Borcherds seminar notes, uncut
Scott Carnahan

Week 1, 26 Jan 2004, The Wightman axioms and how to fix them

Some of you might remember I gave a class a few years ago on the standard model. It ran into a few technical problems, the main one being the fact that I didn't know what I was talking about. I've learned a thing or two since then, and I'm going to try again.

Key insight in studying physics: Physics literature is usually correct, but they use mathematical terms in a different way than mathematicians do. The standard example is that when physicists say "function", they mean "distribution". A slightly less well known example is that when they say "Hilbert space", they mean "some module with a sesquilinear form over a ring of formal power series".

Physics	Mathematics
function	distribution
Hilbert space	some module with a sesquilinear form over a ring of formal power series

In particular, physicists seem to be completely uninterested in completeness, and basically never use it. There is a page in Weinberg's text which gives his definition of Hilbert space, and mentions that mathematicians like to add this extra axiom for some reason.

What is a QFT?

Quantum field theory means different things to different people. It's like the story about the three blind men examining an elephant. The definition should:

1. be rigorous
2. include QED, QCD, standard model, etc. as examples

Naive definition (that doesn't work):

We will do the case of single hermitian scalar field ϕ. Spacetime is $\mathbb{R}^{1,3}$ (with metric $x_0^2 - x_1^2 - x_2^2 - x_3^2$, although people are divided about this). [At this point, A.J. says something about general relativists vs. particle physicists - one side likes positive spacelike slices. Borcherds mentions that both sides can give plenty of good reasons why they are right and the other side is stupid, but they will end up with equivalent theories. Allen makes some obscure comment about getting different Pin groups if spacetime is nonorientable. Borcherds pretends not to hear.]

We want:

1. a space H of states

2. For each $x \in$ spacetime, a hermitian operator $\phi(x)$ on H.

3. The Lorentz group $O_{1,3}(\mathbb{R})$ (or the connected component of the identity) should act on H. It should also act on x, and $v \mapsto \phi(x)v$ should behave well under $O_{1,3}(\mathbb{R})$.

4. H contains a (possibly unique) vacuum vector VAC fixed by rotations and translations.

5. Translations act on H in some nice way. In particular, H should have "positive energy": fix some timelike vector in forward direction - operator is (at least formally) of the form e^{iE} for some self-adjoint E. E should be a **positive** operator (in the sense of its eigenvalues).

Notorious Problem: ϕ is "too singular" in x for this to work. $\lim_{x\to y} \phi(x)\phi(y)$ cannot be defined in any sensible way.

Solution: (Gårding, Wightman) Instead of regarding $\phi(x)$ as an operator-valued function, think of it as an operator-valued **distribution**. We have operators $\phi(f)$ for all $f \in \mathcal{S}(\mathbb{R}^{1,n})$, the Schwarz space. Think of $\phi(f)$ as $\int_{\mathbb{R}^{1,3}} \phi(x)f(x)dx$. [n=3 here, apparently]

Wightman axioms (for a hermitian scalar field)

- Standard references:
  - "CPT, Spin. Statistics, and all that" Streater, Wightman
  - Some book by Jost

1. $\phi(f)$ are unbounded operators: only defined on some dense subspace D of H.

2. We want to compose them: $\phi(f)\phi(g)$ so each $\phi(f)$ should map D into D.

3. Regularity: For $\Phi, \Psi \in D$, $(\Phi, \phi(f)\Psi)$ is a tempered distribution (so $\phi(f)$ is an operator-valued distribution).

4. Lorentz group acts on H with unique fixed vector VAC (up to constants). $f, v \mapsto \phi(f)v$ is invariant under Lorentz transforamtions.

5. H is also acted on by translations of spacetime. Action is "positive energy".

6. $[\phi(f), \phi(g)] = 0$ if the supports of f and g are spacelike separated.

7. D is generated by the action of operators $\phi(f)$ on VAC. So vectors of D are linear combinations of $\phi(f_1)\dots\phi(f_n)VAC$. People use the phrase: "Vacuum is cyclic"

8. Miscellaneous axioms I forgot to write down.

The Wightman axioms work fine for:

1. Free field theories (see later)

2. $1 + 1$ dimensions (see Glim-Jaffe)

3. 1 isolated example in 3 dimensions. This is some degenerate super renormalizable theory.

There are no nontrivial examples known in 4 dimensions. In particular, QED, QCD, and the standard model are not known to satisfy them. They are perturbative theories, so they work over formal power series. The Wightman axioms don't allow for formal power series - they only work over $\mathbb{C}$.

There is a solution, due to Alexander the Great, which is to cheat, and change the problem to one you **can** solve. If the main examples of a theory do not satisfy your axioms, this does **not** indicate a problem with the examples, but with the **axioms**. Thus, we should change the Wightman axioms slightly, to allow perturbative theories.

Advantages:

1. We get rigorous axioms

2. Many theories using Wightman axioms still work for the perturbative case.

3. QED, etc. satisfy these axioms.

Changes to Wightman:

1. Work over a ring $\mathbb{C}[[\lambda]]$ of formal power series instead of over $\mathbb{C}$

2. We can recover H from D as its completion, so we may as well discard the Hilbert space, and just work with D. Instead of a Hilbert space, we work with the module D over $\mathbb{C}[[\lambda]]$. D should have a sesquilinear form with values in $\mathbb{C}[[\lambda]]$. [At this point, Marty asks if this should be the space of smooth vectors or the Gårding subspace of H under a unitary representation of $O_{1,3}$. Borcherds replies that it doesn't really matter, as long as it satisfies the axioms.]

3. Drop the condition that $(,)$ on D should be positive (in some sense). We are mainly interested in the positive definite case, **but** the indefinite case often turns up during the construction of positive definite examples. Standard example: Ghosts.

Other variations (also included in the usual Wightman axioms)

1. Instead of taking ϕ fixed by $O_{1,3}(\mathbb{R})$, can take ϕ in some finite-dimensional real space acted on by $O_{1,3}(\mathbb{R})$ e.g. vector fields. We would have $\phi(f)$ for $\phi \in$ some real representation Φ, f Schwarz.

2. Can allow "fermions". Φ and H become "super" vector spaces.

I'm going to be incredibly rude and stop in the middle of my sentence.

Week 2, 2 Feb 2004, Wightman axioms (continued)

Recap and continue extra structures:

1. Action of $O_{1,d-1}(\mathbb{R})$ on Φ.

2. Fermions. All spaces are graded by $\mathbb{Z}/2\mathbb{Z}$, i.e. "superspaces". These have a sign rule: when you exchange the order of A and B, you must add a sign of $(-1)^{\deg A \deg B}$ e.g. the rule that $[\phi(f), \psi(g)] = 0$ if the supports of f and g are spacelike separated becomes:

$$\phi(f)\psi(g) - (-1)^{\deg\phi \deg\psi}\psi(g)\phi(f) = 0$$

   So for fermionic fields ϕ, ψ, $\phi(f)\psi(g) + \psi(g)\phi(f) = 0$ if supports are spacelike separated. For fermionic fields, we do **not** have an action of $O_{1,d-1}(\mathbb{R})$, but instead one of the connected, simply connected group $\mathrm{Spin}_{1,d-1}(\mathbb{R})$. This group has center of order 2, giving a homomorphism to $O_{1,d-1}(\mathbb{R})$ which is not onto, since the latter group has four connected componenets for $d > 1$ (determinant can be positive or negative, and the direction of time can be changed: spinor norm is positive or negative). Note that weak interactions are **not** invariant under $O_{1,d-1}(\mathbb{R})$.

3. We want to allow "ghosts". The spin-statistics theorem (which can be found early in any book on Wightman axioms) says that bosonic fields are exactly those with integral spin, and fermionic fields are exactly those with half-integral spin. Ghost fields violate this: They can avoid the spin-statistics theorem, because the inner product on the space of states is no longer positive definite. Everything has a ghost grading in $\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z}$, although it doesn't matter which one, since we are only concerned with parity. In a Hilbert space, we expect $(b, a) = \overline{(a, b)}$, but we get $(b, a) = (-1)^{ghost(a)ghost(b)}\overline{(a, b)}$.

[Note added much later: The ghost grading is in fact unnecessary. It is possible to formulate equivalent theories with and without this additional grading, and there is no significant advantage to adding this extra complication.]

A common mistake is to try to identify the ghost grading with the fermionic grading. The problem is that $(a, a) = (-1)^{ghost(a)ghost(a)}\overline{(a, a)}$, so (a, a) is imaginary if a has odd ghost degree. If things of odd fermionic number have imaginary norm, it messes up attempts to construct a positive definite Hilbert space. This leads to weird (by which I mean unmemorable) conventions, like $(a, a) > 0$ if a has even degree and $i(a, a) > 0$ (or < 0) if a has odd degree.

Question: When do you use the fermionic minus sign and when do you use the ghost minus sign?

Before I answer that, I will explain my **greatest discovery**. I have solved the Fundamental Problem of Hilbert space theory: Should the inner product $(,)$ be linear on the **left** or **right**?

The key point of the solution is to think of a Hilbert space H as a **bimodule** over $\mathbb{C}$, by using complex conjugation as a sort of "antipode" like in a Hopf algebra. We **define** the right action of $\mathbb{C}$ on H by $v\lambda = \bar{\lambda}v$ for all $\lambda \in \mathbb{C}, v \in H$.

Similarly, if A is an operator on H, define the **right** action of A on H by $vA = A^\dagger v$ (where † = adjoint). The inner product now becomes a bilinear map of bimodules from $H \times H$ to $\mathbb{C}$, and the action follows a natural associative law:

$$\lambda(u, v) = (\lambda u, v) = (u\bar{\lambda}, v) = (u, \bar{\lambda}v) = (u, v\lambda) = (u, v)\lambda$$

So (u, v) is **left** linear in u, **right** linear in v, and the left and right $\mathbb{C}$-module structures are related by complex conjugation. In particular, for left Hilbert spaces, $(,)$ should be linear in the left variable. You can also make right Hilbert spaces if you're one of those people who happens to like things acting on the right for some reason. Similarly, we have for operators A, $(Au, v) = (uA^\dagger, v) = (u, A^\dagger v)$. Now define:

- $\lambda^\dagger := \lambda$ for $\lambda \in \mathbb{C}$ (view it as a constant-multiplication operator)
- $A^\dagger :=$ usual adjoint for operators
- $v^\dagger := v$ for $v \in H$. The reason you haven't seen this one before is that no one has bothered to come up with a special notation for the identity operator.

Then we have

$$(ab)^\dagger = b^\dagger a^\dagger$$

for a and b complex numbers, vectors, or operators, whenever it makes sense.

What about ghost grading?

Replace $(ab)^\dagger = b^\dagger a^\dagger$ by $(ab)^\dagger = (-1)^{ghost(a)ghost(b)} b^\dagger a^\dagger$ everywhere.

When do you use:

$(-1)^{ghost(a)ghost(b)}$ - when the interchange is done using the adjoint operator

$(-1)^{deg(a)deg(b)}$ - when making things commute or anticommute.

Modified Wightman axioms:

1. Allow ghosts.
2. Work over a formal power series ring instead of $\mathbb{C}$.

Ingredients:

All spaces should be $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ graded by fermion number and ghost number.

1. A ring R that we work over. R can be a formal power series ring in several variables, e.g. $\mathbb{R}[[\lambda]]$. This case covers all interesting examples. Everything also works if R is an inverse limit of finite-dimensional algebras over $\mathbb{R}$. Alternatively, R is the dual of some coalgebra over $\mathbb{R}$.

2. A real vector space Φ of hermitian fields. (Note: elements of $\mathbb{C} \otimes \Phi$ are called "complex fields")

3. A space of states D, a real vector space. This is sometimes written $D[[\lambda]]$. $D[[\lambda]]$ is a module over $\mathbb{R}[[\lambda]]$. Think of D as a dense subspace of some Hilbert space, though it need not be. D should have a sesquilinear inner product $(,)$ satisfying $\lambda(u, v) = (\lambda u, v) = (u, \bar{\lambda} v)$ and $\overline{(u, v)} = (-1)^{ghost(u)ghost(v)}(v, u)$. Assume $(,)$ is nondegenerate. We do **not** assume it is positive definite (though we prefer it to be).

4. An action of the Lie algebra of translations of $\mathbb{R}^{1,d-1}$ on D.

5. An action of $\mathrm{Spin}_{1,d-1}$ on Φ and on D. Technical assumption: ϕ is a sum of finite-dimensional algebraic representations of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$. D is **not** usually a sum of finite-dimensional representations.

6. A Field map: For $\phi \in \Phi$ a real Schwarz function on $\mathbb{R}^{1,d-1}$, we have a self-adjoint operator $\phi(f)$ from $D[[\lambda]]$ to $D[[\lambda]]$ such that:

   - it is linear in ϕ.
   - tempered: $(\Phi, \phi(f)\Psi)$ is a tempered distribution for all $\Phi, \Psi \in D$.

   [The letter Φ is used for two different things here] Remark: $(\Phi, \phi(f)\Psi)$ takes values in $R = \mathbb{R}[[\lambda]]$, so when we say it is a distribution, we just mean the coefficient of any power of λ is a distribution. $\partial(\phi(f)v) = \phi(\partial f)v + \phi(f)\partial v$ for ∂ an infinitesimal translation.

7. A vacuum vector $VAC \in D$, fixed by translations, rotations, and normalized such that $(VAC, VAC) = 1$.

Main Axioms:

1. Causality - $[\phi(f), \psi(g)] = 0$ if f and g are spacelike separated.
2. Positive energy - explanation postponed until next week.
3. Lorentz invariance
4. Vacuum is cyclic - $D[[\lambda]]$ is generated by the action of $\phi(f)$ of VAC.

Week 3, 9 Feb 2004, Free massive fields

Warning: The sign in $\overline{(a,b)} = (-1)^{ghost(a)ghost(b)}(\bar{b}, \bar{a})$ **might** be unnecessary. About 47% of the time I work it out it goes one way, and about 53% of the time it goes the other, so I can't give you an answer until it stabilizes somewhat.

[Note added much later: the ghost grading seems to be unnecessary.]

I said I was going to continue talking about the Wightman axioms, but I got a bit bored of them, so instead I will describe a basic example of a quantum field theory. This is given by a free massive hermitian (bosonic) scalar field.

Massive: opposite of massless

Hermitian: this means $\phi \in \Phi$ (not $\mathbb{C} \otimes \Phi$).

Scalar: The Lorentz group acts trivially on Φ.

Examples: Higgs bosons, scalar mesons.

Main idea for constructing free hermitian fields ϕ:

Write ϕ as a sum of 2 fields, ϕ^+ and ϕ^-. ϕ^+ will be the creation field, and ϕ^- will be the annihilation field - it "kills particles". Note that ϕ^+ and ϕ^- are **not** in Φ.

For historical reasons, ϕ^+ is sometimes called ϕ^-, and ϕ^- is sometimes called ϕ^+. At some point during the early days of quantum physics, someone made a sign error, and it has propogated ever since. It's not clear why anyone would use a minus sign to indicate creation.

Commutation relations:

- $[\phi^+(f), \phi^+(g)] = [\phi^-(f), \phi^-(g)] = 0$.
- $[\phi^+(f), \phi^-(g)]$ = some scalar (in $\mathbb{C}$) depending on f and g. Instead of an arbitrary operator, you end up just multiplying by a constant.
- $\phi^-(f)(VAC) = 0$.

These are basically the characteristic relations of any free field. We need to choose the scalar. $[\phi(f), \phi(g)]$ is some sort of bilinear form on f and g, and should be

Lorentz-invariant and translation invariant, so we try:

$$\int \hat{f}(p)\hat{g}(-p) \times (\text{some Lorentz-invariant measure})d^4p$$

where $\hat{f}$ and $\hat{g}$ are Fourier transforms:

$$\hat{f}(p) := \int e^{2\pi ixp} f(x)dx$$

The measure is almost the simplest possible. [some drawing of light cones and hyperbolic sheets goes here] Choose a positive real m. This is the mass. Take the set of vectors in momentum space ($\mathbb{R}^{1,3}$) satisfying $p^2 = m^2$ and $p > 0$ (in particular, ignore the bottom component of the hyperboloid), and take a Lorentz-invariant measure supported on that set. This measure is unique up to scalars (One could take other measures, e.g. supported on more than one hypersurface - this gives "generalized free field theories" which don't offer significant advantages). The statement that "mass = m" is equivalent to the assertion that the measure has support on (the positive part of) the hypersurface $p^2 = m^2$.

Main problem: need to check causality condition - that $[\phi(f), \phi(g)] = 0$ if f and g are spacelike separated.

$$\begin{aligned}[\phi^+(f) + \phi^-(f), \phi^+(g) + \phi^-(g)] &= [\phi^+(f), \phi^-(g)] - [\phi^+(g), \phi^-(f)] \\ &= \int \hat{f}(p)\hat{g}(-p) - \hat{g}(p)\hat{f}(-p) \times (\text{measure})d^4p \\ &= \int e^{2\pi ixp} f(x)e^{-2\pi iyp} g(y) \times M(p)dpdxdy\end{aligned}$$

Vital: M changes sign under rotations changing the direction of time. Take the invariant measure supported on the positive part of $p^2 = m^2$, and subtract the corresponding measure supported on the negative part.

The condition that the commutator vanishes for $x - y$ spacelike is equivalent to the condition that the Fourier transform of $M(p)$ vanishes on spacelike values. So causality for bosonic scalar fields reduces to the following property of the measure M:

$$\hat{M}(x) = 0 \text{ if } x \text{ is spacelike}$$

M has the following properties:

1. Invariant under time-preserving rotations

2. Changes sign under time-reversing rotations

The formula for M can be given explicitly using Bessel functions, but properties 1 and 2 are sufficient to show that $\hat{M} = 0$ where we want it to.

1. If M transforms as above, so does $\hat{M}$. This is a trivial fact about invariance under group actions.
2. If a measure $\hat{M}$ has these transformation properties, then $\hat{M}$ vanishes on spacelike vectors.

Proof: If M is a function, the result is trivial: For any spacelike vector x, there is a time-reversing rotation σ fixing x. We then have $-M(x) = M(\sigma x) = M(x)$, so $M(x) = 0$. For distributions, one needs to work slightly harder, so this is left as an exercise. The main point is that any rotation-invariant distribution on spacelike vectors depends only on the "radius" $r = \sqrt{-(x,x)}$

Define D_1 to be the space of Schwarz functions on $\{p|p^2 = m^2, p > 0\}$. This is the space of 1-particle states. Let D be the symmetric algebra of D_1. We identify D_1 with the space of states of the form $\phi^+(f)VAC$ (look at $\hat{f}$ on $\{p^2 = m^2, p > 0\}$). By the commutation relations, $\phi^+(f_1)\phi^+(f_2)\ldots\phi^+(f_n)(VAC)$ can be identified with the symmetric algebra product of $f_1, \ldots, f_n$.

How does $\phi^-(g)$ act? It is uniquely determined by:

1. $\phi^-(g)(VAC) = 0$
2. $[\phi^+(f), \phi^-(g)]$ is known

For example:

$$\begin{aligned}\phi^-(g)\phi^+(f_1)\phi^+(f_2)(VAC) &= [\phi^-(g), \phi^+(f_1)]\phi^+(f_2)(VAC)\\ &+ \phi^+(f_1)[\phi^-(g), \phi^+(f_2)](VAC)\\ &+ \phi^+(f_1)\phi^+(f_2)\phi^-(g)(VAC)\end{aligned}$$

and the last term necessarily vanishes.

D is positive-definite, because the inner product $(,)$ is determined by the condition that ϕ^+ is adjoint to ϕ^-. When you take the Fourier transform, something funny happens, but since we're out of time, I won't explain that.

Week 4, 16 Feb 2004, Presidents' Day - no class

The following people were seen trying the (locked) door of room 939 to see if the seminar would happen:

1. Nicolai Reshetikhin

Week 5, 23 Feb 2004, Spinors and Clifford algebras

Today and possibly next week, I will be talking about everything worth knowing about spinors and Clifford algebras. For our purposes, we only need to describe Clifford algebras for $\mathbb{R}^{0,n}$ and $\mathbb{R}^{1,n}$, but out of some kind of bloody-mindedness, we will define them for **all** fields. [AJ: even characteristic 2?] ... including characteristic 2, which is the case that scares people. It turns out that the reason people are scared is that they are using the wrong definitions. We will be doing the characteristic 2 case in order to force ourselves to use the best definition.

Let V be a finite-dimensional vector space over a field F, with quadratic form N. This is a degree 2 map from V to F such that $N(a+b) - N(a) - N(b)$ is bilinear in a and b.

Example: If $(,)$ is a symmetric bilinear form, then $N(a) = (a, a)$ is a quadratic form. If $\mathrm{char}(F) \neq 2$, then quadratic forms are equivalent to symmetric blinear forms. In characteristic 2, quadratic forms are the **right** thing to use, and bilinear forms are **wrong**.

Warning: If $N(a) = (a, a)$, then the bilinear form of N is two times the original one. (You can make quadratic forms from bilinear forms in characteristic 2 also, but they tend to be rather degenerate.) We say V is nondegenerate if the **bilinear** form of N is nondegenerate.

The Clifford algebra $C_V(F)$ of V is the algebra generated by the vector space V with relations $Q(v) = v^2$ for $v \in V$. I'm going to skip the details of the construction, since you know how it goes.

At this point, most people spend time proving that they have dimension 2^n, but I'm not going to bother. Instead, I will just tell what they are for the cases we care about.

Let's calculate $C_V(\mathbb{R})$ for all non-degenerate quadratic forms over $\mathbb{R}$. We write $C_{m,n}(\mathbb{R})$ for the Clifford algebra of $x_1^2 + \ldots + x_m^2 - x_{m+1}^2 - \ldots - x_{m+n}^2$. The first few cases are easy to compute and are as follows:

$$C_{0,0} \cong \mathbb{R} \qquad C_{0,1} \cong \mathbb{C} \qquad C_{1,0} \cong \mathbb{R} \oplus \mathbb{R}$$
$$C_{0,2}(\mathbb{R}) \cong \mathbb{H} \quad C_{1,1}(\mathbb{R}) \cong C_{2,0}(\mathbb{R}) \cong M_2(\mathbb{R})$$

After these, it becomes a bit tiresome to figure them out directly, so we use a lemma:

$$C_{m+2,n}(\mathbb{R}) \cong C_{n,m}(\mathbb{R}) \otimes M_2(\mathbb{R})$$
$$C_{m+1,n+1}(\mathbb{R}) \cong C_{m,n}(\mathbb{R}) \otimes M_2(\mathbb{R})$$
$$C_{m,n+2}(\mathbb{R}) \cong C_{n,m}(\mathbb{R}) \otimes \mathbb{H}$$

Note the order reversal in the first and third formulas.

More generally: Suppose dim U is even, U non-degenerate. Then

$$C_{U\oplus V} \cong C_U(F) \otimes C_{V(-\mathrm{disc}\ U)}(F)$$

where $V(-\mathrm{disc}\ U)$ is the vector space V, with the norms of the vectors given by taking the usual norms and multiplying by minus the discriminant of U.

Sketch of proof: We can assume char $F \neq 2$ since the characteristic 2 case is trivial. Then we can choose orthogonal bases $\gamma_1, \ldots, \gamma_m$ for U and $\gamma_{m+1}, \ldots$ for V. (this always exists in characteristic not 2, and is never possible in characteristic 2 if the form is nondegenerate). Consider elements $\gamma'_{m+k} := \gamma_1\gamma_2 \ldots \gamma_m\gamma_{m+k}$. An easy calculation shows:

1. γ'_{m+k} commutes with γ_i $(i \leq m)$
2. $(\gamma'_{m+k})^2 = -\mathrm{disc}(U)\gamma^2_{m+k}$

and $-\mathrm{disc}(U)$ turns out to be $\gamma_1^2\gamma_2^2 \ldots \gamma_m^2$, so $\gamma_1, \ldots, \gamma_m, \gamma'_{m+1}, \ldots$ satisfy the relations for $C_U \otimes C_{V(-\mathrm{disc}\ U)}$.

Now, to calculate all Clifford algebras over $\mathbb{R}$, we just need to recall how to take tensor products of matrix algebras:

$$\begin{array}{cc} \mathbb{R}\otimes \text{ anything } \cong \text{ anything} & \mathbb{C}\otimes\mathbb{C}\cong\mathbb{C}\oplus\mathbb{C} \\ \mathbb{C}\otimes\mathbb{H}\cong M_2(\mathbb{C}) & \mathbb{H}\otimes\mathbb{H}\cong M_4(\mathbb{R}) \\ M_a(X)\otimes M_b(Y)\cong M_{ab}(X\otimes Y) & M_a(M_b(X))\cong M_{ab}(X) \end{array}$$

[Actually, we can do away with the $\mathbb{C}\otimes\mathbb{C}$, but might want $M_a(X \oplus Y) \cong M_a(X) \oplus M_a(Y)$] So from the lemma, we have for $C_{m,n}(\mathbb{R})$:

$m \backslash n$	0	1	2	3	4	5	6	7	8
0	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	$\mathbb{H}\oplus\mathbb{H}$	$M_2(\mathbb{H})$	$M_4(\mathbb{C})$	$M_8(\mathbb{R})$	$M_8(\mathbb{R}\oplus\mathbb{R})$	$M_{16}(\mathbb{R})$
1	$\mathbb{R}\oplus\mathbb{R}$								
2	$M_2(\mathbb{R})$								
3	$M_2(\mathbb{C})$								
4	$M_2(\mathbb{H})$								
5	$M_2(\mathbb{H}\oplus\mathbb{H})$								
6	$M_4(\mathbb{H})$								
7	$M_8(\mathbb{C})$								
8	$M_{16}(\mathbb{R})$								

The rest of the table is filled in by multiplying the dimension of the matrix algebra by 2 each time you go down and to the right. Each time you go down by 8 or to the right by 8, you multiply the dimension of the matrix algebra by 16. That is a complete description of all useful Clifford algebras.

Now, we write $C_V(F) = C_0(F) \oplus C_1(F)$, where $C_0(F)$ is spanned by products of even numbers of γ_i (basis vectors for V), and $C_1(F)$ is spanned by products of odd numbers of γ_i. This splitting is possible because the defining relations of the Clifford algebra only had terms of even degree, so degree is defined mod 2. $C_0(F)$ is the object of interest for us, since it gives us spinors.

Lemma: $C^0_{U\oplus V} \cong C_{V(\mathrm{disc}(U))}$ for U 1-dimensional and nondegenerate.

Proof: Let γ be a basis for U, and $\gamma_1, \ldots, \gamma_n$ a basis for V. Then $\gamma\gamma_1, \ldots, \gamma\gamma_n$ generate $C^0_{U\oplus V}$, and $(\gamma\gamma_i)^2 = -\gamma_i^2\gamma^2 = \gamma_i^2(-\mathrm{disc}(U))$, etc. The relations are satisfied.

This gives you everything you want to know about the Clifford algebra. Now we define the Clifford group Γ_V. We will find:

$$\begin{array}{ccccccccc}
1 & \to & F^\times & \to & \Gamma_V(F) & \to & O_V(F) & \to & 1 \\
 & & \cup & & \cup & & \| & & \\
1 & \to & \pm 1 & \to & \mathrm{Pin}(F) & \to & O_V(F) & & \\
 & & \| & & \cup & & \cup & & \\
1 & \to & \pm 1 & \to & \mathrm{Spin}(F) & \to & SO_V(F) & &
\end{array}$$

Where $O_V(F)$ is the orthogonal group of V. Note that the last map on the second row is usually not surjective.

We are interested in the Spin group, but it is easier to define the Clifford group and look for Spin inside it.

Definition of $\Gamma_V(F)$: Let $\alpha(v) = v$ for $v \in C^0_V(F)$, and $\alpha(v) = -v$ for $v \in C^1_V(F)$, so α is the unique automorphism of $C_V(F)$ induced by $v \mapsto -v$. (If you are interested in categories, you can think of Clifford algebras as a sort of functor, and α as the image of the inversion automorphism)

$\Gamma_V(F)$ is the set of invertible elements $x \in C_V(F)$ such that $xV\alpha(x)^{-1} \subset V$. (note that $V \subset C_V(F)$.

A common "error" among authors: using $xVx^{-1} \subset V$ instead of $xV\alpha(x)^{-1} \subset V$. It's not really an error since the theory can still be made to work, but it makes things much more difficult. Old books tend to use xVx^{-1} and newer books tend to use $xV\alpha(x)^{-1}$ so people seem to have caught on to this.

First reason: If $x \in V$, $v \mapsto xv\alpha(x)^{-1}$ is the reflection in x: $v \mapsto v - \frac{(v,x)}{N(x)}x$. The old way gives negative reflections, and you should use refletions whenever possible, since they fix almost everything. They are the simplest rotations.

Warning: In characteristic 2, these are often called "orthogonal transvections". The only apparent reason for this is that authors want to make characteristic 2 sound difficult.

[Marco: What kind of action do you get from mixed-type elements in $\Gamma_V(F)$? Borcherds: There turn out not to be any.]

Lemma: If $x \in \Gamma_V(F)$ acts trivially on V, then $x \in F^\times \subset C_V(F)$.

This is an easy exercise, and is false if we use the action $v \mapsto xvx^{-1}$.

Now, we want to show $\Gamma_V(F)$ acts as rotations on V.

Define $N^\alpha(x) = \alpha(x)^T x$, where T is the transpose: $(\gamma_1 \dots \gamma_n)^T = (\gamma_n \dots \gamma_1)$. T is the antiautomorphism of $C_V(F)$ fixing all elements of V. As it happens, T is not all that important - we will just use it in one or two proofs. α is also not terribly important - we will eventually throw it away and replace it with something better.

An easy calculation shows: $N^\alpha(x) \in F^\times$ if $x \in \Gamma_V(F)$. This is because $N^\alpha(x)$ acts trivially on V. Next:

$$\begin{aligned} N^\alpha(xy) &= \alpha(y)^T \alpha(x)^T xy \\ &= \alpha(y)^T N^\alpha(x) y \\ &= \alpha(y)^T y N^\alpha(x) \\ &= N^\alpha(y) N^\alpha(x) \end{aligned}$$

So N^α is a homomorphism from $\Gamma_V(F)$ to $F^\times$. Note that this is **not** a homomorphism from the Clifford algebra, and doesn't map its elements to F.

Since $N^\alpha(v) = -N(v)$ for $v \in V$, we see that $\Gamma_V(F)$ preserves $(-)N$ on V, so $\Gamma_V(F) \to O_V(F)$. $O_V(F)$ is generated by reflections if char $F \neq 2$, and there is a unique counterexample in characteristic 2: $F = \mathbb{F}_2$, $V = A \oplus B$, where both A and B have dimension 2 and quadratic form $x^2 + y^2 + xy$. Reflections generate an index 2 subgroup in this case. However, in **all** cases, the map $\Gamma_V(F) \to O_V(F)$ is **onto**, so we get:

$$1 \to F^\times \to \Gamma_V(F) \to O_V(F) \to 1$$

In particular, the Clifford group is a central extension of $O_V(F)$. Before we go on, we ought to tidy up notation a bit: We can split $\Gamma_V(F) = \Gamma_V^0(F) \cup \Gamma_V^1(F)$, where $\Gamma_V^i(F) = \Gamma_V(F) \cap C_V^i(F)$ This can be checked in all but one case by looking at reflections.

We define the Dickson invariant: a homomorphism $\Gamma_V(F) \to \mathbb{Z}/2\mathbb{Z}$ given by $\Gamma_V^i(F) \to i$. In characteristic not 2, $\det(x) = (-1)^{\text{Dickson invariant}}$, while in characteristic 2, the determinant is always 1, so it carries strictly less information.

The spinor norm is defined as: $N(x) = N^\alpha(x)(-1)^{\text{Dickson invariant}}$. This is a homomorphism $\Gamma_V(F) \to F^\times$. Notice that is coincides with the norm N on $\{v \in V | N(v) \neq 0\}$. It extends the homomorphism, and is uniquely defined by this property.

Week 6, 1 March 2004, Spinors and Clifford algebras, continued

Recall from last time:

1. V, a vector space over a field F, with a quadratic form N (or Q, if I'm not paying attention). This gives rise to a corresponding bilinear form $(u, v) \mapsto N(u+v) - N(u) - N(v)$.

2. A Clifford algebra $C_V(F)$ generated by V with relations $a^2 = N(a)$.

3. The Clifford group $\Gamma_V(F)$ of invertible elements x of the Clifford algebra satisfying $xV\alpha(x)^{-1} \subset V$, where if we have the splitting $C_V(F) = C_V^0(F) \oplus C_V^1(F)$, α is the antiautomorphism which acts as -1 on $C_V^1(F)$ and as 1 on $C_V^0(F)$.

Note: Some authors leave out the α, which leads to problems later on.

T is the antiautomorphism of $C_V(F)$ which acts as identity on V, and such that $(v_1 v_2 \ldots v_n)^T = v_n v_{n-1} \ldots v_1$.

We define the twisted spinor norm N^α be $N^\alpha(x) = \alpha(x)^T x$. This is not too interesting, but it is convenient for proofs. We define the spinor norm N by $N(x) = x^T x$.

We showed N^α is a homomorphism $\Gamma_V(F) \to F^\times$ (**not** a homomorphism on $C_V(F)$), and that $\Gamma_V(F) = \Gamma_V^0(F) \cup \Gamma_V^1(F)$, where $\Gamma_V^i(F) = \Gamma_V(F) \cap C_V^i(F)$. $\Gamma_V(F)$ is generated by V if $\dim(V) > 0$ (except for one case over $F = \mathbb{F}_2$. This is just in case there are any Gabbers out there).

Recall $N(x) = N^\alpha(x)(-1)^{\text{Dickson invariant}(x)}$, where the Dickson invariant is 0 if $x \in \Gamma_V^0(F)$ and 1 if $x \in \Gamma_V^1(F)$. The spinor norm is defined uniquely (for $V \neq 0$) by $N(v) = N(v)$ for all $v \in V$ with $N(v) \neq 0$. Here, the first N is spinor norm, and the last two are the quadratic form. Since these two things actually coincide here, we may as well use the same letter for them.

Note: Many authors define N^α to be the spinor norm, so that spinor norm $=$ $-$norm on V, which introduces a pointless, unmemorable sign.

We get an exact sequence which can be uniquely extended to a diagram:

$$\begin{array}{ccccccccc} 1 & \to & F^\times & \to & \Gamma_V(F) & \to & O_V(F) & \to & 1 \\ & & & \searrow & \downarrow N & & \downarrow N & & \\ & & & & F^\times & \to & F^\times/(F^\times)^2 & \to & 1 \end{array}$$

for all F and all V, where the diagonal arrow is the squaring map $x \mapsto x^2$, and $F^\times/(F^\times)^2$ is the image if $F^\times$ under N.

So we get a spinor norm homomorphism $O_V(F) \to F^\times/(F^\times)^2$ uniquely defined by $N(\text{reflection of } v \in V, N(v) \neq 0)$ is $N(v)$.

For people who like weird and irritating exceptions, you may recall there is one example where Γ is not generated by V. However, the spinor norm homomorphism is still uniquely defined even for this irritating example, since $F^\times$ is trivial.

Example: Suppose V is positive or negative definite over $\mathbb{R}$. In this case, we have $F^\times/(F^\times)^2 \cong \{\pm 1\}$. If V is positive definite, $N(v) = 1$, since the norm of any reflecting vector is positive. Thus spinor norm is identically 1. If V is negative definite, then $N(\text{reflection}) = -1 \in \mathbb{R}^\times/(\mathbb{R}^\times)^2$, so spinor norm = determinant.

Now, suppose V has vectors of positive greater than 0 and of norm less than 0.

For reflections of norm > 0, $\det = -1$ and $N = 1$

For reflections of norm < 0, $\det = -1$ and $N = -1$

so $\det \times N$ maps $O_V(\mathbb{R})$ **onto** $\{-1, 1\} \times \{-1, 1\}$ if V is indefinite. Thus, $O_V(\mathbb{R})$ has at least 4 components (in fact exactly four). $O_V(\mathbb{R})$ has 2 components if V is definite (and if $\dim V > 0$ for all the Gabbers in the audience. Gabber was well-known for pointing out cases involving the zero dimensional vector space or the empty set, and people found it rather annoying, but someone's proof of the Atiyah-Singer index theorem fell apart when Gabber pointed out that the union of two vector spaces is not a vector space.)

For Lorentzian vector spaces, where $V = \mathbb{R}^{n,1}$, this is easy to see. A rotation has spinor norm $> 0/ < 0$ if it fixes/exchanges the two light cones. [drawing of light cones]

Definition: $\text{Pin}_V(F)$ = kernel of the map $N : \Gamma_V(F) \to F^\times$. $x \in F^\times$ and $N(x) = 1$ if and only if $x = \pm 1$, so we get and exact sequence: $1 \to \{\pm 1\} \to \text{Pin}_V(F) \to O_V(F)$. The last map is not onto in general, for example if $V = \mathbb{R}^{n,1}$. (Note that some authors define $\text{Pin}_V(F)$ to be the elements with $N(x) = \pm 1$. This makes $\text{Pin}_V(\mathbb{R}) \to O_V(\mathbb{R})$ **onto** over the reals, but messes things up completely over other fields.)

Subtle point: Pin_V and O_V are algebraic groups over F. This basically means they are functors from the category of F-algebras to the category of groups. The map $\text{Pin}_V \to O_V$ is **surjective** as a map of algebraic groups, but $\text{Pin}_V(F) \to O_V(F)$ need not be surjective.

If you're into Galois cohomology, I'll mention you get a Galois cohomology exact sequence:

$$1 \to \pm 1 \to \text{Pin}_V(F) \to O_V(F) \xrightarrow{N} H^1(\text{Gal}(\overline{F}/F), \pm 1) \to H^1(\text{Gal}(\overline{F}/F), \text{Pin}_V(F)) \to$$

where $H^1(\text{Gal}(\bar{F}/F), \pm 1) \cong F^\times/(F^\times)^2$ if F is perfect [actually unnecessary] and characteristic not 2.

In general, when you have an exact sequence of Galois modules, you get a long exact sequence of Galois cohomology, except this one isn't very long, since the groups are nonabelian.

We define $\text{Spin}_V(F)$ to be the kernel of the Dickson invariant $\text{Pin}_V(F) \to \{0, 1\}$. If $\text{char}(F) \neq 2$, then this is the same as the kernel of $\det : \text{Pin}_V(F) \to O_V(F) \xrightarrow{\det} \{\pm 1\}$.

(Note: $SO_V(F)$ should be defined as elements of Dickson invariant 0, as this also works if $\text{char}(F) = 2$. If you are not working in characteristic 2, it doesn't really matter, but since in characteristic 2 it is just as easy to get the definition wrong as it is to get it right, you might as well use the correct definition.)

We have inclusions: $\text{Spin}_V(F) \subset \Gamma_V^0(F) \subset C_V^0(F) \subset C_V(F)$, and homomorphism $\text{Spin}_V(F) \to SO_V(F)$ which is not always onto. Representations of $C_V^0(\mathbb{R})$ give representations of $\text{Spin}_V(\mathbb{R})$, called spinor representations. I never quite figured out when they are called spin and when they are called half-spin.

Examples: $\text{Spin}_{1,3}(\mathbb{R}) \subset C_{1,3}^0(\mathbb{R})$. Last lecture, we showed how to work out every possible Clifford algebra you could conceivably be interested in, and you will see that $C_{1,3}(\mathbb{R}) \cong M_2(\mathbb{C})$, so $\text{Spin}_{1,3}(\mathbb{R})$ has a 4 dimensional real representation, with a complex structure. We get a homomorphism $\text{Spin}_{1,3}(\mathbb{R}) \to GL_2(\mathbb{C})$. A little more work gives an **isomorphism** $\text{Spin}_{1,3}(\mathbb{R}) \cong SL_2(\mathbb{C})$

The last one is widely used in physics textbooks, and I think it is a bad idea, because if you rely on this accidental isomorphism, it means you can't work out Quantum Field Theory in any dimension other than 4.

Exercises:

$$\begin{aligned}
\text{Spin}_{1,2}(\mathbb{R}) &\cong SL_2(\mathbb{R}) \\
\text{Spin}_{1,5}(\mathbb{R}) &\cong SL_2(\mathbb{H}) \\
\text{Spin}_{2,2}(\mathbb{R}) &\cong SL_2(\mathbb{R}) \times SL_2(\mathbb{R}) \\
\text{Spin}_{3,3}(\mathbb{R}) &\cong SL_4(\mathbb{R})
\end{aligned}$$

4 dimensional orthogonal groups often split like on the third line above, but not always. I want to finish this lecture mainly by giving warnings about standard mistakes.

Warning:

$$\begin{aligned}
O_{m,n}(\mathbb{R}) &\cong O_{n,m}(\mathbb{R}) \\
SO_{m,n}(\mathbb{R}) &\cong SO_{n,m}(\mathbb{R}) \\
\text{Spin}_{m,n}(\mathbb{R}) &\cong \text{Spin}_{n,m}(\mathbb{R}) \\
\text{Pin}_{m,n}(\mathbb{R}) &\ncong \text{Pin}_{n,m}(\mathbb{R})
\end{aligned}$$

The easiest example of this is $\text{Pin}_{1,0}(\mathbb{R}) \cong (\mathbb{Z}/2\mathbb{Z})^2$ and $\text{Pin}_{0,1}(\mathbb{R}) \cong \mathbb{Z}/4\mathbb{Z}$.

Why is $O_V(F)$ not a simple group?

It turns out it is essentially simple, but it has a large amount of rubbish attached to it. We'll just list the possible reasons for a group to be not simple, and if none of them apply, we have a simple group.

1. Dickson invariant $O_V(F) \to \mathbb{Z}/2\mathbb{Z}$ (sometimes called the quasi-determinant or pseudo-determinant)
2. Spinor norm $N : O_V(F) \to F^\times/(F^\times)^2$ (not necessarily onto)

3. $O_V(F)$ may have center ± 1 (may not be in the kernel of Dickson or N)

4. If dim $V \leq 4$, you have lots of special exceptions (especially if F is small, finite)

$\mathrm{Spin}_V(F)/(\mathrm{center})$ is simple if dim $V \geq 5$ and V is nondegenerate. The center can have order 1, 2, or 4, and all of them can occur, if you remember that F can have characteristic 2.

I will now make a table of the most common mistakes, and their consequences. You'll notice that none of them are particularly serious, and that's why people continue to make them. It's like some sort of genetic disease that isn't quite severe enough to keep itself from propagating to the next generation.

Most common mistakes	Consequence
1) Use of (,) instead of N to define $C_V(F)$	Things go wrong if $\mathrm{char}(F) = 2$. Of course, no one in their right mind would study orthogonal groups in characteristic 2, anyway
2) Use of xVx^{-1} instead of $xV\alpha(x)^{-1}$ to define the Clifford group	Several minor irritations: reflection is not always in $\Gamma_V(F)$, and $\Gamma_V(F) \to O_V(F)$ is not always onto
3) Defining spinor norm to be N^α, not N (very common, even among people who get everything else right)	Spurious minus sign: spinor norm = −norm
4) Saying the groups $\mathrm{Spin}_{m,n}(\mathbb{R})$ are simply connected	You'll look kind of stupid, because they are not. $\mathrm{Spin}_{n,0}, \mathrm{Spin}_{0,n}, \mathrm{Spin}_{m,1}$, and $\mathrm{Spin}_{1,m}$ are simply connected for $m \geq 3$. $\mathrm{Spin}_{m,n}(\mathbb{R})$ is **not** simply connected for general m, n: usually, $\pi_1(\mathrm{Spin}_{m,n}(\mathbb{R})) \cong \mathbb{Z}/2\mathbb{Z}$ Warning: the algebraic group $\mathrm{Spin}_{m,n}$ **is** simply connected for $m + n \geq 3$. I won't explain this, becuse we're out of time
5) Defining SO using determinant instead of Dickson invariant	You get the wrong group if $\mathrm{char}(F) = 2$. Again, no one cares.

Week 7, 8 March 2004, Free Quantum Field Theories

Today, I will talk about free quantum field theory, which is generally regarded as the trivial case, although it is not really trivial, and it is the only case people can do in a really satisfactory way.

Quick summary of important examples:

We need to choose:

1. Spacetime $\mathbb{R}^{1,d-1}$
2. A vector space Φ of "fields". I don't quite know how to say what these are, so let's just say they're elements of Φ. Φ is a representation of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$.
3. A mass m, for which there are two cases: $m = 0$, and $m > 0$ (negative mass yields causality violations - generalized fields theories allow sums/integrals over different masses, but they don't seem to be used except as counterexamples).

I will now list the main examples. I have experimental evidence that this list is comprehensive, where experiment consists of reading physics texts to see what needs to be accounted for.

1. Hermitian scalar field: dim $\Phi = 1$, i.e. $\Phi \cong \mathbb{R}$, $m = 0$ or $m > 0$. Note: $m = 0$ goes wrong in dimension $d = 2$.
2. Vector field of mass $m > 0$, $\Phi = \mathbb{R}^{1,3}$. For each possible momentum, the space of possible states has dimension 3, which is less than dim Φ (e.g. in electroweak theory, there are three vector bosons). For a vector field of mass $m = 0$ with $d = 4$, there is only a 2 dimensional space of states of a given momentum (e.g. polarization of photons).
3. Spinor fields: Φ is a sum of spin or half spin representations. For example, Dirac spinors in dimension $d = 4$. Here, Φ is a sum of two copies of the spinor representation of $\mathrm{Spin}_{1,3}(\mathbb{R})$, so dim $\Phi = 8$. We might expect an 8-dimensional space of particles of a given momentum. In fact, the space has only 4 dimensions: electrons with spin up or down, and positrons with spin up or down.

As everyone knows, the remaining dimensions are killed off by the Dirac equation.

[Joel: not sure what is meant by dimension - fields given by vector fields on spacetime? Borcherds: Naïve quantization gives you dim $= \Phi$, but some other requirement has to be satisfied. Marco: Are these the constraints given by the classical equations of motion? Borcherds: Right now, I'm just trying to say that you get fewer fields than you would with naïve quantization. (failed to catch some of this discussion)]

More complications: If $d \equiv 2, 3, 4 \bmod 8$, then we can construct "Majorana spinors" [AJ: pronounced Mah-yor-ah-nah]. In this case, the number of states of a given momentum is **half** that of Dirac spinors. If the dimension is even and mass is 0, we can construct Weyl spinors. If $d \equiv 2 \bmod 8$ and $m = 0$, then we get Weyl-Majorana spinors. It appears that free quantum field theory is a complete mess.

Problem Explain all these cases and exceptions in a uniform way.

Physics texts generally cover these in a case-by-case manner, and a new framework has to be introduced for each new field theory (e.g. introduce the Dirac equation for no reason at all).

We want to state:

1. What data do we need to give to define a QFT?
2. What conditions do these data have to satisfy?

I bet you're all impressed that I know "data" is a plural word. Here is the obvious data:

1. Spacetime $\mathbb{R}^{1,d-1}$
2. Choose $m \geq 0$.
3. Choose a representation Φ of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$.

What else do we need to choose to define a free quantum field theory? Well, I haven't really said what a free quantum field theory is (except a few weeks ago). The distinguishing property is that we can split each field as a sum: $\phi = \phi^+ + \phi^-$ of a "creation" field ϕ^+, and an "annihilation" field ϕ^-, that have "easy" commutation relations.

More precisely:

- $[\phi^+(f), \psi^+(g)] = 0$ for all ϕ, ψ, f, and g, where $[X, Y] = XY - (-1)^{\deg X \deg Y} YX$.
- $[\phi^-(f), \psi^-(g)] = 0$
- $[\phi^-(f), \psi^+(g)]$ is a scalar (depending on ϕ, ψ, f, and g of course).

(In general field theories, the commutator of two fields is some horrendous mess that no one can understand.)

Not really needed now: $\phi^+(f)$ is adjoint to $\phi^-(f)$ for f real.

Remark: note that $\phi^+(f)$ and $\phi^-(f)$ generate a Heisenberg Lie (super)algebra, which is a Lie algebra L containing a 1-dimensional subalgebra Z with $[L, L] = Z$ and $[Z, L] = 0$. The space of states is a representation of this Lie algebra, which is convenient, since the representation theory of Heisenberg algebras is basically completely understood.

The key point is that free field theories give a Heisenberg Lie algebra, and other theories give this huge Lie algebra that no one can understand.

We need to spacify a scalar. This determines the entire QFT. We will define:

$$[\phi^-(f), \phi^+(g)] = \int_p \hat{f}(-p)\hat{g}(p)(\phi,\psi)_p d^4p$$

The $\pm p$ in f and g corresponds to some kind of translation invariance. $(,)$ is a bilinear form on Φ depending on p. $\hat{f}$ is the Fourier transform of f: $\hat{f}(p) = \int_x e^{2\pi ixp} f(x)dx$. Whenever you do a Fourier transform, you have to put a 2π in somewhere, and this seems to be the least worst place to put it.

[Joel: p is in the dual of spacetime? Borcherds: p lies in energy momentum space. As long as spacetime has a nondegenerate bilinear form, you can identify this with spacetime.]

What conditions must $(,)$ satisfy?

1. Invariance under $\mathrm{Spin}_{1,d-1}(\mathbb{R})$: $(\sigma\phi, \sigma\psi)_{\sigma p} = (\phi,\psi)_p$ for $\sigma \in \mathrm{Spin}_{1,d-1}(\mathbb{R})$. This is just Lorentz invariance - not very difficult or surprising.

2. Spectral condition: $(\phi,\psi)_p$ has support in $\{p \in \text{positive cone}\}$. Again this is not terribly difficult.

3. Mass condition: $(\phi,\psi)_p$ has support on $\{p^2 = m^2\}$.

These are the obvious conditions it must satisfy. There are some subtle conditions, which I want to talk about.

Pick p, $p^2 = m^2$, $p > 0$. Put $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$ = elements fixing p. Put $(,) = (,)_p$, so $(,)$ is invariant under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$. Conversely, any such inner product on Φ can be extended to a set of inner products $(,)_p$.

Actually, this isn't quite true. It only works if $m > 0$ or $d \neq 2$. I want to explain why we get this funny exception.

Technical point: It is **not** always possible to extend an invariant measure on $\{p^2 = m^2, p > 0\}$ to all of $\mathbb{R}^{1,d-1}$. It is always possible if $m \neq 0$, as that set is **closed**. It is also always possible if $m = 0$, $d > 2$ (easy to check). If $m = 0$, $d = 2$, it is not possible essentially because the measure becomes too infinite near the origin. It is similar to the following problem: Extend the measure $\frac{1}{|x|}dx$ on $\mathbb{R}\setminus 0$ to $\mathbb{R}$. This is not possible, as $\frac{1}{|x|}$ is too large near 0.

Actually, you can get massless scalar fields if you don't mind certain complications such as non-positive-definite Hilbert spaces, etc.

What condition do we need for causality? This is the hard part of free QFT. It seems the correct condition is not written down anywhere in the physics literature.

Correct condition: Suppose $\theta \in \mathrm{Spin}_{1,d-1}(\mathbb{C})$ with $\theta p = -p$. Then $(\phi,\psi) = (-1)^{\deg\phi\deg\psi}(\theta\psi, \theta\phi)$.

Note by the way, this implies the Lorentz invariance condition, since any rotation is generated by reflections.

Question: Is $(,)$ symmetric or antisymmetric?

Answer: No. In general it is **neither**.

It seems to be hardwired into our brains that bilinear forms are either symmetric or antisymmetric, and Dirac fermions are a natural example of this not being the case. I fell into this trap once and got a terrible headache trying to work out fermions, because when I assumed the form was symmetric I got a contradiction, and when I assumed it was antisymmetric I got a contradiction.

Warning If $(,)$ is neither symmetric nor antisymmetric, the left kernel may not be the same as the right kernel.

In general, $(,)$ will be **degenerate**, as the left kernel may be nonzero. Note: $\theta(\text{left kernel}) = \text{right kernel}$.

Next is to prove this is necessary and sufficient for causality, but that will be done next week, along with innumerable examples.

[Marco: Spinor field has a natural inner product - is this what you use? Borcherds: If you've got a natural inner product on Φ, the first thing you do is throw it away.]

Week 8, 15 March 2004, Examples of free quantum field theories

[Notes by Joel Kamnitzer - thanks, Joel!]

Let p be such that $p^2 = m^2$, $p > 0$. Then by invariance under $\mathrm{Spin}_{1,d-1}(\mathbb{R})$, a choice of inner product $(,)_p$ is the same as a choice of $(,)$ invariant under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$, and a p, provided $m > 0$ and [or?] $d \neq 2$.

If $m = 0$ and $d = 2$, then the integral $\int ...dp$ won't converge. For example, $\frac{1}{|x|}dx$ cannot be extended over 0.

Question: What condition do we need for causality to hold?

For all $\theta \in \mathrm{Spin}_{1,d-1}(\mathbb{C})$ with $\theta(p) = -p$, we need $(\phi, \psi) = (-1)^{\deg\phi \deg\psi}(\theta\psi, \theta\phi)$. In fact, this condition implies the above invariance under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$. Note that there are no elements $\theta \in \mathrm{Spin}_{1,d-1}(\mathbb{R})$ taking p to $-p$.

Proposition $(\phi, \psi) = (-1)^{\deg\phi \deg\psi}(\theta\psi, \theta\phi)$ implies causality.

Fix θ with the above property. Then $\theta\mathrm{Spin}_{1,d-1}(\mathbb{R})$ does not necessarily map Φ into Φ. It may take Φ to $i\Phi$, so $\mathrm{Spin}_{1,d-1}(\mathbb{R}) \cup \theta\mathrm{Spin}_{1,d-1}(\mathbb{R})$ acts on $\mathbb{C} \otimes \Phi$.

Proof We want $[\phi(f), \psi(g)] = 0$ if f and g [have supports that] are spacelike separated. Now,

$$= \int [\phi^+(f), \psi^-(g)] - (-1)^{\deg\phi\deg\psi}[\psi^+(g), \phi^-(f)]$$
$$= \int f(x)g(y)e^{2\pi i(x-y)p}((\phi,\psi)_p - (-1)^{\deg\phi\deg\psi}(\psi,\phi)_{-p})dxdydp$$

so we need to show that

$$(\phi,\psi)_p - (-1)^{\deg\phi\deg\psi}(\psi,\phi)_{-p}$$

has Fourier transform vanishing on spacelike vectors.

Now, a distribution h has Fourier transform vanishing on spacelike vectors if h is invariant under $\mathrm{Spin}_{1,d-1}(\mathbb{R})$, and changes sign under rotations changing the direction of time. Since if h is (invariant measure on $\{p|(p,p) = m^2, p > 0\}-$ invariant measure on $\{p|(p,p) = m^2, p < 0\}$, then $\hat{h}$ vanishes on spacelike vectors, then $\hat{h}\cdot$ (any polynomial) vanishes on spacelike vectors.

So in our case, choose $\theta \in \mathrm{Spin}_{1,d-1}(\mathbb{C})$ with $\theta(p) = -p$. Then

$$\begin{aligned}(\phi,\psi)_p &= (\theta\phi,\theta\psi)_{\theta p}\\ &= (\theta\phi,\theta\psi)_{-p}\\ &= (-1)^{\deg\phi\deg\psi}(\psi,\phi)_{-p}\end{aligned}$$

Hence, we get a free quantum field theory if we have:

1. An action of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$ on Φ.
2. $p > 0$, $p^2 = m^2$
3. A complex-valued bilinear form $(,)$ on $\Phi \otimes \mathbb{C}$, invariant under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$, satisfying $(\phi,\psi) = (-1)^{\deg\phi\deg\psi}(\psi,\phi)$ for all $\theta \in \mathrm{Spin}_{1,d-1}(\mathbb{C})$ with $\theta(p) = -p$.
4. If we want the inner product $(,)$ to be positive definite, then $(\phi,\phi) \geq 0$ for $\phi \in \Phi$.

For example, let Φ be the vector representation $\mathbb{R}^{1,d-1}$ of $\mathrm{Spin}_{1,d-1}$. We need to choose $(,)$. The usual inner product is not positive definite. Now, $\mathrm{Spin}^p_{1,d-1}(\mathbb{R}) \cong \mathrm{Spin}_{0,d-1}(\mathbb{R})$ for $m > 0$, so Φ decomposes as $\mathbb{R} \oplus \mathbb{R}^{d-1}$, where the first factor is the line spanned by p.

We can choose $(,)$ by:

1. $(,)$ nonzero on $\mathbb{R}$, 0 on $\mathbb{R}^{d-1}$. This just gives derivatives of the free scalar field.
2. It can be 0 on $\mathbb{R}$, and nonzero on $\mathbb{R}^{d-1}$. This gives a massive vetor field.

Week 9, 22 March 2004, Spring Break

Week 10, 29 March 2004, Examples of free fermionic quantum field theories

Today, I'm going to talk about more standard examples of free quantum field theories. We will start by recalling the data we need:

- Spacetime $= \mathbb{R}^{1,d-1}$
- A vector p in the positive closed cone of (the dual of) spacetime. $(p,p) = m^2$. m is called **mass**.
- A representation Φ of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$. Φ is real, and usually finite-dimensional (in practice, always a sum of finite-dimensional representations).
- A complex bilinear form $(,)$ on $\mathbb{C} \otimes \Phi$. This is **not** necessarily symmetric, antisymmetric, or real on ϕ. It must satisfy the causality condition: $(\phi, \psi) = (-1)^{\deg\phi \deg\psi}(\theta\phi, \theta\psi)$, where θ is any element of $\mathrm{Spin}_{1,d-1}(\mathbb{C})$ mapping p to $-p$ and satisfying $\theta^* = \tau\theta$, where θ^* is the complex conjugate, and τ is the element of order 2 in the center of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$. The elements θ in fact lie in a coset of $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$. In particular, this condition implies invariance of $(,)$ under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$.
- Extra conditions if $m = 0$, $d = 2$ (omitted).

We might want $(,)$ to be positive semidefinite (implying the space of states is positive definite). This condition just means $(\phi, \phi) \geq 0$ for all $\phi \in \Phi$.

In order to write an example of a free quantum field theory, it suffices to write a representation and a form, so this problem is basically reduced to one of linear algebra. We have already mentioned the examples of scalar fields (where $\Phi = \mathbb{R}$) and massive vector fields (where $\Phi = \mathbb{R}^{1,d-1}$). Now I will mention a massless vector field.

General warning: Massless fields are usually **not** some sort of limit of massive fields as m approaches zero. For example, if $\Phi = \mathbb{R}^{1,d-1}$, a vector representation of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$, and p is a vector of norm 0, then $\mathrm{Spin}^p_{1,d-1}$ is quite different from the case where $p^2 > 0$, where it is $\mathrm{Spin}_{0,d-1}(\mathbb{R})$, which is compact, semi-simple, and all the rest of it. If $p^2 = 0$, you get a semidirect product $\mathbb{R}^{d-2} \cdot \mathrm{Spin}_{0,d-2}(\mathbb{R})$, which in particular is not compact.

This is an important difference. If a compact group G acts on a real vector space Φ, we can always find a positive definite symmetric form on Φ fixed by G, by taking the average (over G) of any positive definite form. If G is noncompact, in general, no such positive definite form exists. In our particular case, $G = \mathrm{Spin}^p_{1,d-1}(\mathbb{R})$, $\Phi = \mathbb{R}^{1,d-1}$, the only G-invariant positive semi-definite form is given by projecting $\mathbb{R}^{1,d-1}$ to $\mathbb{R}$ using $(p, -)$, and taking taking the usual inner product on $\mathbb{R}$. This does **not** give the quantum field theory of a massless vector field. It does give the quantum field theory of **derivatives** of a massless scalar field, which is not what we want e.g. when quantizing the electromagnetic field.

The correct way to quantize massless vectors: take $\Phi = \mathbb{R}^{1,d-1}$ with the **usual** inner product, which is not positive definite.

Problem 1: The resulting space of states is **not** positive definite. This problem is resolved using BRST operators in a rather labor intensive process about which I won't say any more today. This is one of the reasons why gauge theories don't fit into the Wightman axioms. [Marco: How did people resolve this problem e.g. in QED, which came before the discovery of BRST? Borcherds: They used ad hoc methods. AJ: They were choosing representatives of BRST cohomology classes, but didn't know it.]

What I wanted to do today is talk about Dirac fermions and other fermions. We start with the following construction: Let M be **any** complex representation of $\mathrm{Spin}_{1,d-1}(\mathbb{R})$. Suppose it has a Hermitian form $\langle , \rangle$, invariant under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$, with $p^2 > 0$. Form the space $M \oplus M^\dagger$, where $M^\dagger$ is a second copy of M - a sort of complex conjugate. $\dagger$ is the antilinear operator defined by switching: $(\phi, \psi)^\dagger = (\psi, \phi)$. I haven't quite worked out the scalar multiplication action of $\mathbb{C}$ on $M^\dagger$. It might involve complex conjugation, but it might not. Let $\Phi =$ fixed points of $\dagger$. Then Φ is a **real** vector space of the same **real** dimension as M. Define a bilinear form $(,)$ in $M \oplus M^\dagger$ by:

$$
\begin{aligned}
(\phi, \psi^\dagger) &= \langle \phi, \psi \rangle \\
(\phi, \psi) &= (\phi^\dagger, \psi^\dagger) = 0 \\
(\psi^\dagger, \phi) &= (-1)^{\deg\phi \deg\psi}(\theta\phi, \theta(\psi^\dagger))
\end{aligned}
$$

Note that $(,)$ is neither symmetric nor antisymmetric, but satisfies the causality condition essentially by definition. Thus, Φ together with $(,)$ satisfies our conditions for constructing a free quantum field theory.

Dirac fermions come from this construction by taking M to be the complex irreducible representation of the Clifford algebra of spacetime. We need to define $\langle , \rangle$ on M: notice that p acts on M, since p is an element in the Clifford algebra. $p^2 = m^2$, so p has eigenvalues $\pm m$, and we get an eigenspace decomposition $M = M_+ \oplus M_-$, which is invariant under $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$.

Dirac's bilinear form $\langle , \rangle$ is defined to be 0 on M_-, and nonzero (i.e. positive definite) on M_+. This is basically the Dirac equation in a disguised form. Note that this form has some interesting properties that may be disconcerting if you are used to symmetric and antisymmetric forms: M_- is the **left** kernel, and M_+ is the **right** kernel.

$$\mathbb{C} \otimes \Phi = M_+ \oplus M_- \oplus M_+^\dagger \oplus M_-^\dagger$$

If $d = 4$, these are all 2 dimensional complex vector spaces. In the pairing $(,)$, we have $M_- \oplus M_-^\dagger$ making up the left kernel, and $M_+ \oplus M_+^\dagger$ making up the right kernel, and there are pairings $M_+ \leftrightarrow M_-^\dagger$ and $M_+^\dagger \leftrightarrow M_-$, although I might have made some mistakes in working this out.

Note that this gives a **charged** free field, meaning there is an action of $U_1(\mathbb{R})$ on everything. M is complex, so the action is through the unit complex numbers.

It might have occurred to you that this construction is rather complicated and arbitrary, and you might wonder if there are other ways to construct fermion fields. The answer is **yes**. You can get:

- Majorana fermions when $d \equiv 2$, 3, or 4 mod 8.
- Weyl fermions when d is even and $m = 0$.
- Majorana-Weyl fermions when $d \equiv 2$ mod 8 and $m = 0$.

Why do we use Dirac fermion fields rather than others? The main reason is just experimental evidence. From a mathematical standpoint, all four constructions are equally valid. However, each formulation gives some experimentally testable predictions, and the fermions we see behave like Dirac fermions. For a long time, it was thought that neutrinos were Weyl fermions, but a few years ago, neutrinos acquired mass, eliminating that possibility. It is still unknown what kind of fermions neutrinos are, but they are suspected to be Dirac.

In the remaining five minutes I will try to describe the other constructions. This will be necessarily rather sketchy.

Majorana: take Φ to be the **real** representation of the Clifford algebra $C_{d-1,1}(\mathbb{R})$. The tricky part here is that 1 and $d-1$ are **reversed**, so $p^2 = -m^2$. The eigenvalues of the action of p are then $\pm im$, giving an eigenspace decomposition $M \oplus M^\dagger \subset \mathbb{C} \otimes \Phi$, where the spaces are complex conjugates of each other. Choose $\langle, \rangle$ on M to be $\mathrm{Spin}^p_{1,d-1}(\mathbb{R})$-invariant (note this group is isomorphic to $\mathrm{Spin}_{d-1,0}(\mathbb{R})$ which is compact). Then define $(,)$ on $\mathbb{C} \otimes \Phi$ by:

$$(\phi, \psi^\dagger) = \langle \phi, \psi \rangle$$
$$(\phi, \psi) = (\phi^\dagger, \psi^\dagger) = 0$$
$$(\phi^\dagger, \psi) = 0$$

It's not hard to check that the causality condition holds.

I haven't covered Weyl and Majorana-Weyl spinors, so you'll have to do them as an exercise if you are interested.

Next week, I'll stop doing free quantum field theories, and instead I'll explain how Bernstein's work on $\mathcal{D}$-modules can be used to analytically continue the integrals that arise when evaluating Feynman diagrams.

Week 11, 5 April 2004, Holonomic Functions

Today's lecture is really quite independent of quantum field theory, so I'll just say what the motivation is now.

Motivation: In quantum field theory, many Feynman diagrams have values which are given by integrals. The problem is that these integrals often don't converge, and this is the cause of all the problems that crop up involving renormalization and regularization. These integrals depend on parameters (such as the dimension of spacetime) and "converge" if the parameters are sufficiently large or small. I put "converge" in commas, because in fact they often don't converge for **any** value of the parameters, but we'll sort of ignore that. Often you can split up the integral into a part that converges if the parameters are large, and a part that converges if the parameters are small.

Idea: Analytically continue in the parameters to define the integral for the values you want.

The problem of analytically continuing integrals which depend on parameters is a perfectly valid problem outside physics, so we can treat it mathematically.

The basic example which everyone knows is that of the Gamma function:

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt$$

The convergence of this integral depends on the parameter s: it converges for $s > 0$, as the exponential term kills everything at ∞, and we just need t^{s-1} to behave well at 0. To analytically continue, we notice that

$$\Gamma(s-1) = \int_0^\infty e^{-t} t^{s-2} dt$$

and integrate by parts

$$\begin{aligned} &= \int_0^\infty e^{-t} \frac{t^{s-1}}{s-1} dt \\ &= \frac{1}{s-1} \Gamma(s) \end{aligned}$$

Idea: Relate $\Gamma(s-1)$ to $\Gamma(s)$. This gives an analytic continuation for all s.

If you want to see why this works, the key point is that you can get a relation by differentiating:

$$\frac{d}{dt}t^{s+1} = (s+1)t^s$$

Here we've got a polynomial in t which happens to be t. Notice this has the following form:

$$D(t, \frac{d}{dt})p(t)^{s+1} = B(s)p(t)^s$$

Here, $p(t)$ is a polynomial in t which is not terribly exciting in this case. $B(s)$ is a polynomial in s, called the Bernstein polynomial. Here, it is $s+1$. D is a polynomial in t, $\frac{d}{dt}$, and s, i.e. a differential operator.

Bernstein showed that for **any** polynomial p in several variables, we can always find D and B as above, with B not identically zero. [Joel: so we're trying to analytically continue the integral of $e^{-t}t^{s-1}$? Borcherds: we're not actually going to analytically continue it. Joel: What are we going to continue? Borcherds: I'm going to set out tools for continuing whatever we want to continue. AJ: Does Bernstein's theorem apply to rational functions? Borcherds: Can't answer that off the top of my head. Ogg: Probably - the continuation ends up having a $\frac{1}{B(s)}$ in it. Borcherds: Not really sure.]

Second basic example: If $\Re(s) > -1$, then $|x|^s$ is locally integrable, so it defines a distrbution in x.

Problem: Can we extend the definition to all complex s?

Answer: It extends to a meromorphic (distribution-valued) function of s. It has poles at $s = -1, -2, -3, \ldots$

Put $|x|^s = (x^2)^{s/2}$. Then $\frac{d}{dx}(x^2)^{s/2} = (s/2)2x(x^2)^{s/2-1} = sx(x^2)^{s/2-1}$. We should refrain from combining the last terms, since the case $x < 0$ is important. We differentiate again: $\frac{d^2}{dx^2}(x^2)^{s/2} = s(s-1)(x^2)^{s/2-1}$. We can use this to **define** $(x^2)^{s/2-1}$ in terms of $(x^2)^{s/2}$, as the derivative of any distribution is a distribution. Then, $(x^2)^{s/2-1}$ becomes a **meromorphic** distribution-valued function, which picks up poles from the $\frac{1}{s(s-1)}$ factor. In fact, we get poles of order 1 at $s = -1, -2, -3, \ldots$. The concept of distribution-valued meromorphic function may be a bit strange for people who haven't seen it before, so we'll calculate a residue.

Example: What is the residue of the pole at $s = -1$? This will be some **distribution**. Put $s = 1$ in the formula

$$(x^2)^{s/2-1} = \frac{1}{s(s-1)}\frac{d^2}{dx^2}(x^2)^{s^2}$$

The pole is coming from the term $\frac{1}{s-1}$. In general, the residue of $\frac{f(s)}{s-1}$ at $s = 1$ is just $f(1)$, provided f is holomorphic at $s = 1$, so the residue of $(x^2)^{s/2-1}$ at $s = 1$ is just:

$$\begin{aligned} \frac{1}{s}\frac{d^2}{dx^2}(x^2)^{s/2} &= \frac{d^2}{dx^2}(x^2)^{1/2} \\ &= \frac{d^2}{dx^2}|x| \\ &= \frac{d}{dx}g(x) \\ &= 2\delta(s) \end{aligned}$$

where g(x) is the function that is 1 for positive x and -1 for negative x. Note that this means in practice that all attempts to define a distribution $|x|^s$ at $s = -1$ have an ambiguity given by a multiple of $\delta(x)$. I'm not sure if that 2 is supposed to be there, but I won't waste time on that.

Example: We can try to take the constant term of $|x|^s$ at $s = -1$. The problem is, we can also define it as the constant term of $|x|^s g(s)$, where g is holomorphic at $s = -1$, $g(-1) = 1$. These differ by a multiple of $\delta(x)$.

I'm not just doing this to be irritating, because it turns out there is **no canonical** way to define $|x|^{-1}$ as a distribution. In particular, there is no way to do it that gives a distribution that is homogeneous of degree -1. Rescaling the domain yields spurious terms.

Remark: (reason this comes up): Values of Feynman diagrams are meromorphic functions of parameters whose values are **distributions**. This turns upon a much larger and more complicated scale: they cannot be defined in a "canonical" fashion. The ambiguities you get are more or less what physicists call "anomalies". This failure of rescaling symmetry in some massless field theories can be traced back to exactly this phenomenon. [Ogg: Differentials have residues, but functions don't. You need a ds on the end of your expression. Borcherds: Yes, I'm being very sloppy here.]

I will discuss the proof of Bernstein's theorem. The theorem itself is not powerful enough to be useful in quantum field theory, but the techniques he develops are. We start by defining $\mathcal{D}$-modules. People tend to view these as sheaves on a manifold or a ringed topos, but most of the fancy machinery is just for bookkeeping, and the basic ideas are visible in the most basic example.

We consider a module over a ring $K[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$. This is a ring of differential operators, but we don't bother to include any functions to operate on. It is defined by the following relations:

- x_i's commute
- ∂_i's commute

- $[x_i, \partial_j] = 0$ for $i \neq j$
- $\partial_i x_i = 1 + x_i \partial_i$ for all i

Note the analogy between ∂_i and $\frac{\partial}{\partial x_i}$.

Example: Suppose a function f satisfies [is annihilated by] partial differential operators D_1, D_2, …, D_m. We can form the $\mathcal{D}$-module $M = K[x_1, \dots, \partial_n]/\langle D_1, \dots, D_m\rangle$, where this quotient is by the left or right module, and I've forgotten which one it is, but I'm not going to worry about it too much. There is a correspondence between functions f satisfying $D_1 f = D_2 f = \cdots = D_m f = 0$ and monomorphisms of $\mathcal{D}$-modules from M to the space of smooth functions of x_1, …, x_n, so $\mathcal{D}$-modules correspond to systems of partial differential equations. I think the whole of $\mathcal{D}$-module theory was invented by algebraists who wanted to study partial differential equations but didn't like analysis, so they just dispensed with the tiresome nonsense about existence of solutions.

A **holonomic** $\mathcal{D}$-module is one that is "small", meaning there are "lots" of differential operators. Roughly speaking, we should have n independent differential operators in the system of differential operators corresponding to the module. If you've done differential equations, you may have run across the term "overdetermined sytem of PDE's". I've never figured out what an overdetermined system is, but I've got a strong suspicion it has something to do with holonomic $\mathcal{D}$-modules.

[Joel: could you clarify the correspondence?]

$$\begin{aligned} \frac{K[x_1, \dots, x_n, \partial_1, \dots, \partial_n]}{\langle D_1, \dots, D_n\rangle} &\to \{\text{smooth functions}\} \\ 1 &\mapsto \text{some solution of "}D_i f = 0\ \forall i\text{"} \end{aligned}$$

This is not anything deep; it's just a complete triviality.

I haven't given you a terribly precise definition of $\mathcal{D}$-module, since I haven't said what "small" means, or what "lots" of differential operators are. [Soroosh: Which $\mathcal{D}$-modules can you get by taking quotients like that? Borcherds: In fact all holonomic $\mathcal{D}$-modules can be obtained this way.]

How do we measure the **size** of a $\mathcal{D}$-module? We do this by using Hilbert polynomials. If you don't know what these are, I'll be teaching a commutative algebra course in about a year.

Quick summary: Suppose $M = \oplus M_i$ is a finitely generated graded module over a graded ring $K[x_1, \dots, x_n]$ (a polynomial ring, graded by $\deg(x_i) = 1$). We can measure size by describing how $\dim(\Gamma_i)$ depends on i, where $\Gamma_i = \oplus_{j \leq i} M_j$. This is a polynomial in i for i **sufficiently large** (note that it is identically zero for i sufficiently negative). This is called the **Hilbert polynomial** of the module M. Its degree is called the dimension of m. The leading coefficient is of the form $\frac{\text{integer}}{(\text{degree})!}$, and the integer is called the **multiplicity** of M. Since it is already 3:00, I'll postpone the proof that this is a polynomial. Next week's lecture will be essentially a primer on $\mathcal{D}$-modules and Hilbert functions and so on.

Week 12, 12 April 2004, Holonomic functions (continued)

We'll start by recalling what we were doing last week. Our main aim was to study the analytic continuation of Feynman integrals, and we seem to be going about this in a roundabout way.

General fact: integrals of **holonomic** functions can often be analytically continued.

At the end of last lecture I was trying to tell you essentially what holonomic meant. It means it is a solution to a holonomic system of partial differential equations, i.e. an element of $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]/(\text{PDE's})$, where this is a **holonomic** module, whih means it is "small". In order to measure the size of a module, we use the Hilbert polynomial.

Given a finitely generated module M over a (commutative) polynomial ring $k[x_1, \ldots x_n]$, $M := \oplus_{i \in \mathbb{Z}} M_i$, define $\Gamma_i := \oplus_{m \leq i} M_m$, so we have $\Gamma_0 \subset \Gamma_1 \subset \ldots$, and look at the dimension of Γ_i. Clearly, the speed at which this increases is some measure of the size of M. For large i, $\dim(\Gamma_i)$ is a **polynomial** $\chi(M, i)$ in i. This is the Hilbert polynomial. The degree of $\chi(M, i)$ is called the **dimension** of M. $\dim(M)! \times$(leading coefficient) is called the **multiplicity** of M. The factor of $\dim(M)!$ makes this an integer.

You might be wondering why I'm spending so much time talking about stuff that is not so clearly related to quantum field theory. The reason is that this is not really a "quantum field theory" seminar. Rather, it is an "interesting ideas in mathematics" seminar. That is why I'm spending so much time on $\mathcal{D}$-modules and Hilbert polynomials.

Proof that $\dim(\Gamma_i)$ is a polynomial for large i.

We shall use induction on n: For $n = 0$, M is a finitely generated module over k, i.e. a finite-dimensional vector space, so for $i >> 0$, $\dim(\Gamma_i) = \dim(M)$.

Now assume this has been proved for #variables $< n$, and look at the exact sequence:

$$0 \to \ker(x_n) \to M_{i-1} \xrightarrow{x^n} M_i \to M_i/(\text{image of } x_n) \to 0$$

Note that the first and last terms are homogeneous components of finitely generated graded modules over $k[x_1, \ldots, x_{n-1}]$. We have

$$\dim(\ker(x_n)) - \dim(M_{i-1}) + \dim(M_i) - \dim(M - i/(\text{image of } x_n)) = 0$$

This has something to do with Euler-Poincaré characteristics. Then

$$\dim(M_i) - \dim(M_{i-1}) = \dim(M-i/(\text{image of } x_n)) - \dim(\ker(x_n))$$

By our inductive hypothesis, the right hand side is a polynomial in i of degree less than n for large i. If you've got a sequence of numbers whose differences comes from a polynomial of degree less than n, then the sequence comes from a polynomial of degree n. Thus, $\dim(M_i)$ is a polynomial of degree n for large i.

End of proof.

As you can see from the proof, this polynomial always takes integer values at integers. We recall a few facts about polynomial maps $\mathbb{Z} \to \mathbb{Z}$. They do **not** necessarily have integer coefficients. This is a standard mistake everyone makes once in his life out of absent-mindedness. For example, the polynomial $\frac{x(x-1)}{2}$ takes integer values, but does not have integer coefficients. There is a basis over $\mathbb{Z}$ for these polynomials given by:

$$\begin{array}{rcl}\binom{x}{0} & = & 1 \\ \binom{x}{1} & = & x/1! \\ \binom{x}{2} & = & x(x-1)/2! \\ \binom{x}{3} & = & x(x-1)(x-2)/3! \\ \vdots & \vdots & \vdots \\ \binom{x}{n} & = & x(x-1)\cdots(x-(n-1))/n! \end{array}$$

Notice that the last polynomial vanishes as $x = 0, \ldots, n-1$ and takes the value 1 at n. Now, any polynomial of degree n taking integer values at integers is a linear combination of these. It is easy to see by induction that for any integers $p(0), \ldots, p(n)$, we can find an integer linear combination of $\binom{x}{0}, \ldots, \binom{x}{n}$ taking these values, so if p is a polynomial of degree n, then $p-$ (linear combination) is 0 on $0, 1, \ldots, n$, and if a polynomial is degree n with $n+1$ zeroes, then it must be 0.

In particular, if you notice that if $a_n x^n + \cdots + a_0$ takes integer values, then $n!a_n$ is an integer. This explains the factor of $\dim(M)!$ in the formula $\text{mult}(M) = \dim(M)! \times (\text{leading coefficient})$. Thus, $\text{mult}(M)$ is an integer.

One thing I haven't bothered to check that I leave as an exercise because I won't do eveything is ... no, it's an exercise. [Me: What is the exercise? Borcherds: Um. Ogg: The exercise is to figure out what the exercise is.]

Now, what we want to do is apply this to modules over noncommutative rings. Basically, we can turn noncommutative rings into commutative rings if we've got a filtration.

Suppose R is a ring $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$, which is **not** commutative. We filter R: $k = R_0 \subset R_1 \subset \ldots$ where R_i is spanned by all monomials in $x_1, \ldots, x_n, \partial_1, \ldots, \partial_n$ of degree at most i, where the variables x_j and ∂_j are all given degree 1. We look at the associated graded ring $R_0 \oplus R_1/R_0 \oplus R_2/R_1 \oplus \cdots$ This turns out to be a **commutative** polynomial ring on $2n$ generators. Of course, this doesn't work for all noncommutative rings: we need the commutators to live in a lower filtered piece. If M is a module over R generated by $m_1, \ldots, m_k$, we construct the Bernstein filtration of M: We let M_0 be the linear span of $m_1, \ldots, m_k$, and let $M_i := R_i M_0$. Then $R_i M_j \subset M_{i+j}$, and $M_0 \oplus M_1/M_0 \oplus M_2/M_1 \oplus \cdots$ is a module over our graded ring. For a module over this nongraded ring, we've constructed a graded module over a graded ring, so we can now take the Hilbert polynomial of this to get the **dimension** and **multiplicity** of M.

Now I get to the exercise I mentioned before: check that $\dim(M)$ and $\mathrm{mult}(M)$ do not depend on our choice of generators. This basically involves showing that changing the generators changes the Hilbert polynomial by a polynomial of degree less than n. (**Warning**: other coefficients of the Hilbert polynomial **do** depend on the choice of generators, so the other terms are more of a pain to use.)

We can find modules over the commutative ring $k[x_1, \ldots, x_n]$ of any dimension from 0 to n. This is really easy to do: for example we can take the module $k[x_1, \ldots, x_n]/(x_1, \ldots, x_i)$ which has dimension $n - i$.

Bernstein showed that if M is a nonzero finitely generated module over the polynomial ring $k[x_1, \ldots, x_n, \partial_i, \ldots, \partial_n]$, then $n \leq \dim(M) \leq 2n$, where the inequality on the right is just the trivial bound.

[Joel: Does a set of generators of M become a set of generators of the graded module? Borcherds: I think so, but I'm not sure. This is probably one of those amazingly plausible facts that turn out to wrong if you check the details.]

M is called **holonomic** if $M = 0$ or if $\dim(M) = M$.

Proof of Bernstein's theorem. Let B_i be the subspace of $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$ spanned by monomials of degree at most i. I think this was previously called R_i but I've lost track. Our aim is to show that the natural map $B_i \to \mathrm{Hom}(M_i, M_{2i})$ given by multiplication is **injective** if $M_0 \neq 0$ (where $M_i := B_i M_0$, and M_0 is spanned by a set of generators of M). Note that $\dim(B_i)$ is a polynomial of degree $2n$ in i, and $\dim(\mathrm{Hom}(M_i, M_{2i}))$ is a polynomial of degree at most $2 \times \dim(M)$ in i.

For $i = 0$, we have $B_0 = k$, and the statement follows, as $M_0 \neq 0$. This is where we pick up the exceptional case where M is 0.

Now, assume the statement is true for $i - 1$. Choose $a \in B_i$. We notice that $aM_i \subset M_{2i}$ implies $a \notin k$, so $a = cx^\alpha \partial^\beta + \ldots$, which is short for $cx_1^{\alpha_1} \ldots x_n^{\alpha_n} \partial_1^{\beta_1} \ldots \partial_n^{\beta_n}$ for $c \in k$ nonzero. Since $a \notin k$, some x_i or some ∂_i must appear in the leading term for a [under some term ordering]. We assume x_1 appears (other cases are similar). Then $[a, \partial_1] \neq 0$, as ∂_1 reduces the number of x_1s by 1, and hence doesn't kill everything off. $[a, \partial_1] \in B_{i-1}$, as $[B_i, B_j] \subset B_{i+j-2}$. If $aM_i = 0$, then

$$[a, \partial_1]M_{i-1} = a\partial_1 M_{i-1} - \partial_1 a M_{i-1} = 0$$

as both the $\partial_1 M_{i-1}$ and the M_{i-1} on the right side next to the a's are in M_i. By our inductive hypothesis, $[a, \partial_1] = 0$, contradicting our previous conclusion that $[a, \partial_1] \neq 0$. If a contains ∂_1, we use $[a, x_1]$ instead.

End of proof.

As you can see, it is one of those short but tricky proofs. The actual proof is just two lines of algebra, but you have to get the right two lines.

I'll just say what holonomic means. It means "M is as small as possible". If $M = k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]/(\text{some PDE's})$, then M holonomic means (roughly) that we have n "independent" PDE's, so a holonomic function in n variables is roughly one that satisfies n "independent" PDE's.

If you've studied differential equations in the past, no doubt you've noticed that there is a fundamental difference between solutions of ordinary differential equations and solutions of partial differential equations. Solutions of ODE's are rather rigid: they can be determined near a point by finitely many derivatives. On the other hand, solutions of PDE's are incredibly floppy. You get infinite dimensional spaces of solutions, and they are not fixed by a finite amount of data. However, holonomic systems of PDE's behave much like ODE's, in that their solutions have rigid behavior. In some sense, the correct generalization of ODE's is not PDE's, but holonomic systems of PDE's. [AJ: Is there are connection between holonomy and holonomic systems of PDE's? Borcherds: Yes, but I haven't quite figured out what it is. You can take a solution to an ODE and analytically continue it around a singularity to get holonomy. Presumably this can be done with a holonomic system of PDE's also.]

The fundamental property of holonomic modules is that they have **finite length** (analogous to vector spaces having finite dimension). This means there is a sequence $0 \subset M^1 \subset M^2 \cdots \subset M$ where M^i/M^{i-1} is **irreducible**. Note: If $R = k[x]$, then R does **not** have finite length, as we have $R \supset xR \supset x^2R \supset \ldots$.

If $0 \to A \to B \to C \to 0$ is an exact sequence of holonomic modules, then:

1. $\text{mult}(A)\text{mult}(C) = \text{mult}(B)$ (this is easy to prove, since the Hilbert polynomial of B is more or less the product of the Hilbert polynomials of A and C)

2. mult is an integer.

3. $\text{mult} = 0 \Rightarrow \dim < n \Rightarrow$ module is 0 (Bernstein).

These three properties basically imply that holonomic modules have finite length. In fact, the length is at most the multiplicity.

We've pretty much finished the commutative algebra material I meant to cover. Next week, we'll show that the fact that holonomic modules have finite length implies that certain integrals can be analytically continued.

Week 13, 19 April 2004, Holonomic functions (continued)

We'll start by trying to remember what we did in the last seminar.

We have a Hilbert polynomial of a finitely generated graded module over a graded (commutative) ring. This gives rise to the dimension (the degree of the polynomial) and the multiplicity (leading coefficient $\times$ (degree)!, where the last bit makes the multiplicity integral).

For modules over $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$, filter everything, look at graded rings, which become commutative if you've set everything up right. Again, you get dimensions and multiplicities. [Marty: Isn't there another natural grading on that ring? Borcherds: Yes, you get a graded ring structure if you set the degree of ∂_i to be -1 instead of 1, but then you just get a noncommutative graded ring, and there isn't a nice way to get something commutative out of it.]

A module over $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$ is called **holonomic** if it is finitely generated, and of dimension $\leq n$. By a theorem of Bernstein, the dimension of a nonzero finitely generated module is **at least** n. This is the result that makes the theory of $\mathcal{D}$-modules possible.

Alternatively: Define the multiplicity of a holonomic module as above. Bernstein's theorem says that if the coefficient of x^n in the Hilbert polynomial is 0, then the module is 0.

Consequence: Holonomic modules have **finite length**, i.e. we have a chain $0 \subset M_1 \subset M_2 \subset \cdots \subset M_n = M$, where each M_i/M_{i-1} is irreducible. As it happens, the length is independent of the choice of chain, and we have finite length if and only if the module is both noetherian and artinian, for rather boring reasons. Finite length means the modules satisfy almost any finiteness condition you could expect to hold. One can draw an analogy, where finite length modules are to modules over $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$ as finite dimensional vector spaces are to vector spaces.

I'm going to give some easy examples of holonomic modules, and once we're done with those, I'll give a construction of Bernstein which yields somewhat less easy examples.

1. $M = k[x_1, \ldots, x_n]$ as a module over $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$, where ∂_i acts as $\frac{d}{dx_i}$. What is the Hilbert polynomial? We choose 1 as a set of generators and take a filtration:

	Γ_0	$\subset$	Γ_1	$\subset$	Γ_2	$\subset$	$\ldots$	$\subset$	Γ_k
basis	1		$1, x_1, \ldots, x_n$		$1, x_i, x_i x_j$		$\ldots$		$\deg \leq k$
dim	1		$n+1$		$\frac{(n+1)(n+2)}{2}$		$\ldots$		$\frac{(n+1)\cdots(n+k)}{k!} = \binom{n+k}{n}$

   We have $\dim(\Gamma_k) = \frac{k^n}{n!} +$ lower terms in k, so the dimension is n and the multiplicity is 1. This module is holonomic and irreducible since we cannot write 1 as a nontrivial sum of positive integers.

2. $M = k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$. The Hilbert polynomial has degree $2n$, so this module is **not** holonomic. Similarly, we can find modules of dimension between n and $2n$ by killling off some of the ∂_i's: $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]/(\partial_{i+1}, \ldots, \partial_n)$ has dimension $n + i$.

3. Suppose $p_1, \ldots, p_n$ are nonzero polynomials, and we just look at the module given by $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]/(p_1(\partial_1), p_2(\partial_2), \ldots, p_n(\partial_n))$. It is easy to check that the dimension is n, and the multiplicity is $\deg(p_1) \times \deg(p_2) \times \cdots \times \deg(p_n)$. This is actually a module over $k[x_1, \ldots, x_n]$ generated by $\partial_1^{i_1}, \partial_2^{i_2}, \ldots, \partial_n^{i_n}$, where $i_k \leq \deg(p_k)$. This is basically a system of n "independent" differential equations, and this is what holonomic systems ought to look like.

I had earlier said something about these systems being called overdetermined systems of PDE's, and I had slightly misspoken. These are actually "**maximally** overdetermined systems of PDE's," and if anyone can tell me what an overdetermined system is, I'd like to hear about it.

We've finished the trivial examples, and we'll introduce a theorem of Bernstein.

Lemma Suppose M a module over $k[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n]$ is filtered by $\Gamma_0 \subset \Gamma_1 \subset \Gamma_2 \subset \ldots$. Suppose $\dim(\Gamma_i) \leq p_n(i)$ for p_n some polynomial of degree at most n. Then M is holonomic (and in **particular**, finitely generated).

(Note: we did **not** assume M is finitely generated, and in fact it is quite important that we did not, since the main use of this lemma is to show that modules are finitely generated.)

Proof (surprisingly easy)

1. Every finitely-generated submodule of M is holonomic. (obvious)

2. Suppose $0 \subset N_1 \subset N_2 \subset \ldots$ is an increasing sequence of finitely generated submodules. Each N_i is holonomic, so they have a well-defined multiplicity, which is at most $n! \times$ degree of the coefficient of x^n in p_n. Thus, the sequence of N_i's must eventually be **constant**. This implies M is finitely generated. (If not, we could find an infinite strictly increasing sequence of finitely generated submodules.)

We can use this to give examples of holonomic modules. Here is an important example:

Suppose p is **any** nonzero polynomial. Then $k[x_1, \ldots, x_n, p^{-1}]$ is holonomic. If you're into algebraic geometry, you'd say you've localized at p. This is part of a very general statement involving derived categories and complexes of sheaves, saying that direct images of holonomic $\mathcal{D}$-modules are holonomic. However, the fancy language doesn't actually have any new content, except if you wanted to know what a derived category is. Our framework contains all the meat of the ... well, we're in Berkeley, and there are vegetarians ... it contains all the **tofu** of the ...

Let m be the degree of p, and grade M as follows: Let Γ_k be the span of all things of the form f/p^k. We'll call this "stuff of degree at most k". I'm going to write these down, because they are a bit tricky to work out without making a mistake. They are just an easy calculation, or wrong:

$$\begin{aligned} \deg(f) &\leq (m+1)k \\ B_i\Gamma_k &\subset \Gamma_{k+i} \\ \dim\,\Gamma_k &\leq \binom{(m+1)k+n}{n} \end{aligned}$$

(B_i is the span of elements of degree at most i in $k[x_1,\dots,x_n,\partial_1,\dots,\partial_n]$.) By Bernstein's Lemma, this **implies** M is finitely generated and holonomic. (Note: It is far from obvious that M is finitely generated over $k[x_1,\dots,x_n,\partial_1,\dots,\partial_n]$. You might expect that adding higher powers of p^{-1} would continue to increase the size of the module.)

We now have more or less what we need to construct the Bernstein polynomial.

Bernstein polynomial: Suppose p is any polynomial in $x_1,\dots,x_n$. Then there is a **nonzero** polynomial $b(s)$ and a polynomial $D(x_i,\partial_i,s)$ satisfying:

$$b(s)p^s = D(x_i,\partial_i,s)p^{s+1}$$

with s a real variable. [Marty: Does D have to depend on s? Borcherds: Not in trivial examples. Actually, I don't know of any examples where s appears in D, but since I've never seen a proof that D does not depend on s, I strongly suspect that s is necessary in general.]

Proof: We work over the field $K=k(s)$ of rational functions in s. We look at the module $M=K[x_1,\dots,x_n,p^{-1}]p^s$, with x_i and ∂_i acting in the obvious way: $\partial_i(p^s)=sp^{s-1}+\partial_i p=p^s(s\frac{\partial_i p}{p})$. (This makes it clear why we need to invert p.) M is holonomic over $K[x_1,\dots,x_n,\partial_1,\dots,\partial_n]$ (the proof of this is basically the same as for the proof that the ring $k[x_a,\dots,x_n,p^{-1}]$ is holonomic, as the modules are basically the same size, but this one is shifted by p^s.) We look at the increasing sequence of submodules of M generated by $p^s,p^{s-1},p^{s-2},\dots$. Since M is holonomic, this sequence is eventually constant, so $p^{s-k-1}=f(x_i,\partial_i)p^{s-k}$, where f is some polynomial with coefficients in $K=k(s)$. So, $p^s=g(x_i,\partial_i)p^{s+1}$. The coefficients of g are rational functions of s, so we can multiply by a common denominator $b(s)$ to get $b(s)p^s=\text{poly}(x_i,\partial_i,s)p^{s+1}$. $b(s)$ is the notorious Bernstein polynomial.

At this point, it seems pretty clear that you need D to depend on s, as you have very little control over the denominators in g. In general, calculating the Bernstein polynomial is very difficult. There is some kind of small cottage industry of people calculating these using Newton polygons and such.

The example that everyone does is: $p = x_1^2 + x_2^2 + \cdots + x_n^2$. Instead of going through the trouble of calculating Hilbert polynomials, we'll just differentiate:

$$\begin{aligned}
\frac{\partial}{\partial x_i} p^s &= 2x_i s p^{s-1} \\
\frac{\partial^2}{\partial x_i^2} p^s &= 2sp^{s-1} + 4x_i^2 s(s-1)p^{s-2} \\
(\sum_i \frac{\partial^2}{\partial x_i^2}) p^s &= 2snp^{s-1} + 4s(s-1)p^{s-1} \\
(2sn + 4s(s-1))p^{s-1} &= Dp^s
\end{aligned}$$

Hence, we have Bernstein polynomial $b(s-1) = 2sn + 4s(s-1)$, and $D = \sum_i \frac{\partial^2}{\partial x_i^2}$. This doesn't even contain x_i terms, but that is because this is a particularly simple example, I think.

Since I said this could be used to analytically continue integrals, I ought to give you one example of analytic continuation: The integral $\int_0^1 \cdots \int_0^1 p(x)^s dx_1 \cdots dx_n$ converges for $\Re(s) \geq 0$. By Bernstein, this is equal to

$$\frac{1}{b(s)} \int \cdots \int Dp(x)^{s+1} dx_1 \cdots dx_n$$

The differential operator D gives integrals over the $n-1$-dimensional boundary pieces. Inducting on n gives an analytic continuation.

I'll have to leave this as an exercise, since it is already 3:00.

I have no idea what I'm going to do next week.

Week 14, 26 April 2004, Analytic continuation

Today, I will talk about how to analytically continue almost anything. In particular, I will show you how to analytically continue the integrals that come out of a Feynman diagram.

We return to the example from last time, for which my explanation was too sketchy for anyone to understand. We have some polynomial p in n variables x_i, we assume $p(x)$ to be non-negative for $0 \leq x_i \leq 1$, and we wish to continue

$$\int_0^1 \cdots \int_0^1 p(x)^s dx_1 \cdots dx_n$$

which converges for $\Re(s) \geq 0$, to a meromorphic function for **all** complex s. We use the Bernstein polynomial: $b(s)p(x)^s = Dp(x)^{s+1}$, where D is some polynomial in x_i, $\frac{d}{dx_i}$, and s, so

$$\int_0^1 \cdots \int_0^1 p(x)^s dx_1 \cdots dx_n = \frac{1}{b(s)} \int \cdots \int Dp(x)^{s+1} dx_1 \cdots dx_n$$

Note that the $b(s)$ in the denominator gives poles at the zeroes of $b(s)$. Now, suppose D has a term of the form $\frac{\partial}{\partial x_1} \times$(some rubbish). Then

$$\int_0^1 \cdots \int_0^1 \frac{\partial}{\partial x_1} (*) p(x)^{s+1} dx_1 \cdots dx_n = \left[\int_0^1 \cdots \int_0^1 (*) p(x)^{s+1} dx_2 \cdots dx_n \right]_{x_1=0}^{x_1=1}$$

where $(*)$ is the rubbish. We hope the right side is meromorphic in s by induction on n. However, we end up having to integrate things of the form

$$\int \cdots \int p(x)^s (\text{some polynomial in } x_i) dx_1 \cdots dx_n$$

We started out trying to prove a result that was too weak, and the inductive step failed. We should really prove that $\int_0^1 \cdots \int_0^1 p(x)^s \text{poly}(x) dx_1 \cdots dx_n$ can be analytically continued so that our inductive hypothesis is strong enough. Actually, we need to extend this a bit: What about $\int_0^\infty$ or $\int_{-\infty}^\infty$? Well, $\int_0^\infty dx = \int_0^1 dx + \int_1^\infty dx$, and for the second integral, we can change x to $1/x$, so we get $\int_0^1 q(x) x^{(*)} dx$, i.e. we integrate a polynomial times some (negative) power of x. This leads to a more general problem: Can we analytically continue the following integral?

$$\int \cdots \int p_1(x)^{s_1} p_2(x)^{s_2} \cdots p_k(x)^{s_k} p(x) dx_1 \cdots dx_n$$

The answer is **yes**. The proof is basically the same as for the case $k = 1$. The main part is the fact that $k(s_1, \ldots, s_k)[x_1, \ldots, x_n, p_1^{-1}, \ldots, p_k^{-1}] p_1^{s_1} \cdots p_k^{s_k}$ is a holonomic module.

By the same argument as before, we can find Bernstein polynomials $b_1, \ldots, b_k$ in several variables such that

$$b_1(s_1, \ldots, s_k) p_1^{s_1} \cdots p_k^{s_k} = D p_1^{s_1+1} p_2^{s_2} \cdots p_k^{s_k}$$

and so on. This is just like Bernstein polynomials for one variable, except there are more variables. More generally still: $p(x)$ can be **any** holonomic function. We can also let $p_i(x)$ be non-polynomial, but then we need to be more careful dealing with singularities.

Here is a silly proof that Γ can be analytically continued: $\Gamma(s)$ is defined for $\Re(s) > 0$ by $\int_0^\infty t^{s-1}e^{-t}dt$, and the function e^{-t} is holonomic.

There has been a request, that people want to see a Feynman diagram some time this semester, so here is one: [draws two vertices, with an edge between them]

We want to consider Feynman diagrams, like this mess. [slightly more complicated graph] This gives rise to some multidimensional integral in the following way: Designate some vertices of the graph as "internal" [circles some vertices] - these are the ones you integrate over. You integrate the following function of $(x_1, \ldots, x_v)$, where v is the number of internal vertices, and the variables x_i take values in $\mathbb{R}^d$ or $\mathbb{R}^{1,d-1}$ (Since integrals over Lorentz space have even more problems than those over Euclidean space, we'll work with Euclidean space for now):

$$\underbrace{\int \cdots \int}_{\substack{\text{internal} \\ \text{vertices i}}} \prod_{\substack{\text{edges} \\ i \text{ to } j}} \Delta(x_i - x_j) \prod dx_i$$

Where Δ is some function of $x \in \mathbb{R}^d$ called a propagator. It measures the "amplitude" for a particle to go from the point x_i to the point x_j. This integral usually does **not** converge.

If you want to know what an amplitude is, I would suggest you not bother, because I've been trying to understand them for years, and I still have no idea what they are. This seems to be one of the ways that quantum mechanics is beyond human comprehension. [AJ: Feynman's book on QED has a nice explanation of amplitudes. Borcherds: Feynman also said that no one understands quantum mechanics, and that you just get used to it.]

What is Δ?

This is often given by some Bessel function of $\sqrt{(x_i - x_j)^2}$ multiplied by some elementary function. If you've seen these integrals before and none of them had Bessel functions, this is because computations with Feynman diagrams are often done in momentum space rather than position space, and Fourier transforms of Bessel functions are often somewhat simpler. [Marty: Are these K-Bessel functions? Borcherds: Mostly. Marty: Why do the integrals blow up? Is the problem at infinity? Borcherds: There are problems everywhere. They come in two basic types: stuff blowing up at a point, and stuff blowing up at infinity. Ogg: Are the analytic continuations always single-valued? Borcherds: They will be single-valued in the parameters we will introduce.]

It turns out integrals of Bessel functions often can be simplified to something reasonable:

$$\int_0^\infty \alpha^{v-1}\exp(-x^2\alpha - m^2/\alpha)d\alpha = 2\left(\frac{m^2}{x^2}\right)^{v/2} K_v(2\sqrt{m^2x^2})$$

The propagator is usually more or less the left side of the equation, and v here is the dimension of spacetime, although it is often varied for the purposes of analytic continuation. So the Feynman integral is:

$$\underbrace{\int\cdots\int}_{\text{x's}}\underbrace{\int\cdots\int}_{\alpha\text{'s}}\prod \alpha^{(*)}\exp(-x^2\alpha - m^2/\alpha)d\alpha dx$$

where $(*)$ is some number. By introducing these α's, we seem to have turned a complicated integral into an even worse one (note that the argument of the exponential is just schematic, and some factors have been left out).

But, we can now integrate over the x's explicitly.

Reason: The integral is of the form exp(quadratic in x's).

What happens when we actually do this integral? Recall how to do $\int e^{-\text{quadratic in } x} dx$

Lemma: Let $M = \left(\begin{pmatrix} A & B \\ B^t & C \end{pmatrix}\right)$ where A is a $|I| \times |I|$ matrix, I is some set parametrizing the first few rows. Then:

$$\int \exp(-\pi x^t M x)\prod_{i\in I} dx_i = \frac{1}{\sqrt{\det(A)}} \times \sum_{i,j\notin I} \det(M_{I\cup i, I\cup j})$$

where $M_{I\cup i,I\cup j}$ is the matrix given by adding the ith row and jth column to A.

Proof: $\int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1$, plus lots of linear algebra. This basically involves diagonalizing a matrix, and I'll skip that, because the details aren't too terribly interesting.

Technical Lemma: (for evaluating various determinants) Suppose we have a graph Γ (you can think of this as having a Feynman diagram). Suppose we are given a number t_e for each edge e of Γ. We are going to form the following matrix M from Γ.

$$m_{ii} = \sum_{\substack{\text{one end}\\ \text{of } e \text{ is } i}} t_e \qquad m_{ij} = -\sum_{\substack{e \text{ has}\\ \text{ends } i,j}} t_e$$

Example: [draws triangle Γ with vertices 1,2,3 and edges labeled t_{12}, t_{23}, and t_{31}]

$$M = \begin{pmatrix} t_{12}+t_{31} & -t_{12} & -t_{31} \\ -t_{12} & t_{12}+t_{23} & -t_{23} \\ -t_{31} & -t_{23} & t_{31}+t_{23} \end{pmatrix}$$

This is the matrix tha turns up when evaluating the Feynman diagram arising from the graph Γ.

Suppose I and J are subsets of the vertices of Γ, with $|I| = |J| = n$ for some n. By M_{IJ} we mean the $n \times n$ matrix of elements m_{ij}, $i \in I$, $j \in J$.

The problem we want to solve is: Find a nice formula for $\det(M_{IJ})$. The answer, somewhat surprisingly, is that it actually has a reasonably nice formula:

$$\det(M_{IJ}) = \sum_T \pm |T|$$

T is a set of n edges of Γ sith the following property:

1. Every point of I is connected by T to a point **not** in I.
2. Every point of J is connected by T to a point **not** in J.

For example, if I is the whole graph, the determinant is 0. For the above formula, we define:

$$|T| = \prod_{e \in T} t_e$$

Now, I've got to tell you what the sign is. T is a union of trees. Each tree either joins a point of $I - J$ to a **unique** point of $J - I$, or it has no points in $I - J$ or $J - I$. We get an isomorphism from $I - J$ to $J - I$ by connectivity in T. Choose a fixed isomorphism from $I - J$ to $J - I$. The sign $\pm$ is the sign of the permutation of $I - J$ given by composition of one isomorphism with the inverse of the other. Since changing your isomorphism by an odd permutation changes the sign of both sides, it doesn't really matter which isomorphism you choose [not sure I heard this correctly].

Proof Induction on $|I \cap J|$. We actually only need the cases where $I = J$ or $|I - J| = 1$, but the induction involves starting with I and J disjoint, and seeing what happens when vertices are moved into the intersection.

Week 15, 3 May 2004, Analytic continuation (continued)

Last lecture we were in the middle of trying to compute the determinant of a rather hairy looking matrix, so we'll just recall what the matrix is.

Suppose we're given a graph Γ with n vertices, and define a $n \times n$ matrix M_{ij} as follows: choose a variable t_e for each edge e, and either take a sum of singular 2×2 matrices:

$$M = \sum_{\text{edges}} \begin{matrix} \\ i \\ j \end{matrix} \begin{matrix} i & j \\ \begin{pmatrix} t_e & -t_e \\ -t_e & t_e \end{pmatrix} \end{matrix}$$

where the sum is over all vertices i and j, and all edges e from i to j, or we can take one big matrix:

$$M_{ij} = \sum_{\text{edges } e: i \to j} -t_e \text{ if } i \neq j$$

$$M_{ii} = \sum_{\text{one end of } e \text{ is } i} t_e$$

Given two subsets I and J of the same size m, we can form the minor M_{IJ} with rows in I and columns in J.

Problem: What is $\det(M_{IJ})$?

A: It is given by $\sum_T \pm|T|$, where

1. T has m edges.
2. Everything in I is connected by T to something not in I.
3. Everything in J is connected by Y to something not in J.

Note that any graph T with properties 2 and 3 has at least m edges, so these conditions are quite tight. In particular, they force T to be a forest (i.e. a union of trees, a graph with no loops). They also force each point of $I - J$ to be connected to a unique point of $J - I$, so we get an isomorphism of $I - J$ with $J - I$.

$|T| = \prod_{e \in T} t_e$ with sign $\pm$ being the sign of the isomorphism from $I - J$ to $J - I$ in some fixed total order of $I - J$ and $J - I$, although the actual order chosen doesn't change the sign of the determinant.

It's actually rather surprising that the determinant turns out so nicely. One's first reaction is that this is going to be a horrible mess.

Proof by induction on $I \cap J$:

Note: for applications, we only need the cases $I \cap J = \emptyset$ or $I \cap J =$ point.

Suppose $I \cap J = \emptyset$. I and J each have m points, and for each $i \in I$, there is a single edge of T connecting i to something in J. The formula for $\det(M_{IJ})$ turns out to be (essentially) the usual formula for the determinant of $n \times n$ matrices, although Γ may be missing some edges that would contribute some terms.

Now, suppose there is some $i \in I \cap J$. Pick an edge e joining i to some $j \neq i$, and form a new matrix N from M by adding the ith row to the jth row, then the ith column to the jth column. Then $\det(N_{IJ}) = \det(M_{IJ})$ as $i \in I \cap J$, and t_e only occurs in position (i,i) of N:

$$M = \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{pmatrix} t_e + * & -t_e + * \\ -t_e + * & t_e + * \end{pmatrix} \end{array} \qquad N = \begin{array}{c c} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{pmatrix} t_e + * & * \\ * & * \end{pmatrix} \end{array}$$

Here, $* =$ some junk. So, $\det(N_{IJ}) = \det(M_{IJ})$ is **linear** in t_e.

What is the coefficient of t_e? It is given by $\pm$ the minor of N with column and row i removed. What we are left with is the matrix corresponding to the graph Γ/e, meaning we **identify** the vertices i and j and discard the edge e. The fact that we've added the ith row and column to the jth row and column means all the adjacency information in the ith vertex has been transferred to the jth vertex.

By induction, the coefficient of t_e is the sum over forests of Γ/e of size $m-1$ connecting all points of $I-i$ to something not in $I-i$ and so on, which correspond to forests of Γ containing e and satisfying the usual conditions.

This proves the result, as it is true for all edges e.

I should note a couple of slightly remarkable things: For the case $I = J$, we have

$$M = \begin{array}{c c} & \begin{array}{cc} I & J \end{array} \\ \begin{array}{c} I \\ J \end{array} & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \end{array}$$

We notice that if $I = J$, then $\det(A) = \sum |T|$, where the sum is over all forests with $|I|$ edges such that everything in I is connected to something not in I. Things in I should be thought of as "internal vertices", and vertices not in I should be thought of as "external vertices". [Noah: Should there be a $\pm$ in the sum? Borcherds: No. This is the second interesting point.]

You will notice here that there is **no sign**. This is because the sign of any permutation of the empty set is +1. This is because the group of permutations of the empty set is the trivial group. So $\det(A)$ is a polynomial in t_e with **non-negative** coefficients. This is far from obvious if you just look at the terms of the determinant itself. So, $\det(A) \geq 0$ (if we set $t_e \geq 0$).

The same is true if $|I-J| = |J-I| = 1$, since all permutations of a one-element set have sign +1. $\det(M_{IJ})$ has all terms of the same sign.

When the difference is at least two, you no longer have this property. It is rather convenient that the only cases that come up when evaluating Feynman integrals are those for which all coefficients are positive.

Now, recall that all Feynman integrals could be reduced to:

$$\int_\alpha \int_x \alpha^* \exp(-x^2\alpha - \alpha/m^2) d\alpha dx$$

where the α's and x's vary over d-dimensional spacetime. We do the integral over the x's using the fact that

$$\int \exp(-\pi x^t M x) \prod_{i \in I} dx_i = \frac{1}{\sqrt{\det(A)}} \exp \frac{\sum (x_i, x_j) \det(M_{I \cup i, J \cup j})}{\det(A)}$$

I horribly garbled this last time, as I forgot the exponential on the right and the $\det(A)$ in the denominator of the exponential, but it doesn't really matter, and the important term is the $\sqrt{\det(A)}$ in the denominator, where A is given by $M = \left(\begin{pmatrix} A & B \\ B^t & D \end{pmatrix} \right)$.

We use this theorem to integrate over the x's, so the Feynman integral becomes of the form:

$$\int_\alpha \alpha^* \frac{1}{\det(A)^{d/2}} \exp \left(\sum (x_i, x_j) \frac{\det(M_{I \cup i, J \cup j})}{\det(A)} \right) \prod d\alpha_i$$

The i's and j's in this integral correspond to external vertices, i.e. those not in I. In the physical description, the amplitude should depend on the positions of the particles before and after the interaction.

This integral is now of a form that can be analytically continued using Bernstein's ideas:

$$\iiint_{\alpha=0}^{\infty} (\text{non-negative polys in } \alpha)^* \times (\text{reasonable function}) d\alpha$$

The polynomials in α are either powers of α or $\det(A)$, and we know that $\det(A)$ is positive for $\alpha_i \geq 0$, as it is given by a sum over trees, so there is no problem taking positive powers of it.

Note: This initially converges for d negative with large absolute value, and we analytically continue to the dimension of spacetime that we want. This is closely related to what physicists call "dimensional regularization", which involves treating the dimension of spacetime as some arbitrary real number. Unfortunately, the integral usually has poles at $d = 1, 2, 3, 4, \ldots$ and is otherwise regular. It can be rather frustrating to have done all this work and get a pole for your answer. [Ogg: Why don't you take residues? Borcherds: That is the question that mathematicians always like to ask - why not take the constant term? The short reply is that you get the wrong answer.]

In general, renormalization is used to cancel out the terms you get from poles. This is a rather bizarre process by which terms are inserted that depend on the coupling constant, so that the resulting poles exactly cancel out the poles you get from the analytic continuation. The main justification for why it works is that the answers you get agree with experiment to some absurd number of significant figures.

Well, that concludes the seminar for today, and for the term. There is another day of classes next week, but I'm rather exhausted.