Math 273A: Topics in Mathematical Physics

Lecturer: Professor Sourav Chatterjee

Notes by: Andrew Lin

Autumn 2024

Introduction

This course will focus on **Yang-Mills theories**, specifically (discrete) lattice gauge theories. The Clay problem about mass gap is still open and connected to these ideas, and we'll talk about what is known and what remains to be known. There are too many things to choose from, but luckily we'll go through a sequence of two quarters and spread out the content accordingly.

All course logistics are found in the syllabus page (which is now available to us).

1 September 23, 2024

Before we can answer the question of "what Yang-Mills is," we'll have to introduce various preliminary materials from differential geometry and probability and measure theory. We'll assume the proofs (and if we're curious, we can look up the results on our own) so that we can define the Yang-Mills action and Euclidean Yang-Mills theories. From there, we'll construct the simplest example ($U(1)$ gauge theory, corresponding to electromagnetism and Maxwell's theory).

The first background knowledge we'll need is the idea of **tensor product of vector spaces**:

Definition 1

Suppose $V_1, \cdots, V_k$ are vector spaces over a field $\mathbb{F}$. The **tensor product** $V_1 \otimes \cdots \otimes V_k$ is the vector space over $\mathbb{F}$ defined as follows: consider the Cartesian product $V_1 \times \cdots \times V_k$, and let L be the vector space with that set $V_1 \times \cdots \times V_k$ as its basis. (A general element of L is then of the form $\sum_{i=1}^m a_i(v_{i_1}, \cdots, v_{i_k})$.) We then define an equivalence relation as follows: let R be the subspace of L spanned by all elements of the form (for u, v elements of $V_1 \times \cdots \times V_k$)

$$(u_1, \cdots, u_{i-1}, u_i + v_i, u_{i+1}, \cdots, u_k) - (u_1, \cdots, u_k) - (u_1, \cdots, u_{i-1}, v_i, u_{i+1}, \cdots, u_k)$$

or of the form (for some $a \in \mathbb{F}$)

$$(u_1, \cdots, u_{i-1}, au_i, u_{i+1}, \cdots, u_k) - a(u_1, \cdots, u_k).$$

Then $V_1 \otimes \cdots \otimes V_k$ is the quotient L/R; the image of $(v_1, \cdots, v_k)$ in the quotient is denoted $v_1 \otimes \cdots \otimes v_k$.

This is a rather abstract definition, but we'll see some clearer examples soon. In particular, the definition directly

tells us that we must have the "linearity" relations in each coordinate

$$u_1 \otimes \cdots \otimes u_{i-1} \otimes (u_i + v_i) \otimes u_{i+1} \cdots \otimes u_k = u_1 \otimes \cdots \otimes u_k + u_1 \otimes \cdots \otimes u_{i-1} \otimes v_i \otimes u_{i+1} \cdots \otimes u_k,$$

$$u_1 \otimes \cdots \otimes u_{i-1} \otimes au_i \otimes \cdots \otimes u_k = a(u_1 \otimes \cdots \otimes u_k),$$

and in fact there are no other relations.

Example 2

Suppose V is a finite-dimensional vector space of dimension n, and $e_1, \cdots, e_n$ is a basis for V. Then we can produce a basis of $V^{\otimes k}$, namely $\{e_{i_1} \otimes \cdots \otimes e_{i_k}\}$, where $1 \leq i_1, \cdots, i_k \leq n$ are arbitrary (not necessarily distinct) integers; in particular the dimension of the space is n^k.

More specifically, the elements of $(\mathbb{R}^n)^{\otimes k}$ may be written as linear combinations of the form

$$\sum_{1 \leq i_1, \cdots, i_k \leq n} a_{i_1, \cdots, i_k} e_{i_1} \otimes \cdots \otimes e_{i_k},$$

and ignoring those basis vectors, what determines this element is this k-dimensional array of numbers $a_{i_1, \cdots, i_k}$ (so if $k = 2$, this is an $n \times n$ matrix; if $k = 3$, this is an $n \times n \times n$ box, and so on).

We'll be using tensor products everywhere, so it's good to be familiar; one useful way to think about the tensor product of two vector spaces $U \otimes V$ is as follows. If $u_1, \cdots, u_m$ form a basis for U and $v_1, \cdots, v_n$ form a basis for V, then an element of their tensor product is of the form

$$\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} u_i \otimes v_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} u_i \right) \otimes v_j,$$

using our linearity relations. So in fact **elements of $U \otimes V$ are like linear combinations of basis elements of V, with coefficients in U instead of $\mathbb{F}$** – this is a perspective we'll come back to later when discussing differential forms, where the coefficients of our forms come from Lie algebras rather than real or complex numbers.

Now that we've learned about tensor powers, we'll turn to our next useful construction, **exterior powers**:

Definition 3

Let V be a vector space over a field $\mathbb{F}$. For any $k \in \mathbb{N}$, the tensor power $V^{\otimes k}$ is a quotient space; the exterior power is a further quotient space defined as follows. Let S be the subspace of $V^{\otimes k}$ spanned by all elements of the form $v_1 \otimes \cdots \otimes v_k$ such that $v_i = v_{i+1}$ for some $1 \leq i \leq k-1$. Then the ***k*th exterior power of *V***, denoted $\Lambda^k(V)$, is the quotient space $V^{\otimes k}/S$; the image of $v_1 \otimes \cdots \otimes v_k$ in $\Lambda^k(V)$ is denoted $v_1 \wedge \cdots \wedge v_k$ (we call $\wedge$ the **wedge product** or **exterior product**).

Notice in particular that (first by definition of the quotient, then by distributivity)

$$\begin{aligned} 0 = (v_1 + v_2) \wedge (v_1 + v_2) \wedge v_3 \wedge \cdots \wedge v_k & \\ = v_1 \wedge v_1 \wedge v_3 \cdots \wedge v_k + v_1 \wedge v_2 \wedge v_3 \cdots \wedge v_k & \\ + v_2 \wedge v_1 \wedge v_3 \cdots \wedge v_k + v_2 \wedge v_2 \wedge v_3 \cdots \wedge v_k, & \end{aligned}$$

but the first and last terms on the right-hand side are zero, and thus "swapping two adjacent coordinates gives us a negation;" more generally, for any permutation $\pi \in S_k$ we find that

$$v_{\pi(1)} \wedge \cdots \wedge v_{\pi(k)} = \operatorname{sgn}(\pi) v_1 \wedge \cdots v_k.$$

This means that whenever $v_i = v_j$ for some i, j, we can use a permutation to bring them side-by-side and thus the whole thing evaluates to zero.

Example 4

If V is again a finite-dimensional vector space of dimension n with basis $e_1, \cdots, e_n$, then a basis of $\Lambda^k(V)$ is given by "increasing indices"

$$\{e_{i_1} \wedge \cdots \wedge e_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}.$$

By convention we define $\Lambda^0(V) = \mathbb{F}$, and notice that for all $k > n$ we have $\Lambda^k(V) = 0$ (since we will always get repeated indices in this perspective). Thinking now about how to represent this in terms of matrices or arrays, we now have

$$\Lambda^k(\mathbb{R}^n) = \left\{ \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{i_1, \cdots, i_k} e_{i_1} \wedge \cdots \wedge e_{i_k} : a_{i_1, \cdots, i_k} \in \mathbb{R} \right\},$$

so the elements are again determined by these numbers $a_{i_1, \cdots, i_k}$ but only with increasing indices. Thus we can think of this as a k-tensor but also satisfying the relation

$$a_{i_{\pi(1)}, \cdots, i_{\pi(k)}} = \operatorname{sgn}(\pi) a_{i_1, \cdots, i_k}.$$

So for $k = 2$ this is like "an antisymmetric $n \times n$ matrix," and more generally this is called the **antisymmetric property**.

Our next definition (still working towards defining the Yang-Mills action) is that of an **exterior algebra** or **Grassmann algebra** of a vector space.

Definition 5

Let V be a vector space over a field $\mathbb{F}$. The **tensor algebra** $T(V)$ is the direct sum $\oplus_{k=0}^{\infty} V^{\otimes k}$ (where $V^{\otimes 0} = \mathbb{F}$ by convention), equipped with the following product (also denoted $\otimes$). For any $v_1 \otimes \cdots \otimes v_k \in V^{\otimes k}$ and $w_1 \otimes \cdots \otimes w_\ell \in V^{\otimes \ell}$, we define their product to be the element

$$(v_1 \otimes \cdots \otimes v_k) \otimes (w_1 \otimes \cdots \otimes w_\ell) = v_1 \otimes \cdots \otimes v_k \otimes w_1 \otimes \cdots \otimes w_\ell$$

in the tensor product $V^{\otimes (k+\ell)}$.

We need to prove that this is well-defined (since there are multiple ways to represent an element of a tensor product as linear combinations), but it is, and this allows us to tensor together things between different tensor products.

Definition 6

With the notation in the definition above, let I be the two-sided ideal in this algebra generated by elements of the form $v \otimes v$ for $v \in V$. More explicitly, we can write

$$\mathcal{I} = \left\{ \sum_{i=1}^{m} a_i t_{i,1} \otimes v_i \otimes v_i \otimes t_{i,2} : m \geq 0, a_i \in \mathbb{F}, t_{i_1}, t_{i_2} \in T(V), v_i \in V \right\}.$$

The **exterior algebra** or **Grassmann algebra** is then the quotient $\Lambda(V) = T(V)/\mathcal{I}$; we let the image of $v_1 \otimes v_k$ be written as $v_1 \wedge \cdots \wedge v_k$, as usual.

It turns out that as a vector space, we do indeed have the alternate characterization $\Lambda(V) = \oplus_{k=0}^{\infty} \Lambda^k(V)$, with product (from $\Lambda^k(V) \times \Lambda^\ell(V)$ to $\Lambda^{k+\ell}(V)$) given by

$$(u_1 \wedge \cdots \wedge u_k) \wedge (v_1 \wedge \cdots \wedge v_\ell) = u_1 \wedge \cdots \wedge u_k \wedge v_1 \wedge \cdots \wedge v_\ell.$$

Thus the exterior algebra is a way of putting together exterior powers together using a multiplication between those spaces.

We've been talking so far about a general field $\mathbb{F}$, but we'll basically always use $\mathbb{R}$ in this class, and that's because we want to work with **smooth manifolds**. In short, a smooth manifold is a space that "locally looks like an open subset of $\mathbb{R}^n$:"

Definition 7

An n-dimensional **smooth manifold** M equipped with an **atlas** $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in \mathcal{A}}$ is a second countable Hausdorff topological space such that

1. each U_α is an open subset of M (this makes sense because we have a topology),
2. each ϕ_α is a homeomorphism betwen U_α and some open subset $\phi_\alpha(U_\alpha) \subseteq \mathbb{R}^n$,
3. $M = \bigcup_{\alpha \in \mathcal{A}} U_\alpha$, and
4. if U_α and U_β intersect, then $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$ between open sets in $\mathbb{R}^n$ is a smooth (C^∞) map.

We call elements (U, ϕ) of the atlas **charts**; we say that (U, ϕ) is **around** a point x if $x \in U$.

Two atlases are compatible if their union is also an atlas; for any atlas we can take the union of all atlases compatible with it, and this determines a differentiable structure of a manifold (so that's where the "uniqueness" comes in). But this is just a background detail.

Definition 8

Let M and N be two manifolds (from now on, "manifold" will mean "finite-dimensional smooth manifold"). A map $f : M \to N$ is **smooth** if for all $x \in M$, there is some chart (U, ϕ) around x and some chart (V, ψ) around $f(x)$ such that $\psi \circ f \circ \phi^{-1} : \phi(U \cap f^{-1}(V)) \to \psi(V)$ is smooth.

(In other words, smoothness needs to be checked by "pulling back to charts.") We'll continue next time with many more of these definitions – vector bundles, fiber bundles, cotangent spaces, and so on!

2 September 25, 2024

We'll continue with preliminaries in differential geometry today – last time, we defined smooth maps between (smooth, finite-dimensional) manifolds in terms of smoothness on charts (converting to a statement about smooth maps on Euclidean space).

Definition 9

The set of all smooth maps from M to $\mathbb{R}$ is denoted by $C^\infty(M)$.

Definition 10

A smooth map $f : M \to N$ between manifolds is a **diffeomorphism** if it is a bijection and the inverse $f^{-1} : N \to M$ is also smooth.

We'll sometimes need to deal with Cartesian products of manifolds, which are defined in the "natural way:"

Definition 11

Let M be a manifold with atlas $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in \mathcal{A}}$. The **Cartesian product** $M \times M$ has a natural manifold structure with atlas $\{(U_\alpha \times U_\beta, \phi_\alpha \times \phi_\beta)\}_{\alpha,\beta \in \mathcal{A}}$, where we define the charts "coordinate-wise" as

$$(\phi_\alpha \times \phi_\beta)(x, y) = (\phi_\alpha(x), \phi_\beta(y)).$$

In particular, if M has dimension n, then $M \times M$ will have dimension $2n$.

We'll also be integrating on our manifolds, so we will also need the notion of **orientability**. Recall that if U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^n$ is a smooth map, the **Jacobian** of $f = (f_1, \cdots, f_n)$ is defined to be

$$J(x) = \det\left(\left(\frac{\partial f_i}{\partial x_j}\right)_{1 \le i,j \le n}\right),$$

Definition 12

Let M be a (smooth) n-dimensional manifold. We say that M is **orientable** if there exists an atlas $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in \mathcal{A}}$ of M such that for any $\alpha, \beta \in \mathcal{A}$ with $U_\alpha \cap U_\beta \neq \varnothing$, the map $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$ has positive Jacobian everywhere.

Intuitively, we form M by gluing together open subsets of $\mathbb{R}^n$ together, and we want to do so in a way that gives us a well-defined outward direction everywhere (for example, the Mobius strip does not have this property, but the sphere does). We'll need this property because integration on manifolds should not depend on the particular charts that we use, and unless we have orientability we will run into problems.

We'll now move to some more complicated constructions which are important for Yang-Mills theories:

Definition 13

A **smooth fiber bundle** is a 4-tuple (E, B, π, F) satisfying the following conditions:

1. The **total space** E, **base space** B, and **typical fiber** F are all smooth manifolds.
2. The **projection map** π is a smooth surjective map from E to B.
3. For each $x \in B$, there is an open neighborhood U of x and a diffeomorphism $\phi : \pi^{-1}(U) \to U \times F$ such that $\pi(\phi^{-1}(u, y)) = u$ for all $u \in U$ – we call this a **local trivialization**.

This is a very abstract definition, but the picture to have in mind is that we have some manifold B and we put a copy of F on top of each point in B; the actual space E is the collection of all such fibers together, and the projection map takes any point and maps us to the point in B that the fiber comes from. So condition 3 is sort of like saying that "the preimage of x looks like F," but it's imposing an additional condition on how these copies of F must smoothly vary in x – locally everything looks like a Cartesian product, but globally things may twist.

Example 14 (Trivial fiber bundle)

Let B and F be any two manifolds. Then letting $E = B \times F$ (defining this similarly to how we defined $M \times M$ above in Definition 11) and having $\pi : E \to B$ defined by $\pi(x, y) = x$, we see that (E, B, π, F) is a smooth fiber bundle. Indeed, for any x, we can let the neighborhood U just be B itself (so $\pi^{-1}(U)$ is all of $E = B \times F$) and let the diffeomorphism $\phi : \pi^{-1}(U) \to U \times F$ be the identity map.

This indeed satisfies all of our requirements – indeed, $\pi(\phi^{-1}(u,y)) = \pi(u,y) = u$. The reason we care about fiber bundles instead of just working directly with Cartesian products is that the additional "global twisting" can be helpful, and we often want to study how this twisting propagates:

Definition 15

Let (E, B, π, F) be a fiber bundle (sometimes we just write E for brevity). A **smooth section** of this fiber bundle is a smooth map $s : B \to E$ such that $\pi(s(x)) = x$ for all $x \in B$.

Referring to the previous picture again, we have a copy of our fiber F at each base point $x \in B$; a section chooses a point $s(x)$ from each fiber so that it varies smoothly.

Definition 16

We say that a fiber bundle is a **vector bundle** if

1. there is some n such that the **fibers** $\pi^{-1}(\{x\})$ are $\mathbb{R}$-vector spaces of dimension n,
2. the typical fiber F is $\mathbb{R}^n$, and
3. each local trivialization $\phi : \pi^{-1}(U) \to U \times \mathbb{R}^n$ has the property that $\phi^{-1}(x, \cdot)$ is a **linear** isomorphism between the vector spaces $\pi^{-1}(\{x\})$ and $\mathbb{R}^n$ for all $x \in U$.

In other words, these local trivializations give not just diffeomorphisms but linear isomorphisms. This is one of the two types of bundles we'll deal with in this course (the other is what is called a **principal bundle**, which is more complicated).

Example 17

Consider a particle in 3-dimensional space. Then the velocity of the particle is another 3-vector, so (position, velocity) can be thought of as a vector bundle (this is what's called a **phase space** for the particle).

Definition 18

Let $(E_1, B, \pi_1, \mathbb{R}^m)$ and $(E_2, B, \pi_2, \mathbb{R}^n)$ be two vector bundles over the same base space B. The **tensor product** of the two vector bundles $(E, B, \pi, \mathbb{R}^m \otimes \mathbb{R}^n)$ is defined as follows:

1. The total space E is the set of points $(x, u \otimes v)$ for $x \in B$, $u \in \pi_1^{-1}(\{u\})$, and $v \in \pi_2^{-1}(\{x\})\}$.
2. The projection map is given by $\pi(x, u \otimes v) = x$.
3. The local trivializations are defined as follows: for each $x \in B$, we can find an open set $U \ni x$ and local trivializations ϕ_1, ϕ_2 of E_1, E_2 (here we can take U to be the intersection of the open sets U_1, U_2 from those trivializations). Remembering that ϕ_1 is a map $\pi_1^{-1}(U) \to U \times \mathbb{R}^m$ and ϕ_2 is a map $\pi_2^{-1}(U) \to U \times \mathbb{R}^n$, write $\phi_1(s) = (u, \tilde{\phi}_1(s))$ and $\phi_2(t) = (u, \tilde{\phi}_2(t))$ (for any $s \in \pi_1^{-1}(u)$ and $t \in \pi_2^{-1}(u)$). We then define

$$\phi(u, s \otimes t) = (u, \tilde{\phi}_1(u, s) \otimes \tilde{\phi}_2(u, t))$$

for any $u \in U, s \in \pi_1^{-1}(u)$, and $t \in \pi_2^{-1}(u)$, and then we extend to the whole space by linearity.

In words, we cannot take tensor products of manifolds, so we just take tensor products of the fibers, and we combine the local trivializations from the two bundles together. One natural place in which these bundles come up is the following:

Definition 19 (Informal)

Let M be an n-dimensional manifold. A **tangent vector** at a point $x \in M$ is defined in the following way: a **curve** in M is a smooth map $\gamma : (-\varepsilon, \varepsilon) \to M$ for some $\varepsilon > 0$, viewing $(-\varepsilon, \varepsilon)$ as a 1-dimensional manifold. Consider all curves γ such that $\gamma(0) = x$, and consider the "derivative" via evaluation by smooth maps; we say two curves are equivalent if every f has the same derivative. An equivalence class is then a tangent vector.

Putting together the spaces of tangent vectors then gives us a vector space at each point; putting these together gives us a tangent bundle. We'll see the definition in more detail on Friday!

3 September 27, 2024

Last time, we defined the tensor product of two vector bundles in Definition 18 – we may look there to recall the details. Our first goal today is to define some specific such examples of vector bundles that will come up in this course – we'll first elaborate a bit more on Definition 19:

Definition 20

Let M be an n-dimensional manifold. For any $x \in M$ and $\gamma : (-\varepsilon, \varepsilon) \to M$ a smooth function with $\gamma(0) = x$ (we say the curve "passes through x at time 0"), take any $f \in C^\infty(M)$. The **derivative of f at x in the direction γ** is the number $\frac{d}{dt} f(\gamma(t))\big|_{t=0}$. We say that two curves γ_1, γ_2 passing through x at time 0 are **equivalent** if $\frac{d}{dt} f(\gamma_1(t))\big|_{t=0} = \frac{d}{dt} f(\gamma_2(t))\big|_{t=0}$ for all smooth $f \in C^\infty(M)$ (basically, the two curves go in the same direction).

An equivalence class of the above form is called a **tangent vector at x**, and the set of tangent vectors at x is called the **tangent space at x**, denoted $T_x M$.

It turns out that $T_x M$ has a natural vector space structure, and it always has dimension n (the same as the manifold) over $\mathbb{R}$. We can construct a basis of this vector space as follows: take a chart (U, ϕ) around x, and define the curves

$$\gamma_i(t) = \gamma^{-1}(\phi(x) + te_i)$$

for $1 \leq i \leq n$, where e_i are the standard basis vectors. (In other words, each of these curves is traveling in one of the straight-line coordinates on the chart.) Indeed, $\gamma_i(0) = x$ for all i, and we'll let $\frac{\partial}{\partial x_i}$ denote the corresponding tangent vector to γ_i – note that this is **not intrinsic** and depends on the chart chosen. So for any γ, (it turns out) there are unique numbers $a_1, \cdots, a_n$ such that

$$\frac{d}{dt} f(\gamma(t))\bigg|_{t=0} = \sum_{i=1}^{n} a_i \frac{\partial f}{\partial x_i}$$

for all $f \in C^\infty(M)$, where $\frac{\partial f}{\partial x_i} = \frac{d}{dt} f(\gamma_i(t))\big|_{t=0}$. Thus the tangent vector v corresponding to γ is exactly specified by these arbitrary real numbers $(a_1, \cdots, a_n)$, and thus we formally write $v = \sum_{i=1}^{n} a_i \frac{\partial}{\partial x_i}$.

Definition 21

With the notation in the definition above, the **cotangent space** $T_x^* M$ is the dual space of $T_x M$.

In other words, we choose a chart (U, ϕ) around x to determine the basis vectors $\frac{\partial}{\partial x_i}$; the corresponding dual basis is denoted dx_i (meaning that $dx_i\left(\frac{\partial}{\partial x_j}\right) = \delta_{ij}$ is 1 if $i = j$ and 0 otherwise).

Definition 22

The **tangent bundle** of a manifold M is the vector bundle

$$TM = \{(x, v) : x \in M, v \in T_x M\}.$$

In other words, TM is the total space, the base space is M, and the projection map is $\pi((x, v)) = x$.

If $\{(U_\alpha, \phi_\alpha)\}_{\alpha \in \mathcal{A}}$ is an atlas of M, then $V_\alpha = \{(x, v) : x \in U_\alpha, v \in T_x M\}$ are open sets of the total space. Then for any $x \in U_\alpha$ and $v \in T_x M$, there are unique $a_1, \cdots, a_n$ with $v = \sum_{i=1}^n a_i \frac{\partial}{\partial x_i}$, so then we can define $\psi_\alpha : V_\alpha \to \mathbb{R}^n \times \mathbb{R}^n$ via

$$\psi_\alpha((x, v)) = (\phi_\alpha(x), (a_1, \cdots, a_n)).$$

This definition means $\{(V_\alpha, \psi_\alpha)\}_{\alpha \in \mathcal{A}}$ is an atlas making TM into a $2n$-dimensional manifold (much like the phase space example from last lecture). The local trivializations then require a bit more work to write down, but we won't do that here.

Definition 23

The **cotangent bundle** of a manifold is the vector bundle $T^*M = \{(x, v) : x \in M, v \in T_x^* M$, where we construct the bundle analogously to above.

We may also construct exterior powers $\Lambda^k(T_x^* M)$ of the contangent spaces; specifically, putting them together again yields a vector bundle

$$\Lambda^k(T^*M) = \{(x, v) : x \in M, v \in \Lambda^k T_x^* M.\}$$

We'll come back to this later when we think more about differential forms.

Definition 24

Let M be a (smooth) manifold. A (smooth) **vector field** on M is a (smooth) section of the tangent bundle TM. We denote the set of smooth vector fields on M by $\mathfrak{X}(M)$.

Unpacking this definition, a vector field is a map $f : M \to TM$ such that $\pi(f(x)) = x$ for all $x \in M$; thus, each point on the manifold is mapped to something in its tangent space, and we do so in a way that varies smoothly with x. Writing $f(x) = (x, X(x))$, we can thus specify a vector field via this map X (that is, we really only need the second coordinate of the map f). It's important that we use the language of a tangent bundle here, because we can't think of X as a map from M into some other fixed space – the image changes with x. But intuitively we should really just think of it as a gradually changing tangent vector.

For any $X \in \mathfrak{X}(M)$, we get an action on $C^\infty(M)$ defined by

$$X(f)(x) = X(x)(f).$$

In other words, for any tangent vector $v \in T_x M$ which we write as $\sum a_i \frac{\partial}{\partial x_i}$, this tangent vector "acts on the function" $v(f) = \sum a_i \frac{\partial f}{\partial x_i}$ by taking the directional derivative. So X maps from $C^\infty(M)$ to $C^\infty(M)$, where given a function, we get out another function encoding the directional derivative in the direction specified by X.

Definition 25

Let $X, Y \in \mathfrak{X}(M)$ be two vector fields. The map $[X,Y] : C^\infty(M) \to C^\infty(M)$, called the **Lie bracket** of X and Y, is defined by

$$[X,Y](f) = X(Y(f)) - Y(X(f)).$$

Lemma 26

For any $X, Y \in \mathfrak{X}(M)$, there is some $Z \in \mathfrak{X}(M)$ such that $Z(f) = [X,Y](f)$ for all $f \in C^\infty(M)$; in other words, the Lie bracket is also another vector field.

Proof. The typical idea of proofs in this area is to take local coordinates (from a chart) and work on them directly. In local coordinates (suppressing the "evaluation at x"), we may write $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_n)$, meaning that $X(x) = (X_1(x), \cdots, X_n(x))$ for some functions depending on the chart. We then have

$$\begin{aligned}
[X,Y](f) &= X(Y(f)) - Y(X(f)) \\
&= X\left(\sum Y_i \frac{\partial f}{\partial x_i}\right) - Y\left(\sum X_i \frac{\partial f}{\partial x_i}\right) \\
&= \sum_{i=1}^n X_i \frac{\partial}{\partial x_i}\left(\sum_{j=1}^n Y_j \frac{\partial f}{\partial x_j}\right) - \sum_{i=1}^n Y_i \frac{\partial}{\partial x_i}\left(\sum_{j=1}^n X_j \frac{\partial f}{\partial x_j}\right),
\end{aligned}$$

remembering that we can think of these as "regular partial derivatives" now that we're working in local coordinates. But by the product rule we see that the second-derivative terms cancel out because the order of partial derivatives doesn't matter:

$$\begin{aligned}
[X,Y](f) &= \sum_{i=1}^n \sum_{j=1}^n \left(X_i \frac{\partial Y_j}{\partial x_i}\frac{\partial f}{\partial x_j} + X_i Y_j \frac{\partial^2 f}{\partial x_i \partial x_j} - Y_i \frac{\partial X_j}{\partial x_i}\frac{\partial f}{\partial x_j} - Y_i X_j \frac{\partial^2 f}{\partial x_i \partial x_j} \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \left(X_i \frac{\partial Y_j}{\partial x_i} - Y_i \frac{\partial X_j}{\partial x_i} \right) \frac{\partial f}{\partial x_j}.
\end{aligned}$$

So if we define the red terms to be Z_j, then putting those together gives us a tangent vector, and that's exactly what $Z = [X,Y]$ is. □

Definition 27

Let G be a group that is also an n-dimensional manifold. We say that G is a **Lie group** if the maps $(g,h) \mapsto gh$ (from $G \times G$ to G) and $g \mapsto g^{-1}$ (from G to G) are smooth.

In other words, Lie groups are objects such that the group and manifold structure are compatible with each other.

Example 28

The group of all $n \times n$ invertible complex matrices $GL(n, \mathbb{C})$ is a Lie group under matrix multiplication. Similarly, the groups $U(n)$ of $n \times n$ unitary matrices and $SU(n)$ of $n \times n$ unitary matrices with determinant 1 are also Lie groups.

Note that even though our matrices are complex-valued, we're always thinking of these as manifolds under $\mathbb{R}$. So for example, $GL(n, \mathbb{C})$ is a $2n^2$-dimensional manifold, since for any invertible matrix a small neighborhood is also

invertible. (Next time, we'll define the Lie algebra of a Lie group, and that's often the easiest way to compute the exact dimension of the corresponding manifold.)

4 September 30, 2024

We'll continue discussing Lie groups and Lie algebras today – last time, we defined a Lie group to be a manifold and also a group, such that the operations of group multiplication are smooth with respect to the manifold structure.

Definition 29
Let G be a Lie group. The **Lie algebra** of G, denoted $\mathfrak{g}$, is the tangent space of G at its identity element.

As usual, this is a pretty abstract definition, but we'll come to some more concrete examples later (and also see why it's called an "algebra") through the following concept:

Definition 30
We call Lie subgroups of $GL(n, \mathbb{C})$ **matrix Lie groups**.

We are interested in such Lie groups because for any such matrix group $G \subseteq GL(n, \mathbb{C})$, the Lie algebra $\mathfrak{g}$ can be represented as a subspace of the $\mathbb{R}$-vector space $M(n, \mathbb{C})$ of $n \times n$ complex matrices. Indeed (recalling the definition), a tangent vector at the identity $e = I_n \in G$ is an equivalence class of smooth curves $\gamma : (-\varepsilon, \varepsilon) \to G$ such that $\gamma(0) = I_n$, and it turns out this equivalence class is exactly determined by the coordinate-wise derivative $\gamma'(0) \in M(n, \mathbb{C})$. (So the Lie algebra is exactly the set of possible derivatives if our curves lie in G.)

Example 31
Suppose $G = U(n)$ is the group of unitary $n \times n$ matrices, and $\gamma : (-\varepsilon, \varepsilon) \to U(n)$ is a curve passing through the identity at time 0. We wish to understand restrictions on $\gamma'(0)$.

By definition of unitarity, we have for all t that $\gamma(t)\gamma(t)^* = I$, so taking the coordinate-wise derivative, we have

$$0 = \frac{d}{dt}\left(\gamma(t)\gamma(t)^*\right) = \gamma'(t)\gamma(t)^* + \gamma(t)(\gamma'(t))^*,$$

so evaluating at $t = 0$ yields $0 = \gamma'(0) + \gamma'(0)^*$, meaning that $\gamma'(0)$ must be a **skew-Hermitian matrix**. Thus the Lie algebra $\mathfrak{u}(n)$ is a subspace of the vector space of skew-Hermitian matrices, and in fact we claim it is equal to that space. indeed, for any skew-Hermitian matrix A (meaning $A^* = -A$), we may construct the (smooth) curve $\gamma : \mathbb{R} \to U(n)$ defined by

$$\gamma(t) = e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k.$$

We can check that this sum converges, and by the spectral theorem this is a unitary matrix for all t (because iA is Hermitian, so it has a spectral decomposition with all real eigenvalues, so A has one with all imaginary eigenvalues, meaning e^{tA}'s eigenvalues all lie on the unit circle). Furthermore $\gamma(0) = I$ and $\gamma'(0) = A$, so this indeed verifies that all of the skew-Hermitian matrices are in the Lie algebra.

Remark 32. *We may use similar reasoning to compute Lie algebras for other matrix groups. For example, we can check that the Lie algebra of $SU(n)$ is the set of all skew-Hermitian matrices with trace zero. And this construction of an "exponential map" is much more general than the example that we just did.*

Turning back to the abstract case (so G is a general Lie group), if we now consider any curve $\gamma : (-\varepsilon, \varepsilon) \to G$ with $\gamma(0) = e$ and take any $g \in G$, we may define another curve $\eta : (-\varepsilon, \varepsilon) \to G$ via $\eta(t) = g\gamma(t)$ (that is, we multiply the entire curve by g). So now $\eta(0) = g$, so this gives us a way of pushing forward a tangent vector X at the identity to a tangent vector at some other element. We call this new tangent vector at g the **pushforward of the tangent vector X at e to g** and denote it $X(g)$. For example, if G is a matrix group and $X \in \mathfrak{g}$ is any Lie algebra element, then the pushforward is just gX.

In particular, given any $X \in \mathfrak{g}$, we get a pushforward of X at every point in G; putting these all together induces a smooth vector field on G, which we also call X.

Definition 33

Let $X, Y \in \mathfrak{g}$ be two elements of a Lie algebra. Viewing X and Y as vector fields, let $Z = [X, Y]$ be their Lie bracket (as defined in Definition 25). Then the value of Z at e is again an element of the Lie algebra, which we also denote by $[X, Y]$.

We'll mostly be working with matrix Lie groups, and for any matrix Lie group the Lie bracket is exactly the commutator

$$[X, Y] = XY - YX,$$

where XY is now just regular matrix multiplication. But in any case, this is where the name "algebra" comes from – the product $X, Y \mapsto [X, Y]$ defines a product on $\mathfrak{g}$, and it satisfies the **Jacobi identity**

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

This can be verified for matrix groups by plugging in $[X, Y] = XY - YX$, and more generally we use the definition in terms of vector fields.

We'll now turn to a bit of representation theory (of Lie groups and Lie algebras); physicists say that "any particle corresponds to a representation of the Poincaré group," but we won't dive into that too much here.

Definition 34

Let G be a Lie group and V be a vector space over $\mathbb{R}$ or $\mathbb{C}$, and let $GL(V)$ denote the group of invertible linear operators on V (that is, linear maps from V to V). A **representation of G in V** is a group homomorphism $\rho : G \to GL(V)$, meaning that $\rho(e) = I$ and $\rho(gh) = \rho(g)\rho(h)$.

(This definition automatically implies that $\rho(g)^{-1} = \rho(g^{-1})$.)

Example 35

Let G be a matrix Lie group (so it is a subset of $GL(n, \mathbb{C})$). Then the **fundamental representation** of G is the map which sends g to itself (viewed as an element of $GL(\mathbb{C}^n)$, meaning that $\rho(g)x = gx$.

Example 36

Let G be a Lie group with identity element e, and let $\mathfrak{g}$ be its Lie algebra. Recalling that the Lie algebra is itself a vector space, the **adjoint representation** of G is the map $\mathrm{Ad} : G \to GL(\mathfrak{g})$ defined in the following way. For any $g \in G$ and any $X \in \mathfrak{g}$, let $\gamma : (-\varepsilon, \varepsilon) \to G$ be a smooth curve in the equivalence class that X represents. Then the curve $\eta(t) = g\gamma(t)g^{-1}$ is also a curve passing through e at time 0, so it corresponds to some other tangent vector Y. We say that $\mathrm{Ad}(g)(X) = Y$, and we sometimes write Ad_g instead of $\mathrm{Ad}(g)$.

It turns out this is a linear map (on X, not on G), and in words $\mathrm{Ad}(g)$ turns an equivalence class of vector fields into another one via conjugation by g. And in the case where G is a matrix Lie group, we just have $\mathrm{Ad}(g)(X) = gXg^{-1}$, so it's easy to check that we do have a representation here. Indeed, we verify that

$$\mathrm{Ad}(h)(\mathrm{Ad}(g)(X)) = h(gXg^{-1})h^{-1} = (hg)X(hg)^{-1} = \mathrm{Ad}(hg)(X),$$

so we indeed have a group homomorphism.

Definition 37

Let G be a Lie group with Lie algebra $\mathfrak{g}$, and let V be a vector space over $\mathbb{R}$. A **representation of $\mathfrak{g}$ in V** is an algebra homomorphism $\rho : \mathfrak{g} \to \mathfrak{gl}(V)$. In other words, ρ is a linear map which also respects the Lie bracket, so that $\rho([X, Y]) = [\rho(X), \rho(Y)]$ (where the left Lie bracket is on $\mathfrak{g}$ and the right one is on $\mathfrak{gl}(V)$).

Note that $\mathfrak{gl}(V)$ is the set of linear maps from V to V but now **not** requiring invertibility.

Example 38

Closely connected to Example 36, the **adjoint representation** of a Lie algebra $\mathfrak{g}$ is the map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ defined by

$$\mathrm{ad}(X)(Y) = [X, Y].$$

Like above, we often write ad_X in place of $\mathrm{ad}(X)$.

In other words, $\mathrm{ad}(X)$ is a map from $\mathfrak{g}$ into $\mathfrak{g}$ which is itself linear (since the Lie bracket is linear). The main thing to check is that we satisfy the condition for being a homomorphism – indeed, we have

$$\begin{aligned}
\mathrm{ad}_{[X,Y]}(Z) &= [[X, Y], Z] \\
&= [X, [Y, Z]] - [Y, [X, Z]] \\
&= \mathrm{ad}_X([Y, Z]) - \mathrm{ad}_Y([X, Z]) \\
&= \mathrm{ad}_X(\mathrm{ad}_Y(Z)) - \mathrm{ad}_Y(\mathrm{ad}_X(Z)) \\
&= [\mathrm{ad}_X, \mathrm{ad}_Y](Z)
\end{aligned}$$

by the Jacobi identity (and also noting that $[X, Y] = -[Y, X]$).

5 October 2, 2024

Today's topic will be **differential forms** – we'll then confirm these objects with the definitions we've made before to later define connections and gauge fields.

Definition 39

Let M be an n-dimensional (smooth) manifold. A **differential 0-form** on M is an element of $C^\infty(M)$ (that is, a smooth zero-dimensional function). For $1 \leq k \leq n$, a **differential k-form** is a smooth section of $\Lambda^k T^*M$. We write $\Omega^k(M)$ for the space of all k-forms on M.

Recall that $\Lambda^k T^*M$ is a vector bundle over M, where the total space is of the form $\{(x, v) : x \in M, v \in \Lambda^k T^*_k M$. (The projection map sends (x, v) to x, and the rest is defined as we typically do so for the cotangent bundle.) A

smooth section is then a map which assigns to each $x \in M$ some point (x, v) so that the "second coordinate varies smoothly."

The picture here becomes more clear if we write everything in local coordinates: any k-form ω can be written in the form $\omega(x) = (x, \tilde{\omega}(x))$, and by abuse of notation we'll just write $\tilde{\omega}$ as ω (dropping the first coordinate). We then write

$$\omega(x) = \sum_{1 \le i_1 < \cdots < i_k \le n} \omega_{i_1, \cdots, i_k}(x) dx_{i_1} \wedge \cdots \wedge dx_{i_k},$$

where these coefficients $\omega_{i_1, \cdots, i_k}(x)$ are a set of $\binom{n}{k}$ functions that determine the differential k-form. (But remember that these functions depend on our choice of coordinates and of charts – we cannot just say that a k-form is a collection of functions.)

What we've discussed so far is a "real-valued" form, but we'll now generalize to something beyond what we might typically see in a differential geometry course:

Definition 40

Let M be an n-dimensional (smooth) manifold, and let $\mathfrak{g}$ be a Lie algebra. For $1 \le k \le n$, a **$\mathfrak{g}$-valued k-form** is a smooth section of the bundle $(M \times \mathfrak{g}) \otimes \Lambda^k T^*M$. (0-forms are defined identically to the real-valued case.) We write $\Omega^k(M, \mathfrak{g})$ for the space of $\mathfrak{g}$-valued k-forms on M.

Here, $M \times \mathfrak{g}$ is the trivial bundle (so we have a copy of $\mathfrak{g}$ at every point in M) over M, and we take the tensor product of that bundle with $\Lambda^k T^*M$. So the base space M stays the same, and so a typical element of this bundle is of the form $(x, \sum_{i=1}^m A_i \otimes v_i)$, where $A_i \in \mathfrak{g}$ and $v_i \in \Lambda^k T_x^* M$ (and where m is arbitrary).

To describe what a smooth section is, it's easiest to first write out the charts of the bundle to understand the topology; it turns out that in local coordinates, an element $A \in \Omega^k(M, \mathfrak{g})$ can be written as

$$A = \sum_{1 \le i_1 < \cdots < i_k \le n} A_{i_1, \cdots, i_k} \otimes (dx_{i_1} \wedge \cdots \wedge dx_{i_k}).$$

So indeed here, we're taking a linear combination of the basis elements $(dx_{i_1} \wedge \cdots \wedge dx_{i_k})$ of the cotangent space, and our coefficients $A_{i_1, \cdots, i_k}$ are now Lie algebra-valued. (But again, if we change coordinates, those functions will change.) For notational convenience, we often write $A = \sum A_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ and drop the tensor product.

For noncommutative Lie groups, this really is an interesting generalization to the real-valued case, and some complications with **curvature** will come up in such examples.

We're now ready to show why we've gone to such trouble to do all of these definitions in detail. If we want to do integration on a manifold, it may seem like we can do so on charts and carry everything out on subsets of $\mathbb{R}^n$ instead, but **different coordinate charts will give us different values**. So we cannot just get a well-defined integral given a smooth function on a smooth manifold, and one fundamental advantage of differential forms is that they do indeed have well-defined integrals.

Example 41

Let M be a smooth n-dimensional manifold, and let $\omega \in \Omega^n(M)$ be a (real-valued) n-form, meaning we can write it in local coordinates as $\omega(x) = a(x) dx_1 \wedge \cdots \wedge dx_n$. We wish to integrate ω over M.

We can first consider the simple case where ω vanishes outside of a single coordinate chart (U, ϕ). Then we can **define** (note that we don't have any "dx" written after the integral

$$\int_M \omega = \int_{\phi(U)} a(\phi^{-1}(x)) dx.$$

In other words, we pull (U, ϕ) back to $\mathbb{R}^n$, and we integrate the function a. This quantity turns out to be coordinate-independent because of how we set up our definitions:

Lemma 42
The definition of $\int_M \omega$ above does not depend on the choice of (U, ϕ), as long as ω indeed vanishes outside the chart.

Proof. Suppose (V, ψ) is another chart such that ω vanishes outside of V. Then ω also vanishes outside $U \cap V$, so we can just work on the intersection (and we can assume it's nonempty, otherwise the integral is zero). We now have $U \cap V$ mapped to $\phi(U \cap V)$ and $\psi(U \cap V)$, which are both subsets of $\mathbb{R}^n$.

For $x \in U \cap V$, let $dx_1, \cdots, dx_n$ be the standard basis of $T_x^* M$ in the chart (U, ϕ) and $dx_1', \cdots, dx_n'$ be the standard basis of $T_x^* M$ in the other chart (V, ψ). (So $T_x^* M$ is the same space in both cases, but we're choosing different bases of that vector space depending on where the coordinate axes map onto M via the charts.) Then the relation between these bases (basically the Jacobian matrix) can be written

$$dx_i = \sum_{j=1}^{n} \frac{\partial (\phi \circ \psi^{-1})_i}{\partial x_j} dx_j'.$$

But recall that locally we defined $\omega(x) = a(x) dx_1 \wedge \cdots \wedge dx_n$, and substituting in the above equation yields

$$\omega(x) = a(x) \left(\sum_{j=1}^{n} \frac{\partial (\phi \circ \psi^{-1})_1}{\partial x_j} dx_j' \right) \wedge \cdots \wedge \left(\sum_{j=1}^{n} \frac{\partial (\phi \circ \psi^{-1})_n}{\partial x_j} dx_j' \right).$$

But wedge products are distributive across addition, so we may rewrite this (by picking a j_i for each of the terms in the wedge product)

$$\omega(x) = a(x) \sum_{1 \le j_1, \cdots j_n \le n} \left(\prod_{i=1}^{n} \frac{\partial (\phi \circ \psi^{-1})_i}{\partial x_{j_i}} \right) dx_{j_1}' \wedge \cdots \wedge dx_{j_n}'.$$

But now remember that $dx_{j_1}' \wedge \cdots \wedge dx_{j_n}' = 0$ if the j_is are not distinct, and otherwise it is equal to $\operatorname{sgn}(j) dx_1' \wedge \cdots \wedge dx_n'$ (where $\operatorname{sgn}(j)$ is the sign of the permutation $(j_1, \cdots, j_n)$). Thus the above equation simplifies to

$$\omega(x) = a(x) \left(\sum_{j \in S_n} \operatorname{sgn}(j) \frac{\partial (\phi \circ \psi^{-1})_1}{\partial x_{j_1}} \cdot \cdots \cdot \frac{\partial (\phi \circ \psi^{-1})_n}{\partial x_{j_n}} \right) dx_1' \wedge \cdots \wedge dx_n',$$

and now everything in blue is exactly the Jacobian of $\phi \circ \psi^{-1}$ by the determinant expansion formula. Thus the two definitions indeed agree, because by the usual change of variables formula we have

$$\begin{aligned} \int_{\phi(U \cap V)} a(\phi^{-1}(x)) dx &= \int_{\psi(U \cap V)} a(\phi^{-1}(\phi \circ \psi^{-1}(y))) J(y) dy \\ &= \int_{\psi(U \cap V)} a(\psi^{-1}(y)) J(y) dy, \end{aligned}$$

as desired. □

So the point is that we "make up" for the Jacobian term when we integrate n-forms instead of smooth functions, and that's what allows us to get a well-defined integral. And now we can finish the definition of integration:

Definition 43

Let $\omega \in \Omega^n(M)$ be an n-form. Suppose that we can find a finite collection of charts $(U_1, \phi_1), \cdots, (U_N, \phi_N)$ such that ω vanishes outside of the union $\bigcup_{i=1}^N U_i$. (For example, this is always possible if we have a compact manifold.) Then we can find smooth nonnegative functions $\rho_1, \cdots, \rho_N \in C^\infty(M)$ such that ρ_i vanishes outside of U_i for all i and $\sum_{i=1}^N \rho_i$ is identically 1 on $\bigcup_{i=1}^N U_i$ (we call this a **partition of unity**; such a partition turns out to always exist). Then we define

$$\int_M \omega = \sum_{i=1}^N \int_M \rho_i \omega.$$

The idea is that we can always multiply a k-form by a smooth function ρ_i to get another k-form, and now $\rho_i\omega$ is only nonvanishing on a single chart. And since the ρ_is sum to 1 (so adding up the $\rho_i\omega$s gives us a "total" of ω), this is a sensible definition, and we can indeed prove that the integral is well-defined (it doesn't depend on the partition of unity or the charts we chose).

We next discuss other operations on our differential forms that will be useful later on:

Definition 44

For $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$, the wedge product of the differential forms ω and η is the $(k+\ell)$-form defined pointwise by

$$(\omega \wedge \eta)(x) = \omega(x) \wedge \eta(x).$$

Recall from Definition 6 that both $\omega(x)$ and $\eta(x)$ are elements of the exterior algebra $\Lambda(T_x^*M)$, so it makes sense to do this operation pointwise. For example, the product of a smooth function and a k-form can be thought of as the wedge product of a 0-form and a k-form. It turns out that if ω and η vary smoothly, so does $\omega \wedge \eta$, so this is indeed a valid $(k+\ell)$-form. This notion will also come up when we later define curvature.

Remark 45. *On a Riemannian manifold, we are able to define a volume form (Definition 62), and when we integrate smooth functions we're really multiplying those functions by this volume form. But we don't typically have this structure on an abstract manifold, which makes things more complicated.*

There is also a way to differentiate k-forms to get a $(k+1)$-forms. What's tricky is that everything has to be **not dependent on coordinates**, so every definition (for example the wedge product above) also needs to be verified to be intrinsic. We'll discuss this next time!

6 October 4, 2024

We'll start today by defining the **exterior derivative**; the naming is just meant to indicate that this object is related to the exterior algebra.

Definition 46

Let M be a (smooth) n-dimensional manifold, and let f be a k-form on M for some $0 \leq k \leq n$. The **exterior derivative** df of f is a $(k+1)$-form, defined as follows. If f is a 0-form (that is, a smooth function on M), then df is the "gradient vector" $\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$ in local coordinates (we can show that this is an intrinsic definition, so we get the same object with a different choice of charts). If f is an n-form, then $df = 0$ (because the space of $(n+1)$-forms is trivial). And for any $1 \leq k \leq n-1$, write $f = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ in local coordinates. Then we must have distributivity

$$df = \sum_{1 \leq i_1 < \cdots < i_k \leq n} d\left(f_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}\right),$$

and in order to (1) be linear, (2) be consistent with the $k = 0$ definition, (3) satisfy the **Liebniz rule** $d(f \wedge g) = (df) \wedge g + (-1)^k f \wedge (dg)$ for any k-form f and ℓ-form g, and (4) have $ddf = 0$ for all f (this is the **Poincaré lemma**), we must in fact make the definition

$$df = \sum_{1 \leq i_1 < \cdots < i_k \leq n} df_{i_1, \cdots, i_k} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Writing this last equation out in terms of the $k = 0$ definition, we have

$$\begin{aligned} df &= \sum_{1 \leq i_1 < \cdots < i_k \leq n} \left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \right) \wedge \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} \\ &= \sum_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \end{aligned}$$

Now if $i \in \{i_1, \cdots, i_k\}$, then $dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} = 0$ because two of the terms being wedged together are equal. And if i is strictly between i_{j-1} and i_j for some j, then $dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} = (-1)^{j-1} dx_{i_1} \wedge \cdots \wedge dx_{i_{j-1}} \wedge dx_i \wedge dx_{i_{j+1}} \wedge \cdots \wedge dx_{i_k}$. Thus df becomes now a sum over all sets of $(k+1)$ distinct indices, and if we think about all the ways we can get any such set of indices to come out of this sum we will find that

$$df = \sum_{1 \leq i_1 < \cdots < i_{k+1} \leq n} \left(\sum_{j=1}^{k+1} (-1)^{j+1} \frac{\partial f_{i_1, \cdots, i_{j-1}, i_{j+1}, \cdots, i_{k+1}}}{\partial x_{i_j}} \right) dx_{i_1} \wedge \cdots \wedge dx_{i_{k+1}},$$

where $f_{i_1, \cdots, i_{j-1}, i_{j+1}, \cdots, i_{k+1}}$ omits the jth index i_j so that it is indexed by k total indices. And again, we can show that (through a messy calculation) this definition is also intrinsic.

Example 47

Let's consider the calculation when $k = 1$, meaning that $f = \sum_{i=1}^{n} f_i dx_i$ is a 1-form.

Then unpacking the definition, we have

$$df = \sum_{1 \leq i < j \leq n} \left(\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) dx_i \wedge dx_j,$$

because in the parenthetical term we first omit the first index and then the second index, and in the latter case we get a $(-1)^{2-1}$ factor. And we can now verify the Poincaré lemma for 0-forms: if $f \in \Omega^0(M)$, then $df = \sum \frac{\partial f}{\partial x_i} dx_i$, and

thus plugging in $f_i = \frac{\partial f}{\partial x_i}$ yields

$$ddf = \sum_{1 \le i < j \le n} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i} \right) dx_i \wedge dx_j = 0,$$

since the parenthetical term is always zero. In fact this is the general strategy for showing that $ddf = 0$ for general k-forms – it's always due to the mixed partial derivatives canceling out.

Remark 48. *Intuitively, d is adjoint to an operator which is meant to indicate the "boundary of a manifold," and the point is that even for a manifold with boundary "the boundary of the boundary is still always empty." In the discrete setting, in place of a manifold we have $\mathbb{Z}^d$, and a k-form is like assigning a number to each **k-cell** (where a 0-cell is a point, a 1-cell is a line segment, a 2-cell is a square, and so on). Supposing that the numbers are all 0 or 1, we may then think of the set of cells valued 1 as forming a surface, and there is an operator called the "coderivative" which goes down in dimension and shows the boundary.*

The definitions for $\mathfrak{g}$-valued forms is basically analogous to the real-valued case:

Definition 49

Note: this definition is elaborated on in the next lecture for more clarification; see Definition 55. Let $\mathfrak{g}$ be a matrix Lie algebra, and suppose $A = \sum_{1 \le i_1 < \cdots < i_k \le n} A_{i_1, \cdots, i_k} dx_1 \wedge \cdots \wedge \cdots dx_{i_k}$ is the local representation of a $\mathfrak{g}$-valued k-form. Similarly, suppose we have a $\mathfrak{g}$-valued ℓ-form $B = \sum_{1 \le j_1 < \cdots < j_\ell \le n} B_{j_1, \cdots j_\ell} dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}$. Then the **wedge product** of A and B may be defined locally to be the $(k+\ell)$-form

$$A \wedge B = \sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} A_{i_1, \cdots, i_k} B_{j_1, \cdots j_\ell} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

Definition 50

Again letting A be a $\mathfrak{g}$-valued k-form with the notation as above, we may define the **exterior derivative** of A via

$$dA = \sum_{1 \le i_1 < \cdots < i_k \le n} \left(\sum_{i=1}^{n} \frac{\partial A_{i_1, \cdots, i_{j-1}, i_{j+1}, \cdots, i_{k+1}}}{\partial x_{i_j}} (-1)^{j-1} \right) dx_{i_1} \wedge \cdots \wedge dx_{i_{k+1}}.$$

We'll now define a certain kind of fiber bundle which will be a central object in the remainder of this course:

Definition 51

A **smooth principal bundle** is a smooth fiber bundle (P, M, π, G) where G is a Lie group which has a **smooth right action** on P (that is, a smooth map $P \times G \to P$) satisfying the following properties:

1. it **preserves each fiber**, so that if $p \in \pi^{-1}(\{x\})$ for some $x \in M$, then $pg \in \pi^{-1}(\{x\})$ for all $g \in G$,
2. it is **free,** meaning that for any non-identity element g we have $pg \neq p$ for all $p \in P$,
3. it is **transitive on each fiber**, meaning that if $p, q \in \pi^{-1}(\{x\})$ for some x, then there is some $g \in G$ with $q = pg$.

This is basically saying that "each fiber is acting like the group itself," and we will be exclusively interested in the case of **trivial principal bundles** $P = M \times G$, where $\pi(x, g) = x$ for all $x \in M, g \in G$, and the local trivialization is

given by the single chart (U, ϕ) where $U = M$ and $\phi : M \times G \to M \times G$ is the identity map. The group action on this trivial bundle is then simply

$$(x, g)h = (x, gh).$$

The idea is that we want to define **connections** on principal bundles, and this definition can be quite complicated for nontrivial bundles (but we don't need to do so for our purposes) and thus we'll only do it for the trivial case:

Definition 52

Let $P = M \times G$ be a trivial principal bundle, and let $\mathfrak{g}$ be the Lie algebra of G. A **principal G-connection on M** is a $\mathfrak{g}$-valued 1-form on M; we also call these **gauge fields**.

The point of a connection is to "perform parallel transport" to connect one fiber to another, taking a Lie group element on a fiber to one on another fiber. (This will become more relevant when discussing **Wilson loop observables** later on.) So for a trivial bundle this transport is relatively simple, but for a nontrivial bundle we have to think about the intrinsic differences between the fibers, and a principal connection is no longer a $\mathfrak{g}$-valued 1-form. (Somehow this parallel transport map can be thought of as a "Wick-ordered integral along a curve," where we make sense of the quantity $e^{\int f}$ for a matrix-valued f. But that's something we'll discuss in greater generality later.)

Definition 53

Let A be a principal G-connection on M. The **curvature** of A is the $\mathfrak{g}$-valued 2-form

$$F = dA + A \wedge A.$$

This will really be the object of interest; the integral of the "curvature form squared" will turn out to be the Yang-Mills action, but we'll need some more time to really make that precise! Intuitively, curvature measures how different of a value we get when parallel transporting a group element and returning back to where we started.

7 October 7, 2024

We'll start today by elaborating more on some of the constructions from last week. For any (smooth) manifold M and any finite-dimensional vector space V over $\mathbb{R}$, we can define the **V-valued k-form** on M to be a smooth section of the vector bundle $(M \times V) \otimes \Lambda^k T^*M$. So a typical V-valued k-form is given by

$$\omega(x) = \sum_{i=1}^{m} v_i(x) \otimes \omega_i(x),$$

where $v_i(x)$ are vectors in the vector space, and $\omega_i(x)$ are elements of the kth exterior power of the tangent space. But we can get rid of the dependence of the v_is on x by letting $v_1, \cdots, v_m$ be a (fixed) basis of V and thus write

$$\omega(x) = \sum_{i=1}^{m} v_i \otimes \omega_i(x)$$

with $\omega_1, \cdots, \omega_m$ real-valued k-forms. From there, we define the exterior derivative as follows: we know how to define the exterior derivative of a real-valued k-form, so since the v_is are fixed we may define

$$d\omega = \sum_{i=1}^{m} v_i d\omega_i.$$

And we can indeed still verify that eveerything here is intrinsic, so it doesn't depend on the choice of coordinates.

Example 54

This in particular covers the case of $\mathfrak{g}$-valued forms – we let $V = M(m, \mathbb{C})$ be the set of all $m \times m$ complex matrices, which is a $2m^2$-dimensional vector space over $\mathbb{R}$. The basis elements are then the matrices with a single 1 or a single i entry and zeros everywhere else. So indeed our exterior derivative essentially "operates coordinate-wise."

If we now further assume that we have a product structure for V, so that V is an algebra over $\mathbb{R}$, then we may expand out any two differential forms

$$\omega = \sum_{i=1}^{m} v_i \omega_i, \quad \eta = \sum_{i=1}^{m} v_i \eta_i$$

(note that ω can be a k-form and η an ℓ-form), and then we define

$$\omega \wedge \eta = \sum_{i,j=1}^{m} v_i v_j \omega_i \wedge \omega_j.$$

So with matrices under matrix multiplication, this gives us one definition, but if we consider matrices under the Lie product, that gives us another definition:

Definition 55

Returning to Definition 49, first let the set of all matrices $M(m, \mathbb{C})$ be an algebra under ordinary matrix multiplication. If A, B are two $M(m, \mathbb{C})$-valued differential forms, then writing out (notice that now the coefficients vary with x but the dx_is and dx_js don't)

$$A = \sum_{1 \le i_1 < \cdots < i_k \le n} a_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}, \quad B = \sum_{1 \le j_1 < \cdots < j_\ell \le n} B_{j_1, \cdots j_\ell} dx_{j_1} \wedge \cdots \wedge dx_{j_\ell},$$

then indeed as before we get the wedge product

$$A \wedge B = \sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} A_{i_1, \cdots, i_k} B_{j_1, \cdots j_\ell} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

However, $M(m, \mathbb{C})$ is also an algebra under the Lie product $[A, B] = AB - BA$, and under that definition it will give us a **different** wedge product which is written in either of the two notations

$$[A \wedge B] = [A, B] = \sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} [A_{i_1, \cdots, i_k}, B_{j_1, \cdots j_\ell}] dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

Lemma 56

Let ω and η be $M(m, \mathbb{C})$-valued k-forms and ℓ-forms, respectively. With the definition above, we have

$$[\omega \wedge \eta] = \omega \wedge \eta - (-1)^{k\ell} \eta \wedge \omega.$$

In particular, we see that $[\omega \wedge \omega] = 0$ if k is even, and $[\omega \wedge \omega] = 2\omega \wedge \omega$ if k is odd (and that's why in the Yang-Mills theories we often have various factors of 2; for example the curvature form in Definition 53 may be written also as $dA + \frac{1}{2}[A \wedge A] = dA + \frac{1}{2}[A, A]$. This can become even more confusing, because if $\mathfrak{g}$ is a matrix Lie algebra, then we

can always write $A = \sum_{i=1}^n A_i dx_i$ (recall that A is a $\mathfrak{g}$-valued 1-form), and we can evaluate

$$dA = \sum_{1 \le i < j \le n} \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) dx_i \wedge dx_j.$$

We can also evaluate explicitly that

$$A \wedge A = \sum_{i,j=1}^n A_i A_j dx_i \wedge dx_j,$$

but now $dx_i \wedge dx_j = -dx_j \wedge dx_i$ and thus we may write this as the **ordered** double sum

$$A \wedge A = \sum_{1 \le i < j \le n} (A_i A_j - A_j A_i) dx_i \wedge dx_j.$$

Thus the curvature form is actually

$$F = \sum_{1 \le i < j \le n} \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} + [A_i, A_j] \right) dx_i \wedge dx_j,$$

and there is now no $\frac{1}{2}$ appearing here even though this expression is in reality equal to $dA + \frac{1}{2}[A, A]$. So it's good to be aware that this comes from the fact that we're only summing over $i < j$.

Proof of Lemma 56. Write $\omega = \sum_{1 \le i_1 < \cdots < i_k \le n} \omega_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, and similarly write $\eta = \sum_{1 \le j_1 < \cdots < j_\ell \le n} \eta_{j_1, \cdots j_\ell} dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}$. Then taking the Lie algebra wedge product yields

$$[\omega \wedge \eta] = \sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} [\omega_{i_1, \cdots, i_k}, \eta_{j_1, \cdots j_\ell}] dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell},$$

and now we break this up by the definition of the Lie bracket as

$$\sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} \omega_{i_1, \cdots, i_k} \eta_{j_1, \cdots j_\ell} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell} - \sum_{\substack{1 \le i_1 < \cdots < i_k \le n, \\ 1 \le j_1 < \cdots < j_\ell \le n}} \eta_{j_1, \cdots j_\ell} \omega_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

Now the first term is exactly $\omega \wedge \eta$, while the second term is exactly $(-1)^{k\ell} \eta \wedge \omega$ because we need $k\ell$ swaps to get all of the dx_js in front of all of the dx_is. □

Adding further to potential confusion, if we now let V be a vector space and consider points $x \in \Lambda^k(V)$ and $y \in \Omega^\ell(V)$, we know that $x \wedge y = (-1)^{k\ell} y \wedge x$ (by the same reasoning as in the proof we just did, because we can write x and y in terms of basis elements). So more generally if we have real-valued forms $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$, we have $\boxed{\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega}$. But in contrast, if $\omega \in \Omega^k(M, M(m, \mathbb{C}))$ and $\eta \in \Omega^\ell(M, M(m, \mathbb{C})$ are now **matrix-valued** forms, the wedge product with respect to the **Lie bracket** will instead be $\boxed{[\omega \wedge \eta] = -(-1)^{k\ell} [\eta \wedge \omega]}$, since if $T_1, \cdots, T_{2m^2}$ is a basis of $M(m, \mathbb{C})$, we may write $\omega = \sum_i T_i \omega_i$ and $\eta = \sum_j T_j \eta_j$, and then

$$[\omega \wedge \eta] = \sum_{i,j} [T_i, T_j] (\omega_i \wedge \eta_j) = -(-1)^{k\ell} \sum_{i,j} [T_j, T_i] (\eta_j \wedge \omega_i)$$

because the Lie bracket is now anticommutative. So it's good to be careful about all of these signs depending on the context and the definitions!

Lemma 57

Let $\mathfrak{g}$ be a Lie algebra and M be a smooth manifold, and let $\omega \in \Omega^k(M, \mathfrak{g})$ be a $\mathfrak{g}$-valued k-form. Then

$$[\omega \wedge [\omega \wedge \omega]] = 0.$$

Note that this relation is not true if we're not taking the Lie algebra wedge product – in fact, this triple wedge product comes up in the Chern-Simons theory.

Proof. This is a consequence of the Jacobi identity – if $T_1, \cdots, T_m$ is a (vector space) basis of $\mathfrak{g}$, which we also call a **set of generators**, we may write $\omega = \sum_{a=1}^m T_a \omega_a$ for real-valued k-forms ω_a. (Notice that we're choosing a basis of $\mathfrak{g}$ rather than of the real-valued forms here, because it will be more convenient for our calculations.) We then expand out

$$[\omega \wedge [\omega \wedge \omega]] = \sum_{a,b,c=1}^m [T_a, [T_b, T_c]](\omega_a \wedge (\omega_b \wedge \omega_c)).$$

We know that $[T_a, [T_b, T_c]] + [T_b, [T_c, T_a]] + [T_c, [T_a, T_b]] = 0$, so we can instead write this expression above as

$$-\sum_{a,b,c=1}^m [T_b, [T_c, T_a]](\omega_a \wedge (\omega_b \wedge \omega_c)) - \sum_{a,b,c=1}^m [T_c, [T_a, T_b]](\omega_a \wedge (\omega_b \wedge \omega_c)).$$

We can use associativity of the wedge product now (this is clear because we're working with real forms, so the operation starts with a product which is associative and then quotienting) so that we can swap terms around to match up the indices. In the first sum we swap ω_a and ω_b, then ω_a and ω_c, and each swap gives a factor of $(-1)^{k^2}$ so overall the sign doesn't change. Similarly, in the second sum we swap ω_b and ω_c, then ω_a and ω_c. This yields

$$-\sum_{a,b,c=1}^m [T_b, [T_c, T_a]](\omega_b \wedge (\omega_c \wedge \omega_a)) - \sum_{a,b,c=1}^m [T_c, [T_a, T_b]](\omega_c \wedge (\omega_a \wedge \omega_b)),$$

and now each of the two sums on the right-hand side is the same as the original $[\omega \wedge [\omega \wedge \omega]]$. Thus $3[\omega \wedge [\omega \wedge \omega]] = 0$, proving the claim. □

We're now ready to discuss Stokes' theorem for differential forms; we'll do it just for manifolds without boundary here, though there is a general theory for manifolds with boundary. We'll elaborate on this next time, but the idea is as follows: let M be a (smooth) n-dimensional manifold without boundary, and let ω be a real-valued $(n-1)$-form on M. Under some mild conditions, we then have that $\int_M d\omega = 0$.

8 October 9, 2024

We'll begin today with Stokes' theorem. In various textbooks, it can be a bit hard to find the following general form (since often the statement is restricted to compact manifolds):

Theorem 58

Let M be a smooth n-dimensional manifold (without boundary), and let ω be a real-valued $(n-1)$-form on M whose derivatives are integrable. Suppose that $\{(U_i, \phi_i) : 1 \leq i \leq N\}$ is a collection of charts along with a partition of unity $\rho_1, \cdots, \rho_N$ (so pointwise they add to 1, and each ρ_i vanishes outside U_i), such that

1. ω vanishes outside $U_1 \cup \cdots \cup U_N$, and
2. for all i and for all sequences of points $x_k \to \partial U_i$ (that is, for all $K \subseteq U_i$, x_k is eventually outside of K), $\rho_i(x_k)\omega(x_k) \to 0$.

Then $\int_M d\omega = 0$.

Note that these two conditions are automatically satisfied for all ω if M is compact, since we may take any finite atlas and pick ρ_i so that ρ_i vanishes outside some compact set $K_i \subseteq U_i$ (thus as x_k approaches the boundary it'll exit K_i eventually). But on the other hand, if $M = \mathbb{R}^n$ we can take our atlas to just be the single chart $(U, \phi) = (\mathbb{R}^n, \mathrm{id})$. Then condition (1) is vacuously satisfied, and condition (2) holds as long as $\lim_{|x|\to\infty} \omega(x) = 0$.

Proof sketch. In local coordinates, we may write the $(n-1)$-form ω as

$$\omega = \sum_{1 \leq i_1 < \cdots < i_{n-1} \leq n} \omega_{i_1, \cdots, i_{n-1}} dx_{i_1} \wedge \cdots \wedge dx_{i_{n-1}}.$$

There are always $(n-1)$ indices in this sequence, so we can equivalently write this in terms of the index being left out:

$$\omega = \sum_{i=1}^{n} f_i dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots dx_n.$$

Then taking the exterior derivative yields

$$d\omega = \sum_{i=1}^{n} (-1)^i \frac{\partial f_i}{\partial x_i} dx_1 \wedge \cdots \wedge dx_n,$$

and we must show that this integrates to zero. The rest can be worked out with a partition of unity, but for simplicity we'll just show how it works if ω vanishes outside U for some single chart (U, ϕ) and further vanishes near ∂U (this is condition (2) above). The general case then works if we replace ω with $\rho_i \omega$.

Writing out the integral, we then have

$$\int_U d\omega = \sum_{i=1}^{n} (-1)^i \int_U \frac{\partial f}{\partial x_i} dx_1 \wedge \cdots \wedge dx_n.$$

We will show that each inner integral is zero. Indeed, by definition of the integral of an n-form on a chart, we have

$$\int_U \frac{\partial f}{\partial x_i} dx_1 \wedge \cdots \wedge dx_n = \int_{\phi(U)} \frac{\partial}{\partial y_i} (f \circ \phi^{-1})(y) dy,$$

where $\phi(U)$ is now a subset on $\mathbb{R}^n$. Now if i is fixed, and $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$ are all fixed, we can consider the integral only over x_i, specifically

$$\int_{V_i(y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)} \frac{\partial}{\partial y_i} (f \circ \phi^{-1})(y) dy_i,$$

where $V_i = \{y_i : (y_1, \cdots, y_n) \in \phi(U)\}$. Here V_i is an open set (because U is open), so it is a countable union of disjoint open intervals. Consider such a decomposition by connected components. Now for any **bounded** interval (a, b), we have $\int_a^b u'(y) dy = u(b) - u(a)$ by the fundamental theorem of calculus, and if $u = f \circ \phi^{-1}$ with all other coordinates

fixed, $u(b) = u(a) = 0$ and thus the contribution is zero. On the other hand, if we have a **unbounded** interval (a, ∞) or $(-\infty, a)$ we need condition (2) to ensure that the function again goes to zero.

Thus then if we integrate over all of the other variables, we'll still have zero, proving the theorem. (Note here that we do need Fubini's theorem for this to hold so that the coordinate-by-coordinate integration is allowed and the result doesn't depend on the order.) □

Next, we'll discuss how to equip our manifolds with a metric:

Definition 59

Let M be a smooth manifold, and let S be the set of all triples (x, u, v) with $x \in M$ and $u, v \in T_xM$. We can make S a manifold in the same way that we made TM a manifold (using a pair of tangent vectors instead of just a single one). A **Riemannian metric** on M is a smooth map $S \to \mathbb{R}$ such that for any $x \in M$, the map $g(x, \cdot, \cdot) : T_xM \times T_xM \to \mathbb{R}$ is a positive definite inner product on the vector space T_xM, and a **Riemannian manifold** is a manifold equipped with such a metric.

The fact that the map g is smooth tells us that the "inner product on the tangent spaces varies smoothly." And we can write this in a local coordinate system as well: given a chart (U, ϕ), we can represent g as a map from U to the set of $n \times n$ positive definite matrices, which we may write as a matrix of functions

$$g(x) = (g_{ij}(x))_{i,j=1}^{n}$$

for $g_{ij}(x) = g(x, \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j})$. Of course, these functions g_{ij} depend on our choice of coordinates.

Example 60

The tangent space at each point in $M = \mathbb{R}^n$ is then another copy of $\mathbb{R}^n$, so using the standard inner product makes $\mathbb{R}^n$ into a Riemannian manifold (and then $g(x)$ is just the identity matrix at each point x).

On the other hand, if we consider the manifold S^2 (the 2-sphere), we can embed it in $\mathbb{R}^3$ and think of the tangent space as the actual tangent plane at that point in $\mathbb{R}^3$. Then we can take the inner product by viewing the tangent vectors as elements of $\mathbb{R}^3$; this is called the **round metric**.

Definition 61

Let (M, g) be a Riemannian manifold. The **volume form** on M is a particular n-form given by

$$\omega = \sqrt{\det(g(x))}\, dx_1 \wedge \cdots \wedge dx_n$$

in local coordinates, where $g(x)$ is thought of as the matrix of values $g_{ij}(x)$. We sometimes write this form as $d\text{vol}_g$.

We can indeed verify that this is an intrinsic definition, and this now allows us to do general integration:

Definition 62

Let (M, g) be a Riemannian manifold, and let $f \in C^\infty(M)$ be a smooth function (0-form). Then letting ω denote the volume form on M, we define

$$\int_M f = \int_M f\omega.$$

All of our work will be done on Riemannian manifolds, because we do have a lot of tools that will be useful going forward. One of them is the following construction:

Definition 63

Let (M, g) be a Riemannian manifold (assume dimension n unless otherwise stated), and let $\eta \in \Omega^k(M)$ be a k-form. The **Hodge dual** of η is the $(n-k)$-form $*\eta$, defined as follows. (We call $*$ the **Hodge star operator**.) For any $i_1, \cdots, i_n \in \{1, \cdots, n\}$, we define the **Levi-Civita symbol** $\varepsilon_{i_1, \cdots, i_n}$ to be 0 if there is a repeated coordinate and $\operatorname{sgn}(i_1, \cdots, i_n)$ otherwise. Then we can write the local representation of η as

$$\begin{aligned}\eta &= \sum_{1 \le i_1 \le \cdots \le i_k \le n} \eta_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k} \\ &= \frac{1}{k!} \sum_{1 \le i_1, \cdots, i_k \le n} \eta_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k},\end{aligned}$$

where we define $\eta_{i_{\pi(1)}, \cdots, i_{\pi(k)}} = \operatorname{sgn}(\pi) \eta_{i_1, \cdots, i_k}$ and $\eta_{i_1, \cdots, i_k} = 0$ if $i_1, \cdots, i_k$ are not distinct. Then we may define

$$(*\eta)_{j_{k+1}, \cdots j_n} = \frac{\sqrt{\det(g)}}{k!} \sum_{\substack{1 \le i_1, \cdots, i_k \le n \\ 1 \le j_1, \cdots j_k \le n}} \varepsilon_{j_1, \cdots j_k} g^{i_1 j_1} g^{i_k j_k} \eta_{i_1, \cdots, i_k},$$

where (g^{ij}) is the inverse of the matrix g. (See next lecture for some additional details!)

9 October 11, 2024

We introduced the Hodge star operator, a useful computational tool, last time – let's now elaborate on the definition. For an n-dimensional manifold (M, g) and a real **or matrix**-valued k-form η, we may write $\eta = \sum_{1 \le i_1 < \cdots < i_k \le n} \eta_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$. We may then extend this sum to all sets of indices $(i_1, \cdots, i_k) \in [n]^k$ by defining $\eta_{i_{\pi(1)}, \cdots, i_{\pi(k)}} = \operatorname{sgn}(\pi) \eta_{i_1, \cdots, i_k}$ for any permutation π (and 0 if there is a repeated index), which allows us to write our form as

$$\eta = \frac{1}{k!} \sum_{1 \le i_1, \cdots, i_k \le n} \eta_{i_1, \cdots, i_k} dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

We then define the dual of η similarly as

$$*\eta = \frac{1}{(n-k)!} \sum_{1 \le j_{k+1}, \cdots j_n \le n} *\eta_{j_{k+1}, \cdots j_n} dx_{j_{k+1}} \wedge \cdots \wedge dx_{j_n},$$

where these coefficients take the form

$$(*\eta)_{j_{k+1}, \cdots j_n} = \frac{\sqrt{\det(g)}}{k!} \sum_{\substack{1 \le i_1, \cdots, i_k \le n \\ 1 \le j_1, \cdots j_k \le n}} \varepsilon_{j_1, \cdots j_k} g^{i_1 j_1} g^{i_k j_k} \eta_{i_1, \cdots, i_k}$$

for g^{ij} the entries of the inverse matrix g^{-1}. So in other words, we choose $j_{k+1}, \cdots, j_n$ to be distinct, and this definition also ensures that permuting the indices for $*\eta_{j_{k+1}, \cdots j_n}$ behaves similarly as $\eta_{i_1, \cdots, i_k}$.

Remark 64. *We'll see this later on, but in the lattice case a k-form corresponds to studying the k-cells of the lattice (so points, lines, squares, cubes, and so on) – it's a real value attached to each k-cell. In this framework, the Hodge star operator takes a k-cell, considers its dual (for example in three dimensions, the dual of a point is the cube at its center, and the dual of a line segment is the square intersecting it in the middle, and so on) and applies that function*

to the dual cell.

We'll need the following two facts about this operation:

Lemma 65
Let η be a k-form. Then $**\eta = (-1)^{k(n-k)}\eta$.

In particular, we'll need to use this to identify the critical points of the Yang-Mills action.

Proof sketch. Let $\gamma = *\eta$ and $\theta = *\gamma$. For any indices $\ell_1 < \cdots < \ell_k$, we have (by the formula)

$$\theta_{\ell_1,\cdots,\ell_k} = \frac{\sqrt{\det(g)}}{(n-k)!} \sum_{\substack{1\le j_{k+1},\cdots,j_n\le n \\ 1\le \ell_{k+1},\cdots,\ell_n\le n}} \varepsilon_{\ell_{k+1},\cdots,\ell_n,\ell_1,\cdots,\ell_k} g^{j_{k+1}\ell_{k+1}}\cdots g^{j_n\ell_n}\gamma_{j_{k+1},\cdots j_n}.$$

Now plugging in the formula for $\gamma_{j_{k+1},\cdots j_n}$, the factor $(-1)^{k(n-k)}$ comes from moving the indices in the blue part, since each of k indices needs to move in front of $(n-k)$ other terms. And then if we substitute everything back in we will see that the $\eta_{\ell_1,\cdots,\ell_k}$ terms will come up, exactly recovering the original k-form. □

Lemma 66
Let η, γ be real or matrix-valued k-forms on a manifold M. Then letting ω denote the volume form of M, we have

$$\eta \wedge *\gamma = \langle \eta, \gamma\rangle \omega$$

for the smooth function $\langle \eta, \gamma\rangle$ (in $C^\infty(M)$ or $C^\infty(M, M(m,\mathbb{C}))$) given by

$$\langle \eta, \gamma\rangle = \sum_{1\le i_1<\cdots<i_k\le n} \eta_{i_1,\cdots,i_k}\gamma^{i_1,\cdots,i_k}, \text{ where } \gamma^{i_1,\cdots,i_k} = \sum_{1\le j_1,\cdots j_k\le n} g^{i_1 j_1}\cdots g^{i_k j_k}\gamma_{j_1,\cdots j_k}$$

(note that in the first sum above we have increasing indices, while in the second we sum over all indices).

This convention for "raising indices" is the usual convention for Riemannian manifolds, and $\langle \eta, \gamma\rangle$ is somehow an inner product after raising indices. (For example, if η and γ are one-forms, then we may write this in matrix form as $\eta^T g^{-1}\gamma$.)

Proof. The left-hand side is given in coordinates by

$$\begin{aligned}\eta \wedge *\gamma &= \frac{1}{k!(n-k)!} \sum_{\substack{1\le i_1,\cdots,i_k\le n \\ 1\le j_1,\cdots j_k\le n}} \eta_{i_1,\cdots,i_k}(*\gamma)_{j_{k+1},\cdots j_n} dx_{i_1}\wedge\cdots\wedge dx_{i_k}\wedge dx_{j_{k+1}}\wedge\cdots\wedge dx_{j_n} \\ &= \frac{\sqrt{\det g}}{(k!)^2(n-k)!} \sum_{\substack{1\le i_1,\cdots,i_k\le n \\ 1\le j_1,\cdots j_k\le n}} \sum_{1\le i'_1,\cdots,i'_k\le n} g^{i'_1 j_1}\cdots g^{i'_k j_k}\eta_{i_1,\cdots,i_k}\gamma_{i'_1\cdots i'_k}\varepsilon_{j_1,\cdots j_n} dx_{i_1}\wedge\cdots\wedge dx_{i_k}\wedge dx_{j_{k+1}}\wedge\cdots\wedge dx_{j_n}.\end{aligned}$$

We may now rewrite the wedge product part $dx_{i_1}\wedge\cdots\wedge dx_{i_k}\wedge dx_{j_{k+1}}\wedge\cdots\wedge dx_{j_n}$ as $\varepsilon_{i_1,\cdots,i_k\, j_{k+1},\cdots j_n} dx_1\wedge\cdots\wedge dx_n$, since either there is a repeat in the indices or we just have all of $x_1,\cdots,x_n$ in some order. Thus we may write the above expression as (now summing over **increasing** indices in our sums, getting rid of the factorial terms)

$$\left(\sum_{1\le i_1<\cdots<i_k\le n}\sum_{1\le i'_1\le\cdots\le i'_k\le n}\sum_{\pi\in S_k} g^{i'_1 i_{\pi(1)}}\cdots g^{i'_k i_{\pi(k)}}\operatorname{sgn}(\pi)\eta_{i_1,\cdots,i_k}\gamma_{i'_1\cdots i'_k}\right)\omega,$$

where recall the definition $\omega = \sqrt{\det g}\, dx_1 \wedge \cdots \wedge dx_n$ of the volume form. But now for each $\pi \in S_k$, we can write out

$$\begin{aligned}\gamma^{i_{\pi(1)},\cdots,i_{\pi(k)}} &= \sum_{1\le j_1,\cdots j_k\le n} g^{i_{\pi(1)}j_1}\cdots g^{i_{\pi(k)}j_k}\gamma_{j_1,\cdots j_k}\\ &= \sum_{1\le j_1,\cdots j_k\le n} g^{i_1 j_{\pi^{-1}(1)}}\cdots g^{i_k j_{\pi^{-1}(k)}}\gamma_{\pi^{-1}(j_1),\cdots,\pi^{-1}(j_k)}\\ &= \operatorname{sgn}(\pi^{-1})\gamma_{i_1,\cdots,i_k},\end{aligned}$$

and substituting this back in yields the result by the definition of our "inner product." □

With this, we finally have everything we need to define the Yang-Mills action:

Definition 67

Let $P = M \times G$ be a trivial principal bundle where M is a Riemannian manifold and G is a matrix Lie group contained in $U(m)$ for some m (so that the Lie algebra $\mathfrak{g}$ consists of skew-Hermitian matrices). Recall that a gauge field A on M is a $\mathfrak{g}$-valued 1-form, meaning that it may be written in local coordinates as $A = \sum_{i=1}^n A_i dx_i$. Let F be the curvature form of A, which we may recall is given by $dA + A \wedge A = dA + \frac{1}{2}[A \wedge A]$ – write this as $F = \sum_{1\le i<j\le n} F_{ij} dx_i \wedge dx_j$. Then by the above lemma, we have $F \wedge *F = \langle F, F\rangle \omega$, where $\langle F, F\rangle$ is the smooth matrix-valued function given by

$$\langle F, F\rangle = \sum_{1\le i<j\le n} F_{ij}F^{ij}, \text{ where } F^{ij} = \sum_{1\le k,\ell\le n} g^{ik}g^{j\ell}F_{k\ell}$$

(remember that F_{ij} and F^{ij} are skew-Hermitian matrices here). Then $\operatorname{Tr}(F \wedge *F) = \operatorname{Tr}(\langle F, F\rangle)\omega$ is a real-valued n-form, where we take the trace of the matrix at each point and multiply it by the volume form (the trace is intrinsic, so this is well-defined). The **Yang-Mills action** is then given by

$$S(A) = -\int_M \operatorname{Tr}(F \wedge *F).$$

At each point, $\operatorname{Tr}\langle F, F\rangle$ measures the "size" of F, so this is telling us something about the curvature of the connection A.

Example 68

Take $M = \mathbb{R}^n$. We can then let g be the identity matrix everywhere, so $F^{ij} = \sum_{k,\ell} g^{ik}g^{j\ell}F_{k\ell}$ is just F_{ij} and the raising and lowering of indices does not change the entries.

So then we can write out

$$\langle F, F\rangle = \sum_{1\le i<j\le n} F_{ij}^2,$$

and recall that $F = dA + A \wedge A$ and that we have already explicitly computed $F_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} + [A_i, A_j]$. Since our A_is are skew-Hermitian, so are their partial derivatives, and furthermore

$$[A_i, A_j]^* = (A_iA_j - A_jA_i)^* = A_j^*A_i^* - A_i^*A_j^* = A_jA_i - A_iA_j = [A_j, A_i] = -[A_i, A_j]$$

and thus the commutator is also skew-Hermitian. Thus F_{ij} is skew-Hermitian and thus all eigenvalues are purely imaginary; this means the trace of F_{ij}^2 will be **negative**. So that explains the negative sign in the definition of the Yang-Mills action, and note that for skew-Hermitian matrices the trace is basically the sum of the squared norms of

the entries. So indeed we're somehow measuring the size of the matrix itself here, and it's only positive if the matrix is zero.

Yang-Mills connections are then the minima or critical points of this action – we'll discuss the start of that a bit. In mathematics, we've studied extensively the, but if we want to put a probability measure with density e^{-cS} on connections, that's something more difficult and we don't know how to do that beyond a few examples. The point of **Euclidean or quantum Yang-Mills theories** is to try to work towards that!

In our next few lectures, we'll understand how to get to the Yang-Mills equation via variational methods (perturbing our connection by a little bit and setting directional derivatives equal to zero). We'll then see how the four Maxwell's equations governing electromagnetism are actually the same as the critical-point equations for $U(1)$ Yang-Mills. And the weak and strong force turn out to work very similarly but instead with the matrix groups $SU(2)$ and $SU(3)$!

10 October 14, 2024

We defined the Yang-Mills action last week – for a $\mathfrak{g}$-valued 1-form A on a Riemannian manifold (M, g), the action is given by $S(A) = -\int_M \mathrm{Tr}(F \wedge *F)$ for the curvature form $dA + \frac{1}{2}[A, A]$. In local coordinates, we explicitly wrote out (if $A = \sum_i A_i dx_i$ for matrices A_i)

$$F = \sum_{1 \le i < j \le n} F_{ij} dx_i \wedge dx_j, \quad F_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} + [A_i, A_j].$$

Writing out the Yang-Mills action in coordinates is a bit messy, but we can do it and we find that on a single chart (U, ϕ) (and then we need to patch the integral together over all of M)

$$\int_U \mathrm{Tr}(F \wedge *F) = \frac{1}{2} \int_{\phi(U)} \mathrm{Tr} \left(\sum_{1 \le i,j,k,l \le n} g^{ik} g^{j\ell} F_{ij}(\phi^{-1}(x)) F_{k\ell}(\phi^{-1}(x)) \right) \sqrt{\det g(\phi^{-1}(x))} dx$$

where as usual g^{ij} is the (i, j) entry of the inverse of the matrix g. And in the special case where $M = \mathbb{R}^n$ and g is the Euclidean metric, this expression simplifies to

$$S(A) = \int_{\mathbb{R}^n} \mathrm{Tr} \left(\sum_{1 \le i < j \le n} F_{ij}(x)^2 \right) dx,$$

and thus we're taking the sum of the squares of the skew-Hermitian (if our matrix group is a subgroup of the unitary group) matrices $F_{ij}(x)$ and integrating their trace.

Today we'll think about critical points, finding connections for which the variational derivatives vanish in all directions. To do that, we need to understand what happens to S under perturbations, and the following notion (a generalization of the exterior derivative) is useful for that:

Definition 69

Let A be a $\mathfrak{g}$-valued 1-form and ω be a $\mathfrak{g}$-valued k-form (for some $0 \le k \le n-1$). The **covariant derivative** of ω with respect to A is

$$d_A \omega = d\omega + [A \wedge \omega] = d\omega + [A, \omega].$$

In particular, if $A = 0$ or A commutes with everything, then this is the usual exterior derivative.

Lemma 70 (Bianchi identity)
For any gauge field A with curvature form F, we have $d_A F = 0$.

(Two of the four Maxwell's equations for electromagnetism turn out to be special cases of this identity, and the other two come from a similar identity which we will discuss after this proof.)

Proof. Recall the Leibniz rule $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$ for any k-form ω and ℓ-form η. We claim that for any $\mathfrak{g}$-valued k-form ω and $\mathfrak{g}$-valued ℓ-form η, the Leibniz rule is still valid (now using the Lie algebra wedge product). Indeed, let $T_1, \cdots, T_m$ be generators of the Lie algebra $\mathfrak{g}$. Writing $\omega = \sum_{i=1}^m T_i \omega_i$ and $\eta = \sum T_i \eta_i$ for real-valued k-forms ω_i and ℓ-forms η_i, we then find that

$$[\omega \wedge \eta] = \sum_{1 \le i,j \le n} [T_i, T_j] \omega_i \wedge \eta_j \implies d[\omega \wedge \eta] = \sum_{1 \le i,j \le n} [T_i, T_j] d(\omega_i \wedge \eta_j),$$

because the T_is and T_js are not varying over the manifold, and then we just apply the Leibniz rule to the right-hand side. Therefore because A is a 1-form we have

$$d[A \wedge A] = [dA \wedge A] - [A \wedge dA].$$

Now recall that if ω is a $\mathfrak{g}$-valued k-form and η is a $\mathfrak{g}$-valued ℓ-form we have $[\omega \wedge \eta] = -(-1)^{k\ell}[\eta \wedge \omega]$ (the extra negative sign coming from the commutator) and thus $[dA \wedge A] = -(-1)^2 [A \wedge dA]$ and thus the right-hand side is $-2[A \wedge dA]$. This means that

$$\begin{aligned} F = dA + \frac{1}{2}[A \wedge A] \implies dF &= ddA + \frac{1}{2} d[A \wedge A] \\ &= 0 + \frac{1}{2}(-2[A \wedge dA]) \\ &= -[A \wedge dA] \end{aligned}$$

(by the same logic as above, the identity $d^2 = 0$ also holds for $\mathfrak{g}$-valued forms), and therefore

$$\begin{aligned} d_A F &= dF + [A \wedge F] \\ &= -[A \wedge dA] + [A \wedge (dA + \frac{1}{2}[A \wedge A])] \\ &= \frac{1}{2}[A \wedge [A \wedge A]], \end{aligned}$$

and this last expression is zero by Lemma 57. □

Definition 71
Let A be a gauge field, and let η be a $\mathfrak{g}$-valued k-form for $1 \le k \le n$. We define the $(k-1)$-form

$$d_A^* \eta = (-1)^{n(k-1)+1} * d * \eta + (-1)^{n(k-1)+1} * [A \wedge * \eta] = (-1)^{n(k-1)+1)} * (d * \eta + [A \wedge * \eta]).$$

In particular, we may rewrite this as $(-1)^{n(k-1)+1} * d_A * \eta$.

Notice that $*\eta$ is an $(n-k)$-form, so $d*\eta$ is an $(n-k+1)$-form, and thus $*d*\eta$ is an $n-(n-k+1) = (k-1)$-form; similarly we can calculate that $[A \wedge *\eta]$ is an $(n-k+1)$-form and thus $*[A \wedge *\eta]$ is an $n-(n-k+1) = (k-1)$-form as well. (So this expression is indeed something in $\Omega^{k-1}(M, g)$.)

When $A = 0$, the first term on the right (sometimes denoted δ) is the adjoint of the usual d operator; this is sometimes called the **coderivative operator** or **codifferential**. This idea of proving that d_A and d_A^* are "adjoint" is encoded by the following fact:

Lemma 72

Let η be a $\mathfrak{g}$-valued k-form and γ be a $\mathfrak{g}$-valued $(k+1)$-form, where $0 \leq k \leq n-1$. If the $(n-1)$-form $\eta \wedge *\gamma$ satisfies the criteria for Stokes' theorem (in particular this always holds for compact manifolds), then

$$\int_M \mathrm{Tr}(\langle d_A\eta, \gamma\rangle)\omega = \int_M \mathrm{Tr}(\langle \eta, d_A^*\gamma\rangle)\omega.$$

(The inner products both make sense because on the left we have $(k+1)$-forms and on the right we have k-forms.)

Proof. The formula for d_A^* basically drops out of the following calcluation. We have

$$\begin{aligned}\int_M \mathrm{Tr}(\langle d_A\eta, \gamma\rangle)\omega &= \int_M \mathrm{Tr}\left(d_A\eta \wedge *\gamma\right) \\ &= \int_M \mathrm{Tr}\left((d\eta + [A \wedge \eta]) \wedge *\gamma\right)\end{aligned}$$

and now by distributivity this simplifies to

$$\boxed{\int_M \mathrm{Tr}(d\eta \wedge *\gamma) + \int_M \mathrm{Tr}([A \wedge \eta] \wedge *\gamma)}.$$

Dealing first with the first term, we use Stokes' theorem and find that (we discussed it for real-valued forms, but it again similarly applies to $\mathfrak{g}$-valued forms)

$$\begin{aligned}0 &= \int_M \mathrm{Tr}(d(\eta \wedge *\gamma)) \\ &= \int_M \mathrm{Tr}(d\eta \wedge *\gamma + (-1)^k \eta \wedge d*\gamma),\end{aligned}$$

and therefore we can replace the first term with $-(-1)^k \int_M \mathrm{Tr}(\eta \wedge d*\gamma)$. And for the second term, we use that for any $\omega \in \Omega^k(M, \mathfrak{g})$ and $\eta \in \Omega^\ell(M, \mathfrak{g})$ we have $[\omega \wedge \eta] = \omega \wedge \eta - (-1)^{k\ell}\eta \wedge \omega$, so

$$[A \wedge \eta] = A \wedge \eta - (-1)^k \eta \wedge A;$$

substituting this back into the second term, we find that

$$\int_M \mathrm{Tr}([A \wedge \eta] \wedge *\gamma) = \int_M \mathrm{Tr}\left((A \wedge \eta) \wedge *\gamma - (-1)^k(\eta \wedge A) \wedge *\gamma\right),$$

and this time we want to use that $\mathrm{Tr}(\omega \wedge \eta) = (-1)^{k\ell}\mathrm{Tr}(\eta \wedge \omega)$ (no additional negative sign here because we have no brackets). So associativity of the wedge product, along with swapping A and $\eta \wedge *\gamma$ in the red expression above, yields

$$\begin{aligned}&\int_M \mathrm{Tr}\left((-1)^{n-1}\eta \wedge (*\gamma \wedge A) - (-1)^k \eta \wedge (A \wedge *\gamma)\right) \\ &= -(-1)^k \int_M \mathrm{Tr}\left(\eta \wedge (A \wedge *\gamma) - (-1)^{n-k-1}\eta \wedge (*\gamma \wedge A)\right) \\ &= -(-1)^k \int_M \mathrm{Tr}\left(\eta \wedge [A \wedge *\gamma]\right).\end{aligned}$$

Thus plugging everything back in to the boxed expression above, we find that it is equal to

$$\begin{aligned}&-(-1)^k\int_M \mathrm{Tr}\,(\eta\wedge d*\gamma)-(-1)^k\int_M \mathrm{Tr}\,(\eta\wedge[A\wedge *\gamma])\\&=\int_M \mathrm{Tr}\,(\eta\wedge(-(-1)^k d*\gamma-(-1)^k[A\wedge *\gamma]))\end{aligned}$$

and we just wish to show that the blue part is equal to $*d_A^*\gamma$. But this is just a consequence of the fact that $**\eta=(-1)^{k(n-k)}\eta$ for a k-form η, so this exactly recovers the expression for d_A^* in our definition above (since $k^2\equiv k \mod 2$, and remembering that γ is a $(k+1)$-form rather than a k-form). □

Definition 73

Let $A\in\Omega^1(M,g)$ and $B\in\Omega^1(M,g)$ be two gauge fields. Then the **directional derivative** of S at A in the direction B is

$$\nabla_B S(A)\lim_{t\to 0}\frac{S(a+tB)-S(A)}{t},$$

provided the limit exists (for example this always exists if B is compactly supported). We say that A is a **Yang-Mills connection** if its directional derivatives are zero for all compactly supported B.

Theorem 74 (Yang-Mills equation)

The above directional derivative has an explicit formula for compactly supported B:

$$\nabla_B S(A)=-2\int_M \mathrm{Tr}(\langle B,d_A^*F_A\rangle)\omega,$$

where F_A is the curvature form of A. Thus A is a Yang-Mills connection if and only if $d_A^*F_A=0$.

11 October 16, 2024

We'll discuss the Yang-Mills equation today – we're back to the setting where M is a (not necessarily compact) n-dimensional manifold. We want to understand what happens when we perturb our gauge field A by a little bit; specifically, for some compactly supported gauge field B, we want to study $S(A+tB)-S(A)$. This difference is well-defined even if A does not have finite Yang-Mills action, since we can define

$$S(A+tB)-S(A)=-\int_M\left(\mathrm{Tr}(F_{A+tB}\wedge *F_{A+tB})-\mathrm{Tr}(F_A\wedge *F_A)\right),$$

and even if each trace does not integrate to something finite on M, the difference of the traces is zero outside a compact set, so this will indeed make sense.

Furthermore, we can show that $\lim_{t\to 0}\frac{S(A+tB)-S(A)}{t}$ actually exists; we claimed last lecture in Theorem 74 that

$$\lim_{t\to 0}\frac{S(A+tB)-S(A)}{t}=-2\int_M \mathrm{Tr}(\langle B,d_A^*F_A\rangle).$$

We'll prove this now:

Proof of Theorem 74. We have pointwise that

$$\begin{aligned} F_{A+tB} &= d(A+tB) + \frac{1}{2}[(A+tB)\wedge(A+tB)] \\ &= dA + tdB + \frac{1}{2}[A\wedge A] + \frac{1}{2}t[A\wedge B] + \frac{1}{2}t[B\wedge A] + \frac{1}{2}t^2[B\wedge B] \end{aligned}$$

by linearity. Now $[\omega\wedge\eta] = -(-1)^{k\ell}[\eta\wedge\omega]$ for any (Lie algebra-valued) k,ℓ-forms η,ω, so in particular $[A\wedge B] = -(-1)^1[B\wedge A]$ and thus the blue terms simplify to $[A\wedge B]$. Thus

$$F_{A+tB} = F_A + tdB + \frac{1}{2}[A\wedge B] + O(t^2) = F_A + td_AB + O(t^2),$$

so now we plug back into the Yang-Mills action. Recall that for real-valued k-forms η,γ, $\int_M \eta\wedge *\gamma = \int\langle\eta,\gamma\rangle\omega$, with the inner product having form in coordinates $\sum_{i_1,\cdots,i_k j_1,\cdots j_k}\eta_{i_1,\cdots,i_k}\gamma_{j_1,\cdots j_k}g^{i_1j_1}\cdots g^{i_kj_k}$; the same derivation works for matrix-valued k-forms and thus we actually see that $\mathrm{Tr}(\langle\eta,\gamma\rangle) = \mathrm{Tr}(\langle\gamma,\eta\rangle)$ – that is, we can swap the terms in the inner product if we take the trace. So applying that to our calculation here, we have (we can omit the volume form because it's implicitly part of the definition of the integral of a smooth function)

$$\begin{aligned} S(A+tB) &= -\int_M \mathrm{Tr}(\langle F_{A+tB}, F_{A+tB}\rangle) \\ &= -\int_M \mathrm{Tr}(\langle F_A, F_A\rangle + t\langle F_A, d_AB\rangle + t\langle d_AB, F_A\rangle + O(t^2)) \\ &= -\int_M \mathrm{Tr}(\langle F_A, F_A\rangle + 2t\langle F_A, d_AB\rangle + O(t^2)), \end{aligned}$$

and therefore

$$S(A+tB) - S(A) = \int_M -2t\mathrm{Tr}(\langle F_A, d_AB\rangle) + O(t^2)) \implies \lim_{t\to 0}\frac{S(A+tB)-S(A)}{t} = -2\int_M \mathrm{Tr}(\langle d_AB, F_A\rangle).$$

(our functions are smooth and compactly supported, so we can use the dominated convergence theorem to swap the limit and the integral). And thus indeed A is a critical point if and only if $\int_M \mathrm{Tr}(\langle d_AB, F_A\rangle) = 0$ for all compactly supported B, and by the property of the adjoint this occurs if and only if $\int_M \mathrm{Tr}(\langle B, d_A^*F_A\rangle) = 0$. Finally, this happens if and only if $d_A^*F_A = 0$ – for this last step, the idea is that for any $x\in M$ we can take a sequence of Bs approaching $d_A^*F_A(x)\delta_x$. □

We'll now discuss more about these critical points (sometimes called **instantons**) and then do the example of Minkowski space. Notice that $A = 0$ is always a solution to the Yang-Mills equation, and also if the Lie algebra is abelian (so that $[X,Y] = 0$ for all $X,Y\in\mathfrak{g}$), then for any $A = df$ for a smooth function $f\in C^\infty(M,\mathfrak{g})$, we have in coordinates that

$$A = \sum_i A_i dx_i = \sum_i \frac{\partial f}{\partial x_i}dx_i,$$

and then we have $F_A = dA + \frac{1}{2}[A\wedge A] = dA = ddf = 0$ and thus A will again be a solution.

The simplest example of a nontrivial solution in Riemannian space is pretty hard to find – supposedly there cannot be one in a finite action (meaning that S_A is finite) for the Euclidean metric, though it's unclear where the proof is.

Example 75

Let $M = \mathbb{R}^4$ (with g the identity matrix) and $G = U(1)$, so that $\mathfrak{g} = i\mathbb{R}$. We'll work out the Yang-Mills equations explicitly in this case.

For any connection $A\in\Omega^1(M,\mathfrak{g})$ and F its curvature form, we have $F_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}$ (because we have a commuting

Lie algebra). We then have (recalling the definition of the Hodge star operator from Definition 63, and that we have $n = 4, k = 2$ here)

$$*F_{ij} = \frac{1}{2!} \sum_{\substack{1 \le k, \ell \le 4 \\ 1 \le k', \ell' \le 4}} \varepsilon_{k',\ell',i,j} g^{kk} g^{\ell\ell'} F_{k\ell},$$

and since g is the identity matrix this simplifies to just

$$*F_{ij} = \frac{1}{2} \sum_{1 \le k, \ell \le 4} \varepsilon_{k,\ell,i,j} F_{k\ell},$$

and thus for any $i < j$ we must have k, ℓ the other two elements, either arranged in increasing order or decreasing order, but then $F_{k\ell}$ is the negative of $F_{\ell k}$. Therefore we actually end up having

$$*F_{ij} = \operatorname{sgn}(\pi) F_{k\ell}, \quad \{k, \ell\} = \{1, 2, 3, 4, \} \setminus \{i, j\}, \quad \pi = (k, \ell, i, j) \in S_4.$$

We can thus write out explicitly that

$$*F_{12} = \operatorname{sgn}(3, 4, 1, 2) F_{34} = +F_{34},$$

$$*F_{13} = \operatorname{sgn}(2, 4, 1, 3) F_{24} = -F_{24},$$

and similarly we'll find

$$*F_{14} = F_{23}, \quad *F_{23} = F_{14}, \quad *F_{24} = -F_{13}, \quad *F_{34} = F_{12}.$$

Plugging these all in, we thus find the exterior derivative of $*F$ after some calculations; for example the coefficient of $dx_2 \wedge dx_3 \wedge dx_4$ is $\frac{\partial *F_{34}}{\partial x_2} - \frac{\partial *F_{24}}{\partial x_3} + \frac{\partial *F_{23}}{\partial x_4}$, and there are three other coefficients to compute. But the Yang-Mills equations for this case are just $d * F = 0$, which end up just being the set of second-order differential equations

$$\sum_{j=1}^{4} \frac{\partial F_{ij}}{\partial x_j} = 0, \quad i = 1, 2, 3, 4.$$

An example of a nontrivial solution to this is achieved by taking $A_2 = A_3 = A_4 = 0$, so that $F_{23} = F_{24} = F_{34} = 0$ and $F_{1j} = -\frac{\partial A_1}{\partial x_j}$. The Yang-Mills equations then hold if and only if $\sum_{j=2}^{4} \frac{\partial^2 A_1}{\partial x_j^2} = 0$, so any A_1 harmonic in x_2, x_3, x_4 (and depending on x_1 in any way we'd like) works. (Though notice that any such nonlinear A must grow and thus has infinite $S(A)$.)

Maxwell's equations are then basically this example but in Minkowski space (we will end up with trigonometric functions), so somehow special relativity was already present even before Einstein! And that's what we'll discuss next. We first need to introduce what such a space actually is (and all of modern physics has to be invariant under Lorentz transforms, so it needs tis concept to make sense):

Definition 76

Let M be a (smooth, n-dimensional) manifold, and let p, q be nonnegative integers with $p + q = n$. Letting $S = \{(x, u, v) : x \in M, u, v \in T_x M\}$ with manifold structure as in Definition 59, a **pseudo-Riemannian metric** on M is a smooth map $g : S \to \mathbb{R}$ such that for any $x \in M$, the map $g(x, \cdot, \cdot) : T_x M \times T_x M \to \mathbb{R}$ is a **bilinear form** on $T_x M$ **with** p **positive and** q **negative eigenvalues** (when thought of as a matrix); we call (p, q) the **signature** of the metric. A **psuedo-Riemannian manifold** is a manifold equipped with such a metric.

We'll be specifically interested in the particular case

$$\mathbb{R}^{1,3} = \mathbb{R}^4 \text{ with psuedo-Riemannian metric } g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

sometimes we also make the first coordinate $+$ and the others $-$ depending on the context, but either way this is called **Minkowski space**, and $(p, q) = (3, 1)$ is the **Lorentz signature**. This is the example we'll discuss next lecture – we'll see how everything so far can be changed for pseudo-Riemannian metrics, since we really haven't used the positivity of the metric in very many places other than the $\sqrt{\det g}$ part.

12 October 18, 2024

We'll start today by discussing all of the changes we need to make when going from Riemannian manifolds to psuedo-Riemannian ones:

- In the definitions of the Hodge star operator (Definition 63) and the volume form (Definition 61), we must replace $\sqrt{\det g}$ with $\sqrt{|\det g|}$.
- In particular, this means that we now have

  $$** \eta = (-1)^{k(n-k)+q}\eta,$$

  and similarly the definition of the adjoint operator now looks like

  $$d_A^* \eta = (-1)^{n(k-1)+1+q} * d_A * \eta.$$

- Critical points of the Yang-Mills action are still the connections satisfying $d_A^* F_A = 0$, since the proof that we must have $\int_M \mathrm{Tr}(\langle B, d_A^* F_A \rangle) = 0$ for all compactly supported B still works. The only fact that we really need is that the matrix $g(x)$ is nonsingular.

We'll really only be concerned for now about $\mathbb{R}^{1,3}$ (that is, $\mathbb{R}^4$ with the Lorentz metric), but it's good to know how things work in general. And now we're ready to dive into electromagnetism and Maxwell's equations:

Example 77

In physics, electromagnetic waves are described in terms of a pair of vector fields $(\vec{E}, \vec{B})$ on $\mathbb{R}^{1,3}$, meaning that $\vec{E}(x, t)$ and $\vec{B}(x, t)$ are vectors in $\mathbb{R}^3$ and the maps $\vec{E}, \vec{B} : \mathbb{R}^{1,3} \to \mathbb{R}^3$ are smooth. In the absence of matter (such as electrons and positrons), **Maxwell's equations** read (taking **natural units** so that the speed of light is 1)

$$\nabla \cdot \vec{E} = 0 \quad \text{(Gauss's law of electricity)}, \quad \nabla \cdot \vec{B} = 0 \quad \text{(Gauss's law of magnetism)},$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \text{(Faraday's law of induction)}, \quad \nabla \times \vec{B} = +\frac{\partial \vec{E}}{\partial t} \quad \text{(Ampere's law in vacuum)},$$

where recall that for a function $f = (f_1, f_2, f_3) : \mathbb{R}^{1,3} \to \mathbb{R}^3$, parameterizing $\mathbb{R}^{1,3}$ as (t, x, y, z) or (x_0, x_1, x_2, x_3), we define the **divergence** $\nabla \cdot f = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3}$ and the **curl** $\nabla \times f = \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)$. We'll deduce these equations from $U(1)$ Yang-Mills theory on $\mathbb{R}^{1,3}$.

Remark 78. *Even with charge and current density, there does exist a $U(1)$ gauge theory "with sources" corresponding to Maxwell's equations. But those equations assume that the matter is "just there" affecting the electromagnetic fields, even though the effect goes both ways. Including that whole thing means we need an action with everything together – this is much more complicated and thus we're avoiding it for now.*

Taking $M = \mathbb{R}^{1/3}$ and $G = U(1)$, a gauge connection $A \in \Omega^1(M, \mathfrak{g})$ is a $\mathfrak{g}$-valued 1-form, where $\mathfrak{g} = i\mathbb{R}$ (because it's a 1×1 skew Hermitian matrix, which means the entry is pure imaginary). Thus we may write

$$A = \sum_{j=0}^{3} A_j dx_j$$

for smooth functions $A_j : \mathbb{R}^{1,3} \to i\mathbb{R}$, or equivalently we can think of A as a map from $\mathbb{R}^{1,3}$ into $(i\mathbb{R})^4$. The idea now is to write $A = (-i\phi, i\vec{A})$ (where $A_0 = -i\phi$ and $i\vec{A} = (A_1, A_2, A_3)$), so that ϕ is a real-valued function called the **scalar potential** and $\vec{A}$ is an $\mathbb{R}^3$-valued function called the **vector potential**. We then set (remember that gradient and curl are ignoring the time coordinate)

$$\vec{E} = -\nabla\phi - \frac{\partial \vec{A}}{\partial t}, \quad \vec{B} = \nabla \times \vec{A}.$$

We may ask if this is one-to-one, but we'll see when discussing gauge transformations that it's not. This corresponds to how $\vec{E}$ and $\vec{B}$ can be measured, but ϕ and $\vec{A}$ cannot and thus are not physically meaningful until we do this kind of quotienting operation.

Theorem 79

Let $\vec{A}$ and $(\vec{E}, \vec{B})$ be as above. Then $(\vec{E}, \vec{B})$ satisfies Maxwell's equations if and only if A is a Yang-Mills connection (that is, $d_A^* F_A = 0$). Conversely, if $(\vec{E}, \vec{B})$ is an electromagnetic field satisfying Maxwell's equations and (a technical condition – **see the remark below**) for all t, there is some $\varepsilon > 0$ such that $|\vec{B}(t,x)|$, $\left|\frac{\partial \vec{B}}{\partial t}\right|$, and $|\vec{E}(t,x)|$ are all $O(|x|^{-1-\varepsilon})$ as $|x| \to \infty$, then $(\vec{E}, \vec{B})$ is generated by some Yang-Mills connection A via $\vec{E} = -\nabla\phi - \frac{\partial \vec{A}}{\partial t}$ and $\vec{B} = \nabla \times \vec{A}$.

Remark 80. *After class, Professor Chatterjee noted that these technical conditions aren't necessary – in particular, in the proof of this converse direction, we can instead use* **Poincaré's lemma**, *which states for any smooth real-valued k-form ($1 \le k \le n$) ω on $\mathbb{R}^n$ with $d\omega = 0$, there is some $(k-1)$-form η with $d\eta = \omega$, and we may use that to construct $\vec{A}$ rather than the method described below. We talk about this more in the next lecture.*

Proof. Let $A \in \Omega^1(\mathbb{R}^{1,3}, \mathfrak{u}(1))$, write $F = F_A$, and let $\vec{E}, \vec{B}$ be generated by ϕ and $\vec{A}$ as above. We claim that the Bianchi identity $d_A F = 0$, which is **always** true, is equivalent to the second and third laws $\nabla \cdot \vec{B} = 0$ and $\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$, and that the Yang-Mills equation $d_A^* F = 0$ is equivalent to the other two laws $\nabla \cdot \vec{E} = 0$ and $\nabla \times \vec{B} = \frac{\partial \vec{E}}{\partial t}$. Indeed, we can write the 2-form F as

$$F = \sum_{j=1}^{3} F_{0j} dt \wedge dx_j + \sum_{1 \le i < j \le 3} F_{ij} dx_i \wedge dx_j.$$

We have $F_{0j} = \frac{\partial A_j}{\partial t} - \frac{\partial A_0}{\partial x_j}$, and since $\vec{E} = -\nabla\phi - \frac{\partial \vec{A}}{\partial t}$ this right-hand side is exactly $-iE_j$, where E_j is the jth component of the electric field. Similarly, the magnetic field terms come from the coordinates of $\vec{B} = \nabla \times \vec{A}$ matching up with those of $F_{ij} dx_i \wedge dx_j$:

$$\boxed{\vec{B} = \nabla \times \vec{A} = (-iF_{23}, iF_{13}, -iF_{12})}.$$

Thus the $3+3$ coordinates of the electric and magnetic fields correspond to the $\binom{4}{2} = 6$ components of the curvature

form. Therefore because we have an abelian Lie algebra, we get the 3-form

$$d_A F = dF = \sum_{1\le i<j\le 3}\left(\frac{\partial F_{ij}}{\partial t} - \frac{\partial F_{0j}}{\partial x_i} + \frac{\partial F_{0i}}{\partial x_j}\right) dt\wedge dx_i\wedge dx_j + \left(\frac{\partial F_{23}}{\partial x_1} - \frac{\partial F_{13}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_3}\right) dx_1\wedge dx_2\wedge dx_3.$$

Since this must be zero, each coefficient is zero and we indeed recover those second and third Maxwell's equations (see the boxed equation for $\vec{B}$ above, and also see that $\frac{\partial F_{ij}}{\partial t} = \frac{\partial F_{0j}}{\partial x_i} - \frac{\partial F_{0i}}{\partial x_j}$ is exactly saying that $-\frac{\partial \vec{B}}{\partial t} = \nabla\times\vec{E}$. We call these two the **homogeneous Maxwell's equations**.

For the other two (which we call the **inhomogeneous Maxwell's equations**), we first need to compute the Hodge star operator

$$*F_{ij} = \frac{1}{2}\sum_{0\le k,\ell\le 3}\sum_{0\le k',\ell'\le 3}\varepsilon_{k,\ell,i,j}g^{kk'}g^{\ell\ell'}F_{k\ell}.$$

We've computed this last lecture except with g being the identity matrix; now because g has a -1 in the top left entry we instead have

$$*F_{01} = F_{23} = iB_1,\quad *F_{02} = -F_{13} = iB_2,\quad *F_{03} = F_{12} = iB_3,$$
$$*F_{12} = -F_{03} = iE_3,\quad *F_{13} = F_{02} = -iE_2,\quad *F_{23} = -F_{01} = iE_1.$$

In particular, the Yang-Mills equation $d_A^* F = 0$ implies that $d*F = 0$ (since there's no commutator part), and thus if we replace F with $*F$ we can go through all of the calculations again. The analysis then proceeds exactly the same, but $\vec{E}$ is replaced by $-\vec{B}$ and $\vec{B}$ is replaced by $\vec{E}$; this is indeed the form of the first and fourth equations, as desired. (With the Euclidean metric we would have instead gotten $\frac{\partial\vec{E}}{\partial t} = -\nabla\times\vec{B}$, so we do need the Minkowski metric for all of the physics to work out!)

Finally, we'll just sketch the proof of the converse direction. Suppose $(\vec{E},\vec{B})$ satisfy the **homogeneous** Maxwell equations. It's a fact (called the Helmholtz decomposition) that if $\vec{F}$ is a smooth vector field on $\mathbb{R}^3$ such that $|\vec{F}(x)| = O(|x|^{-1-\varepsilon})$, then we can decompose $\vec{F}(x) = -\nabla\phi + \nabla\times\vec{A}$, where we have the formulas

$$\phi(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{\nabla\cdot\vec{F}(y)}{|y-x|}dy,\quad \vec{A}(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{\nabla\times\vec{F}(y)}{|y-x|}dy.$$

So in particular, if the divergence of $\vec{F}$ is identically zero, then $\phi = 0$ (meaning $\vec{F}$ is the curl of a vector field), and if the curl is zero, then $\vec{A} = 0$ (meaning $\vec{F}$ is the gradient of a scalar field). Thus the condition $\nabla\cdot\vec{B} = 0$ means $\vec{B}$ is the curl of some field which we make our $\vec{A}$, and the curl of $\vec{E} + \frac{\partial\vec{A}}{\partial t}$ is zero and thus this is the gradient of some scalar field, which we'll make $\nabla\phi$. □

Next week, we'll discuss some solutions of these equations and then get more into gauge transformations!

13 October 21, 2024

We'll start by elaborating a bit on how Remark 80 lets us prove the converse direction. If $(\vec{E},\vec{B})$ is an electromagnetic field satisfying Maxwell's equations, then we can define the 2-form ω on $\mathbb{R}^4$ ($\mathbb{R}^{1,3}$, but the metric doesn't matter for this) via

$$\omega = \sum_{0\le i<j\le 3}\omega_{ij}dx_i\wedge dx_j,$$

where we set $\omega_{0j} = -E_j$ for $j = 1,2,3$ and $\omega_{23} = -B_1$, $\omega_{13} = B_2$, $\omega_{12} = -B_3$. Then Maxwell's equations imply that $d\omega = 0$ (by direct computation) and thus there is some η with $d\eta = \omega$; we define the $U(1)$-valued 1-form $A = \sum_{j=0}^3 A_j dx_j$ by $A_j = i\eta_j$, and we can verify that this indeed is a Yang-Mills connections.

Remark 81. *Maxwell's equations were not originally formulated by just a single person – this kind of framework was only realized to be a critical point of the Yang-Mills action by Weyl. But it's a good point to get us started with generalizing this to more complicated gauge groups.*

Remark 82. *The gauge theory framework will be the same even if we do everything on a different manifold than $\mathbb{R}^{1,3}$; however, we won't have the Poincaré lemma anymore and we also need to change Maxwell's equations and use the Yang-Mills critical point characterization directly.*

We won't do the full proof of the Poincaré lemma, but we'll do a quick sketch for 1-forms:

Proof sketch of Poincaré lemma (from Remark 80). Write $\omega = \sum_{i=1}^n \omega_i dx_i$ for smooth functions ω_i (remember we're working in $\mathbb{R}^n$ so we don't need to worry about local coordinates). Since $d\omega = 0$, this means that $\frac{\partial \omega_j}{\partial x_i} - \frac{\partial x_j}{\partial \omega_i} = 0$ for all i, j. Thus we can produce an η explicitly as

$$\eta(x) = \int_0^1 \left(\sum_{i=1}^n x_i \omega_i(tx) \right) dt.$$

Indeed, we can differentiate under the integral and find

$$\begin{aligned} \frac{\partial \eta}{\partial x_j} &= \int_0^1 \left(\omega_j(tx) + \sum_{i=1}^n t x_i \partial_j \omega_i(tx) \right) dt \\ &= \int_0^1 \left(\omega_j(tx) + \sum_{i=1}^n t x_i \partial_i \omega_j(tx) \right) dt, \end{aligned}$$

but now by the product and chain rule this bracket term is exactly $\frac{\partial}{\partial t}(t\omega_j(tx))$, so integrating that will get us $\omega_j(x)$. More generally, for k-forms we have to write

$$\eta(x) = \sum_{1 \le j_1 < \cdots < j_{k-1} \le n} \sum_{i=1}^n \left(\int_0^1 t^{k-1} x_i \omega_{i, j_1, \cdots j_{k-1}}(tx) dt \right) dx_{j_1} \wedge \cdots \wedge dx_{j_{k-1}}$$

and do a more complicated substitution. □

We'll now return to Maxwell's equations and consider some sample solutions for illustration purposes:

Example 83

Let $A_0 = A_1 = A_2 = 0$ and take $A_3(t, x) = i \sin(x_2 - t)$. Recalling that $A = (-i\phi, i\vec{A})$, where $\vec{E} = -\nabla\phi - \frac{\partial \vec{A}}{\partial t}$, we find that (because $\vec{A}_3 = \sin(x_2 - t)$ and all other coordinates of the potential are zero) $\boxed{\vec{E} = (0, 0, \cos(x_2 - t))}$. Meanwhile, $\vec{B} = \nabla \times \vec{A} = \left(-i\frac{\partial A_3}{\partial x_2}, i\frac{\partial A_3}{\partial x_1}, 0 \right)$ and thus $\boxed{\vec{B} = (\cos(x_2 - t), 0, 0)}$.

We only need to check that the inhomogeneous Maxwell's equations hold, since the other two always will by the Bianchi identity – indeed $\nabla \cdot \vec{E} = 0$ and $\nabla \times \vec{B} = \frac{\partial \vec{E}}{\partial t}$. So the point is that we have an electromagnetic wave traveling at the speed of light in the y-direction, where $\vec{E}$ is pointed in the z-direction and $\vec{B}$ is pointed in the x-direction. So this means A is indeed a Yang-Mills connection.

Our last topic of this part of the course will be **gauge transformations**:

Definition 84

Let $M \times G$ be a trivial principal bundle, where M is a (smooth) n-dimensional manifold with a pseudo-Riemannian metric and G is a **matrix** Lie group contained in $M(m, \mathbb{C})$. Let $\mathcal{A} = \Omega^1(M, \mathfrak{g})$ denote the set of gauge fields (here $\mathfrak{g}$ is the Lie algebra of G).

Let $\mathcal{G} = C^\infty(M, G)$ be the set of smooth functions from M into G, which we can think of as a subset of matrix-valued 0-forms $\Omega^0(M, M(m, \mathbb{C}))$. We call these the **gauge transformations**. Defining a group structure on $\mathcal{G}$ via pointwise multiplication, the product does indeed preserve the smoothness, so $\mathcal{G}$ acts on $\mathcal{A}$ via (for any $A \in \mathcal{A}, U \in \mathcal{G}$)

$$A^U = U^{-1}AU + U^{-1}dU.$$

Here dU is taken in the sense of viewing U as a matrix-valued 0-form, and $\mathfrak{g}$ is thought of as a subset of $M(m, \mathbb{C})$ so that $U^{-1}AU$ is a wedge product of a 0-form, a 1-form, and a 0-form, and similarly $U^{-1}dU$ is the wedge product of U^{-1} with dU.

So A^U is again a 1-form; we claim that it is indeeed an element of the Lie algebra. In coordinates, this definition says that if $A = \sum_{i=1}^n A_i dx_i$, then

$$U^{-1}AU = \sum_{i=1}^n U^{-1}A_i U dx_i, \quad U^{-1}dU = \sum_{i=1}^n U^{-1}\frac{\partial U}{\partial x_i}dx_i,$$

and we add these together to get A^U – our goal is to prove that both of those coefficients are in the Lie algebra so that A^U is also a $\mathfrak{g}$-valued 1-form.

Indeed, thinking about this in terms of curves, if $A_i(x) \in \mathfrak{g}$ at some point $x \in M$, then there exists a curve $\gamma : (-\varepsilon, \varepsilon) \to G$ such that $\gamma(0) = I$ and $\gamma'(0) = A_i(x)$. If we then consider the curve $\eta(t) = U(x)^{-1}\gamma(t)U(x)$, then similarly we have $\eta(0) = I$ and $\eta'(t) = U(x)^{-1}A_i(x)U(x)$, and thus $U^{-1}A_iU$ is indeed also in the Lie algebra.

And for the other term, suppose (W, ϕ) is a chart with $x \in W$. Defining the curve

$$\gamma(t) = U(x)^{-1}U(\phi^{-1}(\phi(x) + te_i))$$

so that on the Euclidean plane we move along a coordinate direction, we have $\gamma(0) = U(x)^{-1}U(\phi^{-1}(\phi(x))) = U(x)^{-1}U(x) = I$, and by the chain rule we have

$$\gamma'(t) = U(x)^{-1}\left.\frac{d}{dt}U(\phi^{-1}(\phi(x) + te_i))\right|_{t=0},$$

but this blue part is exactly our definition of $\frac{\partial U}{\partial x_i}$ (see the discussion after Definition 20). Thus $U^{-1}\frac{\partial U}{\partial x_i}$ comes from the derivative of a curve and is thus part of the Lie algebra as well.

Lemma 85

The action of gauge transformations is indeed a group action, meaning that $(A^U)^V = A^{UV}$.

Proof. We may compute

$$\begin{aligned}(A^U)^V &= V^{-1}A^U V + V^{-1}dV \\ &= V^{-1}(U^{-1}AU + U^{-1}dU)V + V^{-1}dV \\ &= (UV)^{-1}A(UV) + (UV)^{-1}(dU)V + V^{-1}dV.\end{aligned}$$

We wish to show that the last two terms here are equal to $(UV)^{-1}d(UV)$, but by the Leibniz rule we have

$$(UV)^{-1}d(UV) = (UV)^{-1}((dU)V + U(dV))$$

because the Liebniz rule for k-forms has a $(-1)^k$ factor but we have 0-forms here, and then $(UV)^{-1}UdV$ is exactly just $V^{-1}dV$ as desired. □

Definition 86

With the notation in the previous definition, we say that two gauge fields $A, B \in \mathcal{A}$ are **gauge equivalent** if $B = A^U$ for some $U \in \mathcal{G}$. This is an equivalence relation because we have a group action, and the quotient space $\mathcal{A}/\mathcal{G}$ is called the space of **gauge orbits**.

We mentioned that $\vec{E}$ and $\vec{B}$ are the quantities we may measure rather than ϕ and $\vec{A}$ – in general the idea is that **gauge orbits are the actual physical space**. But for nonabelian groups, we may get some complicated nonlinear quotient and thus things may be not so easy to understand as in the Maxwell's equations case.

Theorem 87

Let $A \in \mathcal{A}$ and $U \in \mathcal{G}$, and let $B = A^U$. Then $F_B = U^{-1}F_A U$, and therefore if $S(A)$ is finite then $S(B) = S(A)$. That is, the Yang-Mills action is a **gauge-invariant observable**.

We'll do this (messy) computation at the start of next lecture!

Remark 88. *After this discussion concludes, we'll be trying to take these classical into quantum systems and path integrals; we'll start by quantizing particles (and those can be done rigorously) and then see how quantizing fields can be a huge challenge. (There aren't any mathematically rigorously quantum field theories constructed to this day.)*

14 October 23, 2024

We'll prove today that the Yang-Mills action remains invariant under the action of the gauge group:

Proof of Theorem 87. Recall that $B = U^{-1}AU + U^{-1}dU$. We have by definition of the curvature form that

$$\begin{aligned} F_B &= dB + B \wedge B \\ &= d(U^{-1}AU + U^{-1}dU) + (U^{-1}AU + U^{-1}dU) \wedge (U^{-1}AU + U^{-1}dU). \end{aligned}$$

We'll do the two terms here separately. First, the dB term simplifies by the Leibniz rule as

$$\begin{aligned} dB = d(U^{-1}AU + U^{-1}dU) &= d(U^{-1}AU) + d(U^{-1}dU) \\ &= (d(U^{-1})) \wedge AU + U^{-1}d(AU) + d(U^{-1})dU + U^{-1}ddU. \end{aligned}$$

The last term is zero because $d^2 = 0$, and to proceed further we use that

$$U^{-1}U = I \implies 0 = d(U^{-1}U) = d(U^{-1})U + U^{-1}dU$$

and thus rearranging we see that $d(U^{-1})U = -U^{-1}dU$, and since U is just a matrix that is multiplied component-wise we can rewrite this as $\boxed{d(U^{-1}) = -U^{-1}(dU)U^{-1}}$. Plugging this back in, we find that (now also using the Leibniz rule

on $d(AU)$ but now with the 1-form A)

$$\boxed{dB = -U^{-1}(dU)U^{-1} \wedge AU + U^{-1}((dA)U - A \wedge (dU)) - U^{-1}(dU) \wedge U^{-1}(dU)}.$$

Meanwhile, the other term simplifies as

$$\begin{aligned} B \wedge B &= (U^{-1}AU + U^{-1}dU) \wedge (U^{-1}AU + U^{-1}dU) \\ &= U^{-1}AU \wedge U^{-1}AU + U^{-1}dU \wedge U^{-1}AU + U^{-1}AU \wedge U^{-1}dU + U^{-1}dU \wedge U^{-1}dU. \end{aligned}$$

But the last term here now cancels out with the last term of the boxed dB expression, and the second term here cancels with the first term of dB. Additionally, the U and U^{-1} around the wedge product cancel out in the first and third terms and thus we can rewrite them as $U^{-1}A \wedge AU$ and $U^{-1}A \wedge dU$, respectively (at which point the latter term cancels out with one of the remaining terms from dB). Therefore if we plug everything back in, the only terms that remain are

$$dB + B \wedge B = U^{-1}(dA + A \wedge A)U = U^{-1}F_A U,$$

as desired. Finally, to show that the Yang-Mills action is invariant, we show that the smooth functions $\mathrm{Tr}(\langle F, F\rangle)$ and $\mathrm{Tr}(\langle F', F'\rangle)$ are equal, where $F = F_A$ and $F' = F_B$. Recall that in local coordinates, we have the expression (coming from the identity $F \wedge *F = \langle F, F\rangle\omega$)

$$\langle F, F\rangle = \sum_{1 \le i,j,k,\ell \le n} g^{ik}g^{j\ell}F_{ij}F_{k\ell},$$

and now if we apply the trace to both sides and note that $F' = U^{-1}FU$ implies $F'_{ij} = U^{-1}F_{ij}U$ for all i, j, we see

$$\begin{aligned} \mathrm{Tr}(\langle F', F'\rangle) &= \sum_{1 \le i,j,k,\ell \le n} g^{ik}g^{j\ell}\mathrm{Tr}(U^{-1}F_{ij}UU^{-1}F_{k\ell}U) \\ &= \sum_{1 \le i,j,k,\ell \le n} g^{ik}g^{j\ell}\mathrm{Tr}(U^{-1}F_{ij}F_{k\ell}U) \\ &= \sum_{1 \le i,j,k,\ell \le n} g^{ik}g^{j\ell}\mathrm{Tr}(F_{ij}F_{k\ell}UU^{-1}) \\ &= \sum_{1 \le i,j,k,\ell \le n} g^{ik}g^{j\ell}\mathrm{Tr}(F_{ij}F_{k\ell}) \\ &= \mathrm{Tr}(\langle F, F\rangle), \end{aligned}$$

completing the proof. □

We'll show one last result for now, specifically seeing how all of this theory works out for electromagnetic fields:

Theorem 89

Two $U(1)$ connections A and B on $\mathbb{R}^{1,3}$ yield the same electric and magnetic fields $(\vec{E}, \vec{B})$ if and only if A and B are gauge-equivalent. Thus, quotienting by gauge-equivalence here really gives us the physical objects of interest.

Proof. First suppose that $B = A^U$ for some gauge transformation $U \in \mathcal{G}$. Then U is a smooth map from $\mathbb{R}^{1,3}$ to $U(1)$, or equivalently we can say that there is a smooth map $\psi : \mathbb{R}^{1,3} \to \mathbb{R}$ such that $U(x) = e^{i\psi(x)}$. We then have

$$\begin{aligned} U^{-1}dU &= \sum_{j=0}^{3} e^{-i\psi(x)}\frac{\partial}{\partial x_j}e^{i\psi(x)}dx_j \\ &= i\sum_{j=0}^{3}\frac{\partial\psi}{\partial x_j}dx_j, \end{aligned}$$

and now because our gauge group is abelian, $U^{-1}AU = A$ and thus we actually have the form $\boxed{B = A + id\psi}$ for any gauge transformation. Therefore

$$F_B = dB + \frac{1}{2}[B, B] = dB = dA + idd\psi = dA = F_A$$

and thus the pair $(\vec{E}, \vec{B})$ will indeed be the same for A and B (since the electric and magnetic fields are expressed in terms of the entries of the curvature form).

For the converse direction, suppose A and B generate the same $(\vec{E}, \vec{B})$. Then again by the correspondence we have $F_A = F_B$, but this means $d(A - B) = 0$ and thus by the Poincaré lemma $A - B = d\eta$ for some $\mathfrak{u}(1)$-valued 0-form η. So writing $\eta = i\psi$ for some **real**-valued 0-form, we indeed have $B = A + id\psi$ and thus we have a gauge transformation. □

For the next part of the course, we'll try to describe as precisely as possible "what a quantum field theory is" – many attempts have been made but it's quite difficult to do anything universally accepted and complete. For this, we'll start from the basics and explain things from the physics, specifically **Lagrangian mechanics**. This will explain where the whole concept of an action comes from in the first place.

Definition 90

Suppose we have a system of N particles in $\mathbb{R}^3$ with masses $m_1, \cdots, m_N$. Say that when these particles are at locations $x_1, \cdots, x_N$, the system has potential energy $V(x_1, \cdots, x_N)$ (for example due to gravity). A **trajectory** of the system is a smooth map $\gamma : (-\infty, \infty) \to (\mathbb{R}^3)^N$, specifying the locations $\gamma(t) = (\gamma_1(t), \cdots, \gamma_N(t))$ of the particles at time t. In the absence of any outside influences (such as external forces), the **action** of the trajectory is formally given by the integral of the Lagrangian

$$S(\gamma) = \int_{-\infty}^{\infty} \left(\frac{1}{2} \sum_{i=1}^{N} m_i (\dot{\gamma}_i(t))^2 - V(\gamma(t)) \right) dt,$$

where $\dot{\gamma}_i = \frac{d\gamma_i}{dt}$.

In other words, the Lagrangian is the kinetic energy **minus** the potential energy, and the action integrates this over all time. (It'll be clear soon why this is a reasonable quantity to consider.) This is only a formal integral because it may not converge, but it's still useful – just like we can find critical points for the Yang-Mills action, we can find critical points here by noting that if we change γ by a finite amount (only in a localized region), then the difference in S is well-defined. More precisely, suppose η is a smooth trajectory which is zero outside a bounded time interval. Then the quantity $S(\gamma + \varepsilon\eta) - S(\gamma)$ is well-defined for all real ε, because the difference of the integrands is only nonzero on a bounded interval, and thus we say that γ is a **critical point of the action** if the directional derivative

$$\lim_{\varepsilon \to 0} \frac{S(\gamma + \varepsilon\eta) - S(\gamma)}{\varepsilon}$$

is zero for all such η.

So now let's calculate what those critical points are: we have

$$S(\gamma + \varepsilon\eta) = \int_{-\infty}^{\infty} \left(\frac{1}{2} \sum_{i=1}^{N} m_i (|\dot{\gamma}_i(t) + \varepsilon\dot{\eta}_i(t)|^2) + V(\gamma(t) + \varepsilon(\eta(t)) \right) dt,$$

and if we take the derivative of this with respect to ε at $\varepsilon = 0$, then we find that

$$\frac{\partial}{\partial \varepsilon} S(\gamma + \varepsilon \eta)\Big|_{\varepsilon=0} = \int_{-\infty}^{\infty} \left(\sum_{i=1}^{N} m_i \dot{\gamma}_i(t) \cdot \dot{\eta}_i(t) - \sum_{i=1}^{N} \eta_i(t) \partial_i(t) \partial_i V(\gamma(t)) \right) dt.$$

But now we can apply integration by parts to the first sum, noting that η vanishes outside a bounded interval, to rewrite this as

$$-\int_{-\infty}^{\infty} \sum_{i=1}^{N} (m_i \eta_i(t)(m_i \ddot{\gamma}_i(t) + \partial_i V(\gamma(t)))\, dt,$$

where note that $\partial_i V$ is a three-dimensional vector. This can be zero for all η if and only if we in fact have $m_i \ddot{\gamma}_i = -\partial_i V$ for all t and for all i, and this is exactly **Newton's second law** $F = ma$.

Remark 91. *In general the critical points of the action may not be the* **minimizers** *of the action, and particularly in quantum mechanics those other points do matter as well. But in this case this does in fact minimize the action.*

Drawing the analogy, the trajectory of the system here takes the role of the 4-component spacetime gauge field A in electromagnetism. So here we define an action $S(\gamma)$ integrated over time instead of the action $S(A)$ integrated over $\mathbb{R}^4$, but we can think of $A(t, \cdot)$ as being the value of field at each time slice (evolving over time), and so solving for the critical points tell us how that field evolves.

So if we have this **classical picture**, we want to do a certain **quantization** in which we move to a **quantum picture** instead. This can be done in various ways (for example through the Hamiltonian, which is a function which can be converted to an operator), but we'll talk about the **path integral approach** here.

Example 92

We stick with the N-particle system defined above. In (nonrelativistic) quantum mechanics, instead of specifying the locations of the particles as a point in $(\mathbb{R}^3)^N$, we instead specify the locations via a **state** $\psi : (\mathbb{R}^3)^N \to \mathbb{C}$, whose interpretation is as follows. For any $\psi \in L^2((\mathbb{R}^3)^N)$, the **probability density function** for the system is

$$\frac{|\psi(x)|^2}{\int_{(\mathbb{R}^3)^N} |\psi(y)|^2 dy}.$$

So in particular the system is not deterministically at a particular location – we have probabilities for the particles appearing in various regions. The reason for specifying this whole system via ψ rather than via the probability density itself is that there are particular kinds of processes that don't have stochastic analogs, so probability densities don't evolve in a nice Markovian way but ψ does.

15 October 25, 2024

We discussed last time that the action corresponding to a particle takes the form $\int_{-\infty}^{\infty} \mathcal{L}(\gamma(t))dt$, where the Lagrangian $\mathcal{L}$ is $\frac{1}{2}\sum_i m_i |\dot{\gamma}_i(t)|^2 - V(\gamma(t))$, and the critical points give us the possible trajectories. For example if $V = 0$ we have a free particle in space, and thus any straight line is a critical point and they are all valid trajectories. But we can also restrict our action by only considering trajectories that satisfy for example $\gamma(t_0) = x_0$ (requiring that the particle is at a particular point at a certain time), and then the critical points will be the set of possible trajectories that also satisfy the additional constraint. We may additionally require $\gamma(t_0) = x_0$ and $\gamma(t_1) = x_1$ for $t_0 < t_1$, and then in the case of the free particle we now only have one possible trajectory.

Moving now to quantum systems, we introduced last time the idea of a **quantum state**, which is a function $\psi \in L^2((\mathbb{R}^3)^N)$ whose interpretation is that $\frac{1}{Z}|\psi(x)|^2$, normalized to have total integral 1, is the probability density function for the state.

Fact 93

The **path integral formulation** states that if the system is in state ψ_0 at some time 0, then the state at time $t > 0$ is given by "weighted averages of the values of ψ, with weightings given by not-very-well-defined complex numbers"

$$\psi_t(x) = \frac{\int_{\gamma:\gamma(t)=x} \psi_0(\gamma(0)) \exp\left(\frac{i}{\hbar} S(\gamma)\right) \mathcal{D}\gamma}{\int_{\gamma:\gamma(t)=x} \exp\left(\frac{i}{\hbar} S(\gamma)\right) \mathcal{D}\gamma}$$

where $\hbar$ is Planck's constant (usually taken to be 1, but the classical limit being recovered if we take $\hbar \to 0$).

The oddest part of this is the "integration" $\int_{\gamma:\gamma(0)=x}$, which is meant to be an integral over all paths $\gamma : [0, t] \to (\mathbb{R}^3)^N$ which satisfy $\gamma(t) = x$. Then $\mathcal{D}\gamma$ is meant to be "Lebesgue measure" or "uniform measure on all such paths," and $S(\gamma)$ is the action of γ, which in our case (since we restrict our paths to only be defined on $[0, t]$) is

$$S(\gamma) = \int_0^t \left(\frac{1}{2} \sum_i m_i |\dot{\gamma}_i(s)|^2 - V(\gamma(s)) \right) ds.$$

(And we could replace this with any other action to dictate a different way the system evolves over time.) The problem is that such an integral over all paths **is not well-defined rigorously**, and furthermore we're actually integrating some oscillatory things and thus $\psi_t(x)$ becomes a weighted average of $\psi_0(\gamma(0))$ over the paths that reach x at time t.

Remark 94. *The reason for this imaginary number in the exponential is really due to Feynman – we'll derive Schrodinger's equation from this formulation, so in some sense this is a general prescription for something that came out of the blue. In particular, Feynman was interested in quantum electrodynamics, and it's not quite clear how you'd write down Schrodinger's equation directly in that case because the states are even more complicated.*

The idea here is that if $\hbar \to 0$ and we are back in the classical picture, then if γ is far away from a critical point and the oscillations will cancel out; hence, the system concentrates on states which satisfy the classical equation and minimize the action.

Remark 95. *It turns out that the map that takes ψ_0 to ψ_t is a unitary operator (so it preserves the L^2 norm) and that we get a group of unitary operators satisfying $U(t)U(s) = U(t + s)$. Unfortunately, we still don't have a general understanding of how to construct such things, so we'll do an example where we can do the exact calculations and then go to functional analysis using which we can set up a general framework that possibly captures what we want (though no one has done it yet).*

Example 96

To do a simple case, consider a single free particle on the real line of mass 1. Then for any trajectory $\gamma : [0, t] \to \mathbb{R}$, we have the action

$$S(\gamma) = \frac{1}{2} \int_0^t \dot{\gamma}(s)^2 ds.$$

(Note: this is corrected later in Fact 104 for the general case, though things work out okay here.) Our goal is then, given some $\psi_0 \in L^2(\mathbb{R})$, to define the quantity

$$\psi_t(x) = \frac{\int_{\gamma:\gamma(t)=x} \psi_0(\gamma(0)) \exp\left(\frac{i}{\hbar} S(\gamma)\right) \mathcal{D}\gamma}{\int_{\gamma:\gamma(t)=x} \exp\left(\frac{i}{\hbar} S(\gamma)\right) \mathcal{D}\gamma}.$$

The obvious idea is to discretize: let $0 = t_0 < t_1 < \cdots < t_n = t$ be a sequence of time points (we don't need to discretize space because there are no spatial derivatives), and let γ now be a discrete time trajectory $\gamma(t_0), \cdots, \gamma(t_n)$, where we are given $\gamma(t_n) = x$. If we write $x_j = \gamma(t_j)$, then $\mathcal{D}\gamma$ can now be $dx_0 dx_1 \cdots dx_{n-1}$ (that is, we do uniform integration over all possible values of the path at these discrete times), and now instead of the action $S(\gamma)$ above we just discretize the derivative:

$$S(\gamma) \to \frac{1}{2}\sum_{j=0}^{n-1}\left(\frac{\gamma(t_{j+1}) - \gamma(t_j)}{t_{j+1} - t_j}\right)^2 (t_{j+1} - t_j).$$

So in particular, we're freely allowing the values at t_j to fluctuate, and if we had a $-S(\gamma)$ instead we'd get a probability measure on the space of paths and the limit gives us a well-defined probabilistic object, which is Brownian motion constrained to be x at time t. But that's nowhere differentiable, and we even lose continuity if we go to higher dimensions, so something is really not in our favor here. But we can work out what this becomes: the **numerator** of $\psi_t(x)$ becomes

$$I_n = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \psi_0(x_0) \exp\left(\frac{i}{2\hbar}\sum_{j=0}^{n-1}\frac{(x_{j+1} - x_j)^2}{s_j}\right) dx_0 \cdots dx_{n-1},$$

where $s_j = t_{j+1} - t_j$ and $x_n = x$, and our goal is to evaluate this integral. First note that even if ψ_0 is very nice, this integral is not even absolutely convergent (since the exponential has absolute value 1), so we need to do some regularization. We may define

$$I_n(\varepsilon) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \psi_0(x_0) \exp\left(\frac{i-\varepsilon}{2\hbar}\sum_{j=0}^{n-1}\frac{(x_{j+1} - x_j)^2}{s_j}\right) dx_0 \cdots dx_{n-1},$$

and for any $\varepsilon > 0$ this is now integrable, so we can evaluate this exactly and then take $\varepsilon \to 0$. This is an improper integral, so there are many ways to take the limit and different ways may give rise to different answers; this way turns out to be one that does give the right answer. (It's not clear exactly what the class of sensible procedures that gives the right answer is, but at the least this one does.)

Being a little less ambitious, instead of taking $\psi_0 \in L^2(\mathbb{R})$, we'll take a dense subset and consider the subset $\mathcal{F}$ of such functions ψ_0 such that the Fourier transform satisfies $\hat{\psi}_0 \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. In particular this means we may express

$$\psi_0(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}_0(p) e^{-ipx} dp$$

for $\hat{\psi}_0 \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and thus

$$I_n(\varepsilon) = \frac{1}{2\pi}\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{\psi}_0(p) \exp\left(-ipx_0 + \frac{i-\varepsilon}{2\hbar}\sum_{j=0}^{n-1}\frac{(x_{j+1} - x_j)^2}{s_j}\right) dx_0 \cdots dx_{n-1} dp.$$

Proceeding with evaluating this integral, we first do a change of variable: letting $y_j = x_{j+1} - x_j$ for all $0 \le j \le n-1$ (the corresponding matrix has determinant 1), this means we may rewrite the integral in the product form

$$\begin{aligned} I_n(\varepsilon) &= \frac{1}{2\pi}\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{\psi}_0(p) \exp\left(-ip\left(-x + \sum_{j=0}^{n-1} y_j\right) + \frac{i-\varepsilon}{2\hbar}\sum_{j=0}^{n-1}\frac{y_j^2}{s_j}\right) dy_0 \cdots dy_{n-1} dp \\ &= \int_{\mathbb{R}} \hat{\psi}_0(p) e^{ipx} \prod_{j=0}^{n-1}\left(\int_{-\infty}^{\infty} \exp\left(-ipy + \frac{i-\varepsilon}{2\hbar}\frac{y^2}{s_j}\right) dy\right) dp. \end{aligned}$$

These inner integrals can now be calculated easily via the following lemma, which we'll prove next time:

Lemma 97

Suppose $a, b \in \mathbb{C}$ with $a = re^{i\theta}$ for $r > 0$ and $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ (meaning that a is in the right half-plane). Then the following integral is absolutely convergent:

$$\int_{-\infty}^{\infty} e^{-ax^2-bx}\,dx = \frac{e^{b^2/4a}e^{-i\theta/2}}{2\sqrt{\pi r}}.$$

Substituting these values into our inner integrals, we have $a = \frac{\varepsilon - i}{2\hbar s_j}$ (no dependence on p) and $b = ip$ (and in fact b is what will matter to us most), so

$$\frac{b^2}{4a} = -\frac{p^2}{2(\varepsilon - i)}\hbar s_j = -\frac{p^2(\varepsilon + i)}{2(1+\varepsilon^2)}\hbar s_j.$$

The other parts $\frac{e^{-i\theta/2}}{2\sqrt{\pi r}}$ will actually cancel out between the numerator and denominator, and thus what matters for us is that

$$\int_{-\infty}^{\infty} \exp\left(-ipy + \frac{i-\varepsilon}{2\hbar s_j}y^2\right) dy = \exp\left(-\frac{p^2(\varepsilon+i)}{2(1+\varepsilon^2)}\hbar s_j\right) \text{ (a factor not depending on } p).$$

Therefore we get (we only care about evaluating the integral over p now, and now we have to sum over s_j because each of those contribute a factor and $\sum_i s_i = t$)

$$I_n(\varepsilon) = \frac{1}{2\pi} C \int_{-\infty}^{\infty} \hat{\psi}_0(p) e^{ipx} \exp\left(-\frac{p^2(\varepsilon+i)}{2(1+\varepsilon^2)}\hbar t\right) dp.$$

By the same calculation, the denominator will end up just being exactly this C, since we had everything except the ψ_0 term and thus we basically have $b = 0$ instead of $b = ip$. So our path integral formulation tells us that our ε-approximation is

$$\psi_{t,\varepsilon}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}_0(p) e^{ipx} \exp\left(-\frac{p^2(\varepsilon+i)}{2(1+\varepsilon^2)}\hbar t\right) dp,$$

and as $\varepsilon \to 0$ this converges to

$$\psi_t(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}_0(p) e^{ipx} \exp\left(-\frac{-ip^2\hbar t}{2}\right) dp.$$

Thus the Fourier transforms satisfy the equation

$$\hat{\psi}_t(p) = \hat{\psi}_0(p) \exp\left(-\frac{ip^2\hbar t}{2}\right);$$

the evolution of a free particle is that we just get a phase shift depending on p, and therefore $U(t) : \psi_0 \to \psi_t$ is a unitary map for $\psi_0 \in \mathcal{F}$. We may then observe (by Plancherel) that U is an isometry, meaning that $\langle U(t)\psi, U(t)\phi\rangle_{L^2} = \langle \psi, \phi\rangle_{L^2}$; thus we may extend U to $L^2(\mathbb{R})$ via approximation. And the formula indeed shows that $U(s+t) = U(s)U(t)$ for all $s, t \in \mathbb{R}$, so we do have a group of isometries.

16 October 28, 2024

We'll start by quickly discussing the integration lemma from last time:

Proof of Lemma 97. We evaluate this via contour integral. Instead of integrating over the real line, integrate over the line passing through $-\frac{b}{2a}$ parallel to the real line; this can be done because e^{-az^2-bz} is an entire function and thus integrating around a big rectangle between those two lines (from real part $-L$ to L) yields zero, and the "outside

edges" will have negligible contribution as $L \to \infty$ because $\mathrm{Re}(a) > 0$. Thus

$$\int_{-\infty}^{\infty} e^{-ax^2-bx}dx = \int_{-\infty}^{\infty} e^{-a(x-b/(2a))^2-b(x-b/(2a))}dx,$$

and this shift helps us because it essentially "completes the square" and leaves us with $e^{b^2/(4a)}\int_{-\infty}^{\infty} e^{-ax^2}dx$. To evaluate this remaining integral, we note that

$$\int_{-\infty}^{\infty} e^{-ax^2}dx = 2\int_0^{\infty} e^{-ax^2}dx = \int_0^{\infty} e^{-a(te^{-i\theta/2})^2}e^{i\theta/2}dt$$

where in the last step we've done the change of contour where instead of integrating outward along the ray at angle 0, we integrating along the ray at $-\frac{\theta}{2}$ (by tracing out a sector of radius R between those arcs and taking $R \to \infty$), removing the oscillatory part. (This trick of changing the contour so our oscillatory integral is no longer oscillatory is pretty common when we do such calculations.) Thus we just have $e^{i\theta/2}\int_0^{\infty} e^{-rt^2}dt$, and now this last integral is the usual Gaussian integral. □

So summarizing our situation now, we have a Hilbert space $\mathcal{H} = L^2(\mathbb{R})$, and we have a group of unitary operators $(U(t))_{t\in\mathbb{R}}$ where each $U(t)$ is unitary and $U(s+t) = U(s)U(t)$ for all s, t. (We can check that $U(0)$ is indeed the identity; we call this a one-parameter group of unitary operators.) These unitary operators have the property that if ψ is the state of our system at time t, then $U(s)\psi$ is the state of the system at time $t+s$.

Fact 98

There's one important technical condition from the mathematical point of view here – $(U(t))_{t\in\mathbb{R}}$ is a **strongly continuous** group, meaning that for all states $\psi \in \mathcal{H}$ the map $t \mapsto U(t)\psi$ is continuous (in t).

This can be easily checked, since U was defined on a dense subset of the Hilbert space with explicit formulas, and then we can prove that there's enough to do a continuous extension. Indeed,

$$\widehat{U(t)\psi} = \hat{\psi}(p)e^{-ip^2\hbar t/2},$$

and thus if we look at the L^2 distance between $U(t)\psi$ and $U(s)\psi$ it's the same as taking the L^2 distance between the Fourier transforms, and thus

$$\begin{aligned}||U(t)\psi - U(s)\psi||_{L^2} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{U(t)\psi}(p) - \widehat{U(s)\psi}(p)|^2 dp \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{\psi}(p)|^2 \left|e^{-ip^2\hbar t/2} - e^{-ip^2\hbar s/2}\right| dp,\end{aligned}$$

and by the dominated convergence theorem this goes to zero as $t \to s$, since $\hat{\psi}$ is square-integrable and the remaining absolute value term is bounded by 2, and the function goes to zero pointwise as $t \to s$.

Remark 99. *Notice that what we're doing here is* **convolving by the heat kernel but with imaginary variance***: if we write $\psi_t = U(t)\psi$, then we can write*

$$\psi_t(x) = \int_{-\infty}^{\infty} \psi(y)\frac{1}{\sqrt{4\pi i t}}e^{-\frac{(y-x)^2}{4it}}dy$$

Again, this is well-defined by first considering good enough ψ and then extending via uniform continuity – the isometry gives us everything.

This all has been a **prototype of a quantum system**: any quantum system has a Hilbert space in which its states live, and it has a strongly continuous one-parameter group of unitary operators which describe the evolution of the

system. (And while we don't often hear about this "strongly continuous" condition in physics, it really is crucial.) And the probabilistic approach to quantum mechanics is to reconstruct the Hamiltonian using probability theory and then using that to generate the unitary group.

Example 100

Let's consider again the free particle on the real line; we'll find the Hamiltonian for this system.

Define $\mathcal{D} \subseteq L^2(\mathbb{R})$ to be the set of all $\psi \in L^2(\mathbb{R})$ such that the limit

$$\lim_{t\to 0} \frac{U(t)\psi - \psi}{-it}$$

exists in L^2 (so not pointwise, but in the sense of convergence of functions). Then we define $H\psi$ to be this limit for any $\psi \in \mathcal{D}$. The general definition is below:

Definition 101

Suppose we have a strongly continuous one-parameter group of unitary operators $(U(t))_{t\in\mathbb{R}}$ on some Hilbert space $\mathcal{H}$. For any x in the set

$$\mathcal{D} = \left\{ x \in \mathcal{H} : \lim_{t\to 0} \frac{U(t)x - x}{-it} \text{ exists} \right\},$$

we let $Hx = \lim_{t\to 0} \frac{U(t)x-x}{-it}$, and we call the operator H the **Hamiltonian** or the **infinitesimal generator** of $(U(t))_{t\in\mathbb{R}}$.

Note that H is linear, and it's only defined on a subspace, not all of $\mathcal{H}$. We'll be able to prove next that (even with no other conditions) $\mathcal{D}$ is actually dense in $\mathcal{H}$, so the Hamiltonian is defined on a dense subspace, and then it turns out that H will be self-adjoint (note that this is not the same thing as Hermitian when we're only densely defined). From there, we'll show (Stone's theorem) that there's a one-to-one correspondence between self-adjoint operators in a Hilbert space and such strongly continuous unitary groups.

Fact 102

Rigorously, what physicists call **Wick rotation** (going from Euclidean to Minkowski space) is to build a probabilistic system, extract its infinitesimal generator (an operator), and then under suitable show it's self-adjoint and thus use Stone's theorem to build the quantum system. And the key thing connecting this all together is showing that H is self-adjoint.

Lemma 103

With the notation in the above defintion, $\mathcal{D}$ is dense in $\mathcal{H}$.

Proof. We claim that for any $x \in \mathcal{H}$, the map $t \mapsto U(t)x$ is uniformly continuous on any bounded time interval (this is not immediately clear just from strong continuity because our map is Hilbert space-valued). For this (unpacking some definitions and using contradiction) it suffices to show that for any bounded interval $[a, b] \to \mathbb{R}$ and for any sequences $s_n, t_n \in [a, b]$ with $s_n - t_n \to 0$, w have $U(s_n)x - U(t_n)x \to 0$. Indeed, pass to a subsequence so that t_n, s_n both

converge to t, since it suffices to show that a further subsequence always goes to 0. Then

$$\begin{aligned} U(s_n)x - U(t_n)x &= U(s_n - t_n)U(t_n)x - U(t_n)x \\ &= (U(s_n - t_n) - I)U(t_n)x \\ &= (U(s_n - t_n) - I)U(t)x + (U(s_n - t_n) - I)(U(t_n)x - U(t)x). \end{aligned}$$

The first term goes to zero by strong continuity, and for the second term notice that $U(s_n - t_n) - I$ has norm bounded by 2 and thus

$$||(U(s_n - t_n) - I)(U(t_n)x - U(t)x)|| \leq 2||U(t_n)x - U(t)x||,$$

and the right-hand side goes to 0 again by strong continuity. Thus the claim is proved.

So now if we let $f : \mathbb{R} \to \mathbb{R}$ be any smooth function with compact support (we can build such real-valued functions), then using the same process as for the construction of Riemann integrals we may define the quantity

$$\left(\int_{-\infty}^{\infty} f(t)U(t)dt \right) x$$

for any x. Specifically if f vanishes outside an interval $[a, b]$, we may take a partition $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$, and we define

$$I_{\mathcal{P}}(f, x) = \sum_{j=0}^{n-1} (t_{j+1} - t_j) f(t_j) U(t_j) x,$$

which is some element of the Hilbert space. By the same reasoning as for Riemann integrals, $I_{\mathcal{P}}(f, x)$ does have a limit as the mesh size (the maximum of $t_{j+1} - t_j$) goes to 0, since $t \mapsto U(t_j)x$ is a **uniformly** continuous function on bounded intervals.

So now if we fix some $x \in \mathcal{H}$, we define

$$x_f = \int_{-\infty}^{\infty} f(t)U(t)x dt,$$

and we claim that this "smoothed point" x_f is in $\mathcal{D}$. Indeed, we may show by retracing the steps of the calculation above that

$$U(s)x_f - x_f = \int_{-\infty}^{\infty} f(t)(U(s) - I)U(t)x dt,$$

and this now simplifies to

$$\begin{aligned} \int_{-\infty}^{\infty} f(t)U(s+t)x dt - \int_{-\infty}^{\infty} f(t)U(t)x dt &= \int_{-\infty}^{\infty} f(t-s)U(t)x dt - \int_{-\infty}^{\infty} f(t)U(t)x dt \\ &= \int_{-\infty}^{\infty} (f(t-s) - f(t))U(t)x dt. \end{aligned}$$

This means that

$$\frac{U(s)x_f - x_f}{-is} = \int_{-\infty}^{\infty} \frac{f(t-s) - f(t)}{-is} U(t)x dt,$$

and after taking limits, while we cannot apply the dominated convergence theorem we see that $\frac{f(t-s)-f(t)}{-is}$ is nice enough to move the limit inside the integral, since for all compactly supported g we have the estimate $||x_g|| \leq \int_{-\infty}^{\infty} |g(t)| \cdot ||U(t)x|| dt$. Thus the limit works out to

$$\lim_{s \to 0} \frac{U(s)x_f - x_f}{-is} = -i \int_{-\infty}^{\infty} f'(t)U(t)x dt,$$

meaning that $x_f \in \mathcal{D}$. And now we let our smooth functions approach the Dirac mass at 0 by letting f be a smooth

compactly supported nonnegative function integrating to 1 and defining $f_n(t) = nf(nt)$; we then see that $x_{f_n} \to x$ as $n \to \infty$. □

This is the first step (and the easiest) of Stone's theorem; next lecture we'll work out what this Hamiltonian looks like explicitly for the free particle, and then we'll move towards the proof and general statement.

17 October 30, 2024

Fact 104

We'll start off with a correction to Example 96 – it's true that we want to define $\psi(t,x)$ to be the numerator $\int_{\gamma:\gamma(t)=x} \psi_0(\gamma(0)) \exp\left(\frac{i}{\hbar}S(\gamma)\right) \mathcal{D}\gamma$ divided by some normalizing constant Z, but this constant Z **should depend only on** t. The point though is that the choice of the constant is immaterial, since the state of a quantum system isn't described by an element of a Hilbert space but by a "ray" of the form $\{\alpha x : \alpha \in \mathbb{C} \setminus 0\}$. So for any two elements x, y with $x = \alpha y$ for nonzero $\alpha \in \mathbb{C}$, we in fact have x and y representing the same state. So the normalization factor doesn't actually matter because we get the same evolution anyway, and indeed two equivalent vectors give the same probability measure.

Note however that two states which are not constant multiples of each other may still have the same probability measure (for example if you multiply by a phase which is different at different values of x).

We typically want to choose Z so that the evolution is norm-preserving – it doesn't have to be, but remember that for the free particle we chose a particular one so that we do get this property (so that the numerator and denominator canceled out). And we can always make such a choice for a general system.

Turning now to a general quantum system, suppose we have a strongly continuous one-parameter group of unitary operators on a Hilbert space $\mathcal{H}$. Let's be a bit more precise about what "unitary" means:

Definition 105

Let $U : \mathcal{H} \to \mathcal{H}$ be a linear operator. We say that U is **unitary** if it preserves the inner product, meaning that $\langle Ux, Uy\rangle = \langle x, y\rangle$ for all x, y, and additionally U is surjective. Equivalently, we may write that $U^*U = UU^* = I$.

This surjective assumption is necessary in the infinite-dimensional case, and it corresponds to also requiring that $UU^* = I$ in addition to $U^*U = I$. For example, consider the set of sequences ℓ^2. Then the right shift operator $(a_1, a_2, a_3, \cdots) \mapsto (0, a_1, a_2, \cdots)$ preserves inner products but is not surjective, so it's not unitary – we just say that it's an **isometry** if only the former is satisfied.

Last time, we showed that if we define

$$\mathcal{D} = \left\{ x : \lim_{t\to 0} \frac{U(t)x - x}{it} \text{ exists} \right\},$$

then $\mathcal{D}$ is dense and we can define H on this dense subset $\mathcal{D}$. For any $\boxed{x \in \mathcal{D}}$, we can then define $x(t) = U(t)x$ for all $t \in \mathbb{R}$, and notice that we may write the following quotient in two different ways:

$$\frac{x(t+s) - x(t)}{s} = \frac{U(t+s)x - U(t)x}{s} = \frac{U(s)U(t)x - U(t)x}{s} = \frac{U(t)(U(s)x - x)}{s}.$$

But the last expression here converges to $-iU(t)Hx$ as $s \to 0$ because $U(t)$ is a bounded linear operator, and so as $s \to 0$ the third expression also converges to some limit $-iHU(t)x = -iHx(t)$; this means that $\boxed{U(t)x \in \mathcal{D}}$ as well.

Therefore we have (equating the left and right sides as $s \to 0$)

$$\frac{d}{dt}x(t) = -iHx(t) \implies \boxed{i\frac{dx}{dt} = Hx(t)}.$$

Example 106

Let's analyze this equation for a free particle: recall that at a point p, we had $\widehat{U(t)\psi}(p) = \hat{\psi}(p)e^{-ip^2\hbar t/2}$.

In particular, this means that if we write $\psi_t = U(t)\psi$, then for any $\psi \in \mathcal{F}$ (recall that this is a **subset** of the domain $\mathcal{D}$, where the Fourier transforms $\hat{\psi}$ of the functions are in $L^1 \cap L^2$ – actually, this is how we prove that it is a subset) we have

$$\psi_t(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}(p)e^{-ip^2\hbar t/2}e^{ipx}dp.$$

By the dominated convergence theorem, we then have (we take pointwise derivatives but it'll turn out to be the same as derivative in L^2)

$$H\psi(x) = \frac{1}{-i}\frac{\partial}{\partial t}\psi_t(x)\Big|_{t=0} = \frac{1}{-i}\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}(p)e^{-ip^2\hbar t/2}e^{ipx}\left(-\frac{ip^2\hbar}{2}\right)dp.$$

But this extra factor we've gotten would also be what we got if we took two x-derivatives), so in fact

$$H\psi(x) = -\frac{\hbar}{2}\frac{\partial^2}{\partial x^2}\left(\frac{1}{-i}\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}(p)e^{-ip^2\hbar t/2}e^{ipx}dp\right) = -\frac{\hbar}{2}\frac{\partial^2\psi}{\partial x^2},$$

leading us to the usual Schrodinger equation with mass 1

$$\boxed{i\frac{\partial\psi}{\partial t} = -\frac{\hbar}{2}\frac{\partial^2\psi}{\partial x^2}}.$$

Note that this equation makes sense only for a subset of $\mathcal{H}$, but we can extend the **actual** evolution $\psi_t = U(t)\psi$ via the isometry to the full Hilbert space.

Example 107

This equation does become more complicated if we add a potential V or have multiple particles. For example, if we have N particles in one dimension in a potential field V with mass 1 (if we wanted to have particles in three dimensions, we can just treat each coordinate separately as its own particle), then our Hilbert space is now $\mathcal{H} = L^2(\mathbb{R}^N)$, and the Schrodinger equation now reads

$$i\frac{\partial\psi}{\partial t} = -\hbar\Delta\psi + V(x)\psi(x).$$

In a vague sense, we can think of fields as being made up of infinitely many particles, where the height is moving along the real line at each point. But the path integral calculation will be hard – we had to do an exact computation to deal with the oscillatory integral, and while we can do this again if V is for example quadratic (the harmonic oscillator), in general solutions may not exist. So it's difficult to answer the question of "when do solutions to the Schrodinger equation exist" and it's hard to explain how we get from the path integral formulation in general, as well as when we can extend things to all of L^2. This all lies at the heart of quantum mechanics!

But to proceed, we'll need to learn some functional analysis and then some probability. And the things from the beginning of the class will come back eventually, too.

Definition 108

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{C}$ with inner product $\langle x, y\rangle$ (which is linear in the first argument and complex linear in the second argument; note that this convention is usually flipped in physics and this would be denoted $\langle y|x\rangle$ instead). A **densely defined linear operator A on $\mathcal{H}$ with domain $\mathcal{D}(A)$** is a linear map $A : \mathcal{D}(A) \to \mathcal{H}$, where $\mathcal{D}(A)$ is a dense subspace of $\mathcal{H}$; note that we do not assume continuity.

For example, we can define the "second derivative operator" on $L^2(\mathbb{R})$ on a dense subspace (so that the right-hand side of the Schrodinger equation makes sense); specifically, this is the set of twice continuous, differentiable functions in L^2 whose second derivative is also in L^2. We'll now describe how to define the adjoint of such an operator:

Definition 109

For A a densely defined linear operator as above, its adjoint A^* is defined as follows: let $\mathcal{D}(A^*)$ be the set of all $x \in \mathcal{H}$ such that there is some $z \in \mathcal{H}$ satisfying $\langle Ay, x\rangle = \langle y, z\rangle$ for all $y \in \mathcal{D}$. Then we define $A^*x = z$.

We can check that this operator is indeed well-defined (so z is unique) and linear, and $\mathcal{D}(A^*)$ is a subspace (though it's not clear if it's a dense subspace).

Definition 110

A densely defined linear operator A is **symmetric** or **Hermitian** (for operators these are the same thing) if for all $x, y \in \mathcal{D}(A)$, we have $\langle Ax, y\rangle = \langle x, Ay\rangle$.

Example 111

Let $\mathcal{H} = L^2(\mathbb{R})$, let $\mathcal{D}(A)$ be the set of all compactly supported smooth functions, and let $A\psi = \psi''$ be the second derivative operator. Then A is symmetric by integration by parts (since $\langle A\psi, \phi\rangle = \int \psi''\phi = -\int \psi'\phi = \int \psi\phi'' = \langle \psi, A\phi\rangle$ for compactly supported smooth functions ψ, ϕ).

Lemma 112

If A is symmetric, then $\mathcal{D}(A) \subseteq \mathcal{D}(A^*)$, A^* is equal to A when restricted to $\mathcal{D}A()$, and so in particular A^* is also densely defined.

Proof. Suppose A is symmetric. Then for any $x \in \mathcal{D}(A)$ we have for all $y \in \mathcal{D}(A)$ that $\langle Ay, x\rangle = \langle y, Ax\rangle$ and thus $x \in \mathcal{D}(A^*)$ with $A^*x = Ax$. □

Definition 113

A densely defined operator A on $\mathcal{H}$ is **self-adjoint** if it is symmetric and **additionally** $\mathcal{D}(A^*) = \mathcal{D}(A)$.

So if we start with the example above where A is the second derivative operator, then the domain of A^* is actually bigger than that of A; we only say that we're self-adjoint if we cannot get anything bigger. The question of "how large we can make the domain before we stop getting more things" is a difficult one in general! (Though if A is defined on all of $\mathcal{H}$, symmetry does imply self-adjointness.)

Remark 114. *Notice that we still aren't imposing continuity, and in general the operators we will discuss will not be continuous operators. For example, there's no way we can guarantee that if f_n converges in L^2 that the second derivatives do as well.*

Next time we'll prove a basic criterion for self-adjointness which is more useful than this definition, and then we'll show that the spectrum of such an operator is always real (which is important for the mass gap problem)

18 November 1, 2024

Today, we'll start with some preparation towards proving criteria for self-adjointness:

Definition 115

Let $\mathcal{H}$ be a Hilbert space, and let $S \subseteq \mathcal{H}$ be any subset (not necessarily a subspace). The **orthogonal complement** of S is the set

$$S^\perp = \{x \in \mathcal{H} : \langle x, y\rangle \quad \forall y \in S\}.$$

We can check that $S^\perp$ is always a subspace, and in fact it's always a closed subspace (by continuity of the inner product).

Lemma 116

Let A be a linear operator on $\mathcal{H}$ with domain $\mathcal{D}(A)$ (not necessarily dense). Then $\text{range}(A)^\perp = \ker(A^*)$.

Proof. We just verify inclusions on both sides. If $x \in \text{range}(A)^\perp$, then $\langle Ay, x\rangle = 0$ for all $y \in \mathcal{D}(A)$, but because $0 = \langle y, 0\rangle$, this means that x is in the domain of A^* with $A^*x = 0$ by definition (recall Definition 109).

Conversely, if $x \in \ker(A^*)$ (in particular meaning that $x \in \mathcal{D}(A^*)$), then $A^*x = 0$. Therefore $\langle y, A^*x\rangle = 0$ for all $y \in \mathcal{D}(A)$, and this quantity is also $\langle Ay, x\rangle$ and thus x is in the orthogonal complement of $\text{range}(A)$, as desired. □

Definition 117

Let A be a (linear) operator defined on a domain (which is always a subspace) $\mathcal{D}(A) \subseteq \mathcal{H}$. We say that A is **closed** if its **graph** $\mathcal{G}(A) = \{(x, Ax) : x \in \mathcal{D}(A)\}$ is a closed subset of $\mathcal{H} \times \mathcal{H}$ (under the product topology).

Lemma 118

The adjoint A^* of any operator A is a closed operator.

Proof. Define an inner product on $\mathcal{H} \times \mathcal{H}$ via

$$\langle (x, y), (w, z)\rangle = \langle x, w\rangle + \langle y, z\rangle.$$

Then we can check that $\mathcal{H} \times \mathcal{H}$ is a Hilbert space under this inner product, and the topology under that inner product is exactly the product topology. Defining $S = \{(-Ax, x) : x \in \mathcal{D}(A)\}$, it thus suffices to prove that the graph $\mathcal{G}(A^*)$ is the orthogonal complement of S (which means it is closed).

To prove that claim, notice that a point (x, y) belongs to $\mathcal{G}(A^*)$ if and only if $x \in \mathcal{D}(A^*)$ and $y = A^*x$, which occurs if and only if $\langle Az, x\rangle = \langle z, y\rangle$ for all $z \in \mathcal{D}(A)$ (by the definition of the adjoint). This occurs if and only if

$0 = -\langle x, Az\rangle + \langle y, z\rangle$ for all $z \in \mathcal{D}(A)$, which occurs if and only if $\langle (x,y), (-Az, z)\rangle = 0$ for all $z \in \mathcal{D}(A)$; this occurs if and only if $(x,y) \in S^\perp$ (since the latter argument in the inner product is an arbitrary element of S). □

We now come to the main theorem used to prove Stone's theorem (among other things):

Theorem 119

Let A be a densely defined symmetric linear operator on a Hilbert space $\mathcal{H}$. Then the following are equivalent:

1. A is self-adjoint,
2. $\ker(A^* \pm iI) = \{0\}$ and A is closed,
3. $\text{range}(A \pm iI) = \mathcal{H}$.

(Recall that for a symmetric linear operator $A = A^*$ on $\mathcal{D}(A)$, but it's possible that A^* has a larger domain than A.) Here $A^* + iI$ and $A^* - iI$ are operators from $\mathcal{D}(A^*) \to \mathcal{H}$ (with $(A^* + iI)x = A^*x + ix$ and $(A^* - iI)x = A^*x - ix$), and similar for $A + iI$ and $A - iI$.

This probably seems like a rather strange result – the idea is that adding a constant times iI makes our operator invertible, and we'll use this to show that self-adjoint operators have real spectrum. And it may look abstract, but (for example) this machinery and Brownian motion will make constructing solutions to the Schrodinger equation for a general potential much easier.

Proof. **For (1) implies (2)**, recall that self-adjoint A implies that $\mathcal{D}(A) = \mathcal{D}(A^*)$ and $A = A^*$. We know that A^* is always closed by Lemma 118 and thus A is closed as well.

For the other part of the claim, suppose we have some x such that $(A^* + iI)x = 0$, meaning that $A^*x = -ix$. Then

$$i||x||^2 = i\langle x, x\rangle = \langle x, -ix\rangle = \langle x, A^*x\rangle = \langle Ax, x\rangle$$

(last step because x is in the domain of A^*, but that's the same as the domain of A so this is all defined). But now this last expression is $\langle A^*x, x\rangle$ because $A = A^*$, and thus it also simplifies to $\langle -ix, x\rangle = -i||x||^2$; hence $||x|| = 0$ and $x = 0$ so that the kernel is trivial. The argument for $A^* - iI$ is similar.

For (2) implies (3), by Lemma 116 we know that $\text{range}(A + iI)^\perp = \ker(A^* - iI)$ (since the adjoint of iI is $-iI$). But we assumed that the right-hand side is trivial, which implies that $\text{range}(A + iI)^\perp$ is dense in $\mathcal{H}$. To show that it is in fact all of $\mathcal{H}$, it suffices to show that $\text{range}(A + iI)$ is closed. But since A is a closed operator by assumption, the set of points $\{(x, Ax) : x \in \mathcal{D}(A)\}$ is closed in $\mathcal{H} \times \mathcal{H}$. Supposing we have a sequence of points x_n such that $(A + iI)x_n \to y$, we will show that x_n is Cauchy; indeed,

$$||(A + iI)x_n - (A + iI)x_m||^2 = ||(A + iI)(x_m - x_n)||^2 = ||A(x_n - x_m)||^2 + ||(x_n - x_m)||^2,$$

since for any z we have $||(A+iI)z||^2 = ||Az||^2 + ||z||^2 + \langle Az, iz\rangle + \langle iz, Az\rangle$, and the last two terms cancel out because $\langle Az, iz\rangle = -i\langle Az, z\rangle = -i\langle z, Az\rangle = -\langle iz, Az\rangle$, using that A is symmetric. Since the left-hand side approaches zero, we must have $x_n \to x$ and $Ax_n \to z$, and therefore we must have $z = Ax$ because the graph is closed. So $y = (A+iI)x$ and thus $\text{range}(A + iI)$ is closed; therefore it is all of $\mathcal{H}$. And again the argument for $A - iI$ is similar.

Finally, **for (3) implies (1)**, we just need to show that $\mathcal{D}(A^*) \subseteq \mathcal{D}(A)$. Indeed, take any $x \in \mathcal{D}(A^*)$ and let $y = A^*x - ix$. By assumption, there is some $z \in \mathcal{D}(A)$ with $y = Az - iz$. Since $z \in \mathcal{D}(A)$, we have $Az = A^*z$, and thus

$$A^*x - ix = A^*z - iz \implies x - z \in \ker(A^* - iI).$$

But $\ker(A^* - iI) = \ker((A+iI)^*)$ is the orthogonal complement of $\text{range}(A+iI)$, so it is trivial; thus $x = z$ and therefore x must be in $\mathcal{D}(A)$ as well. □

Fact 120

For the one-dimensional free particle with $H = -\frac{\hbar}{2}\frac{d^2}{dx^2}$, the domain of H turns out to be

$$\mathcal{D}(H) = \left\{\psi \in L^2(\mathbb{R}) : \int_{-\infty}^{\infty} p^4 |\hat{\psi}(p)|^2 < \infty\right\}.$$

H then acts on ψ via $\widehat{H\psi}(p) = \frac{\hbar p^2}{2}\hat{\psi}(p)$.

We won't go through the details of this during lecture, but we do need it because we'll construct the free particle Hamiltonian in a different way using probability and Stone's theorem later on, and we do need to know that we get the same generator.

Definition 121

Let A be a densely defined operator on a Hilbert space $\mathcal{H}$. We say that A is **invertible** if the map $A : \mathcal{D}(A) \to \mathcal{H}$ is a bijection. (It can be checked that for any invertible A, the inverse $A^{-1} : \mathcal{H} \to \mathcal{D}(A)$ is always a linear operator as well.)

An element $\alpha \in \mathbb{C}$ is in the **resolvent set** $\rho(A)$ if $\alpha I - A$ is invertible and the inverse map $R(\alpha) = (\alpha I - A)^{-1}$ is bounded. (We call $R(\alpha)$ the **resolvent of A at α**.) The **spectrum** of A is the set $\sigma(A) = \mathbb{C} \setminus \rho(A)$.

Even though we're not imposing any continuity on A, by definition we have that for all $\alpha \in \rho(A)$, $R(\alpha) : \mathcal{H} \to \mathcal{D}(A)$ is a bounded (hence continuous) linear operator.

Theorem 122

Let A be a densely defined self-adjoint operator. Then $\sigma(A) \subseteq \mathbb{R}$, and for all $\alpha \in \mathbb{C} \setminus \mathbb{R}$ we have the quantitative bound $||R(\alpha)|| \leq \frac{1}{|\text{Im}(\alpha)|}$.

In particular, if A is self-adjoint, then $\alpha I - A$ is invertible for any nonreal α and we get a bound on its norm, even though we have no continuity in the definition of the operator (for example for the second derivative operator above). We'll prove this next time!

19 November 4, 2024

We'll start today by proving the theorem we stated last time:

Proof of Theorem 122. Suppose $\lambda \in \mathbb{R} \setminus \{0\}$. Then if A is self-adjoint, $\lambda^{-1}A$ is self-adjoint as well (directly from the definition), and therefore by Theorem 119, $\ker((\lambda^{-1}A)^* + iI)$ is trivial and $\text{range}(\lambda^{-1}A + iI)$ is all of $\mathcal{H}$, so by rescaling this is the same as saying that $\text{range}(A + i\lambda I)$ is all of $\mathcal{H}$. But since λ^{-1} is real and A is self-adjoint, $(\lambda^{-1}A)^*$ is just $\lambda^{-1}A$. Thus we also have that $\ker(\lambda^{-1}A + iI)$ is trivial and thus $\ker(A + i\lambda I)$ is trivial; this means $A + i\lambda I : \mathcal{D}(A) \to \mathcal{H}$ is a bijection (since it's both an injection and a surjection).

But now again from the definition, if A is self-adjoint so is $A + \mu I$ for any real number μ. Thus for all $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R} \setminus \{0\}$, $A + (\mu + i\lambda)I$ is a bijection between $\mathcal{D}(A)$ and $\mathcal{H}$. Now to show that all such $\alpha = \mu + i\lambda$ are in the

resolvent set, we just need to show that the inverse map is bounded. For this, fix any such α and note that for any $x \in \mathcal{D}(A)$ we have (by expanding out the inner product)

$$\begin{aligned}\boxed{||(\alpha I - A)x||^2} &= |\alpha|^2||x||^2 - \alpha\langle x, Ax\rangle - \overline{\alpha}\langle Ax, x\rangle + ||Ax||^2 \\ &= (\mu^2 + \lambda^2)||x||^2 - 2\mu\langle x, Ax\rangle + ||Ax||^2 \\ &\geq (\mu^2 + \lambda^2)||x||^2 - 2|\mu|\,||x||\,||Ax|| + ||Ax||^2\end{aligned}$$

by Cauchy-Schwarz. But this last expression is equal to $\lambda||x||^2 + (|\mu|||x|| - ||Ax||)^2 \geq \boxed{\lambda^2||x||^2}$, and this inequality is exactly what we need to finish the proof. Indeed, for any $y \in \mathcal{H}$, we have $x = R(\alpha)y = (\alpha I - A)^{-1}y$ (since we've already proven that the inverse is defined everywhere) and the above inequality can be rewritten as $||y||^2 \geq \lambda^2||R(\alpha)y||^2 \implies ||R(\alpha)y|| \leq \frac{1}{|\lambda|}||y||$. So $||R(\alpha)|| \leq \frac{1}{|\lambda|}$, proving the quantitative bound and that the inverse map is indeed bounded. □

We're now ready for the main result of this "chapter," which gives us a one-to-one correspondence between generators and strongly continuous one-parameter groups:

Theorem 123 (Stone)

Let $(U(t))_{t\in\mathbb{R}}$ be a strongly continuous one-parameter group of unitary operators on a Hilbert space $\mathcal{H}$, and define the domain

$$\mathcal{D} = \left\{x \in \mathcal{H} : \lim_{t\to 0} \frac{U(t)x - x}{it} \text{ exists}\right\};$$

on this set $\mathcal{D}$, call that limit Hx. Then $\mathcal{D}$ is dense in $\mathcal{H}$ (we proved this already in Lemma 103), and H (with domain $\mathcal{D}$) is a self-adjoint operator. Conversely, for any self-adjoint operator $H : \mathcal{D}(H) \to \mathcal{H}$ defined on a dense subspace, there exists a unique strongly continuous unitary group $(U(t))_{t\in\mathbb{R}}$ such that H arises from it in this manner (with $\mathcal{D} = \mathcal{D}(H)$).

Since any quantum system is a group of this form, this is a foundational result for quantum mechanics. One problem, though, is that the most important systems in quantum mechanics are quantum field theories (used in the Standard Model), and except for the toy models like the free field or ϕ^4 fields (in three dimensions), we don't even know what the Hilbert space is and thus can't even get started on building these operators. And even the ϕ^4 theory is quite deep and the Hilbert space isn't even explicitly described; it's done via a complicated process (which we'll see later on in our course).

Start of proof. **For the forward direction**, it just remains (Lemma 103) to prove that H is self-adjoint. For this, we must first show that H is symmetric and then that one of the equivalent criteria in Theorem 119 also holds. We've actually already proved in some sense that H is symmetric – for any $x, y \in \mathcal{D}$ (since everything is in the domain there's no issue with any of the following quantities being defined), we have

$$\begin{aligned}\langle Hx, y\rangle &= \left\langle \lim_{t\to 0} \frac{U(t)x - x}{-it}, y \right\rangle \\ &= \lim_{t\to 0} \left\langle \frac{U(t)x - x}{-it}, y \right\rangle,\end{aligned}$$

since the map $z \mapsto \langle z, y\rangle$ is continuous and thus we can interchange the limit and the inner product. This further

simplifies to

$$\lim_{t\to 0}\frac{1}{-it}\left(\langle U(t)x,y\rangle-\langle x,y\rangle\right)=\lim_{t\to 0}\frac{1}{-it}\left(\langle x,U(t)^*y\rangle-\langle x,y\rangle\right)$$
$$=\lim_{t\to 0}\frac{1}{-it}\left(\langle x,U(-t)y\rangle-\langle x,y\rangle\right)$$

since $U(t)U(-t)=U(0)=I$ (since $U(0)^2=U(0)$ and then we can multiply both sides by $U(0)^{-1}$) and thus $U(t)^*=U(t)^{-1}=U(-t)$. But now we can bring the limit back inside the inner product now in the other argument (now using conjugate linearity), and this simplifies to $\left\langle x,\lim_{t\to 0}\frac{U(-t)y-y)}{+it}\right\rangle=\langle x,Hy\rangle$, as desired.

We'll now show that range$(H\pm iI)=\mathcal{H}$; this is easiest to do by producing a right inverse. Using Riemann sums in the same way as before, we may define the operator

$$R=\int_0^\infty e^{-t}U(t)dt$$

(which we'll show is a map from $\mathcal{H}$ into $\mathcal{D}$) and we claim that in fact $(H-iI)(iR)x=x$ for all $x\in\mathcal{H}$. Indeed, for any nonzero s, we can move the operator inside the integral and use the group property

$$\begin{aligned}(U(s)-I)Rx&=\int_0^\infty e^{-t}(U(s)-I)U(t)x\\&=\int_0^\infty e^{-t}U(s+t)xdt-\int_0^\infty e^{-t}U(t)xdt\\&=\int_s^\infty e^{-(t-s)}U(t)xdt-\int_0^\infty e^{-t}U(t)xdt\\&=e^s\int_s^\infty e^{-t}U(t)xdt-\int_0^\infty e^{-t}U(t)xdt\\&=(e^s-1)\int_s^\infty e^{-t}U(t)xdt-\int_0^s e^{-t}U(t)xdt\\&=(e^s-1)\left[Rx-\int_0^s e^{-t}U(t)xdt\right]-\int_0^s e^{-t}U(t)xdt.\end{aligned}$$

We now want to divide both sides by $-is$. This yields

$$\frac{(U(s)-I)Rx}{-is}=\frac{e^s-1}{-is}\left[Rx-\int_0^s e^{-t}U(t)xdt\right]-\frac{\int_0^s e^{-t}U(t)xdt}{-is},$$

and as $s\to 0$, $\frac{e^s-1}{-is}$ approaches i and the blue part goes to 0 because its norm is bounded by $|s|$. Finally, the red term is $sx+o(s)$ because we may write it as $sx+\int_0^s e^{-t}(U(t)x-x)dt+\int_0^s(e^{-t}-1)xdt$; the latter integral can be explicitly computed, and by strong continuity $U(t)x\to x$ as $t\to 0$ so the former integral is bounded by $sC(s)$ for some $C(s)\to 0$ as $s\to 0$. Therefore

$$HRx=\lim_{s\to 0}\frac{(U(s)-I)Rx}{-is}=iRx-ix,$$

so in particular this shows that Rx is in the domain because the left-hand side exists as $s\to 0$. And rearranging the terms we get $(H-iI)Rx=-ix$ and thus $(H-iI)(iR)x=x$, as desired. And to do a similar proof for $(H+iI)$, we let $S=\int_0^\infty e^{-t}U(-t)dt$ and show that $(H+iI)(-iS)x=x$; thus we've verified the conditions required for self-adjointness and completed one direction of the proof.

The converse direction is more difficult: let H be a self-adjoint operator on $\mathcal{H}$ with dense domain $\mathcal{D}$. We must construct the group of unitary operators with H as its generator; we can use the spectral theorem but that would require proving it, and it's generally better to do a direct proof. In particular, if we want to later understand the solutions of the Schrodinger equation, we'll use probability to construct H and the converse direction will be used to

generate the solution. The idea is that for any n, the map

$$T_n = \left(I + \frac{iH}{n}\right)^{-1} = -in((-in)I - H)^{-1}$$

exists (because this is the resolvent at $-in$) with $||T_n|| \leq n \cdot \frac{1}{|-in|} = 1$ by our quantitative bound, and we'll define an approximation to the unitary group via, for any $t \geq 0$,

$$U_n(t) = e^{nt(T_n - I)} = \sum_{k=0}^{\infty} \frac{(nt)^k}{k!}(T_n - I)^k = e^{-nt}\sum_{k=0}^{\infty}\frac{(nt)^k}{k!}T_n^k.$$

This makes sense because the exponential of a bounded operator is always well-defined pointwise, and in fact the norm of U_n is bounded by $e^{-nt}\sum_{k=0}^{\infty}\frac{(nt)^k}{k!} = 1$. So these operators are defined everywhere, and we will obtain $U(t)$ as the limit $\lim_{n\to\infty} U_n(t)$. Indeed,

$$U_n(t) = e^{nt(T_n - I)} = \exp\left(nt\left(\left(I + \frac{iH}{n}\right)^{-1} - I\right)\right)$$

and then we can do a Taylor expansion and show that it approaches e^{itH}; the key point is that this particular approximation is exactly the right one to get us a bounded operator. □

20 November 6, 2024

We'll continue the proof of Stone's theorem today:

Continuation of proof of Theorem 123. Recall that our goal (in progress last time) was to define the operator $U(t) = e^{-itH}$ and thus get a one-parameter family of strongly continuous operators. We will do this by first defining the approximation $U_n(t) = e^{nt(T_n - I)}$ and then obtaining U as the pointwise limit of U_n as $n \to \infty$. First, we'll show a preparatory lemma:

Lemma 124

For any m, n, we have $T_m T_n = T_n T_m$, and in fact T_n maps into $\mathcal{D}$ for all n (meaning it is a bijection between $\mathcal{H}$ and $\mathcal{D}$).

Proof of lemma. The iea is that "the operators commute without the inverse, so they do with the inverse as well." Let $y = T_n x$ and $z = T_m y$. Then $x = (I + \frac{i}{n}H)y$ and $y = (I + \frac{i}{m}H)z$, so

$$\begin{aligned} x &= \left(I + \frac{i}{n}H\right)\left(I + \frac{i}{m}H\right)z \\ &= \left(I + \frac{i}{m}H\right)\left(I + \frac{i}{n}H\right)z \end{aligned}$$

which means that (taking inverses) $T_m T_n x = z = T_n T_m x$, as desired. And the latter claim then follows just by the definition of H. □

Since we know that T_n maps into $\mathcal{D}$, it makes sense to apply H afterward, and we then have the following other result (which is proved in exactly the same way as the previous one, showing that the order can be interchanged):

Lemma 125

T_n commutes with H for all n.

This then allows us to study our approximations U_n and how they interact with our H operators:

Lemma 126

$U_n(t)$ commutes with T_m and H for any m, n, t. Moreover $U_n(t)$ maps $\mathcal{D}$ into $\mathcal{D}$, and $U_n(t) - I$ maps all of $\mathcal{H}$ into the domain $\mathcal{D}$ (though we won't actually need this latter fact).

Proof of lemma. Since H is self-adjoint, recall that it is closed (by Theorem 119), which will be useful when we need to make arguments involving continuity. Commutativity with T_m is trivial because our operators are all norm-bounded by 1 and thus we can exchange the commuting T_ms and T_ns. By definition, we have

$$U_n(t) = e^{-nt} \sum_{k=0}^{\infty} \frac{(nt)^k}{k!} T_n^k,$$

and for any $x \in \mathcal{D}$ we have

$$\begin{aligned} U_n(t)Hx &= e^{-nt} \sum_{k=0}^{\infty} \frac{(nt)^k}{k!} T_n^k Hx \\ &= e^{-nt} \sum_{k=0}^{\infty} \frac{(nt)^k}{k!} H T_n^k x \end{aligned}$$

by the commutativity of T_n and H (though since H is not continuous, we can't bring it completely out of the sum). To understand what this last expression means, we can write it as the limit of partial sums $\lim_{m\to\infty} Hy_m$, where $y_m = e^{-nt} \sum_{k=0}^{m} \frac{(nt)^k}{k!} T_n^k x$ (for any finite partial sum we can indeed bring the H outside). But as $m \to \infty$ we have $y_m \to U_n(t)x$, and $Hy_m \to U_n(t)Hx$, so by closedness of the operator this means $HU_n(t)x = U_n(t)Hx$, showing commutativity with H on $\mathcal{D}$.

To extend this argument to all $x \in \mathcal{H}$, we can define y_m as above again and note that $y_m - x \in \mathcal{D}$ for all m (even though T_n maps into the domain, we have to exclude the $k = 0$ term in the sum). Then repeating the argument, we see that $H(y_m - x)$ converges to $U_n(t)Hx - Hx$, while $y_m - x$ converges to $U_n(t)x - x$. And again since H is a closed operator, we have that $U_n(t) - I$ maps from $\mathcal{H} \to \mathcal{D}$ and that it commutes with H. □

Returning to Stone's theorem now, we can rewrite

$$\begin{aligned} T_n\left(I + \frac{i}{n}H\right) = I \implies T_n\left(I + \frac{i}{n}H\right)x = x \quad \forall x \in \mathcal{D} \\ \implies \boxed{T_nHx = in(T_n - I)x} \quad \forall x \in \mathcal{D}. \end{aligned}$$

Thus for all $x \in \mathcal{D}$, $||T_nx - x|| = \frac{1}{n}||T_nHx|| \leq \frac{1}{n}||Hx||$ (because T_n has norm bounded by 1), and as $n \to \infty$ the right-hand side goes to zero. Thus $T_nx \to x$ as $n \to \infty$, and we can even strengthen this past the domain $\mathcal{D}$. Specifically, for any $x \in \mathcal{H}$ and any $\varepsilon > 0$, we can find $y \in \mathcal{D}$ with $||y - x|| < \varepsilon$, and

$$\limsup ||T_nx - x|| \leq \limsup ||T_nx - T_ny|| + \limsup ||T_ny - y|| + ||y - x|| \leq 2||x - y|| + 0 = 2\varepsilon.$$

Since ε is arbitrary this then shows that $T_nx \to x$ for all x. (This shouldn't be surprising, since T_n is a small perturbation of the identity.) So the main remaining thing is to take any $x \in \mathcal{D}$ and look at the difference

$$U_m(t)x - U_n(t)x = \int_0^t \frac{d}{ds}\left(U_m(t-s)U_n(s)x\right) ds$$

(since the thing in parentheses is $U_m(t)$ at $s = 0$ and $U_n(t)$ at $s = t$). Towards, computing this blue derivative, we

note that

$$\frac{d}{dt}U_n(t) = \frac{d}{dt}\left(e^{nt(T_n - I)}\right) = n(T_n - I)U_n(t)x$$

(we can move the derivative inside the infinite sum because everything is bounded), and using the boxed identity above this can be rewritten as $-iT_nHU_n(t)x$. Applying that to our blue derivative, we can work out (by taking difference quotients and taking $\varepsilon \to 0$) that

$$\int_0^t \frac{d}{ds}\left(U_m(t-s)U_n(s)x\right) ds = \int_0^t \left(iT_mHU_m(t-s)U_n(s)x - iU_m(t-s)T_nHU_n(s)x\right) ds.$$

Since x is in the domain, all of our commutation relations work out, and we can move H and T_m to the end to get overall that

$$U_n(t)x - U_m(t)x = \int_0^t iU_m(t-s)U_n(s)(T_m - T_n)Hx ds.$$

Taking norms on both sides and using the triangle inequality, we thus have (because all U_ns are norm-bounded by 1)

$$||(U_n(t) - U_m(t))x|| \leq \int_0^t ||(T_m - T_n)Hx|| ds = t||(T_m - T_n)Hx||.$$

So as $m, n \to \infty$ the right-hand side goes to zero for any fixed t, and thus $U_n(t)x$ is a Cauchy sequence and thus has a limit for every $x \in \mathcal{D}$. And in particular, there exists some $U(t)x \in \mathcal{H}$ with $U_n(t)x \to U(t)x$, and $||U_n(t)x - U(t)x||$ goes to 0 **uniformly on** $x \in \mathcal{D}$ **over** t **on any bounded interval**. So now to extend to all x in $\mathcal{H}$ is easy, since our U_ns are norm-bounded by 1, and the usual triangle inequality denseness argument shows that $U_n(t)x \to U(t)x$ for all $x \in \mathcal{H}$, as desired.

So at this point we've constructed $U(t)$, and it's easy to verify that it's a linear map and that $||U(t)|| \leq 1$. There are just some other properties left now: since $U_n(0) = I$ for all n, we also have $U(0) = I$, and since $U_n(s+t) = U_n(s)U_n(t)$ for all n, the same will be true in the limit because

$$||U_n(s+t)x - U_n(s)U_n(t)x|| = 0 \implies ||U(s+t)x - U(s)U(t)x|| = 0.$$

So we do have a one-parameter group of unitary operators, and for all $x \in \mathcal{H}$ we have strong continuity property for $t \mapsto U(t)x$ because of the bolded property above (in other words, that $\sup_{0 \leq s \leq t} ||U_n(s)x - U(s)x|| \to 0$ as $n \to \infty$ for any t) and the fact that $t \mapsto U_n(t)x$ is continuous. There are only two main things remaining to show now: we still need to show that U is unitary (because we've only defined this for $t \geq 0$ so far), and we need to show that H is in fact the generator. For this, define

$$\mathcal{D}' = \left\{x : \lim t \downarrow 0 \frac{U(t)x - x}{-it} \text{ exists}\right\},$$

and let that limit be called $H'x$ when it does exist. We wish to show that $\mathcal{D}' = \mathcal{D}$ and that $H' = H$. We'll actually finish out the rest of this proof later (after student presentations, which will take up the next two weeks), but the point is to define $U_n(-t)$ for $t > 0$ in the exact same way but with $I - \frac{i}{n}H$ instead of $I + \frac{i}{n}H$. □

21 November 11, 2024

Today is the first day of class presentations:

Stanislav Krymskii

This paper is called "Correlation decay for finite lattice gauge theories at weak coupling." Yang-Mills theories study Lie-algebra valued connections on a Riemmanian manifold and attempt to construct a measure on them. However, the set of connections is infinite-dimensional, so it's difficult to define a Lebesgue measure. Instead, the idea is to study the **lattice theory**, and another simplification we can make is to make the group G itself finite. We assign an element of the Lie group G to every edge under the reversal, and as Balaban shows, this ensures ultraviolet stability in 3 and 4 dimensions.

The paper we'll discuss focuses on **finite** pseudo-Lie groups G, but even this simplification shows some new phenomena. Discrete models with nonabelian symmetry require topology and knot theory because of the existence of "defects," and we'll discuss the calculation of expectations and (exponentially decaying) correlations.

The idea is that in the nonabelian group case, we face some difficulties. What we can try to do is to expand the partition function into clusters, where disjoint defects "appear independently," but algebraic topology will still be required.

Given a finite group G and a finite lattice Λ, we assign an element $\sigma_{xy} \in G$ to every edge so that $\sigma_{xy}\sigma_{yx} = \mathrm{id}$. We can then define a character χ and define its action to be

$$S_\Lambda(\sigma) = \sum_{p \in \Lambda_2} \mathrm{Re}(\chi(1) - \chi(\sigma_p))$$

over "plaquettes" (unit squares) p, where $\sigma_p = \sigma_{x_1x_2}\sigma_{x_2x_3}\sigma_{x_3x_4}\sigma_{x_4x_1}$. We then define $\Delta_G = \min_{g \neq 1} \mathrm{Re}(\chi(1) - \chi(\sigma_p))$, which will be a quantity of interest.

Theorem 127

Suppose f is a conjugacy-invariant function. Then if B_1, B_2 are two rectangles a distance at least L apart, then $|\operatorname{Cov}(f_1, f_2)|$ decays exponentially with L.

The idea is to construct a bijection so that $\mu^{\otimes 2}$, the joint distribution of two independent configurations, is invariant, but

$$f_1(\Sigma_1)f_2(\Sigma_2) = (f_1 f_2)(T_1(\Sigma_1, \Sigma_2)),$$

and defining this bijection on all but an exponentially-small measure set.

Here's where algebraic topology comes into play – we represent Σ as a homomorphism from $\pi_1(\Lambda_1)$ of loops to G. Given any such ψ, there are $|G|^{|\Lambda_0|-1}$ configurations that yield it, and we can thus represent homomorphisms equivalence classes as **gauge** equivalence classes. We then have a "canonical representation" of each homomorphism by fixing a particular tree Λ_1 and requiring edges along that tree to take values id.

There are some technical details that are worth noting as well:

- We say that two sets of plaquettes are well-separated by a rectangle B (that is, a 4-dimensional box) if one set is inside the 2-complexes of B, the other is in the complement, and neither intersects the boundary. The point is that any plaquette set can be separated into knots, and each plaquette only has at most $(10^{24})^m$ knots of size m containing it.
- So if two sets P_1, P_2 are well-separated, then we can find a bijection between morphisms supported at $P_1 \cup P_2$ and **pairs** of morphisms supported in P_1 and P_2, and this also splits the energy up into two summands.
- If we have lots of sets that are well-separated, we can biject into morphisms supported on separate P_is; this allows us to construct the swapping map by defining things on the set E of pairs (ψ_1, ψ_2) such that $\mathrm{supp}(\psi_1) \cup$

$\mathrm{supp}(\psi_2) \cup B_1 \cup B_2$ can be separated into well-separated knots. (The map is a measure-preserving involution and thus a bijection.)

- What's left is to check that the "bad cases" (bad homomorphisms) are rare. Indeed, each knot $K \supset P_0$ appears with small probability, and knots that render two far-away sets P_1, P_2 unsplittable are unlikely to appear because they just also include a lot of plaquettes in between them. (Here's where the fact that $\Delta_G > 0$ matters – it penalizes the appearance of such a knot.) Since our swapping map deals with any splittable knot, this is all we need to do.

Remark 128. *Note that throughout this, β is taken to be large, and we'll need to estimate a partition function to estimate probabilities. (The "partition function entropy term" does not cause significant contributions.)*

Fred Rajasekaran

This is the first of three student presentations on the paper "Random surfaces and lattice Yang-Mills" by Cao, Park, and Sheffield. We'll start by introducing and motivating the model, defining lattice gauge theories and why they're important, and explaining what Wilson loops are and why we care.

We'll begin with Euclidean Yang-Mills: in class we've been talking a lot about quantum Yang-Mills with the Minkowski metric, but that's difficult to do because of these oscillatory terms. So often we'll transform to Euclidean Yang-Mills, and we do so via something called a "Wick rotation." The heuristic idea of that is as follows: given a quantum Yang-Mills theory with the Minkowski metric $\mathbb{R}^{1,3}$, we get a corresponding Euclidean theory with the standard metric $\mathbb{R}^4$, and it's called a "rotation" in the sense that we send time t to "imaginary time" $-i\tau$ to deal with the negative sign in front of the metric. This formalism then turns oscillatory integrals into exponentially decaying ones, which should remind us of Gibbs measures; we can thus convert quantum mechanics into statistical mechanics. Then we can get back to the quantum theory if we can show that the Euclidean theory satisfies certain axioms, and thus it often makes sense just to work with the Euclidean case.

In our paper, we consider the gauge group $G = U(N)$, and let $\mathcal{A}$ be a gauge field (that is, a skew-symmetric matrix-valued 1-form). The Euclidean Yang-Mills action is then given by (typically we want $d = 4$, but we can do any dimension)

$$S_{YM}(A) = -\int_{\mathbb{R}^d} \sum_{i,j=1}^{d} \mathrm{Tr}(F_{ij}(x)^2)dx,$$

and the associated theory is the probability measure (ill-defined, but we'll talk about how to make it make sense with a discretization)

$$d\mu(\mathcal{A}) = \frac{1}{Z} \exp\left(-\frac{1}{4g^2} S_{YM}(\mathcal{A}) d\mathcal{A}\right).$$

Here g is the **coupling strength** (just some parameter) and $d\mathcal{A}$ is an "infinite-dimensional Lebesgue measure" which assigns to every point in $\mathbb{R}^d$ a unitary matrix according to the Haar measure. This $d\mathcal{A}$ doesn't actually make sense, but we can discretize it so that it does make sense.

Remark 129. *There are ways to define uncountable products of measures, but somehow it's very hard to make those things work here because Z isn't summable and other issues come up too.*

Going now from Euclidean to lattice Yang-Mills, here's how the model works:

- We work on a finite lattice $\Lambda \in \mathbb{Z}^d$ so that computations are finite. Sometimes we can send the lattice to infinity and use a compactness argument to get a limiting measure.
- Let E_Λ^+ be the set of **positively oriented edges**, meaning that the edge points in $+e_i$ for some basis vector e_i.

- Let P_Λ be the set of **plaquettes**, which are 2-dimensional squares.
- Lattice Yang-Mills is then a probability measure on the ways to assign unitary matrices to the edges of our lattice. For each $e \in E_\Lambda^+$, we assign a matrix $U_e \in U(N)$, and then for any negatively oriented edge it must satisfy $U_{e^{-1}} = U_e^{-1}$. Let G_Λ be the set of all such configurations (ways to assign unitary elements overall).

For any plauqette $p = e_1e_2e_3e_4$, we then define

$$U_p = U_{e_1}U_{e_2}U_{e_3}U_{e_4};$$

note that this may depend on the ordering of the elements, but we only care about trace and that's invariant under cyclic shifts. So now we're ready to define the associated action, which we call the **Wilson action** (because Wilson did a lot of work on this): given a configuration $U \in G_\Lambda$, we define

$$S_\Lambda(U) = \sum_{p \in P(\Lambda)} \mathrm{Re}(\mathrm{Tr}(I - U_p)),$$

and the associated probability measure is

$$d\mu_{\Lambda,\beta} = \frac{1}{Z} \exp\left(-\beta S_\Lambda(U)\right) \prod_{e \in E_\Lambda^+} dU_e,$$

where Z is an overall normalization, the exponential term penalizes configurations with a high Wilson action, and $\prod dU_e$ is the product normalized Haar measure (think of it as a measure on the unitary matrices which is invariant under the action of the group).

First of all, notice that this is a weighted measure of the same kind as for Euclidean Yang-Mills theory, so to see that this actually makes sense we just need to match up $S_\Lambda(U)$ with $S_{YM}(A)$. But we'll discuss this next quarter – the main idea is just to discretize $\mathbb{R}^d$ as $\varepsilon\mathbb{Z}^d$ for a small constant ε, and then given a gauge field $A = \sum_{j=1}^d A_j dx_j$, define on the edge $(x, x + \varepsilon e_j)$ the matrix $U(x, x + \varepsilon e_j) = \exp(\varepsilon A_j(x))$ and use the same inverse rule as usual. We can then show that $\sum_p \mathrm{Re}(\mathrm{Tr}(I - U_p))$ is approximately equal to a rescaling ($\frac{\varepsilon^{4-n}}{4}$) of the Yang-Mills action).

Fact 130

This isn't just a toy model where things become easier – when doing quantum mechanics, a lot of the time we have an easily solvable problem and we must do a perturbative approach. But here everything is exactly solvable and we can do analysis, and we may derive interesting things about our theory. Physicists actually use this to predict things, and it's much easier to compute than with the continuum model.

We'll finish by discussing the main observables associated to our lattice. These observables will be the main content of the paper that we'll see next time:

Definition 131

Given a loop $\gamma = (e_1, e_2, \cdots, e_n)$, we define the **Wilson loop variable**

$$W_\gamma = \mathrm{Tr}(U_{e_1}U_{e_2}\cdots U_{e_m})$$

and denote its expectation with respect to the lattice gauge theory $\langle W_\gamma \rangle_{\Lambda,\beta}$.

We care about these Wilson loop expectations, since any abstract theory needs things that can be measured and Wilson loops are such an example – they're gauge-invariant. (This has to do with the fact that you go in a loop, so

any conjugation gets canceled out.) And physically, we can relate this to the concept of **quark confinement** – quarks make up protons and other hadrons, and a key fact is that you never see quarks on their own (they're always with antiquarks). Mathematically, this says that for a large loop, the Wilson loop expectation is very small. Finally, Wilson loops encode concepts related to correlation length and mass gap: the correlation $\langle W_{\gamma_1} W_{\gamma_2}\rangle - \langle W_{\gamma_1}\rangle\langle W_{\gamma_2}\rangle$ and how it depends on distances can show that there is a mass gap (but again, the physics is for another time).

Very briefly, the whole point of this paper is that given a collection of loops $s = (s_1, \cdots, s_n)$, we'll write exact ways for computing the value of $\langle W_{s_1} W_{s_2} \cdots W_{s_n}\rangle_{\Lambda,\beta}$ as a sum over surfaces. So in particular this allows us to recover those correlations mentioned above.

22 November 13, 2024

Ruochuan Xu and I presented on this day, but I unfortunately don't have notes because I focused on presentation delivery and listening to the other presentation. We mostly covered Sections 2 and 3 of the paper and also mentioned a few aspects of Sections 4 and 5.

23 November 15, 2024

Hannah Nabavi

Today's talks will be based on the paper "Langevin Dynamics for the 2D Yang-Mills measure" by Chandra, Chevyrev, Hairer, and Shen. We'll first set up the conceptual framework for the paper, specifically focusing on **stochastic quantization, stochastic Yang-Mills equation, and the statement of main results and methods**.

This paper focuses on the $d = 2$ case, and we'll use a different approach (in particular **not** focusing on lattice Yang-Mills). Our goal is to make sense of the measures $e^{-\beta S_{\mathrm{YM}}(A)}$ for some connection A, and the idea is to look at the Langevin dynamics and prove that we have an invariant measure.

Langevin dynamics are a method used to sample from probability measures of the form $e^{-\beta\mathcal{E}(x)}$ (in particular including the Yang-Mills measure). Langevin dynamics then take values in a connection space, but we'll first illustrate the simpler case of $\mathbb{R}^d$. In Euclidean space, we want to sample from measures of the form $e^{-\beta V(x)}$, and Langevin dynamics take the form (this is an SDE)

$$dx = -\nabla V(x)dt + dW,$$

where x is a function $[0,1] \to \mathbb{R}^d$ and dW is spacetime white-noise. Once we find a solution x, we know it has an invariant measure $e^{-\beta V(x)}$, and we wish to specialize this to the case of stochastic Yang-Mills heat-flow in two dimensions.

Turning now to the connections space, our goal is now instead to sample from something of the form $e^{-\beta S_{\mathrm{YM}}}$; we'll first need to define the corresponding Langevin dynamics, which now take the form

$$\partial_t A_t = -\nabla_A S_{\mathrm{YM}}(A) + \xi$$

for ξ a $\mathfrak{g}$-valued spacetime white noise. We'll later solve the equation, and in later work by the authors it was proved that the invariant measure exists and is unique, so that we can let that be the Yang-Mills measure.

Looking at the stochastic equation above, we need to understand what $-\nabla_A S_{\mathrm{YM}}(A)$ actually means – recall that we have

$$S_{\mathrm{YM}}(A) = \frac{1}{2}||F(A)||^2$$

where F is the curvature form of A, where $F(A) = dA + \frac{1}{2}[A \wedge A]$. By an argument similar to the derivation of the Yang-Mills equation, we then find that

$$\nabla_A S_{\mathrm{YM}}(A) = D_A^* F(A),$$

and so we're just trying to solve $\partial_t A_t = -d_A^* F(A) + \xi$. We'll show that we have a Markov process (Joon) and that we have gauge-invariance (Yanxin), and these topics are broken down into three stages:

- First, we define an appropriate state space for the Markov process.
- There's basically "two problems" with this equation: it's non-elliptic and singular (the regularity is $-\kappa < 0$, so solutions are in the distributional sense). Thus, we have to consider a modified version of the equation via the "DeTurck trick," and we'll also need to make use of the theory of regularity structures.
- The above will make use of mollifier approximations which breaks gauge covariance, but then in the limit we again recover gauge covariance again.

Joon Lee

We'll now focus on the difficulties of showing existence of a solution to the SPDE. First, we can rewrite

$$-D_A^* F_\xi = (-d^* dA + \cdots) + \xi,$$

and the $d^* dA$ term is what causes problems because we have a non-elliptic operator. The idea is that solutions to elliptic PDEs do not have super rough solutions – recall that for an element g in our gauge group, the action of g on A is given by

$$g \cdot A = gAg^{-1} - (dg)g^{-1},$$

and so if our space is split up into different gauge orbits, then a solution to our SPDE above moves us between these orbits. But the condition of gauge-invariance is very restrictive – we want our Yang-Mills measure to take the form $d\mu_{\mathrm{YM}} = Z^{-1} \exp\left(-S_{\mathrm{YM}}(A)\right)$ "dA" with dA some ambiguous Lebesgue measure on connections, and since S_{YM} is invariant under the action of the gauge group, if a path (A_t) will be a solution, then so is (A_t^g), and that pointwise action may be very non-smooth.

Later down the line, the machinery we want to use requires that our SPDE is either elliptic or at least parabolic, and we'll get around this with a strategy originally used for the Ricci flow. If we take any two connections on the same gauge orbit, the probability measure should assign the same value to them, so morally we can think of our situation as "defining on the space of gauge orbits" instead. So this suggests that instead of solving this PDE, we should find a nicer PDE which admits a solution $(\tilde{A}_t)$ such that $\tilde{A}$ is A subject to some gauge action (possibly time-dependent). This is kind of what we call "gauge-fixing," and because we only care about the measure it really just matters that we get a solution up to this quotienting.

To understand what that modification is, DeTurck's observation was to study

$$\partial_t A_t = -D_A^* F(A) + {}_A H(A) + \xi,$$

where $H \in C^\infty(T^2, g)$ is some smooth functional which we can think of as a 0-form. The idea is that if we look at the tangent space of these gauge orbits, they'll be objects of the form "d_A(0-form)," so our modification is just "changing things within a single gauge orbit" at a single time t. In our case, there's a well-motivated answer for what to make our H: our problematic term was $-d^* dA$ previously, but if we instead had $-d^* dA - d d^* A$ then we would just have the Laplacian and that makes things explicitly elliptic. So it makes sense to set $H(A) = d^* A$, and now we're instead

solving the equation $\partial_t A_t = \Delta A + (\cdots) + \xi$, which turns out to be a parabolic equation.

Turning now to the "other problem," this space-time white noise component is a bit irregular and we'll need some heavy machinery to deal with it. We don't have time to talk about what they are, but we'll write out our equation in coordinates and then tell us what the theorem regularity structures gives us (via mollifier approximations). There will be a few terms that are particular to the result that we get, and we'll explain what those mean. Consider the coordinate form of our equation

$$\partial_t A_i = \Delta A_i + \xi_i + [A_j, 2\partial_j A - i - \partial_i A_j + [A_j, A_i]].$$

Theorem 132

For any mollifier $X(t,x)$ (two coordinates corresponding to time and space respectively) and $C \in \mathcal{L}(\mathfrak{g},\mathfrak{g})$ a linear operator from the space of 1-forms to itself, consider the SPDEs parameterized by $\varepsilon > 0$

$$\partial_t A_i = \Delta A_i + X^\varepsilon * \xi + C A_i + [A_j, 2\partial_j A - i - \partial_i A_j + [A_j, A_i]]$$

Then these have solutions in some sensible state space (which we'll state), and they converge in probability as $\varepsilon \to 0$ to a stochastic process in the same state space. (Here note that our ε scaling is $X^\varepsilon = \varepsilon^{-4}(\varepsilon^{-2}t, \varepsilon^{-1}x)$, and we're in two dimensions so it still rescales correctly. Specifically we need a "non-anticipative mollifier" for the stuff here to work.)

What's characteristic of regularity structures arguments is that we have to **add these counter-terms** to cancel out singularities. What's remarkable is that we have the same counterterm C for all ε, and it can be literally any operator. These solutions are not independent of our choice of mollifier for now, but we'll show (in Yanxin's presentation via regularity structures) that there exists a unique choice of C such that the stochastic process is actually independent of our mollifier, so there is a nice canonical choice and additionally we actually get gauge invariance in the limit.

We'll close by talking about the state space that we care about: we want to quotient the space of connections by the action of the gauge group. So we will define a space of connections such that quotienting will result in a complete metric space, and we want Wilson loop observables to be well-defined. We also (just so things are mathematically nice) hope for properties in relation to what we're more familiar with. So for any $\alpha \in \left(\frac{2}{3}, 1\right)$ (measuring the regularity of our initial condition), there is a sensible space of distributional $\mathfrak{g}$-valued 1-forms, which we denote Ω^1_α, which achieves these properties. In particular, quotienting this space by the action of the gauge group does give us a complete metric space, and these spaces are subspaces of Holder-Besov spaces.

Remark 133. *The regularity here is similar to that of the Gaussian free field – in that setting, we can do line integrals in two dimensions but not three dimensions. One of the primary obstacles in higher dimensions is that we have less freedom in the counter-term, and thus showing that we have a limiting object require arduous computations. There are likely other more fundamental physical issues (even for going from the compact torus to $\mathbb{R}^2$), but those are harder to see.*

24 November 18, 2024

Zihao Wang

Today's presentations will focus on Professor Chatterjee's paper "Wilson loops in Ising lattice gauge theory." We'll first focus on the introduction of lattice gauge theories as discrete approximations.

As in previous papers, our goal is to study Euclidean Yang-Mills theories, but lattice gauge theories are discretizations that can be studied. These theories depend on the gauge group and the coupling constant, and our goal is to study Wilson loop expectations. The idea of this paper is that it gives a first-order approximation for weakly coupled 4D lattice gauge theory with gauge group $\mathbb{Z}_2$, and this (in 2018) was the first explicit calculation in the weak coupling regime for any **four-dimensional** theory.

Introducing some notation, let B_N be the cube of lattice points $[-N, N]^4 \cap \mathbb{Z}^4$, and consider the set of configurations $\{-1, 1\}^{E_N}$ on all eedges of B_N. Ising lattice gauge theory then assigns a probability measure $\mu_{N,\beta}$ on configurations proportional to

$$e^{-\beta H_N(\sigma)}, \text{ where } \beta = \frac{1}{g^2}, H_N(\sigma) = -\sum_{p \in P_N} \sigma(p)$$

is a sum over all plaquettes p and $\sigma_p = \sigma_{e_1}\sigma_{e_2}\sigma_{e_3}\sigma_{e_4}$ (each either $+1$ or -1). So we want a lot of the plaquettes to have $\sigma_p = +1$ rather than -1.

The object of interest as usual is the expected value of $W_\gamma = \prod_{e \in \gamma} \sigma_e$ under our measure $\mu_{N,\beta}$. (Notice that W_γ can now take on values either 1 or -1.)

Theorem 134

There exists some β_0 such that

$$\left| \langle W_\gamma \rangle_\beta - e^{-2\ell e^{-12\beta}} \right| \leq C_1 \left(e^{-2\beta} + \sqrt{\frac{\ell_0}{\ell}} \right)^{C_2},$$

where ℓ is the number of edges and ℓ_0 is the number of corner edges. This is tight when β is large.

In particular, this means that $\ell \sim e^{12\beta}$ is the interesting regime, and also we can show that (though this is not obvious) Wilson loop expectations are always nonnegative.

To prove this result, we'll need a toolbox of ideas related to algebraic topology. Much like in the $\mathbb{R}^n$ case, we have dx_1 through dx_n but now corresponding to the n edges coming out of x in the positive direction, and we think of a k-cell as the wedge product $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$. We then say that a G-valued k-form is a G-valued function on the set of positively oriented k-cells (so in our case, it would only take on values 1 or -1).

We also define the exterior derivative

$$df(x) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{1 \leq i \leq n} \partial_i f_{i_1, \cdots, i_k}(x) dx_i \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k};$$

the Poincaré lemma in the discrete setting says that d is a surjective map from the set of $(k-1)$-forms onto the set of **closed** k-forms f (meaning that $df = 0$). (We say that f is **exact** if $f = dg$ for some g; every exact form is closed because $d^2 = 0$.) There's also an analogous Poincaré lemma for the adjoint map δ.

One important fact is that in $n = 4$ we have a correspondence between plaquettes p and their duals $*p$ (since the dual of a k-cell is an $(n-k)$-cell). This means that given a collection of plaquettes (a surface), we also get a dual surface.

One important fact (via the Poincaré lemma) is that for any closed loop γ there is at least one surface P with $\gamma = \partial P$, so loops are like 1-forms and $\delta\gamma = 0$. In particular, this means that we actually have

$$W_\gamma = \prod_{p \in P} \sigma_p$$

for such a surface P. We can also construct an Ising lattice gauge theory on the **dual** lattice (now with zero boundary

condition), and using Fourier expansion for $\mathbb{Z}_2$ (which works for general abelian groups), we can then get the result

$$\left\langle \prod_{p\in P} \sigma_p \right\rangle_\beta = \left\langle e^{-2\lambda \sum_{p\in *P} \sigma_p} \right\rangle_\lambda^*$$

with $\lambda = -\frac{1}{2}\log\tanh\beta$. And so as $\beta \to \infty$ we have $\lambda \to 0$, so this gets us from "weak coupling" to "strong coupling" and thus we can use the strong coupling techniques (such as exponential correlation decay) to prove our result. Importantly, the idea is that σ_ps have exponential decay of correlations in our weak coupling, not σ_es.

One final concept we'll introduce before diving into the proof is the idea of a **vortex**, which is a surface P where $*P$ is closed and connected and $\sigma_p = -1$ for each $p \in P$. The smallest vortex is the set of 6 plaquettes that contain a given edge e; its dual is then the set of 6 plaquettes that form a cube. We'll need the following two facts:

- The set of plaquettes with $\sigma_p = -1$ forms a disjoint union of vortices.
- If Q is a surface and P is a vortex "inside the surface," then the number of common plaquettes $|P \cap Q|$ is even.

Hongrui Chen

We'll now focus on more background and work towards the proof of the main theorem. First we'll consider correlation decay:

Theorem 135

Let f be a function supported on a set of plauqettes P contained in some subcube B, such that the distance between P and ∂B is at least ℓ. Then if $\beta \geq \beta_0$, we have

$$|\langle F\rangle_{N,\beta} - \langle f\rangle_{B,\beta}\rangle| \leq C_f j^3 (C_0 e^{-2\beta})^{C_0 \ell}.$$

As mentioned, the idea of this proof is to use duality and transfer calculations to the strong coupling regime. And one application is that this implies the existence of a thermodynamic limit: for all finitely supported f we have the limit $\langle f\rangle_{N,\beta} \to \langle f\rangle_\beta$ on the "infinite lattice."

The next result tells us that negative sets are rare:

Theorem 136

For any finite set of plaquettes P, we have (let μ_β be the limiting distribution on the infinite lattice)

$$\mathbb{P}(\sigma_p = -1 \text{ for all } p \in P) \leq C_P e^{-2\beta|P|}.$$

The idea is that we approximate the infinite lattice by a finite lattice B, and then we estimate the partition function via $Z_B(\beta) \geq e^{\beta|P(B)|}$ (from the ground state) and note that the contribution from having all $\sigma_p = -1$ can only contribute $C_P e^{\beta(|P(B)-2|P|)}$, and here C_P depends only on the size of B. And in particular if P is a vortex of size m, it turns out $C_P \leq C_m$ because the dual of a vortex is connected and thus can be put into some box as well whose size is dependent on $|*P|$.

In particular, this yields the following corollaries which will be useful in the main theorem's proof:

Corollary 137

The probability under μ_β that p is contained in a vortex of size m is bounded from above by $C_m e^{-2\beta m}$, and the probabilty that a vortex of size at least m intersects a set of j plaquettes is bounded from above by $C_m j e^{-2\beta m}$.

We can now turn to the proof of the main theorem. We'll basically just consider the case where $\ell e^{-12\beta}$ is smaller than some constant C, in which case we will prove that

$$\left|\langle W_\gamma\rangle_\beta - (\tanh 6\beta)^\ell\right| \le Ce^{C\alpha}\left(e^{-2\beta} + \sqrt{\frac{\ell_0}{\ell}}\right).$$

(The other cases can be done quickly; this is the most interesting one.) Since our loop γ is closed, we can write it as ∂Q for some surface Q. Then $W_\gamma = (-1)^N$, where N is the number of negative plaquettes in Q; equivalently we can write $N = \sum_{P\in V} |P \cap Q|$, where V is the set of vortices intersecting Q.

We can now count the parity of N in the following way:

- Drop all vortices that are large, namely $|P| \ge 25$. Indeed, large vortices are rare, so the probability of having one of them is bounded by $C\ell^4 e^{-50\beta} = c\alpha^4 e^{-2\beta}$ (here is where we use the regime that we're considering).
- Drop the vortices that are far from ∂Q, since if they are contained in a 4D box not intersecting the boundary of Q, then $|P \cap Q|$ is even and thus it doesn't contribute to the parity. Note that because we've already dropped the large vortices, these boxes will also be of finite size.
- Now we can drop all vortices with $|P| \ge 7$ (all except the smallest ones), since we only need to consider the **boundary** of Q instead of the whole surface Q and the size is now bounded by $C\ell$. Thus the probability is now bounded by $\ell e^{-14\beta} \le \alpha e^{-2\beta}$.
- So now the smallest vortices left correspond to some edges e; the internal edges e contribute an even number of edges so we just need to consider vortices centered on **edges on the loop**.
- Finally, we further get rid of the corner edges – this is where the $\sqrt{\frac{\ell_0}{\ell}}$ factor comes from; since we get an extra contribution of $\ell_0 e^{-12\beta}$.

Thus we now are left with counting the parity of the number of non-corner edges with $P(e)$ all negative, the point now is that the number $\tanh 6\beta = \frac{e^{-6\beta}}{e^{6\beta}+e^{-6\beta}}$ is the conditional probability that the edges at the vortex e all yield negative $\sigma(p)$s conditional on everything else, and we have conditional independence because no two non-corner edges belong to the same plaquette (flipping the edge e would make them all positive instead of all negative). And the number of edges for which the plaquettes adjacent to e are all negative or all positive is approximately ℓ, since any negative plaquette is contained in a vortex.

25 November 20, 2024

Nan Sheng

We'll discuss today the paper "Rigorous solution of strongly coupled strongly $SO(N)$ lattice gauge theory in the large N limit" (by Professor Chatterjee). The paper specifically demonstrates what's called the string-gauge duality (also called the **AdS-CFT equivalence**), which is an interesting result in physics which links a field theory to some kind of string theory so that we can do certain computations and carry them over.

To recap, our setting is as follows: we consider a large sublattice Λ of $\mathbb{Z}^d$, and place a matrix element in the Lie group $G = SO(N)$ at each edge. Our goal is then to study the Yang-Mills measure $\frac{1}{Z}e^{-\beta S(U)}dG_\Lambda(U)$, where dG_Λ is product Haar measure and the Yang-Mills action S is the sum of $\mathrm{Re}(\mathrm{Tr}(I - U_p))$ over all plaquettes p (where U_p is the product of the matrix elements along the edges bounding the plaquette).

Remark 138. *In the physics context, plaquettes are the "smallest field operators," and all physical quantities of interest (like ψ or $\psi^\dagger$) can be represented as an expansion in terms of those plaquettes.*

As usual, we're curious about the expectation of the Wilson loops $W_\gamma = \mathrm{Re}(\mathrm{Tr}(U(e_1)\cdots U(e_n)))$ under the measure $d\mu_{\Lambda,\beta}$, where $\gamma = (e_1, \cdots, e_n)$ is a particular loop. Physically speaking, this is connected to quark confinement, since if we increase the size of the loop (in the strongly coupled case) we should get something related to $e^{-\beta \mathrm{Area}(\gamma)}$, which should decay very quickly. This paper then produces a two-step process for computing $\langle W_\gamma \rangle$.

First, we write down a "master loop equation," which is similar to the Schwinger-Dyson equations in quantum field theory. The idea is best illustrated with an example: suppose we have three loops ℓ_1, ℓ_2, ℓ_3 and we encode them as a string $s = (\ell_1, \ell_2, \ell_3)$. Letting $\phi(s) = \frac{1}{N^3}\langle W_{\ell_1} W_{\ell_2} W_{\ell_3}\rangle$, the master loop equation will relate such values of ϕ in terms of "lower order terms." Specifically, we get

$$(N-1)|s|\phi(s) = -\sum_{s'\in T^+}\phi(s') + \sum_{s'\in T^-}\phi(s') - N\sum_{s'\in S^+}\phi(s') + N\sum_{s'\in S^-}\phi(s')$$
$$- N\beta\sum_{s'\in D^+}\phi(s') + N\beta\sum_{s'\in D^-}\phi(s') - \frac{1}{N}\sum_{s'\in M^+}\phi(s') + \frac{1}{N}\sum_{s'\in M^-}\phi(s'),$$

where T, S, D, M are certain operations on strings called (positive and negative) "twisting, splitting, deformation, and merging." We'll describe these operations now:

- The M^+ and M^- operations merge two loops if they cross over the same edge – they can either be merged "in the same direction" (keeping the number of edges the same) or "in opposite directions" (lowering the number of edges, and also reversing the orientation of one of the loops for consistency).
- The D^+ and D^- operations do something similar but merging a loop with an adjacent plaquette – they essentially add a square into or remove a square from a loop.
- The S^+ and S^- operations take a loop that crosses over the same edge twice and separates it into two different loops in the "inverse way" as M^+ and M^-.
- The T^+ and T^- operations also deal with a single loop crossing over the same edge twice, and they reverse orientations in some way.

So the point of this recursive relation is that we can think of "evolving time from t to $t+\delta t$ from a string S to a string S', and this can be a kind of string theory – the string is a collection of Wilson loops that evolve in time. An example graphically is the "pair of pants diagram", in which time evolves to merge two loops into one. (So we must consider all possible trajectories, and the probability of each trajectory occurring is dictated by the coefficients.)

We'll give a very quick proof sketch of how this master loop equation can be derived: the basic idea is that we do a local expansion via the $SO(N)$ group and calculate some "integration-by-parts" integrals like $\int_{SO(N)}(g\Delta f + (\nabla g, \nabla f)$ in local charts. (The point is that we end up with some derivatives, and we haven't actually used the large N or the structure of the group very much yet.)

If we now push this to the large N limit (sometimes called the 't Hooft scaling) and replace $\beta \mapsto N\beta$, the idea is that each W_ℓ is a trace over $N\times N$ matrices and thus is bounded by N; thus $\phi_N(s)$ is still bounded by 1 as $N\to\infty$. We can then argue that there is some subsequential limit that converges to some $\phi(s)$, and when we do this we can remove some of the terms in the master loop equation above to get $|s|\phi(s)$ in terms of only **splittings and deformations** of $\phi(s)$. So we must prove that the right-hand side is related to the string if we define some certain weights; we turn out to get

$$\lim_{N\to\infty}\phi_N(s) = \sum_{\chi\in X(s)} w(\chi),$$

where χ is a sequence starting from s and end up at the null loop. These weights w are then defined as a product of $w(s, s')$ terms, which are either $\pm\frac{\beta}{|s|}$ or $\pm\frac{1}{|s|}$ depending on whether we have a deformation or a different kind of operation. (In particular we need β small enough for this to converge via a contraction argument and thus that a **unique** limit exists; in general we just have subsequential limits.) And the derivation of these ws basically comes from a series expansion with respect to β – in particular this function ϕ is analytic.

Remark 139. *This work has also been pushed to $SU(N)$ which is more physically relevant – we can do a $\frac{1}{N}$ expansion and get more accurate estimates. But getting from strong coupling to weak coupling will change things because we go from the area law to a "length law" instead.*

26 November 22, 2024

Maya Chande

The paper for this presentation is Chevyrev's "Yang-Mills measure on the two-dimensional torus as a random distribution." The overall goal is to construct an explicit measure by using a nice space of distributional 1-forms for which we can properly define the set of Wilson loop observables.

Recall that to define a Yang-Mills action, we need a principal bundle ($\mathbb{T}^2$; this analysis turns out to only work for trivial bundles) and a compact connected Lie group G with Lie algebra $\mathfrak{g}$. We want to equip the space of connections $\mathcal{A}$ with a formal Lebesgue measure, and by triviality of the bundle we can identify $\mathcal{A}$ with $\Omega^1(\mathbb{T}^2, \mathfrak{g})$.

The strategy will be to understand the distribution via gauge-invariant observables; for this we'll need to understand the concept of **holonomy**. For a connection A and a piecewise smooth curve γ, the **holonomy** $\mathrm{hol}(A, \gamma)$ around the loop with respect to A can be seen as the solution of a certain ODE

$$y'(t) = y(t)A(\gamma(t))\gamma'(t), \quad y(0) = \mathrm{id}_G;$$

specifically we define the holonomy to be $y(1)$. Once we have this, Wilson loop observables are basically functions of these holonomie, since they depend on $\mathrm{Tr}(\phi\mathrm{hol}(A, \gamma_i))$ for loops γ_i and for some finite-dimensional representation ϕ of G so that the trace makes sense.

What's nice is that these observables play well with 1-forms – there's a correspondence between functions of holonomies and 1-forms. Since we want to define a space of distributional 1-forms, we want to know why we care about these holonomies in the first place:

Theorem 140 (Sengupta, 1992)
Let M be a connected manifold, and let $A_1, A_2 \in \mathcal{A}$ be two connections. Fix a point $x \in M$ in our manifold and two points $u_1, u_2 \in \pi^{-1}(X)$ in the preimage. Suppose we have some positive integer k and $c_1, \cdots, c_k$ some collection of piecewise smooth curves on M based at the point x. Then if f is an Ad-invariant function on G^k (meaning that $f(hg_1h^{-1}, \cdots, hg_nh^{-1}) = f(g_1, \cdots, g_n)$ for all h), and further we have

$$f(\mathrm{hol}_{u_1}(c_1, A_1), \cdots, \mathrm{hol}_{u_1}(c_k, A_1)) = f(\mathrm{hol}_{u_2}(c_1, A_2), \cdots, \mathrm{hol}_{u_2}(c_k, A_2))$$

then $[A_1] = [A_2] = \mathcal{A}/\mathcal{G}$. And further, if this last relation holds then the equality for f holds for all $c_1, \cdots, c_k$.

The idea now is to see Yang-Mills as a **stochastic process** indexed by a nice collection of loops. Here's the main result (so the point is that previously we just had the observables, but this gives us the A that constructs these observables):

Theorem 141

Let G be compact, connected, and simply connected with Lie algebra $\mathfrak{g}$. For all $\alpha \in (\frac{1}{2}, 1)$, there is a random variable A valued in the space $\Omega^1_\alpha(\mathbb{T}^2, \mathfrak{g})$ which satisfies the following: for any $x \in \mathbb{T}^2$, any collection of axis loops $\gamma_1, \cdots, \gamma_n$, and any Ad-invariant f, the quantity

$$f(\text{hol}(A, \gamma_1), \cdots, \text{hol}(A, \gamma_n)$$

is equal in law to to what we would get when computing on the Yang-Mills measure.

Note that to define these axis loops, we're now working in a discrete setting: we have the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, we have a projection map $\pi_{\mathbb{T}^2} : \mathbb{R}^2 \to \mathbb{T}^2$, and Λ_N is a lattice on $\mathbb{Z}^2$ of size 2^{-N} with N a parameter. An **axis loop** is then a piecewise smooth curve $\gamma : [0, 1] \to \mathbb{T}^2$ consisting of putting together steps of the form $x \mapsto x \pm te_1$ or $x \mapsto x \pm te_2$. (The variable A that is defined unfortunately blows up for more general loops.)

We'll also define what this space Ω^1_α actually is: if we pick a point and an adjacent point in the lattice, the directed edge is called a **bond**, and we can also consider the reversal of that bond. We'll build a space which plays well with these reversals: let $\Omega^{1,(N)}$ be the space of functions $A : B_N \to E$ from bonds on Λ_N to some metric space E, such that $A(\alpha) = -A(\overline{\alpha})$ for all bonds α. We then define $\mathcal{U}^{(N)}$ similarly but with functions $U : B_N \to G$ satisfying $U(\alpha) = U(\alpha)^{-1}$ (we can for example get these by exponentiation). Our goal is to build up 1-forms via the data of loops – to do this, we define for any curve $\gamma : [s, t] \to E$ the ***q*-variation**

$$|\gamma|_{q-\text{var}} = \sup_{D \subseteq [s,t]} \left(\sum_{t_i} d(\gamma(t_i), \gamma(t_{i+1}))^q \right)^{1/q},$$

and we define $\mathcal{C}^{q-\text{var}}([s, t], E)$ to be the set of such all functions with bounded q-variation. We say an axis loop $\ell = \{x + ce_\mu : c \in [0, \lambda]\}$ has length λ, and suppose two loops ℓ and $\bar{\ell}$ are parallel and have the same projections onto the two axes. We can then define

$$\rho(\ell, \bar{\ell}) = |\ell|^{1/2} d(\ell, \bar{\ell})^{1/2}$$

We then have a set of nice "additive" functions Ω $A : \mathcal{X} \to E$; let Ω^1 be the set of bounded measurable 1-forms, and define evaluation of an $A \in \Omega^1$ via

$$A(\gamma) = \int_s^t A(\gamma(u)) d\gamma(u)$$

for sufficiently regular loops γ. We can then parameterize our loops via axis line segments and identify an additive function with a corresponding $A \in \Omega^1$ via $A(\ell) = A(\gamma)$, and now we have a bunch of norms on Ω that we can transfer over to Ω^1 via regularity of additive functions on the set of axial line segments.

Yanxin Zhou

We'll finish up the ideas from Hannah and Joon's presentations last week now (Langevin dynamics on 2D Yang-Mills). We previously made sense of the solution space and saw the difficulties of solving the solution; the two main problems are that the solution is non-parabolic (and thus we must use the DeTurck trick, adding a term that doesn't affect our solution if we only care about gauge orbits) and that the stochastic Yang-Mills equation is highly singular (and thus we must make use of regularity structures). To use regularity structures, we need to introduce certain counter-terms C_ε which potentially break gauge invariance; the goal of today's talk is to show that we can indeed choose the counter-terms in a smart way to recover gauge invariance in the limit.

What we'll specifically focus on is intuition for how these C_εs re chosen. We'll first introduce a more general

setting: recall the stochastic differential equation

$$\partial_t U = \delta u = f(U, \Delta U, \xi),$$

where $U : \mathbb{R} \times T^d \to \mathfrak{g}$ is a linear function. We let $U^\epsilon(F)$ be the solution map mapping the initial condition u_0 to the law of the solution with spacetime white noise ξ replaced by a mollified version ξ^ϵ. Our question is whether the limit $\lim_{\epsilon \downarrow 0} U^\epsilon(F)$ exists, and regularity structures says yes – there are a finite number of G_τ and constants C_τ^ϵ such that

$$U(F) = \lim_{\epsilon \to 0} U^\epsilon(F + C_\tau^\epsilon G_\tau),$$

and the solution is independent of our chosen mollification and is solution in F. (In fact, there are explicit expressions for the counter-terms, and the G_τs are of lower order than F.)

But already here there are many degrees of freedom. Remembering that the stochastic Yang-Mills equation reads (after the DeTurck term is added)

$$\partial_t A = \Delta A + A \partial A + A^3 + \xi,$$

our nonlinear terms are $A\partial A + A^3$ (that is our F in this case), and our counter-terms must be within constant, A, A^2, and ∂A. But by the symmetries of the equation, we can make some additional observations:

- The stochastic Yang-Mills equation is invariant under the transformation

  $$x \mapsto -x, \quad A \mapsto -A.$$

  Thus we can rule out A^2, ∂A and also constants.

- Because this equation has no preference over coordinates, A is invariant under permutation of coordinates. Thus the counter-term must actually appear in the form

  $$\partial_t A = (\cdots) + C^{(\epsilon)} A_i,$$

  where $C^{(\epsilon)}$ is a linear operator on $\mathfrak{g}$ and i is a particular coordinate.

We can now argue a bit further: we always take G to be a compact Lie group, which has the property that it can be factored into simple Lie groups. Thus we can assume without loss of generality that G is itself simple. Then any linear map $L : \mathfrak{g} \to \mathfrak{g}$ commuting with the adjoint action must be of the form $\lambda \mathrm{id}$ for some constant $\lambda \in \mathbb{R}$, and we have the following useful property:

Theorem 142
There exists some constant $\hat{C}^{(\epsilon)}$ such that the limit $\lim_{\epsilon \to 0} C^{(\epsilon)} - \hat{C}^{(\epsilon)}$ exists, and $[\hat{C}^{(\epsilon)}, \mathrm{Ad}_g] = 0$ for all $g \in G$.

So $\hat{C}^{(\epsilon)}$ is of the form a constant times the identity, and thus the counter-terms approach a multiple of the identity times A_i.

From here, the final step is to show that there is a **unique** choice $C^{(\epsilon)}$ that gives gauge-invariance. Recall that gauge-invariance in this context is described as follows: if A_0 is a connection and we have some other $g_0 \cdot A_0$ on the same gauge orbit. Then we're saying that we have a time-dependent gauge transformation g_t such that the solution started at $(g_0 \cdot A_0)$ at time t is g_t acting on A_t (the solution at time t when started at A_0).

Intuitively, we first figure out what the g_t should be, and we do this by assuming first that we already know it and seeing what $\overline{A}_t = g_t \cdot A_t$ would look like (making it look similar to the original Yang-Mills equation). The first step of

the calculation requires us to derive

$$\begin{aligned}\partial_t \overline{A}_t &= \partial_t(g_t \dot{A}_t) \\ &= \partial_t(g \cdot A_t) + \partial_t(g_t \cdot A) \\ &= \Big(- D^*_{g_t \cdot A} F(g \cdot A_t) + \mathrm{Ad}_g \xi_\varepsilon - D_{g_t \cdot A}\left[d^*(g \cdot A) + \partial_i(\partial_i g \cdot g^{-1}) + [g \cdot A_i, \partial_i g \cdot g^{-1}] \right] \\ &\qquad + C_\varepsilon(g \cdot A + \partial_i g \cdot g^{-1}) \Big) - D_{g_t \cdot A}(\partial_t g \cdot g^{-1}),\end{aligned}$$

where this C_ε term comes from the renormalization. So if we look at this expression, we find that the blue terms look like what we want in the Yang-Mills equation, and thus we just need to cancel out the remaining terms. But there are a few issues – the mollified noise no longer has the same law as the spacetime white noise (intuitively, the Ad_g action does not commute with the mollifier), and the remaining terms all have the form "exterior derivative of something" except the C_ε one. But luckily there turns out to be a "miracle calculation" that shows that a unique C_ε does indeed work in the limit $\varepsilon \to 0$, since the contribution of $C_\varepsilon(\partial_i g \cdot g^{-1})$ will compensate for the adjoint term $\mathrm{Ad}_g \xi_\varepsilon$ so that we get the same law as the spacetime white noise. (Basically, the two different issues fix each other.)

27 December 2, 2024

We'll finish up the proof of Stone's theorem today – recall that we're doing the converse direction here, where H is a self-adjoint operator defined on some dense subspace $\mathcal{D}$ of our Hilbert space $\mathcal{H}$, and the goal is to show that there exists a unique strongly continuous one-parameter group of unitary operators $(U(t))_{t \in \mathbb{R}}$ whose generator is H. In other words, under certain conditions (self-adjointness, which is slightly stronger than being symmetric), there is a solution to the Schrodinger equation.

We won't go through all of the details, but we've already constructed $U(t)$ for $t \geq 0$ and said that we'll construct them similarly for $t < 0$. Let's review the main ideas: we constructed the approximations

$$U_n(t) = e^{-nt} e^{nt(I + in^{-1}H)^{-1}},$$

which make sense because H is self-adjoint and thus the map $I + in^{-1}H : \mathcal{D} \to \mathcal{H}$ is a bijction with continuous inverse, meaning we can expand out a well-defined power series. (So basically "shifting" by any nonreal multiple of the identity gives us nice properties.) And this is indeed an approximation of $U(t) = e^{-itH}$ because as $n \to \infty$ we have $(I + in^{-1}H)^{-1} \approx I - in^{-1}H$, so the e^{nt}s will cancel out.

We then did some maneuvering to show that for all $t \geq 0$, we have $U_n(t)x \to U(t)x$ for all $x \in \mathcal{H}$ (by doing some Cauchy sequence estimation via interpolations $U_n(t-s)U_n(s)$), and in fact we can get uniform convergence on bounded intervals. With those bits of information, it's now routine to complete the proof that $U(t)$ is indeed what we want (for example, for $s, t \geq 0$ we have $U_n(s+t) = U_n(s)U_n(t)$, so in the limit we also have $U(s+t) = U(s)U(t)$, hence the semigroup property).

What's left is to show that H is actually an infinitesimal generator (this is where we stopped). We defined $\mathcal{D}'$ to be the set of x such that $\lim_{t \downarrow 0} \frac{U(t)x - x}{-it}$ exists, an we call that limit $H'x$; our goal is then to show that in fact $H' = H$ and $\mathcal{D}' = \mathcal{D}$. If we take $x \in \mathcal{D}$, then

$$U(t)x - x = \lim_{n \to \infty} (U_n(t)x - x) = \lim_{n \to \infty} \int_0^t \frac{d}{ds}(U_n(s)x)ds,$$

and plugging in the formula for the derivative of U_n yields that this is in fact $-i \lim_{n \to \infty} \int_0^t T_n H U_n(s)x ds$. Since x is

in the domain, all of these operators commute, and thus we have

$$U(t)x - x = -i \lim_{n\to\infty} \int_0^t U_n(s) T_n H x ds$$

Using estimates we already have on limits of T_n and U_n, we can then show that this last integral is $-\int_0^t U(s)Hxds$, meaning that U "satisfies the integral equation that it should satisfy". From here it's not too hard to show that $\lim_{t\downarrow 0} \frac{U(t)x-x}{-it}$ exists and is indeed equal to Hx, so $H = H'$ on $\mathcal{D}$ and it just remains to show that $\mathcal{D}' \subseteq \mathcal{D}$.

For this, note that $I + iH$ is an invertible map whose inverse is continuous, and thus for any $y \in \mathcal{D}'$ there is some $x \in \mathcal{D}$ with

$$(I + iH')y = (I + iH)x.$$

So recalling the definition of T_n, we have

$$x = (I + iH)^{-1}(I + iH')y = T_1(I + iH')y,$$

and because T_1 is continuous,

$$\begin{aligned} T_1 H' y &= T_1 \lim_{t\downarrow 0} \frac{U(t)y - y}{-it} \\ &= \lim_{t\downarrow 0} \frac{T_1 U(t) y - T_1 y}{-it} \\ &= \lim_{t\downarrow 0} \frac{U(t) T_1 y - T_1 y}{-it} \end{aligned}$$

because T_1 commutes with each U_n and thus commutes with U as well. In particular, this means $T_1 y \in \mathcal{D}'$ and $H'T_1 y = T_1 H' y$; however, T_1 is the inverse of a map $\mathcal{D} \to \mathcal{H}$ and hence actually $T_1 y \in \mathcal{D}$ and $H'T_1 y = HT_1 y$ as well. Therefore

$$x = T_1(I + iH')y = (I + iH)T_1 y = y$$

and thus $y \in \mathcal{D}$ as well.

We'll now conclude with one more class presentation:

Carina Hong

We'll discuss today the paper "Fractional Gaussian Forms and Gauge Theory" by Cao and Sheffield. We'll first define what these fractional Gaussian forms are: they're parameterized by a value s, and in the scalar-valued case we can think of taking $h = (-\Delta)^{-s/2}W$, where Δ is the Laplacian and W is the white noise on $\mathbb{R}^n$ (so n is another parameter). We can also parameterize using the **Hurst parameter** $H = s - \frac{n}{2}$, which is useful because of the scaling property

$$h(ax) \overset{d}{=} a^H h(x)$$

for $h \in \mathrm{FGF}_s(H^n)$. This H is also shows up in the covariance relation

$$\mathrm{Cov}[(h,\phi_1),(h,\phi_2)] = \iint |x-y|^{2H} \phi_1(x)\phi_2(y)dxdy,$$

which holds only for H positive and non-integer (for instance when $H = 0$ we get what's called the "log-correlated Gaussian field" or LGF for short, and we have to replace $|x-y|^{2H}$ with $-\log|x-y|$ instead).

This is a generalization of some more well-known objects: if $s = 0$ we get white noise, and if $s = 1$ we get the Gaussian free field (so $n = 1$ this is Brownian motion and $n = 2$ we get the 2D Gaussian free field). More broadly

for $s \in (\frac{1}{2}, \frac{3}{2})$, we get the family of "fractional Brownian motion," and for $s = 2$ we get the "membrane model" or "bi-Laplacian Gaussian field." The family of parameters is often visualized in a 2D plot of values (n, s); in different regions (depending on the value of H) the behavior is different.

- When $H > 0$, $\mathrm{FGF}_s(\mathbb{R}^n)$ is in fact almost surely continuous.
- When $H < 0$, it is not defined pointwise and viewed instead as a random generalized function.
- Finally, $H = 0$ is the rather pathological case of the LGF mentioned above.

When we replace scalar values with differential forms, we must now instead consider the quantity

$$(-\Delta)^{-s/2} W_k$$

for W_k a "white noise on k-forms." The interesting connection to our class is how this relates to the lattice gauge theory (the introduction, section 1.5, and section 5.3) discussed in previous presentations. Think of a directed graph with vertex set Λ, where for each oriented edge we assign some element Q_e such that $Q_{e'} = Q_e^{-1}$, and we think of this as a "discrete connection" in the following sense. A connection is a 1-form that takes values in a Lie algebra (which we think of as a finite-dimensional vector space), and given any path on our graph, we look at the edges $p = (e_1, \cdots, e_n)$ along the path and think of $Q_{e_1} \cdots Q_{e_n}$ as integrating the connection along p. (Recall that each Q_e is iid Haar, then exponentially weighted by a product of traces on plaquettes.)

We then get a gauge transformation if we consider $Q_e \mapsto Q_e g$ for some group element g; for any closed loop this doesn't affect the final product. We can further also do this "gauge fixing" in a way so the edges e in a spanning tree of the graph Λ have $Q_e = \mathrm{id}$. The thought is that for low enough temperature our unitary matrices are close to the identity, and since $(I + A)(I + B) \approx I + (A + B)$ we get a continuum connection from the discrete connections.

The point is that in the continuum limit, the "gauge-fixed projections" of fractional Gaussian forms $\mathrm{FGF}_1^1(M)$ on a manifold (onto the space $d^* = 0$) are like the **small-scale limits of $U(1)$ Yang-Mills gauge theory**, because the Lie algebra of $U(1)$ is $\mathbb{R}$. And if we replace $U(1)$ with another Lie group, then the corresponding small-scale limit is the Lie algebra analog of those Gaussian forms.

28 December 4, 2024

Our first goal for today is to finish the proof of Stone's theorem. Picking up where we left off from last lecture that we still need to show U is unitary, meaning that $U(t)^* U(t) = I$ and $U(t)U(t)^* = I$.

For the first fact, we just need to show that $\langle U(t)x, U(t)y\rangle = \langle x, y\rangle$ for all $t \geq 0$ and $x, y \in \mathcal{H}$ (since this means $\langle x, U^*(t)U(t)y\rangle = \langle x, y\rangle$ for all x, y – note that this fact doesn't prove the other fact). This can be done by proving that the derivative of the left-hand side is zero, since it evaluates to $\langle x, y\rangle$ when $t = 0$. Indeed,

$$\begin{aligned}\frac{d}{dt}\langle U(t)x, U(t)y\rangle &= \left\langle \frac{d}{dt}U(t)x, U(t)y \right\rangle + \left\langle U(t)x, \frac{d}{dt}U(t)y \right\rangle \\ &= \langle -iHU(t)x, y\rangle + \langle U(t)x, -iHU(t)y\rangle\end{aligned}$$

because we've already proven that we satisfy the Schrodinger equation $\frac{d}{dt}U(t)x = -iHU(t)x$. And now using conjugate linearity, this simplifies to $-i\langle HU(t)x, U(t)y\rangle + i\langle U(t)x, HU(t)y\rangle$, and now the two terms cancel out because H is symmetric on its domain.

For the other fact, we actually need to define $U(t)$ for negative t as well, and we do so via

$$U(-t) = \lim_{n\to\infty} U_n(-t), \quad U_n(-t) = e^{-nt} e^{nt(I - in^{-1}H)^{-1}}.$$

Lemma 143

For all $t \in \mathbb{R}$, we have $U(t)^* = U(-t)$.

Proof. We claim that for each finite n that

$$((I + in^{-1}H)^{-1})^* = (I - in^{-1}H)^{-1}, \quad ((I - in^{-1}H)^{-1})^* = (I + in^{-1}H)^{-1}.$$

Indeed, we know that $(I - in^{-1}H)^* = I + in^{-1}H$ and thus for any $x, y \in \mathcal{H}$ we have

$$\begin{aligned}
\langle (I + in^{-1}H)^{-1}x, y\rangle &= \langle (I + in^{-1}H)^{-1}x, (I - in^{-1}H)(I - in^{-1}H)^{-1}y\rangle \\
&= \langle (I - in^{-1}H)^*(I + in^{-1}H)^{-1}x, (I - in^{-1}H)^{-1}y\rangle \\
&= \langle (I + in^{-1}H)(I + in^{-1}H)^{-1}x, (I - in^{-1}H)^{-1}y\rangle \\
&= \langle x, (I - in^{-1}H)^{-1}y\rangle.
\end{aligned}$$

So after taking exponentials the same adjoint relation holds for $U_n(t)$, and thus the relation also holds in the limit for $U(t)$. □

With this we thus know that $U(t)U(t)^* = U(-t)^*U(-t) = I$, and since this holds for all t we thus indeed get the second fact. And thus a global solution indeed exists for any self-adjoint operator H (both forward and backward in time).

Uniqueness of the family of operators is also simple to prove: suppose $(U(t))_{t\in\mathbb{R}}$ and $(V(t))_{t\in\mathbb{R}}$ are two groups of unitary operators with the same generator H. Then for any $x \in \mathcal{H}$, we may define $w(t) = U(t)x - V(t)x$. We have $w(0) = 0$ (because $U(0)$ and $V(0)$ are both the identity) and

$$\begin{aligned}
\frac{d}{dt}||w(t)||^2 &= \frac{d}{dt}\langle w(t), w(t)\rangle \\
&= \left\langle \frac{dw}{dt}, w(t)\right\rangle + \left\langle w(t), \frac{dw}{dt}\right\rangle \\
&= \langle -iHw(t), w(t)\rangle + \langle w(t), -iHw(t)\rangle \\
&= 0
\end{aligned}$$

(since if U and V satisfy Schrodinger's equation, so does $U - V$, and also remembering that $w(t)$ is indeed in the domain of H because $U(t)x$ and $V(t)x$ are). So indeed $w(t)$ is constant and thus zero, meaning U and V have the same value for any x.

Remark 144. *Identifying the exact condition needed for self-adjointness here is a rather important achievement, but it's not really known much outside of mathematical physics.*

Going forward (and then looking ahead into the next quarter of the course), we'll be using methods from probability theory to get results in quantum field theory as far as we can before getting into discrete systems. Our whole discussion so far has been towards facilitating that goal, and this next result is what we'll use for that goal.

Definition 145

Let $\mathcal{H}$ be a Hilbert space and let $(P(t))_{t\geq 0}$ (note that we now index by nonnegative time) be a family of (not necessarily continuous) linear maps from $\mathcal{H}$ into $\mathcal{H}$. We say that $(P(t))_{t\geq 0}$ is a **semigroup** if $P(0) = I$ and $P(s + t) = P(s)P(t)$ for all $s, t \geq 0$, and we say that it is **strongly continuous** if for all $x \in \mathcal{H}$, the map $t \mapsto P(t)x$ is continuous.

The idea now is that we previously considered unitary operators, but now we'll consider symmetric ones:

Theorem 146

Let $(P(t))_{t\geq 0}$ be a strongly continuous semigroup of symmetric operators on $\mathcal{H}$. Consider the set

$$\mathcal{D} = \left\{ x : \lim_{t \downarrow 0} \frac{P(t)x - x}{t} \text{ exists} \right\},$$

and whenever this limit exists we call it Hx. Then $\mathcal{D}$ is dense in $\mathcal{H}$, and H is a self-adjoint operator with domain $\mathcal{D}$.

This result is important for us, bcause symmetric semigroups are things that we obtain from stochastic processes – we'll see later on how we do that, but they are basic things like the heat semigroup or OU semigroup (which don't require oscillatory integrals, just the theory of Markov processes). So if we first identify the self-adjoint operator H that generates our quantum system, then generate the semigroup using probability theory whose generator is that H, this theorem tells us H is "indeed correct" and Stone's theorem tells us that a solution of the Schrodinger equation does exist – this is the rigorous version of Wick rotation.

Fact 147

Notice in particular that $U(t) = e^{-itH}$, while $P(t) = e^{-tH}$, so we're indeed "going from real to imaginary time" here. And the point is that the domains of the Hs obtained from the two families of operators are the same when we do this whole process.

We won't do the full proof of this theorem because it's similar to what we've already done, but we'll just note the main points. We need the following two facts from functional analysis (which hold for Banach spaces in general, not just Hilbert spaces):

Lemma 148 (Banach–Steinhaus / Uniform boundedness principle)

Let $\{T_n\}_{n\geq 1}$ be a sequence of bounded linear operators from $\mathcal{H}$ into itself. Suppose that for all $x \in \mathcal{H}$, we have $\sup_n ||T_n x|| < \infty$. Then $\sup_{n\geq 1} ||T_n|| < \infty$.

Lemma 149 (Closed graph theorem)

Let $T : \mathcal{H} \to \mathcal{H}$ be a linear map. If the **graph** of T (that is, the set $\{(x, Tx) : x \in \mathcal{H}\}$ is a closed subset of $\mathcal{H} \times \mathcal{H}$, then T is continuous, and in fact the implication goes both ways.

As a consequence, we have the following useful theorem:

Theorem 150 (Hellinger-Toeplitz)

Any symmetric operator $T : \mathcal{H} \to \mathcal{H}$ is continuous.

Proof. We know that the adjoint T^* of any (densely defined) operator is closed. But since T is symmetric and is defined on all of $\mathcal{H}$, $T^* = T$ and thus T is closed. So by the closed graph theorem its graph is closed and thus T is continuous. □

Turning back to the proof, we claim now that there exist constants c_1, c_2 with

$$||P(t)|| \leq c_1 e^{c_2 t} \quad \forall t.$$

Indeed, first we show that there is some $t_0 > 0$ with $\sup_{0 \leq t \leq t_0} ||P(t)|| < \infty$. If this were false, then we would have a sequence $t_n \downarrow 0$ with $||P(t_n)|| \to \infty$. But for any $x \in \mathcal{H}$ we have $P(t_n)x \to P(0)x = x$ by strong continuity, so $\sup_n ||P(t_n)x||$ must be finite and thus $\sup_n ||P(t_n)||$ is finite by uniform boundedness. Then from here, we may writ any t as a sum of t_0s plus the remainder, and we apply the semigroup property and the fact that $||AB|| \leq ||A|| \, ||B||$ for any two operators A and B.

This exponential bound is quite crucial for our result – for denseness of the domain $\mathcal{D}$, we use the same proof as in Stone's theorem (using strong continuity), but instead of the bound $||U(t)|| \leq 1$ we now use the lemma to show that $\sup_{0 \leq s \leq t} ||P(s)|| < \infty$ is finite for all t. What remains is to prove that the operator H is indeed self-adjoint on the domain $\mathcal{D}$, and it suffices to show that $H + \alpha I$ is self-adjoint for some real number α (since it has the same domain as H). Indeed H is symmetric on $\mathcal{D}$, because for any $x, y \in \mathcal{D}$ we have

$$\begin{aligned}
\langle Hx, y \rangle &= \left\langle \lim_{t \downarrow 0} \frac{P(t)x - x}{-t}, y \right\rangle \\
&= \lim_{t \downarrow 0} \frac{1}{-t} \left(\langle P(t)x, y \rangle - \langle x, y \rangle \right) \\
&= \lim_{t \downarrow 0} \frac{1}{-t} \left(\langle x, P(t)y \rangle - \langle x, y \rangle \right)
\end{aligned}$$

by symmetry of $P(t)$, and then we can retrace our steps back to show that this is also $\langle x, Hy \rangle$. So now it suffices to show (by one of our conditions for self-adjointness) that for some $\alpha \in \mathbb{R}$, the range of $(H + \alpha I + iI)$ and $(H + \alpha I - iI)$ are both all of $\mathcal{H}$. We can do this by producing a right inverse for each operator, but we've done something like this before: we define

$$R = \int_0^\infty e^{is - \alpha s} P(s) ds,$$

which is well-defined if we choose any $\alpha > C_2$ large enough so that it counters the exponential growth in norm of P. We can then indeed show, as before, that R maps into the domain $\mathcal{D}$ and that $(H + \alpha I - iI)Rx = x$; something similar also works for $(H + \alpha I + iI)$.

Example 151

As a basic example for how to use this machinery, we'll go back to the free particle. Let $\mathcal{H} = L^2(\mathbb{R})$, and for any $\psi \in \mathcal{H}$ and $t \geq 0$, we may define

$$P(t)\psi(x) = \mathbb{E}[\psi(x + \sqrt{t}Z)]$$

where Z is a standard normal random variable.

This is well-defined because for $t = 0$ we just get $\mathbb{E}[\psi(x)]$, and for $t > 0$ we can check for absolute integrability:

$$\begin{aligned}\mathbb{E}[|\psi(x+\sqrt{t}Z)|] &\leq \sqrt{\mathbb{E}[\psi(x+\sqrt{t}Z)^2]} \\ &= \sqrt{\int_{-\infty}^{\infty} |\psi(x+\sqrt{t}Z)|^2 \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz} \\ &\leq \sqrt{\int_{-\infty}^{\infty} |\psi(x+\sqrt{t}Z)|^2 dz} \\ &= \sqrt{\frac{1}{\sqrt{t}} \int_{-\infty}^{\infty} |\psi(y)|^2 dy}\end{aligned}$$

by making the change of variables $y = x + \sqrt{t}z$, $dy = \sqrt{t}\, dz$ (in the latter inequality we use that $\frac{e^{-z^2}/2}{\sqrt{2\pi}}$ integrates to 1). And to show that $P(t)\psi$ is in L^2, note that

$$\begin{aligned}||P(t)\psi||_{L^2}^2 &= \int_{-\infty}^{\infty} |P(t)\psi(x)|^2 dx \\ &= \int_{-\infty}^{\infty} \left|\mathbb{E}[\psi(x+\sqrt{t}Z)]\right|^2 dx \\ &\leq \int_{-\infty}^{\infty} \mathbb{E}[|\psi(x+\sqrt{t}Z)]|^2 dx,\end{aligned}$$

and by Fubini's theorem we can swap the expectation and integral to get $\mathbb{E}[\int_{-\infty}^{\infty} |\psi(x+\sqrt{t}Z)|^2 dx$, and the inner integral inside is always the constant $||\psi||_{L^2}^2$ because we're just doing a random shift. This actually turns out to be the heat semigroup, and the generator turns out to be exactly the one we want for the free particle – we'll see that next time! And the nice thing is that this is defined on all of L^2, so there's no issues with defining things on a dense subset like we had previously.

29 December 6, 2024

Last time, we were constructing the Hamiltonian for the free particle using probability via the heat semigroup $P(t)$. We showed that $P(t)\psi$ is well-defined for all $\psi \in L^2$, and in fact $P(t)$ is a map $\mathcal{H} \to \mathcal{H}$. To use this to get our Hamiltonian, we need to prove that we have a semigroup of symmetric operators and that we have strong continuity.

Symmetry is verified via the calculation

$$\begin{aligned}\langle P(t)\psi, \phi\rangle &= \int_{-\infty}^{\infty} P(t)\psi(x)\overline{\phi(x)} dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \psi(x+\sqrt{t}Z)\frac{e^{-z^2/2}}{\sqrt{2\pi}} dz\right) \overline{\phi(x)} dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x+\sqrt{t}Z)\overline{\phi(x)}\frac{e^{-z^2/2}}{\sqrt{2\pi}} dz dx\end{aligned}$$

by Fubini's theorem (because ψ, ϕ are both in $L^2(\mathbb{R})$). Now changing the order of integration and doing the change

of variable $y = x + \sqrt{t}z$ yields

$$\begin{aligned}\langle P(t)\psi, \phi\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(y)\overline{\phi(y-\sqrt{t}Z)}\frac{e^{-z^2/2}}{\sqrt{2\pi}}dydz \\ &= \int_{-\infty}^{\infty}\psi(y)\overline{\left(\int_{-\infty}^{\infty}\phi(y-\sqrt{t}Z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz\right)}dy,\end{aligned}$$

and now after the change of variables $w = -z$ this inner barred term is exactly $\overline{P(t)\phi(y)}$, so we have $\langle\psi, P(t)\phi\rangle$ as desired; thus $P(t)$ is symmetric.

For the semigroup property, we expand out the definition

$$\begin{aligned}P(s)P(t)\psi(x) &= \int_{-\infty}^{\infty} P(t)\psi(x+\sqrt{s}z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x+\sqrt{s}z+\sqrt{t}w)\frac{e^{-w^2/2}}{\sqrt{2\pi}}\frac{e^{-z^2/2}}{\sqrt{2\pi}}dwdz,\end{aligned}$$

which we can interpret as $\mathbb{E}[\psi(x+\sqrt{s}Z+\sqrt{t}W)]$ for Z, W independent standard normals; this is the same thing as $\mathbb{E}[\psi(x)+\sqrt{s+t}Z]$ because the sum of independent Gaussians is another Gaussian with variances added together, so this is indeed $P(s+t)\psi(x)$. Therefore we indeed have $P(s)P(t) = P(s+t)$ for all $s, t \geq 0$, and clearly $P(0) = I$ so the semigroup property is verified.

Finally, for strong continuity, we must show that for all $\psi \in L^2(\mathbb{R})$, the map $t \mapsto P(t)\psi$ is continuous. We will first prove this when the Fourier transform satisfies $\hat{\psi} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. By the Fourier inversion formula and Fubini's theorem, we have

$$\begin{aligned}P(t)\psi(x) &= \int_{-\infty}^{\infty}\psi(x+\sqrt{t}z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \\ &= \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{\psi}(p)e^{-ip(x+\sqrt{t}z)-z^2/2}dpdz \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{\psi}(p)e^{-ipx-p^2t/2}dp,\end{aligned}$$

where we've swapped the order of integration and then calculated the z-integral. Thus we actually have $\widehat{P(t)\psi}(p) = \hat{\psi}(p)e^{-p^2t/2}$, notice that this is analogous to what we got previously with the free particle but now without the i in the exponent. This implies that $t \mapsto P(t)\psi$ is continuous, because if $t_n \mapsto t$ then $P(t_n)\psi$ converges in L^2 to $P(t)\psi$. Furthermore (as we've already shown), $P(t)$ is a contraction, meaning that $||P(t)\psi||_{L^2} \leq ||\psi||_{L^2}$. So proving strong continuity on a dense subspace, combined with the contraction property, gives us strong continuity everywhere by the triangle inequality (with the usual strategy – for any ψ we find some ϕ within ε L^2 distance of ψ).

All of this combined with Theorem 146 gives us the Hamiltonian H that we want – that is, for any $\psi \in L^2(\mathbb{R}) \cap C_0^\infty(\mathbb{R})$,

$$H\psi = \lim_{t\downarrow 0}\frac{P(t)\psi - \psi}{-t} = -\frac{1}{2}\psi'',$$

and H is self-adjoint on its domain. But note that the actual domain for H is the Sobolev space H^1; that is,

$$\mathcal{D} = \left\{\psi \in L^2(\mathbb{R}) : p^2\hat{\psi}(p) \in L^2(\mathbb{R})\right\},$$

and on this domain H is instead more generally defined by $\widehat{H\psi}(p) = -\frac{p^2}{2}\hat{\psi}(p)$.

Next quarter, we'll start by generalizing this type of calculation to something more complicated (getting into

quantum field theory as quickly as possible). For example, if we have a general Schrodinger equation where

$$H\psi(x) = -\frac{1}{2}\psi''(x) + V(x)\psi(x)$$

for some general potential $V(x)$, we want a method to prove that solutions to the equation exist, and to do this we must figure out a stochastic process whose Hamiltonian is given by this H. It's not so easy to describe that anymore – when $V = 0$ we defined the semigroup as a solution of the heat equation, and more generally we need an understanding of Brownian motion so that we can think of $\sqrt{t}Z$ instead as a Brownian motion B_t and thus use the Feynman-Kac formula

$$P(t)\psi(x) = \mathbb{E}[\psi(x + B(t)] \exp\left(-\int_0^t V(x + B(s))ds\right).$$

Then everything will go through again – we'll get a semigroup and it'll be strongly continuous, and the generator will indeed be the H we want, so that Stone's theorem proves that a solution exists. The solution will then formally correspond to what we write down as a Feynman path integral, and that will give us the general rigorous version of Wick rotation. (This will all generalize to multiparticle systems too.)

From there, we will go into fields – fields are like having particles at every point in space, so we'll do this same story with the massive free field.

Fact 152

The basic objective of quantum field theory is to have a framework for relativistic particles. The idea is that if $\psi(t, x)$ is the state of the system at time t an position x, and we do a Lorentz transformation of spacetime, then our theory should have the property that the physics still looks the same (that is, the $\psi(t, x)$ function of four variables will be a different function after transformation, but it should still be a solution to our equation). But the ordinary Schrodinger equation is not relativistic, and so we may try to replace it with another equation but then still run into other problems like locality.

Even with our free particle example above, the solution to our Schrodinger equation looks something like

$$\psi_t(x) = \int_{-\infty}^{\infty} \psi_0(y) \frac{1}{\sqrt{4\pi}} e^{-(y-x)^2/(4i)} dy$$

for well-behaved ψ_0, but this means that "instantaneously we have influences from far away y," which shouldn't be allowed physically. Eventually what people realized is that we do need to use quantum field theory to fix this (so that the evolution of the field is actually local), and we'll see that next quarter.

From there, we'll come back to gauge theory, and the issue is that we still don't know how to do this for gauge theories with non-commuting terms (think $SU(2)$, $SU(3)$). But what we can do is replace $B(t)$ with a discretized random walk and replace our gauge theories with lattice gauge theories, and so we'll also discuss what we know in that area (similarly to the class presentations this quarter).

Remark 153. *The standard recommended reading for more of this material is a book by Glimm and Jaffe (" Quantum Physics: A Functional Integral Point of View"), but it's rather difficult to read. And for lattice gauge theories, there isn't really a standard reference yet, but that's something Professor Chatterjee is working on.*