamma^a \gamma^b e^\mu_a (\nabla_\mu e^\nu_b) \nabla_\nu - \gamma^a \gamma^b e^\mu_a e^\nu_b \nabla_\mu \nabla_\nu = -\gamma^a \gamma^b \nabla_{e^a} \nabla_{e^b}, \tag{433}$$

since $\nabla e^a = 0$. So then

$$\begin{aligned} (i\not{D}_A)^2 &= -\sum_a \gamma^a\gamma^a(\nabla_{e^a})^2 - \frac{1}{2}\sum_{a\neq b}\gamma^a\gamma^b[\nabla_{e^a},\nabla_{e^b}] \\ &= \nabla^\dagger\nabla - \frac{1}{2}\sum_{ab}\gamma^a\gamma^b R_{ab}, \end{aligned} \tag{434}$$

where R_{ab} is the curvature tensor (for each a, b, R_{ab} is an $n\times n$ antisymmetric matrix) for the full connection and $\nabla^\dagger = -\nabla$ is the adjoint of the covariant derivative.[43] We can split the curvature tensor into a gauge part and a spin connection part (since the full connection is $A = \omega\otimes\mathbf{1} + \mathbf{1}\otimes\mathcal{A}$, where $\mathcal{A}$ is the gauge connection, there are no cross terms) as $R_{ab} = R^S_{ab} + R^E_{ab}$, where R^S is the spin part and R^E is the gauge part.

Let's start with the spin part R^S_{ab}: for each a, b, we need to represent this as an $n\times n$ matrix in the spinor representation. Now R^S_{ab} is an antisymmetric matrix[44], and so we can write it as $R^S_{ab} = \sum_{c<d} R_{abcd}A^{cd} = \frac{1}{2}\sum_{cd} R_{abcd}A^{cd}$, where $[A^{cd}]_{ij} = \delta_{cj}\delta_{di} - \delta_{ci}\delta_{dj}$ and R_{abcd} is the Riemann curvature tensor for the spin part of the connection. To find the representation of R^S_{ab} on fermion fields, we need to represent the $\mathfrak{so}(n)$ matrix A^{cd} as a matrix in $\mathfrak{spin}(n)$. We can do this using the map $P_*^{-1} : \mathfrak{so}(n)\to\mathfrak{spin}(n)$ derived in the last diary entry: this gives

$$\begin{aligned} -\frac{1}{2}\sum_{ab}\gamma^a\gamma^b R^S_{ab} &= -\frac{1}{4}\sum_{abcd} R_{abcd}P_*^{-1}(A^{cd}) = -\frac{1}{8}\sum_{abcd}\gamma^a\gamma^b\gamma^c\gamma^d R_{abcd} \\ &= -\frac{1}{8}\sum_a \gamma^a\left(\sum_{b\neq c\neq d}\gamma^b\gamma^c\gamma^d R_{abcd} + \sum_{bd}\left[\gamma^b\gamma^b\gamma^d R_{abbd} + \gamma^b\gamma^d\gamma^b R_{abdb}\right]\right), \end{aligned} \tag{436}$$

where we used the antisymmetry of R_{abcd} in both the first and last pair of indices (so that we can sum over $b\neq c\neq d$, and then get the remaining terms by taking $c = b$ and $d = b$). Now the first term dies by the antisymmetry of the gamma matrices in front of it and the Bianchi identity, which says that $R_{a[bcd]} = 0$. Thus

$$-\frac{1}{2}\sum_{ab}\gamma^a\gamma^b R^S_{ab} = -\frac{1}{8}\sum_{abd}\gamma^a\gamma^d(R_{abbd} - R_{abdb}). \tag{437}$$

Now $R_{abdb} = R_{dbab}$ and $R_{abbd} = R_{dbba}$, so the only terms that survive are those where $a = d$. So, using $R_{abab} = -R_{abba}$,

$$-\frac{1}{2}\sum_{ab}\gamma^a\gamma^b R^S_{ab} = -\frac{1}{4}\sum_{ab} R_{abba} = \frac{1}{4}R. \tag{438}$$

[43] $\nabla^\dagger$ is defined so that $\nabla^\dagger\nabla = -\nabla_\mu\nabla^\mu$ is a positive-definite operator. Indeed, for any compactly-supported ψ,

$$0 \leq \int\langle\nabla\psi,\nabla\psi\rangle = -\int\langle(\nabla)^\mu\nabla_\mu\psi,\psi\rangle = \int\langle\nabla^\dagger\nabla\psi,\psi\rangle. \tag{435}$$

[44] Remember that the matrix R_{ab} tells us how transform a vector into the vector that it becomes after being transported around a rectangle defined by the vector fields $e^a = e^{a\mu}\partial_\mu$ and $e^b = e^{b\mu}\partial_\mu$. A vector and its parallel-transported-around-a-rectangle image differ by a rotation, which is encoded by the matrix R_{ab}. Thus R_{ab} lives in $\mathfrak{so}(n)$, and can be built from a linear combination of the matrices A^{ab} that featured in the last diary entry.

In addition to the spin part, we also have the gauge part. Since the gauge part doesn't act on the spinor indices, there is really no more work to do: we just write it as

$$\sum_{ab} \gamma^a \gamma^b R^E_{ab} = -i \sum_{ab} \gamma^a \gamma^b e^\mu_a e^\nu_b \mathcal{F}_{\mu\nu}, \tag{439}$$

with $\mathcal{F} = d\mathcal{A}$ the curvature for the gauge part.[45] Putting everything together, we have

$$(i\not{D}_A)^2 = \nabla^\dagger \nabla + \frac{1}{4} R + \frac{i}{4} \sum_{ab} [\gamma^a, \gamma^b] e^\mu_a e^\nu_b \mathcal{F}_{\mu\nu}. \tag{440}$$

Now as we established above, $\nabla^\dagger\nabla$ is a positive-definite operator. This means that if the total (gauge + geometric) curvature is everywhere nonzero, there are no solutions to $\not{D}_A \psi = 0$, and hence no zero modes.

As an example, consider fermions on S^2 with $U(1)$ flux such that $\int_{S^2} F = 2\pi m$, as in the last diary entry. Letting $\gamma^1 = X, \gamma^2 = Y$ and $e^{x\theta} = 1, e^{y\phi} = \sin^{-1}\theta$, with $\mathcal{A} = \frac{1-\cos\theta}{2} d\phi$, we find

$$(i\not{D}_A)^2 = -\Delta_A + \frac{1}{2} - \frac{m}{2} Z, \tag{441}$$

where Δ_A is the gauged Laplacian on the sphere: $\Delta_A = (\partial_\mu + i(\omega_\mu + \mathcal{A}_\mu))(\partial^\mu + i(\omega^\mu + \mathcal{A}^\mu))$. This formula shows us that left-handed fermions see positive magnetic flux as negative curvature, while the opposite is true for right-handed fermions (our spinors are $(\psi_L, \psi_R)^T$). Since $-\Delta_A$ is positive definite, zero modes are only possible for non-zero magnetic flux, with the chirality of the zero modes depending on the sign of the flux.

Another fun application of these results is to apply them to the $d, d^\dagger$ differential complex, by taking the dirac operator to be the Laplacian $\Delta = d + d^\dagger$. This tells us that if $R \geq 0$ everywhere, Δ has no zero modes: there are no harmonic forms on a manifold which is everywhere positively curved.

Riemann curvature tensor and parallel transport

Today's diary entry is pretty simple—just doing a simple calculation I hadn't done before. It's a problem in Carroll's GR book.

One way to define the curvature tensor R is via

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, \tag{442}$$

where $\nabla_{[X,Y]} Z = \nabla_{\nabla_X Y} Z - \nabla_{\nabla_Y X} Z$ vanishes if X and Y are coordinate vector fields.

Another way to define the curvature tensor $R(X,Y)$ is as the matrix which relates a vector Z to its image under parallel transport through a rectangle defined by the flow lines

[45] The factor of $-i$ is there because we take the covariant derivative to involve $i\mathcal{A}_\mu$, and there is a further sign coming from the i^2 in $(i\not{D}_A)^2$ that we need to include to get the signs straight.

of the vector fields X and Y: if the difference between Z and its parallel-transported image is δZ, then

$$\delta Z^\rho = R^\rho_{\mu\nu\lambda} X^\mu Y^\nu Z^\lambda. \tag{443}$$

Why are these two definitions the same? (This isn't totally obvious to me; one might have thought that parallel transport along a rectangle would involve terms quartic in covariant derivatives)

◇ ◇ ◇

The proof proceeds in essentially the same way as the proof of the analogous statement for conventional gauge theories. Parallel transporting a vector Z along a curve $x(t)$ means that $Z(t)$ is covariantly constant:

$$[\nabla_{x(t)} Z]^\rho = \dot{x}^\mu(\partial_\mu Z^\rho + \Gamma^\rho_{\mu\lambda} Z^\lambda) = 0, \tag{444}$$

so that along the curve, $\partial_\mu Z^\rho = -\Gamma^\rho_{\mu\lambda} Z^\lambda$. Integrating this along the curve and then iterating, the usual procedure of integrating over simplices leads to

$$\begin{aligned} Z^\rho(t) &= Z^\rho(t_0) - \int_{t_0}^t dt'\, \dot{x}^\lambda \Gamma^\rho_{\mu\lambda} Z^\lambda(t') \\ &= \left[P \exp\left(- \int_{t_0}^t dt'\, \dot{x}^\lambda \Gamma_\lambda \right) \right]^\rho_\mu Z^\mu(t_0), \end{aligned} \tag{445}$$

where we are viewing Γ_λ as a matrix with one upper and one lower index.

To prove the equivalence between the two definitions of the curvature tensor, we need to consider the path t to be a small rectangular closed path such that $x(t) = x(t_0)$. We will let the sides of the rectangle have lengths δa and δb, and will wolog take the sides to be along the x^1 and x^2 axes, respectively. This makes doing the integrals easy, since we can parametrize the paths by x^1 and x^2, letting $t \to x^1, x^2$, depending on which side of the rectangle we are on.

With this setup, we then just need to expand the above equation to second order in $\delta a, \delta b$.[46] Let the four corners of the path be $1, 2, 3, 4$. For each path segment, we need to expand the path-ordered exponential to second order. The second order term is, to our order, the matrix $\delta a^2 \Gamma^\rho_{1\mu}\Gamma^\mu_{1\lambda}/2$ or $\delta b^2 \Gamma^\rho_{2\mu}\Gamma^\mu_{2\lambda}/2$, depending on which part of the path we're on. The first-order terms for the 23 and 34 parts of the path can be expanded in terms of the first-order terms for the 12 and 41 parts of the path by taking a first-order derivative expansion of the Γ matrices. This gives

$$\begin{aligned} Z^\rho(t) = &\left(\delta^\rho_\alpha - \int_1^2 \Gamma^\rho_{1\alpha} + \frac{\delta a^2}{2} \Gamma^\rho_{1\lambda}\Gamma^\lambda_{1\alpha} \right) \left(\delta^\alpha_\sigma - \int_2^3 \Gamma^\alpha_{2\sigma} + \frac{\delta b^2}{2} \Gamma^\alpha_{2\lambda}\Gamma^\lambda_{2\sigma} \right) \\ &\times \left(\delta^\sigma_\beta + \int_1^2 (\Gamma^\sigma_{1\beta} - \delta b \partial_2 \Gamma^\sigma_{1\beta}) + \frac{\delta a^2}{2} \Gamma^\sigma_{1\lambda}\Gamma^\lambda_{1\beta} \right) \left(\delta^\beta_\omega + \int_2^3 (\Gamma^\beta_{2\omega} + \delta a \partial_1 \Gamma^\beta_{2\omega}) + \frac{\delta b^2}{2} \Gamma^\beta_{2\lambda}\Gamma^\beta_{2\omega} \right) Z^\omega(t_0). \end{aligned} \tag{446}$$

[46]This will suffice for the proof if the claim is true, since the first formula for the curvature tensor given above is quadratic in derivatives.

The terms that involve only one integral of a single Γ matrix cancel pairwise, while the terms that have the explicit factors of $\delta a^2/2$ and $\delta b^2/2$ are canceled by terms that are quadtratic products of the integrals involving the single Γ matrices (for example, the terms with the explicit $\delta a^2/2$ dependence are canceled by the the product of the two $\int_1^2 \Gamma^\rho_{1\omega}$ terms).

The only terms that survive are then the ones involving derivatives, and ones that contain products of $\int_1^2$ integrals with $\int_2^3$ integrals. Some of these cancel, but two terms remain and we can then do the integrals to lowest order and get

$$Z^\rho(t) - Z^\rho(t_0) = \delta a \delta b \left(\partial_1 \Gamma^\rho_{2\omega} - \partial_2 \Gamma^\rho_{1\omega} + \Gamma^\rho_{1\lambda}\Gamma^\lambda_{2\omega} - \Gamma^\rho_{2\lambda}\Gamma^\lambda_{1\omega}\right) Z^\omega(t_0). \tag{447}$$

This exactly the transformation we expect from (443), with the vector fields X and Y chosen to be x^1 and x^2, respectively. Indeed, a quick check of the commutator of ∇_X with ∇_Y (no $\nabla_{[X,Y]}$ needed since in this case X, Y are coordinate vector fields) shows that we get the above expression of Christoffel symbols.

Why is the periodic spin structure on S^1 non-bounding?

Today we have a super short diary entry — just explaining the question in the title, in a down-to-earth way.

◇ ◇ ◇

Since Spin(1) $= \mathbb{Z}_2$, and since there are two $\mathbb{Z}_2$ bundles over the circle, the circle has two spin stuctures. One spin structure corresponds to the trivial product bundle, and periodic boundary conditions, while the other corresponds to the "edge of a mobius strip" bundle, and has anti-periodic boundary conditions. It turns out that only the latter defines a spin structure which can be extended into a disk bounded by the circle.

To see this, note that a framing on S^1 is a choice of tangent vector $\pm\hat{\phi}$ at every point on S^1, when the S^1 is thought of as living in the plane. The P spin structure has a definite choice of tangent vector and trivial transition functions, while the AP spin structure is defined with two patches, with a transition on one of the patch overlaps that flips the framing by -1.

To see whether these extend into the disk, we draw the disk as two patches 1 and 2, with their overlap constituting a thin strip running along a diameter of the disk. The spin framing is restricted to be tangent to the disk boundary at the disk boundary, and the transition function g_{12} between the two patches is required to be $\mathbf{1}$ on the left edge of the overlap, and either $\mathbf{1}$ (P spin structure) or $-\mathbf{1}$ (AP spin structure) on the right edge of the overlap.

Now Spin(2) $= U(1)$, so we can draw the spin framing in the interior of the disk as a vector field in the plane. Requiring that the framing be nonzero everywhere on each patch, given the requirement that the framing become tangential on the disk boundary, essentially fixes the framing on each patch to be as shown in Figure 3. Given that the transition function is $\mathbf{1}$ on the left side of the patch overlap, we see that the transition function must become a π rotation at the center of the disk, and then grow to a 2π rotation on the right edge of

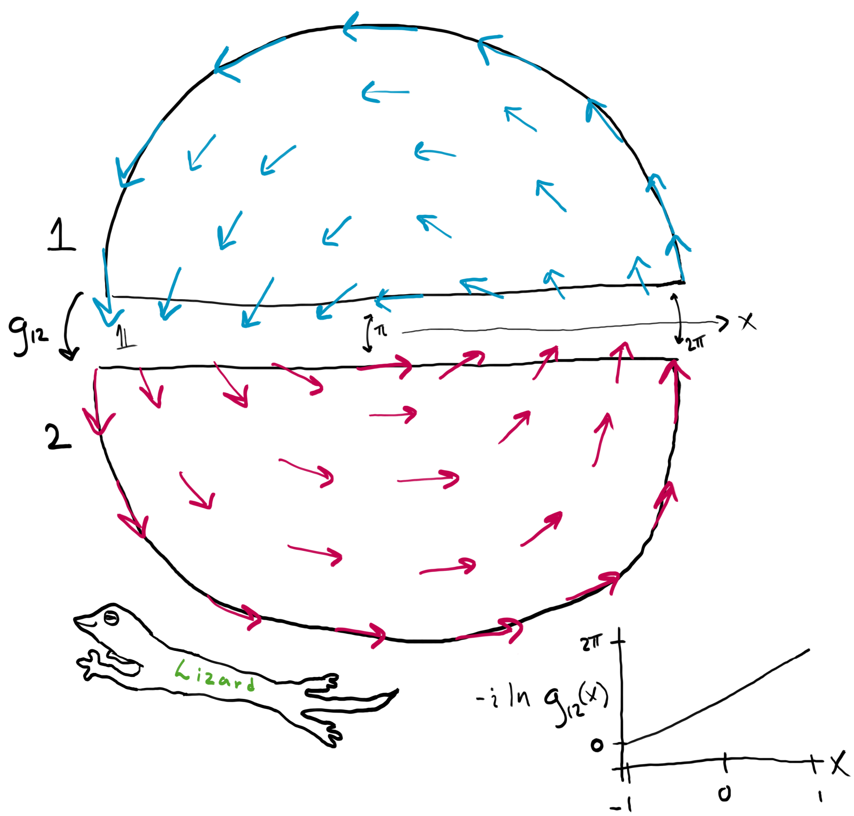


Figure 3: Spin framings on the two halves of the disk, the gluing function between them, and a lizard.

the disk. Now a 2π rotation in Spin(2) is identified with $-\mathbf{1}$, and so the transition function must be $-\mathbf{1}$ on the right edge of the disk (if we walk around the disk boundary, following the framing in the spin bundle, we must walk around twice before we get back to where we started).

This assignment for the transition function g_{12} at each point on the overlap is only compatible if the circle has the AP spin structure, and so we conclude that only the AP spin structurebounds. If we tried to get the P spin structureto bound, we'd run into a problem of defining the spin framing at the center of the disk.

SW transformation and derivatives of matrix exponentials

Today we're just working out some math facts that we need to do SW transformations on Hamiltonians. Nothing fancy, but I thought it'd be nice to have around as a reference.

◇ ◇ ◇

First, we want an expression for $\partial_t e^\Lambda$ which doesn't assume that $[\Lambda, \dot{\Lambda}] = 0$. First, expand the exponential: there are n ways to take the time derivative in the nth term, and so

$$\partial_t e^\Lambda = \sum_{n\in\mathbb{N}} \sum_{k=0}^{n} \frac{1}{(n+1)!} \Lambda^k \dot{\Lambda} \Lambda^{n-k}. \tag{448}$$

To deal with this, we systematically commute the time derivative to the front of each term:

$$\begin{aligned} \Lambda^k \dot{\Lambda} \Lambda^{n-k} &= \Lambda^{k-1}(\dot{\Lambda}\Lambda + [\Lambda, \dot{\Lambda}])\Lambda^{n-k} \\ &= \Lambda^{k-2}(\dot{\Lambda}\Lambda + [\Lambda, \dot{\Lambda}])\Lambda^{n-k} + \Lambda^{k-2}([\Lambda, \dot{\Lambda}]\Lambda + [\Lambda, [\Lambda, \dot{\Lambda}]])\Lambda^{n-k}, \end{aligned} \tag{449}$$

and so on. Let us define the nested commutator

$$\mathcal{C}_n \equiv [\Lambda, [\cdots, [\Lambda, \dot{\Lambda}] \cdots]], \tag{450}$$

where $\mathcal{C}_n$ contains n nested commutators; $\mathcal{C}_0 = \dot{\Lambda}, \mathcal{C}_1 = [\Lambda, \dot{\Lambda}]$, etc. Playing with this a bit, we see that if the derivative occurs in a spot with k powers of Λ before it, then there are k commutation steps needed, and k choose l ways to produce the term $\mathcal{C}_l$. Therefore

$$\partial_t e^\Lambda = \sum_{n\in\mathbb{N}} \sum_{k=0}^{n} \sum_{l=0}^{k} \binom{k}{l} \mathcal{C}_l \Lambda^{n-l}. \tag{451}$$

Now we recall the identity

$$\sum_{k=l}^{n} \binom{k}{l} = \binom{n+1}{l+1}. \tag{452}$$

For a fixed n, consider a fixed l. It will appear in all sums with $k \geq l$. Therefore we can re-arrange the sum by l and write

$$\partial_t e^\Lambda = \sum_{n\in\mathbb{N}} \sum_{l=0}^{n} \binom{n+1}{l+1} \mathcal{C}_l \Lambda^{n-l} = \sum_{n\in\mathbb{N}} \sum_{l=0}^{n} \frac{\mathcal{C}_l}{(l+1)!} \frac{\Lambda^{n-l}}{(n-l)!}. \tag{453}$$

Again consider a fixed l. It appears in every summand with $n \geq l$, and these summands contain $\Lambda^m/m!$, where $m = n - l$ is summed over all $\mathbb{N}$. Therefore we can re-write the sum over n as an exponential, yielding

$$\partial_t e^\Lambda = \sum_{l\in\mathbb{N}} \frac{\mathcal{C}_l}{(l+1)!} e^\Lambda. \tag{454}$$

Another related expression that we'll need is one for $e^\Lambda H e^{-\Lambda}$, which comes up when doing SW transformations. This one is comparatively easy:

$$e^\Lambda H = \sum_{n\in\mathbb{N}} \sum_{k=0}^{n} \frac{1}{n!} \mathcal{N}_k \binom{n}{k} \Lambda^{n-k}, \tag{455}$$

where we have defined

$$\mathcal{N}_k = [\Lambda, [\cdots, [\Lambda, H] \cdots]], \tag{456}$$

where there are k total Λs appearing. To get the expression for $e^\Lambda H$, we used the same strategy: after expanding the exponential we move the H through the Λs until it gets to the left. For the nth term in the expansion, the kth order nested commutator has $\binom{n}{k}$ ways of appearing; hence the RHS of (455). Then using the same strategy as above,

$$e^\Lambda H = \sum_{n\in\mathbb{N}} \sum_{k=0}^{n} \frac{\mathcal{N}_k}{k!} \frac{1}{(n-k)!} \Lambda^{n-k} = \sum_{k\in\mathbb{N}} \frac{\mathcal{N}_k}{k!} e^\Lambda, \tag{457}$$

and so

$$e^\Lambda H e^{-\Lambda} = \sum_{k\in\mathbb{N}} \frac{\mathcal{N}_k}{k!}. \tag{458}$$

Kahler pre-quantization practice

Today we're doing a long elaboration on a homework problem assigned by Greg Moore during his 2019 TASI lectures. We will give a prescription for doing pre-quantization of functions on phase space, and will then apply our machinery to the case where the phase space is S^2, thereby obtaining the Hilbert space for a quantum spinful particle.

◇ ◇ ◇

Generalities

First, some notation. Consider a phase space $\mathcal{P}$ with sympletic form ω. Functions[47] $f : \mathcal{P} \to \mathbb{C}$ are quantized by associating them to $\mathbb{C}$-linear operators $\mathcal{Q}[f]$ which act on $\mathcal{H}$. There are many ways of doing this, and so we need to fix some rules that determine what properties we would like the $\mathcal{Q}[f]$ to satisfy.

First, we will demand that commutation relations of the quantized functions are determined via the Poisson bracket as[48]

$$[\mathcal{Q}[f], \mathcal{Q}[g]] = i\hbar\mathcal{Q}[\{f, g\}], \tag{459}$$

where the Poisson bracket is as usual

$$\{f, g\} = \omega(f, g) = \omega^{ij}\partial_i f \partial_j g. \tag{460}$$

We could consider higher $\hbar$ corrections to (459) (deformation quantization), but for our purposes we will be able to find $\mathcal{Q}[f]$ such that (459) is satisfied exactly.

Now in this scheme, it is in general not possible to have the $\mathcal{Q}[f]$ operators form a linear representation; instead we require the weaker condition

$$||\mathcal{Q}[f]\mathcal{Q}[g] - \mathcal{Q}[f \cdot g]|| = O(\hbar^2). \tag{461}$$

Finally, we require that constants map as

$$\mathcal{Q}[a] = a\mathbf{1}, \tag{462}$$

and that complex conjugation of functions maps to Hermitian conjugation of operators:

$$\mathcal{Q}[f^*] = \mathcal{Q}[f]^\dagger. \tag{463}$$

First, let's do a warmup, where we take $\mathcal{P} = \mathbb{R}^2$. Of course $\mathbb{R}^2 \cong T^*\mathbb{R}$ can be treated by sending $\mathcal{Q}[x^1] = x$, $\mathcal{Q}[x^2] = -i\hbar\partial_x$, but this quantization procedure is rather asymmetric in $\mathbb{R}^2$ and won't generalize to cases where the phase space is not a cotangent bundle, like $\mathcal{P} = S^2$. We will therefore proceed in complex coordinates, with symplectic form $\omega_{z\bar{z}} = -\omega_{z\bar{z}} = \frac{i}{2}$, so that $\omega = dx \wedge dy$. Therefore the Poisson bracket gives

$$\{z, \bar{z}\} = \omega^{z\bar{z}} = -2i, \tag{464}$$

which tells us that we ought to have

$$[\mathcal{Q}[z], \mathcal{Q}[\bar{z}]] = i\hbar\{z, \bar{z}\} = 2\hbar. \tag{465}$$

Consider then the assignment

$$\mathcal{Q}[z] = 2\hbar\bar{\partial} + z/2, \qquad \mathcal{Q}[\bar{z}] = -2\hbar\partial + \bar{z}/2. \tag{466}$$

[47] Really, sections of complex line bundles over $\mathcal{P}$.

[48] Note the sign on the RHS! Many references have $-i\hbar$ instead.

This gives the correct commutator and satisfies $\mathcal{Q}[z]^\dagger = \mathcal{Q}[\bar{z}]$, and so we have successfully quantized the variables on $\mathcal{P}$.

For quantization on the flat plane, we could basically guess the form for the $\mathcal{Q}$ map—for more complicated examples, we will need to have a better system for figuring out what $\mathcal{Q}$ should be. In fact there are several options; quantization is not a unique procedure. The one I've found to be the most useful is to take

$$\begin{aligned}\mathcal{Q}[f] &= -i\hbar X_f^\mu \partial_\mu + \theta_\mu X_f^\mu + f \\ &= -i\hbar \nabla_{X_f} + \theta(X_f) + f \\ &= -i\hbar X_f^\mu D_\mu + f,\end{aligned} \tag{467}$$

where the covariant derivative is $D_\mu = \partial_\mu + i\theta_\mu/\hbar$ and X_f is the Hamiltonian vector field associated with f. Here our conventions are such that

$$X_f^\mu = \omega^{\mu\nu}\partial_\nu f \implies i_{X_f}\omega = -df. \tag{468}$$

Honestly this sign choice is sub-ideal, but this was only realized after finishing the calculations. The $+f$ term here is needed so that $\mathcal{Q}[1] = \mathbf{1}$, while the potential term $\theta(X_f)$ is needed for (459) to be satisfied, as we will see. Note that the sign on the potential is positive—the geometric quantization reference I was reading had the opposite sign, which led to a huge effort to track down where my calculation was going wrong, so use this sign when in doubt.

Note that with this choice of quantization we will never have $\mathcal{Q}[f \cdot g] = \mathcal{Q}[f]\mathcal{Q}[g]$ for f, g with $X_f, X_g \neq 0$, since the RHS is quadratic in derivatives while the LHS is linear. This is okay though, since the term quadratic in derivatives is always $O(\hbar^2)$.

The main reason for choosing this quantization is that it satisfies (459) on the nose (i.e. there are no higher $\hbar$ corrections). One drawback is that we can only ever produce operators that are linear in derivatives, which prevents us from getting formulae like e.g. $\mathcal{Q}[p^2] = -\hbar^2\nabla^2$. Sorting out this issue is a lot of work; I might come back to it at a later point.

Before proving why this quantization works, let's run a sanity check on the plane. We calculate

$$X_z = 2id\bar{z}, \qquad X_{\bar{z}} = -2idz, \qquad \theta = \frac{i}{4}(zd\bar{z} - \bar{z}dz), \tag{469}$$

so that

$$\theta(X_z) = -z/2, \qquad \theta(X_{\bar{z}}) = -\bar{z}/2. \tag{470}$$

Plugging these into (467), we recover the same quantization assignements as written above: for example, $\mathcal{Q}[z] = -i\hbar(2i)\bar{\partial} - z/2 + z = 2\hbar\bar{\partial} + z/2$ ✓.

Now we will prove that this choice of $\mathcal{Q}$ indeed satisfies (459) exactly. In the computation of $[\mathcal{Q}[f], \mathcal{Q}[g]]$, we have two groups of terms. The first comes from the commutator of the covariant derivatives, and gives

$$[X_f^\mu(-i\hbar\partial_\mu + \theta_\mu), X_g^\nu(-i\hbar\partial_\nu + \theta_\nu)] = -\hbar^2(\nabla_{X_f} X_g^\nu - \nabla_{X_g} X_f^\nu)\partial_\nu - i\hbar(\nabla_{X_f}\theta(X_g) - \nabla_{X_g}\theta(X_f)). \tag{471}$$

The second term gives, after taking the derivatives of $\theta(X_{f/g})$,

$$i\hbar(\nabla_{X_f}\theta(X_g) - \nabla_{X_g}\theta(X_f)) = i\hbar\omega(X_f, X_g) + i\hbar\theta(\nabla_{X_f}X_g - \nabla_{X_g}X_f), \tag{472}$$

since $\omega = d\theta$. Now as we will show in tomorrow's diary entry, for any two Hamiltonian vector fields X_f, X_g, we have

$$[X_f, X_g] = -X_{\{f,g\}}, \tag{473}$$

where the minus sign could be fixed if we went back and changed conventions to $i_{X_f}\omega = +df$. Therefore[49]

$$\begin{aligned}[][X_f^\mu(-i\hbar\partial_\mu + \theta_\mu), X_g^\nu(-i\hbar\partial_\nu + \theta_\nu)] &= +\hbar^2\nabla_{X_{\{f,g\}}} + i\hbar\theta(X_{\{f,g\}}) - i\hbar\{f,g\} \\ &= i\hbar(-i\hbar\nabla_{X_{\{f,g\}}} + \theta(X_{\{f,g\}}) - \{f,g\}).\end{aligned} \tag{475}$$

This is almost right, except the curvature contribution $\{f,g\}$ has the wrong sign. This is fixed by the remainder of $[\mathcal{Q}[f], \mathcal{Q}[g]]$, which contains derivatives of f and g. This part is

$$[\mathcal{Q}[f], \mathcal{Q}[g]] \supset -i\hbar(\nabla_{X_f} g - \nabla_{X_g} f) = -i\hbar(X_f^\mu X_g^\nu \omega_{\mu\nu} - X_g^\mu X_f^\nu \omega_{\mu\nu}) = 2i\hbar\{f,g\}. \tag{476}$$

This flips the sign of the last term in (475), and so we get

$$[\mathcal{Q}[f], \mathcal{Q}[g]] = i\hbar(-i\hbar\nabla_{X_{\{f,g\}}} - \theta(X_{\{f,g\}}) + \{f,g\}) = i\hbar\mathcal{Q}[\{f,g\}], \tag{477}$$

as claimed.

$\mathcal{P} = S^2$ and the quantum spin

Now let's specialize to the case where $\mathcal{P} = S^2$, with sympletic form $\omega \propto d\cos\theta \wedge d\phi$. To facilitate calculating Poisson brackets, we will find it useful to work in stereographic projection: looking back at the diary entry on the Hopf fibration, the coordinates in $\mathbb{R}^3$ are determined in terms of the complex coordinates as (here we are projecting from the north pole)

$$X = \frac{z+\bar{z}}{1+|z|^2}, \qquad Y = \frac{z-\bar{z}}{i(1+|z|^2)}, \qquad Z = \frac{1-|z|^2}{1+|z|^2}. \tag{478}$$

In these coordinates, the symplectic form is

$$\omega = \frac{ik\, dz \wedge d\bar{z}}{(1+|z|^2)^2} \implies \omega_{z\bar{z}} = -\omega_{\bar{z}z} = \frac{ik}{(1+|z|^2)^2}, \tag{479}$$

where k is a constant. The normalization here can be determined from

$$\int_{\mathbb{R}^2} \omega = 2k\int \frac{1}{(1+r^2)^2} dx \wedge dy = 2\pi k \int_0^\infty \frac{2r dr}{(1+r^2)^2} = 2\pi k. \tag{480}$$

This normalization is chosen so that, if $k \in \mathbb{Z}$, the Berry connection θ is a canonically normalized $U(1)$ gauge field ($\int \omega \in \overline{\mathbb{Z}}$), meaning that each semiclassical state occupies a phase space volume of 2π. Note that $k \in \mathbb{Z}$ is required in order for integrals like $\oint dx^\mu\, \theta_\mu$

[49] Here we need to pay attention to signs to determine that

$$\omega(X_f, X_g) = \omega_{\mu\nu} X_f^\mu X_g^\nu = X_f^\mu \partial_\mu g = \omega^{\mu\lambda}\partial_\lambda f \partial_\mu g = -\{f,g\}. \tag{474}$$

to make sense in an action (where the contour is around some curve in S^2), and so in what follows k will be an arbitrary integer. k determines the Chern class of the complex line bundle over $\mathcal{P}$, and as we will see later, determines how many states the quantum Hilbert space has.

It will be useful to know the derivatives of these coordinates wrt z and $\bar{z}$. Letting $\sigma \equiv 1+|z|^2$, we find

$$\begin{aligned} dX &= \sigma^{-2}((1-\bar{z}^2)dz + (1-z^2)d\bar{z}), \qquad dY = -i\sigma^{-2}((1+\bar{z}^2)dz - (1+z^2)d\bar{z}), \\ dZ &= -\sigma^{-2}((1+\bar{z})dz + (1+z)d\bar{z}). \end{aligned} \tag{481}$$

One then checks that the Poisson brackets work out appropriately. For example,

$$\begin{aligned} \{X,Y\} &= \omega^{\mu\nu}\partial_\mu X \partial_\nu Y = -\frac{i\sigma^2}{k}(\partial X \bar{\partial} Y - \bar{\partial} X \partial Y) \\ &= -\frac{i}{k\sigma^2}\left[(1-\bar{z}^2)i(1+z^2) - (-i(1+\bar{z}^2)(1-z^2))\right] \\ &= 2\frac{1-|z|^4}{k\sigma^2} = \frac{2}{k}Z, \end{aligned} \tag{482}$$

and in general,

$$\{\widetilde{X}^i, \widetilde{X}^j\} = \epsilon^{ijk}\widetilde{X}^k, \qquad \widetilde{X}^i = \frac{k}{2}X^i, \tag{483}$$

which when quantized will give us the angular momentum algebra.

We now need to identify what the images of the X^i are under quantization. From the above calculation, we require that they satisfy

$$[\mathcal{Q}[X^i], \mathcal{Q}[X^j]] = i\hbar\epsilon^{ijk}\mathcal{Q}[X^k], \tag{484}$$

since the quantization map $\mathcal{Q}$ is linear (sorry for the minus sign! Too late to switch conventions).

First let us identify the Hamiltonian vector fields corresponding to each of the coordinates. This is done straightforwardly using

$$V^\mu_{X_i} = \omega^{\mu\nu}\partial_\nu X_i. \tag{485}$$

Some algebra gives

$$V_X = -\frac{i}{k}(1-z^2)\partial + \frac{i}{k}(1-\bar{z}^2)\bar{\partial}, \qquad V_Y = \frac{1}{k}(1+z^2)\partial + \frac{1}{k}(1+\bar{z}^2)\bar{\partial} \qquad V_Z = \frac{2i}{k}(z\partial - \bar{z}\bar{\partial}) \tag{486}$$

Since $\partial^\dagger = -\bar{\partial}$, all the vector fields are anti-Hermitian as differential operators.

We will also need the symplectic potential. We can take it to be either of the following choices:

$$\theta = \frac{ik\, z d\bar{z}}{1+|z|^2}, \qquad \text{or} \qquad \theta' = \frac{ik}{2(1+|z|^2)}(zd\bar{z} - \bar{z}dz). \tag{487}$$

When we contract with the various Hamiltonian vector fields, we get

$$\theta(V_X) = \sigma^{-1}(\bar{z}|z|^2 - z), \qquad \theta(V_Y) = i\sigma^{-1}(\bar{z}|z|^2 + z), \qquad \theta(V_Z) = 2\sigma^{-1}|z|^2. \tag{488}$$

Similarly,

$$\theta'(V_X) = -\frac{1}{2}(1-|z|^2)X, \qquad \theta'(V_Y) = -\frac{1}{2}(1-|z|^2)Y, \qquad \theta'(V_Z) = i(Z-1). \tag{489}$$

Using these vector fields, we can now write,[50] after some algebra,

$$\begin{aligned} \mathcal{Q}'[X] &= \frac{\hbar}{k}(-(1-z^2)\partial + (1-\bar{z}^2)\bar{\partial}) + \frac{z+\bar{z}}{2}, \\ \mathcal{Q}'[Y] &= -i\frac{\hbar}{k}((1+z^2)\partial + (1+\bar{z}^2)\bar{\partial}) + \frac{z-\bar{z}}{2i} \\ \mathcal{Q}'[Z] &= \hbar(z\partial - \bar{z}\bar{\partial}) + 1. \end{aligned} \tag{492}$$

Or, if we use the other symplectic potential,

$$\begin{aligned} \mathcal{Q}'[X] &= \frac{\hbar}{k}(-(1-z^2)\partial + (1-\bar{z}^2)\bar{\partial}) + \bar{z}, \\ \mathcal{Q}'[Y] &= -i\frac{\hbar}{k}((1+z^2)\partial + (1+\bar{z}^2)\bar{\partial}) + i\bar{z} \\ \mathcal{Q}'[Z] &= \frac{2\hbar}{k}(z\partial - \bar{z}\bar{\partial}) + 1. \end{aligned} \tag{493}$$

The operators which satisfy the usual angular momentum algebra are then

$$J_i \equiv \frac{k}{2}\mathcal{Q}[X_i]. \tag{494}$$

One can check that these satisfy the correct angular momentum algebra; I won't write out the algebra here.

Hilbert space

Having figured out the algebra, let's now figure out the Hilbert space. Like most of geometric quantization, this seems to be more an art than a science. Basically, the space of all functions on $\mathcal{P}$ is too big, so we need to choose a polarization to break it down to get the physical quantum Hilbert space. This is usually done by taking only those sections ψ that

[50] Note for posterity's sake: the following choices

$$\begin{aligned} \mathcal{Q}[X] &= \frac{\hbar}{2}(-(1-z^2)\partial + (1-\bar{z}^2)\bar{\partial}) + \frac{z+\bar{z}}{1+|z|^2}, \\ \mathcal{Q}[Y] &= -i\frac{\hbar}{2}((1+z^2)\partial + (1+\bar{z}^2)\bar{\partial}) - i\frac{z-\bar{z}}{1+|z|^2} \\ \mathcal{Q}[Z] &= \hbar(z\partial - \bar{z}\bar{\partial}) + \frac{1-|z|^2}{1+|z|^2}. \end{aligned} \tag{490}$$

also satisfy the correct commutation relations. These are the operators you get when you neglect the symplectic potential and write

$$\mathcal{Q}[f] = -i\hbar\nabla_{V_f} + f. \tag{491}$$

(also, here ω is normalized so that $\int \omega \in 4\pi\mathbb{Z}$). What's up? Is this a coincidence that this also works?

satisfy $i_X D\psi = 0$, where X is a section of some n-dimensional subbundle P of the full $2n$-dimensional bundle $L \to \mathcal{P}$. For example, in the case where $\mathcal{P}$ is a cotangent space, we might choose P to be the subbundle consisting of all the momenta, so that the sections remaining in the physical Hilbert space would be functions of coordinates only. In the Kahler case, a natural polarization is to choose P to be the subbundle of anti-holomorphic sections, so that the physical Hilbert space contains only the holomorphic sections, $\mathcal{H} = \mathrm{Hol}(L)$.

After we have chosen a polarization, we need to choose a metric—$\mathcal{H}$ then consists of the square-integrable sections ψ such that $i_X D\psi = 0$. This is sometimes a bit tricky—for example, in the case of $\mathcal{P} = \mathbb{R}^2$, if we were to work with the natural momentum polarization mentioned above, we would find no square-integrable sections, since all inner products of such polarized sections would contain the divergent integral $\int_{\mathbb{R}} dp$.

For cases where the phase space $\mathcal{P}$ is a Kahler manifold however, there is a natural inner product: choosing the polarization so that $\mathcal{H}$ consists of holomorphic sections of the complex line bundle $L \to \mathcal{P}$, the inner product is

$$\langle \psi, \eta \rangle = \int \omega\, e^{-\mathcal{K}} \psi^*(z)\eta(z), \qquad \psi, \eta \in \mathrm{Hol}(L). \tag{495}$$

Here $\mathcal{K}$ is the Kahler potential, which in the present setting of $\mathcal{P} = S^2$ is (note the factor of k; most references do not put the k here)

$$\mathcal{K} = k \ln(1 + |z|^2). \tag{496}$$

In general the Kahler potential is such that $i\partial\bar{\partial}\mathcal{K} = \omega$. Indeed, we have

$$\theta = \frac{ik\bar{z}dz}{1 + |z|^2} = i\partial\mathcal{K}, \tag{497}$$

which gives $i\partial\bar{\partial}\mathcal{K} = \omega$ since $d\theta = i\bar{\partial}\partial\mathcal{K} = \omega$. Note that the other choices of symplectic potential, like $i\bar{\partial}\mathcal{K}$, also give the correct symplectic form via $\omega = i\partial\bar{\partial}\mathcal{K}$, but the above choice is preferred, since with this choice $\theta_{\bar{z}} = 0$, meaning that we can define our holomorphic sections by the requirement $D_{\bar{X}}\psi = 0$, where $\bar{X} = \bar{\partial}$.[51]

Let's explain where the $e^{-\mathcal{K}}$ in the inner product comes from. While it initially looks rather mysterious, it actually just comes from requiring the integrand be well-defined when we change between patches on $\mathcal{P}$. Now in general, the local expression for a section ψ_α on different patches U_α, U_β, is

$$\psi_\alpha(z_\alpha) = g_{\alpha\beta}^k \psi_\beta(z_\beta), \tag{498}$$

where $g_{\alpha\beta}$ is the transition function on the line bundle L_1 with Chern class 1 (the bundle that the ψs are sections of is $L = L_1^{\otimes k}$), and is a holomorphic function of z_β (or z_α). Therefore

$$(\psi^*\eta)_\alpha = |g_{\alpha\beta}|^{2k} (\psi^*\eta)_\beta, \tag{499}$$

and so this part of the integrand is not well-defined. Therefore if (495) is to be well-defined, we need to have a compensating $|g_{\alpha\beta}|^{-2k}$ coming from the transformation of $\omega e^{-\mathcal{K}}$.

[51] With this choice, $D_{\bar{X}} = -i\hbar\bar{\partial}$; if we had let $\theta = i\bar{\partial}\mathcal{K}$ then we would have had $D_{\bar{X}} = -i\hbar\bar{\partial} + ikz/(1+|z|^2)$, and so we wouldn't be able to define holomorphic sections by the polarization $D_{\bar{X}}\psi = 0$.

Indeed, such a compensation comes from the relations between the Kahler potentials on overlapping patches:

$$\mathcal{K}_\beta = \mathcal{K}_\alpha + k\left(\ln g_{\alpha\beta} + \ln g^*_{\alpha\beta}\right). \tag{500}$$

We see that this transformation rule leaves the symplectic form invariant and provides exactly the factors needed to ensure that the integrand is well-defined (i.e. independent of the choice of patch we express it in).

Let's check that this is indeed how $\mathcal{K}$ transforms for the case of $\mathcal{P} = \mathbb{CP}^N$. We will parametrize $\mathbb{CP}^N$ by $\{z_1, \dots, z_{N+1}\}/\sim$, where $\sim$ is the usual rescaling relation. The patch U_i is defined to be the subset of $\mathbb{C}^{N+1}$ where $z_i \neq 0$, and we equip it with the coordinates $x_i^j = z_j/z_i$. The Kahler potential on U_i is

$$\mathcal{K}_i = k\ln\left[\sum_{j=1}^{N+1}|x_i^j|^2\right], \tag{501}$$

which reduces to the one written above for S^2 when $N = 1$. Now we see that the transition functions g_{ij} are x_i^j, since $x_i^l = x_i^j x_j^l = g_{ij}x_j^l$. Therefore we may write

$$\mathcal{K}_j = k\ln\left[|x_j^i|^2\sum_{l=1}^{N+1}|x_i^l|^2\right] = \mathcal{K}_i + k(\ln g_{ji} + \ln g^*_{ji}), \tag{502}$$

which shows that $\mathcal{K}_i$ transforms between patches in the way that we claimed[52].

Now, what is the Hilbert space? Suppose that the largest power of z appearing in $\psi(z)$ is z^l. Then the norm of ψ is

$$||\psi|| \sim \int_{\mathbb{R}^2} dz\wedge d\bar{z}\frac{z^{2l}}{(1+|z|^2)^{2+k}} \sim \int_{\mathbb{R}} dr\,\frac{r^{1+2l}}{(1+r^2)^{2+k}}. \tag{503}$$

Therefore[53]

$$||\psi|| < \infty \implies 2+2l < 2(2+k) \implies l < k+1. \tag{504}$$

Therefore normalizable sections can be written as linear combinations of $1, z, \dots z^k$, giving us $k+1$ different basis sections. This tells us that, for $j = k/2$,

$$\dim\mathcal{H} = k+1 = 2j+1, \tag{505}$$

which of course we could have guessed long ago. Therefore the result of geometrically quantizing sections of a $\mathbb{C}$ line bundle over S^2 with Chern number k is the theory of a quantum spin $k/2$ particle. Note that the number of states in the Hilbert space is one greater than the number we would have expected classically: $\dim\mathcal{H} = \int_{\mathcal{P}}\omega + 1$, although

[52] On S^2, sometimes different coordinates are preferred. Instead of working with $\zeta_N = (z_1/z_2,\,1)$ and $\zeta_N = (1,\,z_2/z_1)$, we choose single-component coordinates $z_N = z, z_S = 1/z$. In this case, the transition function is $z_N = g_{NS}z_S$, with $g_{NS} = z_S^{-2}$. However, working with the ζ_Ns is preferred since although they have redundant components (since one of their components is always 1), their definition works for all $\mathbb{CP}^N$.

[53] We can also get this result without analyzing the integral by recalling that the transition function between the north and south patch means that z^l in the N patch gets mapped to z^{k-l} in the S patch. Now in order to not have a pole in the south patch we need $k - l \geq 0$, giving $l < k+1$ as above.

as expected the exact result agrees with the classical one when $k \sim \hbar^{-1} \to \infty$. The extra state in $\mathcal{H}$ in this case comes from the fact that $\mathcal{P}$ is positively curved: in general, positive phase-space curvature means $\dim \mathcal{H}$ is bigger than the semiclassical result, while negative curvature means $\dim \mathcal{H}$ is less. More on this in a future diary entry.

$3j$ and $6j$ symbols for finite groups

In this diary entry, we will briefly review the procedure developed by Hu and Wu for computing the $3j$ and $6j$-symbols for finite groups. We will illustrate an example by considering the quaternion group $\mathbb{Q}_8$. We also mention a way for determining if a finite group is multiplicity-free.

◇ ◇ ◇

For completeness, let us review some of the requisite math stuff. Given a group G, recall that a *unitary representation* is a linear representation ρ of G on a Hilbert space V such that the image of any group element under the representation $\rho(g)$ is a unitary operator. Now, let

$$N_{ij}^k = \mathrm{Hom}(V_k, V_i \otimes V_j) \tag{506}$$

so that

$$V_i \otimes V_j = \bigoplus_k N_{ij}^k V_k. \tag{507}$$

Then since we have a natural associativity automorphism

$$\alpha : V_i \otimes (V_j \otimes V_k) \to (V_i \otimes V_j) \otimes V_k \tag{508}$$

we can derive the vector space isomorphism using (2):

$$F_{kl}^{ij} : \bigoplus_m N_{kl}^m \otimes N_{mi}^j \to \bigoplus_n N_{kn}^j \otimes N_{li}^n. \tag{509}$$

The components of the above mapping are crucial in string-net models, and are called *$6j$-symbols*. Together with the group's character table, they determine the group up to isomorphism. We will usually impose extra tetrahedral symmetry conditions on the $6j$-symbols, and will write the symmetrized versions as G_{kln}^{ijm}.

This is all well and good, but how do we actually compute the $6j$-symbols? First, we need a definition. Given two representations (ρ_1, V_1), (ρ_2, V_2) of G, an *intertwining operator* is a linear operator $O : V_1 \to V_2$ that commutes with the action of G. A *$3j$-symbol* for a triple of representations $(j_1 j_2 j_3)$ is an intertwining operator

$$C_{j_1 j_2 j_3} : V_{j_1} \otimes V_{j_2} \otimes V_{j_3} \to \mathbb{C} \tag{510}$$

that tells us how the tensor product $j_! \otimes j_2 \otimes j_3$ decomposes into a direct sum of irreducible representations. Throughout, we will choose groups so that $j_1 \otimes j_2 \otimes j_3$ is always *multiplicity* *free* – that is, the trivial representation appears at most once is the decomposition of the tensor product. As usual, we set $\delta_{j_1 j_2 j_3} = 1$ if the trivial representation appears in the decomposition of $j_1 \otimes j_2 \otimes j_3$, and $\delta_{j_1 j_2 j_3} = 0$ otherwise.

Each unitary representation (ρ_j, V_j) is associated with a natural *dual representation* (ρ_{j^*}, V_{j^*}). The dual representation comes with an invertible intertwining operator called the *duality operator*:

$$\begin{aligned} \omega : \mathbb{C} &\to V_j \otimes V_{j^*} \\ 1 &\mapsto \sum_{m,n} \Omega^j_{mn} v_m \otimes v_n \end{aligned} \tag{511}$$

where the matrix Ω satisfies $\Omega^\dagger \Omega = id$ and v_m, v_n are the bases of V_j, V_{j^*}. The duality map is unique, and so we must have $(\Omega^j)^T = \alpha_j \Omega^j$ for some number α_j, called the *Frobenius Schur indicator*. The FS indicator is $1, 0$, or -1 depending on wether the representation j is real, complex or quaternionic.

The $3j$-symbols are defined up to a phase, which in general can be a function of the representations j_1, j_2, j_3. We are free to use this phase to impose additional symmetry on the $3j$-symbols, which will simplify our calculations a bit. One such symmetry we demand is the *cyclic condition*, namely that

$$[C_{j_1 j_2 j_3}]_{m_1 m_2 m_3} = \alpha_{j_3} [C_{j_3 j_1 j_2}]_{m_3 m_1 m_2}, \tag{512}$$

which cuts down on the number of $3j$-symbols that we need to calculate explicitly. Notice that if this is to make any sense, we must also require that $\alpha_{j_1} \alpha_{j_2} \alpha_{j_3} = 1$ for any trio of admissible representations (seen by permuting the $3j$-symbols through one full cycle).

One clever method for determining the $3j$-symbols is by averaging a random matrix over all group elements. Let T be some random nonzero tensor, and G be a finite group. We compute

$$\Omega^j_{mn} = \frac{1}{|G|} \sum_{g \in G} [\rho_j(g)]_{mm'} [\rho_{j^*}(g)]_{nn'} T_{m'n'} \tag{513}$$

where m', n' run from 1 to the dimension of the representation. Here we have assumed G is a finite group. If G is a Lie group, we replace the sum with an integration $\int dg$, where dg is the Harr measure. If the Lie group has dimension d, the Harr measure can be found by constructing a translationally-invariant d-form to perform the integration over.

The $3j$-symbols can then be written as

$$[C_{j_1 j_2 j_3}]_{m_1 m_2 m_3} = \frac{1}{|G|} \sum_{g \in G} [\rho_{j_1}(g)]_{m_1' m_1} [\rho_{j_2}(g)]_{m_2' m_2} [\rho_{j_3}(g)]_{m_3' m_3} T_{m_1' m_2' m_3'}. \tag{514}$$

Additionally, we are free to impose the normalization condition

$$1 = \sum_{m_1 m_2 m_3 n_1 n_2 n_3} [C_{j_1 j_2 j_3}]_{m_1 m_2 m_3} [C_{j_3^* j_2^* j_1^*}]_{n_3 n_2 n_1} \Omega^{j_1}_{m_1 n_1} \Omega^{j_2}_{m_2 n_2} \Omega^{j_3}_{m_3 n_3} \tag{515}$$

To illustrate the algorithm with an example, we will compute the $3j$-symbols and $6j$-symbols for the quaternions $\mathbb{Q}_8$. The quaternions have the advantage of having a relatively small order, having representations that are all self-dual, being multiplicity-free, and having

Table 1: Irreducible representations for the Quaternion group $\mathbb{Q}_8$

g	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4
1	1	1	1	1	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
-1	1	1	1	1	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
i	1	1	-1	-1	$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$
$-i$	1	1	-1	-1	$\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$
j	1	-1	1	-1	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$-j$	1	-1	1	-1	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
k	1	-1	-1	1	$\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$
$-k$	1	-1	-1	1	$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$

slightly non-trivial duality operators. The irreducible representations we use are shown in table 1.

The duality maps for the quaternions are

$$\Omega^0 = \Omega^1 = \Omega^2 = \Omega^3 = 1, \quad \Omega^4 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{516}$$

which tell us that $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 1,\ \alpha_4 = -1$. We compute the non-zero independent $3j$-symbols as

$$C_{000} = C_{011} = C_{022} = C_{033} = C_{123} = 1, \tag{517}$$

$$C_{044} = \begin{pmatrix} 0 & -i/\sqrt{2} \\ i/\sqrt{2} & 0 \end{pmatrix}, \qquad C_{144} = \begin{pmatrix} 0 & -i/\sqrt{2} \\ -i/\sqrt{2} & 0 \end{pmatrix}, \tag{518}$$

$$C_{244} = \begin{pmatrix} -i/\sqrt{2} & 0 \\ 0 & -i/\sqrt{2} \end{pmatrix}, \qquad C_{344} = \begin{pmatrix} i/\sqrt{2} & 0 \\ 0 & -i/\sqrt{2} \end{pmatrix}, \tag{519}$$

where the others can be obtained through the cyclic condition. This data allows us to compute the fusion rules for the theory: $\delta_{j_1 j_2 j_3} = 1$ if the corresponding $3j$-symbol is non-zero, and $\delta_{j_1 j_2 j_3} = 0$ otherwise. Notice that there are no nonzero symbols with one or three ρ_4 representations, since $\alpha_4 = -1$ and we have the constraint $\alpha_i \alpha_j \alpha_k = 1$ at any admissible vertex.

We now turn to the (symmetrized) 6j-symbols. We demand tetrahedral symmetry:

$$G^{ijm}_{kln} = G^{mij}_{nk^*l^*} = G^{klm^*}_{ijn^*} = \alpha_m\alpha_n\overline{G^{i^*j^*m^*}_{l^*k^*n^*}} \tag{520}$$

as well as the pentagon identity (derived from the associativity automorphisms between tensor products of four vector spaces):

$$\sum_n \alpha_n d_n G^{mlq}_{kp^*n} G^{jip}_{mns^*} G^{js^*n}_{lkr^*} = G^{jip}_{q^*kr^*} G^{riq^*}_{mls^*} \tag{521}$$

where $d_n = \dim(j_n)$. Finally, we also impose the *orthogonality condition*

$$\sum_n \alpha_n d_n G^{mlq}_{kp^*n} G^{l^*m^*i^*}_{pk^*n} = \frac{\delta_{iq}}{\alpha_i d_i}\delta_{mlq}\delta_{k^*ip}. \tag{522}$$

After a bit of manipulation, we arrive at

$$\begin{aligned} G^{ijm}_{kln} = \sum_{a_i,b,\ldots,a_n,b_n} & \Omega^i_{a_ib_i}\Omega^j_{a_jb_j}\Omega^k_{a_kb_k}\Omega^l_{a_lb_l}\Omega^m_{a_mb_m}\Omega^n_{a_nb_n}\times \\ & C_{ijm;a_ia_ja_m}C_{lm^*k;a_lb_ma_k}C_{k^*j^*n;b_kb_ja_n}C_{n^*i^*l^*;b_nb_ib_l} \end{aligned} \tag{523}$$

where each of a_i, b_i run over the set $1,\ldots,d_i$. Using our expressions for the 3j-symbols, we compute the 6j-symbols as

$$\begin{aligned} &G^{000}_{000} = 1, G^{000}_{111} = 1, G^{000}_{222} = 1, G^{000}_{333} = 1, G^{000}_{444} = -\frac{i}{\sqrt{2}}, \\ &G^{011}_{011} = 1, G^{011}_{233} = 1, G^{011}_{444} = -\frac{i}{\sqrt{2}}, G^{022}_{022} = 1, G^{022}_{133} = 1, \\ &G^{022}_{444} = -\frac{i}{\sqrt{2}}, G^{033}_{033} = 1, G^{033}_{444} = -\frac{i}{\sqrt{2}}, G^{044}_{044} = -\frac{1}{2}, G^{044}_{144} = -\frac{1}{2}, \\ &G^{044}_{244} = -\frac{1}{2}, G^{044}_{344} = -\frac{1}{2}, G^{123}_{123} = 1, G^{123}_{444} = -\frac{1}{\sqrt{2}}, G^{144}_{144} = -\frac{1}{2}, \\ &G^{144}_{244} = \frac{1}{2}, G^{144}_{344} = \frac{1}{2}, G^{244}_{244} = -\frac{1}{2}, G^{244}_{344} = \frac{1}{2}, G^{344}_{344} = -\frac{1}{2}. \end{aligned} \tag{524}$$

Of course, these aren't *all* the non-zero 6j-symbols, but the rest can be obtained using tetrahedral symmetry.

How to determine if a group is multiplicity-free

Given that we assumed a multiplicity-free set of irreducible representations in order to carry out the computations of the 6j-symbols, it would be nice to obtain a quick method for determining which (finite) groups are multiplicity free. This actually turns out be pretty easy.

Let (ρ_1, V_1), (ρ_2, V_2) be two representations of a group G. Let ϱ be the representation defined by

$$\varrho : G \to GL(\hom(V_1, V_2)). \tag{525}$$

Further, denote $\mathrm{Hom}_G(V_1, V_2)$ as all the functions that are invariant under ϱ (i.e. those that satisfy $\varrho(g) \circ f = \rho_2 \circ f \circ \rho_1^{-1}$). Motivated by the procedure outlined above for computing the duality maps, we define the averaging operator

$$A_\rho = \frac{1}{|G|} \sum_{g \in G} \rho(g). \tag{526}$$

Furthermore, let V^G be the space of G-invariant representations in V (I think this is standard notation). Put another way, we can say that $V^G \cong \mathrm{Hom}_G(0, V)$, where 0 is the trivial representation. This means that in order to determine wether or not a set of representations is multiplicity-free, we just need to compute the number $\dim V^G$.

It's easy to check that $A_\rho(V) = V^G$ (note that A_ρ is idempotent), which means that

$$V = \ker A_\rho \oplus \mathrm{im} A_\rho = \ker A_\rho \oplus V^G, \tag{527}$$

which means that we can write

$$\dim V^G = \mathrm{Tr}(A_\rho) = \mathrm{Tr}\left(\frac{1}{|G|} \sum_{g \in G} \rho(g) \right) \tag{528}$$

Recalling the definition of the character of a representation, we get

$$\dim V^G = \frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) \tag{529}$$

Now we apply this way of thinking to the representation ϱ defined earlier, and obtain

$$\dim \mathrm{Hom}_G(V_1, V_2) = \frac{1}{|G|} \sum_{g \in G} \chi_\varrho(g). \tag{530}$$

If we go back and look at the definition of ϱ, we see that is formulated in terms of inverses of representations. This suggests that we make use of the observation that (V_1^* is the dual space of V_1)

$$V_1^* \otimes V_2 \cong \mathrm{Hom}(V_1, V_2). \tag{531}$$

It's then easy to check (and natural to guess) that $\chi_\rho(g^{-1}) = \chi_{\rho^*}(g)$. In particular, this means that

$$\chi_\varrho(g) = \chi_{\rho_1}(g^{-1}) \chi_{\rho_2}(g), \quad \forall g \in G \tag{532}$$

which finally gives us

$$\dim \mathrm{Hom}_G(V_1, V_2) = \frac{1}{|G|} \sum_{g \in G} \chi_{\rho_1}(g^{-1}) \chi_{\rho_2}(g). \tag{533}$$

Thus if want to check if a given set of representations is multiplicity-free, we just need to obtain the character values for each representation, which is usually very easy.

Quantum double and modular matrices for finite groups

In this diary entry, we will overview the construction of the quantum double for a string-net theory based on a given finite group. We'll outline how to use the quantum double to compute the S and T-matrices, and then will explicitly compute some examples for a few simple groups.

Representation theory of the quantum double

The algebraic structure of any discrete gauge theory on a group G is obtained from Drinfeld's quantum double construction. Let $\mathcal{F}(G)$ be the commutative algebra of functions on G, and $\mathbb{C}[G]$ be the cocommutative group algebra of G over the complex numbers. As a vector space, the quantum double is constructed by

$$D(G) = \mathcal{F}(G) \otimes \mathbb{C}[G] \tag{534}$$

The algebra induced by the quantum double describes how the particles in the theory fuse and braid, with its elements consisting of flux-charge pairs. The different particles in the given gauge theory correspond to the set of irreducible representations of $D(G)$, up to conjugacy.

First, we need to fix the notation. Let the conjugacy classes of G be

$$C^A = \{h_1^A \ldots, h_k^A\} \tag{535}$$

and let Z^A be the centralizer of the representative element h_1^A. Although we have some freedom as far as choosing representatives from each equivalence class of G/Z^A goes, the actual choice of representative will not affect the resulting quantum double (up to unitary equivalence), and so we choose the first element in each conjugacy class as a representative for consistency. Finally, we let the set $\{x_1^A, \ldots, x_k^A\}$ be representatives for the quotient G/Z^A, so that $h_i^A = x_i^A h_1^A (x_i^A)^{-1}$.

Each quantum double can be labeled by a set (C^A, α), where C^A is a conjugacy class and α an irreducible representation of the corresponding centralizer Z^A. Since the element h_1^A in any conjugacy class commutes with everything in Z^A by definition, Schur's lemma tells us that its associated irreducible representation must be a constant times the identity matrix:

$$\alpha(h_1^A) = \exp(2\pi i s_{\mathcal{A}})\mathbf{1} \tag{536}$$

where we've associated each quantum double label $\mathcal{A} = (C^A, \alpha)$ with a corresponding spin $s_{\mathcal{A}}$. We usually compile the spins of all the particles in the theory into one matrix, called the T-matrix:

$$T_{\mathcal{AB}} = \delta_{\mathcal{A},\mathcal{B}} \exp(2\pi i s_{\mathcal{A}}) \tag{537}$$

which can alternatively be written as

$$T_{\mathcal{AB}} = \delta_{\mathcal{A},\mathcal{B}} \frac{\mathrm{tr}(\alpha(h_1^A))}{d_\alpha} \tag{538}$$

where d_α is as usual the dimension of the centralizer charge representation of the particle associated with the quantum double label $\mathcal{A}$.

Another important topological observable is the S-matrix. The elements of the S-matrix are proportional to the quantum mechanical amplitude for processes which involve the braiding of two particles. There are many equivalent ways to think about the S-matrix. The S-matrix can be thought of as the matrix which exhibits the diagonalization of the fusion algebra (which can always be diagonalized, as it is commutative and associative). Alternatively, we can think of the S-matrix as the analog of a character table for the quantum double.

There are several ways of obtaining the S-matrix. One is to compute the half-braiding tensors in the theory, and then construct the S-matrix by summing over a product of two half-braiding tensors. This is usually fairly involved however, and for finite groups we are often better off following Verlinde by writing

$$S_{\mathcal{A}\mathcal{B}} = \frac{1}{|G|} \sum_{\substack{h_i^A \in C^A, h_j^B \in C^B \\ h_i^A h_j^B = h_j^B h_i^A}} \mathrm{tr}(\alpha((x_i^A)^{-1} h_j^B x_i^A))^* \mathrm{tr}(\beta((x_j^B)^{-1} h_i^A x_j^B))^* \tag{539}$$

where $\mathcal{A} = (C^A, \alpha)$ and $\mathcal{B} = (C^B, \beta)$ are quantum double labels.

Once we have the S-matrix, we can recover the multiplicity $N_{\mathcal{I}\mathcal{J}}^{\mathcal{K}}$ of the fusion algebra through

$$N_{\mathcal{I}\mathcal{J}}^{\mathcal{K}} = \sum_{\mathcal{L}} \frac{S_{\mathcal{I}\mathcal{L}} S_{\mathcal{J}\mathcal{L}} S^*_{\mathcal{K}\mathcal{L}}}{S_{0\mathcal{L}}} \tag{540}$$

where $\mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{L}$ are quantum double labels, and the 0 index in the denominator corresponds to the trivial quantum double $(1, [1])$. Importantly, we see that this approach to obtaining the topological observables for the theory does not depend on the assumption that the fusion rules are multiplicity-free. It's also a bit less work, depending on the group – we never even had to calculate a $3j$ or $6j$-symbol!

One thing worth mentioning here is that the S and T matrices generate a unitary representation of the modular group $SL(2, \mathbb{Z})$. This leads us to define the charge conjugation operator C as

$$C = (ST)^3 = S^2 \tag{541}$$

so that we have the following set of relations:

$$S^* = CS = S^{-1}, \quad S^T = S, \quad T^* = T^{-1}, \quad T^T = T. \tag{542}$$

C can be thought of as a charge-conjugation map from a particle to its anti-particle pair by $C : (C^A, \alpha) \mapsto (C^{\bar{A}}, \bar{\alpha})$. It's easy to check that $[C, T] = 0$, and so particles and their anti-particle partners carry the same spin.

Examples: $G = \mathbb{Z}_2$

This is all rather abstract, and so we now turn to the discussion of a few concrete examples. The simplest possible example is $\mathbb{Z}_2$ gauge theory, which is just about the only example where it's possible to compute the S-matrix in your head. For this example we'll think in the multiplicative picture with $\mathbb{Z}_2 = \{1, -1\}$.

Since $\mathbb{Z}_2$ is abelian, each element is its own conjugacy class. Of course, the centralizer of each conjugacy class is $\mathbb{Z}_2$, and so we only have two irreducible representations. One is the trivial representation $[1] : 1 \mapsto 1, \ -1 \mapsto 1$, while the other is $[-1] : 1 \mapsto 1, \ -1 \mapsto -1$.

We write each quantum double pair as (conjugacy class, irreducible representation). Working in the basis $\big((1,[1]),(1,[-1]),(1,[-1]),(-1,[-1])\big)^T$ we see that the T-matrix is

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{543}$$

The S-matrix is similarly easy to compute, since when we use equation 539 we don't have to carry out any actual summation. We get

$$S = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \tag{544}$$

How do we interpret this physically? We have four types of particles in the theory. The first type $(1,[1])$ shows no interesting braiding behavior, and creates a trivial quasiparticle. Making the conjugacy class $\to$ flux and irreducible representation $\to$ charge connection, we see that $(1,[-1])$ creates electric charges, $(1,[-1])$ creates magnetic fluxes, and $(-1,[-1])$ creates an electric/magnetic bound state.

Examples: $G = S_3$

Now we move on to the slightly more sophisticated example of S_3. We'll write the elements of S_3 as $\{1,2,3,4,5,6\}$ for simplicity. There are 3 conjugacy classes: $C^1 = \{1\}$ (the identity), $C^2 = \{2,3\}$ (the even permutations), and $C^3 = \{4,5,6\}$ (the others). The corresponding centralizers are $Z_1 = S_3$, $Z_2 = \{1,2,3\} \cong \mathbb{Z}_3$, and $Z_3 = \{1,4\} \cong \mathbb{Z}_2$.

$\mathbb{Z}_2$ has two irreducible representations: $[1]$ and $[-1]$ (as in our last example). S_3 has three: $[1]$ (the trivial one), $[s]$ (the sign representation), and $[2]$ (the standard representation, 2 is for 2-dimensional). $\mathbb{Z}_3$ also has three irreducible representations. Let $\omega = \exp(2\pi i/3)$. The three irreducible representations are the trivial representation, $[\omega]$, which sends $1 \mapsto 1, 2 \mapsto \omega, 3 \mapsto \bar{\omega}$, and $[\bar{\omega}]$, which is the conjugate of $[\omega]$.

All in all, we have 8 quantum double types. For clarity, we present the T-matrix in table 2.

Table 2: T-matrix for S_3

QD type $\mathcal{J}$:	$(C^1,[1])$	$(C^1,[s])$	$(C^1,[2])$	$(C^2,[1])$	$(C^2,[\omega])$	$(C^2,[\bar{\omega}])$	$(C^3,[1])$	$(C^3,[-1]])$
$T_{\mathcal{J}\mathcal{J}}$:	1	1	1	1	ω	$\bar{\omega}$	1	-1

Working in the basis of table 2, we compute the corresponding S-matrix as (we only need to give the upper triangular part by symmetry)

$$S = \frac{1}{6}\begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 \\ & 1 & 2 & 2 & 2 & 2 & -3 & -3 \\ & & 4 & -2 & -2 & -2 & 0 & 0 \\ & & & 4 & -2 & -2 & 0 & 0 \\ & & & & 4 & -2 & 0 & 0 \\ & & & & & 4 & 0 & 0 \\ & & & & & & 3 & -3 \\ & & & & & & & 3 \end{pmatrix} \tag{545}$$

Examples: $G = \mathbb{Q}_8$

As a finial example, we turn to the quaternion group $\mathbb{Q}_8$. The five conjugacy classes and their associated centralizers are provided in table 3. $\mathbb{Q}_8$ has five irreducible representations, while $\langle i\rangle, \langle j\rangle, \langle k\rangle$ each have four. Since the centralizers for C^1, C^2, and C^3 are all structurally the same, we will denote their collective centralizer simply by Z. Finally, since the actual representations are not too important, we just list their associated characters to save space.

Table 3: Magnetic flux types for $\mathbb{Q}_8$

Conjugacy class	Centralizer
$e = \{1\}$	$\mathbb{Q}_8$
$\bar{e} = \{-1\}$	$\mathbb{Q}_8$
$C^1 = \{i, -i\}$	$\langle i\rangle$
$C^2 = \{j, -j\}$	$\langle j\rangle$
$C^3 = \{k, -k\}$	$\langle k\rangle$

Table 4: Character table of the irreducible representations of $\mathbb{Q}_8$

$\mathbb{Q}_8$	e	$\bar{e}$	C^1	C^2	C^3
α^0	1	1	1	1	1
α^1	1	1	1	-1	-1
α^2	1	1	-1	1	-1
α^3	1	1	-1	-1	1
α^4	2	-2	0	0	0

To keep track of all the quantum double types in the theory, it's helpful to introduce symbolic labels for the different particles (here $x \in \{1, 2, 3\}$):

$$\begin{aligned} &1 = (e, \alpha^0), \qquad \bar{1} = (\bar{e}, \alpha^0), \qquad \Sigma_x = (e, \alpha^x), \qquad \widetilde{\Sigma}_x = (\bar{e}, \alpha^x), \qquad \Upsilon = (e, \alpha^4), \\ &\widetilde{\Upsilon} = (\bar{e}, \alpha^4), \qquad \Psi_x = (C^x, \beta^0), \qquad \widetilde{\Psi}_x = (C^x, \beta^2), \qquad \Phi_x = (C^x, \beta^1), \qquad \widetilde{\Phi}_x = (C^x, \beta^3) \end{aligned} \tag{546}$$

Table 5: Character table of the irreducible representations of Z

Z	1	g	-1	$-g$
β^0	1	1	1	1
β^1	1	i	-1	$-i$
β^2	1	-1	1	-1
β^3	1	$-i$	-1	i

We have five conjugacy classes all together, giving four nontrivial magnetic flux sectors. The magnetic flux sectors come in the form of a singlet flux $\bar{e}$ and a triplet of doublet fluxes C^1, C^2, C^3. We also see the existence of four nontrivial pure charges. Again, they come in a $1-3$ split: one doublet charge Υ and three singlet charges labeled by $\Sigma_1, \Sigma_2, \Sigma_3$.

Looking at the centralizers associated with each flux sector allows us to compute the full number of dyons in the theory. All in all, we get 22 particle types. This makes the T and S-matrices gigantic, but we can simplify things a bit if we avoid explicitly writing out the triplet of particles represented by Σ_x, Ψ_x, and Φ_x. The statistics of the particles are

$$\exp(2\pi i s) = \begin{cases} 1 & \text{for } 1, \bar{1}, \Sigma_x, \widetilde{\Sigma}_x, \Upsilon, \Psi_x \\ -1 & \text{for } \widetilde{\Upsilon}, \widetilde{\Psi}_x \\ i & \text{for } \Phi_x \\ -i & \text{for } \widetilde{\Phi}_x \end{cases} \tag{547}$$

We present the S-matrix as a table for clarity, omitting an overall prefactor of 1/8. We also define $\epsilon_{xy} = 2\delta_{x,y} - 1$ to simplify the notation.

Table 6: S-matrix for $\mathbb{Q}_8$

$S \times 8$	1	$\bar{1}$	Σ_x	$\widetilde{\Sigma}_x$	Υ	$\widetilde{\Upsilon}$	Ψ_x	$\widetilde{\Psi}_x$	Φ_x	$\widetilde{\Phi}_x$
1	1	1	1	1	2	2	2	2	2	2
$\bar{1}$		1	1	1	-2	-2	2	2	-2	-2
Σ_y			1	1	2	2	$2\epsilon_{xy}$	$2\epsilon_{xy}$	$2\epsilon_{xy}$	$2\epsilon_{xy}$
$\widetilde{\Sigma}_y$				1	-2	-2	$-2\epsilon_{xy}$	$-2\epsilon_{xy}$	$2\epsilon_{xy}$	$2\epsilon_{xy}$
Υ					4	-4	0	0	0	0
$\widetilde{\Upsilon}$						4	0	0	0	0
Ψ_y							$4\delta_{xy}$	$-4\delta_{xy}$	0	0
$\widetilde{\Psi}_y$								$4\delta_{xy}$	0	0
Φ_y									$4\delta_{xy}$	$-4\delta_{xy}$
$\widetilde{\Phi}_y$										$4\delta_{xy}$

The S-matrix turns out to be equal to its inverse, implying that the charge conjugation operator C acts as the identity. This means that all the particles in $\mathbb{Q}_8$ gauge theory are their own duals.

Finally, we note that the set of eight particles $\{1, \bar{1}, \Sigma_x, \bar{\Sigma}_x\}$ in this theory form a *Lagrangian subgroup*. A Lagrangian subgroup is a group of bosons which have trivial mutual statistics with each other, with every other particle outside the group possessing non-trivial mutual statistics with at least one particle in the group. This data is useful to know because at least In the abelian case, it is known (Lin and Levin 2014) that a topological phase can be realized by a string-net model if and only if it possesses at least one Lagrangian subgroup.

Math needed for thinking about fermionic topological phases from the diagrammatic point of view

Today's diary entry is a compendium of remarks on various things needed to start doing diagrammatic manipulations for fermionic topological phases.

Background math

Supervector spaces and Clifford algebras

In order to think about fermionic topological phases, it's helpful to think about the objects in tensor categories in terms of vector spaces. Since we're working with unitary TQFTs we will be working over $\mathbb{C}$, and so the simple objects in our categories can be regarded as simple $\mathbb{C}$-algebras.

When our theories arise from fermionic degrees of freedom, the main thing we need to do is to keep track of fermion parity. We can do this by splitting all vector spaces in the theory up into a direct sum of their fermion-parity even and fermion-parity odd sectors, writing $V = V^0 \oplus V^1$. This decomposition turns vector spaces into *supervector* spaces, which are just $\mathbb{Z}_2$-graded versions of regular vector spaces. Throughout we will use the operator P as the fermion parity operator, defined by $Pv = v$ if $v \in V^0$ (even fermion parity) and $Pv = -v$ if $v \in V^1$ (odd fermion parity).

When we take the tensor product of two supervector spaces, we must use a super version of the tensor product which respects the $\mathbb{Z}_2$ grading. Specifically, the super version of the tensor product of two supervector spaces (which we will refer to as the tensor product, and will just denote by $\otimes$) works as follows:

$$(V \otimes W)^0 = V^0 \otimes W^0 \oplus V^1 \otimes W^1, \quad (V \otimes W)^1 = V^0 \otimes W^1 \oplus V^1 \otimes W^0. \tag{548}$$

We will use absolute value bars to denote the fermion parity of vectors in supervector spaces. That is, for $v \in V$, we write $|v| = 0$ if $Pv = v$ (i.e. if $v \in V^0$) and $|v| = 1$ if $Pv = -v$ (i.e. if $v \in V^1$). Note that since we want the $\mathbb{Z}_2$ grading to be well-defined on supervector spaces, we are forbidden from adding even fermion parity vectors with odd fermion parity vectors.

Since we're working with a tensor product that respects the fermion parity structure of supervector spaces, it must account for the factors of (-1) that appear when the relative order of two fermionic vectors in a tensor product is switched. This means that the tensor product of two vectors is supercommutative, in the sense that the wedge product is supercommutative:

$$v \otimes w = (-1)^{|v||w|} w \otimes v. \tag{549}$$

This means that sVec (the category of supervector spaces, containing just the vacuum and a fermion) naturally comes equipped with the structure of a braided category, with the supercommutative nature of the tensor product controlling the -1 braiding of the fermion with itself.

If a supervector space V has $\dim V^0 = a$ and $\dim V^1 = b$, we say that V has superdimension $a|b$. We will use the notation $\mathbb{C}^{a|b}$ to denote the supervector space of superdimension $a|b$, so that $\mathbb{C}^{a|b}$ is an $(a+b)$-dimensional vector space with a fermion-parity even generators and b fermion-parity odd generators.

When we tensor $\mathbb{C}^{a|b}$ with $\mathbb{C}^{c|d}$, the graded dimensions behave in the same way that supervector spaces behave when you tensor them together. That is, they behave just like you would expect:

$$\mathbb{C}^{a|b} \otimes \mathbb{C}^{c|d} \cong \mathbb{C}^{ac+bd|ad+bc}. \tag{550}$$

Notice in particular that $\mathbb{C}^{a|b} \otimes \mathbb{C}^{1|0} \cong \mathbb{C}^{a|b}$ and $\mathbb{C}^{a|b} \otimes \mathbb{C}^{0|1} \cong \mathbb{C}^{b|a}$. That is, tensoring with $\mathbb{C}^{1|0} = \mathbb{C}$ does nothing (as it should), and tensoring with $\mathbb{C}^{0|1}$ is equivalent to flipping fermion parity.

The supervector space $\mathbb{C}^{1|1}$ will turn out to play an important role in what follows. This is because it is invariant under fermion-parity flips: $\mathbb{C}^{1|1} \otimes \mathbb{C}^{0|1} \cong \mathbb{C}^{1|1}$. It is also an example of a Clifford algebra, namely the Clifford algebra Cl_1. Clifford algebras will be important for us when we consider theories with Majoranas, and so we will quickly review their definition and basic properties.

The complex Clifford algebras Cl_n are the supervector spaces generated by the number 1 and n *parity-odd* generators $\gamma_1, \dots, \gamma_n$ which satisfy the relations

$$\{\gamma_i, \gamma_j\} = 2Q_{ij}, \tag{551}$$

where $\{\gamma_i, \gamma_j\} = \gamma_i \otimes \gamma_j + \gamma_j \otimes \gamma_i$ and Q_{ij} is some quadratic form. For real Clifford algebras Q_{ij} can generically be written as a diagonal matrix with ± 1s on the diagonal, although our working over $\mathbb{C}$ allows us to set $Q_{ij} = \delta_{ij}$ without loss of generality.

The simplest (other than $Cl_0 = \mathbb{C}$) Clifford algebra is $Cl_1 = \langle 1, \gamma \rangle$, with $\gamma^2 = 1$ and $P\gamma = -\gamma$. Since by definition Cl_1 has one even generator (1) and one odd generator (γ), we have $Cl_1 \cong \mathbb{C}^{1|1}$. In terms of representations, we can choose a representation ρ such that $\rho(1) = \mathbf{1}_{2\times 2}$ and $\rho(\gamma) = \sigma^x$, which is consistent since $\{\sigma^x, \sigma^x\} = 2\mathbf{1}_{2\times 2}$ and $\sigma^x : \mathbb{C}^{1|0} \mapsto \mathbb{C}^{0|1}, \mathbb{C}^{0|1} \mapsto \mathbb{C}^{1|0}$, *i.e.* σ^x is odd. We will see that Cl_1 is the prototypical "Majorana" supervector space that will appear later on.

We will need a few miscellaeous facts about Clifford algebras, which we will list off here. First, tensoring two Clifford algebras gives a larger Clifford algebra whose associated quadratic form is the direct sum of the quadractic forms of the smaller algebras, meaning that the larger Clifford algebras can be built from Cl_1 as $Cl_n \cong Cl_1^{\otimes n}$. Secondly, $Cl_2 \cong \mathrm{End}(\mathbb{C}^{1|1})$, and since the endomorphism rings of matrix algebras always have trivial modules, Cl_2 and $\mathbb{C}$ are Morita equivalent (written $Cl_2 \cong_M \mathbb{C}$), meaning that their simple modules are the same. This will be of use later when we look at constructing quasiparticles and their fusion rules. Since $Cl_n \cong Cl_1^{\otimes n}$, this implies the $Cl_n \cong_M Cl_{n+1}$ of Bott periodicity fame.

Modules and Morita equivalence

A module over an algebra is a vector space whose scalars are drawn from the algebra. That is, it's a way of giving an algebra an action on a vector space. So modules are the algebra analogue of representations: representations construct a way for groups to act on vector spaces, and modules do the same thing for algebras.

A mathematical concept that will be relevant is the notion of *Morita equivalence*. Roughly, Morita equivalence is a way of establishing when two algebras "have the same modules", or when their "representation theory is the same". The technical definition is that two algebras A and B are Morita equivalent (written $A \cong_M B$) when the cateogries of their left modules $\mathrm{Mod}^L(A)$ and $\mathrm{Mod}^L(B)$ are equivalent, although we won't use this definition much.

The reason why the notion of Morita equivalence is useful for us is because quasiparticles in topological phases are identified with simple modules of an algebra **Tube** called the *tube algebra*, which we'll talk about in more detail later. Usually figuring out the algebra structure of **Tube** is straightforward, although computing its simple modules (aka finding an idempotent decomposition of **Tube**) can be tedious. Since Morita equivalent algebras have the same modules, we can often replace a complicated tube algebra or a sub-algebra of a tube algebra with a much simpler but Morita equivalent algebra, which can greatly facilitate the determination of its simple modules.

The canonical example of Morita equivalent algebras are the matrix algebras. It turns out that we actually have $\mathbb{C}(n) \cong_M \mathbb{C}$ for all n (where $\mathbb{C}(n)$ are complex $n \times n$ matrices). This can be seen by using the following proposition:

Let A, B be two algebras, and $\mathcal{E}$ be an $A - B$ bimodule, that is, let $\mathcal{E}$ be a left A-module and a right B-module. Suppose $\mathcal{E}$ is such that

$$\mathcal{E}^* \otimes_A \mathcal{E} \cong B, \qquad \mathcal{E} \otimes_B \mathcal{E}^* \cong A. \tag{552}$$

Then A and B are Morita equivalent.

We will do the proof by setting up the equivalence between the modules of A and B explicitly. Suppose M is a module over A, and N a module over B. Construct the functors $F : \mathrm{Mod}(A) \to \mathrm{Mod}(B)$, $G : \mathrm{Mod}(B) \to \mathrm{Mod}(A)$ as

$$F : M \mapsto \mathrm{Hom}_A(\mathcal{E}, M), \qquad G : N \mapsto \mathcal{E} \otimes_B N. \tag{553}$$

This associates M with a B-module and N with an A-module. We claim that this is an isomorphism, namely that F and G are inverses of one another. If this is true, we must have

$$M \cong \mathcal{E}\otimes_B \mathrm{Hom}_A(\mathcal{E}, M) \cong (\mathcal{E}\otimes_B\mathcal{E}^*)\otimes_A M, \quad N \cong \mathrm{Hom}_A(\mathcal{E}, \mathcal{E}\otimes_B N) \cong (\mathcal{E}^*\otimes_A\mathcal{E})\otimes_B N, \tag{554}$$

but this holds precisely due to our assumptions on $\mathcal{E}$.

The Morita equivalence between $\mathbb{C}(n)$ and $\mathbb{C}$ then follows from setting the bimodule $\mathcal{E}$ to be $\mathcal{E} = \mathbb{C}^n$ in the above construction.

As a useful nonexample, $\mathbb{C}^n \not\cong_M \mathbb{C}^m$ for any $n \neq m$. This can be seen by realizing that the centers of two Morita equivalent algebras must always be the same. This is true for the

$\mathbb{C}(n) \cong_M \mathbb{C}$ example considered above, but of course the centers of $\mathbb{C}^n$ and $\mathbb{C}^m$ are different if $n \neq m$, and so indeed $\mathbb{C}^n \ncong_M \mathbb{C}^m$.

One useful fact that we will exploit in calculations is that two algebras A and B are always Morita equivalent if there exists a supervector space V such that

$$A \cong B \otimes_{\mathbb{C}} \mathrm{End}(V). \tag{555}$$

For a useful example of this, we turn to the Clifford algebras. We will use the fact that

$$Cl_2 \cong \mathrm{End}(Cl_1) \cong \mathrm{End}(\mathbb{C}^{1|1}), \tag{556}$$

which can be seen just by realizing that $\mathrm{End}(Cl_1) \cong Cl_1 \otimes Cl_1 \cong Cl_2$. This means that

$$Cl_{n+2} \cong_M Cl_n \tag{557}$$

for all n, and so Cl_{n+2} and Cl_n always have the same modules (this is Bott periodicity!). In particular, we see that Cl_2 has the same modules as $\mathbb{C}$! This will be very helpful later on when we look at condensing fermions in the Ising theory. There, we will run into subalgebras of **Tube** that are isomorphic to $\mathrm{End}(\mathbb{C}^{1|1})$. By what we've just seen these subalgebras must have the same modules as $\mathbb{C}$, which is much easier to work with than Cl_2. Since the modules of the **Tube** algebra (and its subalgebras) determine the quasiparticles in the theory, which see in particular that subalgebras of **Tube** given by $\mathrm{End}(\mathbb{C}^{1|1})$ must give rise to only a single quasiparticle, since we are free to replace them by the trivial algebra $\mathbb{C}$.

Superfusion categories

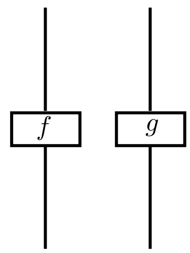
When we go from fusion categories to superfusion categories, the simple objects become associated with supervector spaces, and understanding what happens to them is fairly straightforward. However, the fusion spaces also become superspaces, which for us is very important. In particular, this means that all the Hom spaces (*a.k.a* fusion spaces) carry a $\mathbb{Z}_2$-grading that keeps track of their fermion parity. In particular, $\mathrm{Hom}_{\mathcal{C}}(X,Y)^0$ contains all the morphisms between X and Y that are fermion-parity even, and $\mathrm{Hom}_{\mathcal{C}}(X,Y)^1$ contains all the morphisms that are fermion-parity odd. So, we can think of $\mathrm{Hom}(X,Y)^0$ as being a regular "bosonic" fusion space, while $\mathrm{Hom}(X,Y)^1$ is an odd fusion space in which the fermion parity of the fusion products changes sign. Because of this, we can think of fusion spaces with odd fermion parity as localizing a fermion that lives on the fusion vertex. Especially important for us will be the case when our theories come with objects X such that $\mathrm{Hom}(X,X)^1$ is nontrivial. These objects will be "Majorana" in some sense, and we will discuss them in more detail later. For any morphism $f : X \to Y$, we will write the parity of f as $|f| = 0$ if f is even (preserves fermion parity) and $|f| = 1$ if f is odd (reverses fermion parity).

Without loss of generality, we can consider fusion diagrams built from a tensor product of fusion spaces of the form $\mathrm{Hom}(X \otimes Y, Z)$. Forming a fusion graph requires choosing basis vectors for all these fusion spaces. Following convention, we will write $s_Z^{XY}(\alpha)$ to denote the parity of the basis vector for the fusion space $\mathrm{Hom}(X \otimes Y, Z)$, where $1 \leq \alpha \leq N_Z^{XY}$. For notational simplicity we will initally assume $N_Z^{XY} \leq 1$, so that all the fusion spaces are one-dimensional and have a unique fermion parity (although some of the Fibonacci-like theories we will want to consider later won't satisfy this restriction).

The grading of morphisms is important to keep track of, since morphisms satisfy their own supercommutativity law, called the *superexchange law*, which states that for any four morphisms $f, g, h, k \in \mathcal{C}$, we have

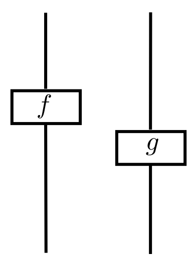
$$(f \otimes g) \circ (h \otimes k) = (-1)^{|g||h|}(f \circ g) \otimes (g \circ k). \tag{558}$$

Let's run through how to see this. First of all, recall that in normal fusion categories, the tensor product of two morphisms is the same as horizontal superposition:

$f \otimes g =$ 

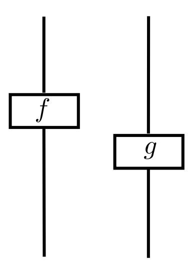
(559)

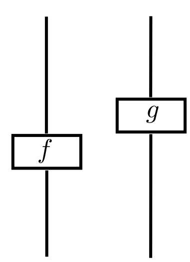
In *superfusion* categories, we need to be more careful, since morphisms (like the guys in the $\mathrm{Hom}(X \otimes Y, Z)$ fusion spaces) can carry nontrivial fermion parity. This means that if we place two odd morphisms side-by-side, the relative positions of the fermions they harbor is ambiguous. This ambiguity is actually really easy to fix diagrammatically: we just write the tensor product of two morphisms in a displaced way, where the first morphism in the tensor product is displaced above the second:

$f \otimes g =$ 

(560)

Moving two morphisms past each other vertically may then result in a minus sign, since if both morphisms have odd fermion parity, moving them past each other is like exchanging two fermions. That is,



$= (-1)^{|f||g|}$


(561)

With this property, it becomes easy to verify the superexchange law.

Finally, let's quickly mention the F-moves (we'll come back to them in more detail later). Recall that we can write them as the basis changes

$$[F^{ijk}_l] : \bigoplus_m \mathrm{Hom}(i \otimes j, m) \otimes \mathrm{Hom}(m \otimes k, l) \cong \bigoplus_n \mathrm{Hom}(i \otimes n, l) \otimes \mathrm{Hom}(j \otimes k, n). \tag{562}$$

The F-moves shouldn't change the fermion parity of the fusion graph, and so must be *even* morphisms. This means that the fermion parity on both sides of the above isomorphism must be the same. To quantify this condition, the even-ness of F means that

$$s_m^{ij} + s_l^{mk} = s_n^{jk} + s_l^{in}, \tag{563}$$

where s_m^{ij} is the parity of the fusion space $\mathrm{Hom}(i \otimes j, m)$ as before. In particular, if $\mathcal{C} = \mathbf{Vec}_G$ for some finite Abelian G, this is equivalent to the 2-cocycle condition, and so we see that such theories are (partially) classified by a choice of cohomology class $s \in H^2(BG, \mathbb{Z}_2)$.

Categorical caveats

A number of formulae that we're used to using when working with fusion categories fail to hold upon passing to superfusion categories, while others go through unchanged.

The most important thing that needs to be modified is the way we write resolutions of the identity. Consider the usual "coproduct" isomorphism

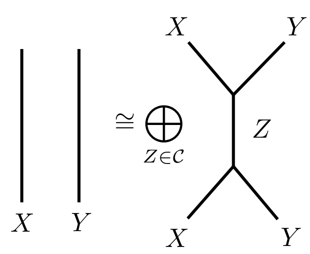
$$\mathrm{Hom}(X, Y) \cong \bigoplus_{Z \in \mathcal{C}} \mathrm{Hom}(X, Z) \otimes \mathrm{Hom}(Z, Y). \tag{564}$$

This amounts to stitching X and Y together by summing over all the simple objects that can connect X to Y.

In string-net diagrams, we usually see (564) used to write the resolution of the identity

$$\mathrm{Hom}(X \otimes Y, X \otimes Y) \cong \bigoplus_{Z \in \mathcal{C}} \mathrm{Hom}(X \otimes Y, Z) \otimes \mathrm{Hom}(Z, X \otimes Y), \tag{565}$$

where X, Y, and Z are all simple objects. Diagrammatically, this looks like



(566)

However, these isomorphisms do not hold as written in more general settings, namely when the simple objects fail to all possess trivial endomorphism algebras. In this scenario, the RHS of (564) will be much bigger than the left, since the "internal" degrees of freedom Z may contribute nontrivially to the size of the RHS. If not all the objects have trivial endomorphism algebras, the big direct sum on the RHS of (564) needs to be replaced with a colimit, although we won't need to go into any detail about what this means mathematically.

The simple way to fix (564) involves changing the type of tensor product we use. Until now, we've tacitly been assuming that all of our tensor products are secretly $\otimes_{\mathbb{C}}$, tensor products over $\mathbb{C}$. This works only when we're working in "regular" categories, where the Hilbert spaces associated with worldlines are always $\mathbb{C}$. However, we run into problems if we have objects with $\mathrm{End}(Z) \not\cong \mathbb{C}$ and continue to use $\otimes_{\mathbb{C}}$. If $\mathrm{End}(Z)$ is bigger than $\mathbb{C}$, then the RHS of (564) is bigger than the LHS, since the internal Z leg of the fusion diagram contributes a larger space to the direct sum. This is unphysical though, since the internal worldlines of freedom Z shouldn't carry any more information than is carried by the incoming and outgoing worldlines (since if this were the case, the sizes of the Hilbert spaces of fusion diagrams would blow up).

We can fix this issue by "modding out" by the Hilbert spaces of the internal worldlines. This can be done by treating $\mathrm{End}(Z)$ as the tensor unit when we tensor the two fusion spaces in (564) together. The correct formula is then

$$\mathrm{Hom}(X,Y) \cong \bigoplus_{Z\in\mathcal{C}} \mathrm{Hom}(X,Z) \otimes_{\mathrm{End}(Z)} \mathrm{Hom}(Z,Y). \tag{567}$$

As a corollary, this means that the superfusion F-symbols should be written as

$$[F_l^{ijk}] : \bigoplus_m \mathrm{Hom}(i\otimes j, m) \otimes_{\mathrm{End}(m)} \mathrm{Hom}(m\otimes k, l) \cong \bigoplus_n \mathrm{Hom}(i\otimes n, l) \otimes_{\mathrm{End}(n)} \mathrm{Hom}(j\otimes k, n). \tag{568}$$

Finally, on another cautionary note, we should really be writing $\mathrm{Hom}_{\mathcal{C}}(X)$, just to distinguish it from $\mathrm{Hom}_{\mathbb{C}}(X)$, and likewise for $\mathrm{End}_{\mathcal{C}}$. Let's focus on $\mathrm{End}_{\mathcal{C}}(X)$. $\mathrm{End}_{\mathcal{C}}(X)$ is the space of all endomorphisms of X that *respect the structure of the category* $\mathcal{C}$. For example, if $\mathcal{C} = \mathrm{Rep}(G)$, $\mathrm{End}_{\mathcal{C}}(X)$ is the space of morphisms that commute with the G-action. This is very different from the more familiar (and much larger!) space of all $\mathbb{C}$-linear maps from X to itself, which is $\mathrm{End}_{\mathbb{C}}(X)$. In particular, we have the familiar $\mathrm{End}_{\mathbb{C}}(X) \cong X^* \otimes X$, but this definitely doesn't hold for $\mathrm{End}_{\mathcal{C}}(X)$! Likewise, X^* is certainly not defined by $\mathrm{Hom}_{\mathcal{C}}(X,\mathbb{C})$. Instead, we can define X^* through

$$\mathrm{Hom}_{\mathcal{C}}(Y, X\otimes Z) \cong \mathrm{Hom}_{\mathcal{C}}(X^*\otimes Y, Z). \tag{569}$$

Of course, all Hom spaces should be understood as $\mathrm{Hom}_{\mathcal{C}}$ spaces unless stated otherwise.

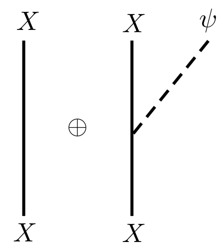
Majorana objects

We might naively think that fusing a physical fermion with a worldline X, which corresponds to doing $X \otimes \mathbb{C}^{0|1}$, would always be an odd morphism on X. That is, we might guess that doing $X \otimes \mathbb{C}^{0|1}$ would always change the grading (fermion parity) of X. However, this may not always be true! Adding a fermion can actually be an even operation, provided that X worldlines are left invariant after tensoring with $\mathbb{C}^{0|1}$. This happens precisely when

$$\mathrm{End}(X) \cong Cl_1. \tag{570}$$

Indeed, since Cl_1 (whose simple module is $\mathbb{C}^{1|1}$) has one even generator and one odd generator, tensoring with $\mathbb{C}^{0|1}$ merely interchanges these two generators, and the result is (oddly) isomorphic to what we started with: $Cl_1 \otimes \mathbb{C}^{0|1} \cong Cl_1$. In terms of representations, tensoring with $\mathbb{C}^{0|1}$ is like multiplying the generators of Cl_1 by σ^x. Since we can take the generators of Cl_1 are $\mathbf{1}_{2\times 2}$ and σ^x, multiplying by σ^x leaves the set of generators unchanged, and so tensoring with $\mathbb{C}^{0|1}$ doesn't do anything. Pictorially, (letting ψ denote a physical fermion)

$\mathrm{End}(X) \cong$  if $\mathrm{End}(X) \cong \mathbb{C}$

$\mathrm{End}(X) \cong$  if $\mathrm{End}(X) \cong \mathbb{C}\ell_1 \cong \mathbb{C}^{1|1}$ (571)

Since objects with $\mathrm{End}(X) \cong Cl_1$ can absorb fermions without changing their grading by way of $Cl_1 \otimes \mathbb{C}^{0|1} \cong Cl_1$ (and as such don't really have a well-defined fermion parity at all), they behave like Majoranas. We thus define an object X to be *Majorana* if $\mathrm{End}(X) \cong Cl_1$. Regular objects with $\mathrm{End}(X) \cong \mathbb{C}$ are simply referred to as *Bosonic*. This distinction is especially useful because all objects must be either bosonic or Majorana – there are no other possibilities.

The proof of this is straightforward: if X is a simple object, it must be simple when regarded as an object in the category (*i.e.* must not admit a direct sum decomposition with more than one nontrivial summand). Schur's lemma then tells us that any morphism between X and itself must be an isomorphism, and so all the elements of $\mathrm{End}(X)$ must be isomorphisms, and hence every element in $\mathrm{End}(X)$ must be invertible. Thus, $\mathrm{End}(X)$ must be a $\mathbb{Z}_2$-graded division algebra – *i.e.*, a $\mathbb{Z}_2$-graded algebra in which every element is invertible.

Then we can realize that $\mathbb{C}$ and $Cl_1 \cong \mathbb{C}^{1|1}$ are the only $\mathbb{Z}_2$-graded division algebras. We won't prove this rigorously, and will just give a plausibility argument. Since we need our division algebra needs to be an algebra over $\mathbb{C}$ and needs to be $\mathbb{Z}_2$-graded, the complex Clifford algebras Cl_n are the only obvious possibilities. Let's look at Cl_1 first. Cl_1 is not a division algebra in the ungraded sense, because if we ignore the grading we can write $(1-\gamma)(1+\gamma) = 0$, even though neither $1+\gamma$ nor $1-\gamma$ is zero. This isn't a problem in the $\mathbb{Z}_2$-graded case, since γ has odd degree while 1 has even degree, and in a $\mathbb{Z}_2$-graded algebra we are forbidden from adding two elements of different degrees. Playing around with the different generators for a while shows that all elements in Cl_1 are invertible, as long as we take into account the constraints from the grading. However, none of the other Clifford algebras $Cl_n, n > 1$ are division algebras: for example, for $n = 2$ we can take the odd generators γ_i in Cl_2 to obey the anticommutation rule $\{\gamma_i, \gamma_j\} = 2\sigma^z$, which means that $(1+\gamma_1\gamma_2)(1-\gamma_1\gamma_2) = 0$, and so Cl_2 is not a division algebra (we can add 1 and $\gamma_1\gamma_2$ since $\gamma_1\gamma_2$ has even parity). Similar arguments rule out the higher Cl_ns, and so Cl_0, Cl_1 are the only obvious possibilities. Thus, our only choices for $\mathrm{End}(X)$ with simple X are $\mathrm{End}(X) \cong \mathbb{C}$ or $\mathrm{End}(X) \cong \mathbb{C}^{1|1} \cong Cl_1$.

The underlying fusion category of a superfusion category

The goal in this section is to switch from a "grading the fusion spaces" picture to a "grading the worldlines" picture. Doing this results in a category with a larger collection of objects which is "bosonic" in some sense. We will call the category obtained from this bosonization procedure the *underlying fusion category* of $\mathcal{C}$ (since grading the worldlines is a procedure that associates every superfusion category with a unique "underlying" fusion category). We will write the underlying fusion category obtained from the superfusion category $\mathcal{C}$ as $\underline{\mathcal{C}}$. The bosonic-ness of this new category is helpful since finding the quasiparticle spectrum can likely be done through the usual tube algebra methods – we will have more to say about the precise relation between the excitations of superfusion categories and their associated underlying fusion categories later.

Objects in $\underline{\mathcal{C}}$ are written as X^a, where $a \in \mathbb{Z}_2$ determines their grading. The naieve rule for fusing objects in $\underline{\mathcal{C}}$ is

$$X^a \otimes X^b = (X \otimes Y)^{a+b}, \tag{572}$$

where on the left hand side $\otimes$ takes place between objects in $\underline{\mathcal{C}}$ and on the right hand side $\otimes$ takes place between objects in $\mathcal{C}$. Instead, we will see that the $a+b$ on the RHS will actually generically be twisted by a 2-cocycle (or generalization thereof for non-Abelian theories).

A very important issue for us is the rule for determining how to relate morphisms in $\underline{\mathcal{C}}$ to morphisms in $\mathcal{C}$, and vice versa. Suppose we have a morphism $f : X \to Y$, $f \in \mathcal{C}$ with parity $|f|$. When we make the switch to $\underline{\mathcal{C}}$, we get a morphism $f^b_a : X^a \to Y^b$, where $|f| = a+b$. We would like to fuse morphisms simply as $f^b_a \otimes g^d_c = (f \otimes g)^{b+d}_{a+c}$, but this leads to inconsistencies, and it turns out that this simple fusion law for morphisms must be replaced with something more twisted.

The correct relation is that tensor products of morphisms in $\underline{\mathcal{C}}$ are defined through tensor products of morphisms in $\mathcal{C}$ by the relation

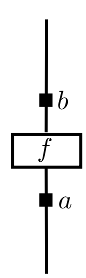
$$f^b_a \otimes g^d_c = (-1)^{(c+d+|g|)a+d|f|}(f \otimes g)^{b+d}_{a+c}. \tag{573}$$

Note that since $|g| = c+d$, we can actually just write

$$f^b_a \otimes g^d_c = (-1)^{d|f|}(f \otimes g)^{b+d}_{a+c}. \tag{574}$$

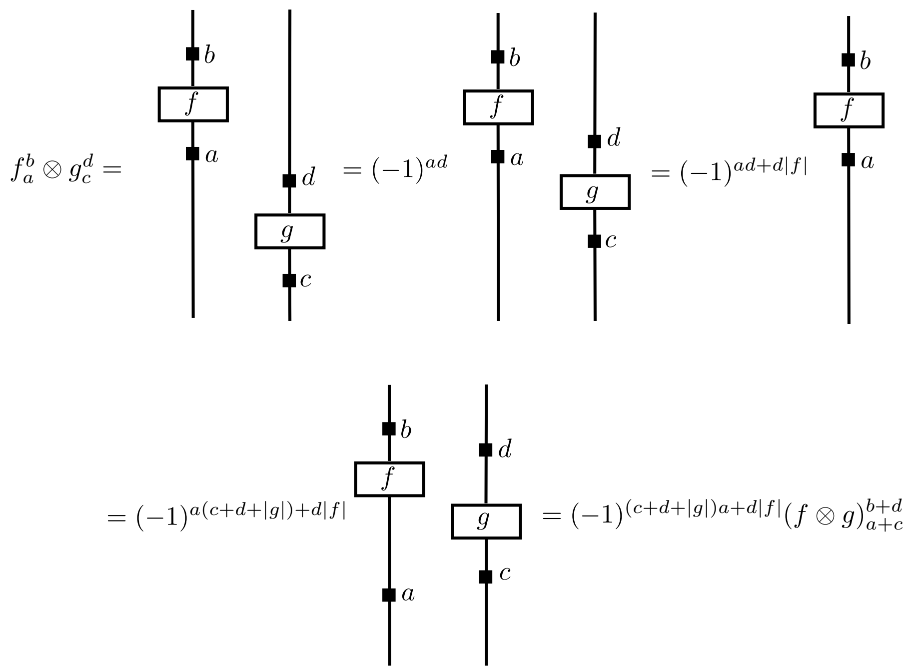
This rule will be super important in what follows, and is what allows us to determine what the F-symbols in $\underline{\mathcal{C}}$ are. However, it looks rather bizarre at first, so let's work out how to see it.

First, we start by updating our diagrammatic notation for morphisms. In particular, worldlines now carry a fermion parity grading, which we keep track of by writing a morphism $f^b_a : X^a \to Y^b$ in $\underline{\mathcal{C}}$ as

$f^b_a =$ 

(575)

where the black squares which keep track of the worldline fermion parities. If the black squares are labelled by the grading 1, they are fermions, and braid with each other when we slide them around on diagrams, while if they are labelled by the grading 0 they are bosons, and we are free to slide them around at will. This notation us allows to derive our above rule for tensoring morphisms in $\underline{\mathcal{C}}$ by the following sequence of diagrams:

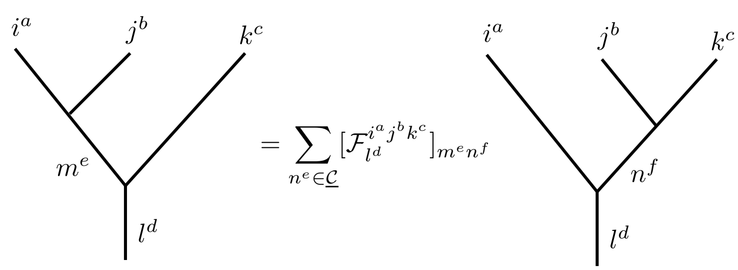


(576)

Deriving the F-symbols in $\underline{\mathcal{C}}$

We are now equipped to derive the F-symbols in $\underline{\mathcal{C}}$, which we will denote as $\mathcal{F}$. Since $\underline{\mathcal{C}}$ is bosonic in the sense that it has no fermions at the fusion spaces, the pentagon identity in $\underline{\mathcal{C}}$ holds exactly, and the $\mathcal{F}$ symbols satisfy the regular pentagon equations (in contrast to the F symbols, which we'll see do not satisfy the pentagon identity in $\mathcal{C}$). We'll ignore Majorana objects for the moment, and might come back to them later (to incorporate them we need to think carefully about Cl_1-valued F-symbols)

With Majorana objects aside, the $\mathcal{F}$ symbols in $\underline{\mathcal{C}}$ are defined by



(577)

Note that only a, b, c are independent. Explicitly, we have $e = a + b + s_m^{ij}$, $d = e + c + s_l^{mk}$, and $f = b + c + s_n^{jk}$, which are consistent constraints since the even-ness of the F-moves forces $s_m^{ij} + s_l^{mk} = s_n^{jk} + s_l^{in}$.

Usher (arXiv:1606.03466) has derived how the $\mathcal{F}$ symbols in $\underline{\mathcal{C}}$ are related to the F symbols in $\mathcal{C}$:

$$[\mathcal{F}^{i^a j^b k^c}_{l^d}]_{m^e n^f} = (-1)^{c s^{ij}_m} [F^{ijk}_l]_{mn}. \tag{578}$$

To prove his result, it helps to write down the algebraic content of the fusion diagrams involved in the previous diagrammatic relation for the $\mathcal{F}$ move. The LHS of (577) is the map

$$(i^a \otimes j^b) \otimes k^c \to m^e \otimes k^c \to l^d. \tag{579}$$

Let $H^{i^a j^b}_{m^e}$ denote a basis vector in the fusion space $\mathrm{Hom}(i^a \otimes j^b, m^e)$ (if we were not assuming multiplicity-free fusion spaces, we would have to choose several different basis vectors). With this notation, the mappings on the LHS of (577) are (in the "time flows downwards" picture)

$$H^{i^a j^b}_{m^e} \otimes \mathbf{1}_{k^c} : (i^a \otimes j^b) \otimes k^c \to m^e \otimes k^c, \qquad H^{m^e k^c}_{l^d} : m^e \otimes k^c \to l^d. \tag{580}$$

We can now use our rule of tensoring morphisms in $\underline{\mathcal{C}}$ (equation 574) to write

$$H^{i^a j^b}_{m^e} \otimes \mathbf{1}_{k^c} = (-1)^{c s^{ij}_m} (H^{ij}_m \otimes \mathbf{1}_{k^c})^{a+b+c}_{a+b+c+s^{ij}_m}, \tag{581}$$

since the parity of the fusion space H^{ij}_m is $|H^{ij}_m| = s^{ij}_m$, by definition. We also trivially have $H^{m^e k^c}_{l^d} = (H^{mk}_l)^{e+c}_{e+c+s^{mk}_l}$.

Now let's look at the RHS of (577). The fusion diagram on the RHS is written as the map

$$i^a \otimes (j^b \otimes k^c) \to i^a \otimes n^f \to l^e, \tag{582}$$

where the mappings are accomplished by

$$\mathbf{1}_{i^a} \otimes H^{j^b k^c}_{n^f} : i^a \otimes (j^b \otimes k^c) \to i^a \otimes n^f, \qquad H^{i^a n^f}_{l^d} : i^a \otimes n^f \to l^e. \tag{583}$$

Turning these into morphisms in $\underline{\mathcal{C}}$, we see that we actually have

$$\mathbf{1}_{i^a} \otimes H^{j^b k^c}_{n^f} = (\mathbf{1}_i \otimes H^{jk}_n)^{a+b+c}_{a+b+c+s^{jk}_n}, \tag{584}$$

since $|\mathbf{1}_{i^a}| = 0$.

Summarizing, we see that translating the LHS of (577) into $\underline{\mathcal{C}}$ morphisms gives us a factor of $(-1)^{c s^{ij}_m}$, and translating the fusion diagram on the RHS can be done for free, as long as we replace $\mathcal{F}$ with F. This means that the only difference between $\mathcal{F}$ and F is the $(-1)^{c s^{ij}_m}$ factor, which is exactly Usher's result.

We also note that Bhardwaj, Gaiotto, and Kapustin have also derived this result for the case when the superfusion category $\mathcal{C}$ is drawn from $\mathbf{Vec}_G$. They have a nice picture of this whole derivation in terms of "pulling out the fermions" from the fusion diagrams: check out figures 19 and 20 of their huge paper on spin-TQFTs (arXiv:1605.01640).

Since the $\mathcal{F}$ symbols satisfy the pentagon relation, we can express them in terms of F to determine that the pentagon relation holds only projectively (*i.e.* up to a sign) in $\mathcal{C}$ if the fusion space gradings s^{ij}_m are nontrivial. Explicitly, we have

$$\sum_{t \in \mathcal{C}} [F^{ijk}_n]_{mt} [F^{itk}_p]_{ns} [F^{jkl}_s]_{tq} = (-1)^{s^{ij}_m s^{kl}_q} [F^{mkl}_p]_{nq} [F^{ijq}_p]_{ms}. \tag{585}$$

The proof is straightforward, if a bit tedious: just take the pentagon relation in $\underline{\mathcal{C}}$ and use $[\mathcal{F}^{i^a j^b k^c}_{l^d}]_{m^e n^f} = (-1)^{c s^{ij}_m} [F^{ijk}_l]_{mn}$ and $s^{ij}_m + s^{mk}_l = s^{jk}_n + s^{in}_l$.

$\mathbf{sVec}_G$ theories for finite Abelian G

Despite the rather restrictive set of examples studied in this section, they are a bit nontrivial and still contain a lot of physics. They have also already been sorta understood recently by Kapustin and Gaiotto from a field theory point of view, but it still might be nice to come up with our own perspective on these theories.

First, we will focus on the case where none of the objects in the theory are Majorana. In the superfusion category $\mathcal{C} = \mathbf{sVec}_G$, the pentagon equation $\delta F = 0$ will not hold if the gradings of the fusion spaces are nontrivial. Because the objects in $\mathcal{C}$ are invertible, we will write $s(g,h)$ instead of $s^{g,h}_{gh}$. The pentagon equation is then

$$(\delta F)(g,h,k,l) = (-1)^{s(g,h)s(k,l)}, \tag{586}$$

or equivalently,

$$\delta F = (-1)^{s \cup s}. \tag{587}$$

The condition (563) on s translates into the 2-cocycle relation, so $s \in H^2(BG, \mathbb{Z}_2)$. The obstruction to satisfying the regular bosonic pentagon equations is governed by the 4-cocycle $s \cup s$, representing the fermion parity of the 4-simplex labelled by (g,h,k,l) which encloses the tetrahedra involved in the definition of the F move. Because of this, we require that $s \cup s$ be trivial when regarded as a four-cochain in $H^4(G, U(1))$, so that our theory is well-definied in strictly (2+1)D. This is a vacuous constraint for the simple examples of $G = \mathbb{Z}_n$, as $H^4(\mathbb{Z}_n, U(1)) = 0$ for all $n > 0$.

Following what Bhardwaj, Gaiotto, and Kapustin have done, let's look at how we go to the "grading the objects" picture by passing from the superfusion category $\mathcal{C} = \mathbf{sVec}_G$ to the bosonized underlying fusion category $\underline{\mathcal{C}}$. The underlying fusion category is bosonic, in the sense that

$$\underline{\mathcal{C}} = \mathbf{Vec}_{\mathcal{G}}, \tag{588}$$

where $\mathcal{G}$ is a finite group determined by the 2-cochain s and the associated short exact sequence

$$1 \to \mathbb{Z}_2 \to \mathcal{G} \to G \to 1. \tag{589}$$

More explicitly, $\mathcal{G}$ is the group consisting of the objects g^a where $g \in G$ and $a \in \mathbb{Z}_2$, with the fusion of two objects determined by the group multiplication law in G, with the $\mathbb{Z}_2$ part twisted by s:

$$g^a \otimes h^b = (gh)^{a+b+s(g,h)}. \tag{590}$$

Using our earlier relation between the F-symbols in $\underline{\mathcal{C}}$ and $\mathcal{C}$, we see that

$$\mathcal{F}(g^a, h^b, k^c) = (-1)^{s(g,h)c} F(g,h,k), \tag{591}$$

or, more concisely

$$\mathcal{F} = (-1)^{s \cup P} F, \tag{592}$$

where $P \in H^1(B\mathcal{G}, \mathbb{Z}_2)$, $P : g^a \mapsto a$ projects onto the fermion parity of its argument. P is related to S through

$$\delta P = s, \tag{593}$$

which can be understood by recalling that $s(g,h)$ measures the difference in fermion parity between the objects $g\otimes h$ and gh, which in $\mathcal{G}$ is precisely measured by the cochain δP. Note that this relation holds in $\mathcal{G}$ (not in G!), where at a more precise level, we are interpreting s in the above formula as the image of $s \in H^2(BG,\mathbb{Z}_2)$ under the inclusion $G \hookrightarrow \mathcal{G}$ induced by the map $g \mapsto g^0$. Explicitly, this just means that $(\delta P)(g^a,h^b) = P(g^a)+P(h^b) - P((gh)^{a+b+s(g,h)}) = a+b-a-b-s(g,h) = s(g,h)$.

This means that the actions of the superfusion and underlying fusion category theories are related by a $\mathbb{Z}_2$ Chern-Simons term:

$$\mathcal{F} = (-1)^{P\cup\delta P} F \tag{594}$$

Likewise, we can write the pentagon identity in $\mathcal{C}$ as holding only up to a "Θ-term" (think $F\wedge F$)

$$\delta F = (-1)^{\delta P \cup \delta P}. \tag{595}$$

We note that the "Θ-term" has to be exact in $H^4(BG,\mathbb{R}/\mathbb{Z})$ if our theory is to be consistent in strictly (2+1)D, since a nontrivial fourth cohomology class would require the presence of nontrivial (3+1)D physics.

Compendium of some category theory facts that are useful for thinking about TQFTs

This diary entry contains a summary of some of the aspects of category theory that play an important role in thinking about TQFTs and topological phases from a formal perspective. It is mainly just a choice selection of definitions and examples drawn from the awesome [2]. It was also written when I was an undergrad; proceed at your own risk.

Prerequisites from category theory

Monoidal categories and functors

The program of "categorization" essentially consists of replacing equals signs by isomorphisms. A category is a categorization of a set, an abelian category is a categorization of an abelian group, and so on. The most fundamental class of categories for us are monoidal categories, which are categorizations of monoids (recall that a monoid is a set with an associative binary operation). More precisely:

Definition 1. A *monoidal category* is a tuple $(\mathcal{C},\otimes,a,\mathbf{1},\iota)$ where $\mathcal{C}$ is an Abelian category, $\otimes : \mathcal{C}\times\mathcal{C}\to\mathcal{C}$ is a bifunctor which provides the binary operation on $\mathcal{C}$, $a : (-\otimes-)\otimes - \to -\otimes(-\otimes-)$ is a natural isomorphism called the associator, $\mathbf{1}$ is the unit object in $\mathcal{C}$, and ι is an isomorphism $\iota : \mathbf{1}\otimes\mathbf{1}\to\mathbf{1}$. We require this data to satisfy two axioms:

a) The maps $R : X\otimes\mathbf{1}\to X$ and $L : \mathbf{1}\otimes X\to X$ are tensor autoequivalences of $\mathcal{C}$.

b) The pentagon identity is satisfied.

Of these, the pentagon axiom is far more important (and we won't really worry about either $\mathbf{1}$ or ι at all). A *monoidal functor* is a functor that preserves the monoidal structure of the categories it maps between. More precisely, a monoidal functor is a pair (F, α) with $F : \mathcal{C} \to \mathcal{D}$ a functor between monoidal categories $\mathcal{C}$ and $\mathcal{D}$, and α a natural transformation (morphism of functors) that gives the tensor structure of F:

$$\alpha_{XY} : F(X) \otimes_{\mathcal{D}} F(Y) \to F(X \otimes_{\mathcal{C}} Y). \tag{596}$$

Deligne's tensor product

Deligne's tensor product is a way to "put two topological phases on top of each other" which will be very useful for us later on, and which we might as well define as part of this prerequisites section. Throughout, we let $\mathcal{C}$ and $\mathcal{D}$ be two finite abelian category, as usual over some field $\mathbb{K}$.

Definition 2. Deligne's tensor product is a bifunctor

$$\boxtimes : \mathcal{C} \times \mathcal{D} \to \mathcal{C} \boxtimes \mathcal{D} : (X, Y) \mapsto X \boxtimes Y \tag{597}$$

which is universal for functors assigning to every $\mathbb{K}$-linear abelian category $\mathcal{E}$ the category of right exact in both variables bilinear bifunctors $\mathcal{C} \times \mathcal{D} \to \mathcal{E}$.

This might seem a little cryptic at first, but it's actually pretty simple. The universality condition for $\boxtimes$ means what it always does – $\boxtimes$ is the "simplest" right exact bilinear bifunctor, meaning that all other right exact bifunctors factor through it. That is, if $\rhd$ is some other right exact bilinear bifunctor $\rhd : \mathcal{C} \times \mathcal{D} \to \mathcal{E}$, then there exists a *unique* right exact bifunctor $\overline{\rhd} : \mathcal{C} \boxtimes \mathcal{D} \to \mathcal{E}$ satisfying $\overline{\rhd} \circ \boxtimes = \rhd$. A silly analogy: If the set of all right exact bilinear bifunctors from $\mathcal{C} \times \mathcal{D}$ to $\mathcal{E}$ is the set of all ways to get from the countryside into a castle with a big moat, then $\boxtimes$ is path the castle's drawbridge. Every path from the countryside into the castle must factor through the drawbridge, and so the drawbridge is 'universal' for the category of paths leading into the castle.

Example 1. Physically, we can interpret $\mathcal{C} \boxtimes \mathcal{D}$ as a topological phase described by $\mathcal{C}$ "stacked on top of" one described by $\mathcal{D}$. In particular, if we stack $\mathcal{C}$ on top of $\mathcal{C}^{op}$, we in some way "average" over the difference between $X \otimes Y$ and $Y \otimes X$, and so $\mathcal{C} \boxtimes \mathcal{C}^{op} = \mathcal{Z}(\mathcal{C})$, where $\mathcal{Z}(\mathcal{C})$ is the center of $\mathcal{C}$ (which also corresponds to the quasiparticle excitations of $\mathcal{C}$). This will be made more precise later on.

Tensor categories

Rigidity

Before we define tensor categories, we need to define the notion of dual objects. This is done in pretty mucht the way we'd expect. Let $\mathcal{C}$ be a monoidal category, and $X \in \mathcal{C}$. We say that X^* is a *left dual* of X if there exist morphisms $d_X : X^* \otimes X \to \mathbf{1}$ (the death morphism) and $b_X : \mathbf{1} \to X \otimes X^*$ (the birth morphism) such that the maps

$$\begin{aligned} &X \to (X \otimes X^*) \otimes X \to X \otimes (X^* \otimes X) \to X, \\ &X^* \to X^* \otimes (X \otimes X^*) \to (X^* \otimes X) \otimes X^* \to X^* \end{aligned} \tag{598}$$

are identity morphisms. This is equivalent to saying that if X has a left dual, we can remove vertical s-shaped wiggles of an X line in any diagram. The notion of a right dual ${}^*X$ is defined in the same way, but with the roles of the birth and death morphisms interchanged. The category $\mathcal{C}$ is called *rigid* if each of its objects have both left and right duals.

A technical but important point: we may not always have $X^{**} = X$. This is the case in $\mathbf{Vec}_{\mathbb{K}}$ and $\mathrm{Rep}_{\mathbb{K}}(G)$, but not in general. The isomorphism between X^{**} and X is called a *pivotal structure* on $\mathcal{C}$, and we write it as $p_X : X \to X^{**}$, which is natural in X and satisfies $p_{X\otimes Y} = p_X \otimes p_Y$. We need to incorporate the pivotal structure into our definition of quantum traces as the factors picked up by killing off closed bubbles in the graph. We are then led to define the *pivotal trace* of a morphism (not just of an object) by

$$\mathrm{ptr}_p(f : X \to X) = b_{X^*}(p_x f \otimes \mathbf{1}_{X^*})d_X \in \mathrm{End}_{\mathcal{C}}(\mathbf{1}) \tag{599}$$

In particular, the quantum dimensions familiar from string-net diagrammatics are defined through the pivotal trance of the identity morphism: $d_X = \mathrm{ptr}_p(\mathbf{1}_X)$ where the pivotal structure is chosen in a canonical way so that the pivotal dimension agrees with the Frobenius-Perron dimension (to be discussed later).

Tensor categories, fusion categories, and tensor functors

We start with some definitions.

Definition 3. A *multitensor category* is a $\mathbb{K}$-linear abelian rigid monoidal category $\mathcal{C}$, where the tensor product bifunctor is bilinear on morphisms. If the unit object is simple (i.e., if $\mathrm{End}_{\mathcal{C}}(\mathbf{1}) \simeq \mathbb{K}$), then $\mathcal{C}$ is just a tensor category. If $\mathcal{C}$ is finite semisimple (multi)tensor category, we call $\mathcal{C}$ a *(multi)fusion category.* If we drop the assumption of rigidity, then $\mathcal{C}$ is called a *multiring category.*

We will essentially always specialize to fusion categories. Some easy examples are $\mathbb{K}-\mathbf{Vec}$ (fusion), $\mathrm{Rep}_{\mathbb{K}}(G)$ (fusion if char $\mathbb{K} = 0$ or $\gcd(\mathrm{char}\,\mathbb{K}, |G|) = 1$), $\mathbf{Vec}_G$, etc.

One very general example of multitensor categories comes from the notion of a groupoid:

Definition 4. A *groupoid* $\mathcal{G}$ is a category where all morhpisms are isomorphisms – that is, $\mathcal{G}$ consists of a set S of objects of $\mathcal{G}$ and a set G of morphisms of $\mathcal{G}$, the image and preimage maps, the composition map, the unit morphism map, and the inversion map.

In particular, any group can be made into a groupoid: this groupoid only has one object, and the endomorphisms of that object are given by the group G. Groupoids also provide the most general examples of multi-tensor categories.

Example 2. Let $\mathcal{G}$ be a groupoid with a set of objects S (with S finite), and let $\mathcal{C}(\mathcal{G})$ be the category of finite dimensional vector spaces graded by the morphisms of G, which we will take to be a finite group. The objects in $\mathcal{C}(\mathcal{G})$ can thus be decomposed as $S \ni X = \oplus_{g\in G} X_g$. The tensor product is defined naturally by

$$(X \otimes Y)_g = \bigoplus_{\{h,k|hk=g\}} X_h \otimes Y_k \tag{600}$$

The unit object in $\mathcal{C}(\mathcal{G})$ also decomposes in a semisimple way as $\mathbf{1} = \oplus_{s\in S}\mathbf{1}_s$. We thus see that $\mathcal{C}(\mathcal{G})$ is a multi-tensor (rather than just tensor) category. In particular, if we take $S = \mathbb{Z}_n$, then $\mathcal{C}(\mathcal{G})$ is a category based on "matrices of vector spaces", and is denoted by $\mathbf{Mat}_n(\mathbf{Vec})$. Note that this is *not* the same as $\mathbf{Mat}_n(\mathbf{Vec}_G)$!

To get $\mathbf{Mat}_n(\mathbf{Vec}_G)$, we need to consider a slightly more general scenario. Let $\{\hat{\mathcal{G}}_i\}$ be the set of isomorphism classes of a groupoid $\mathcal{G}$, and let $G_i = \mathrm{Aut}(g_i)$ for some representative g_i of $\hat{\mathcal{G}}_i$. Then $\mathcal{C}(\mathcal{G})$ is (monoidally equivalent to) $\bigoplus_i \mathbf{Mat}_{|\hat{\mathcal{G}}_i|}(\mathbf{Vec}_{G_i})$.

Just as the definition of monoidal categories led us to define monoidal functors, tensor categories prompt us to define functors between them, which are aptly named *tensor functors*. A tensor functor isn't actually a functor (bad language!) − rather, it's a functor + natural isomorphism pair. The formal definition is as follows:

Definition 5. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor, and $J_{X,Y} : F(X) \otimes_{\mathcal{D}} F(Y) \to F(X \otimes_{\mathcal{C}} Y)$ be a natural isomorphism detailing the tensor structure of F. The pair (F, J) is called a *quasi-tensor functor* if F is exact (i.e., both left exact and right exact), and if $F(\mathbf{1}_{\mathcal{C}}) = \mathbf{1}_{\mathcal{D}}$. If furthermore (F, J) satisfies the monoidal structure axiom (with the associators in $\mathcal{C}$ and $\mathcal{D}$), then (F, J) is just called a tensor functor.

There are tons of examples of (quasi)-tensor functors. They include essentially all forgetful functors: indeed, if $H \subset G$ is a subgroup, then the canonical forgetful functor $\mathrm{Rep}(G) \to \mathrm{Rep}(H)$ is tensor. Similarly, if $f : H \to G$ is a group homomorphism, then the pullback functor $f^* : \mathrm{Rep}(G) \to \mathrm{Rep}(H)$ is tensor (the canonical forgetful functor is an example of this when $f : H \hookrightarrow G$ is the inclusion). As we talked about earlier, if F is the identity functor, then there exists no tensor functor $F : \mathcal{C}_G^{\omega_1} \to \mathcal{C}_G^{\omega_2}$ if ω_1 and ω_2 are not cohomologous, while if they are cohomologous the tensor functors are classified by $H^2(G, \mathbb{K}^\times)$. Since the only problem we run into when ω_1 and ω_2 are not cohomologous is satisfying the monoidal structure axiom, we see that *any* choice of J makes (F, J) a quasi-tensor functor.

Grothendieck rings and FP dimensions

Definition 6. The Grothendieck ring $\mathbf{Gr}(\mathcal{C})$ is the free group generated by isomorphism classes of the simple objects of $\mathcal{C}$. If $\mathcal{C}$ is a (multi)fusion category, $\mathbf{Gr}(\mathcal{C})$ is a (multi)fusion ring.

Any object $X \in \mathcal{C}$ can be decomposed in a series

$$0 = X_0 \subset X_1 \subset \cdots \subset X_n = X, \tag{601}$$

where X_i/X_{i-1} is simple for all i. This series is called the *Jordan-Holder series* of X. We denote by $[X : Y]$ the number of times Y occurs in the set $\{X_i/X_{i+1}\}$. For example, when $\mathcal{C} = \mathrm{Rep}(G)$, the index $[X_1 \otimes X_2 : X_3]$ is the multiplicity of X_3 in the direct sum decomposition of $X_1 \otimes X_2$.

In $\mathbf{Gr}(\mathcal{C})$, we associate each $X \in \mathcal{C}$ with its class $[X] \in \mathbf{Gr}(\mathcal{C})$, given in naturally by

$$[X] = \sum_{j\in \mathrm{Obj}(\mathcal{C})} [X : X_j] X_j. \tag{602}$$

We will often abuse notation and write X instead of $[X]$. The tensor product in $\mathcal{C}$ induces a natural multiplication in $\mathbf{Gr}(\mathcal{C})$ via the formula

$$X_i X_j \equiv [X_i \otimes X_j] = \sum_{k \in \mathrm{Obj}(\mathcal{C})} [X_i \otimes X_j : X_k] X_k. \tag{603}$$

This multiplication rule is called the *fusion rule* on $\mathcal{C}$. It does *not* determine $\mathcal{C}$! As an example, consider theories based on $\mathbf{Vec}_{\mathbb{Z}_n}$. There are n different topologically distinct theories given by the n distinct cocycle generators of $H^3(\mathbb{Z}_N, U(1))$, all of which have the same fusion rules.

Finally, we note that although the fusion rules do not determine $\mathcal{C}$, we can use the behavior of $\mathcal{C}$'s quasiparticle excitations to tell us something about $\mathbf{Gr}(\mathcal{C})$ – in particular, when $\mathcal{C}$ is braided, $\mathbf{Gr}(\mathcal{C})$ is commutative (although the converse need not hold in general).

Now we turn to the concept of Frobenius-Perron dimensions, which are closely related (albeit in a non-obvious way) to the pivotal dimensions discussed earlier.

Definition 7. $\mathrm{fpdim}(X)$ is the largest real eigenvalue of left multiplication by X on the Grothendieck ring $\mathbf{Gr}(\mathcal{C})$.

We can (almost) always choose a pivotal structure on $\mathcal{C}$ so that $\mathrm{fpdim}(X)$ is equal to $\mathrm{pdim}(X)$. In keeping with the notation commonly used in physics, we will ocassionaly write both $\mathrm{fpdim}(X)$ and $\mathrm{pdim}(X)$ simply as d_X when there is no point in distinguishing between the two. Being largest positive eigenvalues of the objects that form the basis of $\mathbf{Gr}(\mathcal{C})$, the d_X's form a one-dimensional representation of the fusion algebra:

$$d_X d_Y = \sum_{Z \in \mathrm{Obj}(\mathcal{C})} [X \otimes Y : Z] d_Z. \tag{604}$$

Proposition 2. *If $F : \mathcal{C} \to \mathcal{D}$ is a quasi-tensor functor, then $\mathit{fpdim}_{\mathcal{C}}(X) = \mathit{fpdim}_{\mathcal{D}}(F(X))$ for all $X \in \mathcal{C}$.*

The proof is straightforward – the (quasi)-tensor structure of F induces a homomorphism in $\mathbf{Gr}(\mathcal{C})$. We then use the fact that the map

$$\chi : [X] \mapsto d_X \tag{605}$$

is a character of $\mathbf{Gr}(\mathcal{C})$ and the standard invariance of characters. We can use this to see that $\mathrm{fpdim}_{\mathcal{C}}(X) = \mathrm{fpdim}_{\mathcal{C}^{op}}(X)$. Later, we will use it to show that the FP-dimensions of a topological phase's quasiparticle excitations are equal to those of the objects of its ground-state.

Example 3. Of course, there are a zillion examples of Grothendieck rings. If $\mathcal{C} = \mathrm{Rep}_k(\mathfrak{sl}_2(\mathbb{C}))$ is the category of representations of $\mathfrak{sl}_2(\mathbb{C})$ truncated at dimension k representations, then the objects in $\mathbf{Gr}(\mathcal{C})$ are irreps with positive integral dimensions $d_X \leq k$, with the fusion rule

$$X_i X_j = \sum_{l=\max(i+j-k.0)}^{\min(i,j)} X_l. \tag{606}$$

The quantum dimensions are given by a formula that looks pretty weird at first glance:

$$\mathrm{fpdim}(X_j) = \frac{q^{j+1} - q^{-j-1}}{q - q^{-1}} \tag{607}$$

with $q = \exp(\pi i/(k+2))$.

This fusion category is usually written as $\mathbf{Ver}_k$. For small k, they can be identified as $\mathbf{Ver}_0 = \mathbb{Z}$, $\mathbf{Ver}_1 = \mathbb{Z}[\mathbb{Z}_2]$, and $\mathbf{Ver}_2 = TY_{\mathbb{Z}_2}$. $TY_{\mathbb{Z}_2}$ is a *Tambara-Yamagami* fusion ring, which we will talk about in the next example.

Considering $\mathbf{Ver}_k$ at different values of k (different "fillings") gives us the origin story of two of our favorite non-abelian theories. For example, $\mathbf{Ver}_2$ is the Ising fusion ring. It of course has two nontrivial objects σ and ψ, with the fusion rules

$$\psi\psi = \mathbf{1},\ \psi\sigma = \sigma\psi = \sigma,\ \sigma\sigma = \mathbf{1} + \psi. \tag{608}$$

Looking at the fusion rules, we can read off $\mathrm{fpdim}(\psi) = 1,\ \mathrm{fpdim}(\sigma) = \sqrt{2}$.

Even $\mathbf{Ver}_3$ is pretty big and unwieldy, but it has a simple subtheory generated by irreps of odd dimension. The basis for this ring consists of $\mathbf{1}$ and one quasiparticle τ, with $\tau\tau = 1 + \tau$. This is of course the Fibonacci fusion ring, and the FP dimension of τ is famously $\mathrm{fpdim}(\tau) = (1+\sqrt{5})/2$.

Example 4. Let G be a finite group. Define the Tambara-Yamagami fusion ring of G by

$$TY_G \equiv \mathbb{Z}G \oplus \mathbb{Z}\sigma \tag{609}$$

which has the fusion rules

$$g\sigma = \sigma g = \sigma,\ \sigma\sigma = \sum_{g\in G} g. \tag{610}$$

The fusion between two elements of G is group multiplication. The quantum dimensions are obviously $\mathrm{fpdim}(g) = 1$ for all $g \in G$ and $\mathrm{fpdim}(\sigma) = \sqrt{|G|}$.

We will now show how the F-symbols are defined, which is old news for most people but is included here for completeness. We let $H^{ab}_c = \mathrm{Hom}_{\mathcal{C}}(a \otimes b, c)$ for simple $a, b, c \in \mathcal{C}$, allowing us to write

$$a \otimes b = \bigoplus_c H^{ab}_c \otimes c. \tag{611}$$

There are two different ways of fusing a triple of objects together:

$$\begin{aligned} (a \otimes b) \otimes c &= \bigoplus_{d,e} H^{ab}_e \otimes H^{ec}_d \otimes e, \\ a \otimes (b \otimes c) &= \bigoplus_{d,f} H^{af}_d \otimes H^{bc}_f \otimes f \end{aligned} \tag{612}$$

The coordinates of the associativity isomorphism between these two equations are called the *F-symbols* of the theory:

$$[F^{abc}_d]_{ef} : H^{ab}_e \otimes H^{ec}_d \to H^{af}_d \otimes H^{bc}_f \tag{613}$$

In most physical applications, we can fix the $[F_d^{abc}]$ to be unitary matrices.

As mentioned earlier, the process of categorification amounts to taking some mathematical structure and replacing all the equalities with isomorphisms. We now have the vocabulary needed to make this more precise for the case of categorified rings:

Definition 8. Given a ring R, its *categorification over* $\mathbb{K}$ is a multifusion category over $\mathbb{K}$ together with an isomorphism $R \to \mathbf{Gr}(\mathcal{C})$ for some category $\mathcal{C}$.

Of course, since fusion rules do not uniquely determine the category, there may be many possible categorifications of a given ring. Is it possible to find all the categorifications of a given ring? In principle, yes. To do so, we need to find all possible solutions to the axioms for a monoidal category (the pentagon and triangle axioms) modulo the group of local vertex gauge transformations given by $\mathrm{Aut}(H_k^{ij})$. Of course, this is very difficult – the pentagon identity usually takes the form of a huge number of over-determined nonlinear equations. However, a complete classification is possible in some scenarios, like the ones considered earlier in this section. For example, if G is finite then the categorifications of G are classified by $\omega \in H^3(G, \mathbb{K}^\times)$ and are identified by $\mathbf{Vec}_G^\omega$. As another example, the Tambara-Yamagami fusion ring TY_G we touched upon earlier is categorifiable if and only if G is abelian.

Graded categories and SETs

Definition 9. A *grading* of a tensor category by a finite group G is a decomposition

$$\mathcal{C} = \bigoplus_{g \in G} \mathcal{C}_g \tag{614}$$

where the $\mathcal{C}_g$'s are tensor subcategories and the tensor product maps $\otimes : \mathcal{C}_g \times \mathcal{C}_h \to \mathcal{C}_{gh}$. The grading is called *faithful* if $\mathcal{C}_g$ is non-empty for all $g \in G$.

If a category $\mathcal{C}$ is faithfully graded in two different ways, then we can always choose a common refinement of the two gradings such that the refinement is still faithful. That is, we can always choose a "smallest" faithful grading of $\mathcal{C}$. We will call such a grading the *universal grading*, and label the associated group by $U_{\mathcal{C}}$.

There is a very natural interpretation of graded categories. We regard the subcategory $\mathcal{C}_e$ as some base topological phase, and interpret the G-grading as a way of incorporating a symmetry group G of $\mathcal{C}_e$ into category-theoretic language. One particularly simple example of this is offered by bosonic symmetry-enriched topological phases (SETs) – we take $\mathcal{C}_e$ to consist of a single bosonic quasiparticle, and each $\mathcal{C}_g$ to consist of a single object $g \in G$.

A useful concept in our upcoming discussion of equivariantization and anyon condensation is the notion of an adjoint category.

Definition 10. Let $\mathcal{C}$ be a tensor category. We define the *adjoint category* to be the smallest tensor subcategory $\mathcal{C}_{ad} \subset \mathcal{C}$ such that $\mathcal{C}_{ad}$ contains all fusion products in the set $\{X \otimes X^* | X \in \mathcal{C}\}$.

In particular, $\mathcal{C}_{ad}$ is a proper tensor subcategory of $\mathcal{C}$ when the quasiparticles in $\mathcal{C}$ are non-Abelian. As a concrete example, recall the Tambara-Yamagami categorization of $\mathbb{Z}_n$, whose objects are those of $\mathbb{Z}_n$ and an object σ satisfying $\sigma \otimes g = \sigma$ for $g \in \mathbb{Z}_n$ and $\sigma \otimes \sigma = \oplus_{g \in \mathbb{Z}_n} g$. We see that in this example $\mathcal{C}_{ad} = \mathbb{Z}_n$, and that $\sigma \notin \mathcal{C}_{ad}$.

We can also think about $\mathcal{C}_{ad}$ as the trivial subcategory $\mathcal{C}_e$ of $\mathcal{C}$. This is because if $X \in \mathcal{C}$ is graded by g, then X^* must be graded by g^{-1}, implying that every fusion product in $X \otimes X^*$ (there may be several if $\mathcal{C}$ has non-abelian quasiparticles) must be in $\mathcal{C}_e$.

The next result is straightforward, but important enough to put as a proposition:

Proposition 3. *Let $\mathcal{C} = \mathrm{Rep}(G)$. Then $\mathcal{C}_{ad} = \mathrm{Rep}(G/Z(G))$, where $Z(G)$ is the center of G.*

In particular, if G is abelian then $\mathcal{C}_{ad}$ is trivial (since $X \otimes X^* = \mathbf{1}$ always – $\mathrm{Rep}(G)$ has no irreps with dimension greater than 1).

Our goal in this section is to start to understand how anyon condensation is described by this sort of mathematics. The next proposition is a good starting point:

Proposition 4. *There is a bijection between subgroups $G \subset U_{\mathcal{C}}$ and tensor subcategories $\mathcal{D} \subset \mathcal{C}$ which contain $\mathcal{C}_{ad}$. The correspondence is given explicitly by*

$$\mathcal{C} \mapsto G_{\mathcal{D}} = \{g \in U_{\mathcal{C}} | \mathcal{D} \cap \mathcal{C}_g \neq 0\}, \quad G \mapsto \mathcal{D}_G = \bigoplus_{g \in G} \mathcal{C}_g \tag{615}$$

The proof is actually pretty easy – just write $\mathcal{C} = \oplus_{g \in U_{\mathcal{C}}} \mathcal{C}_g$ and remember that $\mathcal{C}_e = \mathcal{C}_{ad}$. It's also pretty intuitive – if a tensor subcategory of $\mathcal{C}$ contains $\mathcal{C}_{ad}$ (which functions as the unit object in the corresponding subgroup of $U_{\mathcal{C}}$), its grading naturally inherits a subgroup structure.

On a related note, we have the following:

Proposition 5. *There is a bijection between equivalence classes of faithful gradings of $\mathcal{C}$ and tensor subcategories $\mathcal{C} \subset \mathcal{C}$ with $\mathcal{C}_{ad} \subset \mathcal{D}$ such that $G_{\mathcal{D}} \subset U_{\mathcal{C}}$ is a normal subgroup. The bijection is exhibit by the assignment of $\mathcal{D}$ to the universal grading of $\mathcal{C}$ which grades $\mathcal{D}$ trivially and the assignment of a given grading of $\mathcal{C}$ to its trivial component.*

The proof is done as follows: for every normal $G \subset U_{\mathcal{C}}$, we can grade $\mathcal{C}$ faithfully by $U_{\mathcal{C}}/G$ Since $U_{\mathcal{C}}$ is universal, there is a bijection between normal subgroups of $U_{\mathcal{C}}$ and equivalence classes of faithful gradings of $\mathcal{C}$. We then use the result of the previous proposition to identify these subgroups with the desired tensor subcategories $\mathcal{D} \subset \mathcal{C}$ which contain $\mathcal{C}_{ad}$.

An important remark is that the structure of $U_{\mathcal{C}}$ actually tells us quite a bit about how the quasiparticle excitations of $\mathcal{C}$ behave. For example, if $\mathcal{C}$ is braided (which is the same thing as saying that $\mathbf{Gr}(\mathcal{C})$ is a commutative ring), then $U_{\mathcal{C}}$ must be abelian.

Equivariantization part II

In this section, we revisit equivariantization in a different but (non-obviously) equivalent way. As before, an action of a group G on a multitensor category $\mathcal{C}$ is a monoidal functor

$$T : \mathrm{Cat}(G) \to \mathrm{Aut}_{\otimes}(\mathcal{C}) \tag{616}$$

Also as before, we let $\mathcal{C}^G$ denote the category of G-equivariant objects.

Example 5. We can already see some examples of "electromagnetic duality" (or Rep $-$ **Vec** duality) by considering the equivariantization of **Vec**. The G-action (for finite G) on **Vec** is always trivial, of course. This means that an object in $\mathcal{C}^G$ is labeled solely by a collection of isomorphisms $\{u_g : X \simeq X | g \in G\}$, where X is a vector space. These isomorphisms have to form a 1d (projective) representation to satisfy the definition of equivariance, and so $\{u_g\}$ determine a representation of G, implying that $\text{Rep}(G) \simeq \mathbf{Vec}^G$.

Example 6. More generally, let $N \subset G$ be normal in G. Then we can form a natural action of G/N on $\text{Rep}(N)$ by "coset multiplication". That is, we let $g \in G/N$ act on N by $N \mapsto gN$ (thinking of $\text{Rep}(N)$ as a coset is a bit imprecise, but we'll just deal with it). This takes the unit coset to another coset in the coset decomposition of G, which will generically also permute the elements of N as it does so. That is, we have a functor $T : \mathbf{Cat}(G/N) \to \text{Aut}_\otimes(\text{Rep}N)$. It's easy to check that this functor is monoidal, and so G/N acts on $\text{Rep}(N)$ in the category theory sense.

Since we've found an action $T : \mathbf{Cat}(G/N) \to \text{Aut}_\otimes(\text{Rep}N)$, we can consider the equivariantization $\text{Rep}(N)^{G/N}$. The equivariantized category has objects of the form

$$\text{Obj}(\text{Rep}(N)^{G/N}) = \Big\{(n, \{u_g\} | u_g : T_g(n) \xrightarrow{\sim} n,\ g \in G/N)\Big\}. \tag{617}$$

That is, every object is a pair consisting of one object n of $\text{Rep}(N)$ and one collection of isomorphisms that equivariantize the action on n. But the collection of isomorphisms is really just the same as a representation of G/N, and so we have

$$\text{Rep}(N)^{G/N} = \text{Rep}(G). \tag{618}$$

The last step is a bit subtle, but quite easy to see when G/N is abelian. In this case, we have $\text{Obj}(\text{Rep}(N)^{G/N}) = \{(n, g)\},\ n \in N,\ g \in G/N$, which is just equal to $N \times G/N = G$.

This sort of process is very closely related to the process of gauging symmetries in topological phases, which should be thought of as an "inverse" of anyon condensation. Equivariantization also offers a natural way of understanding the quasiparticle excitations of a theory, through the relation $\mathbf{Vec}_G^G = \mathcal{Z}(\mathbf{Vec}_G)$, with G acting on $\mathbf{Vec}_G$ by conjugation (which will be discussed at greater length later on).

Finally, we note an interesting result about quantum dimensions.

Proposition 6. *Let $F : \mathcal{C}^G \to \mathcal{C}$ be the canonical forgetful functor. Since F is tensor, we have fpdim$((X, u))$ = fpdim$(F((X, u)))$ = fpdim(X).*

In particular, equivariantizing an object does not change its quantum dimension. The proof is straightforward if we realize that any monoidal functor must preserve quantum dimensions. This holds since the map $\chi : [X] \mapsto \text{fpdim}(X)$ is a character on $\mathbf{Gr}(\mathcal{C})$, which is invariant under ring homomorphisms.

We will now discuss a way of formulating the quantum double (or Drinfeld center) construction for phases with finite symmetry groups G in a fairly non-obvious way.

Definition 11. Let G be a finite group with an action on a tensor category $\mathcal{C}$. We define the *crossed product category of $\mathcal{C}$ by G*, denoted as $\mathcal{C} \rtimes G$, by setting $\mathcal{C} \rtimes G = \mathcal{C} \boxtimes \mathbf{Vec}_G$ (which is just an abelian category) with the "twisted" tensor product

$$(X \boxtimes g) \otimes (Y \boxtimes h) = (X \otimes T_g(Y)) \boxtimes gh. \tag{619}$$

Of course, the unit object in $\mathcal{C} \rtimes G$ is $(\mathbf{1} \boxtimes e)$, and the crossed product category inherits the associativity properties of $\mathcal{C}$. $\mathcal{C} \rtimes G$ also has a natural G-grading, where the gth component of the grading is $\mathcal{C} \otimes (\mathbf{1} \boxtimes g)$ (this . When $g = e$, the resulting tensor subcategory is $(\mathcal{C} \rtimes G)_{ad}$. A dumb example of a crossed product category is $\mathbf{Vec} \rtimes G = \mathbf{Vec}_G$. A less dumb example is $\mathbf{Vec}_G \rtimes G$, which is known as the *twisted quantum double* of G, which provides a great way to realize simple topological phases that do not admit a braiding.

If you have already encountered the twisted quantum double model before, you might notice that the way of defining the tensor product in $\mathcal{C} \rtimes G$ is very similar to the fusion rules for the quasiparticle excitations of $\mathbf{Vec}_G$. Indeed, we will show later on that $\mathcal{C} \rtimes G$ is very closely related to $\mathcal{Z}(\mathbf{Vec}_G)$ through a rather interesting form of "electromagnetic" duality. Indeed, the crossed product construction and the equivariantization process are essentially two sides of the same coin—more on this later as well.

The center construction

The center construction $\mathcal{Z}(\mathcal{C})$ is just what you think it is – the categorical analogue of the center of a group. Objects in $\mathcal{Z}(\mathcal{C})$ are essentially objects in $\mathcal{C}$ equivariantized over itself. Of course $\mathcal{C}$ is not a group, so the equivariantization process will not be accomplished by adding representations of $\mathcal{C}$ to make the '$\mathcal{C}$ action' commute with itself. However, the definition of $\mathcal{Z}(\mathcal{C})$ closely parallels that of $\mathcal{C}^G$:

Definition 12. The *center* of a monoidal category $\mathcal{C}$ with associator a is the category $\mathcal{Z}(\mathcal{C})$ whose objects are pairs (Z, c) where $Z \in \mathcal{C}$ and

$$c_X : X \otimes Z \to Z \otimes X, \quad X \in \mathcal{C} \tag{620}$$

is a natural isomorphism such that the following diagram (the hexagon equations) commutes:

Tensoring two objects in $\mathcal{Z}$ is done by $(Z, c) \otimes (Z', c') = (Z \otimes Z', \widetilde{c})$, where $\widetilde{c}_X : X \otimes (Z \otimes Z') \to (Z \otimes Z') \otimes X$ is a natural isomorphism defined so that it respects the associative structure. Tensor categories with c isomorphisms satisfying the previous two diagrams are called *braided*.

Physics stuff

When some anyon model given by a category $\mathcal{C}_e$ has a symmetry G, we can consider the existence of localized point-like defects corresponding to G-symmetry fluxes (G doesn't have to be finite – see Hermele's paper on the flux-fusion anomaly test). These defects act like quasiparticles, but are confined rather than delocalized. In addition, the defects can

As a simple example (yet one of a rather different flavor than the examples considered here), let $G = U(1)$ and let $\mathcal{C}_e$ be the toric code. For simplicity we can break $U(1)$ down to $\mathbb{Z}_2$, and consider imposing π-fluxes Ω_π that function as the symmetry defects in the theory. They behave like quasiparticles, with the fusion rule $\Omega_\pi \otimes \Omega_\pi = m$, and can also influence deconfined quasiparticles that orbit around them in a way that will become more precise soon.

Of course, defect fusion must respect the G-grading:

$$a_g \otimes b_h \in \mathcal{C}_{gh}. \tag{621}$$

This immediately tells us that our notion of braiding will need to be refined when dealing with G-graded categories – when G is not abelian, the fusion rules are not even commutative!

The symmetry group G naturally possesses an action on the original anyon theory, which is a monoidal functor

$$T : \mathbf{Cat}(G) \to \mathrm{Aut}_{\otimes}(\mathcal{C}_e) \tag{622}$$

and describes how the global symmetry G acts on the topological degrees of freedom.

The number of topologically distinct g-defects ($|\mathcal{C}_g|$) is equal to the number of g-invariant topological charges (i.e., those with $T_g(a) = a$), at least when G is finite. We can prove this by realizing that we must have

$$T_g : \mathcal{C}_h \to \mathcal{C}_{ghg^{-1}} \tag{623}$$

in order for T to be compatible with the G-graded fusion rules (i.e., we want $T_g(a_h \otimes b_k)$ to be in the same component of $\mathcal{C}$ as $T_g(a_h) \otimes T_g(b_k)$). Additionally, twisting the incoming legs of a fusion vertex should not affect its fusion outcomes, and so we require that

$$[a_g \otimes b_h : c_{gh}] = [T_g(b_h) \otimes a_g : c_{gh}]. \tag{624}$$

This sort of "twisted" rule should get your spidey senses tingling! We can finish up the proof about the number of topologically distinct g-defects if we realize that (writing N^c_{ab} for $[a \otimes b : c]$)

$$N^1_{a_g b_h} = N^1_{T_g(b_h) a_g} = \delta_{T_g(b_h), \bar{a_g}} \tag{625}$$

and so any topological defect in $\mathcal{C}_g$ must be invariant under the action of $\langle g \rangle \subset G$:

$$T_{g^k}(a_g) = a_g, \ \forall \ k \in \mathbb{Z} \tag{626}$$

Of course, no other b_h will be invariant under T_g, and so indeed $|\mathcal{C}_g|$ is equal to the number of topologically distinct g-invariant defects.

Another useful result tells us that the total quantum dimension of each component $\mathcal{C}_g$ is independent of g:

Proposition 7. *We have*

$$D_g = \sqrt{\sum_{\mathcal{C}_g} d^2_{a_g}} = D_e, \ \forall \ g \in G. \tag{627}$$

The proof is done by straightforward manipulation. From the definition of $1/d_a^2$ as "the probability that $a \otimes a^*$ is vacuum" and the fact that the quantum dimensions must be real, we see that $d_{a_g^*} = d_{a_g}$ for all a_g. We also use the fact that the quantum dimensions form a (G-graded) Grothendieck ring, and so give a 1D representation of the fusion algebra:

$$d_{a_g} d_{b_h} = \sum_{c \in \mathcal{C}_{gh}} N^{c_{gh}}_{a_g b_h} c_{gh}. \tag{628}$$

This allows us to write (choosing an arbitrary $c_g \in \mathcal{C}_g$)

$$d_{a_e} = d_{a_e} d_{c_g}^{-2} d_{c_g} d_{c_g^*} = \sum_{f \in \mathcal{C}_g} d_{c_g}^{-2} N^{f_g}_{a_e c_g} d_{f_g} d_{c_g^*} = \sum_{f \in \mathcal{C}_g, b \in \mathcal{C}_e} d_{c_g}^{-2} N^{f_g}_{a_e c_g} N^{b_e}_{f_g c_g^*} d_{b_e} \tag{629}$$

Which we can then insert into the definition of D_e^2:

$$
\begin{aligned}
D_e^2 = \sum_{a\in\mathcal{C}_e} d_{a_e}^2 &= \sum_{a,b\in\mathcal{C}_e, f\in\mathcal{C}_g} d_{a_e} d_{c_g}^{-2} N_{a_e c_g}^{f_g} N_{f_g \overline{c_g}}^{b_e} d_{b_e} = \sum_{a,b\in\mathcal{C}_e, f\in\mathcal{C}_g} d_{c_g}^{-2} N_{f_g c_g^*}^{a_e} d_{a_e} N_{f_g c_g^*}^{b_e} d_{b_e} \\
&= \sum_{f\in\mathcal{C}_g} d_{c_g}^{-2} d_{f_g}^2 d_{c_g}^2 = \sum_{f\in\mathcal{C}_g} d_{f_g}^2 = D_g^2
\end{aligned}
\tag{630}
$$

Finally, we have to take into account the interaction between the G action and the hom spaces H_c^{ab}. We do this diagramatically by associating to each anyon worldline and hom space a branch cut that points into the page. When anyon worldlines pass over branch cuts, they pick up a phase factor equal to the projective local symmetry action.

As mentioned earlier, these G-crossed theories will not be braided in general. This leads us to define the notion of G-crossed braiding, and gives rise to the *heptagon equations* which replace the role of the hexagon equations (a.k.a. the monoidal structure axiom for the functor $F : \mathcal{C} \to \mathcal{C}_{op}$). We suggestively let $R_{a,b} : a_g \otimes b_h \simeq T_g(b_h) \otimes a_g$ be a collection of isomorphisms, called the *G-braiding.*

The simplest example is, as pointed out earlier, that of bosonic SPTs. In these theories we take $\mathcal{C}_e$ to consist of a single trivial boson with trivial G-action. This implies that there is one g-invariant topological charge (the boson) for each $g \in G$, and so each $\mathcal{C}_g$ consists of a single defect labeled by the associated group element.

Theorem 1. *The equivariantization and de-equivariantization processes establish a bijection between the set of equivalence classes of braided G-crossed fusion categories and the set of equivalence classes of braided fusion categories containing* $\mathrm{Rep}(G)$ *as a symmetric fusion subcategory.*

Recall that "symmetric fusion subcategory" means a braided category where all topological charges are mutually bosonic (i.e., if $c_{X,Y} : X \otimes Y \mapsto Y \otimes X$ is the braiding, then $c_{X,Y} \circ c_{Y,X} = \mathbf{1}_{X\otimes Y}$). This is precisely the criteria a subcategory of anyons must have to be condensable. So translating this result, we have that the number of theories that posses a subalgebra $\mathrm{Rep}(G)$ of anyons which can be condensed is precisely equal to the number of braided G-crossed fusion categories. The bijection is established by realizing that condensing a subcategory of anyons is exactly inverse to gauging (G-crossing) the symmetry group carried by those anyons. Very powerful!

A little bit about Hopf algebras

There are two main ways of approaching Hopf algebras. One is in terms of convolutions over algebras (a la Kassel's book), and one is in terms of endomorphisms of fiber functors (a la Etingof's book). We will choose the latter approach. It is more abstract, but ties in more smoothly with the flavor of the rest of these notes.

Bialgebras

We've already talked about monoidal functors and tensor functors, and saw that they were essentially the same thing. There's a third related type of functor called a fiber functor, which we need to define before beginning our discussion of Hopf algebras in earnest. As was the case with monoidal and tensor functors, it's actually a functor + natural isomorphism pair.

Definition 13. A *quasi-fiber functor* on a ring category $\mathcal{C}$ over the field $\mathbb{K}$ is an exact functor $F : \mathcal{C} \to \mathbf{Vec}$ with $F(\mathbf{1}) = \mathbb{K}$, together with a natural isomorphism

$$J : F(-) \otimes F(-) \to F(- \otimes -). \tag{631}$$

If J satisfies the axioms for being a monoidal functor, we just call F a fiber functor.

Obviously the forgetful functors $\mathbf{Vec}_G \to \mathbf{Vec}$ and $\mathrm{Rep}(G) \to \mathbf{Vec}$ are fiber. $\mathcal{C}_G^\omega \to \mathbf{Vec}$ is always quasi-fiber, but never fiber, which follows from our earlier results on monoidal functors.

Let F be a fiber functor with tensor structure J, as in the definition. Also, define the algebra $H = \mathrm{End}(F)$. By $\mathrm{End}(F)$, we mean "the algebra of functorial endomorphisms of F", by which we mean the set of natural transformations of F to itself. We will define co(stuff) on this algebra in the natural way, by introducing the coproduct $\Delta : H \to H \otimes H$ and counit $\varepsilon : H \to \mathbb{K}$. The definition of the coproduct is kind of a pain. We let

$$\alpha : \mathrm{End}(F) \otimes \mathrm{End}(F) \to \mathrm{End}(F \boxtimes F) \tag{632}$$

be the natural isomorphism associated with Δ. We can then do the coproduct by (for $a \in H$)

$$\Delta(a) = \alpha^{-1}(\widetilde{\Delta}(a)), \tag{633}$$

where $\widetilde{\Delta}(a) \in \mathrm{End}(F \boxtimes F)$ is given component-wise as

$$\widetilde{\Delta}(a)_{X,Y} = J^{-1}_{X,Y} a_{X \otimes Y} J_{X,Y}. \tag{634}$$

The counit is easy: we let $\varepsilon(a) = a_{\mathbf{1}} \in \mathbb{K}$. Δ and ε give H the structure of a coalgebra (after checking associativity, etc.). But H is also an algebra, and being both an algebra and a coalgebra we call it a *bialgebra*.

This is rather abstract, so let's go over an easy example.

Example 7. Let G be a group, and let $\mathcal{C} = \mathbf{Vec}_G$. Let F be the obvious forgetful fiber functor $F : \mathcal{C} \to \mathbf{Vec}$. In this case, the endomorphisms of F are just functions from G to $\mathbb{K}$, and so $H = \mathrm{End}(F) = \mathbf{Fun}(G, \mathbb{K})$. The coproduct needs to map into $H \otimes H = \mathbf{Fun}(G^2, \mathbb{K})$, and so we can define the coproduct as

$$\Delta(f)_{x,y} = f(xy), \quad x, y \in G, \tag{635}$$

where $f \in \mathrm{End}(F)$. The counit is (of course) just $\varepsilon(f) = f(1)$.

Bialgebras are "self-dual" in a natural way (the dual of the product is the coproduct, etc), and so we can look at the dual of $H = \mathbf{Fun}(G, \mathbb{K})$. Just as in regular linear algebra, we have $H^* = \mathbb{K}G$. The coproduct in H^* is $\Delta(g) = g \otimes g$ and the counit is $\varepsilon(g) = 1$, for $g \in G$.

Given H, we can form the category $\mathrm{Rep}(H)$ in the usual way, which inherits the structure of a monoidal category. To tensor two H-modules (representations) X and Y, we just use the regular tensor product of vector spaces. The action of H is defined through the "factor set"

$$\rho_{X\otimes Y}(a) = (\rho_X \otimes \rho_Y)(\Delta(a)), \tag{636}$$

with $a \in H$ and $\rho_X : H \to \mathrm{End}(X)$ (and the same with ρ_Y). We will avoid going down the rabbit hole, and will just state the following theorem (the *Reconstruction Theorem*), which should be reasonable given the last few paragraphs:

Theorem 2. *Let $F : \mathcal{C} \to$ **Vec** be a fiber functor, and H be a bialgebra (over $\mathbb{K}$). Then the assignments*

$$(\mathcal{C}, F) \mapsto H = \mathit{End}(F), \qquad H \mapsto (\mathrm{Rep}(H), \mathit{Forget} : \mathrm{Rep}(H) \to \textbf{\textit{Vec}}) \tag{637}$$

are mutually inverse bijections between pairs $(\mathcal{C}, F)$ (up to tensor equivalence) and isomorphism classes of bialgebras H over $\mathbb{K}$.

Hopf algebras

The extra piece of information we need to define a Hopf algebra is a linear endomorphism called the antipode. Its definition will (probably) seem pretty arbitrary at first, but we'll see eventually why having an antipode makes Hopf algebras nice to work with.

Definition 14. Let $H = \mathrm{End}(F)$ be a bialgebra, with F a fiber functor $F : \mathcal{C} \to \mathbf{Vec}$ and $\mathcal{C}$ a tensor category. The *antipode* is a linear map $S : H \to H$ defined by

$$S(a)_X = a^*_{X^*} \tag{638}$$

where we have used the natural equivalence between $F(X)^*$ and $F(X^*)$.

The antipode should be thought of as a sort of "inverse" in H. Just like like the inverse, it is an anti-homomorphism, in the sense that $S(ab) = S(b)S(a)$ and $\Delta(S(a)) = (S \otimes S)\tau\Delta$, where $\tau(a \otimes b) = b \otimes a$ is the switch map.

Proposition 8. *Let $\mu : H \otimes H \to H$ be the multiplication on H and $\eta : \mathbb{K} \to H$ be the unit map of H. Then the antipode satisfies*

$$\mu \circ (\mathbf{1} \otimes S) \circ \Delta = \mu \circ (S \otimes \mathbf{1}) \circ \Delta = \eta \circ \varepsilon. \tag{639}$$

The proof is surprisingly long (and boring), and so we'll omit it. Another long and boring proof shows that the antipode (if it exists) is unique.

We will also briefly mention an alternative way to define the antipode. Let $(H, \mu, \eta, \Delta, \varepsilon)$ be a bialgebra, and write convolution over H as

$$(f \star g)(x) = \sum_i f(x_{(1)i})g(x_{(2)i}). \tag{640}$$

Then the antipode is the (still unique) endomorphism $S : H \to H$ which satisfies

$$S \star \mathbf{1} = \mathbf{1} \star S = \eta \circ \varepsilon. \tag{641}$$

In any case, we can finally define a Hopf algebra:

Definition 15. A *Hopf algebra* is a bialgebra with an invertible antipode.

Example 8. Let's revisit the example of the bialgebra of functions considered earlier. $\mathbf{Fun}(G, \mathbb{K})$ is a Hopf algebra iff G is a group. The antipode in this case is $S(f)(g) = f(g^{-1})$. We can also look at the "dual" Hopf algebra $\mathbb{K}G$ – the antipode in this case is $S(g) = g^{-1}$, in analogy to the antipode in $\mathbf{Fun}(G, \mathbb{K})$.

The quantum double

The quantum double is a (non-obvious and simpler) reformulation of the center construction $\mathcal{Z}(\mathcal{C})$ for the case when $\mathcal{C}$ is the category of finite-dimensional representations of a Hopf algebra. There are two ways of defining the quantum double. One generalizes the notion of G-crossed categories familiar from previous discussions, and one uses a sort of reduction process from the center of $\mathcal{C}$. We'll start with the former approach.

We talked about the crossed product of a category with a group earlier (it looked like $\mathcal{C} \rtimes G$). We now need to generalize this to a similar operation called the bicrossed product (written as $\bowtie$) which (as the notation suggests) is a generalization of the semidirect product. The complete formulation of bicrossed products of bialgebras is pretty involved, and we won't talk about it. However, we can get a pretty good feel for what bicrossed products are by looking at bicrossed products of groups.

Let G be a group, and H and K be subgroups of G. We assume that for any $x \in G$, there is a a unique pair $(y, z) \in H \times K$ which satisfy $x = yz$. ...

Proposition 9. *Let (H, K) be a matched pair of groups. There is a unique group structure $H \bowtie K$ (the bicrossed product) with the multiplication rule*

$$(x, y)(x', y') = (x(y \cdot x'), y^{x'} y'), \quad x \in H, y \in K. \tag{642}$$

In the proposition, $\cdot$ denotes a group action of K on H and powers denote an action of H on K. Of course, the unit in $H \bowtie K$ is $(1, 1)$.

Example 9. Let H and K act trivialy on each other. In this case, the bicrossed product is isomorphic to the direct product: $H \bowtie K \simeq H \times K$. On the other hand, suppose we keep the trivial action of H on K but let K act on H by group automorphisms:

$$y \cdot (xx') = (y \cdot x)(y \cdot x'), \quad y \cdot 1 = 1. \tag{643}$$

In this case, it's easy to see that $H \bowtie K \simeq H \rtimes K$. Indeed, $H \times \{1\}$ is a normal subgroup of $H \bowtie K$ and the action of K on H is conjugation in $H \bowtie K$.

Generalizing the bicrossed product to Hopf algebras is straightforward, but a pain. We can imagine that all the multiplications get replaced with convolutions – this is indeed what happens. Let's skip the details and rely on intuition from the finite group case to get us through.

As mentioned earlier, Hopf algebras are "self-dual" in a natural way. Namely, given a Hopf algebra $(H, \mu, \eta, \Delta, \varepsilon, S)$, we can define $(H^{op})^*$ as the Hopf algebra $(H^*, \Delta^*, \epsilon, (\mu^{op})^*, \eta, (S^{-1})^*)$ – the coproduct becomes the product, the (opposite) product becomes the coproduct, etc. The quantum double of H is defined as follows:

Definition 16. The *quantum double* $D(H)$ of a Hopf algebra H is the bicrossed product of H and $(H^{op})^*$. That is,

$$D(H) = (H^{op})^* \bowtie H. \tag{644}$$

This line of reasoning allows us to describe the universal R-matrix in a Hopf algebra setting. Let $\{e_i\}$ form a basis of H, with $\{e^i\}$ the corresponding dual basis. Knowing that the R-matrix describes how objects in $D(H)$ braid with each other, it should be intuitively plausible that $R \in D(H) \otimes D(H)$ is defined through

$$R = \sum_i (\mathbf{1} \otimes e_i) \otimes (e^i \otimes \mathbf{1}). \tag{645}$$

Example 10. Let $H = \mathbb{K}G$, with G finite. Let $\{g\}_{g\in G}$ be the basis of H and $\{e_g\}_{g\in G}$ be the corresponding dual basis. The dual algebra $(H^{op})^*$ is then $\mathbb{K}^G$, with multiplication $e_g e_h = \delta_{gh} e_g$. The unit is the average of the dual basis: $1 = \sum_{g\in G} e_g$. The counit, comultiplcation, and the antipode in $(H^{op})^*$ are

$$\Delta(e_g) = \sum_{hk=g} e_h \otimes e_k, \quad \varepsilon(g) = \delta_{g1}, \quad S(e_g) = e_{g^{-1}}. \tag{646}$$

We see that a basis for $D(H)$ is the set $\{e_g h | (h, g) \in G^2\}$. Since $D(H)$ is the bicrossed product of H with itself, we can write multiplication in H through the relation

$$h e_g = e_{hgh^{-1}} h, \tag{647}$$

i.e. the algebra acts on itself by conjugation. The R-matrix in this scenario is just

$$R = \sum_{g\in G} g \otimes e_g. \tag{648}$$

That is, we can interpret $\mathcal{C} \rtimes G$ as a sort of 'half-gauging', since the symmetry fluxes are introduced into the system but remain confined. When we do $\mathcal{C} \bowtie G$, we fully gauge the symmetry and the symmetry fluxes become deconfined.

The quantum double construction descends from the center construction in a natural way. To see this, let H be a Hopf algebra and $\mathcal{C}$ the be category $\mathrm{Rep}(H)$. We use the reconstruction theorem to derive the existence of the canonical fiber functor $F : \mathcal{C} \to \mathbf{Vec}$ with $H = \mathrm{End}(F)$. We then compose this F with the forgetful functor $F' : \mathcal{Z}(\mathcal{C}) \to \mathcal{C}$ to obtain a tensor functor $\widetilde{F} : \mathcal{Z}(\mathcal{C}) \to \mathbf{Vec}$. Again using the reconstruction theorem, we obtain a Hopf algebra $D = \mathrm{End}(\widetilde{F})$. This D is precisely the quantum double of H.

Lefschetz duality

Today we're going to explain the analogue of Poincare duality for manifolds with boundary. Recall that relative cycles $M \in C_\bullet(Y, \partial Y)$ are those cycles whose boundaries are contained

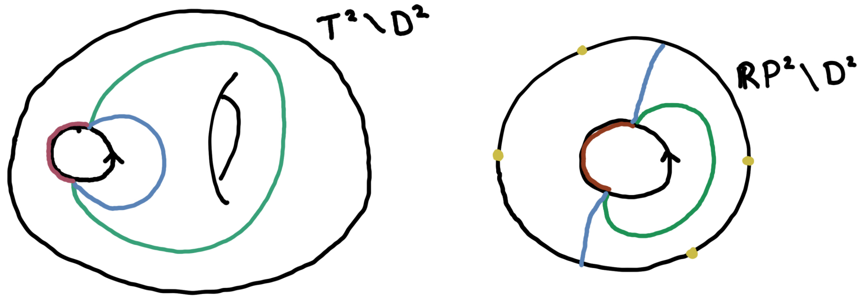


Figure 4: Some clumsily-drawn pictures that might aid in visualization. On the left we have a T^2 with a disk cut out (the boundary of the manifold is the oriented circle). The blue curve is trivial in $H_1(Y, \partial Y)$; the green curve is not. The restrictions of both curves to the boundary are trivial in $H_0(\partial Y)$, as both are boundaries of the red curve, which is contained in ∂Y. The right figure shows basically the same thing for $\mathbb{R}P^2 \setminus D^2$.

within Y, while relative cocycles $\alpha \in C^\bullet(Y, \partial Y)$ are those which evaluate to zero when restricted to ∂Y. The goal will be to show that on a manifold Y of dimension d,

$$H^p(Y, \partial Y; R) \cong H_{d-p}(Y; R), \qquad H^p(Y; R) \cong H_{d-p}(Y, \partial Y; R). \tag{649}$$

◇ ◇ ◇

Just by using intuition from Poincare duality, it is easy to intuit why the above isomorphisms hold. Consider e.g. figure 4. The blue curve on the left is a relative boundary, and it is clear that while its Poincare dual (obtained in the usual way of employing bump functions) 1-cochain is closed, the dual does not vanish on the boundary of $Y = T^2 \setminus D^2$. The green curve is not a relative boundary, and is nontrivial in relative homology—its dual is closed, but again, does not vanish on ∂Y. This tells us that we should be relating relative homology and regular cohomology, *not* relative cohomology. On the other hand, if a cycle is nontrivial in regular homology then it does not end on the boundary, and so its Poincare dual will vanish on the boundary—therefore we should be relating regular homology and relative cohomology.

Now to substantiate these guesses. As in the proof of Poincare duality, the main technical tool that will be used is the cap product. First let's recall what the cap product is, and how we use it to get Poincare duality—for now we take $\partial Y = 0$. The coefficient ring for our simplicial needs will be denoted as R; we will only indicate it explicitly when necessary.

The cap product is a way of integrating q-cochains over $(p > q)$-dimensional chains / submanifolds: the integration eats up q of the dimensions of the chain, and produces a linear combination of $(p-q)$-dimensional chains. It is easiest to describe at the level of simplices. Let $v_0, \dots, v_p$ be a p-simplex and and let $M|_{v_0,\dots,v_q}$ denote the restriction of M to the part of the simplex specified by the coordinates $v_0, \dots, v_q$. Then the cap product acts as

$$\cap : C_p(Y) \times C^q(Y) \to C_{p-q}(Y), \qquad (M \cap \alpha)|_{v_q,\dots,v_p} = \alpha(M|_{[v_0,\dots,v_q})M|_{v_q,\dots,v_p]}, \tag{650}$$

with $M \in C_p(Y), \alpha \in C^q(Y)$, and with the brackets denoting antisymmetrization. One can then check that[54]

$$\partial(M \cap \alpha) = (-1)^{|\alpha|}(\partial M \cap \alpha - M \cap d\alpha). \tag{653}$$

This means that if M is a cycle and α is a cocycle then $M \cap \alpha$ is a cycle. Furthermore, if $M = \partial N$ is a boundary and α is a cocycle, then $M \cap \alpha = (-1)^{|\alpha|}\partial(N \cap \alpha)$ is trivial in homology. Likewise, if $\alpha = d\beta$ is a coboundary then $M \cap \alpha = (-1)^{|\alpha|+1}\partial(M \cap \beta)$ is also trivial in homology. These three facts that the cap product actually extends to an operation on (co)homology:

$$\cap : H_p(Y) \times H^q(Y) \to H_{p-q}(Y). \tag{654}$$

Setting $p = q$ gives the usual integration pairing, while setting $p = d$ with $d = \dim Y$ gives the usual Poincare duality $H^q(Y; R) \cong H_{d-q}(Y; R)$ (since $H_d(Y; R) \cong R$ is generated just by the fundamental class).

Modifying this argument to the case where ∂Y is non-empty is straightforward. Suppose $M \in H_p(Y, \partial Y)$ so that $\partial M = N \subset \partial Y$, and $\alpha \in H^q(Y)$. By the rule (653) for the cap product, we have (up to a minus sign that won't affect things)

$$\partial(M \cap \alpha) = N \cap \alpha \implies M \cap \alpha \in H_{p-q}(Y, \partial Y), \tag{655}$$

but that $M \cap \alpha \notin H_{p-q}(Y)$, in general. Therefore the cap product might give a pairing $H_\bullet(Y, \partial Y) \times H^\bullet(Y) \to H_\bullet(Y, \partial Y)$, but it will not give a pairing if we replace the RHS with $H_\bullet(Y)$.

Now suppose that $M = \partial N + L$, with $L \subset \partial Y$. Then if we still take $\alpha \in H^q(Y)$, we have (again up to a minus sign; we will ignore these annoying signs from now on since they don't affect the argument)

$$M \cap \alpha = (\partial N + L) \cap \alpha = L \cap \alpha + \partial(N \cap \alpha) \sim 0, \tag{656}$$

where the last step means as a class in $H_\bullet(Y, \partial Y)$. Finally, if $\partial M = N \subset \partial Y$ so that $M \in H_\bullet(Y, \partial Y)$ and if $\alpha = d\beta$, then

$$M \cap \alpha = \partial(M \cap \beta) + N \cap \beta \sim 0, \tag{657}$$

which is again trivial in $H_\bullet(Y, \partial Y)$. Putting this paragraph and the last paragraph together, we see that the cap product gives a well-defined pairing

$$\cap : H_p(Y, \partial Y) \times H^q(Y) \to H_{p-q}(Y, \partial Y). \tag{658}$$

[54] Recall that simplicially, boundary operators act by

$$(\partial M)|_{v_0,\dots,v_{p-1}} = \sum_i (-1)^i M|_{v_0,\dots,\widehat{v}_i,\dots,v_p}, \tag{651}$$

while coboundary operators do

$$(d\alpha)(N|_{v_0,\dots,v_{q+1}}) = \sum_i (-1)^i \alpha(N|_{v_0,\dots,\widehat{v}_i,\dots,v_q}) \implies (d\alpha)(N) = \alpha(\partial N). \tag{652}$$

Plugging these formulas into the boundary of $M \cap \alpha$ gives the desired result.

Setting $p = d$ so that the left relative homology group is generated just by the fundamental class $[Y]$, we see that capping with $[Y]$ produces an isomorphism

$$H^q(Y) \cong H_{d-q}(Y, \partial Y). \tag{659}$$

This gives us half of what we wanted to show. Next, we will show that the cap product also gives a well-defined pairing

$$\cap : H_p(Y, \partial Y) \times H^q(Y, \partial Y) \to H_{p-q}(Y). \tag{660}$$

Note that we still have the relative homology on the LHS! This is because we'll need to take $p = d$ to get the second equation of Lefschetz duality, and $H_d(Y)$ is trivial if $\partial Y \neq 0$ (since the fundamental "class" is duh, not closed).

Anyway, let's check that this pairing works. First, if $M \in H_p(Y, \partial Y)$ with $\partial M = N \subset \partial Y$ and $\alpha \in H^q(Y, \partial Y)$, then

$$\partial(M \cap \alpha) = N \cap \alpha = 0, \tag{661}$$

since α, being a relative cochain, kills chains that live in ∂Y. Secondly, if $M = \partial N + L$ with $L \subset \partial Y$, then

$$M \cap \alpha = \partial(N \cap \alpha) + L \cap \alpha \sim 0, \tag{662}$$

where the equivalence is now in $H_{p-q}(Y)$, since the $L \cap \alpha$ term vanishes exactly. Finally, if again $\partial M = N \subset \partial Y$ but now $\alpha = d\beta$ with $\beta \in C^{q-1}(Y, \partial Y)$, then

$$M \cap \alpha = \partial(M \cap \beta) + N \cap \beta \sim 0, \tag{663}$$

since $N \cap \beta = 0$. These three facts mean that the pairing we wrote down in the previous paragraph is indeed correct. Setting $p = d$ as usual then gives us the remaining isomorphism

$$H^q(Y, \partial Y) \cong H_{d-q}(Y). \tag{664}$$

Outer automorphism bootcamp

The physics motivation for this diary entry was trying to learn about what the definition of charge conjugation symmetry actually is. It turns out that $\mathbb{Z}_2^C$ is an outer automorphism of the symmetry group that the fields in question transform under (well, the part of the symmetry group that doesn't include the $\mathbb{Z}_2^C$ factor itself), and hence the discussion of outer automorphisms in today's diary entry.

◇ ◇ ◇

An outer automorphism of G is one that can't be generated via conjugation by elements in G. Often (but not always) people actually define one as an equivalence class of such automorphisms, where we mod out by ones generated by conjugation: if $\mathrm{Inn}(G)$ is the group of inner automorphisms, viz. those generated by conjugation by elements in G, then

$$\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G). \tag{665}$$

This is the definition we'll use.

When we consider the representation theory of G, then $U \in \mathrm{Out}(G)$ may be defined to act by (since we are interested in physics applications, we will only consider unitary representations)

$$U^\dagger \rho_{R'}(g) U = \rho_R(U(g)). \tag{666}$$

The representations here are allowed to be different (although we must at least have $\dim R = \dim R'$), since we do not have $U = R''(g)$ for any representation R''. Now if $U \in \mathrm{Inn}(G)$ is a trivial outer automorphism, then we must have $R = R'$: indeed, if $U(g) = h^{-1}gh$ then we have

$$\rho_{R'}(g) = (R(h)U)^\dagger \rho_R(g)(R(h)U) \implies R' \cong R. \tag{667}$$

However, if $R \cong R'$, then we do not necessarily have $U \in \mathrm{Inn}(G)$ (for example, $\mathrm{Out}(SO(4)) \neq 0$ even though the vector representation must get mapped to itself).

A general math fact is that for a semisimple Lie group G, $\mathrm{Out}(\widetilde{G})$ can be determined from the graph automorphisms of the Dynkin diagram of $\mathfrak{g}$, where $\widetilde{G}$ is the universal cover.[55] Since most dynkin diagrams have at most a $\mathbb{Z}_2$ reflection symmetry (except for $\mathfrak{so}(8)$), an outer automorphism of the symmetry group, if it exists, can usually be interpreted as some sort of $\mathbb{Z}_2$ "charge conjugation" action. From this we can conclude that all the $SU(n)$s for $n > 2$ (whose Dynkin diagrams are just straight lines) have a $\mathbb{Z}_2$ charge-conjugation symmetry, coming from flipping the diagram by exchanging the first and last nodes. The $\mathfrak{so}(2n)$ Dynkin diagrams have a forked split end which gives a $\mathbb{Z}_2$ reflection automorphism, while for $\mathfrak{so}(2n+1)$ we have a straight chain with a double-linked end which gives no nontrivial automorphisms. This tells us that

$$\mathrm{Out}(\mathrm{Spin}(2n)) = \mathbb{Z}_2, \quad \mathrm{Out}(\mathrm{Spin}(2n+1)) = \mathbb{Z}_1. \tag{668}$$

Another important example is $\mathrm{Out}(Sp(n)) = 0$, since the Dynkin diagram again has a double-line that spoils any reflection symmetry.

We will often want to know $\mathrm{Out}(G)$, with $G = \widetilde{G}/\Gamma$ and Γ some discrete subgroup of the center of $\widetilde{G}$. $\mathrm{Out}(G)$ can be determined from $\mathrm{Out}(\widetilde{G})$ as follows: let $U \in \mathrm{Aut}(\widetilde{G})$. Then U will restrict to an automorphism of G if $U(\Gamma) = \Gamma$. This is because then for any $g \in G$, $U(g\Gamma) = U(g)\Gamma$: therefore if we change the preimage by an element in Γ the image also

[55] "Proof": consider a choice of maximal torus T. Then if $g, g' \in T$, the images of g and g' under any automorphism of G must also commute. This means that all automorphisms map maximal tori to maximal tori. Since all maximal tori are equivalent up to an inner automorphism, any outer automorphism can be chosen to preserve the maximal torus. The whole notion of roots and weights and so on comes from choosing a maximal torus, and so the structure of the roots and weights will be preserved by any outer automorphism. Therefore any outer automorphism must act as a graph automorphism on the Dynkin diagram. That these graph automorphisms exhaust all outer automorphisms can be motivated by noting that inner automorphisms all act trivially on the Dynkin diagrams since they do not permute representations.

changes by an element in Γ, and so the induced map on equivalence classes of elements in $\widetilde{G}/\Gamma$ is well-defined. Therefore $\mathrm{Out}(G)$ is the subgroup of $\mathrm{Out}(\widetilde{G})$ that preserves Γ.

If $\Gamma = Z(\widetilde{G})$ then all automorphisms restrict to automorphisms of $\widetilde{G}/\Gamma$, since all automorphisms preserve the center. This means that e.g. for $n > 2$,

$$\mathrm{Out}(SU(n)) = \mathrm{Out}(PSU(n)) = \mathbb{Z}_2, \tag{669}$$

since in that case $\Gamma = \mathbb{Z}_n$ is the whole center. This means that theories with a $PSU(N)$ internal symmetry should still possess a notion of charge conjugation, even though the fundamental representation of $PSU(N)$ is self-dual.

In the case of $SO(2n)$, the kernel of $\mathrm{Spin}(2n) \to SO(2n)$ does not involve the entire center, so we need to be more careful.[56] Now the generator of $\mathrm{Out}(\mathrm{Spin}(2n)) = \mathbb{Z}_2$ exchanges the L and R spin representations,[57] but the kernel of the projection $\pi : \mathrm{Spin}(2n) \to SO(2n)$ is precisely the element which is assigned to $-\mathbf{1}$ by both L and R representations, and so the kernel is left invariant under the $L \leftrightarrow R$ switch. Therefore taking the quotient doesn't actually change the outer automorphism group, and we have

$$\mathrm{Out}(SO(n)) = \begin{cases} \mathbb{Z}_2 & \text{if } n \in 2\mathbb{N},\ (n > 2,) \\ \mathbb{Z}_1 & \text{if } n \in 2\mathbb{N}+1 \end{cases} \tag{671}$$

The outer automorphisms here are the reflections that extend $SO(2n)$ to $O(2n)$, which are e.g. generated by $U = -1 \oplus \mathbf{1}_{2n-1}$. The fact that the reflection action is the same thing as the exchange of the L and R spin representations is because $\bar{\gamma}$ anticommutes with reflections, since they are generated either by γ_μ or $\bar{\gamma}\gamma_\mu$. The reason that $\mathrm{Out}(SO(2n+1)) = 0$ is because in that case we can take $U = -\mathbf{1}$ and so we always have $U^\dagger \rho(g) U = \rho(g)$. Another way of saying this is to say that $O(2n) = SO(2n) \rtimes \mathbb{Z}_2$ but $O(2n+1) = SO(2n+1) \times \mathbb{Z}_2$, so that in the later case the semidirect product is trivial and there are no outer auts. Anyway, note again that if $n \in 2\mathbb{N}$ we have a nontrivial charge conjugation, even though the fundamental representation of $SO(2n)$ is self-dual.

Let's now look at some easy examples. First look at $SU(N > 2)$. The outer automorphism is complex conjugation, and can be constructed quite explicitly by writing $SU(N)$ as a subgroup of $O(2N)$. For any $\mathcal{U} \in SU(N)$, we factor it into real and imaginary parts and then write $i \in \mathbb{C}$ as J:

$$\mathcal{U} = \mathcal{U}_{\mathbb{R}} \otimes \mathbf{1} + \mathcal{U}_{\mathbb{I}} \otimes J, \tag{672}$$

where both $\mathbf{1}$ and J are 2×2. We see that $\mathbb{C}$ conjugation is then performed with the matrix $\mathbf{1}_N \otimes Z$.[58] When we include this, we can write the full group as $SU(N) \rtimes \mathbb{Z}_2^C$, with a generic

[56] $Z(\mathrm{Spin}(4n+2)) = \mathbb{Z}_4$ and $Z(\mathrm{Spin}(4n)) = \mathbb{Z}_2^2$, while $Z(SO(2n)) = \mathbb{Z}_2$.

[57] The case of $\mathfrak{so}(8)$ is an exception because of triality: the symmetries of the $\mathfrak{so}(8)$ Dynkin diagram tell us that

$$\mathrm{Out}(\mathrm{Spin}(8)) \cong S_3. \tag{670}$$

At the level of representations, the triality is possible since both the L and R spinor representations have the same dimension as the vector representation (viz. $2^4/2 = 8$), and so we can find maps which permute these three representations among themselves. However, not all these outer automorphisms descend to $\mathrm{Out}(SO(8))$: anything in $\mathrm{Aut}(SO(8))$ must leave the vector representation invariant, which leaves us with only $\mathrm{Out}(SO(8)) = \mathbb{Z}_2$ coming from exchange of the L and R irreps.

[58] Or the matrix $\mathbf{1}_N \otimes X$. These two choices are related by $\mathbf{1}_N \otimes J$, which is only in $SU(N)$ if $N \in 4\mathbb{Z}$, and so for different values of N, we may have distinct ways of performing the charge conjugation.

element $\mathcal{V} \in SU(N) \rtimes \mathbb{Z}_2^C$ being written as $(\mathcal{U}, n) = \mathcal{U}(\mathbf{1}_N \otimes Z^n)$, $n \in \{0,1\}, \mathcal{U} \in SU(N)$. Then we check that the multiplication law is

$$(\mathcal{U}, n) \cdot (\mathcal{U}', n') = (\mathcal{U}\rho_n[\mathcal{U}'], [n+n']_2), \qquad \rho_n(\mathcal{U}) = \begin{cases} \mathcal{U} & \text{if } n = 0 \\ \mathcal{U}^* & \text{if } n = 1, \end{cases} \tag{673}$$

which is of course befitting of a $\rtimes$ product.

A more interesting example is $SU(2)$: the Dynkin diagram for $SU(2)$ is just a single node, and so according to the above fact, $\mathrm{Out}(SU(2)) = 0$. This of course is consistent with $O(3) = SO(3) \times \mathbb{Z}_2$ implying $\mathrm{Out}(SO(3)) = 0$, which by the above statement means $\mathrm{Out}(SU(2)) = 0$ since $\mathrm{Out}(\widetilde{G})$ is determined by information contained in $\mathfrak{g}$. But it's instructive to examine this more closely without appealing to the Dynkin diagram theorem. For example, naively one might think that a reflection of one of the coordinates in $\mathbb{R}^4$ would generate a nontrivial element of $\mathrm{Out}(SU(2))$ through its action on S^3 (this is what happens for the S^1 case, after all). If $SU(2)$ really has no outer automorphisms, then such a reflection must be generated by conjugation in $SU(2)$. That this is possible is most easily seen by using the presentation

$$SU(2) \ni \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}, \qquad |z|^2 + |w|^2 = 1. \tag{674}$$

A reflection of one of the coordinates maps e.g.

$$R : \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} \mapsto \begin{pmatrix} \bar{z} & w \\ -\bar{w} & z \end{pmatrix}. \tag{675}$$

Letting $w = re^{i\phi}$, one can then check that this is realized by conjugation

$$R : \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} \mapsto U^\dagger \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} U, \qquad U = \begin{pmatrix} 0 & e^{i\phi} \\ -e^{-i\phi} & 0 \end{pmatrix} \in SU(2). \tag{676}$$

So indeed, this reflection is actually an inner automorphism.

Determining the possible group extensions of $\mathbb{Z}_m$ by $\mathbb{Z}$

Today we will examine possible central extensions

$$1 \to \mathbb{Z} \to E \to \mathbb{Z}_m \to 1, \tag{677}$$

We won't make the assumption that E is Abelian, although we will see that it must be. Indeed, we claim that

Theorem 3. *The only groups E fitting into the exact sequence*

$$1 \to \mathbb{Z} \to E \to \mathbb{Z}_m \to 1 \tag{678}$$

are of the form

$$E \cong \mathbb{Z} \times \mathbb{Z}_k, \tag{679}$$

where k is a divisor of m.

If we were to replace the $\mathbb{Z}_m$ in the above with a general finite group G, the statement would then be that $E \cong F \times \mathbb{Z}$, with F some finite group.

Proof. The following lemma gets us most of the way to the proof of the theorem:

Lemma 1. Let E be a group which contains an infinite cyclic central group C of finite index

$$[E : C] = n. \tag{680}$$

Then

$$E \cong F \rtimes \mathbb{Z}, \tag{681}$$

with F finite (but not necessarily of order n).

Proof. Let the quotient group E/C be denoted by G, with $|G| = n$. Consider the SES $1 \to C \to E \to G \to 1$. We claim that multiplication by n in C gives us the diagram

$$\begin{tikzcd} 1 \arrow[r] & C \arrow[r] \arrow[d, "\times n"] & E \arrow[r] \arrow[d, "\psi"] & G \arrow[r] \arrow[d, equal] & 1 \\ 1 \arrow[r] & C \arrow[r] & C \times G \arrow[r] & G \arrow[r] & 1 \end{tikzcd} \tag{682}$$

where the rows are exact. The fact that a homomorphism $A \to B$ induces a commutative diagram of the above form between the SESs $1 \to A \to E \to G \to 1$ and $1 \to B \to E' \to G \to 1$ is in Brown, exercise 1b in chapter 4.3. The reason why the bottom-central term is $C \times G$ is because $|H^2(G; C)| = |G| = n$,[59] so that multiplication by n sends any extension to the trivial extension. Now since the diagram commutes, the left square tells us that the homomorphism ψ acts as the nth power map ("multiplication" by n is the precise terminology here since we aren't assuming E is Abelian) when its domain is restricted to $C \subset E$. We can then form a surjection $\phi : E \to C$ by restricting the image of ψ to $C \subset C \times G$—that is, we form ϕ by simply taking the outputs of ψ and setting all the G-components to the identity (note that just a projection onto the C factor in E isn't a homomorphism unless E is a direct product). We therefore have a surjection $\phi : E \to \mathbb{Z} \cong nC$. The kernel $F \equiv \ker(\phi)$ of this surjection is finite, since the image of ϕ is infinite cyclic, and so $E/\ker(\phi)$ needs to be infinite cyclic; since E has only one independent infinite cyclic generator on account of the index $[E : C]$ being finite, $\ker(\phi)$ cannot contain an infinite cyclic generator, and hence must be finite.

Now we make use of the following:

Proposition 10. *Suppose we have a surjection $\phi : E \to \mathbb{Z}$ with finite kernel. Then E must split as $E \cong F \rtimes \mathbb{Z}$ for some finite F.*

Proof. Since $\ker(\phi)$ is always normal in E, we can construct the extension

$$1 \to \ker(\phi) \to E \to \mathbb{Z} \to 1. \tag{683}$$

[59]This is proved in Brown, chapter 4.3.

We claim that this extension must be trivial (i.e. E must split as a $\rtimes$ product), since $H^2(\mathbb{Z}; \ker(\phi)) = 1$ is trivial. Intuitively, this is because $H^2(G; A)$ classifies projective representations, which are classified by the different ways of fractionalizing the relations among the generators of G. If $G = \mathbb{Z}$ so that there are no such relations to fractionalize, there must consequently be no projective representations. There are a variety of ways to prove this more formally; one is to use $B\mathbb{Z} = U(1)$ so that the group cohomology $H^2(\mathbb{Z}; A)$ is equal to the simplicial cohomology $H^2(U(1); A) = 1$ on account of $B\mathbb{Z} = U(1)$. □

We can put this to use since we have a surjection $\phi : E \to \mathbb{Z}$ derived above. The surjection must therefore split, so that E may be written as

$$E \cong F \rtimes \mathbb{Z} \tag{684}$$

for some finite F. This proves the lemma. □

Now we can finish off the proof of the main theorem. First, we can show that the action in the $\rtimes$ must in fact be trivial, so that in fact $E \cong \mathbb{Z} \times F$. Indeed, this is done by using the fact that the (A-conjugacy classes) of splittings of the split extension $1 \to A \to A \rtimes G \to G \to 1$ are in 1-1 correspondence with elements of $H^1(G; A)$.[60] Applying this to our case, we see that the distinct splittings are classified by $H^1(\mathbb{Z}_m; \mathbb{Z}) = 1$ (this is just saying that $\mathrm{Hom}(\mathbb{Z}_m; \mathbb{Z})$ is trivial), and so evidently no nontrivial semi-direct products are possible. Hence we in fact must have

$$1 \to \mathbb{Z} \xrightarrow{\gamma} \mathbb{Z} \times F \xrightarrow{\alpha} \mathbb{Z}_m \to 1. \tag{685}$$

All that remains is to show that $F = \mathbb{Z}_k$ for some $k|m$. This is easy, though. $\mathbb{Z} \times F$ is generated by the image of 1 under γ and by the pre-image of 1 under α and therefore only has two generators, so that F can only have a single generator and hence $F = \mathbb{Z}_k$ for some k. Restricting α to F gives us a homomorphism $F \to \mathbb{Z}_m$. Since $\mathrm{Hom}(\mathbb{Z}_k, \mathbb{Z}_m) \cong \mathbb{Z}_{\gcd(k.m)}$, F must get sent to the identity if $\gcd(k, m) = 1$, which is a contradiction since $F \not\subset \ker(\gamma)$ because $\ker(\alpha) = \mathrm{Im}(\gamma)$ is torsion-free. Furthermore k cannot be larger than m, since then part of F would necessarily be in $\ker(\alpha)$. Therefore we have determined that the only possible groups that fit into the extension are

$$1 \to \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}_k \to \mathbb{Z}_m \to 1, \qquad k \in \mathbb{N}, \quad k|m. \tag{686}$$

Done! □

Thus the number of possible groups fitting into the extension is $n - \phi(n)$, where ϕ is Euler's totient function. Hold on, though—on one hand, $H^2(\mathbb{Z}_n; \mathbb{Z}) = \mathbb{Z}_n$, which means that there are n inequivalent extensions of $\mathbb{Z}_m$ by $\mathbb{Z}$. However from we have just shown, the number of options for E is just $n - \phi(n) < n$, so evidently we will always get multiple distinct extensions with the same E. As another example, we have $|H^2(\mathbb{Z}_p; \mathbb{Z}_p)| = p$, while we know that $\mathbb{Z}_{p^2}$ and $\mathbb{Z}_p^2$ are the only groups of order p^2, and so in this case the difference between the number of inequivalent extensions and inequivalent groups E can be made arbitrarily large. What's up?

[60] This is just because H^1 gives homomorphisms from G to A (twisted appropriately for a possible group action), which determine how the group law in the $\rtimes$ is taken.

The key here is to remember that the classification of group extensions is a (much) finer classification than that of isomorphism classes of groups that occupy the central slot in the group extension SES. Indeed, the existence of an isomorphism $E \cong E'$ does *not* mean that all extensions with E at the center need to be equivalent: equivalence is a stronger constraint, and requires the commutativity of the diagram one gets when stacking the two SESs on top of eachother and connecting the columns with vertical isomorphisms.

Coming back to the extension problem, we see that we now get the diagram

$$\begin{tikzcd} 1 \arrow[r] & \mathbb{Z} \arrow[r, "f"] \arrow[d, "\times m"] & \mathbb{Z} \times \mathbb{Z}_k \arrow[r, "g"] \arrow[d, "\psi"] & \mathbb{Z}_m \arrow[r] \arrow[d, equal] & 1 \\ 1 \arrow[r] & \mathbb{Z} \arrow[r, "i"] & \mathbb{Z} \times \mathbb{Z}_m \arrow[r, "\pi"] & \mathbb{Z}_m \arrow[r] & 1 \end{tikzcd} \tag{687}$$

Now we can wolog take i to act as an injection into the $\mathbb{Z}$ factor and π to be a projection onto the $\mathbb{Z}_m$ factor. The commutativity of the left square then means that knowing f tells us ψ, and vice versa. Likewise, knowing g also tells us ψ, and in fact, knowledge of any one of f, g, ψ immediately determines the other two. In principle, there are many different choices of ψ, given by $\text{Hom}(\mathbb{Z} \times \mathbb{Z}_k, \mathbb{Z} \times \mathbb{Z}_m) \cong \mathbb{Z} \times \mathbb{Z}_k \times \mathbb{Z}_m$ (since $k|m$ by assumption so that $\text{Hom}(\mathbb{Z}_k, \mathbb{Z}_m) \cong \mathbb{Z}_k$).

Computing $\pi_1[U(n)/\mathbb{Z}_m]$

Today we're computing a fundamental group—this calculation should have been easy (and in any case, we get the obvious answer that one would expect), but took a lot of rather formal-feeling work before I was totally convinced of the answer. Our focus will be rather algebraic.

⋄ ⋄ ⋄

A prerequisite for computing this fundamental group will be showing the rather obvious fact that $\pi_1[U(n)] = \mathbb{Z}$. One algebraic way to do this is to use $\pi_1[PSU(n)] = \mathbb{Z}_n$ and the exact sequence

$$1 \to U(1) \xrightarrow{i} U(n) \xrightarrow{\pi} PSU(n) \to 1. \tag{688}$$

Since $\pi_{k>1}[U(1)] = 1$, we get

$$1 \to \mathbb{Z} \xrightarrow{i} \pi_1[U(n)] \xrightarrow{\pi} \mathbb{Z}_n \to 1. \tag{689}$$

To figure out the central term in this exact sequence, we need to know the map i (here abusing notation slightly to denote i for both the map on groups and the induced map on homotopy groups). We get this by noting that $\pi : e^{i\theta} \mapsto e^{i\theta}\mathbf{1}$. Note that we can't choose π to send $e^{i\theta}$ to something like $e^{i\theta} \oplus \mathbf{1}_{n-1}$ since exactness fixes the image of $U(1)$ to be the thing we're quotienting out by, which is the diagonal $U(1)$. Anyway, as θ winds once around the $U(1)$, the determinant of $i(e^{i\theta})$ winds by $2\pi n$, and hence the map i on homotopy groups is multiplication by n. Therefore $\ker(\partial) = n\mathbb{Z}$, and this forces $\pi_1[U(n)] = \mathbb{Z}$ as expected.

At the risk of being overly didactic here, let's elaborate on what exactly the different elements in $\pi_1[U(n)] = \mathbb{Z}$ are. The non-contractible S^1 inside of $U(n)$ that gives rise to the $\mathbb{Z}$ is of course found by the map $\det : U(n) \to U(1)$. Of course, $\ker(\det)$ is generated by $e^{2\pi i/n}\mathbf{1}_n$. The generator of the fundamental group can thus be given by the following homotopy: first, we consider the path $\gamma_{t<1} = \zeta_n^t \mathbf{1}_n$, where $\zeta_n = e^{2\pi i/n}$ and $t \in [0,1]$. When $t = 1$, we have something with trivial determinant. We then define a second part of the path via

$$\gamma_{1<t\leq 2} = \gamma_{t=1} \cdot \left(\zeta_n^{(t-1)(n-1)} \oplus (\zeta_n^*)^{t-1}\mathbf{1}_{n-1}\right). \tag{690}$$

Of course, this is chosen so that $\gamma_{t=2} = \mathbf{1}_n$, so that we form a closed path for $0 \leq t \leq 2$. Since the $1 \leq t \leq 2$ part of the path doesn't change the determinant, the path γ_t indeed wraps the $U(1)$ of the determinant once, and hence generates the fundamental group.

$\pi_1[U(n)/\mathbb{Z}_m]$ *from explicit constructions of the generators*

First we give a rather explicit construction of the fundamental group—this is likely good enough for physicists, but left me wondering if I was missing anything after doing it. A more careful algebraic approach is given in the next section.

Recall first how one can write down the generator of the fundamental group $\pi_1[PSU(n)] = \mathbb{Z}_n$. We can give a representative of the generator using only diagonal matrices as follows:

$$\gamma_t = \begin{pmatrix} \zeta_n^{-(n-1)t} & & & \\ & \zeta_n^t & & \\ & & \ddots & \\ & & & \zeta_n^t \end{pmatrix}, \qquad \zeta_n \equiv e^{2\pi i/n}. \tag{691}$$

Note that $\det \gamma_t = 1 \,\forall\, t$ as required, and that when we get to $t = 1$ we have $\zeta_n \mathbf{1}_n \sim \mathbf{1}$, so that the path closes.

We claim that γ_t is n-torsion. One rather hand-waving way to see this is to note that the endpoint of γ_t^{n-1}, viz. $\gamma_{t=1}^{n-1} = \gamma_n^{n-1}\mathbf{1}_n$, is the same as the endpoint of the inverse of γ, namely $\gamma_{t=1}^{-1} = \gamma_{t=1}^* = \gamma_t^{n-1}$. Therefore $\gamma^{n-1} = \gamma^{-1}$, and hence the fundamental group is n-torsion (this is a little slick since technically all of these endpoints are identified in $PSU(n)$, so one has to think about momentarily lifting up to $SU(n)$). Of course this conclusion is easily verified by computing the LES in homotopy groups associated with $1 \to \mathbb{Z}_n \to SU(n) \to PSU(n) \to 1$.

Now we go back and look at $U(n)/\mathbb{Z}_m$. One path which definitely appears in the fundamental group is the generator of $U(n)$ projected into $U(n)/\mathbb{Z}_m$; call this path γ^d (for determinant). Another path appearing is one similar to the above path in $PSU(n)$:

$$\gamma_t^S = \begin{pmatrix} \zeta_g^{-(n-1)t} & & & \\ & \zeta_g^t & & \\ & & \ddots & \\ & & & \zeta_g^t \end{pmatrix}, \qquad g \equiv \gcd(n,m). \tag{692}$$

Here the S stands for special unitary, since the determinant is $\gamma_t^S = 1 \,\forall\, t$. When $t = 1$ we have $\gamma_1^S = \zeta_g \mathbf{1}_n \sim \mathbf{1}_n$, since ζ_g is an m-th root of unity (note that taking ζ_m in the above path doesn't work unless $n \in m\mathbb{Z}$; taking g is the best we can do). For the same reason as in the $PSU(n)$ case, this path should have torsion g.

At this point we can conjecture that the homotopy classes of the loop in the phase of the determinant and the γ_t^S loop are enough to generate the fundamental group. Since the loop in the phase of the determinant can't be torsion on account of it coming from a loop in $U(1)$, we are thus led to guess that

$$\pi_1[U(n)/\mathbb{Z}_m] = \mathbb{Z} \times \mathbb{Z}_g, \tag{693}$$

which indeed turns out to be correct. Verifying this is a little tricky, though. For example, the minimal loop in $U(n)$ does not project entirely to the $\mathbb{Z}$ factor in the fundamental group—generically, its projection also contains a certain number of the torsionful γ^S loops. This is basically because of the $1 \leq t \leq 2$ part of the loop in the determinant (described in the previous section), which takes the matrix $\zeta_n \mathbf{1}_n$ back to $\mathbf{1}_n$. If we look back at the form of $\gamma_{1<t\leq 2}$, we see that this part of the path has precisely the form of the γ^S homotopy above. Therefore the projection of the generator of $U(n)$ will pick up some loops in the torsionful $\mathbb{Z}_g$ factor by virtue of this part of the homotopy. To see exactly how all of this works and check our guess for the fundamental group, we will need some more formal algebraic machinery.

More carefully: $\pi_1[U(n)/\mathbb{Z}_m]$ *from algebraic methods*

We now turn to computing $Q \equiv \pi_1[U(n)/\mathbb{Z}_m]$ from a more rigorous algebraic method, which lets us be sure that we aren't missing anything. We start from the relevant part of the LES in homotopy groups coming from the SES $1 \to \mathbb{Z}_m \to U(n) \to U(n)/\mathbb{Z}_m \to 1$, which is [61]

$$1 \to \mathbb{Z} \xrightarrow{\pi} Q \xrightarrow{\partial} \mathbb{Z}_m \to 1. \tag{696}$$

Recall how the two maps above work: π is just the projection of a loop in $U(n)$ to a loop in the quotient. ∂ works by taking a loop γ in Q and inserting it back into $U(n)$. This inserted loop will generically not be closed, with its endpoints differing by an mth root of unity ζ_m^l, $l \in \mathbb{Z}_m$. The map ∂ then assigns $\partial : \gamma \mapsto l$.

In a math diary entry on exact sequences, we show that the only possible choices for Q which fit into the SES (696) are

$$Q \cong \mathbb{Z} \times \mathbb{Z}_k, \qquad k|m. \tag{697}$$

[61] Unfortunately despite spending more time looking at homological algebra nonsense than I really wanted to, the following didn't produce anything useful, but I'm leaving it here for posterity's sake: now in some sense we only care about the torsion part of the Q, which can be accessed by hitting the SES with the Tor functor. For any SES $1 \to A \to B \to C \to 1$, we can tensor with some K and get (see Hatcher)

$$1 \to \mathrm{Tor}[K, A] \to \mathrm{Tor}[K, B] \to \mathrm{Tor}[K, C] \to A \otimes K \to B \otimes K \to C \otimes K \to 1, \tag{694}$$

with $\otimes$ taken over $\mathbb{Z}$. Applying this to the example at hand and setting $K = \mathbb{Z}_l$, we have

$$1 \to \mathrm{Tor}[Q, \mathbb{Z}_l] \to \mathbb{Z}_{(l,m)} \to \mathbb{Z}_l \to \mathbb{Z}_l \otimes Q \to \mathbb{Z}_{(l,m)} \to 1 \tag{695}$$

...

Also note that we have the exact sequence $1 \to U(1) \to U(n)/\mathbb{Z}_m \to PSU(n) \to 1$. The relevant part of the LES in homotopy groups then gives us the SES

$$1 \to \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}_k \to \mathbb{Z}_n \to 1. \tag{698}$$

Applying the same math result tells us that we must have $k|n$ if this is to be a legit SES, so that in fact $k|g$, with $g \equiv \gcd(n, m)$.

Given the SES (696), the results from the math diary also give us a commutative diagram

$$\begin{tikzcd} 1 \arrow[r] & \mathbb{Z} \arrow[r, "\pi"] \arrow[d, "\times m"] & \mathbb{Z} \times \mathbb{Z}_k \arrow[r, "\partial"] \arrow[d, "\psi"] & \mathbb{Z}_m \arrow[r] \arrow[d, equal] & 1 \\ 1 \arrow[r] & \mathbb{Z} \arrow[r, "i"] & \mathbb{Z} \times \mathbb{Z}_m \arrow[r, "p"] & \mathbb{Z}_m \arrow[r] & 1 \end{tikzcd} \tag{699}$$

where the rows are exact. Likewise by using the SES (698), we also have a copy of the above diagram with m replaced by n.

We then just need to figure out the value of k. There are a few ways of doing this; all of them involve using some information about our knowledge of the maps appearing in the above commutative diagram. The easiest is to just count dimensions, using our knowledge of the map π.

We claim that for the diagram (699), we have $x \equiv \pi(1) = (m/g, x_2)$, with x_2 some integer that we won't need to compute for this argument. Indeed, consider the path $\gamma_t(k,l) = (\zeta_m^k \zeta_n^l)^t \mathbf{1}_n$. For any $k, l \in \mathbb{Z}$, the matrix $\gamma_{t=1}$ can be connected to the identity in $U(n)/\mathbb{Z}_m$ without changing the determinant. Therefore $\gamma_t(k,l)$ must correspond to some element in Q with a non-zero value for the $\mathbb{Z}$ factor, which comes from the map $\det : U(n) \to U(1)$ as mentioned above (the full path γ_t may also have a nonzero value for the torsionful $\mathbb{Z}_k$ factor, but this won't be important for us). The matrix $\gamma_{t=1}$ with the smallest phase is found by minimizing $k/m + l/n = (kn + lm)/mn$, which is just g/mn (this γ_t has a det which progresses the smallest distance around the unit circle before becoming equivalent to 1). Therefore the path $\gamma_t^{min} = \zeta_{g/mn}^t$ will correspond to an element of Q with the minimal value of the torsion-free $\mathbb{Z}$ factor, viz. it will be of the form $(1, q)$ for some $q \in \mathbb{Z}_k$. Now the element x corresponds to a homotopy where the phase of the matrices involved increases from 0 to $2\pi/n$. Since a phase increase from 0 to $2\pi g/(mn)$ gives the minimal value of the $\mathbb{Z}$ component of Q, the $\mathbb{Z}$ component of x must be

$$\frac{2\pi/n}{2\pi g/(nm)} = \frac{m}{g}, \tag{700}$$

which implies indeed that $x = (m/g, x_2)$ for some $x_2 \in \mathbb{Z}$.

Now since

$$\langle x \rangle = \pi(\mathbb{Z}) \subset \mathbb{Z} \times \mathbb{Z}_k = \{(pm/g, px_2) \,|\, p \in \mathbb{Z}\}, \tag{701}$$

the index $[\mathbb{Z} \times \mathbb{Z}_k : \pi(\mathbb{Z})]$, which counts the number of cosets of $\pi(\mathbb{Z})$ in $\mathbb{Z} \times \mathbb{Z}_k$, must be mk/g (since each coset is labeled by a pair (a, b) with $a \in \mathbb{Z}_{m/g}, b \in \mathbb{Z}_k$). Given this, we can then apply the relation between the index of a normal subgroup and the order of the associated quotient group: for us, this means that we must have

$$[\mathbb{Z} \times \mathbb{Z}_k : \pi(\mathbb{Z})] = |\mathbb{Z}_m| = m. \tag{702}$$

Since the index is $(m/g)k$, we see that k must be g; hence the fundamental group is indeed $\mathbb{Z} \times \mathbb{Z}_g$.

The diagram (699) and the associated one with $m \leftrightarrow n$ let us check that our intuition about the generators of the fundamental group as discussed in the last section is correct.

First, take the easy case where $n = m$. The projection π in (699) must act as $\pi : 1 \mapsto (1, 0)$ in order for the top row to be exact. This makes total sense: at $t = 1$, the projection of the generator $x = \pi(1)$ of $\pi_1[U(n)]$ is the matrix $\gamma_1^d = \zeta_n \mathbf{1}_n$. The effect of quotienting by $\mathbb{Z}_n$ is simply to make the $1 \leq t \leq 2$ part of the homotopy γ_t^d unnecessary (effects from the quoteitn don't arise for the $0 \leq t \leq 1$ part of the homotopy since that part never passes through any of the $\zeta_n^k \mathbf{1}_n$ matrices that are killed by the quotient). We also see from the top row that $\partial(a, b) = b$ acts as projection onto the second factor. This also makes sense: the $\mathbb{Z}_m$ factor in the fundamental group is generated by the homotopy class of the γ_t^S loop. When this loop is lifted up into $U(n)$, its endpoints differ by precisely an mth root of unity; hence this generator is indeed sent to $1 \in \mathbb{Z}_m$.

The next simplest example is when $g = 1$. In this case we must have $x = \pi(1) = m$ and $\partial(1) = 1$ by exactness.

For example, take $U(2)/\mathbb{Z}_4$. Then we have the diagram

$$\begin{tikzcd} 1 \arrow[r] & \mathbb{Z} \arrow[r, "\pi"] \arrow[d, "\times 4"] & \mathbb{Z} \times \mathbb{Z}_2 \arrow[r, "\partial"] \arrow[d, "\psi"] & \mathbb{Z}_4 \arrow[r] \arrow[d, equal] & 1 \\ 1 \arrow[r] & \mathbb{Z} \arrow[r, "i"] & \mathbb{Z} \times \mathbb{Z}_4 \arrow[r, "p"] & \mathbb{Z}_4 \arrow[r] & 1 \end{tikzcd} \tag{703}$$

Intersection theory on the lattice and the Pontryagin square's purpose in life

The purpose of today's diary entry is to just make a note of the Pontryagin square and its properties that can be used as a reference. Thanks to Ryan Thorngren for some inspiring discussions in this regard.

◇ ◇ ◇

First, some opening / notational remarks: we will let $\Delta_{i_1 \dots i_{k+1}}$ denote a specific k-simplex in the lattice (alias CW complex) X under consideration (we will always be working with a triangulation). For a given numbering of each vertex in X, our convention is that the Δs are always listed with the sequence $i_1 \dots i_{k+1}$ monotonically increasing.

The cup product of a k-cochain a and an l-cochain b evaluated on the simplex $\Delta_{i_1\dots i_{k+l+1}}$ is

$$(a\cup b)_{\Delta_{i_1\dots i_{k+l+1}}} = a(\Delta_{i_1\dots l_{k+1}})b(\Delta_{i_{k+1}\dots i_{k+l+1}}). \tag{704}$$

Note that there is *no* kind of (anti)symmetrization on the RHS. If we were naively trying to write down a discrete analogue of a wedge product, we might have included a sum of some sorts (of sum sorts?) on the RHS—indeed, without the sum, the cup product is obviously not supercommutative on cochains, which seems to be worrying for its purported use as a discrete wedge product. But we will see later on that this lack of supercommutativity is essential for all the geometric properties of the cup product.

More notation: if $a \in C^\bullet(X;R)$ is any (co)chain, then $\widetilde{a} \in C_\bullet(X^\vee;R)$ will denote its Poincare dual.[62] Furthermore if C is any chain or submanifold, then C' will denote a "pushoff" of C, which is a copy of C displaced from C by a small amount in a direction determined by some choice of framing (not sure if this is standard terminology).

Cup products and intersections

A good reference for stuff in this subsection is Ryan Thorngren's thesis. Results on higher cup products originally come from [?], I believe.

As we mentioned above, a very important fact following from the definition (704) is that unlike $\wedge$, $\cup$ is not supercommutative. Instead,

$$a\cup b-(-1)^{|a||b|}b\cup a = (-1)^{|a|+|b|+1}\delta(a\cup_1 b)+(-1)^{|a|+|b|}\delta a\cup_1 b+(-1)^{|b|}a\cup_1\delta b, \tag{706}$$

Here $\cup_1$ is a degree -1 operation that is needed to allow the cup product to geometrically be dual to the intersection product.[63] It won't be important for us to know the exact CW complex description of $\cup_1$; we only point out that the way it appears in the above formula means that $\cup$ is still supercommutative when acting on equivalence classes of cocycles, but not supercommutative on the cochain level, even when cupping two cocycles.

The geometric meaning of the cup product is that it is dual to the intersection product via Poincare duality. That is, $a\cup b = \widetilde{a}\cap\widetilde{b}$, with $\cap$ the signed intersection number. When we want to define $a\cup a$ geometrically, we have to make use of a pushoff $\widetilde{a}'$ of $\widetilde{a}$, so that $a\cup a = \widetilde{a}\cap\widetilde{a}'$. We say "a pushoff" because the topological invariance of the intersection number means we just need to choose any pushoff—in fact, any other chain that is in the same homology class as $\widetilde{a}$ can be used for $\widetilde{a}'$ when computing $a\cup a$. This method of defining $a\cup a$ is of course familiar from the regularization procedure used in CS theory.

[62] Note that we are using the approach to Poincare duality wherein we have an isomorphism

$$C^k(X;R)\cong C_{n-k}(X^\vee;R), \tag{705}$$

with $X^\vee$ the dual lattice. At first sight, this is rather different, even conceptually, from the standard approach where one establishes $H^k(X;R)\cong H_{n-k}(X;R)$ by sending a k-cochain ϕ to $\widetilde{\phi}=[X]\cap\phi$, with $[X]$ the fundamental class. Not only does the latter approach remain on the same lattice, but it also involves extra minus signs coming from the fundamental class. That these two approaches are actually equivalent for the purposes of doing calculations in cohomology and intersection theory and so on is not obvious; see Ryan Thorngren's thesis for a great discussion of this.

[63] $\cup_1$ is also not supercommutative, in the same way: the general supercommutativity formula for $\cup_i$ is found by replacing $\cup$ in the above formula with $\cup_i$ and $\cup_1$ with $\cup_{i+1}$.

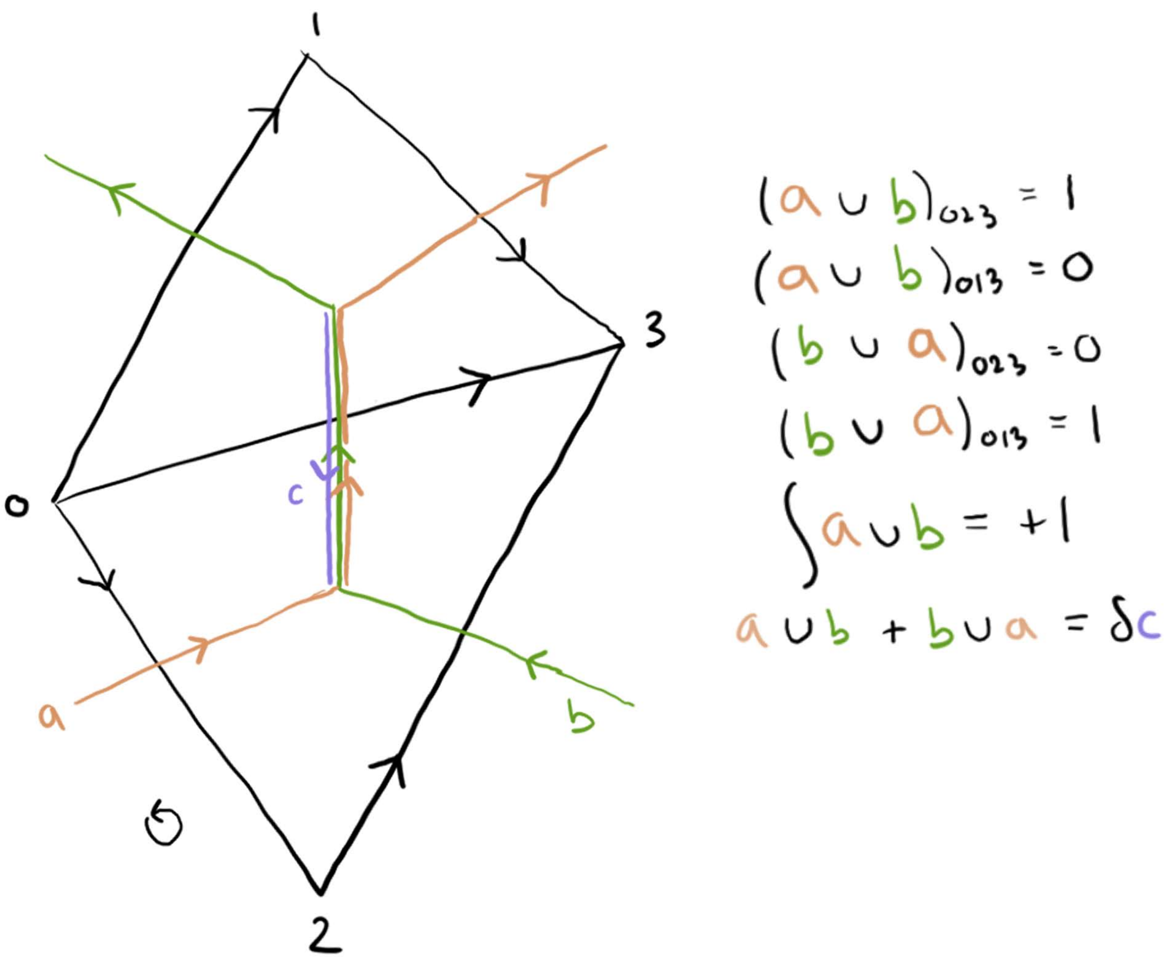


Figure 5: Calculation of the cup product for two 1-cochains a and b in two dimensions; see text for description.

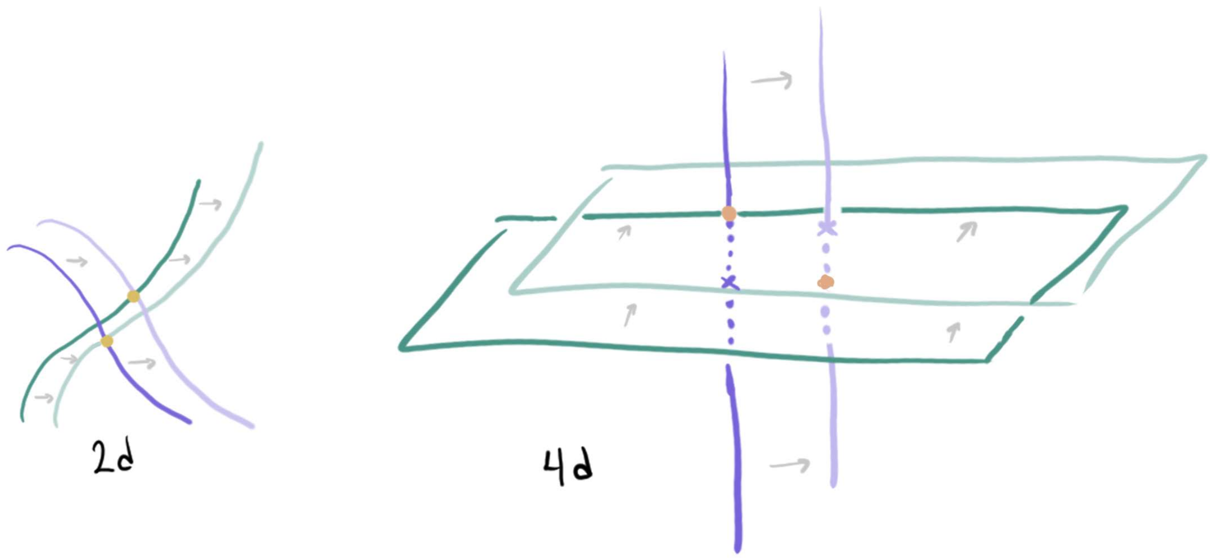


Figure 6: A figure showing the doubling of the explicit self-intersection points for lines in 2D and surfaces in 4D. On the left, the union of the purple and green dark lines is supposed to be a single 1-manifold C that has an explicit self-intersection (retrospectively they should have been the same color; the point is that dark stuff is one manifold and light stuff is a pushoff). The union of the light green and purple lines is a pushoff C'; one sees that it intersects C in two points. The right figure is basically the same thing: the union of the dark green plane + dark purple line (really a plane; the fourth dimension is suppressed) constitute a surface S with an explicit self-intersection. The pushoff S' again intersects S in two points.

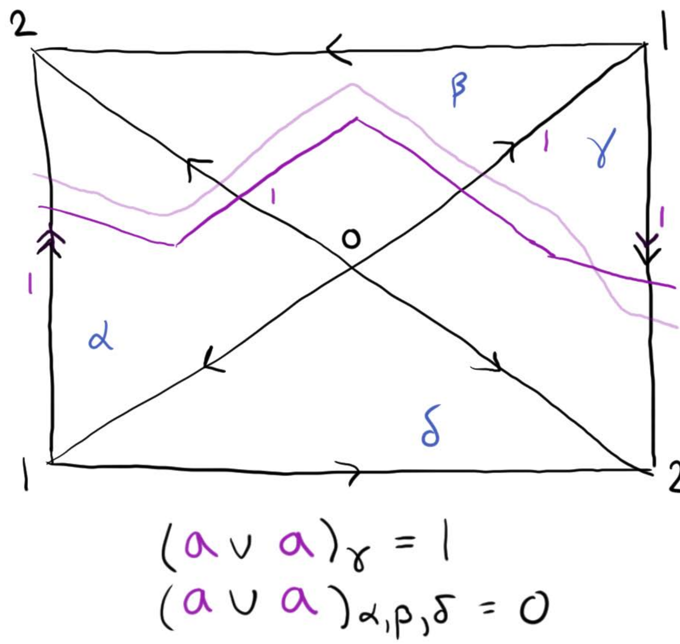


Figure 7: An example of how the $\cup$ product counts the intersection number of a framing SI. Here a is a 1-chain on the mobius strip (the left and right edges are identified as shown), and is Poincare dual to the dark purple line $\widetilde{a}$. The letters $\alpha, \beta, \gamma, \delta$ indicate the four different 2-cells in the triangulation. The light purple line is a pushoff of the dark purple line, and the point of the figure is to show that $\int a \cup a = \widetilde{a} \cap \widetilde{a}'$ (working with mod 2 coefficients as is appropriate for an unorientable manifold). Note that to correctly count explicit SIs, the Pontryagin square would need to be used instead of the $\cup$ product.

To get warmed up, we can consider figure 5, which shows the calculation of the cup product of two cochains whose Poincare duals are shown as orange and green lines on the dual lattice (for example, the cochain a assigns the number 1 to each link crossed by the orange lines; the orientation of the link needed to assign $+1$ and not -1 is determined with reference to a global orientation [the black corkscrew]).[64]

The Poincare duals of a and b intersect transversely,[65] and so we expect the cup product to assign $+1$ to this configuration. Indeed, we find $\int a \cup b = +1$ for this region of the CW complex. While $a \cup b \neq -b \cup a$, their difference is a total derivative, as indicated in the figure. Consequently, $\int a \cup b = -\int b \cup a$ (remember that the integral is done using signs coming from capping with the fundamental class!). This figure also lets us see that reversing the branching structure (flipping the direction of the arrows) reverses the cup product, so that if Δ^R is the cell Δ with reversed branching structure, then

$$(a \cup b)_\Delta = (b \cup a)_{\Delta^R}. \tag{708}$$

A baby example testing the claim that $a \cup a$ is dual to the self-intersection number of $\widetilde{a}$ is shown in figure 7, where we compute $a \cup a$ for a such that $\widetilde{a}$ is the nontrivial cycle of the mobius band. In order to not get $a \cup a = 0$ (since a is a 1-cochain), we need to work with $\mathbb{Z}_2$ coefficients. By looking at the figure, we see that indeed $a \cup a = \widetilde{a} \cap \widetilde{a}' = 1 \mod 2$.

The Pontryagin square is all about self-intersections (SIs), so let's talk about them for a little bit. As mentioned above, the self-intersection of a chain / submanifold C is computed as $C \cap C'$ for some pushoff C'. There are two types of SI that a given chain can have. The first type, which we'll call "explicit", is where the chain intersects itself on account of it being immersed but not embedded in the ambient manifold (e.g. the Klein bottle in $\mathbb{R}^3$ intersects itself explicitly). Figure 6 shows an example of an explicit SI in 2d (left) and 4d (right)—the dark colors represent two parts of a single self-intersecting manifold; the light colors are pushoffs.

The second type, which we'll call a "framing" SI, is one where the chain is embedded, but is such that its pushoffs intersect it due to the way it's framed, with its normal bundle having a nontrivial Euler class (the curve on the mobius band in fig. 7 has a framing SI

[64]For completeness, let's recall how the signs of transverse intersections are determined in general, for two chains / submanifolds N and M of complementary dimensions. One does this by comparing the orientations of the intersecting chains with the orientation from the ambient space X (coming from the fundamental class). Since the intersections we're dealing with are transverse, we can always choose a framing $\mathcal{F}_X$ of X and coordinate systems on N and M such that each basis vector e_i of $\mathcal{F}_X$ is parallel to some basis vector of $\mathcal{F}_N$ or $\mathcal{F}_M$. Let the volume form on some manifold Y be given by $V_Y = e_{i_1} \wedge \cdots \wedge e_{i_Y}$, where in the subscript Y denotes $\dim Y$. To compute the sign of $N \cap M$, we simply compute the sign σ such that

$$\sigma V_N \wedge V_M = V_X. \tag{707}$$

By the supercommutativity of the wedge product, the intersection number is also supercommutative.

[65]Well, not quite. The fact that the lattice forces the duals to be degenerate along a link is the origin of all this non-supercommutativity and higher cup product business. However, in the discussion that follows, when we talk about chains intersecting, we will always be thinking in the continuum, where the intersections can always be wiggled to be transverse and well-defined. We thus always keep in the back of our mind that the "transverse intersections" talked about below really degenerate when put on the lattice. The point of the properties of the cup product is to ensure that this continuum geometric picture for intersections always works out when we restrict to the lattice and calculate with $\cup$s.

number of 1). These types of framing SI always appear in pairs for curves immersed in spin 2-manifolds and surfaces immersed spin 4-manifolds, so we usually have to go to non-spin manifolds to get a chain which has a minimal framing SI.

The Pontryagin square

The purpose of the Pontryagin square is to allow one to do intersection theory with torsion. Algebraically, the Pontryagin square provides us with a squaring operation on cohomology classes in $H^k(X;\mathbb{Z}_m)$:

$$P : H^k(X;\mathbb{Z}_m) \times H^k(X;\mathbb{Z}_m) \to H^{2k}(X;\mathbb{Z}_n), \tag{709}$$

where we will determine what n should be momentarily. The defining algebraic feature of P is that it factors in the way one expects a squaring operation to factor, namely

$$P(a+b) = P(a) + P(b) + 2a \cup b. \tag{710}$$

The defining geometric feature of P (of course this is just another way of formulating its algebraic properties) is that it is dual to the self-intersection pairing. We know from the above that the cup product is dual to the intersection pairing when the coefficient group is $\mathbb{Z}$, and so the difference between $P(a)$ and $a \cup a$ must come from torsion effects.

The whole framework of intersection theory described above is equipped to deal with integer linear combinations of chains, and not $\mathbb{Z}_m$-valued linear combinations. Therefore, in order to do intersection theory with $\mathbb{Z}_m$-valued chains, we will first lift them to $\mathbb{Z}$-valued chains, and then later reduce back to torsionful coefficients after calculating intersection numbers. We will write the integer lift of a chain w as $w_{\mathbb{Z}}$. One is then tempted to conclude that a cohomology operation that is dual to the self-intersection pairing could be constructed by taking the $\cup$ of the integer lifts, and then reducing mod m:

$$\widetilde{a} \cap \widetilde{a}' \xrightarrow{?} [a_{\mathbb{Z}} \cup a_{\mathbb{Z}}]_m, \tag{711}$$

where $[]_m$ is reduction mod m.

The reason why $[a_{\mathbb{Z}} \cup a_{\mathbb{Z}}]_m$ isn't the appropriate intersection pairing dual in the torsionful case can be illustrated by considering the case of $\mathbb{Z}_2$ coefficients. From looking at figure 6, we see that the intersection number for an explicit SI is always even.[66] Therefore the cup product on $\mathbb{Z}_2$ cochains, which maps into $H^{2k}(X;\mathbb{Z}_2)$, is completely blind to explicit SIs. But of course an explicit SI should still be able to be kept track of, if we want a cohomology operation that is truly dual to the self-intersection. This motivates us to look for a cup product-like operation that maps into $H^{2k}(X;\mathbb{Z}_4)$, since then explicit SIs could be properly counted mod 2.[67]

[66] This means that in a precise sense, framing SIs give us SIs that are "smaller" than the more obvious explicit SIs.

[67] The mod 2 case can also be understood in terms of orientations—this is the subject of a nice math overflow answer pertaining to $P(w)$ by Kevin Walker. An intersection number defined for a $\mathbb{Z}_2$-valued chain needs to be well-defined under local changes in the chain's orientation. In figure 6, we can see what happens under local orientation changes for explicit SIs: we make a local change of the immersed manifold that

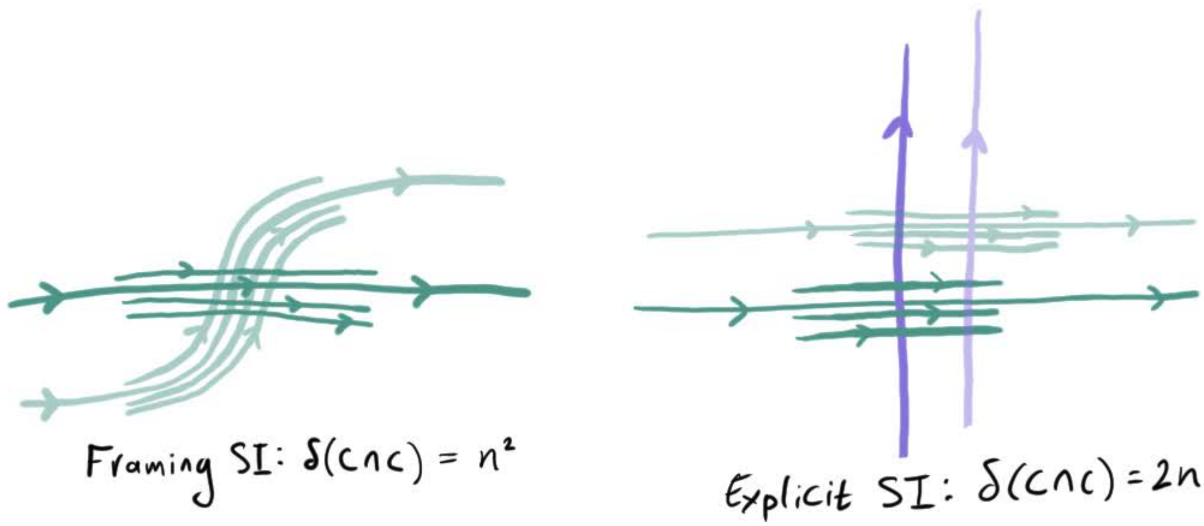


Figure 8: Illustrating the well-definedness of the intersection number for a $\mathbb{Z}_n$-valued chain C (in the figure, $n = 3$). Left: framing SI. Adding a chain in $C_\bullet(X; n\mathbb{Z})$ (the segments of three parallel lines in the figure) to C results in a change of SI number of $\delta(C \cap C) = n^2$ (should really be $\delta(C \cap C')$). Right: explicit SI; the dark green and dark purple lines are two parts of the same chain C; the lighter versions are C'. The change in SI number is now $\delta(C \cap C) = 2n$.

More generally, consider the case of $\mathbb{Z}_m$ coefficients. If $m \in 2\mathbb{Z}$ then we still have the problem that explicit SIs get doubled-counted by $\cup$, so that $\cup$ can't tell the difference between $m/2$ explicit SIs and 0 explicit SIs (we want it to only fail to distinguish between m and 0). In order to properly count these SIs, we need to look for a cohomology operation that maps into $\mathbb{Z}_{2m}$ coefficients. On the other hand if $m \in 2\mathbb{Z}+1$, then since 2 generates $\mathbb{Z}_m$, a minimal explicit SI is m-torsion, as desired. Therefore in the case where m is odd, we can get away with using just $[a_\mathbb{Z} \cup a_\mathbb{Z}]_m$.

Now we ask whether such a cohomology operation can be well-defined on $\mathbb{Z}_m$ cochains. Consider sending

$$a_\mathbb{Z} \mapsto a_\mathbb{Z} + mC, \tag{712}$$

where $C \in C^k(X; \mathbb{Z})$ is any cochain (not necessarily closed!). The effect of doing this on the SI of a is shown in figure 8. Under this shift, a minimal framing SI changes in value by something in $m^2\mathbb{Z}$, while a minimal explicit SI changes by something in $2m\mathbb{Z}$. Therefore the intersection number is well-defined in $\mathbb{Z}_{\gcd(2m,m^2)}$, which is equal to $\mathbb{Z}_m$ if m is odd, and $\mathbb{Z}_{2m}$ if m is even. We also see that the intersection number of a sum of cochains satisfies the square rule (710)—the self-intersection of $a + b$ contains the SI of a, the SI of b, and if a and b intersect, an explicit SI between a and b that comes with a factor of 2.

reverses only the orientation of one part of the manifold at the intersection point (say, we just reverse the orientation of the green part—this is possible because we are changing the orientation of the immersed manifold, and not of the ambient background manifold). This changes the SI number from ± 2 to ∓ 2. On the other hand, framing SI numbers don't change at all, since both the immersed manifold and its pushoff change orientation. Therefore the SI number is well-defined mod 4 in general.

Given the above considerations, we are lead to define $P(w)$ for any $w \in H^\bullet(X;\mathbb{Z}_m)$ by

$$P(w) = \begin{cases} [w_\mathbb{Z} \cup w_\mathbb{Z} + w_\mathbb{Z} \cup_1 \delta w_\mathbb{Z}]_{2m} & m \in 2\mathbb{Z} \\ [w_\mathbb{Z} \cup w_\mathbb{Z}]_m & m \in 2\mathbb{Z}+1 \end{cases} \tag{713}$$

which is the appropriate cohomology dual of the mod m intersection pairing. The mysterious $w_\mathbb{Z} \cup_1 \delta_1 w_\mathbb{Z}$ (remember that $w_\mathbb{Z}$ is only closed mod m, so this does not need to vanish mod $2m$) in the m even case (it's not there for the odd case since it dies when reduced mod m) is needed to ensure that $P(w)$ is well-defined and obeys (710), which we just argued is obeyed by the self intersection number. Indeed, for $m \in 2\mathbb{Z}$, we calculate

$$\begin{aligned} P(w+v) = [w_\mathbb{Z}^{\cup 2} + v_\mathbb{Z}^{\cup 2} + w_\mathbb{Z} \cup_1 \delta w_\mathbb{Z} + v_\mathbb{Z} \cup_1 \delta v_\mathbb{Z} + (1-(-1)^{|w|}) w \cup v \\ - \delta(v \cup_1 w) + (1+(-1)^{|w|}) v \cup_1 \delta w + \{w, \delta v\}_{\cup_1}]_{2m}, \end{aligned} \tag{714}$$

where the asymmetric nature of the last two terms comes from the way we wrote the supercommutativity law. The first line in (714) contains the answer we want, and happily all the terms in the second line vanish mod $2m$ after we mod out by total coboundaries, since the the $\cup_1$ product is supercommutative up to things in $C^\bullet(X; m\mathbb{Z})$, and because terms that are valued in $m^2\mathbb{Z}$ die on account of $m^2 \in 2m\mathbb{Z}$ if m is even. Therefore we indeed have (710). That (710) is satisfied in the case of m odd is easy to check, since in that case we just reduce mod m and all the $\cup_1$ terms coming from the supercommutativity rule die.

Applying this result, we can then check that $P(w)$ is indeed well-defined on $H^\bullet(X;\mathbb{Z}_m)$: first, under $w \mapsto w + mc$, we have

$$P(w) \mapsto P(w) + P(mc) + [2mw \cup c]_n, \tag{715}$$

with $n = m$ or $2m$ as appropriate. The last term dies because of the $2m$, and the second term dies because of the m^2. Finally, one needs to verify that $P(w)$ is closed in $C^{2|w|}(X;\mathbb{Z}_n)$ if w is closed in $C^{|w|}(X;\mathbb{Z}_m)$; this is again shown through the fact that the $\cup_i$s are supercommutative modulo m. Therefore, we have shown that $P(w)$ obeys all the relations we expect of the cohomology dual to the intersection pairing.

The point of all this is to show that there is a precise way to translate the simple geometric intuition one has from thinking about intersection theory into a dual algebraic formalism one can use on cochains. To quote *The Wild World of Four-Manifolds*, "think with intersections, prove with cup products"!

Finally we can talk about the meaning of the discrete term that often appears in topological actions, viz.

$$S_{top} \supset \frac{p}{2n} \int P(w_2). \tag{716}$$

In order for this to be well-defined, we evidently must have $p \in \mathbb{Z}_{2n}$ if n is even, or $p \in 2\mathbb{Z}_{2n}$ if n is odd.

Consider a w_2 surface which has a minimal explicit self-intersection—the picture here would be e.g. a pair of (discrete) magnetic flux lines propagating and then linking at some point in time. The minimal phase that S_{top} can assign to this SI is $e^{2\pi i/n}$, so that n explicit SIs always give a trivial phase.

We could have been forgiven for thinking that S_{top} would always assign nth roots of unity to w_2 surface intersections, but this is only true for explicit SIs: framing SIs instead can contribute a phase that is only half this value, namely $e^{\pi i/n}$. For $n \in 2\mathbb{Z}+1$ there is not a sharp distinction between the two types of SIs vis-a-vis how they appear in S_{top} since both 1 and 2 generate $\mathbb{Z}_n$, but for $n \in 2\mathbb{Z}$ (and for non-spin manifolds, where the framing SIs can come in odd numbers) the distinction is important. Therefore even though the characteristic class w_2 is $\mathbb{Z}_n$-valued, the Pontryagin square provides us with a way (in certain circumstances), to get a well-defined topological action which gives phases that are $2n$-th roots of unity, and it does this by constructing a correct torsionful cohomology dual of the intersection pairing.

Basic stuff on Haar integration

Today we will derive a minimal set of results needed to get acquainted with how to calculate Haar integrals over unitary groups. A nice physics-friendly paper to get intuition for Haar averages is [1].

◇ ◇ ◇

The general goal is to understand how (at least in principle) to compute integrals of the form

$$\mathbb{E}_U[U_{IJ}U^*_{KL}], \tag{717}$$

where $\mathbb{E}_U[\,]$ denotes a Haar average over U, and where we are using multi-index notation

$$U_{IJ} \equiv U_{i_1j_1}\cdots U_{i_nj_n}. \tag{718}$$

We will let $|I|$ denote the order of the $i_1,\dots,i_n$ (with $|I|=|J|$ and $|K|=|L|$ are required in (717)).

The invariance of the Haar measure, viz.

$$\mathbb{E}_U[f(U,U^\dagger)] = \mathbb{E}_U[f(V_1UV_2^\dagger, V_2U^\dagger V_1^\dagger)] \tag{719}$$

for any unitaries $V_{1,2}$, allows us to establish some basic facts about the above expectation values rather quickly. First, it is easy to see that the average (717) vanishes unless $|I|=|K|$. Indeed, the average is invariant under the rotation $U\mapsto e^{i\theta}U$, which means that

$$\mathbb{E}_U[U_{IJ}U^*_{KL}] = e^{i\theta(|I|-|K|)}\mathbb{E}_U[U_{IJ}U^*_{KL}] \implies \mathbb{E}_U[U_{IJ}U^*_{KL}] \propto \delta_{|I|,|K|}. \tag{720}$$

In what follows we will thus assume wolog that $|I|=|K|=|J|=|L|$.

We can also consider the invariance under the rotation

$$U\mapsto V(\theta)UV(\theta'), \tag{721}$$

where $[V(\theta)]_{ij} = e^{i\theta_i}\delta_{ij}$, and likewise for $V(\theta')$. Tracking down indices shows that this means

$$\mathbb{E}_U[U_{IJ}U^*_{KL}] = \exp\left(i\sum_{j=1}^{|I|}(\theta_{I_j}-\theta_{K_j}+\theta'_{L_j}-\theta'_{J_j})\right)\mathbb{E}_U[U_{IJ}U^*_{KL}]. \tag{722}$$

Since we can choose all of the θ_i, θ_i' independently, the expectation value will be nonzero only if I, K and J, L form two pairs related by permutations:

$$\mathbb{E}_U[U_{IJ}U^*_{KL}] \neq 0 \implies I = \sigma(K), \quad J = \sigma'(L), \quad \sigma, \sigma' \in S_{|I|}. \tag{723}$$

In the special case where $|I| = 1$, this gives

$$\mathbb{E}_U[U_{ij}U^*_{kl}] = C\delta_{ik}\delta_{jl}, \tag{724}$$

where C is a constant. The proportionality constant can be fixed for example by calculating

$$\sum_j \mathbb{E}_U[U_{ij}U^*_{kj}] = \sum_j C\delta_{ik} = dC\delta_{ik}. \tag{725}$$

On the other hand, this is equal to $\mathbb{E}_U[\delta_{ik}]$. Since we want this to be equal to δ_{ik}, we evidently must have $C = 1/d$, so that

$$\mathbb{E}_U[U_{ij}U^*_{kl}] = \frac{1}{d}\delta_{ik}\delta_{jl}. \tag{726}$$

However, exact results for large values of $|I|$ quickly get complicated. As an example, we will do the calculation for the case where $|I| = 2$. This will require making use of some simple representation theory tools.

Applying Schur-Weyl duality

The standard mathematical way of tackling the computation of expectation values for generic $|I|$ is to use Schur-Weyl duality. This is reviewed in another diary entry, but basically it is a statement about the representation theory of unitary groups on vector spaces of the form $V^{\otimes n}$. It tells us that

$$V^{\otimes n} \cong \bigoplus_\alpha V_\alpha \otimes \mathbb{S}_\alpha V. \tag{727}$$

Here the sum runs over irreps V_α of S_n, and

$$\mathbb{S}_\alpha V = \mathrm{Hom}_{S_n}(V_\alpha, V^{\otimes n}) \tag{728}$$

either vanishes or is an irrep of $U(d)$, where $d = \dim V$.

The sum over irreps here can make explicit computations rather complicated as n gets large. A simple case, and the only one that we ever really bother with in most cases, is where $n = 2$, which as we will see is relevant for computing Haar averages where $|I| = 2$. Here we only have two irreps for the permutation group, and the above tells us that in this case that the splitting

$$V^{\otimes 2} \cong \mathrm{Sym}^2 V \oplus \mathrm{Alt}^2 V \tag{729}$$

is in fact a splitting into simple modules for the $U(d)$ action.

We will use this to compute the Haar average in the case where $|I| = 2$. Writing the indices out, our goal is to compute

$$\mathbb{E}_U[U_{a'a}U_{b'b}U^*_{c'c}U^*_{d'd}]. \tag{730}$$

Even though it looks like we might need to know the representation theory for S_4 since this looks like the action of something in $\mathrm{End}(V^{\otimes 4})$, we can in fact view the above as the matrix element of something in $\mathrm{End}(V^{\otimes 2})$, where the something is the matrix

$$\mathcal{E}_{abcd} \equiv U|a\rangle\langle c|U^\dagger \otimes U|b\rangle\langle d|U^\dagger \in \mathrm{End}(V^{\otimes 2}). \tag{731}$$

We can then use the above result to compute the average of the above matrix element for each a, b, c, d. First, since we can change variables in the Haar integration, we know that

$$[\mathbb{E}_U[\mathcal{E}_{abcd}], U(d)^{\otimes 2}] = 0. \tag{732}$$

Schur-Weyl duality tells us that in the basis in which the $U(d)^2$ action is block-diagonalized, we can write $\mathbb{E}_U[\mathcal{E}_{abcd}]$ as a $\oplus$ of a symmetric matrix and an anti-symmetric one. Furthermore, the vanishing of the commutator above implies that when $\mathbb{E}_U[\mathcal{E}_{abcd}]$ is written in the basis of the direct sum decomposition, both blocks are proportional to the identity.

To figure out the constants of proportionality, we take a trace on each subspace:

$$\begin{aligned}\mathrm{Tr}\left[\frac{\mathbf{1}\pm\mathcal{S}}{2}\mathbb{E}_U[\mathcal{E}_{abcd}]\right] &= \frac{1}{2}\mathbb{E}_U \mathrm{Tr}[U_{a'a}U^\dagger_{cc'}U_{b'b}U^\dagger_{dd'} \pm U_{b'a}U^\dagger_{cd'}U_{a'b}U^\dagger_{dc'}] \\ &= \frac{\delta_{ac}\delta_{bd}\pm\delta_{ad}\delta_{bc}}{2},\end{aligned} \tag{733}$$

where $\mathcal{S} = \sum_{ab}|ab\rangle\langle ba|$ is the swap operator. Now the symmetric and antisymmetric subspaces $\mathrm{Sym}^2 V$, $\mathrm{Alt}^2 V$ have dimensions $d(d+1)/2$ and $d(d-1)/2$, respectively. Therefore in order for the traces to work out, the constant of proportionality must be such that

$$\mathbb{E}_U[\mathcal{E}_{abcd}] = \mathbf{1}_{\mathrm{Sym}}\frac{2}{d(d+1)}\frac{\delta_{ac}\delta_{bd}+\delta_{ad}\delta_{bc}}{2} \oplus \mathbf{1}_{\mathrm{Alt}}\frac{2}{d(d+1)}\frac{\delta_{ac}\delta_{bd}-\delta_{ad}\delta_{bc}}{2}. \tag{734}$$

We can sum this up by using the projectors onto the two subspaces to write

$$\mathbb{E}_U[\mathcal{E}_{abcd}] = \sum_{\sigma=\pm 1}\frac{\mathbf{1}+\sigma\mathcal{S}}{2d(d+\sigma)}(\delta_{ac}\delta_{bd}+\sigma\delta_{ad}\delta_{bc}), \tag{735}$$

where now $\mathbf{1}$ acts on the full $V^{\otimes 2}$.

We can now get the full Haar average that we originally wanted by evaluating the matrix elements of the above matrix:

$$\begin{aligned}\mathbb{E}_U[U_{a'a}U_{b'b}U^*_{c'c}U^*_{d'd}] &= [\mathbb{E}_U[\mathcal{E}_{abcd}]]_{a'b',c'd'} \\ &= \delta_{a'c'}\delta_{b'd'}\left(\delta_{ac}\delta_{bd}\left(\frac{1}{2d(d+1)}+\frac{1}{2d(d-1)}\right)+\delta_{ad}\delta_{bc}\left(\frac{1}{2d(d+1)}-\frac{1}{2d(d-1)}\right)\right) \\ &+ \delta_{a'd'}\delta_{c'b'}\left(\delta_{ac}\delta_{bd}\left(\frac{1}{2d(d+1)}-\frac{1}{2d(d-1)}\right)+\delta_{ad}\delta_{bc}\left(\frac{1}{2d(d-1)}+\frac{1}{2d(d+1)}\right)\right) \\ &= \frac{1}{d^2-1}\left(\delta_{a'c'}\delta_{b'd'}\delta_{ac}\delta_{bd}+\delta_{a'd'}\delta_{c'b'}\delta_{ad}\delta_{bc}-\frac{1}{d}\left(\delta_{a'd'}\delta_{b'c'}\delta_{ac}\delta_{bd}+\delta_{a'c'}\delta_{b'd'}\delta_{ad}\delta_{bc}\right)\right)\end{aligned} \tag{736}$$

We can do a sanity check on this expression by e.g. setting $b' = c'$ and summing over b', since then the RHS should reduce to δ_{bc} times the second moment expression we calculated above. Indeed,

$$\begin{aligned}\sum_{b'} \delta_{b'c'} \mathbb{E}_U[U_{a'a} U_{b'b} U^*_{c'c} U^*_{d'd}] &= \sum_f \frac{1}{d^2-1} \Big(\delta_{a'f}\delta_{fd'}\delta_{ac}\delta_{bd} + \delta_{a'd'}\delta_{ff}\delta_{ad}\delta_{bc} \\ &\qquad - \frac{1}{d}\left(\delta_{a'd'}\delta_{ff}\delta_{ac}\delta_{bd} + \delta_{a'f}\delta_{fd'}\delta_{ad}\delta_{bc}\right)\Big) \\ &= \frac{d-1/d}{d^2-1}\delta_{a'd'}\delta_{ad}\delta_{bc} \\ &= \delta_{bc}\mathbb{E}_U[U_{a'a}U^*_{d'd}] \quad \checkmark\end{aligned} \tag{737}$$

Primer on Schur-Weyl Duality

Today's diary entry is a very brief recapitulation of Schur-Weyl duality. Fulton and Harris is a good reference, as is the paper [6].

◇ ◇ ◇

First we need to recall some facts about the representation theory of S_n. Since this diary entry is meant to be used as a reference and hence isn't attempting to be pedagogical, we will mostly just recall facts, without explaining them in detail. In the following we will let $\alpha = (\alpha_1, \ldots, \alpha_k)$ denote a partition of n, i.e. a collection of decreasing integers that sum to n. To each such partition one can assign a Young diagram Y_α by giving the jth row α_j boxes. We can then assign to each α the groups

$$\begin{aligned} C_\alpha &= \{g \in S_n \,|\, g \text{ preserves the columns of } Y_\alpha\} \\ R_\alpha &= \{g \in S_n \,|\, g \text{ preserves the rows of } Y_\alpha\} \end{aligned} \tag{738}$$

For each α we can then define the following two elements of the group algebra $\mathbb{C}[S_n]$:

$$A_\alpha = \sum_{g\in R_\alpha} g, \qquad B_\alpha = \sum_{g \in C_\alpha} s(g) g, \tag{739}$$

where $s(g)$ is the sign of g. The product

$$\mathcal{Y}_\alpha = A_\alpha B_\alpha \tag{740}$$

is called the *Young symmetrizer* of α. This construction is important because there is a theorem saying that every irrep[68] V_α of S_n can be uniquely associated to a partition α via

$$V_\alpha = \mathcal{Y}_\alpha \mathbb{C}[S_n]. \tag{741}$$

[68]Here we are using V_α to label representations and not R_α since the latter has unfortunately already been taken. Also, we are always working over $\mathbb{C}$.

This is possible because the C_αs are minimal idempotents in the group algebra.

The two simplest examples come from the Young diagrams with only one row or column. The Young diagram with only one row, coming from $\alpha = (n)$, corresponds to the trivial representation: here $R_\alpha = S_n$ while C_α is trivial so that $\mathcal{Y}_\alpha$ is the uniform sum over all of S_n, and by Peter-Weyl we know that this is the trivial representation. Indeed, it is easy to check that we have $\mathcal{Y}_{(n)} h = \mathcal{Y}_{(n)}$ for any $h \in S_n$.

The Young diagram with only one column is similarly easy: now R_α is trivial while C_α is all of S_n. Thus we have $\mathcal{Y}_{(1,\dots,1)} = \sum_g s(g) g$. By changing variables in the sum, we see that for any $h \in S_n$, we have $\mathcal{Y}_{(1,\dots,1)} h = \mathrm{sgn}(h) \mathcal{Y}_{(1,\dots,1)}$, and so the partition $(1,\dots,1)$ corresponds to the alternating representation.

The other piece of background information that we will need is the double-centralizer theorem:

`Double-centralizer theorem:` Let $\mathcal{A}$ be a semisimple subalgebra of $\mathrm{End}(V)$ for some vector space V, and define

$$\mathcal{B} = \mathrm{End}_{\mathcal{A}}(V). \tag{742}$$

Let $\mathcal{M}_i$ denote the simple modules[69] of $\mathcal{A}$, and define $\mathcal{N}_i \equiv \mathrm{Hom}_{\mathcal{A}}(\mathcal{M}_i, V)$. Then V decomposes as a $\mathcal{A} \otimes \mathcal{B}$ module as

$$V \cong \bigoplus_i \mathcal{M}_i \otimes \mathcal{N}_i, \tag{743}$$

where furthermore the $\mathcal{N}_i$s are isomorphic to the simple modules of $\mathcal{B}$.

Proof. The statement that the above isomorphism holds given the definition of the $\mathcal{N}_i$s is actually the easy part of the proof; the slightly trickier part is showing that the $\mathcal{N}_i$s correspond to the simple modules of $\mathcal{B}$. For the easy part, the claim is that for any $\mathcal{A}$ and V as above, we have

$$V \cong \bigoplus_i \mathcal{M}_i \otimes \mathrm{Hom}_{\mathcal{A}}(\mathcal{M}_i, V). \tag{744}$$

This is proved using Schur's lemma. First, we know that V splits as a direct sum of simple modules of $\mathcal{A}$, so that we can write $V \cong \bigoplus_i \mathcal{M}_i^{n_i}$, where n_i is the multiplicity of the submodule $\mathcal{M}_i$. Then the $\mathrm{Hom}_{\mathcal{A}}$ is, using Schur's lemma,

$$\mathrm{Hom}_{\mathcal{A}}\left(\mathcal{M}_i, \bigoplus_j \mathcal{M}_j^{n_j}\right) \cong \mathbb{C}^{n_i}. \tag{745}$$

Plugging this in to (744) gives the result. The $\mathcal{N}_i = \mathrm{Hom}_{\mathcal{A}}(\mathcal{M}_i, V)$ factors are known as the multiplicity spaces of $\mathcal{M}_i$, since as we have seen they are isomorphic to vector spaces of dimension equal to the multiplicity of $\mathcal{M}_i$ in V. Explicitly, the isomorphism from the RHS of (744) to the LHS is given by $v \otimes \rho \mapsto \rho(v)$.

Now we just have to show that the multiplicity spaces $\mathcal{N}_i$ are in fact the simple modules (irreps) of $\mathcal{B}$. First we show that $\mathcal{B}$ is a direct sum of the $\mathcal{N}_i$: re-writing the first V in the

[69] Think: 'invariant subspaces'.

RHS of $\mathrm{End}(V) = \mathrm{Hom}_{\mathcal{A}}(V, V)$ using (744), we have

$$\begin{aligned} \mathcal{B} &\cong \mathrm{Hom}_{\mathcal{A}}\left(\bigoplus_i \mathcal{M}_i \otimes \mathcal{N}_i, V\right) \\ &\cong \bigoplus_i \mathrm{Hom}_{\mathcal{A}}(\mathcal{N}_i, V \otimes \mathcal{M}_i^*) \\ &\cong \bigoplus_i \mathrm{Hom}(\mathcal{N}_i, \mathrm{Hom}_{\mathcal{A}}(\mathcal{M}_i, V)) \\ &\cong \bigoplus_i \mathrm{End}(\mathcal{N}_i). \end{aligned} \tag{746}$$

This is what we'd expect if the $\mathcal{N}_i$ were simple $\mathcal{B}$ modules. To check that this is indeed the case,[70] we need to show that $\mathcal{B}$ acts on the $\mathcal{N}_i$ transitively (we can get from one element to any other element by acting with $\mathcal{B}$), since this wouldn't be the case if the $\mathcal{N}_i$ were not simple. So to this end, fix two elements $f, \widetilde{f}$ in $\mathcal{N}_i$, Since $\mathcal{M}_i$ is simple, we can wolog characterize $f, \widetilde{f}$ by how they act on a particular element $m \in \mathcal{M}_i$. The action of $\mathcal{B}$ on $\mathcal{N}_i$ is transative simply because we can always define an endomorphism $b \in \mathcal{B}$ such that $b(f(m)) = \widetilde{f}(m)$, with b acting as $\mathbf{1}$ on the space orthogonal to $\mathcal{A}f(m)$.[71]

□

Note that since the $\mathcal{N}_i$ are simple modules of $\mathcal{B}$, we can equally well take a dual perspective, where we define $\mathcal{A}$ by $\mathcal{A} \equiv \mathrm{End}_{\mathcal{B}}(V)$ for $\mathcal{B}$ some semisimple subalgebra of $\mathrm{End}(V)$. The we can write (744) dually as

$$V \cong \bigoplus_i \mathcal{N}_i \otimes \mathrm{Hom}_{\mathcal{B}}(\mathcal{N}_i, V), \tag{747}$$

with the dual multiplicity spaces $\mathrm{Hom}_{\mathcal{B}}(\mathcal{N}_i, V)$ being isomorphic to the simple modules $\mathcal{M}_i$ of $\mathcal{A}$. We can also write this as

$$\mathrm{End}_{\mathrm{End}_{\mathcal{A}}(V)}(V) \cong \mathcal{A}, \tag{748}$$

although the important part of this theorem is the above direct sum decomposition. This result is called the double-centralizer theorem because the centralizer of the centralizer of $\mathcal{A}$ in $\mathrm{End}(V)$ is $\mathcal{A}$ itself. Any subalgebra of $\mathrm{End}(V)$ determines its centralizer, and vice versa.

Now we come to the setting of Schur-Weyl duality, which involves vector spaces of the form $V^{\otimes n}$, where each V factor is isomorphic to $\mathbb{C}^d$. For concreteness, and for applications to physics, we will focus our attention on the case where the subalgebra $\mathcal{A}$ of $\mathrm{End}(V)$ is $U(d)$, which naturally acts diagonally on each factor of the $\otimes$ through matrices of the form $U(d)^{\otimes n}$. S_n also acts naturally by permuting the tensor factors, and the statement of Schur-Weyl duality is that these two actions are actually complete mutual centralizers of one another within $\mathrm{End}(V^{\otimes n})$.[72] Indeed, it is easy to see that $\mathrm{End}_{S_n}(V^{\otimes n})$ is constructed precisely from

[70] Although to some extent this is already obvious from the simplicity of the $\mathcal{M}_i$s.

[71] V splits into $\mathcal{A}f(m) \oplus U$ since $\mathcal{A}v$ is an invariant subspace of V for any $v \in V$.

[72] In this sense each action determines the other, hence the use of the word "duality". Of course though, this cannot be a one-to-one association between irreps of S_n and irreps of $U(d)$, since the latter are infinite (although I think the association does become one-to-one if we take an appropriate $n, d \to \infty$ limit).

matrices of the form $U^{\otimes n}$ with $U \in U(d)$.[73] From the double centralizer theorem, we then have

$$V^{\otimes n} \cong \bigoplus_\alpha V_\alpha \otimes \mathbb{S}_\alpha V, \tag{749}$$

where the sum is over the irreps of S_n V_α, and where $\mathbb{S}_\alpha V$ is fancy notation for the hom space in the double centralizer theorem (depending on the representation α, this may vanish):

$$\mathbb{S}_\alpha V = \mathrm{Hom}_{S_n}(V_\alpha, V^{\otimes n}). \tag{750}$$

By the double-centralizer theorem, we also know that the $\mathbb{S}_\alpha V$ are irreducible representations of $U(d)$. We can determine them semi-explicitly by

$$\begin{aligned} \mathbb{S}_\alpha V &\cong V_\alpha^* \otimes_{\mathbb{C}[S_n]} V^{\otimes n} \\ &\cong V_\alpha \otimes_{\mathbb{C}[S_n]} V^{\otimes n} \\ &\cong \mathcal{Y}_\alpha \mathbb{C}[S_n] \otimes_{\mathbb{C}[S_n]} V^{\otimes n} \\ &\cong \mathcal{Y}_\alpha V^{\otimes n}, \end{aligned} \tag{751}$$

where we have used that all the representations of S_n are self-dual.[74]

This is called a duality because the $U(d)$ and S_n actions determine each other within the space $V^{\otimes n}$. This has implications for quantum data compression and tomography; see e.g. [5] for further reading.

Unfortunately, working out examples is rather involved for all but the easiest cases. For example, take $n = 2$. $\mathcal{Y}_{(2)}$ projects onto those matrices which are symmetric under the interchange of the two $\otimes$ factors. Therefore

$$\mathcal{Y}_{(2)} V^{\otimes 2} = \frac{\mathbf{1} + \mathcal{S}}{2} V^{\otimes 2} = \mathrm{Sym}^2 V, \tag{752}$$

where $\mathcal{S}$ is the swap operator. Similarly, $\mathcal{Y}_{(1,1)}$ comes from the sign representation and so acts as

$$\mathcal{Y}_{(1,1)} V^{\otimes 2} = \frac{\mathbf{1} - \mathcal{S}}{2} V^{\otimes 2} = \mathrm{Alt}^2 V. \tag{753}$$

Just these simple facts are very useful for calculating Haar averages; see the relevant diary entry for more details. We will stop here though, since going further with the examples necessitates getting slightly into the weeds with the representation theory of the symmetric groups.

[73]Of course as written $\mathrm{End}_{S_n}(V^{\otimes n})$ is actually isomorphic to $GL(V, \mathbb{C})$ acting as nth tensor powers. We can however restrict our attention to the compact case where $GL(V, \mathbb{C})$ becomes $U(d)$, and will tacitly do so in what follows.

[74]Proof: because of the (non-canonical) bijection between conjugacy classes and irreps, it is enough to show that all conjugacy classes of S_n are self-dual. This is true if for all $g \in S_n$, there exists an $h \in S_n$ such that $hgh^{-1} = g^{-1}$. But it is straightforward that this is indeed the case for S_n: for example, if g is an exchange of two letters we can take h to be the identity, while if g is a cyclic permutation we can take h to be the element which reverses the order of any word. This shows that all the irreps are self-dual (in fact they are also all $\mathbb{R}$, but this is not important for our purposes.

Smoothing out piecewise defined functions

Today's diary entry is incredibly trivial, and just contains a reminder about a few different ways of smoothing functions.

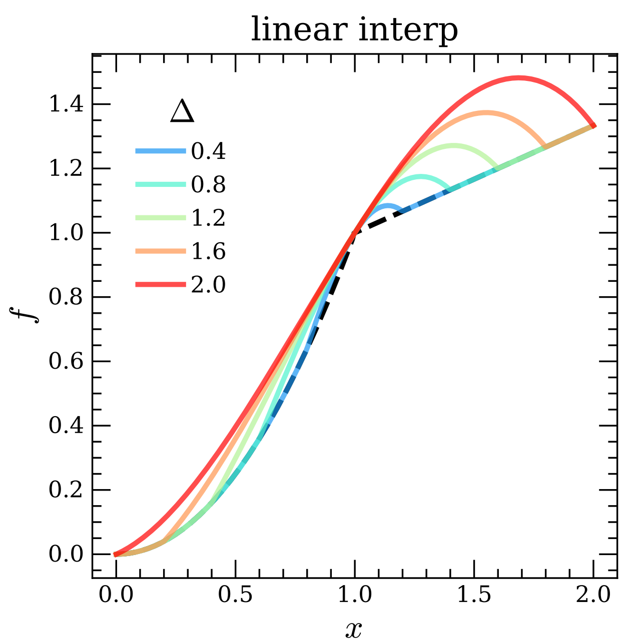
◇ ◇ ◇

In the following we will consider the smoothing of a function which behaves as $f_L(x)$ for $x < x_c$ and as $f_H(x)$ for $x \geq x_c$, with $f_L(x_c) = f_H(x_c)$ and both $f_{L,H}(x)$ smooth. When making plots, for concreteness we will take $f_L(x)$ to be quadratic in x and $f_H(x)$ to be linear, with $x_c = 1$.

It should first of all be clear that a linear interpolation

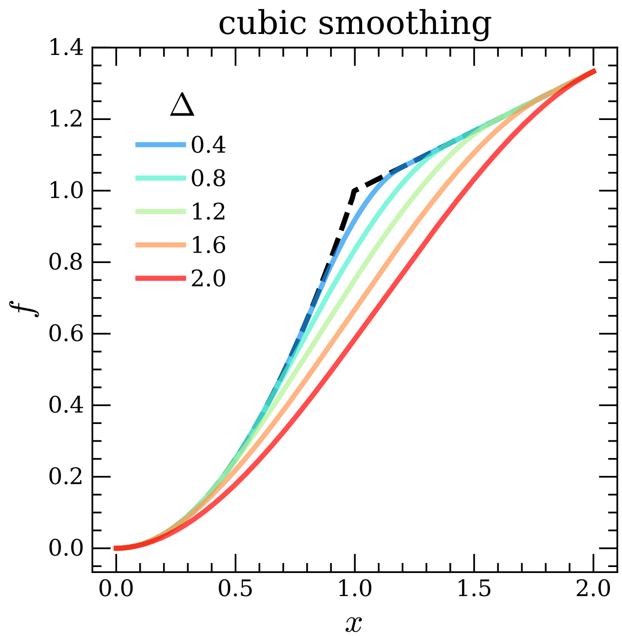
$$f_s(x) = (1 - s(x))f_L(x) + s(x)f_H(x), \qquad s = \begin{cases} 0 & x < x_c - \Delta/2 \\ \dfrac{x - (x_c - \Delta/2)}{\Delta} & x_c - \Delta/2 < x < x_c + \Delta/2 \\ 1 & x > x_c + \Delta/2 \end{cases} \tag{754}$$

is not going to work, because the resulting function is not smooth at $x = x_c \pm \Delta/2$. Indeed, this produces the disgusting



(755)

A cubic spline fit allows us to to better, where in the smoothing region we interpolate between f_L and f_H with a cubic polynomial. Since we have four parameters to tune, we can make both the resulting function everywhere smooth and its derivative everywhere continuous. For example, if we let Δ denote the width of the smoothing region (centered on x_c), then we have

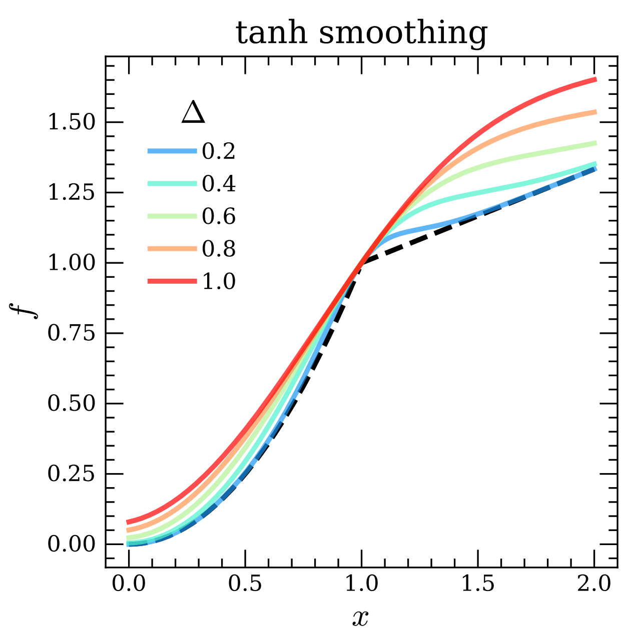


(756)

Of course the resulting function is only $\mathcal{C}^1$, and hence doesn't really count as smooth. In an attempt to do better, we may try doing a smoothened version of the linear interpolation, using some sort of sigmoid function. For example, we may take

$$f_s(x) = \frac{1 - \tanh\left(\frac{x - x_c}{\Delta}\right)}{2} f_L(x) + \frac{1 + \tanh\left(\frac{x - x_c}{\Delta}\right)}{2} f_H(x). \tag{757}$$

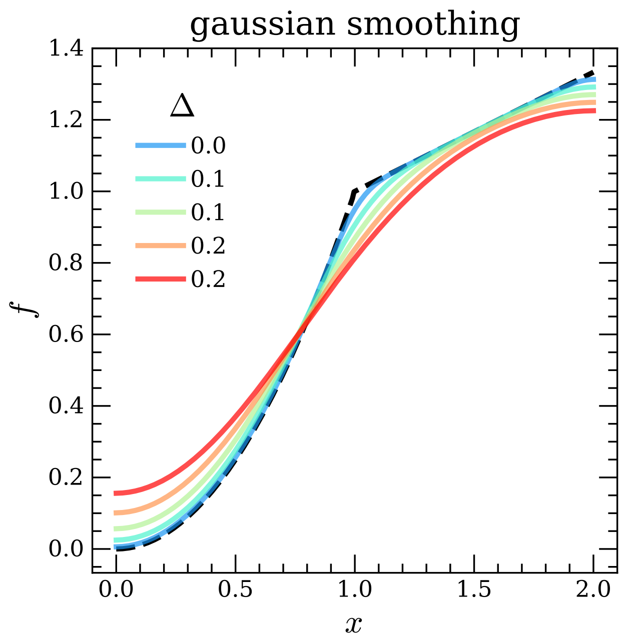
This produces curves which look like



(758)

One problem with all these approaches is that they do not preserve integrals. An approach which does preserve integrals is the canonical way of doing smoothing, viz. convolving with a $\mathcal{C}^\infty$ function. Since the convolution of a $\mathcal{C}^\infty$ function with a $\mathcal{C}^0$ one is $\mathcal{C}^\infty$ (when taking derivatives of the convolution, we can always change variables in the integral so that the derivatives only hit the smooth kernel), the result of the convolution will be $\mathcal{C}^\infty$, and the integral of f will be preserved if the kernel is normalized properly.

For example, consider a Gaussian kernel. If we let Δ denote the standard deviation of the Gaussian, we have



(759)

Of course the disadvantage to doing this are the edge effects which occur when working on a finite domain, which is something that e.g. the cubic spline interpolation doesn't suffer from. For the same reason, within this approach we are not able to smooth only along a finite window without introducing non-analyticitices at the window edges.

Learning about AdS space

The AdS metric in a presentation which makes the $SO(2,d)$ symmetry manifest is

$$ds^2 = -dX_{-1}^2 - dX_0^2 + \sum_{i=1}^{d} dX_i^2, \tag{760}$$

where the coordinates are constrained to cut out a hyperbola:

$$-X_{-1}^2 - X_0^2 + \sum_{i=1}^{d} X_i^2 = R^2. \tag{761}$$

Today's problem statement is as follows:

a) find coordinates in which the metric is the simpler

$$ds^2 = \frac{r^2}{R^2}dx_\mu dx^\mu + \frac{R^2}{r^2}dr^2. \tag{762}$$

b) find a transformation where the $1/r^2$ factor is pulled out front. c) find a formulation of ds^2 with only one timelike coordinate. d) what is the Pensrose diagram for AdS?

◇ ◇ ◇

a) Written as a hyperboloid in $(2, d)$ Lorentzian space, the coordinates of AdS_{d+1} are not all independent: there is one linear relation among them. We want to eliminate this dependency by introducing a new variable r and getting rid of X_{-1} and X_d. Now

$$dX^\mu = \frac{dr}{R}x^\mu + \frac{r}{R}dx^\mu \implies \sum_{i=0}^{d} dX_i dX^i = dX_d^2 + \frac{dr^2}{R^2}x^\mu x_\mu + \frac{r^2}{R^2}dx_\mu dx^\mu + \frac{2rdr}{R^2}dx^\mu x_\mu. \tag{763}$$

Here $\mu = 0, \ldots, d-1$ and the μ indices are raised / lowered with the mostly-positive Minkowski metric. so that

$$ds^2 = -dX_{-1}^2 + dX_d^2 + \frac{1}{R^2}\left(dr^2 x^\mu x_\mu + r^2 dx_\mu dx^\mu + rdrd(x_\mu x^\mu)\right). \tag{764}$$

Using the constraint, we can use

$$x_\mu x^\mu = \frac{R^2}{r^2}\left(X_{-1}^2 - X_d^2 - R^2\right) \tag{765}$$

to re-write this as

$$ds^2 = \frac{r^2}{R^2}dx_\mu dx^\mu - dX_{-1}^2 + dX_d^2 + \frac{dr^2}{r^2}(X_{-1}^2 - X_d^2 - R^2) - 2\frac{dr^2}{r^2}(X_{-1}^2 - X_d^2 - R^2) + \frac{2dr}{r}(X_{-1}dX_{-1} - X_d dX_d). \tag{766}$$

Expressing r in terms of X_{-1} and X_d and simplifying by giving all relevant terms a $1/r^2$ to factor out,

$$\begin{aligned} ds^2 = \frac{r^2}{R^2}dx_\mu dx^\mu + \frac{R^2}{r^2}dr^2 + \frac{1}{(X_d + X_{-1})^2}\Big((dX_{-1} + dX_d)^2(X_d^2 - X_{-1}^2) + 2(dX_d + dX_{-1})(X_{-1}dX_{-1} \\ - X_d dX_d)(X_d + X_{-1}) + (dX_d^2 - dX_{-1}^2)(X_d + X_{-1})^2\Big). \end{aligned} \tag{767}$$

Somewhat amazingly the terms in the big parentheses all cancel, and we get the simple

$$ds^2 = \frac{r^2}{R^2}dx_\mu dx^\mu + \frac{R^2}{r^2}dr^2. \tag{768}$$

b) If we let $z = R^2/r$, then $dr = -\frac{R^2}{z^2}dz$, so that $(R^2/r^2)dr^2 = (z^2/R^2)(R^4 z^{-4} dz^2)$ and the metric goes to

$$ds^2 = \frac{R^2}{z^2}(dz^2 + dx_\mu dx^\mu). \tag{769}$$

c) Now we take

$$X_0 = R\sqrt{1+r^2}\cos t, \quad X_{-1} = R\sqrt{1+r^2}\sin t, \qquad \sum_{i=1}^{d} X_i^2 = (Rr)^2. \tag{770}$$

The last relation means that the spatial X_i coordinates form a sphere of radius Rr. It also means that since e.g.

$$dX_0 = -R\sqrt{1+r^2}\sin t\,dt + R(1+r^2)^{-1/2}\cos t\,r dr, \tag{771}$$

we have

$$dX_0^2 + dX_{-1}^2 = R^2(1+r^2)dt^2 + \frac{R^2r^2}{1+r^2}dr^2. \tag{772}$$

Thus the metric is now

$$ds^2 = -R^2\left((1+r^2)dt^2 + \frac{r^2dr^2}{1+r^2}\right) + R^2dr^2 + r^2R^2d\Omega_{d-1}^2 = R^2\left(-(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + d\Omega_{d-1}^2\right). \tag{773}$$

Note that now the metric only has a single timelike coordinate. Now we let $r = \tan\rho$, so that $1+r^2 = 1/\cos^2\rho$ and $dr^2 = (1+\tan^2\rho)^2 d\rho^2$. Then the metric is

$$ds^2 = R^2(1+r^2)\left(-dt^2 + d\rho^2 + \frac{r^2}{1+r^2}d\Omega_{d-1}^2\right) = \frac{R^2}{\cos^2\rho}\left(-dt^2 + d\rho^2 + \sin^2\rho\, d\Omega_{d-1}^2\right). \tag{774}$$

d) The overall $(R/\cos\rho)^2$ factor in front doesn't affect the casual structure since the later is determined by right rays where $ds^2 = 0$, a condition which doesn't care about the Weyl rescaling of the metric. Thus the Penrose diagram can be determined just by looking at the terms inside of the parentheses. dt^2 gives us a copy of $\mathbb{R}$, and so we get a sort of solid cylinder, where the axis of the cylinder is at $\rho = 0$ ($r = 0$) and the surface of the cylinder is at $\rho = \pi/2$ ($r = \infty$). The surface of this cylinder is a copy of $\mathbb{R} \times S^{d-1}$. Note that this boundary $\mathbb{R} \times S^{d-1}$ lies at finite proper distance away from any point in the interior of the cylinder, despite the fact that the boundary is at $r = \infty$.

Perspectives on the Kelvin circulation theorem in two and three dimensions

This diary was inspired by wanting to understand parts of Dam Son's paper on chiral metric hydrodynamics [3]. The setting will be a general hydrodynamic theory of translation-invariant conserved bosons.

◇ ◇ ◇

The two fundamental hydrodynamic quantities in the present setting are the generators for the two conserved quantities, viz. momentum and particle number. We define the particle number and momentum densities as[75]

$$n(x) = \sum_\alpha \delta(x_\alpha - x), \qquad \pi^i(x) = \sum_\alpha \delta(x_\alpha - x) p^i_\alpha, \tag{775}$$

where the index α labels microscopic particles and where the phase space coordinates are $\{x_\alpha, p_\alpha\}$. We see that

$$\{n(x), n(y)\} = 0 \tag{776}$$

while

$$\begin{aligned} \{n(x), \pi^i(y)\} &= \sum_\alpha \delta(x_\alpha - y) \frac{\partial}{\partial x^i_\alpha} \delta(x_\alpha - x) \\ &= n(y) \frac{\partial}{\partial y^i} \delta(x - y) \end{aligned} \tag{777}$$

and

$$\begin{aligned} \{\pi^i(x), \pi^j(y)\} &= \sum_\alpha \left(p^i_\alpha \delta(x_\alpha - y) \frac{\partial}{\partial x^j_\alpha} \delta(x_\alpha - x) - p^j_\alpha \delta(x_\alpha - x) \frac{\partial}{\partial x^i_\alpha} \delta(x_\alpha - y) \right) \\ &= \left(\pi^i(y) \frac{\partial}{\partial y^j} - \pi^j(x) \frac{\partial}{\partial x^i} \right) \delta(x - y) \end{aligned} \tag{778}$$

The densities n, π^i allow us to generate symmetry actions by integrating them against various functions. For example, define the translation generator

$$Q_\xi = - \int_x \xi^k \pi_k. \tag{779}$$

The claim is that Q_ξ generates translations, in the sense that

$$\{Q_\xi, X\} = \mathcal{L}_\xi X, \tag{780}$$

where $\mathcal{L}_\xi$ is the Lie derivative along the vector field ξ. Indeed, using the above Poisson brackets one checks that[76]

$$\{Q_\xi, n(y)\} = - \int_x \xi^k \{\pi_k(x), n(y)\} = \int_x \xi^k n(x) \frac{\partial}{\partial x^k} \delta(x - y) = -\xi^i \partial_i n - (\partial_i \xi^i) n \tag{781}$$

[75] We could also take the quantum approach of working with commutators, but this is more annoying due to the fact that the ordering of operators matters. This means that e.g. the velocity has to be defined in Weyl quantization like $(\sim jn^{-1} + n^{-1}j)/2$, etc — see Landau's OG paper on HeII. Sometimes working in the QM context is nice since the commutator automatically satisfies the Jacobi identity, which depending on the situation might be a difficult property to satisfy when constructing a PB. Here we will have no such issues however, and it is more convenient to work with PBs from start to finish.

[76] There are almost certainly several minus signs wrong here, as I am being cavalier about index placement even though co/contra-variant indices appear with different signs in the Lie derivative.

and

$$\begin{aligned}\{Q_\xi, \pi^i(y)\} &= -\int_x \xi^k(x) \left(\pi_k(y) \frac{\partial}{\partial y^i} - \pi_i(x) \frac{\partial}{\partial x^k} \right) \delta(x-y) \\ &= -(\pi_k \partial_i \xi^k + \xi_k \partial_k \pi_i + \pi_i \partial_k \xi^k).\end{aligned} \tag{782}$$

The last terms proportional to $\nabla \cdot \boldsymbol{\xi}$ in these expressions arise because n, π^i are *densities*, i.e. they implicitly contain a factor of $\sqrt{|g|}$. To recall why this gives $\nabla \cdot \boldsymbol{\xi}$, remember that

$$\begin{aligned}\frac{1}{\sqrt{|g|}} \delta \sqrt{|g|} &= \frac{1}{2} \delta \mathrm{Tr}(\ln g) = \frac{1}{2} g^{ij} (\xi^k \partial_k g_{ij} + \partial^k \xi_i g_{kj} + \partial^k \xi_j g_{ik}) \\ &= \partial_k \xi^k,\end{aligned} \tag{783}$$

where the last line comes from $\partial_k \mathrm{Tr} g = 0$.

The momentum per particle on the other hand, viz.

$$u^i \equiv \pi^i / n, \tag{784}$$

transforms as a proper vector, without the $\nabla \cdot \boldsymbol{\xi}$ term in the PB with Q_ξ. The PB with π^i which makes this work is

$$\{\pi^i(x), u^j(y)\} = \left(u^i(y) \frac{\partial}{\partial x^j} + \frac{n(x)}{n(y)} \left(u^j(x) - u^j(y) \right) \frac{\partial}{\partial x^i} \right) \delta(x-y). \tag{785}$$

This is easy to check using the fact that the PB is a derivation, so that we have e.g.

$$\{A(f_1, f_2, \dots), B\} = \sum_n \frac{\partial A}{\partial f_n} \{f_n, B\}. \tag{786}$$

The fact that u^i transforms as a vector rather than a vector density is one of the reasons that we sometimes prefer to work with u^i rather than π^i.

These PBs can be used to derive the equations of motion for the particle number and momentum densities. The Hamiltonian will be a function of n and π, and we denote

$$v^i \equiv \frac{\partial H}{\partial \pi^i}, \qquad \mu \equiv \frac{\partial H}{\partial n}. \tag{787}$$

We then have the correct continuity equation[77] (note that in our conventions $d_t \mathcal{O} = \{\mathcal{O}, H\}$)

$$d_t n(y) = \int_x v_i(x) \{n(y), \pi^i(x)\} = -\partial_i (v^i n) = -\nabla \cdot \mathbf{j}, \tag{788}$$

as well as the correct equation of motion

$$\begin{aligned}d_t \pi_i(y) &= \int_x \left(v_j \{\pi^i(y), \pi^j(x)\} + \mu \{\pi^i(y), n(x)\} \right) \\ &= -\partial^j (v_j \pi_i) - \pi^j \partial_i v_j - n \partial_i \mu.\end{aligned} \tag{789}$$

[77]The equations of motion computed here are not comoving, in the sense that $d_t n$ does not include a term $u^i \partial_i n$.

The eom for the momentum per particle is also seen to be

$$d_t u^i(y) = -u^j \partial_i v_j - v^j \partial_j u^i - \partial_i \mu, \tag{790}$$

which accordingly lacks the $\partial_j v^j$ term of the π_i eom. As such, we may write

$$d_t u = -d\mu + \mathcal{L}_v u. \tag{791}$$

To perform a sanity check of these results, consider what happens when we take the Hamiltonian to be (setting the mass of the particles to be $m = 1$)

$$H = \int \left(\frac{\pi^2}{2n} + V(n) \right) \implies \mu = -\frac{u^2}{2} + \partial_n V, \qquad v_i = u_i. \tag{792}$$

The continuity equation then takes the usual form

$$d_t n = -\partial_i (u^i n) \implies D_t n \equiv d_t n + u^i \partial_i n = -n \partial_i u^i, \tag{793}$$

which is the form that one sees in most hydro contexts. The eom for u^i similarly becomes

$$D_t u^i = -\partial_i \partial_n V. \tag{794}$$

◇ ◇ ◇

We are now in a position to discuss the Kelvin circulation theorem. We will first work in two dimensions, which is a special case due to the fact that the vorticity is a scalar. Higher dimensions will be addressed in the next section.

Let us introduce the vorticity

$$\omega \equiv \star du. \tag{795}$$

Recalling that we are in two dimensions, we see that ω is a scalar density (u is a proper vector, but the $\star$ turns it into a density). The claim of Kevlin's circulation theorem is then that

$$I_F = \int_x n \, F(\omega/n) \tag{796}$$

is a conserved quantity, where F is an *arbitrary* function of its argument (which can be arbitrary since ω/n is a scalar, not a scalar density).

A sufficient condition for I_F to be conserved is that it have zero PB with both n and π^i, since these are the quantities that the Hamiltonian H is built out of. Indeed, this turns out to be true. First, it is easy to see that $\{I_F, n(x)\} = 0$, since $\{n, n\} = 0$ and

$$\begin{aligned} \{\omega(x), n(y)\} &= \varepsilon^{ij} \frac{\partial}{\partial x^i} \{u_j(x), n(y)\} \\ &= -\varepsilon^{ij} \frac{\partial}{\partial x^j} \left(\frac{\partial}{\partial x^i} \delta(x-y) \right) \\ &= 0. \end{aligned} \tag{797}$$

Furthermore since ω is a scalar density, it must have a PB with the momentum of

$$\{\omega(x), \pi^i(y)\} = \omega(y)\frac{\partial}{\partial y^i}\delta(x-y), \tag{798}$$

which is easy to check. Using this PB and the fact that the PB is a derivation, we then have (using the notation $z \equiv \omega/n$)

$$\begin{aligned}\{I_F, \pi^i(y)\} &= \int_x \left[F(z(x))\{n(x), \pi^i(y)\} + F'(z(x))\{\omega(x), \pi^i(y)\} - F'(z(x))z(x)\{n(x), \pi^i(y)\}\right] \\ &= -\int_x \left(F(z(x))n(y) + F'(z(x))\omega(y) - F'(z(x))z(x)n(y)\right)\frac{\partial}{\partial x^i}\delta(x-y) \\ &= n\partial_i F + \omega\partial_i F' - n\partial_i(F'z) \\ &= F'(\partial_i\omega - z\partial_i n) - F'n\partial_i z \\ &= 0.\end{aligned} \tag{799}$$

Therefore $\{I_F, H\} = 0$ and I_F is conserved for any F as claimed.

◇ ◇ ◇

In higher dimensions, the usual statement of the circulation theorem is given in terms of the conservation of the integral of the vorticity around closed curves, and is the statement that

$$\mathcal{I}_C = \oint_C u \tag{800}$$

is conserved for all closed curves C. We don't have the freedom of sticking in arbitrary functions of u into the integrand (since magnetic fields are scalars only in two dimensions).

The statement that $\mathcal{I}_C$ is conserved requires a bit of qualification. Certainly if C is fixed in time this cannot be true — if the curve C encloses a vortex at time t, the vortex may very well move to a location not enclosed by C at some later time. Instead, the statement is that $\mathcal{I}_C$ is conserved provided that we take C to be comoving with the fluid flow. Mathematically, this means that $C = C(t)$, with

$$d_t C = \mathcal{L}_v C. \tag{801}$$

The proof that $\mathcal{I}_C$ is conserved is then rather trivial:

$$\begin{aligned}d_t\mathcal{I}_C &= \frac{d}{dt}\int \widehat{C} \wedge u \\ &= \int \left(\mathcal{L}_v\widehat{C} \wedge u + \widehat{C} \wedge (\mathcal{L}_v u - d\mu)\right) \\ &= \int \mathcal{L}_v(\widehat{C} \wedge u),\end{aligned} \tag{802}$$

since $d\widehat{C} = 0$. The using Cartan's magic formula,

$$\begin{aligned}d_t\mathcal{I}_C &= \int (i_v d + d i_v)(\widehat{C} \wedge u) \\ &= \int d i_v(\widehat{C} \wedge u) \\ &= 0,\end{aligned} \tag{803}$$

as $d(\widehat{C} \wedge u) = 0$ on account of $\widehat{C} \wedge u$ being top-dimensional. Note that this proof makes no reference to the fact that we are in three dimensions.

Differential geometry calisthenics

In today's diary entry we are going to be proving a few semi-trivial low-dimensional differential geometry facts, which are solutions to exercises in the first few chapters in Tu's book. We will be using math notation, with $\mathfrak{X}(M)$ denoting C^∞ vector fields on a manifold M.

◇ ◇ ◇

Proposition 11. *Let $c(t) = (x(t), y(t))$ be a curve in the plane. The curvature along c is given by*

$$\kappa = \frac{x''y' - x'y''}{(x'^2 + y'^2)^{3/2}}, \tag{804}$$

where primes denote derivatives wrt t.

Proof. Recall that the curvature is computed as

$$\kappa = \langle d_s T(s), N \rangle, \tag{805}$$

where $T(s)$ is the tangent vector to c in the arc length parametrization and N is the unit normal to the curve. The tangent vector is

$$T(s) = d_s c(t) = \frac{(x', y')}{\sqrt{x'^2 + y'^2}}, \tag{806}$$

where we used

$$s(t) = \int_0^t d\lambda \, ||c'(\lambda)|| \implies d_s t = \frac{1}{\sqrt{x'^2 + y'^2}}. \tag{807}$$

Now we need to differentiate this again with respect to s. After some algebra that we won't write out,

$$d_s T = \frac{1}{(x'^2 + y'^2)^2}(x''y'^2 - x'y'y'', y''x'^2 - y'x'x''), \tag{808}$$

where we have remembered to include a factor of $1/\sqrt{x'^2 + y'^2}$ coming from $d_s t$.

The unit normal to the curve is

$$N = \frac{(y', -x')}{\sqrt{x'^2 + y'^2}}, \tag{809}$$

and so, after some more algebra, we take the inner product and indeed get

$$\kappa = \frac{x''y' - x'y''}{(x'^2 + y'^2)^{3/2}}. \tag{810}$$

□

Example 11. Consider an ellipse with semi-axes a and b. We can then use the above formula for κ to compute

$$\kappa(x, y) = \frac{ey^2 + e^{-1}x^2}{(e^2y^2 + e^{-2}x^2)^{3/2}}, \tag{811}$$

where $e \equiv a/b$.

Proposition 12. *The mean curvature of a surface at a point p is the circular average of the normal curvatures at p.*

Proof. The mean curvature of a surface at a point p is defined as

$$\bar{\kappa} = \frac{\kappa_1 + \kappa_2}{2}, \tag{812}$$

where κ_i are the principal curvatures at p, viz. the eigenvalues of the second fundamental form Π at p. For $X \in \mathfrak{X}(M)$ a unit tangent vector field, the curvature along X is $\Pi(X_p, X_p)$. We may thus average over X by writing X in local coordinates at p as $X(\theta) = e_1 \cos\theta + e_2 \sin\theta$, where the e_i are an orthonormal basis for the tangent space at p. Therefore

$$\frac{1}{2\pi} \int d\theta\, (\cos\theta,\, \sin\theta)_i \Pi^{ij} (\cos\theta,\, \sin\theta)_j = \frac{1}{2\pi}(\pi\delta_{ij})\Pi^{ij} = \frac{1}{2}\mathrm{Tr}[\Pi] = \bar{\kappa}, \tag{813}$$

which is what we wanted to show.

□

Proposition 13. *Let $M \subset \mathbb{R}^3$ be a surface. The Gauss map is a map $\nu : M \to S^2$ which sends*

$$\nu :\, M \ni p \mapsto N_p, \tag{814}$$

with N the unit normal vector field on M. Then we have

$$\int_M |K| = \mathrm{Area}[\nu(M)], \tag{815}$$

where K is the Gaussian curvature. This is a baby step to the GB theorem.

Proof. Work in local coordinates, with orthonormal tangent space basis vectors e_1, e_2. Recall that the Gaussian curvature is

$$K = \det[\langle L(e_i), e_j \rangle], \tag{816}$$

where L is the shape operator. Therefore

$$\int_M |K| = \int_M de^1 \wedge de^2\, |\partial_1 N^1 \partial_2 N^2 - \partial_1 N^2 \partial_2 N^1|. \tag{817}$$

We recognize this as the pullback of the volume form on S^2, so that

$$\int_M |K| = \int_M \nu^*(\mathrm{vol}_{S^2}) = \int_{\nu(M)} \mathrm{vol}_{S^2} = \mathrm{Area}[\nu(M)]. \tag{818}$$

□

Corollary 1. *Any ellipsoid has total curvature* 4π*.*

Indeed, $\nu(M)$ is bijective if M is an ellipsoid, since the normal vector field N is different at every point on the ellipsoid.

Proposition 14. *Let a surface* $M \subset \mathbb{R}^3$ *be parametrized by the function* $\sigma(x, y) = (x, y, h(x, y))$*, for some function* h*. Then the Gaussian curvature at a point* x, y *is given by*

$$K = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1 + h_x^2 + h_y^2)^2}. \tag{819}$$

Proof. First we will need to know the induced metric on M. This is done by computing the pullbacks of the basis vectors ∂_x, ∂_y by σ:

$$E_1 \equiv \sigma_*(\partial_x) = \partial_x + h_x\partial_z, \qquad E_2 \equiv \sigma_*(\partial_y) = \partial_y + h_y\partial_z. \tag{820}$$

Then

$$g_{ij} = \langle E_i|E_j\rangle = \begin{pmatrix} 1 + h_x^2 & h_x h_y \\ h_x h_y & 1 + h_y^2 \end{pmatrix}. \tag{821}$$

By taking the cross product of E_1 and E_2, we see that the unit normal vector field on M is given by

$$N = \frac{(-h_x,\, -h_y,\, 1)}{\sqrt{1 + h_x^2 + h_y^2}}. \tag{822}$$

Now the second fundamental form on M is

$$\langle E_i|\Pi|E_j\rangle = \langle L(E_i)|E_j\rangle, \tag{823}$$

where $L(E_i) = -(E_i \cdot \partial)N$ is the shape operator. Since $\partial\langle N|E_j\rangle = 0$, this is

$$\langle E_i|\Pi|E_j\rangle = \langle N|(E_i \cdot \partial)E_j\rangle. \tag{824}$$

Now

$$(E_i \cdot \partial)E_j = (0, 0, \partial_i\partial_j h). \tag{825}$$

Therefore

$$\langle E_i|\Pi|E_j\rangle = \frac{1}{\sqrt{1 + h_x^2 + h_y^2}} \begin{pmatrix} h_{xx} & h_{xy} \\ h_{xy} & h_{yy} \end{pmatrix}. \tag{826}$$

Now it would actually be a mistake to take the determinant at this point (note that the power in the denominator would not be correct). This is because the vectors E_i are not normalized, while for the Gaussian curvature we always want to compute in an orthogonal basis. One way to get the normalization right (note: I have no idea if this is actually something math people do) is to define the adjoint using the inverse of the metric. We therefore define the vectors

$$|e_i\rangle = |E_i\rangle, \qquad \langle e_i| = g^{ij}\langle E_j|, \tag{827}$$

where as usual the upper indices on g indicate the inverse.

We now take the determinant of Π in the $|e_i\rangle$ basis. This gives

$$K = \det[g^{-1}\,\Pi_E], \tag{828}$$

where Π_E is the expression in (826). Therefore

$$K = \frac{(h_{xx}h_{yy} - h_{xy}^2)/(\sqrt{1+h_x^2+h_y^2})}{(1+h_x^2)(1+h_y^2) - h_x^2 h_y^2} = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1+h_x^2+h_y^2)^2}, \tag{829}$$

which is what we wanted to show.

□

Proposition 15. *The sum of two connections is not a connection.*

Proof. This is trivial, and is just included as a reminder to myself. Recall that a connection is a map

$$\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M), \qquad (X, Y) \mapsto \nabla_X Y, \tag{830}$$

which is linear in the first argument and satisfies

$$\nabla_X(fY) = (Xf)Y + f\nabla_X Y \tag{831}$$

for any function f on M. Linearity in the first argument is why the sum of two connections is not a connection, as

$$(\nabla + \nabla')_{fX} Y = 2(Xf)Y + f(\nabla + \nabla')_X Y. \tag{832}$$

Therefore we may instead get a connection by taking a convex linear combination: if ∇_i are connections then so too is $\widetilde{\nabla}$, where

$$\widetilde{\nabla} = \sum_i t_i \nabla_i, \qquad \sum_i t_i = 1. \tag{833}$$

□

Proposition 16. *Let $M \subset \mathbb{R}^3$ be a smooth surface, and let*

$$\mathrm{pr} : T_p\mathbb{R}^3 \to \mathfrak{X}(M) \tag{834}$$

be the projection map. Then the Riemannian connection on M is given by

$$\nabla_X Y = \mathrm{pr}(D_X Y), \tag{835}$$

where D is the directional derivative in $\mathbb{R}^3$.

Proof. For the proof, we will take for granted the fact that the Riemannian connection is the *unique* connection on M that is torsion free

$$T_\nabla(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = 0 \tag{836}$$

and compatible with the metric (i.e. acts covariantly on inner products)

$$Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle, \tag{837}$$

where $X, Y, Z \in \mathfrak{X}(M)$.

We then just have to check that the projection defined above satisfies these two properties. First for the torsion. The directional derivative in Euclidean space satisfies

$$D_X Y - D_Y X = [X, Y], \tag{838}$$

and as such

$$T_\nabla(X, Y) = \mathrm{pr}([X, Y]) - [X, Y]. \tag{839}$$

Since the bracket is in $\mathrm{End}(\mathfrak{X}(M))$,[78] $\mathrm{pr}([X, Y]) = [X, Y]$, and hence $T_\nabla = 0$.

Now for the metric compatibility. Since $Y \in \mathfrak{X}(M)$, we have

$$\langle \nabla_Z X, Y\rangle = \langle D_Z X - \langle D_Z X, N\rangle N, Y\rangle = \langle D_Z X, Y\rangle, \tag{840}$$

and likewise for the other term. Thus

$$\langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle = \langle D_Z X, Y\rangle + \langle X, D_Z Y\rangle. \tag{841}$$

But now the RHS can be treated as living in $\mathbb{R}^3$. Since the metric in $\mathbb{R}^3$ commutes with the derivatives, the D_Zs pull out of the inner products, and we see that we indeed have metric compatibility.

Thus the projection indeed gives the Riemannian connection. This is the best way of understanding covariant derivatives and connections in gauge theory, where instead of embedding a surface in $\mathbb{R}^3$ we're embedding a base space inside of a fiber bundle. The math is the same though, with the connection providing a way of separating the directional derivative on the full space into horizontal and vertical parts.

□

Proposition 17. *The Lie bracket transforms in the way one expects it to under diffeomorphisms: if $\phi : M \to N$ is a diffeomorphism, then*

$$\phi_*([X, Y]) = [\phi_* X, \phi_* Y] \tag{842}$$

for $X, Y \in \mathfrak{X}(M)$.

Proof. This is just an application of the chain rule. Let $\phi_*(dx^i) = \frac{\partial x^i}{\partial y^a} dy^a$, with i running over coordinates in M and a running over those in N. Then

$$\phi_*(X) = X^i \frac{\partial y^a}{\partial x^i} \partial_a \tag{843}$$

[78] Abstractly, this is one direction of Frobenius's theorem, which states that a distribution is closed under $[,]$ iff it is (locally) tangent to a submanifold (in this case M). A more explicit proof is given in the next proposition.

and likewise for $\phi_*(Y)$, and hence

$$[\phi_*(X), \phi_*(Y)] = X^i \frac{\partial y^a}{\partial x^i} \partial_a \left(Y^j \frac{\partial y^b}{\partial x^j} \partial_b \right) - (X \leftrightarrow Y) \tag{844}$$

The only terms which don't cancel by commutativity of partials are the one where the coefficients of Y get acted on by the derivative and the similar term with $X \leftrightarrow Y$, and so

$$\begin{aligned}
[\phi_*(X), \phi_*(Y)] &= X^i \frac{\partial y^a}{\partial x^i} \frac{\partial x^k}{\partial y^a} (\partial_k Y^j) \frac{\partial y^b}{\partial x^j} \partial_b - (X \leftrightarrow Y) \\
&= X^i \partial_i Y^j \frac{\partial y^b}{\partial x^j} \partial_b - (X \leftrightarrow Y) \\
&= (X^i \partial_i Y^j - Y^i \partial_i X^j) \phi_*(\partial_j) \\
&= \phi_*([X, Y]),
\end{aligned} \tag{845}$$

as claimed. □

Corollary 2. *If X and Y are tangent to a submanifold $M \subset \mathbb{R}^n$, then so too is $[X, Y]$.*

Proof. Consider a local patch U of M. By definition, there is a chart ϕ giving a diffeomorphism between U and $\mathbb{R}^{\dim M}$. In the image of ϕ, vector fields in $\mathfrak{X}(M)$ have components whose last $n - \dim M$ entries are all zero. A basis for the normal complement of the tangent space of $\phi(U)$ is given by those vectors which are nonzero only in the last $n - \dim M$ components. Thus in $\phi(M)$, it is obvious that if a vector N is orthogonal to the tangent space of $\phi(U)$, it will also be orthogonal to the Lie bracket of any two vectors in the tangent space of $\phi(U)$. Applying the above pushforward formula to translate these results back to M itself gives the desired result. □

Conservation of $T^{\mu\nu}$, different kinds of stress tensors, and useful covariant derivative identities

Today we're doing a short calculation showing why $T^{\mu\nu}$ is conserved. There's a nontrivial integration by parts that is normally glossed over which we will explain. We'll also explain the relation between the two different usual ways of calculating $T_{\mu\nu}$ (by varying the fields or by varying the metric), which is annoyingly not really explained correctly in any of my books.

◇ ◇ ◇

The covariant conservation of the stress tensor in a given theory defined on the Riemannian manifold (M, g) follows from the invariance of the action under isometries $f : (M, g) \to (M, f_* g)$. f here is an isometry since distances computed on $M = (M, g)$ match those computed in $f(M) = (f(M), f_* g)$: the line element satisfies

$$ds^2_M(x) = ds^2_{f(M)}(f(x)). \tag{846}$$

These isometries are often called diffeomorphisms by physicists, and here we will adopt this terminology as well (for better or worse).

Consider an infinitesimal flow along a diffeomorphism $f : M \to M$, where the coordinates are mapped as $x^\mu \mapsto x^\mu + \xi^\mu$. The action $\int d^dx\,\sqrt{g}\,\mathcal{L}$ is unchanged under both mapping the coordinates under the diffeomorphism and pulling back the fields along the diffeomorphism, since the combination of both is a reparametrization of our coordinate system (mathematically, it is an isometry), under which all theories whose Lagrangians don't explicitly depend on the coordinates (i.e. all non-pathological theories in physics) are invariant.

After mapping the fields under the diffeomorphism by replacing them with their pullbacks (but not changing the coordinates), the Lagrangian changes as

$$\delta(\sqrt{g}\mathcal{L}) = \left(\mathcal{L}_\xi\phi\frac{\delta}{\delta\phi} + (\mathcal{L}_\xi g)_{\mu\nu}\frac{\delta}{\delta g_{\mu\nu}}\right)(\sqrt{g}\,\mathcal{L}), \tag{847}$$

where $\mathcal{L}_\xi$ is the Lie derivative along ξ, not to be confused with the Lagrangian (sorry!). Here the fields ϕ could be any sorts of fields (well, not spinors, but just for simplicity), so that if $\phi = A_\mu$ is a vector field, $(\mathcal{L}_\xi A)_\mu = \xi^\nu\partial_\nu A + \partial_\lambda\xi^\mu A^\lambda$. Now when we translate all the fields appearing in $\mathcal{L}$ along ξ, the Lagrangian changes by a total derivative, since we are just moving the Lagrangian infinitesimally along the flow:

$$\delta(\sqrt{g}\mathcal{L}) = \partial_\mu(\xi^\mu\sqrt{g}\mathcal{L}). \tag{848}$$

As a very simple check, consider a mass term for a scalar φ. Then

$$\mathcal{L}_\xi\varphi\frac{\delta}{\delta\varphi}(\sqrt{g}\varphi^2) = \sqrt{g}\xi^\mu\partial_\mu\varphi^2, \qquad (\mathcal{L}_\xi g)_{\mu\nu}\frac{\delta}{\delta g_{\mu\nu}}(\sqrt{g}\varphi^2) = \frac{1}{2}\sqrt{g}\varphi^2(g^{\mu\nu}\xi^\lambda\partial_\lambda g_{\mu\nu} + 2\partial\cdot\xi) \tag{849}$$

where we used that the Lie derivative is

$$(\mathcal{L}_\xi g)_{\mu\nu} = \xi^\alpha\partial_\alpha g_{\mu\nu} + \partial_\mu\xi^\alpha g_{\alpha\nu} + \partial_\nu\xi^\alpha g_{\mu\alpha}. \tag{850}$$

Now

$$\partial_\lambda\sqrt{g} = \frac{1}{2}\sqrt{g}g^{\mu\nu}\partial_\lambda g_{\mu\nu}, \tag{851}$$

and so

$$(\mathcal{L}_\xi g)_{\mu\nu}\frac{\delta}{\delta g_{\mu\nu}}(\sqrt{g}\varphi^2) = \varphi^2\partial_\mu(\sqrt{g}\xi^\mu). \tag{852}$$

Therefore the total variation is

$$\delta(\sqrt{g}\mathcal{L}) = \partial_\mu(\sqrt{g}\xi^\mu\varphi^2) \tag{853}$$

as expected.

A less trivial example is something which involves vectors and derivatives, like $\sqrt{g}A_\mu\partial^\mu\varphi$. The variations over the matter fields give

$$\left((\mathcal{L}_\xi A)_\mu\frac{\delta}{\delta A_\mu} + \mathcal{L}_\xi\varphi\frac{\delta}{\delta\varphi}\right)(\sqrt{g}A_\mu\partial^\mu\varphi) = \sqrt{g}\left(\partial^\mu\varphi(\xi^\nu\partial_\nu A_\mu + \partial_\mu\xi^\nu A_\nu) + A_\mu\partial^\mu(\xi^\nu\partial_\nu\varphi)\right), \tag{854}$$

while the variation over the metric produces, after some algebra,

$$(\mathcal{L}_\xi g)_{\mu\nu}\frac{\delta}{\delta g_{\mu\nu}}(\sqrt{g}A_\mu\partial^\mu\varphi) = (\xi^\lambda\partial_\lambda\sqrt{g}+\partial\cdot\xi)A_\mu\partial^\mu\varphi - A^\mu\partial^\nu\varphi\xi^\alpha\partial_\alpha g_{\mu\nu} - A_\lambda\partial_\alpha\xi^\lambda\partial^\alpha\varphi - A_\alpha\partial^\alpha\xi_\lambda\partial^\lambda\varphi. \tag{855}$$

Since $\partial_\alpha g_{\alpha\sigma} = -g_{\mu\nu}\partial_\alpha g^{\nu\lambda}g_{\lambda\sigma}$, the third term on the RHS is actually $+A_\mu\partial_\nu\varphi\xi^\alpha\partial_\alpha g^{\mu\nu}$. Adding up the contributions from the matter and the metric, we see that the last two terms on the LHS of the above equation cancel with two of the terms in the variation of the matter fields, again leaving us with

$$\delta(\sqrt{g}\mathcal{L}) = \partial_\mu(\sqrt{g}\xi^\mu A_\nu\partial^\nu\varphi). \tag{856}$$

Now we can see the general pattern that's at work: the variation over the metric produces derivatives of $\sqrt{g}$, ξ, and any $g_{\mu\nu}$s appearing in $\mathcal{L}$, plus some extra stuff coming from transforming the indices of the $g_{\mu\nu}$s. The variation over the matter fields produces derivatives of thins involving the matter fields, plus some extra stuff coming from the transformation of any vector indices that the matter fields have. The extra stuff from the matter fields and the extra stuff from the variation of the $g_{\mu\nu}$s cancel, since upper-index variations cancel lower-index ones. After the smoke clears, we are left with a total derivative.

Another way to see this is simply to write the Lagrangian as a d-form. If $L = \mathcal{L}\cdot\star 1$, where $\star 1$ is the volume form, then under the variation we have

$$\delta_\xi L = \mathcal{L}_\xi L = (i_\xi d + d i_\xi)L = d(i_\xi L), \tag{857}$$

where we have used Cartan's formula and that L is a top-dimensional form. Therefore the variation of the Lagrangian is indeed always a total derivative.

Since our diffeomorphism must vanish at ∂M, the upshot to the above discussion is that (no $\sqrt{g}$ in the integration measure on the LHS; it will be picked up from varying the action)

$$\int_M d^dx\left((\mathcal{L}_\xi g)_{\mu\nu}\frac{\delta}{\delta g_{\mu\nu}} + \mathcal{L}_\xi\phi\frac{\delta}{\delta\phi}\right)S[\phi, g_{\mu\nu}] = \int_{\partial M} d^{d-1}x^\mu\sqrt{g|_{\partial M}}\xi_\mu\mathcal{L}|_{\partial M} = 0. \tag{858}$$

Now by definition,[79] (we are working in Euclidean signature; hence the sign)

$$\frac{2}{\sqrt{g}}\frac{\delta S}{\delta g_{\mu\nu}} = T^{\mu\nu}. \tag{861}$$

[79]Not really. This is the correct definition to make in most circumstances, but sometimes the action depends explicitly on the vielbeins themselves, instead of just the metric (e.g. when doing fermions or problems with torsion). In this case, the correct definition is rather

$$T^\mu_a = \frac{1}{\det e}\frac{\delta S}{\delta e^a_\mu}. \tag{859}$$

In the presence of e.g. torsion this will not be symmetric.

If the only dependence of S on e is of the form $g_{\mu\nu} = e^a_\mu e^b_\nu\eta_{ab}$, then

$$\frac{1}{\det e}\frac{\delta S}{\delta e^a_\mu} = \frac{1}{\sqrt{g}}\frac{\delta S}{\delta g^{\lambda\sigma}}(\eta^{ad}e^\sigma_d\delta^{\mu\lambda} + \eta^{ca}e^\lambda_c\delta^{\sigma\mu}) \tag{860}$$

If we then define $T^{\mu\nu} = T^\mu_a e^\mu_b\eta^{ab}$ we get the same expression for $T^{\mu\nu}$ as written above.

The conservation of currents associated to global symmetries holds on shell. So, choosing a configuration of matter fields which solves the EOM, we see that

$$\int_M d^dx\,\sqrt{g}\,(\mathcal{L}_\xi g)_{\mu\nu}T^{\mu\nu} = 0, \tag{862}$$

on shell.

To get a conservation law, we need to massage the Lie derivative slightly. First, add and subtract $\Gamma^\lambda_{\alpha\mu}g_{\lambda\nu} + \Gamma^\lambda_{\alpha\nu}g_{\mu\lambda}$ on the RHS of the expression (850) for the Lie derivative. The positive terms combine with the first term above to produce a covariant derivative of $g_{\mu\nu}$, which dies. The negative terms are the connection coefficients needed to turn the derivatives of ξ into covariant ones. Therefore

$$(\mathcal{L}_\xi g)_{\mu\nu} = \nabla_{(\mu}\xi_{\nu)}. \tag{863}$$

Since the stress tensor is symmetric, we then have

$$\int_M d^dx\,\sqrt{g}\,(\nabla_\mu\xi_\nu)T^{\mu\nu} = 0. \tag{864}$$

Now naively we would like to integrate by parts and conclude that T is covariantly conserved. This is correct, but nonzero work is required to demonstrate it. First, we need the product rule

$$\nabla_\mu(T^{\mu\nu}\xi_\nu) = (\nabla_\mu T^{\mu\nu})\xi_\nu + T^{\mu\nu}\nabla_\mu\xi_\nu. \tag{865}$$

The first term on the RHS has two Christoffel symbol terms, which both have minus signs, while the second term on the RHS has one positive Christoffel symbol. Two of these cancel, leaving a single negative-sign Christoffel symbol, which just the right index structure to match the $\Gamma^\mu_{\mu\lambda}T^{\lambda\nu}\xi_\nu$ term on the LHS. Since this is straightforward algebra, and I won't write it out.

Now we need to argue that $\sqrt{g}\nabla_\mu(T^{\mu\nu}\xi_\nu)$ is a total derivative. This is only true because the covariant derivative is acting on a vector; if it was acting on a larger-rank tensor, it would not be a total derivative. Indeed, for any vector V^μ, we have

$$\begin{aligned}\frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}V^\mu) &= \partial_\mu V^\mu + \frac{1}{2}V^\mu g^{\alpha\beta}\partial_\mu g_{\alpha\beta}\\ &= \partial_\mu V^\mu + \frac{1}{2}V^\mu g^{\alpha\beta}(\Gamma^\lambda_{\mu\alpha}g_{\lambda\beta} + \Gamma^\lambda_{\mu\beta}g_{\alpha\lambda})\\ &= \nabla_\mu V^\mu,\end{aligned} \tag{866}$$

where in the first line we used the usual way of differentiating the determinant by writing it as $e^{\frac{1}{2}\mathrm{Tr}\ln g}$, and in the next line we used the metric compatibility of the connection to substitute in some Christoffel symbols for the derivative of the metric. Therefore we indeed have

$$\sqrt{g}\nabla_\mu(T^{\mu\nu}\xi_\nu) = \partial_\mu(\sqrt{g}\,T^{\mu\nu}\xi_\nu). \tag{867}$$

Therefore, taking ξ^μ to vanish at ∂M, we conclude that

$$\int_M d^dx\,\sqrt{g}\,\xi_\nu\nabla_\mu T^{\mu\nu} = 0, \tag{868}$$

and since this must hold for any ξ_ν, we have

$$\nabla_\mu T^{\mu\nu} = 0, \tag{869}$$

as required.

Finally, we can take a look at the relationship between the stress tensor defined here and the "canonical" stress tensor $T_c^{\mu\nu}$ obtained from the Noether procedure. Under $\phi(x^\mu) \mapsto \phi(x^\mu + \xi^\mu)$, the action changes as

$$\delta S = \int_M d^d x \sqrt{g}\, \nabla_\mu \xi_\nu T_c^{\mu\nu}. \tag{870}$$

Since this variation is the one produced by the second term on the LHS of (858), we have, for any matter field configuration (on-shell or off),

$$\begin{aligned} \int_M d^d x \sqrt{g}\, \nabla_\mu \xi_\nu T_c^{\mu\nu} &= \int_M d^d x \sqrt{g}\, (\mathcal{L}_\xi g)_{\mu\nu} \frac{\delta}{\delta g_{\mu\nu}} S[\phi, g_{\mu\nu}] \\ &= \int_M d^d x \sqrt{g}\, \nabla_\mu \xi_\nu T^{\mu\nu}. \end{aligned} \tag{871}$$

Since this must hold for any ξ_ν, we may take the matter fields to be off-shell (so that $T^{\mu\nu}$ has a non-zero divergence), and conclude that

$$\nabla_\mu T_c^{\mu\nu} = \nabla_\mu T^{\mu\nu}, \tag{872}$$

which tells us that the two stress tensors agree up to something whose divergence is zero, namely the components of $d^\dagger B$, where B is a 3-form.

Entropy relations brainwarmer

This diary entry proves some entropy relations that I had to write up as homework for a class.

◇ ◇ ◇

$S_A = S_B$ if ψ_{AB} is pure

Write the pure state $\psi \in \mathcal{H}_A \otimes \mathcal{H}_B$ in self-evident notation as

$$\psi = \sum_{i,j \in \mathcal{H}_A} \sum_{a,b \in \mathcal{H}_B} C_{ia}^* C_{jb} |ia\rangle\langle jb|. \tag{873}$$

Then the RDMs on A and B are

$$\psi_A = C^* C^T, \qquad \psi_B = C^\dagger C. \tag{874}$$

We then do an SVD on C, writing $C = U \Lambda V^\dagger$ for U, V unitary and Λ real-valued. Then

$$\psi_A = U^* \Lambda \Lambda^T U^T, \qquad \psi_B = V \Lambda^T \Lambda V^\dagger. \tag{875}$$

Thus

$$S_A = -\mathrm{Tr}[\Lambda\Lambda^T \log(\Lambda\Lambda^T)] = -\mathrm{Tr}[\Lambda^T\Lambda \log(\Lambda^T\Lambda)] = S_B, \tag{876}$$

as claimed.

Triangle inequality

We claim that if ψ_{ABC} is pure, then $S_A \leq S_B + S_C$. Indeed, using the above fact, we have $S_A = S_{BC}$, so we just need to show the positivity of the mutual information, viz. that

$$I(B:C) = S_B + S_C - S_{BC} \geq 0. \tag{877}$$

This can be done by using the non-negativity of the relative entropy $D(\rho||\sigma)$, together with the fact that $I(B:C) = D(\rho_{BC}||\rho_B \otimes \rho_C)$. The latter fact is a simple consequence of the definition. To prove the former fact, let ρ, σ be two dmats on the same Hilbert space, and let U be the unitary that relates the eigenbasis $|a\rangle$ of σ to the eigenbasis $|i\rangle$ of ρ. Then

$$\begin{aligned} D(\rho||\sigma) &= \sum_i \lambda_i \log(\lambda_i) - \sum_{i,a} |U_{i,a}|^2 \lambda_i \log(\eta_a) \\ &= \boldsymbol{\lambda}^T \cdot \log(\boldsymbol{\lambda}) + \boldsymbol{\lambda}^T M \log(1/\boldsymbol{\eta}), \end{aligned} \tag{878}$$

where the matrix $M_{ia} \equiv |U_{i,a}|^2$ is doubly stochastic, by the unitarity of U. Since the entries of $\boldsymbol{\lambda}$ and $\log(1/\boldsymbol{\eta})$ are positive, the second term is always minimized if M is a single permutation matrix. Thus after a possible re-ordering of the eigenbasis $|a\rangle$,

$$D(\rho||\sigma) \geq \boldsymbol{\lambda}^T \cdot \log(\boldsymbol{\lambda}) + \boldsymbol{\lambda}^T \log(1/\boldsymbol{\eta}) = D(\lambda||\eta), \tag{879}$$

where $D(\lambda||\eta)$ is the relative entropy of the classical probability distributions λ_i, η_i. But this is always positive: using $\log(1/x) \geq 1 - x$ by the convexity of the logarithm,

$$D(\lambda||\eta) \geq \sum_i \lambda_i (1 - \eta_i/\lambda_i) = 1 - 1 = 0. \tag{880}$$

Thus $D(\rho||\sigma) \geq 0$ as claimed (with equality only when $\rho = \sigma$).

Araki-Lieb

We claim that for any ρ_{AB},

$$|S_A - S_B| \leq S_{AB}. \tag{881}$$

This follows from the above triangle equality by using purifications. Let C be a system purifying ρ_{AB}, so that $\rho_{AB} = \mathrm{Tr}_C[\psi_{ABC}]$. Then $S_{AB} = S_C$, and so $S_A \leq S_B + S_C$ reads $S_A - S_B \leq S_{AB}$, proving the claim.

$|S(A|B)| \leq S_A$

By the definition of the conditional entropy, $|S(A|B)| = |S_{AB} - S_B|$. Again purifying with a system C, we thus claim that $|S_C - S_B| \leq S_A$. But this is again true by the above triangle inequality after re-labeling subregions.

$I(A:B|C) \leq 2 \log \min(d_A, d_B)$

The conditional mutual information of a tripartite system is

$$I(A:B|C) = S_{AC} + S_{BC} - S_C - S_{ABC}. \tag{882}$$

Thus from the above inequality on the conditional entropy,

$$\begin{aligned} I(A:B|C) &\leq |S_{AC} - S_C| + |S_{ABC} - S_{BC}| \\ &= |S(A|C)| + |S(A|BC)| \\ &\leq 2S_A \\ &\leq 2\log d_A. \end{aligned} \tag{883}$$

Due to the symmetry $I(A:B|C) = I(B:A|C)$, we can also exchange A and B in the above, giving

$$I(A:B|C) \leq 2\log d_B. \tag{884}$$

The claim thus follows.

Reminder about different types of covariant derivatives and some useful geometric identities

Today's entry is a reminder about various basic differential geometry facts. Of course this is all in math books somewhere, but the notation is never consistent and the results are scattered, and so having a reference is useful.

◇ ◇ ◇

We will be working in the general setting of a spacetime M and an associated vector bundle $E = M \times_\rho V$, where ρ is an action of the structure group $GL(V)$ (or a subgroup thereof). The fibers will usually just be copies of $\mathbb{R}^{\dim M}$. We will use the word "internal" to refer to stuff in V, and "external" to refer to stuff in M.

When doing geometry in noncoordinate bases, it's easy to get confused about the different types of covariant derivatives that get used. This is because there are two connections — the spin connection ω on the internal space and the Christoffel connection (CC) on the tangent space Γ. The spin connection is a 1-form, but transforms as a gauge field under internal rotations. The CC is similar, but transforms as a gauge field under external rotations.

We will define two types of covariant derivatives. The first one, which we will write as ∇, will act covariantly on *all* indices, both vector bundle (internal) and tangent space (external). For example, the defining relation between the ω and the Γ is that the vielbeins are covariantly constant under ∇:

$$\nabla_\mu e^a_\nu = \partial_\mu e^a_\nu + \omega_\mu{}^a{}_b e^b_\nu - \Gamma^\lambda_{\mu\nu} e^a_\lambda = 0. \tag{885}$$

Usually we will want to use this relation to fix Γ in terms of ω and ε, the latter two of which we treat as the fundamental geometric variables.

The second type of covariant derivative we will define is the covariant exterior derivative, which we will write as D (D_ω or d_ω would also be acceptable). This only acts covariantly on the internal indices, and also includes antisymmetrization, so that it is meant to act on forms on spacetime. Acting on a general form X, we write

$$DX = dX + \omega \wedge_{d\rho} X. \tag{886}$$

The notation $\wedge_{d\rho}$ means that ω and X are wedged together in their spacetime indices, with the $GL(V)$ structure of the result determined by the differential of the representation ρ (this second term is also often written $[\omega \wedge X]$). In all of our applications, X will either carry no internal indices, carry a single fundamental index, or be $\mathfrak{gl}(V)$-valued. In the first two cases, $X \wedge_{d\rho} Y = X \wedge Y$. When X is Lie algebra-valued however, some care is needed — we want D to be a derivation, and this means that the Lie bracket gets used: if X and Y are Lie-algebra valued p and q forms, then

$$X \wedge_{d\rho} Y = \frac{1}{p!q!}[X_{[\mathbf{u}}, Y_{\boldsymbol{\nu}]}]dx^{\boldsymbol{\mu}} \wedge dx^{\boldsymbol{\nu}}, \tag{887}$$

where $\boldsymbol{\mu}, \boldsymbol{\nu}$ are composite indices, and where the $[,]$ denotes the commutator. It is straightforward to check that this wedge product obeys the supercommutativity rule (still assuming that both X, Y are Lie algebra-valued)

$$X \wedge_{d\rho} Y = (-1)^{pq+1} Y \wedge_{d\rho} X, \tag{888}$$

which has an extra minus sign coming from the antisymmetry of the Lie bracket. Therefore for two Lie algebra-valued forms,

$$X \wedge_{d\rho} Y = X \wedge Y - (-1)^{pq} Y \wedge X. \tag{889}$$

Since the $\wedge_{d\rho}$ keeps the Lie algebra indices covariant, D obeys the Liebnitz rule wrt $\wedge_{d\rho}$:

$$D(X \wedge_{d\rho} Y) = DX \wedge_{d\rho} Y + (-1)^p X \wedge_{d\rho} DY. \tag{890}$$

For example, the vielbeins carry one fundamental index, and so

$$T^a \equiv De^a = de^a + \omega^a{}_b \wedge e^b. \tag{891}$$

The 2-form curvature R, defined by

$$R \equiv d\omega + \omega \wedge \omega, \tag{892}$$

caries a fundamental and an antifundamental index, and so (of course the following actually vanishes; more on this in a sec)

$$\begin{aligned} DR &= dR + \omega \wedge_{d\rho} R \\ &= dR + \omega \wedge R - R \wedge \omega. \end{aligned} \tag{893}$$

Remember that covariant derivatives map covariant objects to covariant ones (hence the name), but cannot really be applied to non-covariant objects, like the connection itself. Thus it is not in general correct to write $R = D\omega$.[80] Instead, we have

$$R = d\omega + \frac{1}{2}\omega \wedge_{d\rho} \omega, \tag{894}$$

which differs from $D\omega$ by the factor of 1/2. However, the variation of the curvature is indeed a total covariant exterior derivative, as

$$\begin{aligned} \delta R &= d\delta\omega + \delta\omega \wedge \omega + \omega \wedge \delta\omega \\ &= d\delta\omega + \omega \wedge_{d\rho} \delta\omega \\ &= D\delta\omega. \end{aligned} \tag{895}$$

It is straightforward that the square of D is given by wedging with the field strength R, in that

$$D^2 X = R \wedge_{d\rho} X, \tag{896}$$

so that R measures the obstruction to D forming a complex (R is a 2-form, so the degrees work out correctly). Indeed, if X carries zero or one internal index, so that $\omega \wedge_{d\rho} X = \omega \wedge X$, then

$$\begin{aligned} D^2 X &= D(dX + \omega \wedge X) \\ &= \omega \wedge dX + d\omega \wedge X - \omega \wedge dX + (\omega \wedge \omega) \wedge X \\ &= R \wedge X. \end{aligned} \tag{897}$$

On the other hand if X is a Lie algebra-valued q-form, then (note: I haven't seen this in anything I've read but I'm pretty sure it's correct)

$$\begin{aligned} D^2 X &= D(dX + \omega \wedge_{d\rho} X) \\ &= D(dX + \omega \wedge X - (-1)^q X \wedge \omega) \\ &= \omega \wedge dX + (-1)^q dX \wedge \omega + d\omega \wedge X - \omega \wedge dX + \omega \wedge \omega \wedge X + (-1)^q \omega \wedge X \wedge \omega \\ &\qquad - (-1)^q dX \wedge \omega - X \wedge d\omega - (-1)^q \omega \wedge X \wedge \omega - X \wedge \omega \wedge \omega \\ &= (d\omega + \omega \wedge \omega) \wedge X - X \wedge (d\omega + \omega \wedge \omega) \\ &= R \wedge X - X \wedge R \\ &= R \wedge_{d\rho} X. \end{aligned} \tag{898}$$

The Bianchi identities are

$$\begin{aligned} DR &= dR + \omega \wedge_{d\rho} R \\ &= dR + \omega \wedge R - R \wedge \omega \\ &= d(\omega \wedge \omega) + \omega \wedge d\omega - d\omega \wedge \omega \\ &= 0 \end{aligned} \tag{899}$$

and, using the fact about D^2,

$$\begin{aligned} DT &= D^2 e \\ &= R \wedge e. \end{aligned} \tag{900}$$

[80] This works only when the $\omega \wedge \omega$ term vanishes by virtue of the gauge group being Abelian, which for us only happens when the gauge group is $SO(2)$.

1-form propagators in 2+1d

Today's entry is mega-short: we will write down a general propagator for a vector field in 2+1d, to serve as a useful reference.

◇ ◇ ◇

Consider a vector field X with Lagrangian

$$\mathcal{L} = \frac{1}{2} X \wedge \star(\alpha \Pi_T + \beta \Pi_L + \gamma \star d) X, \tag{901}$$

with $\Pi_T = d^\dagger d/\Box, \Pi_L = dd^\dagger/\Box$. We invert this by writing the propagator as

$$D_X = A\Pi_T + B\Pi_L + C \star d. \tag{902}$$

We then use the orthogonality of $\Pi_{T/L}$ as well as (as usual, $\Box = d^\dagger d + dd^\dagger$ is negative-definite)

$$(\star d)^2 = -\Box \Pi_T, \qquad \star d \Pi_T = \Pi_T \star d = \star d, \tag{903}$$

where $\star d$ is viewed as a matrix with vector indices. The sign on this first equation is important, and follows from the fact that when acting on p-forms in D-dimensional Euclidean space, the adjoint of d is

$$d^\dagger = (-1)^{Dp+D+1} \star d \star . \tag{904}$$

For us $D = 3$ and $p = 1$, so that $d^\dagger = - \star d \star$ (alternatively one can just write out $\star d \star d$ explicitly).

This gives the conditions

$$\begin{aligned} \alpha A - \gamma C \Box &= 1 \\ \beta B &= 1 \\ \gamma A + \alpha C &= 0 \end{aligned} \tag{905}$$

so that

$$D_X = \frac{1}{\Box + \alpha^2/\gamma^2} \left(\frac{\alpha}{\gamma^2} \Pi_T - \frac{1}{\gamma} \star d \right) + \frac{1}{\beta} \Pi_L, \tag{906}$$

or in momentum space,

$$D_X^{\mu\nu} = \frac{1}{q^2 - \alpha^2/\gamma^2} \left[-\frac{\alpha}{\gamma^2} \left(\delta^{\mu\nu} - \frac{q^\mu q^\nu}{q^2} \right) + \frac{i}{\gamma} \varepsilon^{\mu\nu\lambda} q_\lambda \right] + \frac{1}{\beta} \frac{q^\mu q^\nu}{q^2}. \tag{907}$$

This is such that D_X is the inverse of the kernel in (901). If we just want to e.g. invert the kernel on coexact forms (viz. those with $\Pi_L X = 0$), we simply need drop the last $1/\beta$ term in the above expression.

As a check, note that this gives the correct topologically massive propagator when we take $\alpha = -\Box/e^2, \gamma = ik/2\pi, \beta = 0$.

Facts about types

The purpose of this diary entry is to introduce types and prove some basic facts about them.

◇ ◇ ◇

Definition

Given a d-letter alphabet $[d]$ and a squence $\mathbf{x} \in [d]^n$, let $N(a|\mathbf{x})$ be the number of occurances of a in $\mathbf{x}$. The *type* (aka emperical probability distribution) $P_{\mathbf{x}}$ of $\mathbf{x}$ is the distrubtion provided by sampling from the string $\mathbf{x}$. We let $\mathcal{P}_n$ denote the set of all possible types on $[d]^n$; equvalently $\mathcal{P}_n$ is the set of probability distributions on $[d]$ whose entries are valued in $\frac{1}{n}\mathbb{N}$. As an example, consider $n = 3$. Then $\mathcal{P}_3$ includes the distributions where $\mathbf{p} = (0,0,1), (1/3,1/3,1/3), (2/3,1/3,0)$, as well as all permutations thereof.

Let

$$\mathcal{T}_p^n \equiv \{\mathbf{x} : P_{\mathbf{x}} = p\} \tag{908}$$

be the set of all length-n strings whose type is the distribution p. Since the number of occurances of x_i in $\mathbf{x}$ is np_i,

$$|\mathcal{T}_p^n| = \binom{n}{np} \equiv \frac{n!}{\prod_{i=1}^d (np_i)!}. \tag{909}$$

Size of $\mathcal{P}_n$

A trivial upper bound for $\mathcal{P}_n$ can be obtained by noting that each symbol in $[d]$ can be assigned a probability in $(1/n)\mathbb{Z}_{n+1}$, which gives $n+1$ possible choices. Since there are only $d-1$ independent assignments of probabilities due to $\sum_i p_i = 1$, we thus have

$$|\mathcal{P}_n| \leq (n+1)^{d-1}. \tag{910}$$

Entropy

For $\mathbf{x} \in \mathcal{T}_p^n$, let

$$p(\mathbf{x}) \equiv \prod_{i=1}^n p(x_i). \tag{911}$$

Grouping the product by symbols, and using that the number of times $x \in [d]$ appears in $\mathbf{x}$ is equal to $np(x)$, we have

$$p(\mathbf{x}) = \prod_{x\in[d]} p(x)^{np(x)} = 2^{n\sum_{x\in[d]} p(x)\log[p(x)]} = 2^{-nH_2(p)}. \tag{912}$$

Probability of a type

For types $p, q \in \mathcal{P}_n$, define
$$p(\mathcal{T}_q^n) = \sum_{\mathbf{x} \in \mathcal{T}_q^n} p(\mathbf{x}). \tag{913}$$

Rearranging the product in $\mathbf{p}(\mathbf{x})$ in the same way as before, and using that the probability $p(\mathbf{x})$ is identical for all $\mathbf{x} \in \mathcal{T}_q^n$ by definition, we have

$$\begin{aligned} p(\mathcal{T}_q^n) &= |\mathcal{T}_q^n| 2^{n \sum_{x \in [d]} q(x) \log[p(x)]} \\ &= |\mathcal{T}_q^n| 2^{-n \sum_{x \in [d]} q(x)(\log[q(x)/p(x)] - \log[q(x)])} \\ &= |\mathcal{T}_q^n| 2^{-n(D(q||p) + H_2(q))}, \end{aligned} \tag{914}$$

where $D(q||p)$ is the relative entropy.

Bounds on $|\mathcal{T}_p^n|$

Gibbs inequality says that the relative entropy $D(q||p) \geq 0$, with equality iff $p = q$ as distributions. Thus for all $p, q \in \mathcal{P}_n$,

$$p(\mathcal{T}_q^n) \leq p(\mathcal{T}_p^n) = |\mathcal{T}_p^n| 2^{-nH_2(p)}. \tag{915}$$

On the other hand, since $\mathcal{T}_p^n$ is a subset of all length-n strings, its total probability is of course upper-bounded by 1. Thus we have

$$p(\mathcal{T}_q^n) \leq |\mathcal{T}_p^n| 2^{-nH_2(p)} \leq 1, \tag{916}$$

for all $p, q \in \mathcal{P}_n$. We can then sum over all $q \in \mathcal{P}_n$, and use the normalization condition $\sum_{q \in \mathcal{P}_n} p(\mathcal{T}_q^n) = 1$ to write

$$2^{H_2(p)} \leq |\mathcal{P}_n||\mathcal{T}_p^n| \leq |\mathcal{P}_n| 2^{nH_2(p)}. \tag{917}$$

Using the above bound on $|\mathcal{P}_n|$ then gives

$$\frac{2^{nH_2(p)}}{(n+1)^{d-1}} \leq |\mathcal{T}_p^n| \leq 2^{nH(p)}. \tag{918}$$

Bound on $p(\mathcal{T}_q^n)$

Pinsker's inquality states that

$$D(q||p) \geq \frac{1}{2\ln 2} ||p - q||_1^2. \tag{919}$$

Thus from (914),
$$p(\mathcal{T}_q^n) \leq |\mathcal{T}_q^n| 2^{-n(||p-q||_1^2/(2\ln 2) + H_2(q))}. \tag{920}$$

Using the right inequality from (918) to plug in for $|\mathcal{T}_q^n|$, we then get, after chaning bases,

$$p(\mathcal{T}_q^n) \leq e^{-\frac{n}{2}||p-q||_1^2}. \tag{921}$$

And an application

As a rather trivial application of the above inequality, let q be the distribution on $\{0,1\}$ with $q(1)=b$, and p be the distribution with $p(1)=a<b$. Intuitively, it should be obvious that the probability (using a coin whose outcomes are governed by p) of getting a sequence in $\mathcal{T}_q^n$ should be exponentially rare by an amount determined by $|b-a|$. Indeed, since in this case $||p-q||^2/2=|b-a|$, we have

$$p(\mathcal{T}_q^n) \leq e^{-n|b-a|}, \tag{922}$$

as expected.

Types and typical sets

We can use types to define a sharper version of typical sets, by allowing $\mathcal{T}_p^n$ to be 'fuzzy' in the space of probability distributions. Define

$$\mathcal{T}_{p,\delta}^n = \bigcup_{q:||p-q||_1 \leq \delta} \mathcal{T}_q^n. \tag{923}$$

For fixed $\delta > 0$, we claim that $p^n(\mathcal{T}_{p,\delta}^n)$ is exponentially close to 1 — i.e. essentially all length-n strings sampled according to p will have emperical distributions that agree with the underlying distribution. Indeed,

$$\begin{aligned} 1-p(\mathcal{T}_{p,\delta}^n) = p\left(\bigcup_{q:||p-q||_1>\delta} \mathcal{T}_q^n\right) &\leq e^{-n\delta^2/2}\left|\bigcup_{q:||p-q||_1>\delta} \mathcal{T}_q^n\right| \\ &\leq e^{-n\delta^2/2}\left|\bigcup_{q\in\mathcal{P}_n} \mathcal{T}_q^n\right| \\ &\leq e^{-n\delta^2/2}(n+1)^{d-1} \\ &= \exp\left(-n\left(\frac{\delta^2}{2}-\frac{d-1}{n}\ln(n+1)\right)\right), \end{aligned} \tag{924}$$

with the log correction in the exponent becoming unimportant at large n.

Trivialities on real representations of finite groups

Today's entry is a rather trivial comment about the counting of real representations of finite groups.

◇ ◇ ◇

We are all familiar with the completeness relation

$$\frac{1}{|G|}\sum_\alpha d_\alpha \mathrm{Tr}[R_\alpha(g)R_\alpha(h^{-1})] = \delta_{g,h}, \tag{925}$$

where the sum is over $\mathbb{C}$ irreps and $d_\alpha = \dim R_\alpha$. When $g = h = \mathbf{1}$, we get the usual sum of squares formula $|G| = \sum_\alpha d_\alpha^2$. This formula is only however true for representations over a vector space whose coefficients are valued in a splitting field for G, which for physics we always will take to be $\mathbb{C}$. In particular, it is *not* true if we require that the vector space V acted on by the $R_\alpha(g)$ is real. A simple example of this is given by the group $\mathbb{Z}_N$. This of course has N 1d irreps over $\mathbb{C}$. Over a real vector space however, it has only one 1d irrep if N is odd, and two 1d irreps if N is even. Clearly then we cannot simply take d_α to be the dimension of the corresponding representation in the real case.

The counting over a real vector space V is done by lifting to $V \otimes \mathbb{C}$, and doing the counting there. Doing this makes use of the fact that every irrep on V lifts to either a) a single irrep on $V \otimes \mathbb{C}$ or b) a direct sum of a complex irrep and its conjugate. We may then write the completeness relation in $V \otimes \mathbb{C}$ as

$$\frac{1}{|G|}\left(\sum_{\alpha\in\mathrm{Rep}_\mathbb{R}} d_\alpha \mathrm{Tr}[R_\alpha(g)R_\alpha(h^{-1})] + \sum_{\alpha\in\mathrm{Rep}_\mathbb{C}} d_\alpha \mathrm{Tr}[(R_\alpha \oplus R^*_\alpha)(g)(R_\alpha \oplus R^*_\alpha(h^{-1}))]\right) = \delta_{g,h}, \tag{926}$$

where the nonstandard notation $\mathrm{Rep}_\mathbb{R}$ means all of the real irreps of G, and $\mathrm{Rep}_\mathbb{C}$ means half of the non-real irreps, with the elements of $\mathrm{Rep}_\mathbb{C}$ and their complex conjugates enumerating all of the non-real irreps. Going back to V, we see that we may write

$$\frac{1}{|G|}\sum_\alpha \frac{d_\alpha}{\mathsf{CIR}(\alpha)}\mathrm{Tr}[R_\alpha(g)R_\alpha(h^{-1})] = \delta_{g,h}, \tag{927}$$

where now the sum runs over only the (real) irreps acting on V, d_α is the dimension of α as a representation acting on V, and the function ('CIR' for 'complex irreducible')

$$\mathsf{CIR}(\alpha) = \begin{cases} 1 & \text{if } \alpha \text{ is irreducible in } V\otimes\mathbb{C} \\ 2 & \text{if } \alpha \text{ splits as } R_\alpha \oplus R^*_\alpha \text{ in } V\otimes\mathbb{C} \end{cases} \tag{928}$$

In practice, $\mathsf{CIR}(\alpha)$ is computed from the characters using the FS indicator

$$\nu(\alpha) \equiv \frac{1}{|G|}\sum_g \chi_\alpha(g^2). \tag{929}$$

α is either $1, 0$ or -1, depending on whether α is real, complex, or quaternionic, respectively (so that e.g. $\mathsf{CIR}(\alpha) = 1$ if $\nu(\alpha) = 1$). The proof of this property of $\nu(\alpha)$ is easy to find online.

Lets look at a simple example. Consider the group $\mathbb{Z}_3$. Over $\mathbb{R}^2$ this group has two irreps, the trivial one A_1 and a single 2d irrep E. In order for the counting to work out, we had better have $\mathrm{CIR}(E) = 2$, and of course we do: in $\mathbb{C}^2$, E acts as $e^{i2\pi n/3} \oplus e^{-i2\pi n/3}$.

Consider instead D_3. This group has order 6, and has three irreps over $\mathbb{R}^2$: two 1d irreps A_1, A_2 (with A_2 the identity on the $\mathbb{Z}_3$ subgroup), and one 2d irrep E. Since $1^2+1^2+2^2 = 6$, evidently in this case we must now have $\mathsf{CIR(E)} = 1$, i.e. E must now be *irreducible* over $\mathbb{C}^2$. And indeed this is the case: while $\mathbb{Z}_3 \supset D_3$ is represented reducibly as before, the reflection is represented as σ^x, and therefore mixes the two 1d factors, rendering E irreducible over $\mathbb{C}^2$. This can of course also be checked by explicitly computing $\nu(E)$.

Integral representation of conjugations of derivations by matrix exponentials

Today we are proving a trivial math result which I noticed myself looking up frequently. We will show that if $\mathcal{D}$ is any derivation (e.g. derivative or commutator), then

$$\mathcal{D}e^X = \int_0^1 dt\, e^{tX}(\mathcal{D}X)e^{(1-t)X}. \tag{930}$$

◇ ◇ ◇

From the form of the answer, we can guess that the right strategy is to break up the exponentials as products. Indeed, we simply write

$$\begin{aligned} \mathcal{D}e^X &= \lim_{N\to\infty} \mathcal{D}\left(1+\frac{X}{N}\right)^N \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N}\left(1+\frac{X}{N}\right)^k (\mathcal{D}X)\left(1+\frac{X}{N}\right)^{N-k}. \end{aligned} \tag{931}$$

We then define $t \equiv k/N$, so that

$$\begin{aligned} \mathcal{D}e^X &= \lim_{N\to\infty} \sum_{t=0}^{1}(1/N)\left(1+\frac{tX}{N}\right)^N (\mathcal{D}X)\left(1+\frac{(1-t)X}{N}\right)^N \\ &\to \int_0^1 dt\, e^{tX}(\mathcal{D}X)e^{(1-t)X} \end{aligned} \tag{932}$$

as claimed. That's all!

Markov chain trivialities

Today's entry is proving a short rather trivial result meant to remind myself of some definitions conerning Markov chains.

◇ ◇ ◇

Proposition 18. *Define a random walk $\mathcal{R}$ on a graph G by the Markov chain with transition elements*

$$\mathcal{R}_{x\to y} = \delta_{y\in\partial x}\frac{1}{\deg(x)}, \tag{933}$$

where x, y are vertices of G. Then $\mathcal{R}$ is mixing if G is connected and non-bipartite. If G is bipartite, then $\mathcal{R}$ has a unique eigenvector with eigenvalue -1. The equilibrium distribution is

$$p_{eq}(x) = \frac{\deg(x)}{2n_E} \tag{934}$$

where n_E is the number of edges. Finally, define the lazy random walk $\mathcal{L}$ through the transition elements

$$\mathcal{L}_{x\to y} = \frac{1}{2}\delta_{x,y} + \frac{1}{2}\delta_{y\in\partial x}\frac{1}{\deg(x)}. \tag{935}$$

$\mathcal{L}$ is ergodic if G is connected, and has the same equilirbrium distribution as $\mathcal{R}$.

Proof. First let us review some definitions, since there seems to be some ambiguity in the literature depending on where one looks. We will be interested in a dynamical system $(X, \mathcal{A}, \mu, M)$, where X is some space, μ a measure on X, $\mathcal{A}$ the collection of all (measurable, i.e. $\{A \subset X \,:\, \mu(A) \neq 0\}$) subsets of X, and M some Markov process acting on $\mathcal{A}$. Then M is said to be ergodic if any subset $A \in \mathcal{A}$ eventuallly visits all of X under the evolution of M:

$$M \text{ ergodic} \implies \bigcup_{t=1}^{\infty} M^t(A) = X,\ \forall\, A \subset \mathcal{A}. \tag{936}$$

Mixing is a stronger property: M is said to be mixing if (roughly speaking) one converges to an equilibrium distribution regardless of one's starting point:

$$M \text{ mixing} \implies \forall\, A, B \subset \mathcal{A},\ \exists\, t \,:\, M^{t'}(A) \cap B \neq \emptyset\ \ \forall t' > t. \tag{937}$$

Thus for us, ergodicity of a Markov chain is equivalent to irreducibility (for every two states x, y, there exists a transition from $x \to y$ with nonzero probability), while the mixing property holds only if the chain is both irreducible and aperiodic (for all starting points x, there exists a time t such the probability of returning to x is nonzero for all times $t' \geq t$).

The irredicibility of $\mathcal{R}$ follows directly from the fact that G is connected and that $\mathcal{R}_{x\to y} \neq 0$ as long as x, y are neighbors. Now for aperiodicity. $\mathcal{R}$ is *not* aperiodic if G is bipartite, since the probability of being on a given sublattice oscillates from 1 to 0 for all times t; thus the return probability is necessarily zero for $t \in 2\mathbb{N}+1$.[81] Consider then a non-bipartite G. Such a G necessarily has closed loops, since trees are bipartite. If all closed loops have even length, then G is bipartite since we can 2-color G by alternating colors along each loop.Therefore there must exist at least one loop L of odd length. For any vertex $x \in G$, the connectivity of G implies that there exists a loop L_x of odd length which starts and ends at x and traverses L once. The probability of traversing L_x is nonzero, and hence the return probability at x is nonzero for all $t' \in 2\mathbb{N}+1$ as long as $t' \geq |L_x|$. Since we always have a length-2 path which leaves x and comes right back on the next time step, the return probability is thus nonzero for all $t' > t$, and $\mathcal{R}$ is aperiodic.

Since $\mathcal{R}$ is mixing, it has a unique equilibirum distribution with eigenvalue 1. That the proposed p_{eq} works is easy to check:

$$\sum_y \mathcal{R}_{x\to y} p_{eq}(y) = \sum_{y\in\partial x} \frac{1}{2n_E} = \frac{\deg(x)}{2n_E}. \tag{938}$$

Now consider the lazy walk $\mathcal{L}$. $\mathcal{L}$ is obviously irreducible, and is aperiodic regardless of whether or not G is bipartite. Indeed, in the bipartite case the nonzero probability of staying

[81] The eigenvector of $\mathcal{R}$ with eigenvalue -1 is $\sum_x (-1)^x |x\rangle$, where $(-1)^x$ alternates signs on the two sublattices.

at a given vertex $\mathcal{L}_{x\to x} \neq 0$ means that there is always a nonzero probability of returning to a given starting point. Note that since $\lambda \in \mathrm{Spec}(\mathcal{R}) \implies |\lambda| \leq 1$, the eigenvalues of $\mathcal{L}$ (viz. $(1+\lambda)/2$) are strictly positive, and linearly related to those of $\mathcal{R}$. Thus $p_{eq}(x)$ is the same for both $\mathcal{L}$ and $\mathcal{R}$. □

A party trick

Today's entry is a fun party trick that I learned from Shankar Balasubramanian. It goes as follows: you pick an arbitrary polynomial $f(x)$ with non-negative integer-valued coefficients, and then by asking for $f(x)$ evaluated on just two points x_1, x_2, I tell you your polynomial.

◇ ◇ ◇

This trick is rather remarkable since I can make an infinite number of choices for both of the points I ask you to evaluate $f(x)$ on. I first ask you for $f(x_1)$, where x_1 is any positive integer. Given $f(x_1)$, I then choose any $x_2 > f(x_1)$, and ask you for $f(x_2)$. If $f(x) = \sum_j c_j x^j$, then all of the c_j must be less than x_2. This means that $\sum_j c_j x_2^j$ is simply $f(x_2)$ written in base x_2! Therefore all I need to do is to write down $f(x_2)$ in base x_2 — the digits of the resulting expression are then the coefficients c_j of $f(x)$. When doing this in one's head it is thus often helpful to choose x_1 to be small (but to not always take $x_1 = 1$, in order to throw people off), and x_2 to be the smallest power of 10 greater than $f(x_1)$.

Representation degeneracies and blocking

Today's entry is a rather parochial problem I needed for research: we will show that for any system with a local Hilbert space acted on by a representation of a finite Abelian group G, one may always block local $\otimes$ factors together to form composite sites, on which G is represented by

$$R_g = \bigoplus_{h\in G} \chi_h(g)^{\oplus N_h}, \tag{939}$$

where the degeneracies $N_h \geq 1$ for all $h \in G$, and the $\chi_h \in G^*$ are linear characters for the irreps.

◇ ◇ ◇

Microscopically, G acts on each site as

$$r_g = \bigoplus_{\alpha\in G} \chi_\alpha(g), \tag{940}$$

where the α run over some subset of G. Since in quantum mechanics we mod out by overall phases, any element g for which $r_g \propto \mathbf{1}$ is represented trivially. We will not allow this to occur, since in this case G is not represented faithfully. Therefore we require that the $\{\alpha\}$ be such that

$$\chi_\alpha(g) = 1 \; \forall \; \alpha \implies g = e. \tag{941}$$

We will call a representation obeying the above equation a 'legit' representation.

Note that we are only ever interested in reducible representations, since irreps act as phases and are hence trivial. As an example, for $G = \mathbb{Z}_N$ a legit representation is

$$r_g = 1 \oplus \zeta_N^g. \tag{942}$$

For $\mathbb{Z}_N^2$, one requires at least three summands, with a legit representation being e.g.

$$r_{\mathbf{g}} = 1 \oplus \zeta_N^{g_1} \oplus \zeta_N^{g_2}. \tag{943}$$

Note how we are always taking the first summand to be the trivial representation; this is done wolog because we can always redefine r_g by an overall g-dependent phase. This is what allows us to use 1 on the RHS of (941).

Proposition 19. *The $\{\alpha\}$ must consist of the identity e together with a set of elements which generate G.*

This is proved with the following lemma:

Lemma 2. For a subgroup $H \subset G$, every irrep of H can be extended to an irrep of G in $|G : H|$ different ways. In particular, there are $|G : H|$ irreps of G whose characters are trivial when restricted to H.

Proof. This is proved using Frobenius reciprocity, which establishes a duality between induced and restricted representations. Let ψ be an irreducible character of H, and χ an irreducible character of G. Let ψ^G be the induced (redicuble) character of ψ in G, and let $\chi|_H$ be the restriction of χ to H. Then Frobenius reciprocity says

$$\langle \psi, \chi|_H \rangle_H = \langle \psi^G, \chi \rangle_G, \tag{944}$$

where $\langle a, b\rangle_K = \frac{1}{|K|}\sum_{g\in K} a^*(g)b(g)$ is the usual inner product. Now $\chi|_H$ is an irredicible character on H, and hence

$$\langle \psi^G, \chi\rangle_G = \delta_{\psi,\chi|_H}. \tag{945}$$

Therefore, the (redicuble) character χ_H^G contains χ as a direct summand iff $\psi = \chi|_H$. We can thus count the number n_ψ of irreducible G-characters χ_g that restrict to ψ as

$$n_\psi = \sum_g \delta_{\psi,\chi_g|_H} = \sum_g \langle \psi^G, \chi_g\rangle_G = \psi^G(e), \tag{946}$$

since $\sum_g \chi_g(h) = |G|\delta_{h,e}$. But $\psi^G(e) = |G|/|H|$ (see e.g. Wikipedia), and hence

$$n_\psi = |G : H| \tag{947}$$

independent of ψ, as claimed. $\square$

We can now prove the proposition. We show that if the $\{\alpha\}$ do not generate G, then they do not form a legit representation, i.e. there is some $e \neq g_* \in G$ for which $r_g = \mathbf{1}$. Indeed, if the $\{\alpha\}$ do not generate G, then they generate some proper subgroup H of G. By the above lemma, we can find $|G|/|H|$ different irreps of G whose characters are trivial on H. Since each irrep of G is identified with a group element, there exist $|G|/|H|$ choices of $g_* \in G$ such that $\chi_\alpha(g_*) = 1$ for all α. Therefore r_g does not generate a legit representation, as claimed.

Finally, suppose r_g is a legit representation. Taking n tensor powers of r_g gives

$$r_g^{\otimes n} = \bigoplus_{\{\alpha_1,\dots,\alpha_n\}} \chi_{\alpha_1+\cdots+\alpha_n}(g). \tag{948}$$

Since the $\{\alpha\}$ must generate G, for large enough n (upper-bounded by $|G|$), $r_g^{\otimes n}$ will contain $\chi_h(g)$ for all $h \in G$. Thus as long as n is large enough all of the degeneracies N_h are positive, as we wanted to show.

Compression with side information

Today's entry is devoted to proving a few facts about data compression using side information.

◇ ◇ ◇

Conditionally typical set

For a distribution p_{XY}, let $J^n_{p,\delta}$ be the jointly typical set,

$$J^n_{p,\delta} \equiv T^n_{p,\delta} \cap (T^n_{p_X,\delta} \times T^n_{p_Y,\delta}). \tag{949}$$

Given y^n, define the conditionally typical set $J(y^n) \equiv J^n_{p,\delta}(y^n)$ by

$$J(y^n) = \{x^n \in X^n \,:\, (x^n, y^n) \in J^n_{p,\delta}\}. \tag{950}$$

If $y^n \in T^n_{p_Y,\delta}$, we can bound $p(x|y)$ (omitting ns for simplicity) as

$$\frac{p(xy)_{min}}{p(y)_{max}} \leq p(x|y) \leq \frac{p(xy)_{max}}{p(y)_{min}}. \tag{951}$$

We can simplify this by using that the length-n elements in the typical set of tolerance δ have probabilities bounded as

$$2^{-n(H(X)+\delta)} \leq p(x) \leq 2^{-n(H(X)-\delta)}, \qquad x \in T^n_{p,\delta}. \tag{952}$$

This allows us to write e.g. ($\exp(x) = 2^x$)

$$\frac{p(xy)_{max}}{p(y)_{min}} \leq \exp\left(-n(H(XY) - H(Y) - 2\delta)\right). \tag{953}$$

Thus for all $x \in J(y)$,

$$\exp\left(-n(H(X|Y)+2\delta)\right) \leq p(x|y) \leq \exp\left(-n(H(X|Y)-2\delta)\right). \tag{954}$$

We then use the rather loose bound

$$|J(y)| \min_{x \in J(y)} p(x|y) \leq 1 \tag{955}$$

to write (assuming the minimum in (954) can always be acheived, which is true for typical sets)

$$|J(y)| \leq \exp\left(n(H(X|Y)+2\delta)\right). \tag{956}$$

compression using shared information

Let $(X,Y) \sim p_{XY}$ be length-n tuples of random variables drawn from p_{XY}. Suppose A knows both (X,Y), while B only holds Y. Suppose further that A wants to transmit X to B using the fewest bits possible. Of course, A could just use a compression scheme for the typical set of X, but in the present case this is suboptimal. Since both parties know Y, then A can account for the correlations between X and Y by encoding according to the conditionally joint typical set: for a given y, each string $x \in J(y)$ is encdoded with a unique key, and all other strings are not transmitted. By (956), transmitting all of the strings in $J(y)$ only requires

$$n(H(X|Y)+\delta) \tag{957}$$

bits (where the δ here is 2δ above, so that $J(y)$ is defined by $(x,y) \in J^n_{p,\delta/2}$). The failure probability is determined by the total weight $p(J)$ of the conditionally typical set. To upper-bound the failure probability, we need to lower-bound $p(J)$. This can be done by using our earlier result (924) for $p(J)$, which followed from Pinsker's inequality:

$$1 - p(J) \leq \exp\left(-n(\delta^2 + o(1))\right). \tag{958}$$

Thus the error ε can be made arbitrarily small as $n \to \infty$, as long as δ is finite.

compression with limited access

Now suppose that A only has access to X (and B still only has access to Y). This is naively much more challenging that the situation considered above; we will however see that in fact it is just as easy. A performs her encoding using a random codebook $E : X^n \to [2^{nR}] \equiv \{1, \dots, 2^{nR}\}$, where R is a rate to be determined. We take E to be uniformly random, and after a particular choice of E is made, the codebook is fixed and given to both A and B. The preimages $E^{-1}(m)$ messages $m \in [2^{nR}]$ are random subsets of X^n with typical size

$$\mathbb{E}_{m \sim U}[|E^{-1}(m)|] = 2^{n(1-R)}. \tag{959}$$

The compression protocol is trivial if $R = 1$, for then A simply directly sends the raw string x to B. A and B's goal is to minimize the value of R for which decoding can be done in an error-free way.

Given a message m and a realization y of Y (with x assumed to be in $J(y)$), B decodes by finding the x such that $E(x) = m$ and $(x, y) \in E^{-1}(m) \cap J(y)$. One way this can fail is if $|E^{-1}(m) \cap J(y)| > 1$: in this case B doesn't know which string to choose. Define the random variable $n(m, y)$ (with $J(y)$ assumed non-empty) by $n(m, y) = |E^{-1}(m) \cap J(y)| - 1$. The expectation value of $n(m, y)$ over messages m is equal to the fraction of the full space of strings occupied by $E^{-1}(m)$ for a typical m (viz. $2^{n(R-1)}$) times the density of points in $J(y)$, which from the above is upper-bounded by $2^{n(H(X|Y)+2\delta-1)}$. Thus

$$\mathbb{E}_{m\sim U}[n] \leq 2^{n(1-R)}2^{n(H(X|Y)+2\delta-1)} = 2^{-n(R-H(X|Y)-2\delta)}, \tag{960}$$

meaning that as long as

$$R > H(X|Y) + 2\delta, \tag{961}$$

this type of error will not occur.

other errors

The other type of error that can occur when decoding is if $E^{-1}(m) \cap J(y)$ is empty. This can happen if E happens to be chosen such that $m \notin \mathrm{im}(E)$. However, since E is a random function with $[\dim \mathrm{preim}(E)/\dim \mathrm{im}(E)] = 2^{n(1-R)}$, the chance of this happening is exponentially small in n as long as $R < 1$ (as can be proved using e.g. the Chernoff bound). Similarly, this type of error could happen if (x, y) are not jointly typical. But again, the probability of this happening is exponentially small in n as long as $H(X|Y)$ is nonzero. Thus as $n \to \infty$, the only requirement on the rate R comes from (961).

Remainders in Taylor's theorem

In this throwback to our high school days we will prove an integral expression for (and then a convenient bound on) the difference between a differentible function $f(x)$ and its pth order Taylor series approximation $f_{p,a}(x)$ near a point a, viz.

$$f_{p,a}(x) \equiv \sum_{k=0}^{p} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{962}$$

A bound on $|f(x) - f_{p,a}(x)|$ came up in discussions one day and I realized I was not immediately sure how to prove it; what follows is one possible approach.

First let us define the difference we are trying to get an explicit expression for:

$$D_{p,a} \equiv f(x) - f_{p,a}(x). \tag{963}$$

We claim

Proposition 20. *The remainder is given by*

$$D_{p,a} = \int_a^x dy\, \frac{f^{(p+1)}(y)}{p!}(x-y)^p. \tag{964}$$

Proof. We will prove this by induction. First note that the zeroth order difference can indeed be written in this form:

$$D_{0,a} = f(x) - f(a) = \int_a^x dy\, \frac{f^{(1)}(y)}{0!}(x-y)^0. \tag{965}$$

Now we show that if $D_{p,a}$ takes this form, so does $D_{p+1,a}$. Note that

$$D_{p+1,a} - D_{p,a} = f_{p,a}(x) - f_{p+1,a}(x) = -\frac{f^{(p+1)}(a)}{(p+1)!}(x-a)^{p+1}. \tag{966}$$

Rearranging,

$$\begin{aligned} D_{p+1,a} &= \int_a^x dy\, \frac{f^{(p+1)}(y)}{p!}(x-y)^p - \frac{f^{(p+1)}(a)}{(p+1)!}(x-a)^{p+1} \\ &= \frac{f^{(p+1)}(a)}{(p+1)!}(x-y)^{p+1}\Big|_{y=a}^{y=x} - \int_a^x dy\, \frac{f^{(p+1)}(y)}{(p+1)!}\frac{\partial}{\partial y}(x-y)^{p+1} \\ &= \int_a^x dy\, \frac{f^{(p+2)}(y)}{(p+1)!}(x-y)^{p+1}, \end{aligned} \tag{967}$$

as claimed.

□

Note that if $g(y)$ is continuous on $[a,x]$ then there exists some $z_{a,x} \in [a,x]$ such that $\int_a^x dy\, g(y) = (a-x)g(z_{a,x})$. Thus we know that

$$D_{p,a} = \frac{f^{(p+1)}(z_{a,x})}{p!}(x - z_{a,x})^p(x-a). \tag{968}$$

This is however rather non-constructive. If we just want a bound on $|D_{p,a}|$, a convenient one is as follows:

$$\begin{aligned} |D_{p,a}| &\le \int_a^x dy\, \left| \sum_{k=0}^{\infty} \frac{f^{(p+k+1)}(x)}{p!k!}(x-y)^{p+k}(-1)^k \right| \\ &\le \sum_{k=0}^{\infty} \frac{|f^{(p+k+1)}(x)|}{p!k!(p+k+1)}|x-a|^{p+k+1} \\ &\le \frac{|f^{(p+1)}(x)|}{(p+1)!}|x-a|^{p+1} \end{aligned} \tag{969}$$

where in the last line we just kept the $k=0$ term. Note that $f^{(p+1)}$ is *not* evaluated at a!

Random facts about random walks on trees

In this entry we will derive some basic results about simple random walks[82] on z-regular trees. We will mostly be interested in determining the asymptotic scaling of return probabilities.

walks on $\mathbb{Z}$

We first consider the simplest case $z = 2$, which of course we expect to behave very differently from $z > 2$. In this case, the walks are simple enough that one does not have to resort to the use of generating functions. Let us examine the number $N(n; L)$ of length-L random walks which start at the origin and end at n. Laziness does not affect any of the asymptotics of the walks, and so for concreteness we will work in the non-lazy case and let both n, L have the same parity. In the large L limit, and defining $\mathsf{n} \equiv n/L$, we have

$$N(n; L) = \binom{L}{(n+L)/2} \approx \sqrt{\frac{2}{\pi L(1-\mathsf{n}^2)}} 2^L \exp\left(\frac{L}{2}\left[(1-\mathsf{n})\ln(1+\mathsf{n}) + (1+\mathsf{n})\ln(1-\mathsf{n})\right]\right) \tag{970}$$

where we used

$$k! \approx \sqrt{2\pi n}(k/e)^k. \tag{971}$$

When $\mathsf{n} = 0$ we get the expected $1/\sqrt{L}$ return probability. We also see that all walks where n is order L^0 (viz. the distance of the walk endpoint is linear in the walk length L) are exponentially rare. When n is order L^β for $0 \leq \beta < 1$, we have $\mathsf{n} \ll 1$ and get

$$N(n; L) \approx \sqrt{\frac{2}{\pi L}} 2^L e^{-L\mathsf{n}^2}. \tag{972}$$

Thus indivudal walks with $\mathsf{n} = O(\sqrt{L})$ (in the CS sense of the notation, e.g. $L^{1/4}$ is $O(\sqrt{L})$) have a $1/\sqrt{L}$ probability of occuring.

One implication of this is that the midpoint of a returning random walk of length L is located a distance from the origin $\alpha\sqrt{L} < n < \beta\sqrt{L}$ for $\alpha, \beta > 0$ and $\alpha, \beta = \Theta(L^0)$ with constant probability:

$$\frac{\sum_{n=\alpha\sqrt{L}}^{\beta\sqrt{L}} N(n; L)^2}{N(0; 2L)} = \Theta(1). \tag{973}$$

This also implies that the difference between the maximum extent of a walk and its mid point distance scales as $\sqrt{L}$ with probability approaching 1 as $L \to \infty$.

[82] viz. walks where all transition probabilities are equal and transitions occur only between neighboring sites.

generating functions for returning walks

The generating function $R(x)$ for returning non-lazy simple RWs on the z-regular tree T_z is easily seen to be

$$R(x) = 1 + zx^2 R(x)B(x), \tag{974}$$

where $B(x)$ is the generating function for returning walks on a version of T_z whose root vertex has degree of only $z-1$. We thus need a recursion relation for $B(x)$, which is

$$B(x) = 1 + x^2(z-1)B(x)^2 \implies B(x) = \frac{1-\sqrt{1-4x^2(z-1)}}{2x^2(z-1)}. \tag{975}$$

Note that $B(x)$ is a function only of x^2, as it should be for a non-lazy walk on account of the T_z's bipartiteness. The first few terms are

$$B(x) = 1 + (z-1)x^2 + 2(z-1)^2x^4 + 5(z-1)^3x^6 + 14(z-1)^4x^8 + \cdots \tag{976}$$

which can be explicitly checked for small values of z with sufficient patience.

We can now use this expression to get $R(x)$, which is

$$\begin{aligned} R(x) &= \frac{1}{1-zx^2B} \\ &= \frac{1}{1-\alpha(1-\sqrt{1-4x^2(z-1)})}, \qquad \alpha \equiv \frac{z}{2(z-1)}, \end{aligned} \tag{977}$$

the first few terms of which are

$$R(x) = 1 + zx^2 + (z^2 + z(z-1))x^4 + (z^3 + 2z^2(z-1) + 2z(z-1)^2)x^6 + \cdots \tag{978}$$

which again may be verified with sufficient patience.

An aside: one should not be alarmed that $R(1)$ is imaginary. If one wants the expected number of times an infinitely long walk returns to the origin, one needs to write down generating functions for probabilities, rather than for number of paths. Since each individual move has an equal probability of $1/z$, this amounts to sending $x \mapsto x/z$, which gives

$$R(x/z) = \frac{2(z-1)}{z-2+z\sqrt{1-(x\rho)^2}}, \tag{979}$$

where

$$\rho \equiv \frac{2\sqrt{z-1}}{z} \tag{980}$$

is the spectral radius of T_z. Sending $x \to 1$ then

$$\langle \text{number of returns} \rangle = \frac{z-1}{z-2}. \tag{981}$$

This appropriately diverges when $z = 2$ but is finite for all $z > 2$, and correctly approaches 1 as $z \to \infty$ (in this limit the random walk only visits the origin at $t = 0$).

Laziness is easy to add. For simplicity, we consider the most natural case where the probability of doing no move is equal to the probability of doing any one particular move. To streamline with our previous notation, define l such that $l = 1$ for this type of lazy walk, while $l = 0$ for a non-lazy walk. Then the generating function is modified by adding a single term as

$$R(x) = 1 + lxR(x) + zx^2R(x)B(x), \tag{982}$$

where $B(x)$ now obeys

$$B(x) = 1 + lxB(x) + x^2(z-1)B(x)^2 \implies B(x) = \frac{1 - lx - \sqrt{1 - lx - 4x^2(z-1)}}{2x^2(z-1)}. \tag{983}$$

return probability

We now want to determine the long-walk asymptotics, in particular the probability for a random walk on the tree to return to the origin (which we expect to go as $1/\sqrt{L}$ if $z = 2$ and to decay exponentiallly fast if $z > 2$, as for $z > 2$ the walk should move away from the origin with constant velocity). To do this we beautify $R(x)$ as

$$\begin{aligned} R(x) &= \frac{1}{1 - lx - zx^2B(x)} \\ &= \frac{(1-\alpha)(1-lx) - \alpha\sqrt{1 - lx - 4x^2(z-1)}}{(1-\alpha)^2(1-lx)^2 - \alpha^2(1 - lx - 4x^2(z-1))}, \end{aligned} \tag{984}$$

which lets us write the series expansion for $R(x)$ as a convolution between a geometric series and the one coming from the expansion of the square root in the numerator.

Because we don't expect the asymptotics of the return probability to be affected by laziness, to simplify things we will take $l = 0$ for now. As a further simplification we will first look at what happens for $z = 2$, which as just mentioned we expect to behave differently than $z > 2$. In this case $\alpha = 1$, so that we in fact have the extremely simple

$$\begin{aligned} R^{(z=2)}(x) &= (1 - 4x^2)^{-1/2} \\ &= \sum_{k=0}^{\infty} \frac{2^{2k}(2k-1)!!}{2^k k!}(2x)^{2k}, \end{aligned} \tag{985}$$

where for the first term we let $(-1)!! \equiv 1$. We now need[83]

$$k!! \approx \sqrt{2k}(k/e)^{k/2} \tag{986}$$

Therefore we can confirm our earlier result that the number $N_L^{(2)}$ of L-length walks on the 2-tree that return to the origin is

$$\begin{aligned} N_L^{(2)} &\approx \delta_{[L]_2,0} 2^{L/2} \frac{\sqrt{2(L-1)}((L-1)/e)^{(L-1)/2}}{\sqrt{\pi L}(L/(2e))^{L/2}} \\ &= \delta_{[L]_2,0}\sqrt{\frac{2e}{\pi}} L^{-1/2} 2^L + \cdots, \end{aligned} \tag{987}$$

[83] Strictly speaking the above is for odd k, for even k the $\sqrt{2}$ is changed to $\sqrt{\pi}$.

where the $\cdots$ are subleading. Since the total number of length-L random walks is 2^L, the return probability $p_{ret}^{(2)}(L)$ for the 2-tree is

$$p_{ret}^{(2)}(L) \approx \delta_{[L]_2,0}\sqrt{\frac{2e}{\pi}}L^{-1/2}, \tag{988}$$

giving the expected scaling.

We will now see how $p_{ret}^{(z)}(L)$ gets exponentially damped upon making $z > 2$. Continuing to let $l = 0$, one can get a hint of this damping from the fact that the coefficient of x^2 in the square root appearing in $R(x)$ is $4(z-1)$, which is less than z^2 for all $z > 2$; this means that the appropriate coefficient of $R(x)$ will scale with an exponential of base less than z, so that $N^{(z>2)}(L)/z^L \to 0$ exponentially fast. To show this more explicitly, we start by writing

$$R(x) = \frac{1 + \frac{z}{2}(\sqrt{1-(\gamma x)^2} - 1)}{1-(zx)^2}, \qquad \gamma \equiv 2\sqrt{z-1}. \tag{989}$$

Then

$$R(x) = \sum_{m=0}^{\infty}(zx)^{2m}\left(1 - \frac{z}{2}\sum_{n=1}^{\infty}\frac{(2n-3)!!}{2^n n!}(\gamma x)^{2n}\right) \tag{990}$$

and so the $2k$th coefficient is

$$R_{2k}^{(z)} = z^{2k} - \sum_{n=1}^{k}\frac{z^{2k-2n+1}\gamma^{2n}(2n-3)!!}{2^{n+1}n!}. \tag{991}$$

Since we know that $R_\infty^{(z>2)}/z^{2k} = 0$, we know that

$$\sum_{n=1}^{\infty}\frac{z^{1-2n}\gamma^{2n}(2n-3)!!}{2^{n+1}n!} = 1. \tag{992}$$

Thus for large k we may write

$$\begin{aligned}\frac{R_{2k}^{(z>2)}}{z^{2k}} &= \sum_{n=k+1}^{\infty}\frac{z^{1-2n}\gamma^{2n}(2n-3)!!}{2^{n+1}n!} \\ &\approx \sum_{n=k+1}^{\infty}\sqrt{\frac{e^3z^2(1-3/2n)}{16\pi}}(n-3/2)^{-3/2}\exp\left(n[-2\ln(z/\gamma)+\ln(1-3/2n)\right) \\ &\approx \int_k^\infty dn\,\sqrt{\frac{z^2e^3}{16\pi}}n^{-3/2}\exp(-2n\ln(z/\gamma)).\end{aligned} \tag{993}$$

In writing down the final expression we have both approximated the sum as an integeral and dropped terms subleading in n — both are valid as $k\to\infty$, although in practice the subleading terms are numerically seen to actually have fairly large effects at relatively large values of $k \lesssim 500$.

The important part for us is that for long walks, the return probability indeed decreases exponentially if $z > 2$, as doing the integral (whose exact expression is written in terms of $\Gamma(1/2, L\ln\gamma)$) gives

$$p_{ret}^{(z>2)}(L) \propto \delta_{[L]_2,0}L^{-3/2}\rho^L, \tag{994}$$

where $\rho = \gamma/z$ is the spectral radius introduced above. That the base of the exponential is ρ follows simply from the definition of the spectral radius, $\rho \equiv \lim_{L\to\infty}(p_{ret}(L))^{1/L}$. Since simple random walks on $\mathbb{Z}$ are recurrent, we accordingly have $\rho(z=2)=1$, while $\rho(z>2)<1$.

expected depth of returning walks

Without the constraint that the walk return to the origin, the endpoint of the walk will concentrate on a depth of $d \sim L(1-2/z)$, since $1-2/z$ is the radial velocity of an unconstrained random walk. How does the expected maximum depth change after we condition on returning to the origin?

One might naively expect that the maximum expected depth would simply be $(L/2)(1-2/z)$, since at any given point one has $1-z$ options to move away from the origin and only 1 option to move towards it. This intuition would suggest that typical walks move ballistically outward until time $L/2$, where the returning constraint makes them "bounce" back inwards towards the origin. This intuition is however *wrong*, as it neglects the large amount of entropy stored in fluctuations in the walk depth.

As a simple way of illustrating this, let us compare the number of returning walks which visit the origin at time $t = L/2$ to those which reach a maximal depth of $L/2$ at $t = L/2$. Keeping only the exponentials,

$$\frac{N_{ret}(x_{t=L/2}=L/2)}{N_{ret}(x_{t=L/2}=0)} \sim \frac{(z-1)^{L/2}}{\gamma^L} \sim 4^{-L/2}, \tag{995}$$

which vanishes exponentially quickly. While our ballistic intuition led us to expect a depth of $(1-2/z)L/2$ rather than $L/2$, these two depths become equal when $z\to\infty$, a limit we can take without affecting our conlusion due to the z-independence of the above equation. We can also falsify the ballistic intuition by noting that the probability of being at the origin at time $t = L/2$ (conditioned on returning at $t = L$) is itself only polynomially (and not exponentially) small, viz.

$$\frac{N_{ret}(x_{t=L/2}=0)}{N_{ret}} \sim L^{-3/2}. \tag{996}$$

Thus conditioning on returning to the origin modifies the behavior of the walks in an exponentially strong way.

Bounds on mixing and relaxation times

In this diary entry we prove two basic facts about the mixing and relaxation times of Markov chains.

Consider an irreducible aperiodic Markov chain with transition matrix M and unique steady state $|\pi\rangle$. We quantify the distance between M^t and $M^\infty = |\pi\rangle\langle\pi|$ in terms of the maximum total variational distance of the probability distribution $(M^t - M^\infty)|w\rangle$, with the maximum taken over initial states $|w\rangle$:

$$d(M^t, M^\infty) \equiv \max_{|w\rangle} ||M^t|w\rangle - |\pi\rangle||_{TV} = \frac{1}{2} \max_{|w\rangle} \sum_{|w\rangle'} |\langle w'|M^t|w\rangle - \langle w'|\pi\rangle| \tag{997}$$

where we used that $||p-q||_{TV} = \frac{1}{2}||p-q||_1$ for probability distributions p, q. The *mixing time* of M is defined as the first time at which $d(M^t, M^\infty)$ drops below some fixed threshold ε:

$$t_{\text{mix}}(\varepsilon) \equiv \min\{t \, : \, d(M^t, M^\infty) < \varepsilon\}. \tag{998}$$

Since the exact value of ε isn't so important, we will fix $t_{\text{mix}} \equiv t_{\text{mix}}(1/3)$. Roughly speaking, the mixing time is reached when the slowest-diffusing initial state $|w_s\rangle$ diffuses out to cover a $\sim 1-\varepsilon$ fraction of the state space. Suppose for example that after t steps, $M^t|w_s\rangle$ is a bump function supported on a fraction f of the state space. Then one easily checks that the mixing time is indeed reached when $f > 1-\varepsilon$.

We begin with the rather obvious result that M cannot mix in a time shorter than the time it takes to traverse state space:

Proposition 21. *Let L be the diameter of the state space in question. Then*

$$t_{\text{mix}} \geq L/2. \tag{999}$$

Proof. We will find it helpful to use an alternate characterization of t_{mix} using the distance measure

$$d'(M^t, M^\infty) \equiv \max_{|w\rangle, |w'\rangle} ||M^t|w\rangle - M^t|w'\rangle||_{TV}, \tag{1000}$$

which makes no explicit reference to the equilibrium distribution. We claim that

$$t_{\text{mix}}(\varepsilon) \geq \min\{t \, : \, d'(M^t, M^\infty) \leq 2\varepsilon\}. \tag{1001}$$

This follows from the fact that

$$d(M^t, M^\infty) \geq \frac{1}{2} d'(M^t, M^\infty), \tag{1002}$$

which is true by virtue of the triangle inequality for the TV:

$$\begin{aligned} \max_{|w\rangle, |w'\rangle} ||M^t|w\rangle - M^t|w'\rangle||_{TV} &\leq \max_{|w\rangle, |w'\rangle} \left(||M^t|w\rangle\pi||_{TV} + ||M^t|w'\rangle - |\pi\rangle||_{TV}\right) \\ &= 2d(M^t, M^\infty), \end{aligned} \tag{1003}$$

so that if $d(M^t, M^\infty) \leq \varepsilon$ then for sure $d'(M^t, M^\infty) \leq 2\varepsilon$.

Now let $|w\rangle, |w'\rangle$ be two states a distance L apart. Then $M^t|w\rangle, M^t|w'\rangle$ have disjoint support for all $t \leq L/2$, and hence $||M^t|w\rangle - M^t|w'\rangle||_{TV} = 1$ for all $t \leq L/2$. □

We now characterize the relaxation time of M in terms of the mixing time. The former is defined by the inverse gap of M, viz.

$$t_{\mathrm{rel}} \equiv \frac{1}{1-\lambda_2}, \tag{1004}$$

where $\lambda_2 < 1$ is the second largest eigenvalue of M. The operational meaning of this

The following proposition lets us bound t_{rel} in terms of t_{mix}, and hence also in terms of L:

Proposition 22. *Assume that the equilibrium distribution of M is uniform, $|\pi\rangle = |1\rangle$ with $|1\rangle \equiv \frac{1}{D}\sum_w |w\rangle$ and D the dimension of state space. Then*[84]

$$t_{\mathrm{rel}} \geq \frac{t_{\mathrm{mix}}(\varepsilon)}{\ln[D/\varepsilon]}. \tag{1005}$$

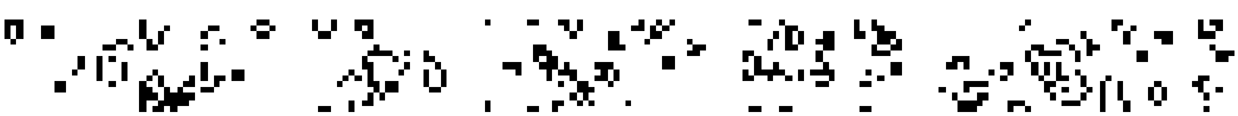
Proof. Let $M = \sum_j \lambda_j |j\rangle\langle j|$ be M's eigendecomposition, with $\lambda_1 = 1$ and $|1\rangle = |\pi\rangle$. Then

$$\begin{aligned}
d(M^t, M^\infty) &= \frac{1}{2}\max_{|w\rangle}\sum_{|w'\rangle}\left(\sum_{j\geq 1}\lambda_j^t\langle w|j\rangle\langle j|w'\rangle - \frac{1}{D}\right) \\
&\leq \frac{1}{2}\max_{|w\rangle}\sum_{|w'\rangle}\sum_{j>1}\lambda_j^t\langle w|j\rangle\langle j|w'\rangle \\
&\leq \frac{1}{2}\max_{|w\rangle}\sum_{|w'\rangle}\lambda_2^t\sqrt{\sum_{j>1}\langle w|j\rangle^2\sum_{k>1}\langle w'|k\rangle^2} \\
&\leq \frac{\lambda_2^t D}{2}
\end{aligned} \tag{1006}$$

since $\sum_{j>1}\langle w|j\rangle^2 = \langle w|(\mathbf{1} - |\pi\rangle\langle\pi|)|w\rangle \leq 1$ for all $|w\rangle$. The claim then follows upon taking $t = t_{\mathrm{mix}}(\varepsilon)$, setting the LHS above equal to $\varepsilon/2$, and using

$$\lambda_2^t = (1 - 1/t_{\mathrm{rel}})^t \leq e^{-t/t_{\mathrm{rel}}}. \tag{1007}$$

□



[84]When π is not uniform I believe D in the log should be replaced by $1/\min_{|w\rangle}\langle w|pi\rangle$.

References

[1] S. Aubert and C. Lam. Invariant integration over the unitary group. *Journal of Mathematical Physics*, 44(12):6112–6131, 2003.

[2] P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik. *Tensor categories*, volume 205. American Mathematical Soc., 2016.

[3] D. T. Son. Chiral metric hydrodynamics, kelvin circulation theorem, and the fractional quantum hall effect. *arXiv preprint arXiv:1907.07187*, 2019.

[4] E. Thomas. On the cohomology of the real grassmann complexes and the characteristic classes of n-plane bundles. *Transactions of the American Mathematical Society*, 96(1):67–89, 1960.

[5] M. Walter. Multipartite quantum states and their marginals. *arXiv preprint arXiv:1410.6820*, 2014.

[6] L. Zhang. Matrix integrals over unitary groups: An application of schur-weyl duality. *arXiv preprint arXiv:1408.3782*, 2014.