Advanced Quantum Field Theory

João F. Melo

Insituut voor Theoretische Fysica,
KU Leuven,
Celestijnenlaan 200D, B-3001 Leuven, Belgium

E-mail: joao.melo@kuleuven.be

Contents

0 Motivation, Bureaucracies and Conventions

When you first learn Quantum Field Theory it is customary to canonically quantise a field, following a very similar playbook to ordinary quantum mechanics. The trick is to reduce fields to many (in fact an infinite number of) harmonic oscillators and then use perturbation theory to add interactions. It is challenging and there are many subtleties and details, but all is well and good. This is the version of QFT that was known in the 1930s.

However, just going a bit further in perturbation theory you find that the result is infinite! And you would not be alone. The physicists that developed this early form of QFT also were unable to get rid of these unwelcome guests. It took decades until an ad-hoc procedure to "cancel" the infinities was found: renormalisation. But, despite working unreasonably well, it left a sour taste. Feynman himself (one of the pioneers in these methods) called it a "dippy process".

Now we know better. Since the 1970s, with the work of Wilson, Weinberg, Kadanoff and many others, we understand where the infinities come from, and why the "dippy process" of renormalisation actually works.

This is the primary goal of this course, develop the tools necessary to tackle the issue of infinities, unravel the mysteries of renormalisation and then reap its benefits.

Unfortunately, our time is very limited and quantum field theory is inconceivably vast. We could have four times as many lectures and still have things left to say. The emphasis will therefore not be on exhaustiveness but depth. We will focus on the simpler examples, examining them deeply, so that the concepts are clear, and the foundations solid when you wish to understand the more complicated (and more interesting) cases.

Our course will be divided in three parts.

Firstly, we need to build the tools that we will need to use to tackle the questions we have set out to answer. In particular, we will develop the *path integral*, which is an alternative formulation of quantum field theories based on the Lagragian instead of the Hamiltonian.

In the second part we put this formalism to good use and construct the *renormalisation group*. We shall go deeply on the theory of a single scalar field and really understand where the infinities come from, what to do about them, and what does that tells us about how theories change as we probe them at different energy scales.

Finally, we use the results and understanding from the second part to figure out how *symmetries* play a role in quantum field theories. In particular we will see how they might or might not survive quantisation.

If there is time we might cover some extra topics regarding gauge theories and/or non-perturbative phenomena.

0.1 Bibliography

There will be lecture notes (these ones!) with all the material covered in the lectures. If you wish to follow a different perspective of the same topics here is a list of suggested references (which were in large part the origin behind these notes).

Books

- Michael Peskin and Dan Schroeder, "An Introduction to Quantum Field Theory" (Part 1 covers the basics, Part 2 is the most relevant for this course, Part 3 is mostly beyond the scope of the course except for the chapter on Anomalies)
- Anthony Zee, "Quantum Field Theory in a Nutshell" (Mostly for conceptual aspects)
- Matthew Schwarz, "Quantum Field Theory and the Standard Model" (Different perspective, but very good for individual topics)
- Mark Srednicki, "Quantum Field Theory" (Different perspective, but very good for individual topics)
- Steven Weinberg, "The Quantum Theory of Fields" (Only for the brave of heart. Unique perspective on the topic with insights not found elsewhere. Volumes 1 and 2 are a superset of what I will be covering, but they contain *a lot* more and are in general more advanced)

Lecture Notes

- Hugh Osborn, "Advanced Quantum Field Theory" (Closest to the course)
- David Tong, "Statistical Field Theory" (For a different perspective on RG flows)
- David Tong, "Quantum Field Theory" (Review of the basics which will be assumed)
- David Tong, "Gauge Theories" (By and large advanced topics beyond the scope of the course, but the chapter on Anomalies is very good and relevant)
- Timo Weigand, "Quantum Field Theory I+II" (Covers both the basics and most of what this course will cover)

0.2 Conventions

We use natural units where $c = \hbar = k_{\mathrm{B}} = 1$. This means we only have one dimension left which can be either mass, energy, length, time, or temperature. We'll prefer the use of energy and more specifically electron volt (eV). They are related by

$$[\text{mass}] = [\text{energy}] = [\text{length}]^{-1} = [\text{time}]^{-1} = [\text{temperature}] \tag{0.1}$$

We'll denote Euclidean signature vectors with boldface, e.g. $\boldsymbol{x} = (x^1, x^2, x^3, \dots, x^d)$, and the spatial indices by Latin letters starting from the middle of the alphabet $i, j, k, \dots = 1, 2, 3, \dots, d$. Lorentzian signature vectors will have no special notation, e.g. $x = (x^0, \boldsymbol{x})$, where we use the "mostly plus" convention: $\eta = \mathrm{diag}(-1, 1, 1, 1, \dots)$, the spacetime indices will be Greek letters starting from the middle of the alphabet $\mu, \nu, \dots = 0, 1, 2, 3, \dots, d$. Any other kind of index such as internal indices will be denoted by Latin or Greek letters starting from the beginning of the alphabet and corresponding vectors will be underlined.

Einstein's summation convention for repeated indices will be used unless explicitly stated. The number of spacetime dimensions will be denoted by $D = d + 1$.

For Fourier transforms we use the following conventions:

$$\tilde{f}(k) = \int \mathrm{d}x\, f(x) \mathrm{e}^{-\mathrm{i}kx} \tag{0.2}$$

$$f(x) = \int \frac{\mathrm{d}k}{2\pi} \tilde{f}(k) \mathrm{e}^{\mathrm{i}kx} \tag{0.3}$$

Putting it simply, all factors of 2π appear in k (or p) integrals. In particular this gives the following normalisation for the momentum states

$$\langle \boldsymbol{p} | \boldsymbol{p}' \rangle = (2\pi)^d \delta^{(d)}(\boldsymbol{p} - \boldsymbol{p}') \tag{0.4}$$

$$\int \frac{\mathrm{d}^d p}{(2\pi)^d} |\boldsymbol{p}\rangle\langle\boldsymbol{p}| = \mathbb{1} \tag{0.5}$$

$$\langle \boldsymbol{x} | \boldsymbol{p} \rangle = \mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{x}} \tag{0.6}$$

The Dirac-γ matrices are defined as follows

$$\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu} \tag{0.7}$$

The minus sign on the RHS comes from the choice of metric sign convention. We could have chosen otherwise by essentially multiplying all γ^μ by a factor of i (which is the choice Weinberg makes). Our choice means that our expressions look more similar to books written with the opposite sign convention, the only differences will be in minus signs with a metric rather than factors of i whenever a γ matrix is involved.

The Dirac adjoint is then defined as

$$\bar{\psi} = \psi^\dagger \gamma^0 \tag{0.8}$$

And we shall use the Feynman slash notation as well

$$\not{p} = \gamma^\mu p_\mu \tag{0.9}$$

Finally, throughout the text we'll merrily swap between Heisenberg and Schrödinger pictures. To simplify our notation, instead of keeping track of H and S subscripts we'll simply make the time dependence of operators/states explicit, i.e.

$$O(t) = \mathrm{e}^{\mathrm{i}Ht} O(0) \mathrm{e}^{-\mathrm{i}Ht}, \qquad O \equiv O(0) \tag{0.10}$$

$$|\psi, t\rangle = \mathrm{e}^{-\mathrm{i}Ht} |\psi, 0\rangle, \qquad |\psi\rangle \equiv |\psi, 0\rangle \tag{0.11}$$

Part I

The Path Integral

1 Deriving the path integral

We start by deriving an alternative formulation of QFT. It is in general very useful to have multiple perspectives on the same topic, but for the purposes of understanding renormalisation it is crucial to know of the path integral.

As we any quantum theory we first need to define what is the theory we are considering, and what is the observable we wish to calculate.

Regarding the theory, we will focus on a single scalar field for simplicity. The Hamiltonian is therefore given by

$$H = \int \mathrm{d}^d x \left(\frac{1}{2}\Pi^2 + \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \frac{1}{2}m^2\phi^2 + V_{\text{int}}(\phi, \Pi) \right) \tag{1.1}$$

where Π is the momentum conjugate to ϕ and V_{int} is some interaction potential.

Regarding observables we will focus our attention on time-ordered expectation values of a single scalar field:

$$G^{(n)}(x_1, \ldots, x_n) = \langle \Omega | \, \mathcal{T}(\phi(x_1) \ldots \phi(x_n)) \, | \Omega \rangle \tag{1.2}$$

where $|\Omega\rangle$ is the vacuum state (that is, the lowest energy eigenstate of the Hamiltonian) and $\mathcal{T}$ denotes time-ordering, putting earlier times on the right.

We have chose these observables for two reasons. Firstly, they are quite simple to calculate and manipulate and we are after the simplest possible examples that still illustrate the physics we wish to describe. Secondly, they have particular experimental relevance. Due to the LSZ reduction formula, we can relate time-ordered correlators to scattering amplitudes, which in turn can be converted to cross-sections and decay rates. These are the measurements of choice for scattering experiments in a particle collider such as the LHC.

The details of converting time-ordered correlators to experimental results lie beyond the scope of this course. You can check out, for example, Timo Weigand's notes for those details. For our purposes it merely suffices to say that we will compute time-ordered correlators and that they are experimentally relevant.

1.1 From time-ordered correlators to the transition amplitude

To simplify matters we will Fourier transform with respect to the spatial variables which effectively reduces our system to ordinary quantum mechanics. To convert back to D-dim we just need to make $m \to E_{\boldsymbol{p}} = \sqrt{\boldsymbol{p}^2 + m^2}$ and integrate[1] over $\boldsymbol{p}$. We therefore write:

$$H = \frac{1}{2}\Pi^2 + \frac{1}{2}m^2\phi^2 + V_{\text{int}}(\phi) \tag{1.3}$$

We shall initially assume the potential does not include derivative interactions for simplicity. We will go back to this point later.

[1] We are intentionally being a bit cavalier. This reduction is only straightforward if it was an ordinary sum. This continuous integration is precisely the origin of all the difficulties with field theories.

Without loss of generality let us assume $t_1 < t_2 < \cdots < t_n$. Writing the time evolution operators explicitly and using the fact the Hamiltonian is time-independent we get

$$G^{(n)}(t_1, \ldots, t_n) = \langle \Omega | \, \mathrm{e}^{-\mathrm{i}H(t_f - t_n)} \phi \mathrm{e}^{-\mathrm{i}H(t_n - t_{n-1})} \cdots \mathrm{e}^{-\mathrm{i}H(t_2 - t_1)} \phi \mathrm{e}^{-\mathrm{i}H(t_1 - t_0)} \, | \Omega \rangle \tag{1.4}$$

We have introduced some spurious $t_0 < t_1$ and $t_f > t_n$. Given that the vacuum state is an eigenstate of the Hamiltonian they will only contribute as an overall constant $\mathrm{e}^{-\mathrm{i}E_\Omega(t_f - t_0)}$, which, if we normalise $G^{(n)}$ appropriately, will drop out of the final calculation[2].

Inserting $n+2$ copies of the identity $\int \mathrm{d}\phi \, |\phi\rangle \langle\phi|$ we find

$$\begin{aligned} & G^{(n)}(t_1, \ldots, t_n) = \\ & = \int \left(\prod_{j=0}^{n+1} \mathrm{d}\phi_j \right) \langle \Omega | \phi_{n+1} \rangle \, \langle \phi_{n+1} | \, \mathrm{e}^{-\mathrm{i}H(t_f - t_n)} \, | \phi_n \rangle \, \phi_n \, \langle \phi_n | \, \mathrm{e}^{-\mathrm{i}H(t_n - t_{n-1})} \, | \phi_{n-1} \rangle \cdots \\ & \cdots \phi_1 \, \langle \phi_1 | \, \mathrm{e}^{-\mathrm{i}H(t_1 - t_0)} \, | \phi_0 \rangle \, \langle \phi_0 | \Omega \rangle \end{aligned} \tag{1.5}$$

where we have converted every instance of the field operator ϕ into the appropriate eigenvalue ϕ_j.

The basic unit in this expression is $\langle \phi_j | \, \mathrm{e}^{-\mathrm{i}H(t_j - t_{j-1})} \, | \phi_{j-1} \rangle$. This is known as the *transition amplitude*: it is the amplitude for finding that the field has a value[3] ϕ_{j-1} at $t = t_{j-1}$ and then later finding that it has the value ϕ_j at $t = t_j$.

1.2 Expressing the transition amplitude as a sum over paths

The first trick is to divide the interval $[t_{j-1}, t_j]$ in N slices of size $\Delta t = \frac{t_j - t_{j-1}}{N}$ and at the end take the limit $N \to \infty$. We then have

$$\begin{aligned} & \langle \phi_j | \, \mathrm{e}^{-\mathrm{i}H(t_j - t_{j-1})} \, | \phi_{j-1} \rangle = \langle \phi_j | \left(\mathrm{e}^{-\mathrm{i}H\Delta t} \right)^N | \phi_{j-1} \rangle = \\ & = \int \left(\prod_{k=1}^{N-1} \mathrm{d}\phi_{(k)} \right) \langle \phi_j | \, \mathrm{e}^{-\mathrm{i}H\Delta t} \, | \phi_{(N-1)} \rangle \, \langle \phi_{(N-1)} | \, \mathrm{e}^{-\mathrm{i}H\Delta t} \, | \phi_{(N-2)} \rangle \cdots \langle \phi_{(1)} | \, \mathrm{e}^{-\mathrm{i}H\Delta t} \, | \phi_{j-1} \rangle \end{aligned} \tag{1.6}$$

where, in going to the last line, we once more inserted the identity between each Hamiltonian. It seems like we started, however, now we can take advantage of the limit $N \to \infty$ and approximate Δt to be small. To leading order in Δt the Baker-Campbell-Hausdorff (BCH) formula gives

$$\mathrm{e}^{-\mathrm{i}H\Delta t} \approx \exp\left(-\mathrm{i}\Delta t \frac{1}{2} \Pi^2 \right) \exp\left(-\mathrm{i}\Delta t \left(\frac{1}{2} m^2 \phi^2 + V_{int}(\phi) \right) \right) \tag{1.7}$$

[2] Alternatively, we could shift H such that $E_\Omega = 0$.

[3] If we were genuinely in a quantum mechanical setting, rather than using quantum mechanics as a stepping stone for quantum field theory "finding that the field has value ..." would become "find the particle at position ...".

The second trick is inserting the identity in the form $\int \frac{\mathrm{d}\Pi}{2\pi} |\Pi\rangle\langle\Pi|$, which gives

$$\begin{aligned}\left\langle\phi_{(k+1)}\right| \mathrm{e}^{-\mathrm{i}H\Delta t} \left|\phi_{(k)}\right\rangle &= \int \frac{\mathrm{d}\Pi_{(k)}}{2\pi} \left\langle\phi_{(k+1)}\middle|\Pi_{(k)}\right\rangle \left\langle\Pi_{(k)}\right| \exp(-\mathrm{i}H\Delta t) \left|\phi_{(k)}\right\rangle \approx \\ &\approx \int \frac{\mathrm{d}\Pi_{(k)}}{2\pi} \left\langle\phi_{(k+1)}\middle|\Pi_{(k)}\right\rangle \left\langle\Pi_{(k)}\middle|\phi_{(k)}\right\rangle \cdot \\ &\quad \cdot \exp\left(-\mathrm{i}\Delta t\left(\frac{1}{2}\Pi_{(k)}^2 + \frac{1}{2}m^2\phi_{(k)}^2 + V_{int}\big(\phi_{(k)}\big)\right)\right) \end{aligned} \tag{1.8}$$

where we went from operators to eigenvalues by acting with $\exp\big(-\mathrm{i}\Delta t\big(\frac{1}{2}m^2\phi^2 + V_{int}(\phi)\big)\big)$ on the right and $\exp\big(-\mathrm{i}\Delta t\frac{1}{2}\Pi^2\big)$ on the left.

Using $\langle\Pi|\phi\rangle = \exp(-\mathrm{i}\Pi\phi)$ we get

$$\begin{aligned}\left\langle\phi_{(k+1)}\right| \mathrm{e}^{-\mathrm{i}H\Delta t} \left|\phi_{(k)}\right\rangle \approx \int \frac{\mathrm{d}\Pi_{(k)}}{2\pi} \exp\Big(\mathrm{i}\Pi_{(k)}\big(\phi_{(k+1)} - \phi_{(k)}\big)\Big)\cdot \\ \cdot \exp\left(-\mathrm{i}\Delta t\left(\frac{1}{2}\Pi_{(k)}^2 + \frac{1}{2}m^2\phi_{(k)}^2 + V_{int}\big(\phi_{(k)}\big)\right)\right) \end{aligned} \tag{1.9}$$

Computing this integral is fairly straightforward. First we complete the square by defining

$$x = \Pi_{(k)} - \frac{\phi_{(k+1)} - \phi_{(k)}}{\Delta t} \tag{1.10}$$

to get

$$\begin{aligned}\left\langle\phi_{(k+1)}\right| \mathrm{e}^{-\mathrm{i}H\Delta t} \left|\phi_{(k)}\right\rangle \approx \exp\left(-\mathrm{i}\Delta t\left(-\frac{(\phi_{(k+1)} - \phi_{(k)})^2}{2\Delta t^2} + \frac{1}{2}m^2\phi_{(k)}^2 + V_{int}\big(\phi_{(k)}\big)\right)\right)\cdot \\ \cdot \int \frac{\mathrm{d}x}{2\pi} \exp\left(-\mathrm{i}\Delta t\frac{1}{2}x^2\right) \end{aligned} \tag{1.11}$$

This would be an ordinary Gaussian integral, however, the coefficient is pure imaginary so it is worth revising how to compute (the Cauchy principal value of) this integral. We write a contour in the complex place $C = C_1 + C_2 + C_3$ such that

$$C_1 : z = x,\ x \in [0, R] \tag{1.12}$$

$$C_2 : z = R\mathrm{e}^{\mathrm{i}\theta},\ \theta \in [0, \theta_2] \tag{1.13}$$

$$C_3 : z = r\mathrm{e}^{\mathrm{i}\theta_2},\ r \in [R, 0] \tag{1.14}$$

for $R > 0$, $\theta_2 < 0$, also note that $\Delta t > 0$.

Due to the absence of poles we have:

$$0 = \int_C \mathrm{d}z \exp\left(-\mathrm{i}\frac{\Delta t}{2}z^2\right) = \tag{1.15}$$

$$\begin{aligned} &= \int_0^R \mathrm{d}x \exp\left(-\mathrm{i}\frac{\Delta t}{2}x^2\right) + \int_0^{\theta_2} \mathrm{d}\theta\, \mathrm{i}R\mathrm{e}^{\mathrm{i}\theta} \exp\left(-\mathrm{i}\frac{\Delta t}{2}R^2\mathrm{e}^{2\mathrm{i}\theta}\right) + \\ &\quad + \int_R^0 \mathrm{d}r\, \mathrm{e}^{\mathrm{i}\theta_2} \exp\left(-\mathrm{i}\frac{\Delta t}{2}r^2\mathrm{e}^{2\mathrm{i}\theta_2}\right) \end{aligned} \tag{1.16}$$

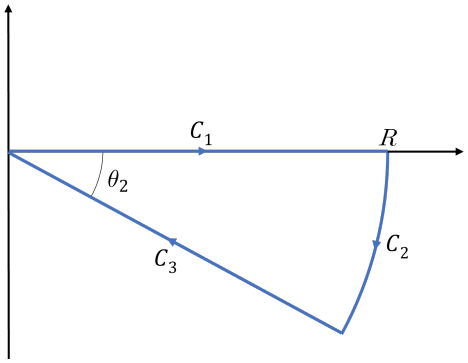


Figure 1: Integration contour in the complex z plane. The idea is to choose θ_2 such that we are able to compute C_3, and that the integral along C_2 will vanish as $R \to \infty$.

The trick is now to choose θ_2 such that the integral along C_2 vanishes as $R \to \infty$ and the integral along C_3 is an ordinary Gaussian integral with a negative coefficient. It is not too hard to see that $\theta_2 = -\frac{\pi}{4}$ is the correct choice. We therefore conclude:

$$\int_0^\infty \mathrm{d}x \exp\left(-\mathrm{i}\frac{\Delta t}{2}x^2\right) = \mathrm{e}^{-\mathrm{i}\frac{\pi}{4}} \int_0^\infty \mathrm{d}r \exp\left(-\frac{\Delta t}{2}r^2\right) = \sqrt{\frac{\pi}{2\mathrm{i}\Delta t}} \tag{1.17}$$

and therefore:

$$\left\langle \phi_{(k+1)} \right| \mathrm{e}^{-\mathrm{i}H\Delta t} \left| \phi_{(k)} \right\rangle \approx \frac{\exp\left(-\mathrm{i}\Delta t\left(-\frac{(\phi_{(k+1)}-\phi_{(k)})^2}{2\Delta t^2} + \frac{1}{2}m^2\phi_{(k)}^2 + V_{int}\left(\phi_{(k)}\right)\right)\right)}{\sqrt{2\pi\mathrm{i}\Delta t}} \tag{1.18}$$

Putting all of these pieces together we find

$$\left\langle \phi_j \right| \mathrm{e}^{-\mathrm{i}H(t_j - t_{j-1})} \left| \phi_{j-1} \right\rangle =$$

$$= \lim_{N\to\infty} \int \left(\prod_{k=1}^{N-1} \frac{\mathrm{d}\phi_{(k)}}{\sqrt{2\pi\mathrm{i}\Delta t}} \right) \exp\left(-\mathrm{i}\Delta t \sum_{k=0}^{N-1} \left(-\frac{(\phi_{(k+1)} - \phi_{(k)})^2}{2\Delta t^2} + \frac{1}{2}m^2\phi_{(k)}^2 + V_{int}\left(\phi_{(k)}\right) \right) \right) \tag{1.19}$$

with the understanding that $\phi_{(0)} = \phi_{j-1}$ and $\phi_{(N)} = \phi_j$.

But how do we interpret this expression? Assuming momentarily that we can take the limit inside of the integral, the interpretation becomes a bit clearer.

The integrand is easier, the first term becomes a time derivative, and the sum becomes an integral. All in all it becomes,

$$\mathrm{e}^{\mathrm{i}S[\phi]} \text{ with } S[\phi] = \int_{t_{j-1}}^{t_j} \mathrm{d}t \left(\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}m^2\phi^2 - V_{\mathrm{int}}(\phi) \right) \tag{1.20}$$

which is just the classical action for our field.

We are then left with a measure which we will denote by

$$\mathcal{D}\phi = \lim_{N\to\infty} \prod_{k=1}^{N-1} \frac{\mathrm{d}\phi_{(k)}}{\sqrt{2\pi\mathrm{i}\Delta t}} \tag{1.21}$$

We have sliced the time interval and then for each time slice we are integrating over all possible values the field can take at that time slice. All subject to the requirement that at times t_{j-1} and t_j the field takes the values ϕ_{j-1} and ϕ_j respectively. We can think of this as integrating over all possible paths the field can take in configuration space in between those two points.

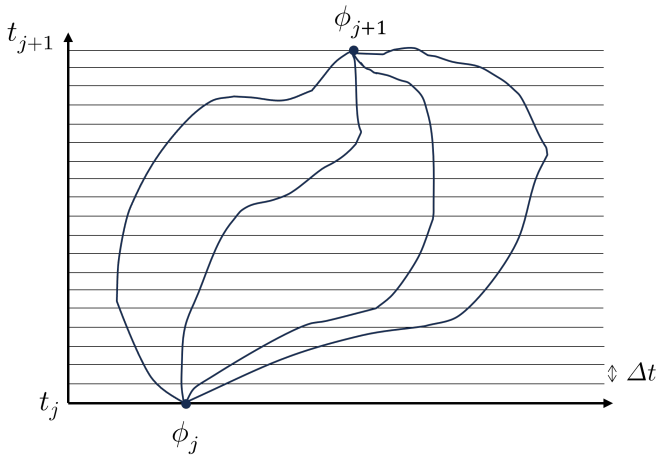


Figure 2: In (1.22) we are integrating over all possible paths that can lead from field value ϕ_j at t_j to ϕ_{j+1} at t_{j+1}

This interpretation is what gives this formalism its name, the *path integral*:

$$\langle\phi_j|\, \mathrm{e}^{-\mathrm{i}H(t_j-t_{j-1})}\, |\phi_{j-1}\rangle = \int_{\phi(t_{j-1})=\phi_{j-1}}^{\phi(t_j)=\phi_j} \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S[\phi]} \tag{1.22}$$

If we had included derivative interactions there would be two changes to the above. Firstly, in (1.8) we would have issues with ordering, unless all the Π were to the left of all ϕ. Secondly, the contour integral would not have been as straightforward as (1.11) would have contained terms with higher powers of x. However, as long as the integral was finite, the only change that could bring is a different V_{int} at the end and a different normalisation for $\mathcal{D}\phi$. As long as we are careful to normalise work correlators appropriately, we can take (1.20) as the starting point for defining V_{int}. Just remember that it might not be the same as the one that appears in the Hamiltonian.

You may be rightly worried regarding the validity of this limiting procedure. In fact, it can be shown that the measure as defined in (1.21) is not a Lebesgue measure in the usual sense, and further that it seems like we have to include non-differentiable field configurations, which would also ruin (1.20). Nevertheless, for the case of ordinary quantum mechanics (that is $D = 1$), there are ways to show that it all works as advertised[4]. However, for higher dimensions, when we have an infinite number of degrees of freedom, the procedure does not work. This is the source of the infamous infinities of quantum field theory.

[4]For a simple demonstration of these details you can check out section 3.2 of https://www.damtp.cam.ac.uk/user/dbs26/AQFT/chap3.pdf

1.3 The initial state

We still need to plug (1.22) into (1.5). To do this we just need to note a few things. Firstly, in the functional integration, the fields now depend on time, the boundary conditions stipulate that $\phi(t_j) = \phi_j$. Therefore the factors of ϕ_j can be replaced by $\phi(t_j)$. Secondly, integrating over field configurations between t_{j-1} and t_j, then integrating over field configurations between t_j and t_{j+1} while matching the field values at t_j, and finally integrating over all possibilities at this matching time, is the same as integrating over all field configurations between t_{j-1} and t_{j+1}.

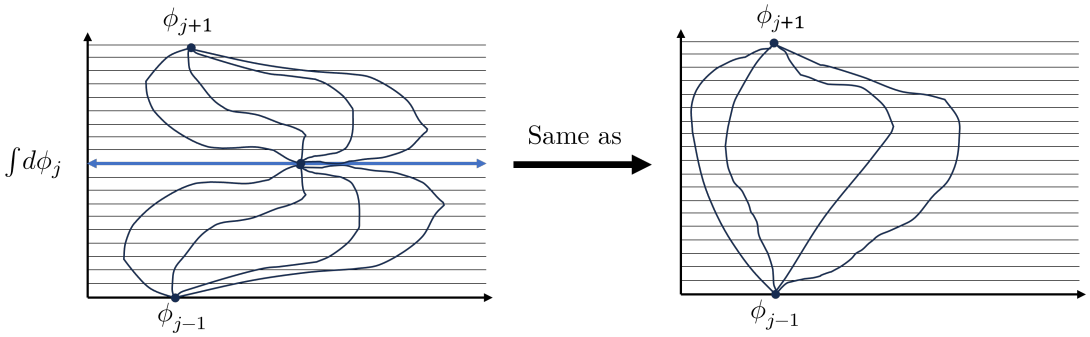


Figure 3: Integrating over all paths between ϕ_{j-1} and ϕ_j, then from ϕ_j to $\phi j + 1$, while integrating over all possible ϕ_j is the same as integrating over all paths between ϕ_{j-1} and ϕ_{j+1} directly

Therefore plugging (1.22) into (1.5) gives us

$$G^{(n)}(t_1, \ldots, t_n) = \int \mathrm{d}\phi_0 \, \mathrm{d}\phi_f \, \langle \Omega | \phi_f \rangle \, \langle \phi_0 | \Omega \rangle \int_{\phi(t_0)=\phi_0}^{\phi(t_f)=\phi_f} \mathcal{D}\phi \; \mathrm{e}^{\mathrm{i}S[\phi]} \prod_{j=1}^{n} \phi(t_j) \tag{1.23}$$

Now we have to sort out the initial and final states. If we pick some arbitrary t_0 and t_f it is quite hard to calculate the position representation of the vacuum state including interactions. However, we can use our freedom to choose them and take the limits $t_0 \to -\infty$ and $t_f \to \infty$ to simplify things.

We can write, expanding in the eigenvalues of the Hamiltonian $|E_n\rangle$,

$$\langle \phi_f | \, \mathrm{e}^{-\mathrm{i}Ht_f} = \sum \langle \phi_f | E_n \rangle \, \langle E_n | \, \mathrm{e}^{-\mathrm{i}E_n t_f} \tag{1.24}$$

If we multiply the Hamiltonian by $(1 - \mathrm{i}\epsilon)$, then in the limit $t_f \to \infty$ all contributions except for the vacuum (by definition, the lowest energy state) are suppressed. If we therefore multiply the Hamiltonian by $(1 - \mathrm{i}\epsilon)$ and take the limit $\epsilon \to 0^+$, regardless of the field boundary conditions at t_0 or t_f we are computing vacuum expectation values.

It is straightforward (exercise!) to check that repeating the above reasoning using instead the Hamiltonian $H_\epsilon = (1 - \mathrm{i}\epsilon)H$ we get the same path integral but with the action

$$S_\epsilon[\phi] = \int \mathrm{d}t \left(\frac{1}{2} \frac{\dot{\phi}^2}{1 - \mathrm{i}\epsilon} - \frac{1}{2}(1 - \mathrm{i}\epsilon) m^2 \phi^2 - (1 - \mathrm{i}\epsilon) V_{\mathrm{int}}(\phi) \right) \tag{1.25}$$

After all of this, we can write to final path integral expression.

$$G^{(n)}(t_1, \ldots, t_2) = \frac{\int \mathcal{D}\phi \ \mathrm{e}^{\mathrm{i}S\epsilon[\phi]} \prod_{j=1}^{n} \phi(t_j)}{\int \mathcal{D}\phi \ \mathrm{e}^{\mathrm{i}S_\epsilon[\phi]}} \tag{1.26}$$

where we are integrating over all times from $-\infty$ to ∞ and we included the denominator to help guarantee we aren't being sloppy with all our normalisations and ensure we maintain $\langle \Omega | \Omega \rangle = 1$.

Note how $\phi(t)$ are now ordinary c-number functions when inside the path integral. They are no longer operators in a Hilbert space. This implies at all the factors of $\phi(t_j)$ commute with one another! The structure of the path integral itself forces the outcome to be time-ordered, it is not optional. If we wished to compute a correlator with a different ordering we would need to build a different path integral, *e.g.* the Schwinger-Keldysh path integral.

2 Computing the path integral

Evaluating an integral over all possible paths seems like a very daunting task, but luckily for us, Feynman found a way to draw diagrams that organise the calculations for us, at least in perturbation theory.

2.1 The Generating Functional

The path integral is an integral over all functions and its integrand itself is a functional (the action), for this reason, it is sometimes also called a "functional integral". Therefore, before we dive in, it will be useful to set some notation and basics of doing calculus in a functional (*i.e.* infinite dimensional) space.

In this case, our vectors are functions $\phi(t)$, and the argument t behaves similarly to an index i. We therefore write the inner product:

$$\phi \cdot \chi = \phi_i \chi^i \longrightarrow \phi \cdot \chi \equiv \int \mathrm{d}t\, \phi(t)\chi(t) \tag{2.1}$$

A linear operator will act as

$$\chi_i = K_i{}^j \phi_j \longrightarrow \chi(t) = \int \mathrm{d}t'\, K(t,t')\phi(t') \tag{2.2}$$

Therefore the identity operator is the Dirac-δ:

$$\mathrm{Id}(t,t') = \delta(t-t') \tag{2.3}$$

so that we write the inverse of an operator satisfying

$$\int \mathrm{d}t''\, K(t,t'')K^{-1}(t'',t') = \delta(t-t') \tag{2.4}$$

We define the functional derivative by the usual power expansion

$$F[\phi + h] = F[\phi] + \int \mathrm{d}t\, \frac{\delta F}{\delta \phi(t)} h(t) + o(h) \tag{2.5}$$

which obeys the fundamental rule

$$\frac{\partial \phi_j}{\partial \phi_i} = \delta^j{}_i \longrightarrow \frac{\delta \phi(t)}{\delta \phi(t')} = \delta(t-t') \tag{2.6}$$

Using this, it is straightforward to show that

$$-i\frac{\delta}{\delta J(t)} \mathrm{e}^{\mathrm{i}\phi\cdot J} = \phi(t)\mathrm{e}^{\mathrm{i}\phi\cdot J} \tag{2.7}$$

Therefore, we can compute the correlation function via the introduction of sources:

$$G^{(n)}(t_1,\ldots,t_2) = \int \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S_\epsilon[\phi]} \prod_{j=1}^{n} \phi(t_j) = \prod_{j=1}^{n} \left(-\mathrm{i}\frac{\delta}{\delta J(t_j)}\right) \int \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S_J[\phi]}\Bigg|_{J=0} \tag{2.8}$$

where

$$S_J[\phi] = S_\epsilon[\phi] + \int \mathrm{d}t\, J\phi \tag{2.9}$$

In this way we have reduced the problem of calculating $G^{(n)}(t_1,\ldots,t_2)$ to the problem of calculating the *generating functional*:

$$Z[m, V_{\text{int}}, J] = \int \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S_J[\phi]} \tag{2.10}$$

We can also repeat the trick to get rid of the potential

$$Z[m, V_{\text{int}}, J] = \exp\left(-\mathrm{i}\int \mathrm{d}t\,(1-\mathrm{i}\epsilon) V_{\text{int}}\left(-\mathrm{i}\frac{\delta}{\delta J}\right)\right) \int \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S_{0,J}[\phi]} \tag{2.11}$$

where

$$S_{0,J}[\phi] = \int \mathrm{d}t \left(\frac{1}{2}\frac{\dot{\phi}^2}{1-\mathrm{i}\epsilon} - \frac{1}{2}(1-\mathrm{i}\epsilon)m^2\phi^2 + J\phi\right) \tag{2.12}$$

Now the only functional integral we need to compute is a Gaussian integral with a linear term, which is much easier.

2.2 Evaluating the free particle generating functional

To compute this Gaussian integral we generalise the trick of completing the square we used in (1.10). Let us first schematically think of a finite dimensional integral where ϕ_a is an ordinary n-dim vector and the action is given by:

$$S = \frac{1}{2}\phi_a M_{ab}\phi_b + J_a\phi_a \tag{2.13}$$

Note that we choose M_{ab} to be symmetric without loss of generality.

We then change variables to

$$\Phi_a = \phi_a + \left(M^{-1}\right)_{ab} J_b \tag{2.14}$$

Two things are worthy of note: first, because M is symmetric then so is its inverse M^{-1}; secondly, there is no Jacobian factor from this change of variables.

Plugging this into the action we get:

$$S = \frac{1}{2}\Phi_a M_{ab}\Phi_b - \frac{1}{2}J_a(M^{-1})_{ab}J_b \tag{2.15}$$

Our new integration variable is Φ, and that integral is an ordinary Gaussian integral. We have traded the linear term with a constant term quadratic in the sources.

In analogy to the above we then define

$$\Phi(t) = \phi(t) + \int \mathrm{d}t'\, G(t-t')J(t') \tag{2.16}$$

for G the functional inverse for the quadratic term in the Lagrangian[5]. Because the quadratic term is a differential operator, it's functional inverse is the *Green's function*:

$$\left(-\frac{1}{1-\mathrm{i}\epsilon}\partial_t^2 - (1-\mathrm{i}\epsilon)m^2\right)G(t-t') = \delta(t-t') \tag{2.17}$$

The boundary conditions are more or less arbitrary[6]. This equation is very easy to solve using Fourier transforms:

$$\tilde{G}(\omega) = \frac{1-\mathrm{i}\epsilon}{\omega^2 - (1-\mathrm{i}\epsilon)^2 m^2} \tag{2.18}$$

And now we see the role of the ϵ more clearly. It serves to shift the would be pole at $\omega^2 = m^2$ away from the real axis. This way, in the limit $\epsilon \to 0^+$ we know what is the correct contour to choose around the poles when integrating over momenta. In the limit, the actual details of the ϵ dependence won't matter, just how to go around the poles. We will then follow what is common in textbooks and write more simply

$$\tilde{G}(\omega) = \frac{1}{\omega^2 - m^2 + \mathrm{i}\epsilon} \tag{2.19}$$

the factor of $(1-\mathrm{i}\epsilon)$ in the potential is completely unimportant so we will drop it.

Let us first consider the case when $V_{\text{int}} = 0$. Plugging (2.16) into (2.11) we then get:

$$Z[m,J] = \mathcal{N}\exp\left(-\frac{\mathrm{i}}{2}\int \mathrm{d}t\,\mathrm{d}t'\, J(t)G(t-t')J(t')\right) \tag{2.20}$$

where

$$\mathcal{N} = \int \mathcal{D}\Phi\ \mathrm{e}^{\frac{\mathrm{i}}{2}\Phi\cdot G^{-1}\Phi} \tag{2.21}$$

where we are using the functional dot product notation introduced earlier, and defining G^{-1} to be the kinetic operator, so that,

$$G^{-1}\Phi = \left(-\frac{1}{1-\mathrm{i}\epsilon}\partial_t^2 - (1-\mathrm{i}\epsilon)m^2\right)\Phi(t) \tag{2.22}$$

This $\mathcal{N}$ factor is independent of the sources J and therefore is a mere normalisation which won't affect the final answer. Nevertheless, it is not too hard to calculate it. We just have to remind ourselves of how Gaussian integrals work in ordinary finite dimensional vector spaces:

$$\int \mathrm{d}^n v\, \mathrm{e}^{\frac{\mathrm{i}}{2}v^a M_{ab} v^b} = \prod_{i=1}^n \int \mathrm{d}u_i\, \mathrm{e}^{\frac{\mathrm{i}}{2}u_i^2\lambda_i} = \prod_{i=1}^n \sqrt{\frac{2\pi\mathrm{i}}{\lambda_i}} = \sqrt{\frac{(2\pi\mathrm{i})^n}{\det M}} \tag{2.23}$$

[5]Time-translation invariance means that G is a function of $t-t'$ rather than t and t' separately

[6]In particular, we need to be able to integrate by parts without boundary terms, and we need to have non-zero overlap with the vacuum wave-functional.

Where in the first equality we used the fact any real symmetric matrix is diagonalisable to convert the n-dimensional integral to a product of one-dimensional integrals; then we used the results from the previous section to compute the complex Gaussian integral; and finally, expressed the product of eingenvalues in terms of the determinant.

In analogy with this expression (and delegating the $(2\pi\mathrm{i})^n$ to the normalisation of the functional measure) we write

$$\mathcal{N} = \frac{1}{\sqrt{\det G^{-1}}} \tag{2.24}$$

where we define the functional determinant as the product of its eigenvalues[7]. In what follows we will mostly ignore this normalisation as it is common between the numerator and the denominator of (1.26).

In conclusion, expanding the exponential in (2.20) we get

$$Z[m,J] = 1 - \frac{\mathrm{i}}{2}\int \mathrm{d}t\,\mathrm{d}t'\, J(t)G(t,t')J(t') - \frac{1}{8}\left(\int \mathrm{d}t\,\mathrm{d}t'\, J(t)G(t-t')J(t')\right)^2 + O(J^4) \tag{2.25}$$

and therefore

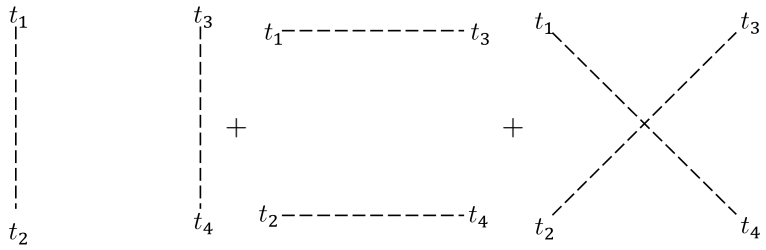
$$\langle\Omega|\phi(t_1)|\Omega\rangle = 0 \tag{2.26}$$

$$\langle\Omega|\mathcal{T}(\phi(t_1)\phi(t_2))|\Omega\rangle = \mathrm{i}G(t_1-t_2) \tag{2.27}$$

$$\langle\Omega|\mathcal{T}(\phi(t_1)\phi(t_2)\phi(t_3))|\Omega\rangle = 0 \tag{2.28}$$

$$\begin{aligned}\langle\Omega|\mathcal{T}(\phi(t_1)\phi(t_2)\phi(t_3)\phi(t_4))|\Omega\rangle = &- G(t_1-t_2)G(t_3-t_4) - G(t_1-t_3)G(t_2-t_4) \\ &- G(t_1-t_4)G(t_2-t_3)\end{aligned} \tag{2.29}$$

We can begin to represent these results diagrammatically, *e.g.* for the 4-point function we could draw:



where each line[8] is associated with a factor of $\mathrm{i}G(t-t')$ with corresponding arguments. From here onwards we will rescale G so that it agrees with the 2-point function (and hence what appears in the diagrams), by defining

[7] Keen-eyed readers will note that this determinant will in general diverge. This is a glimpse of the infinities we will have to contend with later.

[8] Throughout we shall use the convention that scalar fields are denoted with dashes and fermions with continuous lines. Gauge fields would be denoted with wavy lines.

$$\Delta_F(t-t') = \mathrm{i}G(t-t') = G^{(2)}(t,t') \tag{2.30}$$

called the *Feynman propagator*.

In general, it is not too hard to show that, for the $2n$-point function, we just have to draw all possible ways to connect the $2n$ points with n lines and sum them over. This result is known as *Wick's theorem*. And these are the simplest example of *Feynman diagrams*. Next we shall see how to generalise this result for interacting theories.

2.3 Feynman diagrams for the full interacting theory

Including interactions, can write (up to normalisation)

$$Z[m, V_{\text{int}}, J] = \exp\left(-\mathrm{i}\int \mathrm{d}t\, V_{\text{int}}\left(-\mathrm{i}\frac{\delta}{\delta J(t)}\right)\right)\exp\left(-\frac{1}{2}\int \mathrm{d}t\,\mathrm{d}t'\, J(t)\Delta_F(t-t')J(t')\right) \tag{2.31}$$

From here onwards it is a matter of expanding both exponentials and doing combinatorics. However, we can simplify our lives a bit by using the following result

Claim.

$$G\left(\frac{\delta}{\delta J}\right)F(J) = \left.F\left(\frac{\delta}{\delta\phi}\right)G(\phi)\mathrm{e}^{\phi\cdot J}\right|_{\phi=0} \tag{2.32}$$

Proof. Exercise! Tip: first try to prove this for ordinary functions, then extend to vectors and only after think about how it applies for function spaces. □

Now we apply this new result to (2.31) to get:

$$Z = \left.\exp\left(\frac{1}{2}\int \mathrm{d}t\,\mathrm{d}t'\,\frac{\delta}{\delta\phi(t)}\Delta_F(t-t')\frac{\delta}{\delta\phi(t')}\right)\exp\left(\mathrm{i}\int \mathrm{d}t\,(-V_{\text{int}}(\phi)+J(t)\phi(t))\right)\right|_{\phi=0} \tag{2.33}$$

Now we're ready to expand both exponentials and see what we get. A generic term in the expansion, where we are looking at the p^{th} term in the first exponential and the n^{th} term of the second exponential will (schematically) look like:

$$\left.\frac{1}{2^p p!}\Delta_F^p\frac{\delta^{2p}}{\delta\phi^{2p}}\frac{\mathrm{i}^n}{n!}(-V_{\text{int}}+J\phi)^n\right|_{\phi=0} \tag{2.34}$$

Given the source J appears in the expression above in the same way as the interaction potential we don't need to consider them separately, so, just temporarily, let us bring the source inside the potential as a linear term in V_{int} by defining

$$V_{\text{total}} = V_{\text{int}} - J\phi \tag{2.35}$$

Because we are setting $\phi = 0$ at the end of the day, all derivatives must act on all powers of V_{total}, thereby cancelling the $n!$ in the denominator. Moreover, we must have a sufficient number of derivatives to act on every power of the fields in each term of V_{total} until there are no factors of ϕ left.

Just for illustrative purposes, let us think of a quartic interaction potential:

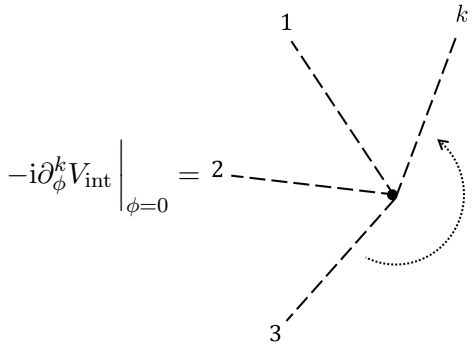
$$V_{\text{int}} = \frac{\lambda_4}{4!}\phi^4 \tag{2.36}$$

Then, for each power of V_{int} in the expansion we would require 4 derivatives with respect to ϕ, so it would look like

$$\frac{\delta}{\delta\phi(t_1)}\frac{\delta}{\delta\phi(t_2)}\frac{\delta}{\delta\phi(t_3)}\frac{\delta}{\delta\phi(t_4)}\int \mathrm{d}t\,\frac{-\mathrm{i}\lambda_4}{4!}\phi^4(t) = -\mathrm{i}\lambda_4\delta(t_4-t_3)\delta(t_4-t_2)\delta(t_4-t_1) \tag{2.37}$$

Note that, because there was a single time integration to begin with, despite the derivatives being wrt to ϕ evaluated at different times, in the end they must all at the same point. In very much the same way that we imagined Δ_F as a *line* connecting two points in time, we can therefore think of these derivatives acting on V_{int} as generating *vertices* with as many lines as factors of ϕ for each term in V_{int}:

$$-\mathrm{i}\partial_\phi^k V_{\text{int}}\Big|_{\phi=0} =$$



(2.38)

where we denoted the derivatives as partial derivatives given they are acting on V_{int} directly, rather than $-\mathrm{i}\int \mathrm{d}t\, V_{\text{int}}$.

From this discussion, it is now clear that the source (which is just a linear term in V_{total}) will correspond to a vertex with a single line and a factor of $\mathrm{i}J(t)$.

$$\mathrm{i}J(t) \bullet\text{----------}\cdot \tag{2.39}$$

Going back to (2.34), due to the Leibniz rule, the derivatives will generate all possible vertices in n powers of V_{total} with a total of $2p$ lines, where we must label each line and vertex uniquely. The propagators will then pair up these lines. Finally we integrate over all time variables.

To get a correlation function, we have to at the end differentiate with respect to $\mathrm{i}J(t)$ and then set J to zero. Therefore, we can calculate (time-ordered) correlators directly with diagrams by associating each field inside the correlation function to an external point whose associated time variable is not integrated over:

$$t \bullet\text{----------}\cdot \tag{2.40}$$

The number of such external points is fixed and should *not* include any extra single line vertices coming from the sources as we have set $J = 0$.

We're almost almost there! We just have to pay attention to the factors in front of the diagrams.

The complication is despite the fact all these diagrams will be distinct if we take into account their labelling, some of them will be completely identical up to relabelling. We say they are *topologically* identical. But, since in the end the internal times are integrated over, the labelling doesn't actually matter and they will give exactly the same contribution. The question is how many diagrams are actually topologically identical.

What made the diagrams identical was the contraction with the propagators, which are symmetric under exchange of the two ends, and are all identical. This means that, at the end, we can freely swap two ends of the same propagator, or exchange two propagators. The terms/diagrams that give the same contribution are the ones who are mapped to one another by this kind of action.

This action forms a discrete group G with size $|G| = 2^p p!$. The factor of 2^p comes from the flipping of the propagators' ends, and the factor of $p!$ comes from the exchange. This is the original factor in front of this term! Take one diagram and call it x. The set of diagrams that can be reached via an element of G from that diagram is called the *orbit* of the group O_x. This is the set of diagrams that give the same contribution. Then the factor you need to include is $|O_x|/|G|$.

Although we are technically done, this is not what is usually taught in textbooks on quantum field theory. To get to the most common phrasing we need to use the *orbit-stabiliser theorem*. We just need one more definition, the stabiliser S_x of the diagram x is the set of elements of G that leave x unchanged as a labelled diagram[9],

Theorem. *Let G be a group acting on a set X, let x be an element of X, let O_x be the orbit of x and let S_x be the stabiliser of x. Then $|O_x||S_x| = |G|$*

This means we can write the factor $|O_x|/|G|$ as $1/|S_x| = 1/S$. This is what physicists usually call the *symmetry factor* of the diagram. To calculate it you need to find the number of actions you can perform on a labelled diagram that leave it completely unchanged (including the labelling), where you can either flip the propagator's ends or exchange it with another propagator.

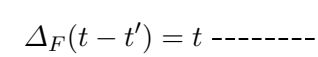
To summarise the *Feynman rules* are as follows:

1. For every field inside the correlation function you associate an external point.

2. Draw every possible diagram (up to the order required) connecting the external points.

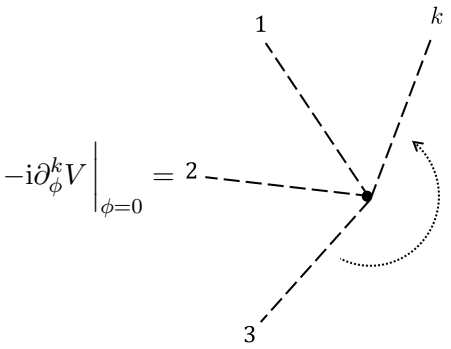
3. Associate a number to each diagram using the following rules:

   - A line joining two points, called a *propagator*, carries a factor of



[9] If you are unfamiliar with this theorem, worry not, we will not have any similar reasoning in what follows. You can also find a very pedagogical explanation on https://gowers.wordpress.com/2011/11/09/group-actions-ii-the-orbit-stabilizer-theorem/

- You join several internal lines in a *vertex*. A vertex with k lines carries a factor of



- Integrate over internal times
- Divide by the symmetry factor, which is the number of actions you can perform on a labelled diagram, that is, one where every end of the each propagator is labelled uniquely, which leave it completely unchanged (including the labelling). The actions are flipping a propagator's ends or exchanging it with another propagator.

2.3.1 Feynman rules in Fourier space

To simplify calculations, it is often a good idea to perform calculations in Fourier space. While the above formalism does hold, we can be a bit clever and change it very slightly.

Notice how the Feynman propagator can be written with a single variable:

$$\tilde{\Delta}_F(\omega) = \frac{\mathrm{i}}{\omega^2 - m^2 + \mathrm{i}\epsilon} \tag{2.41}$$

while a local potential now becomes an integral over several variables (with an overall Dirac-δ):

$$\int \mathrm{d}t\, \phi(t)^4 \longrightarrow \int \prod_{i=1}^{4} \left(\frac{\mathrm{d}\omega_i}{2\pi} \tilde{\phi}(\omega_i) \right) \delta\left(\sum_{j=i}^{4} \omega_j \right) \tag{2.42}$$

To exploit this structure, instead of associating momenta/frequency to points like in position space we instead associate it to lines.

$$\tilde{\Delta}_F(\omega) = \quad \xrightarrow{\quad\omega\quad} \tag{2.43}$$

The overall Dirac-δ in vertices means that the sum of momenta at each vertex should be zero, and hence, momenta is conserved along the diagram. An external point just becomes an external line (with a corresponding factor of the propagator of course). At the end rather than integrating over all points we instead integrate over all undetermined momenta, that is, all momenta that runs around in a loop.

2.4 Connected correlation functions and vacuum bubbles

What we have done above is technically complete. To calculate a (time-ordered) correlation function we just have to include *all* diagrams with a certain number of external points. However, there is some structure we take advantage of and simplify our lives. In particular, we will take into account the *connectedness* of a given diagram.

A diagram is defined as *connected* if all lines are linked together and *disconnected* if there are parts of the diagram which are not linked to any other section.

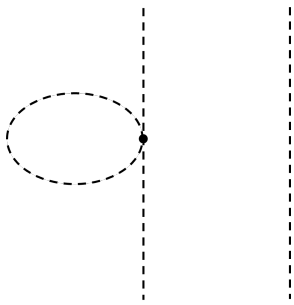


Figure 4: Example of a disconnected diagram

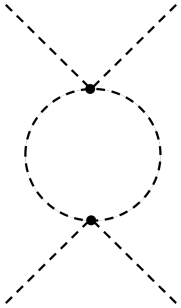


Figure 5: Example of a connected diagram

From the Feynman rules, the value associated to a disconnected diagram is the product of all its connected components. Let the set of all connected diagrams be denoted by $\{\Gamma_a\}$, and the symmetry factor of the diagram Γ_a be denoted $S_{\Gamma_a} \equiv S_a$. We interpret the product of two diagrams $\Gamma_a\Gamma_b$ to be the disjoint union of the two, i.e. the disconnected diagram consisting of one copy of Γ_a and one copy of Γ_b. Then any disconnected diagram is defined by a set of numbers $\{n_a\}$, with $n_a \in \mathbb{N}_0$, which specify how many copies of the diagram Γ_a it contains. The symmetry factor of the disconnected diagram $\Gamma = \prod_a \Gamma_a^{n_a}$ is

$$S_\Gamma = \prod_a n_a! S_a^{n_a} \tag{2.44}$$

This is just a product of all the symmetry factors, times a factor of $n_a!$ to account for the exchange of identical diagrams. Let $F(\Gamma)$ be the factor coming from the Feynman rules associated with a given diagram Γ then

$$F(\Gamma) = \prod_a F(\Gamma_a)^{n_a} \tag{2.45}$$

Then we can write the generating functional as

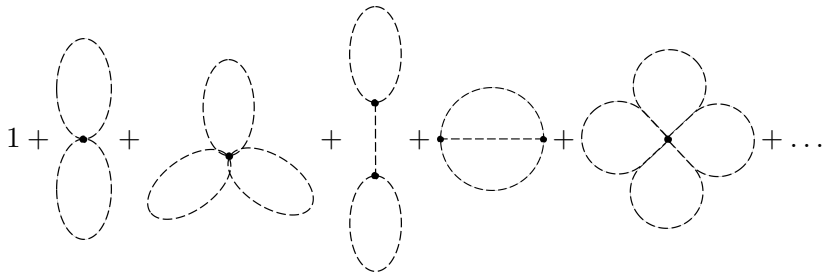
$$\begin{aligned} Z[J] &= \sum_{\Gamma \in \text{disconnected}} \frac{F(\Gamma)}{S_\Gamma} = \sum_{\{n_a\}} \prod_a \frac{F(\Gamma_a)^{n_a}}{n_a! S_a^{n_a}} = \prod_a \sum_{n_a=0}^{\infty} \frac{1}{n_a!} \left(\frac{F(\Gamma_a)}{S_a} \right)^{n_a} = \\ &= \prod_a \exp\left(\frac{F(\Gamma_a)}{S_a} \right) = \exp\left(\sum_a \frac{F(\Gamma_a)}{S_a} \right) = \exp\left(\sum_{\Gamma \in \text{connected}} \frac{F(\Gamma)}{S_\Gamma} \right) \end{aligned} \tag{2.46}$$

There are two consequences of this result.

Firstly, we can use this to finally deal with the normalisation of (1.26). The denominator in that expression (called the *partition function* in analogy with statistical physics):

$$\int \mathcal{D}\phi \; \mathrm{e}^{\mathrm{i}S_\epsilon[\phi]} = Z\Big|_{J=0} \equiv Z_0 \tag{2.47}$$

can be expressed in terms of Feynman diagrams with no external points:



(2.48)

which, due to their appearance are called *vacuum bubbles*.

Given they have no external lines or points, they can never be connected to a diagram with an external point. Therefore, in terms of connected diagrams they form a separate class of diagrams, and we can write in (2.46):

$$Z[J] = \exp\left(\sum_{\Gamma\in\text{connected bubble}} \frac{F(\Gamma)}{S_\Gamma} + \sum_{\Gamma\in\text{connected not bubble}} \frac{F(\Gamma)}{S_\Gamma}\right) = \tag{2.49}$$

$$= \exp\left(\sum_{\Gamma\in\text{connected bubble}} \frac{F(\Gamma)}{S_\Gamma}\right) \exp\left(\sum_{\Gamma\in\text{connected not bubble}} \frac{F(\Gamma)}{S_\Gamma}\right) = \tag{2.50}$$

$$= Z_0 \exp\left(\sum_{\Gamma\in\text{connected not bubble}} \frac{F(\Gamma)}{S_\Gamma}\right) \tag{2.51}$$

Therefore, the *normalised* correlation functions can be obtained by calculating all diagrams that do *not* contain any vacuum bubbles. From this point onwards we will always assume both $Z[J]$ and $G^{(n)}$ have been appropriately normalised.

Finally, given the total correlation function is completely determined by the connected diagrams we can cut the corner and calculate the connected correlation functions $G^{(n)}_{\text{con}}$ directly. To do this we define the *Wilsonian effective action* $W[J]$ by:

$$\mathrm{i}W[J] = \log Z[J] \tag{2.52}$$

From (2.46) it is easy to see that $\mathrm{i}W[J]$ is obtained by summing over all *connected* diagrams, and for that reason, it is the generating functional for connected correlation functions:[10]

[10]It may seem odd that we added a factor of i to the definition of W, this is to make it appear more similarly to the original action, hence also its name. In Part II we will see how, in some cases, this very computation does give rise to an action in the ordinary sense.

$$G^{(n)}_{\text{con}}(t_1,\ldots,t_n) = \prod_{j=1}^{n}\left(-\mathrm{i}\frac{\delta}{\delta J(t_j)}\right)\mathrm{i}W[J]\bigg|_{J=0} \tag{2.53}$$

2.5 1-particle irreducible correlation functions

We might have a good handle on connectedness but there is still more structure left to exploit. Take the following example for a 2-loop correction to the 2-point function:

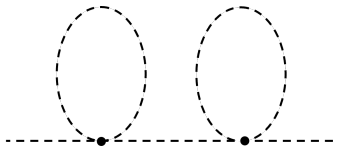


Figure 6: Example of connected but not 1-particle irreducible diagram

Sure it's connected, but still, it is clearly two copies of the same basic diagram joined up with a propagator. There must be some what of only calculating that copy once and then using that to get the full answer. This is topic of this section.

We define a diagram to be *1-particle irreducible* (1PI for short) if it is connected and it remains connected even if we remove any 1 internal line. The example above is not 1PI because if we remove the middle propagator it becomes disconnected. On the other hand, the diagram bellow is 1PI because no matter what line we remove it is still connected.

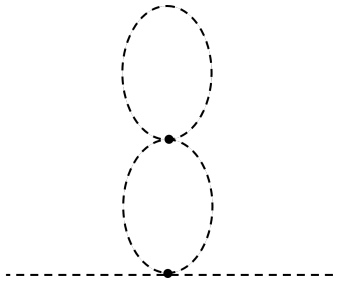


Figure 7: Example of 1PI diagram

If a diagram is connected but not 1PI then that means that there must be at least 1 internal line that once removed makes the diagram disconnected. If you think of the rest of the diagrams as just blobs, then there cannot be any other line that connects those two blobs, because otherwise removing that first internal line would not have made the diagram disconnected. Now we repeat, if either or both of the blobs isn't 1PI on its own then there is one internal line that separates into two other smaller blobs. We divide the diagram in this manner until all blobs that remain are 1PI.

The full diagram must be a tree diagram of these 1PI blobs. There cannot be loops because that would mean that there are two lines connecting a blob but then removing either of those lines would not make the diagram disconnected. In conclusion, a generic connected diagram must be composed of a tree diagram where each vertex is a 1PI diagram[11].

[11] It is tough to put this reasoning into text. The best way to convince yourself that this is true is by

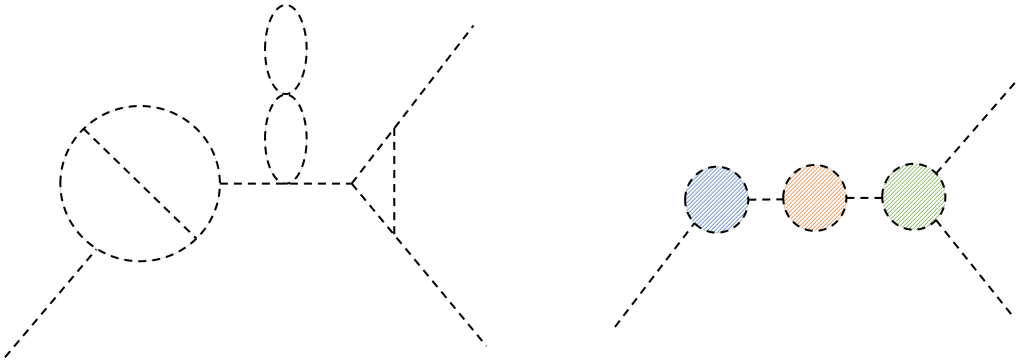


Figure 8: Even a rather complicated diagram, if connected, can be converted to a tree-diagram where each vertex is a 1PI diagram (each blob of differing colours corresponds to a different 1PI diagram)

This means that if we want to find out what object would be the generating functional for these diagrams then we are looking for a functional whose tree diagrams (when put inside of a path integral) would give the *complete* dynamics.

So, first of all, we need to learn how to count loops. To do this, let us briefly return to the usual path integral, but we reinsert factors of $\hbar$:

$$e^{\frac{i}{\hbar}W[J]} = \int \mathcal{D}\phi \; e^{\frac{i}{\hbar}S_J[\phi]} \tag{2.54}$$

Each vertex would receive a factor of $\hbar^{-1}$, but, because we take the inverse of the kinetic term for the propagator, each line would instead have a factor of $\hbar$, and overall $\hbar$ from the definition of $W[J]$[12]. In total a diagram with I lines and V vertices would have a factor of:

$$\hbar^{I-V+1} = \hbar^{L} \tag{2.55}$$

where we used Euler's formula $L = I - V + 1$.[13]

This means we can use $\hbar$ to count loops! And most importantly, if we compute everything in the $\hbar \to 0$ limit the only surviving contributions are the tree-level diagrams. But in the $\hbar \to 0$ limit, the path integral will be dominated by stationary points of S_J[14].

$$\frac{\delta S_J[\phi]}{\delta \phi} = 0 \tag{2.56}$$

These are just the classical equations of motion, albeit with a non-vanishing source. Tree-level diagrams can therefore be computed from the classical action evaluated on a solution to its equations of motion. For this reason we will sometimes refer to the tree-level diagrams as the *semi-classical* contributions and loops as the *quantum* corrections. Note however that the tree-level is still very much quantum mechanical. From a classical

drawing lots of examples and just staring at them for a long while.

[12]This overall factor is, once more, to make it agree with the way the original action S is defined.

[13]Note that for a generic Feynman diagram the number of faces is $L + 1$ because we must also count the outside of the diagram as a "face".

[14]For ordinary integrals this would be known as the "stationary phase approximation"

standpoint alone you would never evaluate $\mathrm{e}^{\mathrm{i}S_J}$ and then take derivatives wrt J. It is merely that we can compute these quantum observables via a classical calculation, hence *semi*-classical.

Therefore we are looking for something whose equations of motion actually compute the full quantum dynamics. Due to this, the 1PI generating function is also called the *quantum effective action*, and we shall denote it by $\Gamma[\varphi]$. Let us then see what form it takes.

We introduce an extra loop counting parameter g independent from $\hbar$ (which has once more been set to 1) and define a new object $W_\Gamma[J,g]$ by

$$\exp\left(\frac{\mathrm{i}}{g}W_\Gamma[J,g]\right) = \int \mathcal{D}\varphi \ \exp\left(\frac{\mathrm{i}}{g}\left(\Gamma[\varphi] + \int \mathrm{d}t\, J(t)\varphi(t)\right)\right) \tag{2.57}$$

By definition the tree-level approximation of the RHS must be equal to original the generating functional $W[J]$. Therefore, expanding in powers of g we find

$$W_\Gamma[J] = \sum_{L=0}^{\infty} g^L W_\Gamma^{(L)}[J] \tag{2.58}$$

for

$$W_\Gamma^{(0)}[J] = W[J] \tag{2.59}$$

All in all, at leading order in g, using the fact the RHS is dominated by the stationary points of the integrand, we find (up to normalisation):

$$\exp\left(\frac{\mathrm{i}}{g}W[J]\right) = \exp\left(\frac{\mathrm{i}}{g}\left(\Gamma[\varphi_J] + \int \mathrm{d}t\, J(t)\varphi_J(t)\right)\right) \tag{2.60}$$

where φ_J is the solution to the equations of motion provided by Γ:

$$\left.\frac{\delta\Gamma[\varphi]}{\delta\varphi(t)}\right|_{\varphi=\varphi_J} = -J(t) \tag{2.61}$$

and so

$$W[J] = \Gamma[\varphi_J] + \int \mathrm{d}t\, J(t)\varphi_J(t) \tag{2.62}$$

which is a Legendre transform.

The Wilsonian effective action is thus the Legendre transform of the quantum effective action. Inverting this is straightforward, we write

$$\Gamma[\varphi] = W[J_\varphi] - \int \mathrm{d}t\, J_\varphi(t)\varphi(t) \tag{2.63}$$

where now φ is our variable, and J_φ is defined via the fact the LHS is independent of J_φ:

$$\left.\frac{\delta W[J]}{\delta J(t)}\right|_{J=J_\varphi} = \varphi(t) \tag{2.64}$$

That is, it is the value of the current such that the vacuum expectation value of the field ϕ attains a value φ.

To summarise, all connected diagrams can be computed from tree-diagrams where each vertex is a 1-particle irreducible diagram. The generating functional for those 1-particle irreducible diagrams, also called the quantum effective action, is the Legendre transform of the Wilsonian effective action. Its equations of motion generate the full quantum dynamics of the theory, justifying its name.

Remembering the Feynman rules for the vertices we can see that the n-point 1PI vertex will be given by:

$$\Gamma^{(n)}(t_1, \ldots, t_n) = \prod_{j=1}^{n} \left(\frac{\delta}{\delta\varphi(t_j)} \right) \mathrm{i}\Gamma[\varphi] \Bigg|_{\varphi=0} \tag{2.65}$$

3 The fermionic path integral

So far we have only deal with path integrals for a single real scalar field and in $D = 1$. Generalising to multiple scalars or higher dimensions is fairly straightforward (in theory at least). Just need to convert some ts into xs add a few indices. There are two cases however which deserve special attention: fermions and gauge fields. Dealing with gauge fields requires a strong understanding of how symmetries interact with the path integral so we will leave this case for the end. In this last section of Part I will learn how to handle fermionic fields.

3.1 Grassmann calculus

When we were deriving a path integral for fermions it was important to introduce the eingenstates $|\phi\rangle$ and their respective eigenvalues ϕ. The issue with fermions is that they obey *anti*-commutation relations, for example, for a Dirac field $\Psi_\alpha(\boldsymbol{x})$:

$$\{\Psi_\alpha(\boldsymbol{x}), \Psi_\beta(\boldsymbol{y})\} = 0 \tag{3.1}$$

so if we tried constructing an eigenstate:

$$\Psi_\alpha(\boldsymbol{x})\,|\psi_\alpha(\boldsymbol{x})\rangle = \psi_\alpha(\boldsymbol{x})\,|\psi_\alpha(\boldsymbol{x})\rangle \tag{3.2}$$

the anti-commutation relation would imply

$$\psi_\alpha(\boldsymbol{x})\psi_\beta(\boldsymbol{y}) + \psi_\beta(\boldsymbol{y})\psi_\alpha(\boldsymbol{x}) = 0 \tag{3.3}$$

But no ordinary complex numbers can satisfy this.

To resolve this matter we introduce a new type of numbers, called *Grassmann* numbers. These are a set of elements $\{\theta^a\}$ obeying the algebra

$$\theta^a\theta^b = -\theta^b\theta^a \quad \text{and} \quad \theta^a\phi^b = \phi^b\theta^a \quad \text{for all } \phi^b \in \mathbb{C} \tag{3.4}$$

The eigenvalues of a fermionic field are defined as taking values in this new set of numbers. Let us then explore some of the properties of these numbers.

Firstly, it is worthy of note that for any a we have $(\theta^a)^2 = 0$ (no sum on a). Therefore, any function of a finite set of Grassmann numbers has a finite expansion:

$$f(\theta) = a + b_a\theta^a + \frac{1}{2}c_{ab}\theta^a\theta^b + \cdots + \frac{1}{n!}f_{a_1\ldots a_n}\theta^{a_1}\ldots\theta^{a_n} \tag{3.5}$$

where n is the number of Grassmann numbers and all coefficients are totally anti-symmetric.

Derivatives are defined in the usual way. For example, taking the function f above:

$$\frac{\partial f}{\partial\theta^a} = b_a \tag{3.6}$$

It satisfies all the usual properties:

$$\frac{\partial \theta^a}{\partial \theta^b} = \delta^a_b \tag{3.7}$$

$$\frac{\partial c}{\partial \theta^a} = 0 \;\forall\; c \in \mathbb{C} \tag{3.8}$$

$$\frac{\partial}{\partial \theta_a}(af(\theta) + bg(\theta)) = a\frac{\partial f}{\partial \theta^a} + b\frac{\partial f}{\partial \theta^a} \tag{3.9}$$

And additionally, that it anti-commutes with the variables themselves, *i.e.*

$$\frac{\partial}{\partial \theta^a}\Big(\theta^c\theta^d\Big) = \delta^c_a\theta^d - \theta^c\delta^d_a \tag{3.10}$$

Integration is a bit more subtle. We will define integration in analogy to $\int_{-\infty}^{\infty} \mathrm{d}x$, we will not find any analogue of an indefinite integral. So first, we impose that we wish to be able to integrate by parts, and therefore:

$$\int \mathrm{d}\theta\, \frac{\partial f(\theta)}{\partial \theta} = 0 \tag{3.11}$$

which, given any function of a single Grassmann variable is linear, that implies

$$\int \mathrm{d}\theta\, c = 0 \quad \forall c \in \mathbb{C} \tag{3.12}$$

The only thing left to specify is the integral of θ itself (using once more that all functions of a single Grassmann variable are linear). We just normalise it to 1:

$$\int \mathrm{d}\theta\, \theta = 1 \tag{3.13}$$

This notion of integration is known as *Berezin integration*. And you will note that it acts in exactly the same way as the derivative!

$$\int \mathrm{d}\theta\, f(\theta) = \frac{\partial f(\theta)}{\partial \theta} \tag{3.14}$$

For multiple integrals, we define in the usual way via nested integration. But pay attention that the ordering matters!

$$\mathrm{d}\theta^a\, \mathrm{d}\theta^b = -\,\mathrm{d}\theta^b\, \mathrm{d}\theta^a \tag{3.15}$$

We will then define

$$\mathrm{d}^n\theta = \mathrm{d}\theta^n \ldots \mathrm{d}\theta^2\, \mathrm{d}\theta^1 \tag{3.16}$$

so that

$$\int \mathrm{d}^n\theta\, \theta^1 \ldots \theta^n = 1 \tag{3.17}$$

and therefore

$$\int \mathrm{d}^n\theta\, \theta^{a_1} \ldots \theta^{a_n} = \epsilon^{a_1 \ldots a_n} \tag{3.18}$$

Finally we come to the strangest property of Berezin integration, the behaviour under change of variables. Let us consider the following integral, where $\theta'^a = A^a{}_b\theta^b$, for A some complex, invertible matrix:

$$\begin{aligned}\int \mathrm{d}^n\theta\, \theta'^{a_1}\theta'^{a_2}\cdots\theta'^{a_n} =& A^{a_1}{}_{b_1}A^{a_2}{}_{b_2}\cdots A^{a_n}{}_{b_n}\int \mathrm{d}^n\theta\, \theta^{b_1}\cdots\theta^{b_n} = \\ =& A^{a_1}{}_{b_1}A^{a_2}{}_{b_2}\cdots A^{a_n}{}_{b_n}\epsilon^{b_1\ldots b_n} = \\ =& \det(A)\epsilon^{a_1\ldots a_n} = \det(A)\int \mathrm{d}^n\theta'\, \theta'^{a_1}\theta'^{a_2}\cdots\theta'^{a_n}\end{aligned} \tag{3.19}$$

therefore we have

$$\mathrm{d}^n\theta = \det(A)\,\mathrm{d}^n\theta' \tag{3.20}$$

which is exactly the opposite of the usual formula for bosonic variables!

We shall finish off this overview with some Gaussian integrals. First for 2 sets of Grassmann variables $\{\theta^a\}$ and $\{\bar\theta^a\}$ (which could correspond to one complex Grassmann variable):

$$\begin{aligned}\int \mathrm{d}^n\bar\theta\,\mathrm{d}^n\theta\,\mathrm{e}^{-\bar\theta^a M_{ab}\theta^b} =& \det(M)\int \mathrm{d}^n\bar\theta\,\mathrm{d}^n\theta'\,\mathrm{e}^{-\bar\theta^a\theta'_a} \\ =& \det(M)\left(\int \mathrm{d}\bar\theta^b\,\mathrm{d}\theta^a\,\mathrm{e}^{-\bar\theta^a\theta^b}\right)^n = \quad \text{(no sum)} \\ =& \det(M)\left(\int \mathrm{d}\bar\theta^b\,\mathrm{d}\theta^a\left(1+\theta^a\bar\theta^b\right)\right)^n = \det(M)\end{aligned} \tag{3.21}$$

where in going to the last line we used the fact only the term with on θ and one $\bar\theta$ would contribute to the integral. Note how, once more, $\det(M)$ appears in the numerator whereas in a bosonic integral it would appear in the denominator.

For a single Grassmann variable the answer is slightly more complicated, let $n = 2m$,

$$\begin{aligned}\int \mathrm{d}^n\theta\,\mathrm{e}^{-\frac{1}{2}\theta^a A_{ab}\theta^b} =& \frac{1}{2^m m!}\int \mathrm{d}^n\theta\, A_{a_1a_2}\ldots A_{a_{n-1}A_n}\theta^{a_1}\ldots\theta^{a_n} = \\ =& \frac{1}{2^m m!}A_{a_1a_2}\ldots A_{a_{n-1}A_n}\epsilon^{a_1\ldots a_n} \equiv \mathrm{Pf}(A)\end{aligned} \tag{3.22}$$

where we have the defined the *Pfaffian* of a matrix: $\mathrm{Pf}(A)$. One can show (Exercise!) that $(\mathrm{Pf}(A))^2 = \det(A)$.

Finally we consider the case of interest: a Gaussian integral with a linear term.

$$Z = \int \mathrm{d}^n\bar\psi\,\mathrm{d}^n\psi\exp\left\{\mathrm{i}\left(\bar\psi^a M_{ab}\psi^b + \bar\eta_a\psi^a + \bar\psi^a\eta_a\right)\right\} \tag{3.23}$$

where ψ^a, $\bar\psi^a$, η^a, and $\bar\eta^a$ are Grassmann and M_{ab} is a complex invertible matrix.

We first complete the square by defining:

$$\chi^a = \psi^a + (M^{-1})^{ab}\eta_b \tag{3.24}$$

$$\bar{\chi}^a = \bar{\psi}^a + \bar{\eta}_b(M^{-1})^{ba} \tag{3.25}$$

plugging this change of variables and noting that there is no Jacobian factor it is straightforward to derive that

$$Z = \left(\int \mathrm{d}^n\bar{\chi}\,\mathrm{d}^n\chi\,\mathrm{e}^{\mathrm{i}\bar{\chi}^a M_{ab}\chi^b}\right)\exp\left\{-\mathrm{i}\bar{\eta}_a(M^{-1})^{ab}\eta_b\right\} \tag{3.26}$$

3.2 Fermionic feynman rules

After having developed the technology of Grassmann calculus it is fairly straightforward, if somewhat laborious, to derive the fermionic path integral. One must simply be careful regarding the definition of the field eigenstates and their respective completeness relation. In the very end the outcome is exactly as expected, except for the fact our fields (and respective sources) are Grassmann-valued:

$$Z[\bar{\eta},\eta] = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi\exp\left\{\mathrm{i}\int \mathrm{d}^D x\left(\bar{\psi}\left(\mathrm{i}\not{\partial} - m\right)\psi - V_{\mathrm{int}}\left(\bar{\psi}_\alpha\psi_\beta\right) + \bar{\eta}\psi + \bar{\psi}\eta\right)\right\} \tag{3.27}$$

where $\not{\partial} = \gamma^\mu\partial_\mu$, and the Dirac-$\gamma$ matrices are defined as follows:

$$\{\gamma^\mu,\gamma^\nu\} = -2\eta^{\mu\nu} \tag{3.28}$$

we also have that $\bar{\psi} = \psi^\dagger\gamma^0$, but in the context of the path integral we shall always deal with $\bar{\psi}$ as an independent Grassmann variable so the precise relation is not very important. We have written V_{int} specifically as a function of $\bar{\psi}_\alpha\psi_\beta$ to emphasise that, despite the possibility of having complicated γ matrices in the interaction terms, both $\bar{\psi}$ and ψ must be present in every term as this is the only way to make a Lorentz-invariant and Grassmann-even action.

An alternative point of view (which will become more familiar in Part II) is that if we will wish to write the most generic action consistent with the symmetries of our theory. Then, because (3.27) is the most generic action we can write down for a single Dirac fermion, whatever the Hamiltonian formalism gives us, must be a special case of (3.27).

To compute correlators we need to be careful because of the anticommuting nature of Grassmann derivatives. There is an extra minus sign when taking the derivatives with respect to the sources:

$$\begin{aligned}&\langle\Omega|\,\mathcal{T}(\psi_{\alpha_1}(x_1)\dots\psi_{\alpha_n}(x_n)\bar{\psi}_{\beta_1}(y_1)\dots\bar{\psi}_{\beta_m}(y_m))\,|\Omega\rangle = \\ &= \prod_{i=1}^{n}\left(-\mathrm{i}\frac{\delta}{\delta\bar{\eta}_{\alpha_i}(x_i)}\right)\prod_{j=1}^{m}\left(\mathrm{i}\frac{\delta}{\delta\eta_{\beta_j}(y_j)}\right)Z[\bar{\eta},\eta]\end{aligned} \tag{3.29}$$

this is because

$$\frac{\delta}{\delta\eta(x)}\int \mathrm{d}^D y\,\bar{\psi}(y)\eta(y) = -\bar{\psi}(x) \tag{3.30}$$

Now we have to repeat the procedure to derive the fermionic Feynman rules. Because every term in the action is Grassmann-even, there are no extra minus signs in deriving the analogue of (2.31). Let us then calculate the quadratic path integral:

$$Z[\bar{\eta},\eta] = \int \mathcal{D}\bar{\psi}\mathcal{D}\psi \exp\left\{ \mathrm{i}\int \mathrm{d}^D x \left(\bar{\psi}(\mathrm{i}\not{\partial} - m)\psi + \bar{\eta}\psi + \bar{\psi}\eta\right)\right\} \tag{3.31}$$

We complete the integral as previously:

$$\chi(x) = \psi(x) + \int \mathrm{d}^D y\, S(x-y)\eta(y) \tag{3.32}$$

$$\bar{\chi}(x) = \bar{\psi}(x) + \int \mathrm{d}^D y\, \bar{\eta}(y) S(y-x) \tag{3.33}$$

for[15]

$$(\mathrm{i}\not{\partial}_x - m)S(x-y) = \delta^{(D)}(x-y) \tag{3.34}$$

$$-\left(\frac{\partial}{\partial x^\mu} S(y-x)\mathrm{i}\gamma^\mu + mS(y-x)\right) = \delta^{(D)}(y-x) \tag{3.35}$$

which can be solved by

$$S(x-y) = \int \frac{\mathrm{d}^D p}{(2\pi)^D} \mathrm{e}^{\mathrm{i}p\cdot(x-y)} \frac{\not{p} - m}{p^2 + m^2 - \mathrm{i}\epsilon} \tag{3.36}$$

where we have included the $\mathrm{i}\epsilon$ prescription to ensure we are calculating vacuum expectation values.

In the end we get (up to normalisation):

$$Z[\bar{\eta},\eta] = \exp\left\{-\mathrm{i}\int \mathrm{d}^D x\, \mathrm{d}^D y\, \bar{\eta}(x) S(x-y)\eta(y)\right\} \tag{3.37}$$

The 2-point function is then given by:

$$\langle\Omega|\, \mathcal{T}(\psi_\alpha(x)\bar{\psi}_\beta(y))\, |\Omega\rangle = \mathrm{i}S_{\alpha\beta}(x-y) \tag{3.38}$$

Similarly to what we did for scalars, we will define the *Feynman propagator* to be equal to the 2-point function, and hence what actually appears in Feynman diagrams:

$$S_F(x-y) = \mathrm{i}S(x-y) \tag{3.39}$$

To get Feynman diagrams, we need the analogues for (2.31) and (2.33). We first write

$$Z[\bar{\eta},\eta] = \exp\left(-\mathrm{i}\int \mathrm{d}^D x\, V_{\mathrm{int}}\left(\frac{\delta}{\delta\eta_\alpha}\frac{\delta}{\delta\bar{\eta}_\beta}\right)\right) \exp\left(-\int \mathrm{d}^D x\, \mathrm{d}^D y\, \bar{\eta}(x) S_F(x-y)\eta(y)\right) \tag{3.40}$$

Generalising (2.32) to Grassmann-variables is fairly straightforward especially because both exponentials are Grassmann-even, we just need to be careful that both of them mix $\bar{\eta}$ and

[15] These two equations are not independent, in fact solving the first one automatically solves the second equation. However, both are needed for the derivation, so it is useful to explicitely write them out.

η and also be careful about the ordering of the sources. Being extremely careful about the minus signs we end up with:

$$Z[\bar{\eta},\eta] = \exp\left(-\int \mathrm{d}^D x\, \mathrm{d}^D y\, \frac{\delta}{\delta\psi(x)} S_F(x-y) \frac{\delta}{\delta\bar{\psi}(y)}\right) \exp\left(-\mathrm{i}\int \mathrm{d}^D x\, V_{\mathrm{int}}(\bar{\psi}_\alpha \psi_\beta)\right) \exp\left(\mathrm{i}\int \mathrm{d}^D x\, (\bar{\eta}\psi + \bar{\psi}\eta)\right)\Big|_{\psi=\bar{\psi}=0} \tag{3.41}$$

The structure is exactly the same as in (2.33) so the gist will be the same. There will be lines with a propagator, vertices with the derivatives of the potential with the same factors of i, and sources for external points. There are a couple of differences which are important to distil.

First, note that the propagator is no longer symmetric under exchange of x and y. This is due to the distinction between ψ and $\bar{\psi}$. Our lines will therefore have an orientation, we shall denote this by an arrow which will from from $\bar{\psi}$ to ψ.

$$(S_F)_{\alpha\beta}(x-y) = y,\beta \longrightarrow x,\alpha \tag{3.42}$$

This explains the absence of the factors of 2 in front of the propagator. This also means the symmetry factors will be slightly different, now there is no symmetry under flipping the propagators ends. The only allowed action is exchange of propagators. However, the combinatorics with the stabiliser and the orbits is unchanged.

This ordering is also clear in the definition for our correlation functions (3.29). We are defining correlators to have all $\bar{\psi}$ on the right and all ψ on the left, and correspondingly for the derivatives wrt the sources. However, we often have to consider diagrams that are identical except for exchange of the final points. These correspond to the derivatives "jumping" over each other to act in a different order. To fix the corresponding minus signs we define a standard order for Feynman diagrams: all points corresponding to ψ are on the right of the diagram with arrows point to them, all $\bar{\psi}$ are to the left of the diagram with arrows pointing away from them. We order the points from first to last.

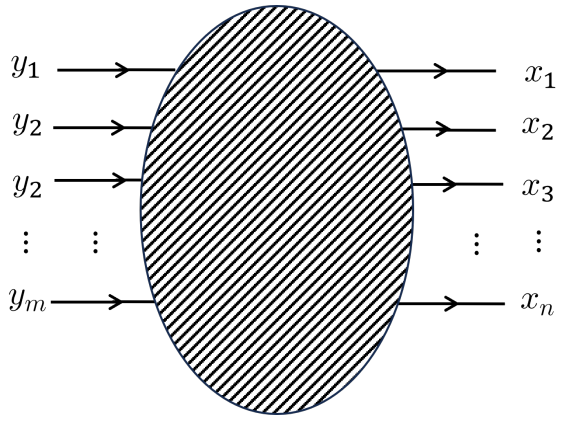


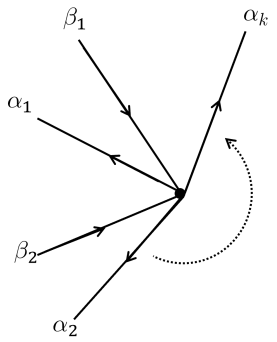
Figure 9: Standard order for external fermions

Any diagram which is an odd permutation of this standard order gets an extra minus sign.

You may be worried about having potentially introduced an overall minus sign. However, the diagrams are just a representation of the equations. We have *defined* the diagram in the standard order to correspond to the contribution with no extra minus sign. Defining a standard order would have changed the map from diagram to expression but not the final expression.

We will write the potential powers of $\bar{\psi}_\alpha \psi_\beta$ always in that order (with potentially some matrices in the middle) as the standard ordering[16]. Then we define the derivatives as acting in powers of $\frac{\delta}{\delta\psi_\beta(x)}\frac{\delta}{\delta\bar{\psi}_\alpha(y)}$. Once more this is just to decide on the sign of every vertex, we are shoving any potential minus signs from this choice of vertex into the propagators.

$$-\mathrm{i}\prod_{i=1}^{k}\left(\frac{\partial}{\partial\psi_{\beta_i}}\frac{\partial}{\partial\bar{\psi}_{\alpha_i}}\right)V_{\mathrm{int}}\Big|_{\psi=\bar{\psi}=0} = \tag{3.43}$$



Finally we worry about the minus signs in the propagators. It is a bit fiddly, but the conclusion is quite simple: we just need to add an extra minus for every fermion loop.

Let us consider every type of line that can occur in turn:

External ψ line For a line like this to happen the $\bar{\psi}$ derivative in

$$-\frac{\delta}{\delta\psi(x)}S_F(x-y)\frac{\delta}{\delta\bar{\psi}(y)} \tag{3.44}$$

needs to act on a $\bar{\psi}$ of a vertex:

$$\bar{\psi}_\alpha M_{\alpha\beta}\psi_\beta \tag{3.45}$$

This does not add any extra minus signs as it only had to potentialy jump over Grassmann-even terms in other vertices or propagators.

And the ψ derivative from (3.44) needs to act on

$$\mathrm{i}\bar{\eta}\psi \tag{3.46}$$

This needs to jump over the ψ in the vertex and the $\bar{\eta}$ in the source, these are two jumps, so no extra minus sign.

In the end we get:

$$-\mathrm{i}(S_F)_{\alpha\beta}(x-y)M_{\beta\gamma}(y)\psi_\gamma(y)\bar{\eta}_\alpha(x)\cdots \tag{3.47}$$

[16] If someone gives a potential which is not written in this form you need to shuffle the fermions and collect the appropriate minus signs.

The factor of $-\mathrm{i}$ in the derivative wrt $\bar{\eta}$ in (3.29) combines with the $-\mathrm{i}$ to give an overall minus sign. But in order for the derivative to act, it needs to jump over the $\psi_\gamma(y)$ remaining (and potentially other Grassmann-even terms), which cancels the minus sign and yielding no overall minus sign[17].

External $\bar{\psi}$ line The reasoning is very similar to the above. But now the $\bar{\psi}$ derivative in (3.44) acts on

$$\mathrm{i}\bar{\psi}\eta \tag{3.48}$$

after potentially jumping over Grassmann-even vertices and/or propagators, so no extra minus sign.

However, the ψ derivative acts on (3.45), this jumps over a $\bar{\psi}$ yielding an overall minus sign.

In the end we get:

$$\mathrm{i}\bar{\psi}_\gamma(x)M_{\gamma\alpha}(x)(S_F)_{\alpha\beta}(x-y)\eta_\beta(y)\cdots \tag{3.49}$$

And now it's the factor of i in the derivative wrt η in (3.29) that combines with the i to give an overall minus sign. But, once more, that derivative needs to jump over $\bar{\psi}_\gamma(x)$ cancelling that minus sign. In the end, once more, there are no minus signs.

Internal line outside of a loop Now we are connecting two separate vertices. We want (3.44) to act on two adjacent copies of (3.45), where each derivative acts on a different copy:

$$\left(-\frac{\delta}{\delta\psi(x)}S_F(x-y)\frac{\delta}{\delta\bar{\psi}(y)}\right)\bar{\psi}_\alpha M_{\alpha\beta}\psi_\beta\bar{\psi}_\gamma M_{\gamma\delta}\psi_\delta \tag{3.50}$$

where the $\bar{\psi}$ derivative acts on the second copy and the ψ derivative acts on the first.

The $\bar{\psi}$ derivative just needs to jump over Grassmann-even terms and then the first copy, which is also Grassmann-even, yielding no minus sign. However, the ψ derivative, only jumps over the first $\bar{\psi}$ of the first copy yielding a minus sign, which cancels the overall minus sign in (3.44) (and (3.50)) and so no overall minus sign.

Closed fermion loop For a closed Fermion loop we need several lines like described above but we also need an extra line that connects two disparate vertices:

$$\cdots\left(-\frac{\delta}{\delta\psi(x)}S_F(x-y)\frac{\delta}{\delta\bar{\psi}(y)}\right)\bar{\psi}_{\alpha_1}M_{\alpha_1\beta_1}\psi_{\beta_1}\cdots\bar{\psi}_{\alpha_n}M_{\alpha_n\beta_n}\psi_{\beta_n} \tag{3.51}$$

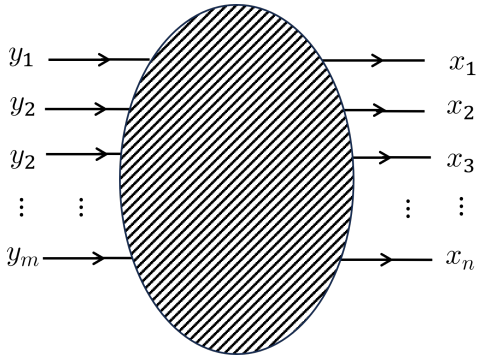
where now the $\bar{\psi}$ derivative acts on the first copy, and the ψ derivative acts on the last.

The first derivative doesn't give any minus sign. The ψ derivative has to jump over ψ_{β_1}, then all the Grassmann-even terms of the remaining vertices and propagators in the loop, and then finally $\bar{\psi}_{\alpha_n}$ from the last vertex to finally act on ψ_{β_n}. Overall it's an even number of jumps and therefore no extra minus sign. We remain with the overall minus sign that originally comes from (3.44) yielding the predicted extra minus sign for fermion loops.

In summary, the Feynman rules for fermionic fields are as follows:

[17] Yes, I know, fermions are gross.

1. For every field inside the correlation function you associate an external point. Points associated with ψ are put on the right hand side of the diagram with arrows pointing towards them. Points associated with $\bar{\psi}$ are put on the left hand side of the diagram with arrows pointing away from them. All of these lines must be horizontal and not cross. Any odd permutation from this standard ordering gets a minus sign.



2. Draw every possible diagram (up to the order required) connecting the external points.

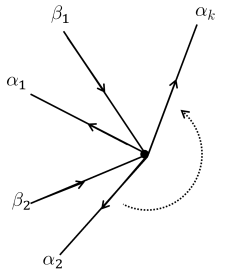
3. Associate a number to each diagram using the following rules:

   - A line joining two points, called a *propagator*, carries a factor of

     $$(S_F)_{\alpha\beta}(x-y) = y,\beta \longrightarrow x,\alpha$$

   - You join several internal lines in a *vertex*. A vertex with $2k$ lines carries a factor of

     $$-\mathrm{i}\prod_{i=1}^{k}\left(\frac{\partial}{\partial\psi_{\beta_i}}\frac{\partial}{\partial\bar{\psi}_{\alpha_i}}\right)V_{\mathrm{int}}\Big|_{\psi=\bar{\psi}=0} =$$

     where once more inward pointing lines are associated with factors of ψ and outward pointing lines to factors of $\bar{\psi}$.

   - Integrate over internal position/momenta
   - Divide by the symmetry factor, which is the number of actions you can perform on a labelled diagram, that is, one where every end of the each propagator is labelled uniquely, which leave it completely unchanged (including the labelling). The only allowed action is exchanging it with another propagator.
   - Every fermion loop receives an additional minus sign

Part II

The Renormalisation Group

4 The Wilsonian Renormalisation Group

Now that we have have a grasp on the path integral we can tackle our main question: how to deal with the infinities of quantum field theory? We begin in this section with the Wilsonian approach, first developed for *statistical* field theories but incredibly insightful for the quantum case as well.

Unless otherwise stated we will be considering a single scalar in 4 dimensions and with a quartic interaction potential:

$$S[\phi] = \int \mathrm{d}^D x \left(-\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 \right) \tag{4.1}$$

4.1 My first infinity

We shall begin with the simplest case, the 2-point function. Using the definitions from Section 2.5, it is straightforward (Exercise!) to show that the connected 2-point function is given by the functional inverse of the quadratic term in the quantum effective action:

$$G^{(2)}_{\text{con}}(x_1, x_2) = (-\mathrm{i})^2 \frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J x_2} \mathrm{i}W[J]\Big|_{J=0} = -\left(\frac{\delta}{\delta\varphi(x_1)} \frac{\delta}{\delta\varphi(x_2)} \mathrm{i}\Gamma[\varphi]\Big|_{\varphi=0} \right)^{-1} \tag{4.2}$$

which makes sense, after all, the propagator is given by the inverse of the quadratic term in the action.

At tree-level, $\Gamma[\varphi]$ is just given by the classical action, hence (and going to Fourier space):

$$\Gamma^{(2)}_{\text{tree}}(p) = -\mathrm{i}\big(p^2 + m^2\big) \tag{4.3}$$

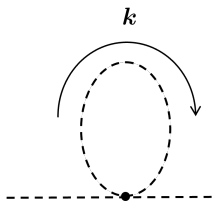
The total 1PI 2-point function is then given by

$$\Gamma^{(2)}(p) = -\mathrm{i}\big(p^2 + m^2\big) + \Gamma^{(2)}_{\text{loops}}(p^2) \tag{4.4}$$

and therefore, the connected 2-point function is:

$$G^{(2)}_{\text{con}}(p) = \frac{-\mathrm{i}}{p^2 + m^2 + \mathrm{i}\Gamma^{(2)}_{\text{loops}}(p^2)} \tag{4.5}$$

Now we start calculating, the 1-loop 1PI correction is:

$\Gamma^{(2)}_{\text{1-loop}} =$ 

$$= -\frac{\lambda}{2} \int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{1}{k^2 + m^2 - \mathrm{i}\epsilon} \tag{4.6}$$

we first do the integration over k_0 to get

$$= \frac{\lambda}{4} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{\sqrt{\mathrm{i}\epsilon - \boldsymbol{k}^2 - m^2}} \underset{\epsilon \to 0^+}{=} -\frac{\mathrm{i}\lambda}{4} \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{\sqrt{\boldsymbol{k}^2 + m^2}} \tag{4.7}$$

where in the last step we took the limit $\epsilon \to 0^+$ as normal.

As $|\boldsymbol{k}| \to 0$ the integral is fine, but as $|\boldsymbol{k}| \to \infty$ the integral grows like:

$$|\boldsymbol{k}|^{d-2} \tag{4.8}$$

therefore the integral is infinite!

Note how the integration of k_0 had no issues whatsoever, it was only the spatial integration that diverged. This makes perfect sense, without the spatial integration we are in ordinary quantum mechanics, for which there are no infinities and everything works fine[18]. It is really the fact we are summing over a continuum of degrees of freedom at short distances that gives us problems.

In order to move forward we must have a handle on this theory. We cannot keep manipulating these infinities forward and backward. In the same way we temporarily sliced the time interval in Chapter 1 to derive the path integral, let us cut off these UV[19] degrees of freedom at some scale Λ_0. So that the integral in (4.7) becomes

$$\frac{\mathrm{i}\lambda}{4} \int_{|\boldsymbol{k}|<\Lambda_0} \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{1}{\sqrt{\boldsymbol{k}^2 + m^2}} \tag{4.9}$$

But there is an issue here: now we are treating time and space completely differently. There is still a long way to go but it seems like if we start like this we will be abandoning Lorentz invariance from the get go and we would have to work hard to recover it. However, we are physicists, so there is a trick: Wick rotation.

The $\mathrm{i}\epsilon$ prescription shifted the poles of the propagator from $\pm m$ to[20] $\pm(m - \mathrm{i}\epsilon)$

We can certainly integrate over the real axis as we did before, but, we can also deform the contour and get the same answer as long as we don't hit any poles. In particular we can rotate the contour a full 90° anti-clockwise by defining a new integrating variable such that

$$\mathrm{i}k_0^{\mathrm{new}} = k_0 \tag{4.10}$$

and integrating over *real* k_0^{new}.

Note how, in terms of the new variable the inner product becomes:

$$k^2 = (k_0^{\mathrm{new}})^2 + \boldsymbol{k}^2 \tag{4.11}$$

[18] There is a caveat to this statement. It is only true because we have no derivative interactions. Derivatives in the vertex would correspond to factors of k_μ in the numerator, which also obviously include factors of k_0. Enough of those would cause the integral over k_0 to not converge.

[19] In analogy with electrodynamics, it is common parlance to use the term ultraviolet (UV) interchangeably with high energy/momentum, and similarly infra-red (IR) to mean low energy/momentum.

[20] Ignoring positive constants in front of ϵ and terms of $O(\epsilon^2)$

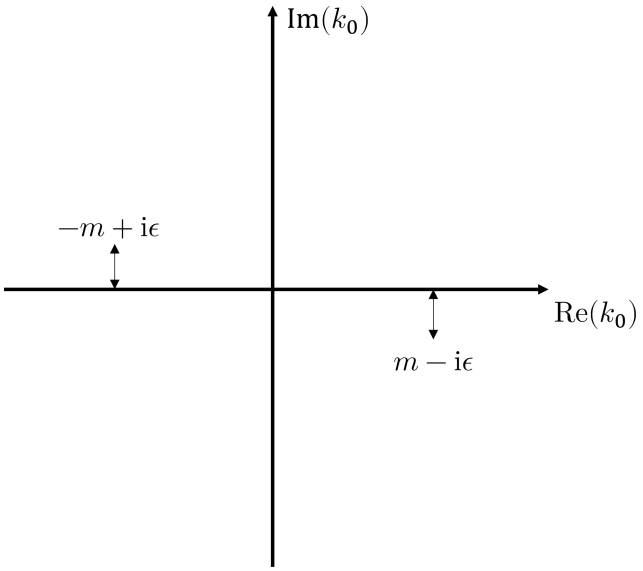


Figure 10: Location of the poles in the propagator in the complex k_0 plane

which is the *Euclidean* inner-product!

Using this integration contour is tantamount to working with a theory in Euclidean space[21] rather than Minkowski spacetime. In terms of position space we write the *Euclidean time* as:

$$\tau = \mathrm{i}t = x^D \tag{4.12}$$

This is called a *Wick rotation*.

The advantage of this formalism is that we can now impose a cut-off on all D variables while still respecting Euclidean *rotational* symmetry:

$$|\boldsymbol{k}_E| < \Lambda_0 \tag{4.13}$$

Euclidean rotational symmetry came from Lorentz invariance in Minwkoski spacetime. The hope is that, if we do all computations in Euclidean space using the above regulator, when we convert back to Lorentzian signature at the end, we still obtain a Lorentz invariant answer.

If all the integrals converged then this procedure is entirely legitimate and does give the correct answers. But, it might seem we actually really want to use this for divergent integrals, in which case the validity of the Wick rotation comes into question. However, a divergent integral is ill-define anyways. Our whole idea at this point is to just have some sort of way to assign a finite value to the integral so that we can manipulate it more easily and hopefully manage to extract the relevant physics. We shall worry in later chapters

[21] You will often physicists use the phrase "Euclidean signature", in opposition to Lorentzian signature. This is especially common in contexts where we are not using the flat metric. However, mathematicians would reserve the term "Euclidean" to flat space, referring to a generic metric with positive signature as having *Riemannian* signature.

how we could have made different choices here and how the final answer depends on these choices.

In summary, our procedure to assign a finite value to the integral (called a *regulator*) is as follows:

1. Wick rotate to Euclidean space
2. Impose $|\boldsymbol{k}_E| < \Lambda_0$ to perform integrals
3. Wick rotate back to Minkowski spacetime

Rather than Wick rotating on every individual integral we could also start with a theory living in Euclidean space from the get-go. Using Euclidean time, the action becomes:

$$S_J[\phi] = \int \mathrm{d}^D x \left(\frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \frac{1}{2}m^2\phi^2 + V_{\text{int}}(\phi) + J\phi \right) \tag{4.14}$$

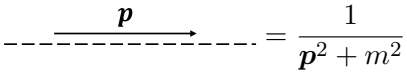
with corresponding generating functional:

$$Z[J] = \int \mathcal{D}\phi \; \mathrm{e}^{-S_J[\phi]} \tag{4.15}$$

Notice how the exponential in the path integral is now *decaying* rather than *oscillatory*. In addition to the above symmetry considerations, using Euclidean space will make it easier to have a convergent path integral, rendering our manipulations more mathematically justified.

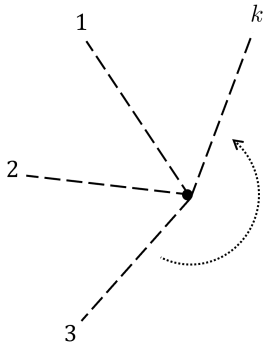
The new Feynman rules are:

- Propagator:



$$= \frac{1}{\boldsymbol{p}^2 + m^2}$$

- Vertex:



$$= -\partial_\phi^k V\Big|_{\phi=0}$$

And (4.5) becomes

$$G^{(2)}_{\text{con}}(\boldsymbol{p}^2) = \frac{1}{\boldsymbol{p}^2 + m^2 - \Gamma^{(2)}_{\text{loops}}(\boldsymbol{p}^2)} \tag{4.16}$$

4.2 Integrating out UV modes

In the previous section we attempted a very simple loop calculations and immediately got infinity out of it. This naturally isn't desirable so we introduced a cut-off. All our fields now have finite support in momentum space:

$$\tilde{\phi}(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda_0 \\ \tilde{\phi}(\boldsymbol{p}) & |\boldsymbol{p}| < \Lambda_0 \end{cases} \tag{4.17}$$

The momentum modes above Λ_0 do not exist at all, we are incapable of answering questions which require momenta higher than that. And of course even for calculations below the cut-off our answer will in general depend on this extra parameter. Our path integral is *defined* with a choice of regulator. Without it we get infinite answers and we cannot even define what we mean by the functional integration.

This whole procedure might be raising some eyebrows. Sure we can get finite answers with this new theory but at what cost? Now we have this extra parameter, but is it physical or just a mathematical crutch? Maybe if we set Λ_0 to be far above our scales of interest it won't matter *too* much? Or should we wish that it drops out in the end entirely, just serving as an intermediate step to define the integrals? Or perhaps it will *not* drop out and actually contribute to *physical* measurements. Would it imply a sort of fundamental discretisation scale of the universe? But we could have used a slightly different cut-off or regulator procedure, which one is right?

All of these questions are perfectly legitimate, and, by the end of Part II of these notes we shall have answered most (if not all) of them. But let us not get ahead of ourselves. First we shall take on a more exploratory point of view and see what we find out. Only after we have gathered some intuition and perspective will we come back and address our concerns.

Let us consider the case when our momenta of interest $\boldsymbol{p}$ are far below the cut-off: $|\boldsymbol{p}| \ll \Lambda_0$. Then there must be a cut-off Λ that, despite being smaller than Λ_0 it still satisfies $|\boldsymbol{p}| \ll \Lambda$. Surely it must be possible for those two theories to yield the same answers, at least for low enough momenta.

With this in mind, we start off in the original theory defined with cut-off Λ_0 and we spilt our field into two components: the useless/high energy/UV modes, ϕ^+, that have support between Λ and Λ_0 i.e.

$$\tilde{\phi}^+(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda_0 \\ \tilde{\phi}(\boldsymbol{p}) & \Lambda < |\boldsymbol{p}| < \Lambda_0 \\ 0 & |\boldsymbol{p}| < \Lambda \end{cases} \tag{4.18}$$

and the useful/low energy/IR modes, ϕ^-, that have support below Λ,

$$\tilde{\phi}^-(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda \\ \tilde{\phi}(\boldsymbol{p}) & |\boldsymbol{p}| < \Lambda \end{cases} \tag{4.19}$$

We also split the sources in the same way to get $J^{\pm}$ which source $\phi^{\pm}$ respectively. To codify the statement that ϕ^+ cannot appear in the external correlators we set J^+ to 0. The generating functional then looks like:

$$Z = \int \mathcal{D}\phi^- \mathcal{D}\phi^+ \; \mathrm{e}^{-S_{\Lambda_0}[\phi^+ + \phi^-] - \int J^- \phi^-} \stackrel{!}{=} \int \mathcal{D}\phi^- \; \mathrm{e}^{-S_{\Lambda}[\phi^-] - \int J^- \phi^-} \tag{4.20}$$

where in the last line we expressed our wish of writing the generating functional solely as an integration over the low energy modes.

To do this we define the *low-energy effective action* as,

$$\mathrm{e}^{-S_{\Lambda}[\phi^-]} = \int \mathcal{D}\phi^+ \mathrm{e}^{-S_{\Lambda_0}[\phi^+ + \phi^-]} \tag{4.21}$$

Note that the $\int J^- \phi^-$ term in the exponential is completely unaffected by this integral.

This path integral looks very scary, but we can make some progress. In momentum space, the action looks like[22]:

$$S[\tilde{\phi}] = \int \frac{\mathrm{d}^D p}{(2\pi)^D} \frac{1}{2} \tilde{\phi}(-\boldsymbol{p})(\boldsymbol{p}^2 + m_0^2)\tilde{\phi}(\boldsymbol{p}) + \frac{\lambda_0}{4!} \int \prod_{i=1}^{4} \left(\frac{\mathrm{d}^D p_i}{(2\pi)^D} \tilde{\phi}(\boldsymbol{p}_i) \right) \delta^{(D)} \left(\sum_{i=1}^{4} \boldsymbol{p}_i \right) \tag{4.22}$$

Note how, in the quadratic term both fields have the same momentum. Therefore, given ϕ^+ and ϕ^- have non-intersecting support in momentum space we conclude there cannot be any quadratic mixing between the two fields, that is:

$$S_{\Lambda_0}[\phi^+ + \phi^-] = S_{\text{free}}[\phi^+] + S_{\text{free}}[\phi^-] + S_{\text{int}}[\phi^+, \phi^-] \tag{4.23}$$

where

$$S_{\text{free}}[\phi] = \int \mathrm{d}^D x \, \frac{1}{2} \left((\boldsymbol{\nabla}\phi)^2 + m_0^2 \phi^2 \right) \tag{4.24}$$

$$S_{\text{int}}[\phi^+, \phi^-] = \int \mathrm{d}^D x \, \frac{\lambda_0}{4!} (\phi^+ + \phi^-)^4 \tag{4.25}$$

Now, since $S_{\text{free}}[\phi^-]$ is independent of ϕ^+, we can pull it out of the integral. Defining $\Delta S_{\Lambda}[\phi^-] = S_{\Lambda}[\phi^-] - S_{\text{free}}[\phi^-]$, we get

$$\mathrm{e}^{-\Delta S_{\Lambda}[\phi^-]} = \int \mathcal{D}\phi^+ \mathrm{e}^{-S_{\text{free}}[\phi^+] - S_{\text{int}}[\phi^+, \phi^-]} \tag{4.26}$$

which has the same form as the Wilsonian effective action[23]! Except now instead of sources we have ϕ^-. We then conclude that, at least in perturbation theory, we can obtain ΔS_{Λ} by evaluating connected Feynman diagrams with ϕ^- playing the role of external sources.

The outcome of these calculations will then generically look like a series expansion in powers of ϕ^-. The connectedness of the diagrams will mean each term will get a Dirac-δ enforcing momentum conservation. However, we might still get a free function of the

[22] We have added a subscript 0 to all our constants to emphasise they are defined at cut-off Λ_0

[23] Indeed this is the origin of the name, from here onwards I will use the terms Wilsonian effective actions and low-energy effective action interchangeably

momenta involved. Nevertheless, we can also expand these functions in powers of the momenta[24]. In position space, the overall Dirac-δ will mean we can express everything in the usual way with an single spatial integration and local interaction terms; the powers of momenta will turn into derivatives of the spatial fields. All in all we get something that schematically looks like this:

$$S_\Lambda[\phi^-] = \int \mathrm{d}^D x \left(\frac{1}{2} Z'_\phi (\boldsymbol{\nabla}\phi^-)^2 + \frac{1}{2} m'^2 (\phi^-)^2 + \frac{\lambda'}{4!} (\phi^-)^4 + \dots \right) \tag{4.27}$$

This really is just an ordinary action for the field ϕ^-, albeit with potentially infinitely many terms. It has couplings which are potentially completely different from the ones for the field ϕ! But we can't quite compare them yet. Our couplings are defined in terms of an expansion of the action in powers of the field, but the field itself and the integration over spacetime both depend on the cut-off. So we might be comparing apples to oranges. To ensure we are using compatible definitions of the couplings we must fix the spacetime integration and the normalisation of our fields.

To fix the spacetime integral we rescale our spacetime variables to ensure the cut-off is the same:

$$\boldsymbol{x} \to \boldsymbol{x}' = \frac{\Lambda}{\Lambda_0} \boldsymbol{x}, \quad \boldsymbol{p} \to \boldsymbol{p}' = \frac{\Lambda_0}{\Lambda} \boldsymbol{p} \tag{4.28}$$

which ensures that $|\boldsymbol{p}| = \Lambda \Rightarrow |\boldsymbol{p}'| = \Lambda_0$.

Fixing the normalisation of the fields is quite different depending on the particular theory and context at hand. In our case the easiest way to do it is to fix the coefficient of the kinetic term to be exactly $1/2$ by defining

$$\phi'(\boldsymbol{x}') = \sqrt{Z_\phi} \phi^-(\boldsymbol{x}) \tag{4.29}$$

where $Z_\phi = \left(\frac{\Lambda_0}{\Lambda}\right)^{D-2} Z'_\phi$. This factor is commonly called the *field strength renormalisation*[25].

The final action is

$$S_\Lambda[\phi'] = \int \mathrm{d}^D x' \left(\frac{1}{2} (\boldsymbol{\nabla}'\phi')^2 + \frac{1}{2} m^2(\Lambda) \phi'^2 + \frac{\lambda_4(\Lambda)}{4!} \phi'^4 + \dots \right) \tag{4.30}$$

This whole procedure is called the *renormalisation group*[26], or RG for short. To summarise, the three steps of RG are:

1. Integrate out the high energy degrees of freedom
2. Fix the spacetime integral by rescaling the spacetime variables
3. Fix the normalisation of the fields

[24] Or more precisely in powers of $\boldsymbol{p}/\Lambda$ which we have assumed to be small.

[25] In fact it is most often called the "wavefunction" renormalisation, but this is a bit of a misnomer since it has nothing to do with wavefunctions.

[26] This name is atrocious. We are not normalising anything neither once nor twice and the action is not group as it does not have an inverse, but it would be criminal of me to call it by any other name.

Let us think about what we have done for a bit. We started off with a certain theory, which was capable of describing momenta below Λ_0 and in particular below $\Lambda < \Lambda_0$. Because no correlator would involve the modes in between the two cut-offs we could change the order of integration in the path integral and integrate over ϕ^+ first. Then we changed the variables of integration inside the generating functional ($\boldsymbol{x}$ and ϕ) to ensure we have compatible definitions for the couplings. All of these are perfectly allowed manipulations of integrals. None of these change the final answer, in fact that was the whole point! Ignoring sources for a moment, we can legitimately write:

$$Z(\Lambda_0, g_{0,a}) = Z(\Lambda, g_a(\Lambda)) \tag{4.31}$$

where $g_{0,a}$ and $g_a(\Lambda)$ schematically represent all the couplings and parameters for cut-offs Λ_0 and Λ respectively.

I stress again, the two theories are not merely related, they are *exactly* the same.

Imagine the vast, *vast* space of all possible couplings g_a. Each point on this space corresponds to a different choice of g_a and hence, to a different choice of *theory*.

This is a wild space, extremely infinite dimensional. Including everything you could possible think of, and even things you wouldn't think of. But, quite remarkably, our innocent little RG has uncovered *structure* in this space. Different points on this space are *not* wholly unrelated. Some of them describe *exactly* the same theory, just at different cut-offs. What we have uncovered is that there are *flow lines* in this space, parametrised by Λ, our cut-off:

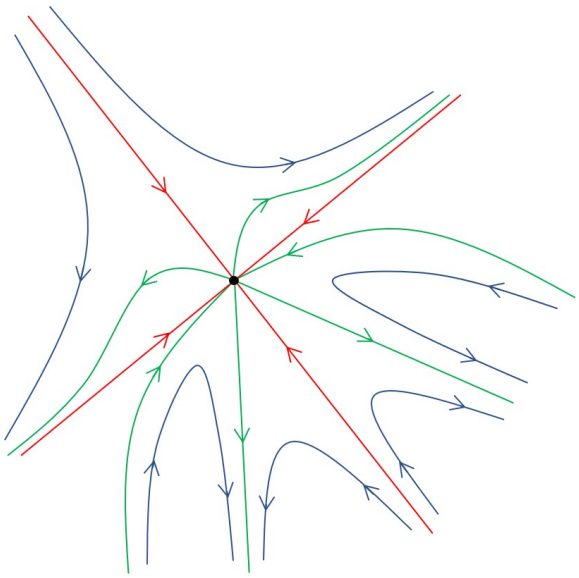


Figure 11: RG flows in theory space

Each and every one of these lines corresponds to the *same* theory and as we flow along them we are lowering the associated cut-off.

In the quest for understanding the infinities of quantum field theory we stumbled across what is quite possibly the deepest and most ambitious programme of all: carving out the

space of all possible theories and understanding its structure.

I also really want to stress that this programme is not merely an abstract theoretical construction. It yields quantitative and *measurable* results. From critical exponents in second order phase transitions to the dependence of the fine structure constant on the scattering energy. All of these have precise theoretical predictions which are accurately born out by experiment.

It will take a while before we can relate this to observables, after all, so far the cut-off could still be just a mathematical crutch. But there is much to reap from RG, and even though we will only scratch the surface, that will already be quite rich.

4.3 Callan-Symanzik equations

What we wrote above is the integral version of the RG flow. However, we can also write differential equations that describe the same process.

First, starting from (4.31), it is pretty straightforward to differentiate with respect to Λ to get:

$$\Lambda\frac{\mathrm{d}Z(\Lambda,g_a(\Lambda))}{\mathrm{d}\Lambda}=\left(\Lambda\frac{\partial}{\partial\Lambda}+\beta_a\frac{\partial}{\partial g_a}\right)Z(\Lambda,g_a(\Lambda))=0 \tag{4.32}$$

where we substituted $\beta_a=\Lambda\frac{\partial g_a}{\partial\Lambda}$, called the β-function of the coupling g_a.

Just looking at this equation gives a different outlook into RG. By varying the cut-off we introduce some change (the first term), but we can cancel that change by varying the couplings as well (the second term) ending up with the same theory at the end.

To figure out what happens for correlators we introduce sources:

$$\begin{aligned}G^{(n)}(\Lambda_0;\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)=&\prod_{j=1}^{n}\left(-\frac{\delta}{\delta J^-(\boldsymbol{x}_j)}\right)Z[J^-;\Lambda_0,g_{0,a}]\Big|_{J^-=0}=\\ =&\prod_{j=1}^{n}\left(-\frac{\delta}{\delta J^-(\boldsymbol{x}_j)}\right)\int\mathcal{D}\phi\ \mathrm{e}^{-S_{\Lambda_0}[\phi]-\int\mathrm{d}^Dx J^-(\boldsymbol{x})\phi^-(\boldsymbol{x})}\Big|_{J^-=0}=\\ =&\prod_{j=1}^{n}\left(-\frac{\delta}{\delta J^-(\boldsymbol{x}_j)}\right)\int\mathcal{D}\phi^-\ \mathrm{e}^{-S_{\Lambda}[\phi^-]-\int\mathrm{d}^Dx J^-(\boldsymbol{x})\phi^-(\boldsymbol{x})}\Big|_{J^-=0}\end{aligned} \tag{4.33}$$

where in the last line we used (4.21) to perform the first step of RG. The integration over $\boldsymbol{x}$ inside the action and the source term are completely independent so we are free to only change the $\boldsymbol{x}$ integral inside the action when we perform steps 2 and 3 of RG:

$$\begin{aligned}G^{(n)}(\Lambda_0;\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)=&\prod_{j=1}^{n}\left(-\frac{\delta}{\delta J^-(\boldsymbol{x}_j)}\right)\int\mathcal{D}\phi'\ \mathrm{e}^{-S_{\Lambda}[\phi']-\int\mathrm{d}^Dx J^-(\boldsymbol{x})Z_\phi^{-\frac{1}{2}}\phi'\left(\frac{\Lambda}{\Lambda_0}\boldsymbol{x}\right)}\Big|_{J^-=0}=\\ =&\int\mathcal{D}\phi'\ \mathrm{e}^{-S_{\Lambda}[\phi']}Z_\phi^{-\frac{n}{2}}\prod_{j=1}^{n}\phi'\left(\frac{\Lambda}{\Lambda_0}\boldsymbol{x}_j\right)=\\ =&Z_\phi^{-\frac{n}{2}}G^{(n)}(\Lambda;\boldsymbol{x}'_1,\dots,\boldsymbol{x}'_n)\end{aligned} \tag{4.34}$$

We can also write this equation as a differential equation, the *Callan-Symanzik equation*:

$$\left(\Lambda\frac{\partial}{\partial\Lambda}+\beta_a\frac{\partial}{\partial g_a}+n\gamma_\phi\right)G^{(n)}\left(\Lambda;\frac{\boldsymbol{x}_1}{\Lambda},\ldots,\frac{\boldsymbol{x}_n}{\Lambda}\right)=0 \tag{4.35}$$

where

$$\gamma_\phi=-\frac{1}{2}\Lambda\frac{\partial\log Z_\phi(\Lambda)}{\partial\Lambda} \tag{4.36}$$

is a sort of β-function for the field rescaling, also called the *anomalous dimension*.

5 Computing RG flows

The best way to get a feel for what an RG flow is actually doing is by sitting down and start calculating. We will proceed in perturbation theory, order by order, however, there are formulations of RG that go beyond perturbation theory.

5.1 Scaling

We will start with the basics: what happens for a quadratic theory? In this case, step 1 is trivial, it just gives a normalisation constant. We then get:

$$S_\Lambda[\phi^-] = \int_{|\boldsymbol{p}|<\Lambda} \frac{\mathrm{d}^D p}{(2\pi)^D} \frac{1}{2} \tilde{\phi}^-(-\boldsymbol{p})\big(\boldsymbol{p}^2 + m_0^2\big)\tilde{\phi}^-(\boldsymbol{p}) \tag{5.1}$$

Step 2 gives us:

$$S_\Lambda[\phi^-] = \int_{|\boldsymbol{p}'|<\Lambda_0} \frac{\mathrm{d}^D p'}{(2\pi)^D} \left(\frac{\Lambda}{\Lambda_0}\right)^D \frac{1}{2} \tilde{\phi}^-(-\boldsymbol{p}) \left(\left(\frac{\Lambda}{\Lambda_0}\right)^2 \boldsymbol{p}'^2 + m_0^2\right) \tilde{\phi}^-(\boldsymbol{p}) \tag{5.2}$$

Therefore, step 3, means:

$$Z_{\tilde{\phi}} = \left(\frac{\Lambda}{\Lambda_0}\right)^{D+2}, \text{ and } \tilde{\phi}'(\boldsymbol{p}') = \left(\frac{\Lambda}{\Lambda_0}\right)^{\frac{D+2}{2}} \tilde{\phi}^-(\boldsymbol{p}) \tag{5.3}$$

The final action is (dropping primes)

$$S_\Lambda[\phi'] = \int_{|\boldsymbol{p}'|<\Lambda_0} \frac{\mathrm{d}^D p'}{(2\pi)^D} \frac{1}{2} \tilde{\phi}'(-\boldsymbol{p}') \left(\boldsymbol{p}'^2 + \left(\frac{\Lambda}{\Lambda_0}\right)^2 m_0^2\right) \tilde{\phi}'(\boldsymbol{p}') \tag{5.4}$$

We don't generate any new terms, but the mass gets rescaled as

$$m^2(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^2 m_0^2 \tag{5.5}$$

or, in terms of the β-function:

$$\beta_{m^2} = \Lambda \frac{\mathrm{d}m^2(\Lambda)}{\mathrm{d}\Lambda} = -2m^2(\Lambda) \tag{5.6}$$

The coefficient on the RHS is negative, which means m^2 *increases* in the IR as we lower the cut-off.

Note that for $m^2 = 0$ this whole procedure has no effect. We say that this is a *fixed point*. This particular one is called the *Gaussian fixed point*.

So far this isn't very interesting, so let us add interactions. At leading order, even with interactions, step 1 is trivial so we once more skip to step 2. Consider a generic interactions with n powers of the field and r derivatives:

$$\int \prod_{j=1}^{n} \left(\frac{\mathrm{d}^D p_j}{(2\pi)^D} \boldsymbol{p}_j^{r_j} \tilde{\phi}(\boldsymbol{p}_j)\right) \lambda_{0,n,r} \delta^{(D)}\left(\sum_{i=1}^{n} \boldsymbol{p}_i\right) \tag{5.7}$$

such that $\sum_i r_i = r$.

Then, on step 2 we get a factor of $\left(\frac{\Lambda}{\Lambda_0}\right)^{nD}$ from the measure, $\left(\frac{\Lambda_0}{\Lambda}\right)^{D}$ from the Dirac-δ, and $\left(\frac{\Lambda}{\Lambda_0}\right)^{r}$ from the derivatives. The third step gives $\left(\frac{\Lambda_0}{\Lambda}\right)^{n\frac{D+2}{2}}$. The end result is:

$$\lambda_{n,r}(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^{\left(1-\frac{n}{2}\right)D+n-r} \lambda_{0,n,r} \Rightarrow \beta_{\lambda_{n,r}} = -\left(\left(1-\frac{n}{2}\right)D + n - r\right)\lambda_{n,r} \tag{5.8}$$

This is similar to the mass β-function, what changes is the coefficient in front of the β-function, and especially whether it's positive or negative, Table 1 summarises some of the possible behaviours.

	$D=1$	$D=2$	$D=3$	$D=4$	$D=5$	$D=6$
ϕ^2	-2	-2	-2	-2	-2	-2
ϕ^4	-3	-2	-1	0	1	2
ϕ^6	-4	-2	0	2	4	6
ϕ^8	-5	-2	1	4	7	10

Table 1: Coefficients in front of the β-functions for powers of scalar fields in various spacetime dimensions

We can see that for high dimensions all interactions shrink in the IR[27]. This suggests they are not very important, even if we suspect they are there for a very high cut-off, we can always just run it a bit down and kill all interactions. With this in mind, we call couplings which shrink in the IR *irrelevant*.

For low dimensions the situation inverts. We can still add enough derivatives to make any interaction irrelevant, but there is an infinite number of interactions which grow in the IR. We cannot get rid of these interactions using RG so we call them *relevant*. Naively this infinite number of relevant interactions makes everything strongly coupled and thus very complicated, however, at this low number of dimensions we have a plethora of other tools available which will be beyond the scope of our treatment.

The 3- and 4-dimensional cases are somewhere in between. We have very few relevant interactions, an infinite number of irrelevant interactions, and a few which appear to not change under RG. The ones with coefficient zero we call *marginal*. To figure out their fate we would need to go to higher order in perturbation theory, then they might turn out to truly be constant, *i.e.* *truly* marginal, or to secretly be relevant or irrelevant, which we would call *marginally (ir)relevant*.

This is why when you were studying QFT you spent so long with ϕ^4 theory. For 4 dimensions it is actually the most generic theory at low energies! All other interaction terms quickly die down as we flow to the IR.

One point is in order. The β-functions we have derived are actually just a consequence of the mass dimension of the interactions we have added. Denoting the mass dimension by square brackets as is customary it is easy to see that by looking at the gradient term:

[27] Note how adding derivatives will just make the coefficient *more* positive.

$$[\phi] = \frac{D-2}{2}, \quad , \quad [\tilde{\phi}] = -\frac{D+2}{2} \tag{5.9}$$

and therefore

$$[\lambda_{n,r}] = \Big(1 - \frac{n}{2}\Big) D + n - r \tag{5.10}$$

which is precisely the coeficient in front of the β-function.

Before we move on, it will be quite instructive to derive this fact more directly. The trick is to express everything in terms of dimensionless couplings. We define them by[28]

$$\lambda_a = \Lambda^{[\lambda_a]} g_a \tag{5.11}$$

Note how we are using the current cut-off of the theory for consistency.

Let us apply the three steps of RG in this notation. We start with

$$\lambda_{0,a} = \Lambda_0 g_{0,a} \tag{5.12}$$

Now we want to do Step 1 of RG, but, to be able to use the previous calculations, we will also part of Step 3. Namely we will already rescale ϕ by Z'_ϕ which came from Step 1. After this, we get new couplings but also a new cutoff:

$$\lambda'_a = \Lambda g'_a \tag{5.13}$$

Now we do the rest of RG: Step 2 plus the rest of Step 3. But this is exactly the calculation we did before, we know that yields $\left(\frac{\Lambda_0}{\Lambda}\right)^{[\lambda_a]}$, so we write

$$\lambda_a(\Lambda) = \Lambda_0 g'_a \tag{5.14}$$

The final steps or RG are just changing the cutoff factor back to Λ_0! Or in equations, using the notation we were using for the dimensionful couplings:

$$g'_a = g_a(\Lambda) \tag{5.15}$$

In terms of the β-function we can write

$$\beta_{\lambda_a}(\Lambda) = \Lambda \frac{\mathrm{d}\lambda_a}{\mathrm{d}\Lambda} = \Lambda_0^{[\lambda_a]} \Lambda \frac{\mathrm{d}g_a}{\mathrm{d}\Lambda} = \Lambda_0^{[\lambda_a]} \beta_a(\Lambda) \tag{5.16}$$

The upshot is that we only have to track the change in the dimensionless coupling to find out the β-function.

This may seem like a huge simplification, but, in practice, sometimes it can be tricky to figure out all of the factors. In this language, neglecting step 1 means

$$\lambda_{0,a} = \lambda'_a \Leftrightarrow \Lambda^{[\lambda_a]} g_a = \Lambda_0^{[\lambda_a]} g_{0,a} \Leftrightarrow g_a(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^{[\lambda_a]} g_{0,a} \tag{5.17}$$

[28] We are using an index a to generalise from the specific basis $\lambda_{n,r}$ we considered earlier to possible linear combinations of these couplings.

making it obvious that

$$\beta_a = -[\lambda_a] g_a \tag{5.18}$$

which is precisely the result we had previously.

So far we have only considered what happens close to the Gaussian fixed point, but this is in fact fairly generic behaviour.

At an arbitrary fixed point RG flow does not do anything so the β-functions vanish. Around this point, they will generically be non-zero but we can linearise them and then diagonalise in the space of all couplings to arrive at

$$\beta_a = -\Delta_a g_a \tag{5.19}$$

In the vicinity of this other fixed point we can still talk of marginal, relevant and irrelevant couplings classifying if deforming in that direction will bring us away or back towards our fixed point of origin. More explicitely:

- $\beta_a > 0 \Leftrightarrow \Delta_a < 0$: irrelevant deformation
- $\beta_a < 0 \Leftrightarrow \Delta_a > 0$: relevant deformation
- $\beta_a = 0 \Leftrightarrow \Delta_a = 0$: marginal deformation

Quite importantly, it is fairly generic that there will be an infinite number of irrelevant operators and only a handful of marginal or relevant operators (except of course in the case of $D = 1$ or 2 where we will have an infinite number of relevant operators). RG flow is highly convergent. Any random deformation will quickly converge on what is usually called the *renormalised trajectory* which only has marginal and relevant interactions. This behaviour is illustrated in figure 12.

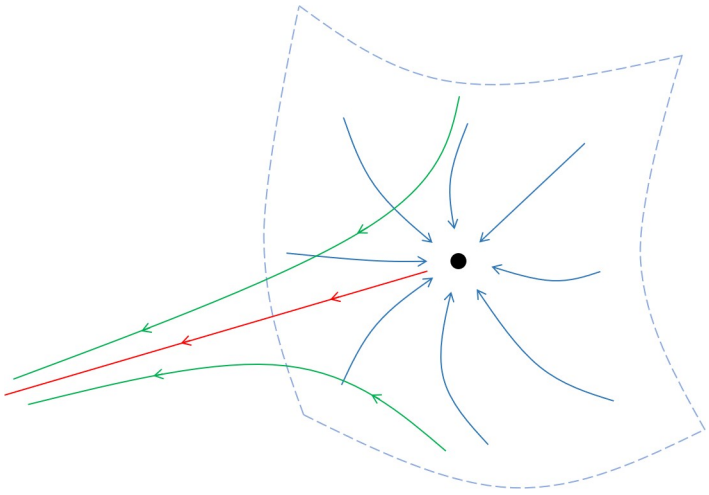


Figure 12: The flow around a generic fixed point. Only a handful of relevant interactions will bring us away from the fixed point along the renormalised trajectory. Most other points will quickly converge on this trajectory.

This is one of the beauties of RG, it gives us a consistent way to get rid of the vast majority of terms we could think about and allows us to only focus on a few. This will ultimately mean that theories with very disparate origins will end up looking very similar in the deep IR which is the phenomenon of *universality*.

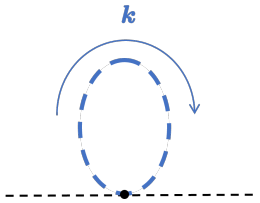
In analogy to this result, around a generic fixed point, the coefficients Δ_a are often called the *scaling dimensions* of the operators. The difference between this scaling dimension and the ordinary mass dimension[29] is often called the *anomalous dimension*. The fact they don't agree was a big mystery for a long time because it naively implied there was a breakdown of dimensional analysis! It just turns out that we have an extra scale to play with, the cut-off, which breaks the assumptions of that analysis.

5.2 The fate of ϕ^4 in $D = 4$

To determine the fate of the ϕ^4 interaction in $D \leqslant 4$ we need to go higher in perturbation theory. However, in $D \leqslant 3$ we will have relevant interactions which will quickly grow and completely ruin our perturbative calculations. At the end we will mention some of the tricks used to deal with lower dimensions, but for now we will restrict our attention to $\lambda\phi^4$ in $D = 4$.

5.2.1 1-loop quadratic terms

The leading order contribution with two external fields is (we are denoting the UV modes with thick blue lines):



$$= -\frac{\lambda_0}{2}\int_\Lambda^{\Lambda_0} \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2+m_0^2} = -\lambda_0\frac{\mathrm{Vol}(S^3)}{2(2\pi)^4}\int_\Lambda^{\Lambda_0}\mathrm{d}k\,\frac{k^3}{k^2+m_0^2} =$$

$$= -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - \Lambda^2 + m_0\log\left(\frac{\Lambda^2+m_0^2}{\Lambda_0^2+m_0^2}\right)\right) \tag{5.20}$$

This means[30]

$$m'^2 = m_0^2 + \frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - \Lambda^2 + m_0\log\left(\frac{\Lambda^2+m_0^2}{\Lambda_0^2+m_0^2}\right)\right) \tag{5.21}$$

Knowing that $g_2 = \Lambda^{-2}m'^2$, we could use (5.21) directly to calculate the beta function. However, it's a lot easier to go directly from the momentum integral, using the fact that

$$\Lambda\frac{\partial}{\partial\Lambda}\int_\Lambda^{\Lambda_0}\mathrm{d}k\,f(k) = -\Lambda f(\Lambda) \tag{5.22}$$

[29] This is also, somewhat insultingly, called the *engineering dimension*

[30] There is an annoying minus sign coming from the fact the Wilsonian action is defined as *minus* the logarithm of the partition function

Either way, the end result is,

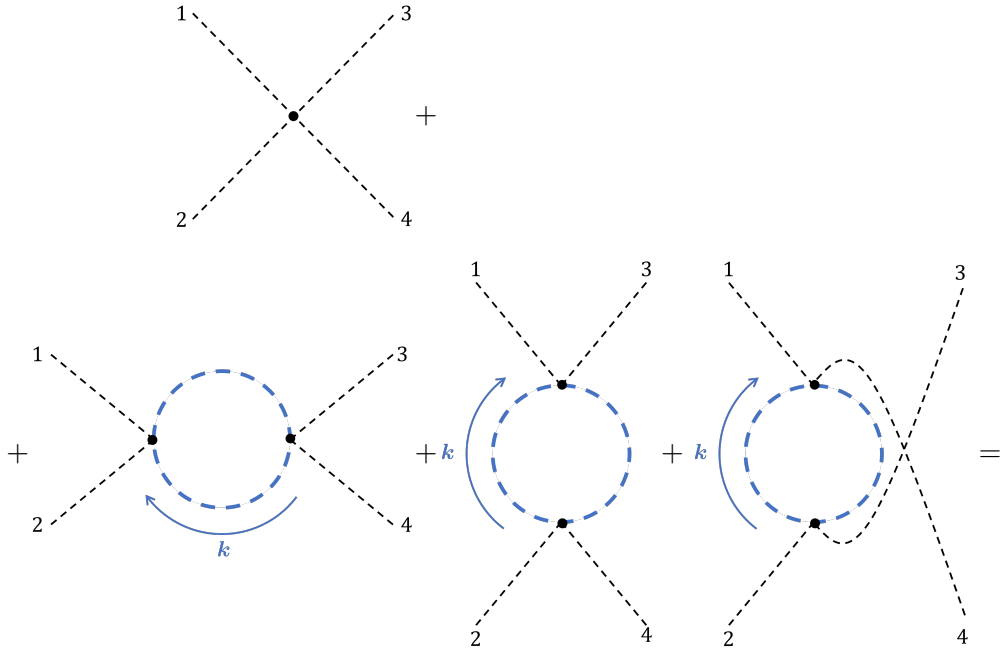
$$\beta_2 = \Lambda\frac{\partial g_2}{\partial\Lambda} = \Lambda\frac{\partial}{\partial\Lambda}\left(\Lambda^{-2}m'^2\right) = -2g_2 - \frac{\lambda_0}{16\pi^2}\frac{1}{1+\Lambda^{-2}m_0^2} = -2g_2 - \frac{1}{16\pi^2}\frac{g_4}{1+g_2} \tag{5.23}$$

where in the final equality we used the fact that, to leading order, we can substitute $g_4 = \lambda_0$ and $g_2 = \Lambda^{-2}m_0^2$, which is legitimate since our calculations are only valid up to that order anyway.

The mass term is still relevant but the extra term demonstrates one of the quirks of RG. We cannot just set $g_2 = 0$ at some high energy scale and just hope RG will hit the Gaussian fixed point. In fact, the quartic interaction will on its own generate a mass term and make us flow along the renormalised trajectory. The "natural" value for the mass is $m^2 \sim g_4\Lambda^2$ which will typically be rather large. To get a mass smaller than this at low energies we would need to "fine tune" m_0^2 to precisely cancel this contribution. Of course there is nothing stopping us from doing so, but it does beg the question of how and why there would be such a precise cancellation.

5.2.2 1-loop quartic terms

The leading order contributions with four external fields are (we shall use conventions were the external momenta are all pointing inwards):



$$\begin{aligned} = -\lambda_0 + \frac{\lambda_0^2}{2}\int_\Lambda^{\Lambda_0}\frac{\mathrm{d}^4k}{(2\pi)^4}\Bigg(&\frac{1}{\boldsymbol{k}^2+m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_2)^2+m_0^2} + \\ &+\frac{1}{\boldsymbol{k}^2+m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_3)^2+m_0^2} + \\ &+\frac{1}{\boldsymbol{k}^2+m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_4)^2+m_0^2}\Bigg) \end{aligned} \tag{5.24}$$

This time the integral depends on external momenta, we need to stop and think about that this means. These integrals are all completely finite, we can just expand in powers of the external momenta to end up with terms that look like $\sim (\tilde{\phi}^-)^2 p^2 (\tilde{\phi}^-)^2$ which in real space correspond to $\sim (\phi^-)^2 \partial^2 (\phi^-)^2$. These terms are all irrelevant so they won't matter too much as they will quickly be swamped by the scaling contribution. This highlights a common tug-of-war in RG calculations: you will always generate all terms that are consistent with the symmetries imposed, you cannot just consider a handful of terms and be done with it. However, the semi-classical scaling will come to the rescue and kill off most of those terms.

The contribution to the original ϕ^4 term with no derivatives is obtained by setting all external momenta to zero:

$$\begin{aligned} &-\lambda_0 + \frac{3\lambda_0^2}{2}\int_\Lambda^{\Lambda_0} \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{(\boldsymbol{k}^2 + m_0^2)^2} = -\lambda_0 + \lambda_0^2 \frac{3\mathrm{Vol}(S^3)}{2(2\pi)^4}\int_\Lambda^{\Lambda_0} \mathrm{d}k \, \frac{k^3}{(k^2+m_0^2)^2} = \\ &= -\lambda_0 + \lambda_0^2 \frac{3}{32\pi^2}\left(\frac{m_0^2}{m_0^2+\Lambda_0^2} - \frac{m_0^2}{m_0^2+\Lambda^2} + \log\left(\frac{m_0^2+\Lambda_0^2}{m_0^2+\Lambda^2}\right)\right) \end{aligned} \tag{5.25}$$

therefore

$$\lambda' = \lambda_0\left(1 - \lambda_0 \frac{3}{32\pi^2}\left(\frac{m_0^2}{m_0^2+\Lambda_0^2} - \frac{m_0^2}{m_0^2+\Lambda^2} + \log\left(\frac{m_0^2+\Lambda_0^2}{m_0^2+\Lambda^2}\right)\right)\right) \tag{5.26}$$

The beta function is then

$$\beta_4 = \Lambda \frac{\mathrm{d}g_4}{\mathrm{d}\Lambda} = \Lambda \frac{\mathrm{d}\lambda'}{\mathrm{d}\Lambda} = \frac{3}{16\pi^2}\frac{\Lambda^4 \lambda_0^2}{(\Lambda^2+m_0^2)^2} = \frac{3}{16\pi^2}\frac{g_4^2}{(1+g_2)^2} \tag{5.27}$$

where in the last line we once again approximated $g_4 = \lambda_0$ and $g_2 = \Lambda^{-2} m_0^2$.

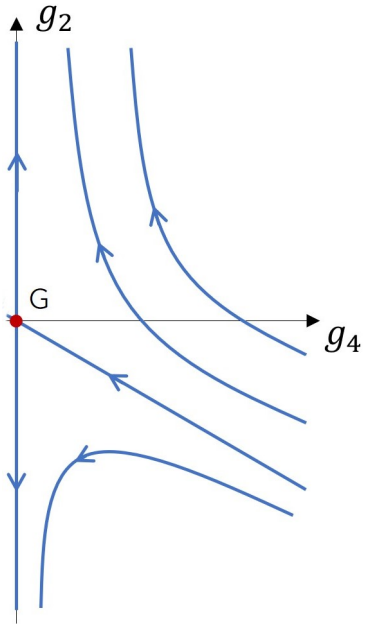


Figure 13: RG flows for $D = 4$ in the (g_2, g_4) plane

The coefficient is positive, so the interaction is in fact *marginally irrelevant*, it still dies off in the IR. In some sense $D = 4$ behaves just like $D > 4$: all interactions are irrelevant, in the IR we will always end up with a free-theory. However, the ϕ^4 term decays much much slower in $D = 4$, rather than polynomial decay we have only a logarithmic decay. Therefore, in physical models it is usually still worth it to include.

5.3 * Lower dimensions

For lower dimensions the story is much more involved. In $D = 1, 2$ we can actually solve the model exactly, in the former by reducing to quantum mechanics (tunnelling will prevent the existence of a fixed point), and in the latter by using very powerful 2 dimensional conformal symmetry techniques to uncover an infinite number of fixed points.

For $D = 3$ we have a dirty trick: we expand in $D = 4 - \epsilon$. This may seem completely crazy at first (because it kind of is), but there a logic to it. The path integral is hard to define for non-integer D but the β-functions can easily be extended to real D. If ϵ is small enough we can actually use the $D = 4$ results with minimal modifications. Then we just have to use some fancy resummation techniques (or shamelessly set ϵ to 1, which works embarrassingly well) to reach $D = 3$. Let us see how this works in practice.

The β-functions for generic D are:

$$\beta_2 = -2g_2 - \frac{\text{Vol}(S^{D-1})}{32\pi^4}\frac{g_4}{1+g_2} \tag{5.28}$$

$$\beta_4 = (D-4)g_4 + \frac{3\text{Vol}(S^{D-1})}{32\pi^4}\frac{g_4^2}{(1+g_2)^2} \tag{5.29}$$

To leading order, we can just take $\text{Vol}(S^{D-1}) = 2\pi^2$ to obtain

$$\beta_2 = -2g_2 - \frac{1}{16\pi^2}\frac{g_4}{1+g_2} \tag{5.30}$$

$$\beta_4 = -\epsilon g_4 + \frac{3}{16\pi^2}\frac{g_4^2}{(1+g_2)^2} \tag{5.31}$$

The remarkable thing about these beta functions is that there is an additional fixed point beyond the Gaussian fixed point, which is called the *Wilson-Fisher fixed point*. To find it, we set the beta functions to zero and solve the resulting system of equations:

$$0 = -2g_2^* - \frac{1}{16\pi^2}\frac{g_4^*}{1+g_2^*} \tag{5.32}$$

$$0 = -\epsilon g_4^* + \frac{3}{16\pi^2}\frac{g_4^{*2}}{(1+g_2^*)^2} \tag{5.33}$$

To leading order in ϵ, the solution for $g_2^* \neq 0$ and $g_4^* \neq 0$ is

$$g_2^* = -\frac{1}{6}\epsilon \tag{5.34}$$

$$g_4^* = \frac{16\pi^2}{3}\epsilon \tag{5.35}$$

Since $g_4^* \sim O(\epsilon)$, this fixed point is at weak coupling and our whole analysis is consistent. Now that we've found a fixed point we should analyse the flow around it. To do that we linearise the beta functions around this point, i.e. we set

$$g_2 = g_2^* + \delta g_2 \tag{5.36}$$

$$g_4 = g_4^* + \delta g_4 \tag{5.37}$$

and expand the beta functions up to linear order in $\delta g_2\,, \delta g_4\,,$ and ϵ. The end result is

$$\Lambda \frac{\mathrm{d}}{\mathrm{d}\Lambda}\begin{pmatrix}\delta g_2\\ \delta g_2\end{pmatrix} = \begin{pmatrix}-2+\frac{\epsilon}{3} & -\frac{1}{16\pi^2}(1+\frac{\epsilon}{6})\\ 0 & \epsilon\end{pmatrix}\begin{pmatrix}\delta g_2\\ \delta g_4\end{pmatrix} \tag{5.38}$$

The eigenvalues of this matrix are $\Delta_2 = -2 + \frac{\epsilon}{3}$ and $\Delta_4 = \epsilon$. Since ϵ is small $\Delta_2 < 0$ and $\Delta_4 > 0$ which means we have one relevant and one irrelevant direction. The relevant direction is the mass direction as usual, the irrelevant direction is some combination of g_2 and g_4 which you can calculate, but it's exact form is not very important. The resulting flow diagram is as follows:

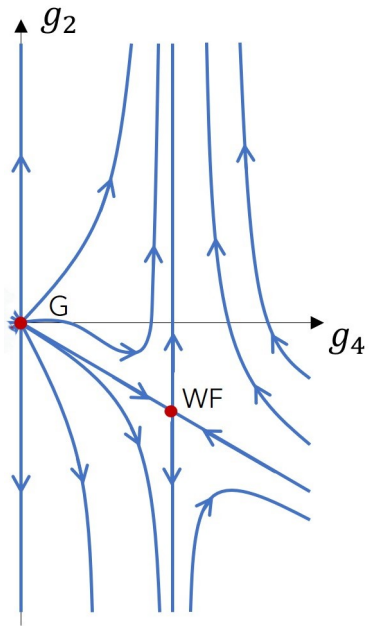


Figure 14: RG flows for $D = 4 - \epsilon$ in the (g_2, g_4) plane

This flow diagram is much richer that we had previously, and although the exact details are only valid for small ϵ, we can hope that the broad structure holds all the way down to $D = 3$ (and independent calculations show that it does!). In summary, we have two fixed points: the Gaussian fixed point, this is the familiar one, it's non-interacting, and, in this projection to the $g_2 - g_4$ plane, it is completely unstable, all deformations lead away from it; and the Wilson-Fisher fixed point, which is an interacting fixed point, and has one stable, and one unstable direction. What can we do with this?

Starting from the Gaussian fixed point we can, for example, flow directly in the g_2 direction. In this trajectory, the theory is always free, we say this is *trivial*. Another possibility is that we turn on the deformation that hits exactly the WF fixed point in the IR, now the theory is interacting in the IR. However, it is a free theory in the UV, for that reason, we say that this theory is *asymptotically free*. If we turn on a direction very close to this one, we will miss the WF fixed point. However, we will very rapidly focus on the renormalised trajectory that corresponds to its mass deformation. If we start on the WF fixed point on the other hand, we can turn on its irrelevant direction which brings us back to the fixed point, and is boring. Alternatively, we can turn on it's relevant direction, g_2, this theory is emanating from a fixed point in the UV but it is an interacting fixed point, we call theories such as this *asymptotically safe*. Any other direction will quickly focus on the renormalised trajectory.

This section is a bit off course from the rest of the notes but it is important for a few reasons. First, it is much richer than in $D = 4$, we can start to see the concepts of asymptotic freedom and asymptotic safety which are very important to categorise and understand quantum field theories. Secondly, it actually describes an actual physical model: the Ising model[31]. This model exhibits a second order phase transition, which is scale invariant, and therefore correspond to a fixed point. Given the Gaussian fixed point is unstable, in the real world, we will only be able to measure the WF fixed point. Correctly explaining the experimental results around this fixed point was one of the first great successes of the RG programme.

[31] This is a simple model for a magnet. Interestingly it is in the same universality class as the critical point for water.

6 Connecting with observables: counterterms and infinities

We began this Part II by realising we need to cut off modes with large momenta. Then we explored a bit the structure behind theories with a cutoff such as that and found that we can define a flow in the space of all theories: RG flow. This tells us that certain theories are actually the same except they correspond to different cutoffs.

However, up to this point, all of this is somewhat academic. The cutoff seems completely artificial and it is very unclear how to make the connection to physical observables. This is precisely the point of this chapter, along the way we shall also understand the meaning of the original methods of "cancelling the infinities".

6.1 Flowing from big to small

Let us revisit the 2-point function. With a cut-off Λ_0, in Euclidean space, and for $D = 4$, the 1-loop 1PI correction becomes:

$$\Gamma^{(2)}_{\text{1-loop}} = -\frac{\lambda_0}{2}\int^{\Lambda_0}\frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2+m_0^2} = -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m_0^2\log\left(1+\frac{\Lambda_0^2}{m_0^2}\right)\right) \tag{6.1}$$

With the cutoff, the integral is perfectly finite, but we are not out of trouble yet. Our original idea was that the cutoff would be much larger than any experimental scales. What's more, given the outcome is infinite without a cutoff, then it must grow as the cutoff gets big, in our case it grows like Λ_0^2. But perturbation theory can only be valid if this 1-loop result is smaller than the tree-level answer! If Λ_0 is too big then it might have regulated the infinity but it is still destroying our perturbative expansion.

Introducing a cutoff only changed 'infinite' into 'big', but 'big' is still bad enough to ruin perturbation theory. Nevertheless, the tools from the previous chapters will tell us how to rescue our perturbative expansion.

Note that the calculation above was independent of $\boldsymbol{p}^2$. We can interpret the result by saying that there is a mismatch between the physical mass and the quadratic term in the action[32]:

$$m^2_{\text{phys}} = m_0^2 + \frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m_0^2\log\left(1+\frac{\Lambda_0^2}{m_0^2}\right)\right) \tag{6.2}$$

And now the solution becomes more clear. The whole point of the construction of the last two chapters was that as we flow with RG to lower cut-offs the theory is *exactly* the same. Physical questions have exactly the same answer wherever we might be in the RG flow. This means we aren't stuck with a high cutoff. We can flow along the RG to as low a cutoff as we want, as long as it's still above the scales of interest. The only catch is that the original couplings like m_0 and λ_0 will have to change. But who cares? What we measure is the final answer, we would always have to fit the correlators to find the couplings. They are not sacred constants that we have to abide by, we can perfectly send an experimentalist

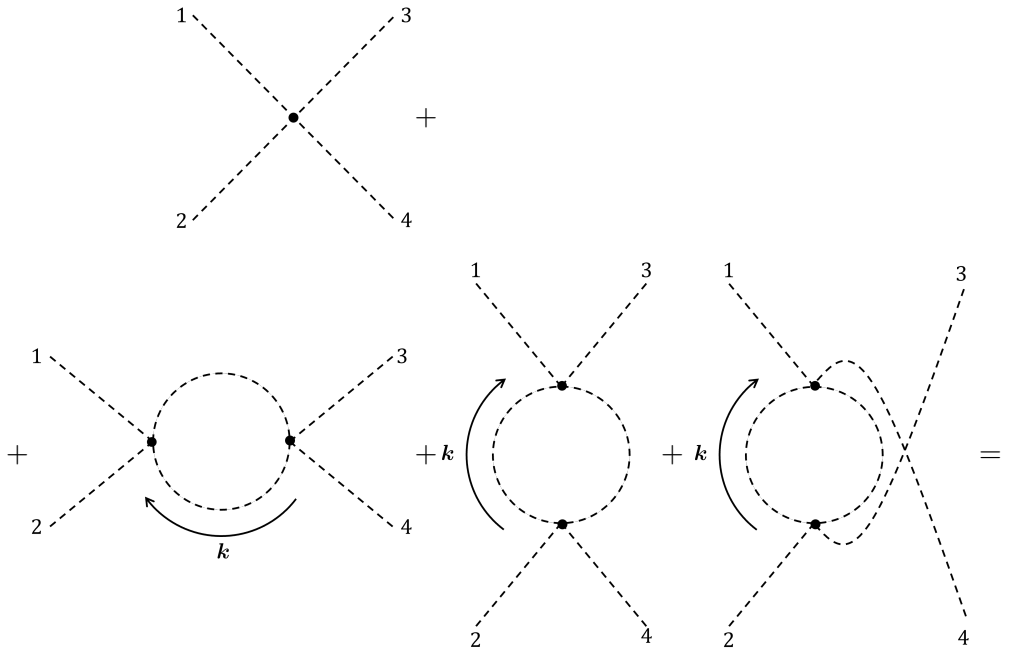
[32] In fact, one can show in very general grounds that the pole of the 2-point function (or equivalently the zero of the 1PI version) is the physical mass by resorting to the Källén–Lehmann spectral representation, you can check out more details in Timo Weigand's notes.

$$m_{\text{phys}}^2 = m'^2 + \frac{\lambda'}{32\pi^2}\left(\Lambda^2 - m'^2 \log\left(1 + \frac{\Lambda^2}{m'^2}\right)\right) \tag{6.3}$$

where we tell them that Λ is the maximum momenta they can handle and tell them to give you m'. The only concern you might have is that ϕ^2 is a relevant coupling so m' will be potentially much bigger than m_0, however, we are not treating that coupling perturbatively so it is not crucial that it remains small.

This case might still be a bit too trivial. After all, what you will measure experimentally is m_{phys}, not m' or Λ. The discussion above was essentially just a pat on our backs to allow us to trust perturbation theory. Let us look at the 4-point function to arrive at something a bit less trivial,

The diagrams that contribute are:



$$\begin{aligned} = -\lambda_0 + \frac{\lambda_0^2}{2}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\Bigg(& \frac{1}{\boldsymbol{k}^2 + m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_2)^2 + m_0^2} + \\ & + \frac{1}{\boldsymbol{k}^2 + m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_4)^2 + m_0^2} + \\ & + \frac{1}{\boldsymbol{k}^2 + m_0^2}\frac{1}{(\boldsymbol{k}+\boldsymbol{p}_1+\boldsymbol{p}_3)^2 + m_0^2}\Bigg) \end{aligned} \tag{6.4}$$

Where, in the first integral, we should integrate over $\{\boldsymbol{k}|\ |\boldsymbol{k}| < \Lambda_0 \wedge |\boldsymbol{k}+\boldsymbol{p}_{12}| < \Lambda_0\}$, for $\boldsymbol{p}_{ij} = \boldsymbol{p}_i + \boldsymbol{p}_j$ because these are the momenta that appear in the propagators. Similar comments apply to the remaining integrals.

These integrals are much harder to do because we cannot just get rid of the external momentum dependence. But we don't need the full expression to get the main picture. We

are potentially concerned with the large Λ_0 regime. Doing some rough scaling, we see that we have three powers of $\boldsymbol{k}$ on top and four on the bottom. For $\Lambda_0^2 \gg m_0^2$ and $\Lambda_0^2 \gg \boldsymbol{p}_{ij}^2$ and in the region of large $\boldsymbol{k}$ our integrals look like

$$\int^{\Lambda_0} \frac{\mathrm{d}k}{k} \sim \log(\Lambda_0) \tag{6.5}$$

which is a slower growth but still worrying. Nevertheless, we can once more choose a smaller Λ_0 (and corresponding λ_0) to tame this behaviour. Alternatively you can also think that we started with a larger cutoff and then we flowed to a small one using RG. The end result is the same.

The lower end of the integral is also under control. Massaging the expression $|\boldsymbol{k} + \boldsymbol{p}_{ij}| < \Lambda_0$ we get (Exercise!):

$$\cos\theta < \frac{\Lambda_0^2}{2|\boldsymbol{p}_{ij}||\boldsymbol{k}|} - \frac{|\boldsymbol{p}_{ij}|}{2|\boldsymbol{k}|} - \frac{|\boldsymbol{k}|}{2|\boldsymbol{p}_{ij}|} \tag{6.6}$$

where $\cos\theta$ is the angle between $\boldsymbol{k}$ and $\boldsymbol{p}_{ij}$.

From this you show that if $\Lambda_0 < \frac{|\boldsymbol{p}_{ij}|}{2}$ (and $|\boldsymbol{k}| < \Lambda_0$) then the RHS is smaller than -1 and therefore the constraint cannot hold. In the end this means there is an effective lower bound on the integral, because the integration from 0 to $\frac{|\boldsymbol{p}_{ij}|}{2}$ vanishes.

We can only potentially have issues if the lowest $\frac{|\boldsymbol{p}_{ij}|}{2}$ is very different from the highest $|\boldsymbol{p}_i|$ because then the smallest possible cut-off would still be too big. In fact a similar issue could also arise in the 2-point function when $\boldsymbol{p}^2 \gtrsim \frac{32\pi^2 m_0^2}{\lambda_0}$. This is an important lesson, RG can only *improve* perturbation theory it might not *save* it. If we have an intrinsic large separation of scales we need different techniques.

Let us step back and look at the big picture. The crux of the argument is that by using RG flows we can actually choose whatever cutoff we want, the catch is that different choices of cutoffs will correspond to different values of the couplings. If we go back to theory space again

INSERT PICTURE

What we want is not to find which 'point' we're at, we want to find which 'flow line' we're at. This is because if two different points are on the same flow line then they actually correspond to the same theory, it's just that the cutoffs are different. To do this in practice we have to choose a particular value for the cutoff and then fit our experimental data to (6.4) and (6.2) to get the values for m_0 and λ_0. In essence, we are cutting through theory space at constant Λ_0

INSERT PICTURE

And by finding the values for m_0 and λ_0 we are finding the point we are at on this cut, which in turn will tell us which flow line corresponds to experiment.

What our discussion regarding 'bigness' and 'infinity' is getting at is that for large enough values of the cutoff it is essentially impossible to preserve perturbation theory. However, by choosing as small a cutoff as possible we are giving as much chance for our perturbation theory to be valid as we can. Of course, it might still not work, the measured

value of λ_0 might just be big on its own, but a big cutoff would have ensured a failure of perturbation theory, even if λ_0 was small.

If we later need an experiment that is performed at a different energy scale we can use RG flows to change the cutoff to a more appropriate value and improve perturbation theory as much as possible.

With all of this in mind we can finally understand where the infinities were coming from. By insisting on using no cutoff we were making a cut at $\Lambda_0 = \infty$, which is a very singular point in the RG flow! Usually couplings will be infinite at infinite cutoffs which is what led people to try to find a way for these two infinities to cancel and yield a consistent answer. Now we know this completely unnecessary, we can cut theory space at a finite Λ_0 and parametrise the flow with that. No infinities in sight, and we have a completely consistent result.

Nevertheless, you might still be somewhat worried about the arbitrariness in our cutoff procedure. Clearly we could have used a different regulating procedure, how would they be related? There are a couple of answers one can give to ease these concerns.

Firstly, there are two kinds of different choices which are obviously fine. Any choice that would correspond to a change of basis in theory space or to a reparametrisation of the flow. Quite clearly this is analogous to a coordinate or gauge choice and won't affect the physics (but it might greatly affect the difficulty of the computations!).

Secondly, in the next chapter we will find a procedure that will allow us to *demand* that two choices of regulators give the same answer. It is a bit forceful but it is the most common practice.

Thirdly, under certain assumptions one can actually prove semi-rigorously that different correlators are physically equivalent. The methods involved are a bit beyond the scope of these lectures [33]. The result is that any two different smooth regulators (*i.e.* ones that smoothly dampen high momenta modes) yield exactly the same value for all correlators in the $\Lambda_0 \to \infty$ limit, assuming that that limit is well behaved.

6.2 Generalising RG by directly matching observables

Let us just look at one more example, the 6-point function. We haven't introduced a ϕ^6 term so this is a slightly different case.

Just doing some rough scaling on the 1-loop 6-point function we see that we have 3 propagators, each giving a k^{-2}; and one integral giving a k^3, so in total

$$\textbf{INSERT DIAGRAM} \sim \int^{\Lambda_0} \frac{\mathrm{d}k}{k^3} \sim \frac{1}{\Lambda_0^2} \tag{6.7}$$

which isn't 'big' at all!

This behaviour is really quite special. If we for example do this calculation in 6 dimensions rather than 4, then we have

$$\textbf{INSERT DIAGRAM} \sim \int^{\Lambda_0} \frac{\mathrm{d}k}{k} \sim \log(\Lambda_0) \tag{6.8}$$

[33] If you are curious, it involves Exact RG to deal with smooth regulators CITE

which grows logarithmically.

In principle, with our previous philosophy, despite ϕ^6 being irrelevant, we should still have introduced it and parametrised its flow. Being irrelevant just means it shrinks in the IR, but maybe it was actually really really big in the UV so right now it is still measurable. However, it seems like if we just hadn't introduced it nothing would go wrong. It might be somewhat overkill to bring the cut-off down for the 6-point correlator if it was small to begin with. At least in 4 dimensions that is.

This suggest there is still some structure we are missing and, perhaps more importantly, not abusing. We need to consider more generic ways to regulate and parametrise the RG flow and to figure out some sort of 'minimal' regulation that will not bother with regulating diagrams which were not big to begin with.

We have been taking a constructive approach to RG. Starting from things we understand and gradually building upon them to get the structure. Now that we have used that structure to do its purpose, we can figure out what were the key elements and what were just artefacts of our starting point. In this way we shall be able to consider more generic RG flows which will help us doing calculations in practice (and understand how most people do these calculations).

The main accomplishments our construction had were turning 'infinity' into 'big', and then turning 'big' into 'small'. The way to turn 'infinity' into 'big' was via the introduction of a regulator; whereas turning 'big' into 'small' is a bit more restrictive, the key aspect being respecting equations (4.31) and (4.35) which ensure the physical observables stay the same as we mess with the regulator to improve our perturbative expansion.

Given we potentially need to change the regulator by a lot to improve perturbation theory, it's probably a bit more general to start with (4.31) rather than the differential versions.

So, what do we want? We have to start with the partition function $Z(\Lambda_0, g_{0,a})$, with a given regulator (we are calling it Λ_0 but it doesn't have to be a cutoff). This theory is perfectly finite but still has some 'big' terms in the perturbative expansion. To fix this we need to find the function $g_a(\Lambda)$ such that different choices of Λ give the same physics and that some of those choices rescues perturbation theory.

As an example, let us a look at the corrections for the physical mass. From the Λ_0 theory with a hard cutoff we get:

$$m^2_{\text{phys}} = m^2(\Lambda_0) + \frac{\lambda(\Lambda_0)}{2}\int^{\Lambda_0} \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2 + m^2(\Lambda_0)} \tag{6.9}$$

The trick to proceed is to expand the functions $m^2(\Lambda_0)$ and $\lambda(\Lambda_0)$ in powers of $\lambda_0 = \lambda(\Lambda_0)$.

$$m^2_{\text{phys}} = m_0^2 + \delta m_0^2 + \frac{\lambda_0}{2}\int^{\Lambda_0} \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2 + m_0^2} + O(\lambda_0^2) \tag{6.10}$$

where m_0 does not depend on λ_0, and $\delta m_0^2 = O(\lambda_0)$ is the first order term in the expansion of $m(\Lambda_0)$.

This expression is still not respecting perturbation theory because the integral is still too big. So we repeat the calculation with a smaller cutoff Λ to get:

$$m^2_{\text{phys}} = m'^2 + \delta m'^2 + \frac{\lambda'}{2}\int^{\Lambda} \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2 + m'^2} + O(\lambda'^2) \tag{6.11}$$

and now we *match* order by order[34].

At order λ_0^0 we get $m_0 = m'$.

At order λ_0^1 we get

$$\delta m'^2 - \delta m_0^2 = \frac{\lambda_0}{2}\int_{\Lambda}^{\Lambda_0} \frac{\mathrm{d}^4k}{(2\pi^4)}\frac{1}{\boldsymbol{k}^2 + m_0^2} \tag{6.12}$$

which is precisely the result we obtained earlier when integrating out the UV modes!

Note how we could only constrain the difference between the first order corrections. This behaviour is generic, it comes from the fact we could always have absorbed part of the first order terms into the definition of m_0. This freedom is equivalent to the integration constant one would get if we were to solve (4.35) instead. Because of this, we can bundle the difference into a single quantity δm^2, called the *counterterm*. We just have to either choose $\delta m_0^2 = 0$ or $\delta m'^2 = 0$.

In some sense, when we were integrating out, we were choosing $\delta m_0^2 = 0$. The reasoning was that we had a fixed starting point and we wanted to figure out what was the value for $m^2(\Lambda)$ (or alternatively $\delta m'^2$) that would describe the same physics.

But we can also take the alternative point of view by setting $\delta m'^2 = 0$. In this way we are instead asking the question: "How can we redefine the original coupling such that the 'bigness' is removed without changing the physics". Of course these two ways of formulating the question are equivalent, but the re-framing will be useful in figuring out the extra structure we were missing earlier.

This method of computing RG is called *matching* because we are directly matching the observables at the two scales, rather than constructively integrating out from one to the other. This method has some big advantages.

Firstly, we can give one of the answers to the question we asked at the end of the last section: now we can use two different schemes at the two different scales. There is nothing stopping me from using a smoother cut-off or similar in (6.11), we just have to match the observables and move on, being careful to impose (4.35) with respect to Λ if we need to dial the cut-off to another relevant scale.

So now, by directly matching the observables at the two scales, we can not only use an arbitrary scheme but we can even use a different scheme at each scale. We can even use this to forcefully relate two different schemes by directly imposing that they give the same physical answers.

Secondly, we can actually use it to simplify the IR calculations! This may come as a surprise, it seems like we need S_Λ to even use this method, and if we have S_Λ we already have the IR observables. However, as we will see shortly, by judiciously choosing our IR

[34]The 4-point equivalent of this reasoning implies, at leading order, $\lambda' = \lambda_0$.

scheme and the value of Λ_0 we can both minimize the amount of knowledge we need of the IR and greatly simplify the UV integrals. This is rather tricky the explain in abstract, so let us a look at a few examples of how to implement this.

6.2.1 Counterterms in diagrams

So far, our understanding of the counterterm method is that we calculate observables with regulator Λ and Λ_0. Then we take the couplings to be functions of the cutoffs and we expand those functions perturbatively. For example:

$$m^2(\Lambda_0) = m_0^2 + \lambda_0 \delta^{(1)} m_0^2 + \lambda_0^2 \delta^{(2)} m_0^2 + \dots \tag{6.13}$$

Rather than expanding the couplings at the very end, we can instead expand them directly in the action[35]:

$$S_{\Lambda_0}[\phi] = S_{\mathrm{B}}[\phi] + S_{\mathrm{CT}}[\phi] \tag{6.14}$$

where

$$S_{\Lambda_0}[\phi] = \int \mathrm{d}^D x \left(\frac{1}{2} Z_\phi(\Lambda_0) (\boldsymbol{\nabla}\phi)^2 + \frac{1}{2} m(\Lambda_0)^2 \phi^2 + \frac{\lambda(\Lambda_0)}{4!} \phi^4 \right) \tag{6.15}$$

$$S_{\mathrm{B}}[\phi] = \int \mathrm{d}^D x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi)^2 + \frac{1}{2} m_0^2 \phi^2 + \frac{\lambda_0}{4!} \phi^4 \right) \tag{6.16}$$

$$S_{\mathrm{CT}}[\phi] = \int \mathrm{d}^D x \left(\frac{1}{2} \delta Z_\phi \, (\boldsymbol{\nabla}\phi)^2 + \frac{1}{2} \, \delta m^2 \, \phi^2 + \frac{\delta\lambda}{4!} \phi^4 \right) \tag{6.17}$$

Where we call S_{B} the *bare* action and S_{CT} the *counterterm* action.

The assumption implicit in this splitting is that the bare action are merely integration constants[36], and the counterterms hold all of the cutoff/regulator dependence.

Additionally, the whole point is that the counterterms contain powers of the coupling which controls perturbation theory (in our case λ). This means they will always contribute at one order higher than its original coupling. So for example, tree level diagrams with $\delta^{(1)} m^2$ are at the same order as 1-loop diagrams with no counterterms. Diagrams with 2-loops without counterterms will come at the same order as 1-loop diagrams with $\delta^{(1)} m^2$ and tree-level diagrams with $\delta^{(2)} m^2$.

In diagrams, the counterterm action would yield the following additional Feynman rules:

INSERT DIAGRAMS WITH COUNTERTERMS

[35] We are taking the $\delta m'^2 = 0$ point of view and therefore defining $\delta m_0^2 = \delta m^2$ with similar relations for the other couplings.

[36] As we saw in the previous section, these integration constants will generically just be the values for the couplings in the IR action S_Λ. This fact has lead to some confusion as to whether the bare action is the IR action or the UV. But it is neither, despite sharing the same couplings as the IR action, its cutoff is still the UV cutoff Λ_0, it is just the part of the UV action that we have arbitrarily separated to isolate the dependence on the cutoff and organise perturbation theory.

After we have computed the Feynman diagrams using S_{Λ_0} up the order required we fix the counterterms by matching with our desired IR result. Let us now look at a few examples.

6.2.2 On-shell scheme

This is a scheme for the IR observables. Rather than defining a full S_Λ, we instead just fix the value of all our correlators at a prescribed external momenta configuration. One advantage of this scheme is that everything is clearly measurable, and therefore the final answer has manifest physical significance.

For the 2-point function the requirements are slightly different[37]. We define m'^2 to be the equal to the physical mass m^2_{phys}, that is, the pole in the 2-point function. Further, to fix δZ_ϕ we demand that the residue of this pole is still unity (as it is in the free theory).

$$m_0^2 - \Gamma^{(2)}_{\text{loops}}(-m^2_{\text{phys}}) = m^2_{\text{phys}} \tag{6.18}$$

$$\left.\frac{\partial \Gamma^{(2)}_{\text{loops}}(p^2)}{\partial p^2}\right|_{p^2=-m^2_{\text{phys}}} = 0 \tag{6.19}$$

Note that we are evaluating p^2 on a negative value, this of course will only make sense in Lorentzian signature, so we have to Wick rotate back to implement this scheme.

In total we have for the 1PI 2-point function:

DRAW THE DIAGRAMS!

Plugging things in, we can satisfy these requirements by choosing,

$$\delta Z_\phi = 0 \tag{6.20}$$

$$\delta m^2 = m^2_{\text{phys}} - m_0^2 - \frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m_0^2 \log\left(1 + \frac{\Lambda_0^2}{m_0^2}\right)\right) \tag{6.21}$$

As advertised, we can choose $m_0^2 = m^2_{\text{phys}} = m^2$ to get the requirement $\Gamma^{(2)}_{\text{loops}}(-m^2) = 0$ and

$$\delta m^2 = -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m^2 \log\left(1 + \frac{\Lambda_0^2}{m^2}\right)\right) \tag{6.22}$$

In this case the counterterm is tuned to exactly cancel the 1-loop contributions, leaving a fairly trivial result. This is an accident of the simplicity of this theory. Namely in the fact that the 1-loop diagram did not depend on the external momentum. Generically, we would expect a non-trivial field strength renormalisation, and that the counterterm would only cancel the loop contributions at that particular energy.

[37] This is a consequence of the fact we have a non-perturbative understanding of the 2-point function via the Kallen-Lehman spectral representation.

For the 4-point function we define the coupling in terms of the (connected) 4-point function at a particular value of the external momenta.

$$\lambda' = \lambda_{\text{phys}} = G^{(4)}_{\text{con}}(s^\star, t^\star, u^\star) \tag{6.23}$$

where we have introduced the usual Mandelstam variables:

$$s = -p_{12}^2, \qquad t = -p_{14}^2, \qquad u = -p_{13}^2 \tag{6.24}$$

where we define $p_{12} = p_1 + p_2$. Note that we can perfectly define those variables in Euclidean signature, the only catch is that they will always be negative. They will only have the usual physical significance once we Wick rotate back to Lorentzian signature.

The Feynman diagrams that contribute in this case are:

DRAW THE DIAGRAMS!

Now we take advantage of the freedom of choice of Λ_0 and we compute the loop integral in the $\Lambda_0 \to \infty$ limit before imposing this condition. To do that, we'll need the following trick due to Feynman (which you can check by explicit computation of the RHS):

$$\frac{1}{AB} = \int_0^1 \mathrm{d}x \, \frac{1}{(xA + (1-x)B)^2} \tag{6.25}$$

which allows us to combine two propagators, e.g.

$$\begin{aligned}
\frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_{12})^2 + m^2} &= \int_0^1 \mathrm{d}x \, \frac{1}{\Big[x\Big((\boldsymbol{k} + \boldsymbol{p}_{12})^2 + m^2\Big) + (1-x)(\boldsymbol{k}^2 + m^2)\Big]^2} = \\
&= \int_0^1 \mathrm{d}x \, \frac{1}{\big[\boldsymbol{k}^2 + m^2 + x\big(2\boldsymbol{k}\cdot\boldsymbol{p}_{12} + \boldsymbol{p}_{12}^2\big)\big]^2} = \\
&= \int_0^1 \mathrm{d}x \, \frac{1}{\big[(\boldsymbol{k} + x\boldsymbol{p}_{12})^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\big]^2}
\end{aligned} \tag{6.26}$$

We then define $\boldsymbol{l} = \boldsymbol{k} + x\boldsymbol{p}_{12}$ to write

$$\begin{aligned}
&\int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_2)^2 + m^2} = \\
&= \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{1}{\big[\boldsymbol{l}^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\big]^2}
\end{aligned} \tag{6.27}$$

Remember that we should integrate over $\{\boldsymbol{k} |\ |\boldsymbol{k}| < \Lambda_0 \wedge |\boldsymbol{k} + \boldsymbol{p}_{12}| < \Lambda_0\}$, which would be a very complicated x dependent region in terms of $\boldsymbol{l}$. However, for $|\boldsymbol{p}_{12}| \ll \Lambda_0$ these two regions are almost identical. Hence, we will approximate and integrate $\boldsymbol{l}$ over the region $|\boldsymbol{l}| < \Lambda_0$. Therefore we write

$$\int_0^1 \mathrm{d}x \int^{\Lambda_0} \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{1}{\left[\boldsymbol{l}^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} =$$

$$= \int_0^1 \mathrm{d}x \, \frac{\mathrm{Vol}(S^3)}{(2\pi)^4} \int_0^{\Lambda_0} \mathrm{d}l \, \frac{l^3}{\left[l^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} =$$

$$= \int_0^1 \mathrm{d}x \, \frac{1}{16\pi^2} \left(\log\left(\frac{\Lambda_0^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2}{m^2 + x(1-x)\boldsymbol{p}_{12}^2} \right) + \frac{m^2 + x(1-x)\boldsymbol{p}_{12}^2}{\Lambda_0^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2} - 1 \right) \tag{6.28}$$

Doing similar manipulations with the other three integrals, and keeping only terms to leading order in $\frac{1}{\Lambda_0}$ since those are the only ones we have reliably computed in our approximation, we get for the full 4-point function:

$$G_{\mathrm{con}}^{(4)}(s,t,u) = \lambda_0 + \delta\lambda - \frac{\lambda_0^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)s} \right) + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)t} \right) + \right.$$
$$\left. + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)u} \right) - 3 \right) + O\left(\frac{1}{\Lambda_0} \right) \tag{6.29}$$

We can now impose our on-shell condition[38]

$$\delta\lambda = \lambda_{\mathrm{phys}} - \lambda_0 + \frac{\lambda_0^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)s^\star} \right) + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)t^\star} \right) + \right.$$
$$\left. + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)u^\star} \right) - 3 \right) \tag{6.30}$$

The full correlator, after plugging in the chosen counterterm is:

$$G_{\mathrm{con}}^{(4)}(s,t,u) = \lambda - \frac{\lambda^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{m^2 - x(1-x)s^\star}{m^2 - x(1-x)s} \right) + \log\left(\frac{m^2 - x(1-x)t^\star}{m^2 - x(1-x)t} \right) + \right.$$
$$\left. + \log\left(\frac{m^2 - x(1-x)u^\star}{m^2 - x(1-x)u} \right) \right) \tag{6.31}$$

There are a few key points to make regarding this method. Firstly, all parameters have a very clear physical meaning, they're the values of correlators at a chosen momentum configuration. They're not just coefficients that you have to fit to.

Secondly, we can now see that there is a greater physical significance to the RG flow. Up to this point the subtraction scale Λ was purely theoretical. It had some physical significance as the highest momentum scale we can reach but now the physical interpretation is a lot more direct.

[38] Once again we can choose $\lambda_{\mathrm{phys}} = \lambda_0 = \lambda$

The subtraction scale $s^\star$ is an actual experimental scale[39]. The β-function in this scheme is capturing the dependence of the correlator with the energy. In this scheme, because everything is defined in terms of experimental scales it is easier to see that we can actually *measure* β-functions. The RG flows we derived last time go from a neat theoretical construction to capturing the energy dependence of physical observables.

On related note, our calculations are valid as long as λ is small and s, t, u are sufficiently close to $s^\star$, $t^\star$, $u^\star$. This suggests the following trick. Take the difference between s, t, u and $s^\star$, $t^\star$, $u^\star$ to be infinitesimal and compute the derivative, *i.e.* the β-function. Then, rather than using our expansion from the loop integrals, we solve that differential equation. In this way we can actually resum some contributions from higher loops into a compact form, thereby slightly improving our perturbative expansion.

For example, in the massless limit our calculations yield:

$$G^{(4)}_{\text{con}}(s,t,u) = \lambda - \frac{\lambda^2}{32\pi^2}\left(\log\left(\frac{s^\star}{s}\right) + \log\left(\frac{t^\star}{t}\right) + \log\left(\frac{u^\star}{u}\right)\right) \tag{6.32}$$

In the centre of mass frame, we have

$$s = 4\mu^2, \quad t = -2\mu^2(1-\cos\theta), \quad u = -2\mu^2(1+\cos\theta) \tag{6.33}$$

where μ is the centre of mass energy.

Imposing that

$$\mu\frac{\mathrm{d}G^{(4)}_{\text{con}}}{\mathrm{d}\mu} = \left(\mu\frac{\partial}{\partial\mu} + \beta_\lambda\frac{\partial}{\partial\lambda}\right)G^{(4)}_{\text{con}} = 0 \tag{6.34}$$

we find, to leading order in λ:

$$\beta_\lambda = \frac{3\lambda^2}{16\pi^2} \tag{6.35}$$

which agrees with our previous calculation!

But now we can actually solve this differential equation directly to get

$$\lambda(\mu) = \frac{16\pi^2\lambda^\star}{16\pi^2 + 3\lambda^\star\log\left(\frac{\mu^\star}{\mu}\right)} \tag{6.36}$$

If we expand this equation out we get exactly the result we had before. However, this now encodes even more information, we have resummed part of the perturbative expansion[40]. This procedure to improve perturbation theory by solve the β-functions is called *RG improvement*. Of course this idea is only useful if we can actually solve the differential equations.

To hammer that we have gone beyond ordinary perturbation theory note that at our solution for $\lambda(\mu)$ blows up at the scale

[39] The remaining Mandelstam variables $t^\star$ and $u^\star$ encode the scattering angle rather than the energy scale of the experiment.

[40] In particular we have resummed what are called the *leading logs* you can find more details in CITE

$$\mu^\infty = \mu^\star \mathrm{e}^{-\frac{16\pi^2}{3\lambda^\star}} \tag{6.37}$$

which cannot be expanded in powers of $\lambda^\star$! This singularity is called a *Landau pole* and it signals the breakdown of perturbation theory at exponentially higher energies[41]. We can use μ^∞ to write our solution as

$$\lambda(\mu) = \frac{16\pi^2}{3\log\left(\frac{\mu^\infty}{\mu}\right)} \tag{6.38}$$

The fact we have traded a dimensionless parameter, λ, for a dimensionful parameter, μ^∞, is called *dimensional transmutation* and was a puzzle for a long time. For us it is no puzzle at all, we know that λ is always defined at a particular energy, we actually traded the pair $(\lambda^\star, \mu^\star)$ for the pair (∞, μ^∞). The puzzle is only there if we insist on parametrising RG flows in terms of very singular points.

Finally, we can see how the counterterm method allowed us to use the $\Lambda_0 \to \infty$ limit to actually compute the loop integrals at non-zero external momenta and end up with a final answer which is perfectly physical. The price we have to pay is that the counterterms are infinite, meaning that the UV action S_{Λ_0} has infinite coefficients. It is only through our more careful analysis that we can see how those two infinities balance each other to give a finite and physical final answer.

6.2.3 Dimensional regularisation and minimal subtraction

All the calculations we have done so far have been very well motivated from a physical standpoint and made sense mathematically[42]. However, that pedagogy has a price, computations were often quite laborious. In the last section we managed to simplify things slightly by resorting to the $\Lambda_0 \to \infty$ limit, but for more complicated theories this would still not have been enough. Additionally, a hard cutoff may break certain symmetries we would like to preserve, like the gauge invariance of QED. To this end we introduce a new kind of regulator: *dimensional regularisation*, which is on a weaker mathematical footing but is much more convenient for practical calculations.

The idea comes from noticing that the loop integrals only diverge for certain dimensions. For example, in $D = 4$ our 4-point loop integral diverges, but in $D = 3$ it would instead scale like $\sim \frac{1}{\Lambda_0}$ and converge. Therefore, if we keep the dimension free rather than setting it to 4, we must be able to compute the integral. For example, for the 2-point function the integral is of the form

$$\int \mathrm{d}k \, \frac{k^{D-1}}{k^2 + m^2} \tag{6.39}$$

by using the change of variables

$$u = \frac{m^2}{k^2 + m^2} \tag{6.40}$$

[41] Just to get an idea of high ludicrously high these energies are, QED also has a Landau pole, but it's at 10^{286} GeV! Our perturbation theory is safe for a long time.

[42] Except of course the $D = 4 - \epsilon$ expansion

we get

$$\int_0^\infty \mathrm{d}k\, \frac{k^{D-1}}{k^2+m^2} = \frac{m^{D-2}}{2}\int_0^1 \mathrm{d}u\, u^{-\frac{D}{2}}(1-u)^{\frac{D}{2}-1} = \frac{m^{D-2}}{2}\Gamma\left(1-\frac{D}{2}\right)\Gamma\left(\frac{D}{2}\right) \tag{6.41}$$

where in the last step we used the Euler-Beta function

$$B(s,t) = \int_0^1 \mathrm{d}u\, u^{s-1}(1-u)^{t-1} = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)} \tag{6.42}$$

Remembering that the Γ-function has poles on all non positive integers we can see that our divergence as $\Lambda_0 \to \infty$ has been replaced by a pole at $D = 4$.

And now comes the trick. We first compute all integrals without specifying D. All our divergences will become poles for certain integer values of D. We then analytically continue to complex D which then renders our integrals finite. Our analogue of the cutoff Λ_0 becomes $\frac{1}{D-4}$. We end by imposing whatever subtraction scheme we wish in the IR (*e.g.* the on-shell scheme we discussed in the last section) and the counterterms will cancel the terms that blow up as $D \to 4$, rendering our final answer finite when $D = 4$, and arguably physically significant.

Let us see how this works in practice.

The master formula to compute integrals in dim-reg is the following (Exercise?):

$$\int_0^\infty \mathrm{d}k\, \frac{k^{a-1}}{(k^2+M^2)^b} = M^{a-2b}\frac{\Gamma\left(b-\frac{a}{2}\right)\Gamma\left(\frac{a}{2}\right)}{2\Gamma(b)} \tag{6.43}$$

Together with the formula for the volume of a sphere in generic dimensions[43]:

$$\mathrm{Vol}(S^{D-1}) = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \tag{6.44}$$

For the 2-point function, not much changes. We could do the integral anyways, but the outcome now is:

$$\begin{aligned}\Gamma^{(2)}_{\text{1-loop}} &= -\frac{\lambda_0}{2}\int \frac{\mathrm{d}^D k}{(2\pi)^4}\frac{1}{\boldsymbol{k}^2+m_0^2} = -\frac{\lambda_0}{2}\frac{\mathrm{Vol}(S^{D-1})}{(2\pi)^D}\int_0^\infty \mathrm{d}k\, \frac{k^{D-1}}{k^2+m_0^2} = \\ &= -\frac{\lambda_0 m_0^{D-2}}{2(4\pi)^{\frac{D}{2}}}\Gamma\left(1-\frac{D}{2}\right)\end{aligned} \tag{6.45}$$

Imposing on-shell scheme still just cancels this contribution completely, *i.e.*

$$\delta m^2 = -\frac{\lambda_0 m_0^{D-2}}{2(4\pi)^{\frac{D}{2}}}\Gamma\left(1-\frac{D}{2}\right) \tag{6.46}$$

which in fact does have a pole at $D = 4$.

For the 4-point function we gain a bit more. We first repeat the Feynman trick by defining $\boldsymbol{l} = \boldsymbol{k} + x\boldsymbol{p}_{12}$ to write

[43] Here is a quick proof: $(\sqrt{\pi})^d = \prod_{i=1}^n \int \mathrm{d}x_i\, \mathrm{e}^{-x_i^2} = \mathrm{Vol}(S^{d-1})\int_0^\infty \mathrm{d}r\, r^{d-1}\mathrm{e}^{-r^2} = \frac{1}{2}\mathrm{Vol}(S^{d-1})\Gamma(d/2)$

$$\int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_2)^2 + m^2} =$$
$$= \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^D l}{(2\pi)^D} \frac{1}{\left[\boldsymbol{l}^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} \tag{6.47}$$

but now the integration region doesn't change! We were integrating over $\boldsymbol{k} \in \mathbb{R}^D$ and therefore we need to integrate over $\boldsymbol{l} \in \mathbb{R}^D$. This is the big simplification that dim-reg provides.

Additionally, we can identify our final integral with our master formula for

$$M^2 = m^2 + x(1-x)\boldsymbol{p}_{12}^2, \quad a = D, \quad b = 2 \tag{6.48}$$

which means we obtain

$$\int_0^1 \mathrm{d}x \left(m^2 + x(1-x)\boldsymbol{p}_{12}^2\right)^{\frac{D}{2}-2} \frac{\Gamma\left(2 - \frac{D}{2}\right)}{(4\pi)^{\frac{D}{2}}} \tag{6.49}$$

and putting everything together we get

$$G_{\mathrm{con}}^{(4)}(s,t,u) = \lambda_0 + \delta\lambda - \frac{\lambda_0^2}{2} \frac{\Gamma\left(2 - \frac{D}{2}\right)}{(4\pi)^{\frac{D}{2}}} \int_0^1 \mathrm{d}x \left(\left(m^2 + x(1-x)s\right)^{\frac{D}{2}-2} + \right.$$
$$\left. + \left(m^2 + x(1-x)t\right)^{\frac{D}{2}-2} + \left(m^2 + x(1-x)u\right)^{\frac{D}{2}-2} \right) \tag{6.50}$$

However, note that in generic D, the coupling constant λ_0 will not be dimensionless. Hence, we shall add an auxiliary mass scale μ to be able to write everything in terms of the dimensionless coupling g_0:

$$G_{\mathrm{con}}^{(4)}(s,t,u) = g_0 + \delta g - \frac{g_0^2 \mu^{4-D}}{2} \frac{\Gamma\left(2 - \frac{D}{2}\right)}{(4\pi)^{\frac{D}{2}}} \int_0^1 \mathrm{d}x \left(\left(m^2 + x(1-x)s\right)^{\frac{D}{2}-2} + \right.$$
$$\left. + \left(m^2 + x(1-x)t\right)^{\frac{D}{2}-2} + \left(m^2 + x(1-x)u\right)^{\frac{D}{2}-2} \right) \tag{6.51}$$

Where we have rescaled $G^{(4)}$ to make it dimensionless as well. Note how we did not have to take any limits or approximations, to compute the integrals.

We can now impose the on-shell condition again.

$$\delta g = \frac{g^2 \mu^{4-D}}{2} \frac{\Gamma\left(2 - \frac{D}{2}\right)}{(4\pi)^{\frac{D}{2}}} \int_0^1 \mathrm{d}x \left(\left(m^2 + x(1-x)s^\star\right)^{\frac{D}{2}-2} + \right.$$

$$+\left(m^2+x(1-x)t^\star\right)^{\frac{D}{2}-2}+\left(m^2+x(1-x)u^\star\right)^{\frac{D}{2}-2}\Bigg) \tag{6.52}$$

The full correlator, after plugging in the chosen counterterm is:

$$\begin{aligned}G^{(4)}_{\text{con}}(s,t,u) =& g-\frac{g^2\mu^{4-D}}{2}\frac{\Gamma\left(2-\frac{D}{2}\right)}{(4\pi)^{\frac{D}{2}}}\int_0^1 \mathrm{d}x\Bigg(\left(m^2+x(1-x)s\right)^{\frac{D}{2}-2}-\left(m^2+x(1-x)s^\star\right)^{\frac{D}{2}-2}+\\ &+\left(m^2+x(1-x)t\right)^{\frac{D}{2}-2}-\left(m^2+x(1-x)t^\star\right)^{\frac{D}{2}-2}+\\ &+\left(m^2+x(1-x)u\right)^{\frac{D}{2}-2}-\left(m^2+x(1-x)u^\star\right)^{\frac{D}{2}-2}\Bigg)\end{aligned} \tag{6.53}$$

Now we would like to take the limit $D\to 4$, however, this seems like it would be a bit tricky. The integrands are clearly 0 when $D=4$ and the Γ function has a pole when $D=4$, presumably they cancel, but we would like the final expression.

To do that we need an extra fact about the Γ function. Its asymptotic expansion near 0 is given by

$$\Gamma(\epsilon)\sim\frac{1}{\epsilon}-\gamma+O(\epsilon)\quad\text{as }\epsilon\to 0^+ \tag{6.54}$$

where $\gamma\approx 0.577216$ is the Euler-Mascheroni constant.

This expansion implies

$$\Gamma\left(2-\frac{D}{2}\right)\sim\frac{2}{4-D}-\gamma+O(D-4) \tag{6.55}$$

Additionally we can write

$$A^{\frac{D-4}{2}}-B^{\frac{D-4}{2}}=\frac{D-4}{2}\log\left(\frac{A}{B}\right)+O((D-4)^2) \tag{6.56}$$

The term proportional to $D-4$ in (6.56) cancels the $D=4$ pole in (6.55) yielding a finite answer as $D\to 4$ as expected. Taking this limit and using our expansions as above yields the same result as we had previously (noting that when $D=4$, $g=\lambda$):

$$\begin{aligned}G^{(4)}_{\text{con}}(s,t,u)=\lambda-\frac{\lambda^2}{32\pi^2}\int_0^1\mathrm{d}x\Bigg(&\log\left(\frac{m^2-x(1-x)s^\star}{m^2-x(1-x)s}\right)+\log\left(\frac{m^2-x(1-x)t^\star}{m^2-x(1-x)t}\right)+\\ &+\log\left(\frac{m^2-x(1-x)u^\star}{m^2-x(1-x)u}\right)\Bigg)\end{aligned} \tag{6.57}$$

Our auxiliary μ dropped out completely in favour of our physical scales $s^\star$, $t^\star$, and $u^\star$ and we recovered the previous result exactly. After all, we are using the same subtraction scheme, just a different regulator. Given we are removing the regulator at the end, it stands to reason that the final answer should be the same. An explicit example of two different regulating schemes yielding the same physical answer.

For a theory with only scalars it is somewhat subjective whether or not dimensional regularisation actually made our calculations easier. However, our regulator never got rid of momenta above a certain value, which means we can preserve far more symmetries of our original theory. This will be crucial for gauge theories for which a hard cutoff breaks gauge invariance (which we can never do), but dim-reg doesn't.

6.2.4 Minimal subtraction

There is one final scheme worthy of note. One whose physical significance is quite opaque but calculationally it is very easy to implement, which in turn means it is very widely used: minimal subtraction.

Its definition is incredibly simple. Instead of fixing the value of the correlator at a certain momenta, we just subtract the poles of our Γ functions by hand. For example:

$$\delta m^2 = \frac{\lambda m^2}{16\pi^2(D-4)} \tag{6.58}$$

$$\delta g = \frac{3g^2}{16\pi^2(D-4)} \tag{6.59}$$

It is also customary to remove the constants like γ or $\log(4\pi)$ which generically appear. This scheme is called *modified minimal subtraction* or $\overline{\text{MS}}$.

$$\delta m^2 = \frac{\lambda m^2}{32\pi^2}\left(\frac{2}{D-4} - \gamma + \log(4\pi)\right) \tag{6.60}$$

$$\delta g = \frac{3g^2}{32\pi^2}\left(\frac{2}{D-4} - \gamma + \log(4\pi)\right) \tag{6.61}$$

For the 2-point function we get:

$$\Gamma^{(2)}_{\text{1-loop}} = \frac{\lambda m^2}{32\pi^2}\left(1 + \log\left(\frac{\mu^2}{m^2}\right)\right) \tag{6.62}$$

Now the 1-loop result isn't completely canceled!

For the 4-point function we get:

$$G^{(4)}_{\text{con}}(s,t,u) = \lambda + \frac{\lambda^2}{32\pi^2}\int_0^1 \mathrm{d}x \left(\log\left(\frac{m^2 - x(1-x)s}{\mu^2}\right) + \log\left(\frac{m^2 - x(1-x)t}{\mu^2}\right) + \right.$$
$$\left. + \log\left(\frac{m^2 - x(1-x)u}{\mu^2}\right)\right) \tag{6.63}$$

In both cases, the auxiliary constant μ remains in the final answer. It now serves the role of the subtraction scale (previously Λ or $s^\star$). However its physical interpretation is quite unclear. We can still define β-functions wrt μ as previously and the validity of perturbation theory will still demand that μ^2 is close to the experimental scale s.

6.3 The BPHZ theorem: renormalisable vs non-renormalisable

Notice how, with the schemes we introduced, it was perfectly enough to renormalise λ and m and all infinities were taken care of. No need to introduce a counterterm for ϕ^6. In fact, even with cut-offs, looking at (6.4) and (6.5) we can see that the $\Lambda_0 \to \infty$ limit is independent of the external momenta. As $\boldsymbol{k}$ gets big we can neglect the addition of $\boldsymbol{p}_{ij}$. So, it stands to reason that we only need to cancel the $\boldsymbol{p}_i = 0$ divergence and that will be enough to render the correlator finite for all other values of external momenta, this would mean choosing:

$$\delta\lambda = \frac{3\lambda_0^2}{2}\int_{\Lambda}^{\Lambda_0} \frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{(\boldsymbol{k}^2+m_0^2)} \tag{6.64}$$

But note the difference with what we obtained by just integrating out completely from Λ_0 to Λ. Now it won't be true that the new 4-point correlator will look like (6.4) but with Λ instead of Λ_0 everywhere. This is only true at $\boldsymbol{p}_i = 0$. In order to recover the previous result we would need to introduce counterterms for higher derivative 4-point correlators such as $\phi^2\nabla^2\phi^2$. Which is perfectly in line with what we found in Chapter 5, when integrating out we do find corrections for all other couplings, they just happened to be irrelevant. What we see here is that correcting all of this is not entirely necessary, just using (6.64) is enough to render everything finite. The behaviour we found with on-shell and dim reg seems to be generic.

However, this may be a bit too hasty. We have just checked at 1-loop and for a couple of correlators. Maybe at higher loop level and/or for higher order correlators something goes wrong and we actually do need all of the counterterms to make everything consistent. It turns out that it depends. For some theories (including ϕ^4 in 4 dimensions) we can indeed only worry about a finite set of counterterms, for others (like ϕ^4 in 6 dimensions) we actually have to include an infinite number of counterterms to make everything consistent. Checking this is true at all loop orders is quite involved, but it can be summarised in the following theorem due to Bogoliubov, Parasiuk, Hepp, and Zimmermann:

Theorem (BPHZ)**.** *Quantum Field Theories can be classified into three categories:*

- ***Super-Renormalisable****: All couplings have positive mass dimension*
- ***Renormalisable****: All couplings have either positive or vanishing mass dimension*
- ***Non-Renormalisable****: At least one coupling has negative mass dimension*

These three categories will have different types of divergences and therefore their behaviour is different:

- ***Super-Renormalisable****: There are only a finite number of divergent diagrams and therefore with a finite number of counterterms we can absorb all divergences*
- ***Renormalisable****: There are an infinite number of divergent diagrams but only a finite number of divergent amplitudes. We can introduce a finite number of counterterms that can be tuned order by order to absorb all divergences. The counterterms at a given order will exactly cancel the divergent sub-diagrams at higher orders*

- ***Non-Renormalisable**: There are an infinite number of divergent amplitudes, and therefore we need an infinite number of counterterms to absorb all divergences*

And there we have it. For ϕ^4 theory in 4-dimensions we only have relevant and marginal operators so we can deal with a finite number of counterterms: δZ_ϕ, δm^2 and $\delta\lambda$. We will need to adjust them order by order but we will never need to add more. We could perfectly add more, as we did when we integrated out, but, for practical calculations, we don't have to. But this is very special to 4-dimensions, in higher dimensions the quartic coupling will be irrelevant and at sufficiently high loop/correlator order we will need to add more and more counterterms. Once we put even a single irrelevant operator we will eventually have to deal with all of them.

At first sight it can seem like theories with irrelevant operators are doomed. Given we have to add all of them it seems like we have to do an infinite number of experiments before we can calculate anything. However, an attentive reader will notice that at a given loop order we will actually only need a finite number of counterterms to fix a given amplitude. At 1-loop, no higher order coupling will be able to contribute to the 2-point function for example. So maybe there is some way to write it down such that we can do calculations, we just know that at higher precision there will be some run-away phenomenon and things will get very complicated very fast. But even then that seems optimistic. After all, we just assumed every coupling was small, but we don't necessarily have any constraints on how big g_4 is in relation to g_6, perhaps we are so unlucky that $g_6^2 \sim g_4$ and 2-loops of g_6 will come at the same order as one loop of g_4. This would mean we would have to fix all couplings before we even know which ones are important and how to write down a perturbative expansion. Very damning indeed.

These problems seemed so unsurmountable that for many decades physicists took 'non-renormalisable' as synonymous with 'inconsistent' and struggled very hard to only write down renormalisable theories. It turns out this is wrong. There is indeed a hierarchy among the different couplings and not only are non-renormalisable theories perfectly consistent they are actually far more useful than renormalisable theories.

The hint comes from the original Wilsonian integrating out. In that case there seems to be some hierarchy between the couplings, after all, we called couplings with negative mass dimension 'irrelevant' because if we flow down to the IR they would shrink very rapidly. Therefore, if, for reasons so far unknown, at a very high scale Λ_0 all couplings are comparable in size, then, at the lower scale Λ the flow under RG would dictate that irrelevant couplings will shrink, and not only that but there will be a hierarchy and higher order couplings will be smaller and smaller.

But this may seem a bit dubious. We have been working under the assumption that Λ_0 is somewhat unphysical. It was crutch, something we needed to introduce to make sense of a continuous path integral. You may think the language of infinities is not that silly in the end. If we really manage to parametrise the flow at low energies, and remove all the bigness/infinities, we can perfectly take the limit $\Lambda_0 \to \infty$. Everything is finite and even small. No problems there. If the theory is renormalisable at least.

But remember the Landau pole from (6.37), it seems like we cannot go to infinitely

high energies because our couplings blow up at the finite energy μ^∞! So it seems like this theory actually doesn't make sense as a continuum theory. At some point it becomes non-perturbative and we would have to change our approach. The fact we could take $\Lambda_0 \to \infty$ was a mere artefact of perturbation theory. And if the theory doesn't have a good continuum limit (or at least not one we know how to describe) then maybe there is an actual honest-to-goodness *physical* cut-off, Λ_{EFT} which in the previous case would be (at least close to) μ^∞. We then say we are dealing with an 'Effective Field Theory' (EFT for short) which is only valid below a certain energy scale.

We have managed to argue for the existence of at least *some* physical cut-offs, but we still don't have a hierarchy. Who knows the relative size of all the couplings at this very complicated scale Λ_{EFT} when our perturbative description breaks down. In general, this is all we have got. Theory space is very wild, all sorts of things might happen.

However, it seems like it is actually very generic that all couplings are $O(1)$ at the EFT scale. There is no formal proof that such a behaviour is generic, but we can take as an assumption. The EFT assumption is that all couplings are $O(1)$ at some scale Λ_{EFT} which then implies that for experimental scales much lower than that, irrelevant couplings are dampened according to their mass dimension, creating a hierarchy among them.

Therefore, in order to define an EFT you need an action, a renormalisation scheme, and an assumed hierarchy. This last one is commonly called a *power counting scheme*.

The upshot is that this extra assumption (which is absolutely crucial to deal with irrelevant couplings) will gives more information which we can then exploit. As soon as we can measure the first irrelevant coupling, we can then estimate Λ_{EFT}, and therefore the scale at which all couplings become $O(1)$ and perturbation theory breaks down. At that scale some new physics must arise! Either, some extra degree of freedom, or some strongly coupled phenomena. And these techniques work! Historically we had predictions for the mass of the W and Higgs bosons way before we actually managed to produce them. By measuring the irrelevant couplings left behind we could be certain new physics was on the horizon.

Non-renormalisable theories, when understood as an EFT, are actually *more* predictive than renormalisable ones! Even though technically we could already do this for marginally irrelevant couplings. The difference is that when we measure an irrelevant coupling Λ_{EFT} will be polynomially higher than our experimental scales rather than exponentially higher.

And there we have it. From regulating nasty infinities away. To carving out flows in the space of theories to make that regulator consistent. Then using said flows to recover perturbation theory. And finally unveiling the structure of effective field theories which can use irrelevant couplings to predict new particles before we even produce them. These are the fruits were can bear from understanding the interplay of different scales in quantum field theory.

Part III

Symmetries in Quantum Field Theory

7 Global symmetries and their anomalies

Except for a brief foray into the fermionic path integral, so far in our course we have focused on scalar fields. This was important to make sure we really understand the subtleties behind the path integral and renormalisation. But it is time we move forward and bring in the second big idea in physics: symmetry. In particular, in this chapter we will focus on global symmetries, what are their consequences for field theories and why they may not survive quantisation.

7.1 Noether's theorem in classical field theory

Let us begin by reviewing Noether's theory in classical field theory. Consider a generic infinitesimal transformation of an arbitrary field ϕ^a:

$$\delta\phi^a = \epsilon^b X^a_b(\phi, \partial\phi) \tag{7.1}$$

where ϵ^b is a constant and small parameter and a is a generic index.

For a classical field, this transformation is a symmetry if the action is invariant, that is:

$$\delta S = 0 \tag{7.2}$$

which means that the Lagrangian density $\mathcal{L}$ transforms by a total derivative:

$$\delta\mathcal{L} = \partial_\mu\Big(\epsilon^b K^\mu_b\Big) \tag{7.3}$$

But we can just calculate $\delta\mathcal{L}$ to obtain

$$\begin{aligned}\delta\mathcal{L} =& \frac{\partial\mathcal{L}}{\partial\phi^a}\delta\phi^a + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^a)}\partial_\mu\phi^a = \\ =& \left(\frac{\partial\mathcal{L}}{\partial\phi^a} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^a)}\right)\delta\phi^a + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^a)}\delta\phi^a\right)\end{aligned} \tag{7.4}$$

The first term vanishes if the equations of motion are satisfied. Equating the second term of (7.4) with the RHS of (7.3) we get:

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^a)}\delta\phi^a - \epsilon^b K^\mu_b\right) = 0 \tag{7.5}$$

defining the current:

$$\mathcal{J}^\mu_a = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^b)}X^a_b(\phi, \partial\phi) - K^\mu_a \tag{7.6}$$

For ϵ^a constant, we can then use (7.5) to conclude

$$\partial_\mu\mathcal{J}^\mu_a = 0 \tag{7.7}$$

so that the current is conserved when the equations of motion are satisfied.

There is also another way to derive the existence of a conserved current, which will be useful later. When the equations of motion are satisfied, the action must be invariant under *any* transformation of the field, in fact, that was how we derived the equations of motion. Therefore, even for a non-constant ϵ^a it must still be true that $\delta S = 0$ for the transformation defined in (7.1) for non-constant ϵ^a. However, when the equations of motion are not satisfied, then δS is only zero for constant ϵ^a. Therefore, the transformation of the action without imposing equations of motion must be proportional to the derivative of ϵ^a:

$$\delta S = \int \mathrm{d}^D x \, (\partial_\mu \epsilon^a) \mathcal{J}'^{\mu}_{\ a} \tag{7.8}$$

integrating by parts we get

$$\delta S = -\int \mathrm{d}^D x \, \epsilon^a \partial_\mu \mathcal{J}'^{\mu}_{\ a} \tag{7.9}$$

Imposing the equations of motion allows us to set $\delta S = 0$ for arbitrary ϵ^a which implies

$$\partial_\mu \mathcal{J}'^{\mu}_{\ a} = 0 \tag{7.10}$$

and therefore there exists a conserved current.

This second reasoning did not explicitly construct the conserved current, it merely showed that a conserved current must exist. The conclusion is nonetheless the same, the existence of a continuous symmetry implies the existence of a conserved current[44].

7.2 Ward identities in quantum field theory

In the quantum theory we have to perform path integrals. Therefore, in order for a transformation to be a symmetry it does not suffice that the action is invariant under the symmetry. The integration measure must also be invariant:

$$\mathcal{D}\phi = \mathcal{D}\phi' \tag{7.11}$$

In particular, to define the measure we need to regulate our theory. Therefore, in order for the symmetry to have any hopes of surviving quantisation, the regulator we use must not break the symmetries we want to preserve. When there is *no* regulator that preserves the symmetry we say that that symmetry is *anomalous*. In this section we will assume the measure is invariant and deduce the analogues of Noether's theory for the quantum theory. In the next section we will delve into a particularly interesting example of an anomalous symmetry.

The first consequence is the simplest. Let us consider the expectation value of the product of a number of operators $\mathcal{O}(\phi)$ which are functions of the fields ϕ^a. Under the transformation $\phi \to \phi'$ they transform like $\mathcal{O}(\phi) \to \mathcal{O}(\phi')$. We can then write

$$\int \mathcal{D}\phi \ \mathrm{e}^{\mathrm{i}S[\phi]} \mathcal{O}_1(\phi(x_1)) \dots \mathcal{O}_n(\phi(x_n)) = \int \mathcal{D}\phi' \ \mathrm{e}^{\mathrm{i}S[\phi']} \mathcal{O}_1(\phi'(x_1)) \dots \mathcal{O}_n(\phi'(x_n))$$

[44] An interesting point is that the conserved current is not unique. We could for example multiply $\mathcal{J}$ by any constant and it would still be conserved. Or similarly add a function f^μ_a satisfying $\partial_\mu f^\mu_a$. This means there is no guarantee that $\mathcal{J}$ and $\mathcal{J}'$ match, only that they must both be conserved.

$$= \int \mathcal{D}\phi \; \mathrm{e}^{\mathrm{i}S[\phi]} \mathcal{O}_1(\phi'(x_1)) \dots \mathcal{O}_n(\phi'(x_n)) \tag{7.12}$$

Where in the first line we just did a change of variables in the integral, which cannot change anything, and in going to the second line we used the invariance of the action and the measure under the symmetry transformation $\phi \to \phi'$.

All in all, we can write[45]

$$\langle \mathcal{O}_1(\phi(x_1)) \dots \mathcal{O}_n(\phi(x_n)) \rangle = \langle \mathcal{O}_1(\phi'(x_1)) \dots \mathcal{O}_n(\phi'(x_n)) \rangle \tag{7.13}$$

This is known as *Ward-Takahashi identities*. It states that any correlator must also be invariant under symmetry transformations. For example, the scalar theories we have been considering are invariant under $\phi \to -\phi$ which then means

$$\langle \mathcal{O}_1(\phi(x_1)) \dots \mathcal{O}_n(\phi(x_n)) \rangle = \langle \mathcal{O}_1(-\phi(x_1)) \dots \mathcal{O}_n(-\phi(x_n)) \rangle \tag{7.14}$$

and therefore any correlator which has an odd power of the field must vanish.

Similarly translation invariance $\phi(x) \to \phi(x-a)$ implies

$$\langle \mathcal{O}_1(\phi(x_1)) \dots \mathcal{O}_n(\phi(x_n)) \rangle = \langle \mathcal{O}_1(\phi(x_1+a)) \dots \mathcal{O}_n(\phi(x_n+a)) \rangle \tag{7.15}$$

so correlators can only depend on the differences $x_i - x_j$.

These results may seem quite intuitive but remember that it is not sufficient for the action to be invariant under the symmetry transformation. The measure also needs to be invariant.

For continuous symmetries we have slightly more structure. Including sources we write

$$Z[J] = \int \mathcal{D}\phi \; \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x \, J\phi\right) \to \int \mathcal{D}\phi' \; \exp\left(\mathrm{i}S[\phi'] + \mathrm{i}\int \mathrm{d}^D x \, J\phi'\right) = Z[J] \tag{7.16}$$

Note how the change in field variable cannot affect $Z[J]$ as it is merely a change of variables in the path integral.

Using (7.1) with non-constant ϵ, (7.9), and (7.11) we get[46]

$$\begin{aligned} Z[J] &= \int \mathcal{D}\phi \; \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x \, J\phi\right) \exp\left(-\mathrm{i}\int \mathrm{d}^D x \, \epsilon^a \left(\partial_\mu \mathcal{J}_a^\mu - J_b X_a^b\right)\right) \approx \\ &\approx \int \mathcal{D}\phi \; \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x \, J\phi\right) \left(1 - \mathrm{i}\int \mathrm{d}^D x \, \epsilon^a \left(\partial_\mu \mathcal{J}_a^\mu - J_b X_a^b\right)\right) \end{aligned} \tag{7.17}$$

Note that the first term in the last line (the "1") is just the original $Z[J]$! We therefore conclude that

$$\int \mathcal{D}\phi \; \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x \, J\phi\right) \int \mathrm{d}^D x \, \epsilon^a \left(\partial_\mu \mathcal{J}_a^\mu - J_b X_a^b\right)) = 0 \tag{7.18}$$

[45] We are being lazy and using $\langle \dots \rangle$ to mean $\langle \Omega | \, \mathcal{T}(\dots) \, | \Omega \rangle$ as usual.

[46] We are dropping the distinction between $\mathcal{J}$ and $\mathcal{J}'$ we had in the previous section.

The previous expression is true for all J so we can now play a bit with the source to find different consequences for $\mathcal{J}$. First we can just set $J = 0$, in which case we find out that

$$\langle \partial_\mu \mathcal{J}_a^\mu \rangle = 0 \tag{7.19}$$

which is the usual quantum mechanical logic that expressions that hold when we impose the classical equations of motion are true under an expectation value.

We can also differentiate first with respect to J before setting it to zero to obtain

$$\partial_\mu \left\langle \mathcal{J}_a^\mu(x)\phi^b(y) \right\rangle = \delta^{(D)}(x-y) \left\langle X_a^b \right\rangle \tag{7.20}$$

Repeating this process we arrive at

$$\partial_\mu \left\langle \mathcal{J}_a^\mu(x)\phi^{b_1}(y_1)\dots\phi^{b_n}(y_n) \right\rangle = 0 \quad \forall x \neq y_i \tag{7.21}$$

if x coincides with any of the y^i there will be some term proportional to X on the RHS, as was the case in (7.20).

These expressions are also collectively known as *Ward identities*, We can sum them up as an operator equation

$$\partial_\mu \mathcal{J}_a^\mu = 0 \tag{7.22}$$

which should be understood as saying that the operator $\partial_\mu \mathcal{J}_a^\mu$ vanishes when inserted in any correlator as long as it's position does not coincide with any other fields in the correlator. Which is precisely the statement we had before.

7.2.1 Slavnov-Taylor identities

There is one further consequence of continuous global symmetries which is worthy of note: the transformation of the quantum effective action. Let us take ϵ to be constant from the start. Then both the action and the measure are just invariant and we have, instead of (7.17)

$$\begin{aligned} Z[J] &= \int \mathcal{D}\phi \ \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x\, J\phi\right) \exp\left(\mathrm{i}\int \mathrm{d}^D x\, \epsilon^a J_b X_a^b\right) \approx \\ &\approx \int \mathcal{D}\phi \ \exp\left(\mathrm{i}S[\phi] + \mathrm{i}\int \mathrm{d}^D x\, J\phi\right)\left(1 + \mathrm{i}\int \mathrm{d}^D x\, \epsilon^a J_b X_a^b\right) \end{aligned} \tag{7.23}$$

but as before, the first term is just $Z[J]$ once more and we thefore conclude

$$\int \mathrm{d}^D x\, \epsilon^a J_b \left\langle X_a^b(\phi) \right\rangle_J = 0 \tag{7.24}$$

where now the expectation value is done in the presence of the source J.

But remember (2.61), we can write J in terms of the derivative of the quantum effective action Γ:

$$\left.\frac{\delta\Gamma[\varphi]}{\delta\varphi^a(t)}\right|_{\varphi=\varphi_J} = -J_a(t)$$

and therefore

$$\int \mathrm{d}^D x\, \epsilon^a \left.\frac{\delta\Gamma[\varphi]}{\delta\varphi^a(t)}\right|_{\varphi=\varphi_J} \left\langle X_a^b(\phi)\right\rangle_J = 0 \tag{7.25}$$

But this expression in turn means that the quantum effective action is invariant under the transformation

$$\varphi^a \to \varphi^a + \epsilon^b \left\langle X_b^a(\phi)\right\rangle_{J_\varphi} \tag{7.26}$$

which involves expectation values of the classical symmetry transformation.

These are the *Slavnov-Taylor identities* and they detail what is the form of the symmetry which actually survives quantisation.

In the case of a linear transformation, that is, one in which X takes the form

$$X_b^a = c_b^a + \int \mathrm{d}^D x\, d_b^a \phi^b \tag{7.27}$$

we have

$$\langle X_b^a\rangle_{J_\varphi} = c_b^a + \int \mathrm{d}^D x\, d_b^a \left\langle \phi^b\right\rangle_{J_\varphi} \tag{7.28}$$

remembering (2.64) we have

$$\left\langle \phi^b\right\rangle_{J_\varphi} = \left.\frac{\delta W[J]}{\delta J_b(t)}\right|_{J=J_\varphi} = \varphi^b(t) \tag{7.29}$$

and therefore

$$\langle X_b^a\rangle_{J_\varphi} = c_b^a + \int \mathrm{d}^D x\, d_b^a \varphi^b = X_b^a(\varphi) \tag{7.30}$$

So, for linear symmetries, the quantum effective action is invariant under the *same* transformation as the classical action.

These considerations are especially relevant for renormalisation. If the quantum effective action is invariant under a different symmetry than the classical action then there may be some 1PI diagrams which are divergent (and non-zero) for which we cannot find a counterterm that could cancel that divergence. For linear symmetries however we do not have to worry. For non-linear symmetries we would need a more careful analysis to understand whether we can get rid of all infinities without breaking any of the original symmetries.

7.3 The chiral anomaly

There are many examples of anomalies in quantum field theory. In this section we will delve into the simplest case, which happens to also be quite important for particle physics. It concerns the theory of a single massless Dirac fermion coupled to a $U(1)$ gauge field in 4 dimensions:

$$S = \int \mathrm{d}^4 x \left(\bar\psi \mathrm{i}\not{D}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right) \tag{7.31}$$

where

$$D_\mu = \partial_\mu - \mathrm{i}gA_\mu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{7.32}$$

are the usual gauge covariant derivative and field strength tensor.

This theory as two global symmetries. First there is a vector like phase rotation

$$\psi \to \mathrm{e}^{-\mathrm{i}\alpha}\psi \tag{7.33}$$

$$\bar{\psi} \to \mathrm{e}^{\mathrm{i}\alpha}\bar{\psi} \tag{7.34}$$

for real α. The corresponding Noether current is (Exercise!)

$$\mathcal{J}_V^\mu = \bar{\psi}\gamma^\mu\psi \tag{7.35}$$

Recall that we can decompose a Dirac fermion in terms of two Weyl fermions:

$$\psi = \begin{pmatrix} \chi \\ \xi^\dagger \end{pmatrix} \tag{7.36}$$

In terms of these Weyl fermions, the above symmetry rotates them with opposite phases:

$$\chi \to \mathrm{e}^{-\mathrm{i}\alpha}\chi \tag{7.37}$$

$$\xi \to \mathrm{e}^{\mathrm{i}\alpha}\xi \tag{7.38}$$

The second symmetry instead rotates the Weyl fermions with the *same* phase:

$$\chi \to \mathrm{e}^{+\mathrm{i}\alpha}\chi \tag{7.39}$$

$$\xi \to \mathrm{e}^{+\mathrm{i}\alpha}\xi \tag{7.40}$$

We can write this symmetry in terms of the original Dirac fermion by use of the 5^{th} γ matrix:

$$\psi \to \mathrm{e}^{-\mathrm{i}\alpha\gamma^5}\psi \tag{7.41}$$

$$\bar{\psi} \to \bar{\psi}\mathrm{e}^{-\mathrm{i}\alpha\gamma^5} \tag{7.42}$$

where $\gamma^5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$. This symmetry is only present because the fermion is massless and it is a consequence of the fact that $\left\{\gamma^5, \gamma^\mu\right\} = 0$ (a mass term would not have the γ^μ present in the derivative term which is needed to flip the sign of the exponential). The corresponding Noether current (Exercise!)

$$\mathcal{J}_A^\mu = \bar{\psi}\gamma^\mu\gamma^5\psi \tag{7.43}$$

this is an axial vector which motivatives calling this the *axial symmetry*.

From the previous discussion you might expect that both currents are conserved. However, the axial symmetry is actually anomalous and its divergence is instead given by:

$$\partial_\mu \mathcal{J}_A^\mu = -\frac{g^2}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \tag{7.44}$$

This expression was first derived by computing certain Feynman diagrams. Nevertheless, and quite remarkably, it is actually *exact* in g, there are no higher order corrections. We shall then take a different approach which does not rely on any perturbative expansion. This is known as the *Fujikawa method*.

We shall consider the path integral over the Dirac field only, leaving the gauge field integration to do later:

$$Z[A] = \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\, \mathrm{e}^{\mathrm{i}S} \tag{7.45}$$

This integral is actually quadratic! Remembering (3.21) we can schematically write the path integral as a functional determinant:

$$Z[A] \sim \det\left(\mathrm{i}\not{D}\right) \tag{7.46}$$

First note that there is a dependence on A_μ inside of the covariant derivative $\not{D}$ so this result is more than just a normalisation factor. Further, to actually compute this we would need to choose our regulator. we will come back to this point later.

Taking inspiration from the classical calculation we shall consider a position dependent axial rotation:

$$\psi(x) \to \mathrm{e}^{-\mathrm{i}\alpha(x)\gamma^5}\psi(x) \tag{7.47}$$

$$\bar{\psi}(x) \to \bar{\psi}(x)\mathrm{e}^{-\mathrm{i}\alpha(x)\gamma^5} \tag{7.48}$$

This is still merely a change in the integration variables in (7.45) so $Z[A]$ cannot change. It is nevertheless straightforward to show that the action changes as:

$$S \to S + \int \mathrm{d}^4x\, \mathcal{J}^\mu_A(x)\partial_\mu\alpha(x) = S - \int \mathrm{d}^4x\, \alpha(x)\partial_\mu\mathcal{J}^\mu_A(x) \tag{7.49}$$

If the measure were invariant this would imply $\partial_\mu\mathcal{J}^\mu_A(x) = 0$ as in the previous section. Let us then calculate said Jacobian. Taking functional derivatives with respect to ψ and $\bar{\psi}$ in (7.47) and (7.48) respectively we find the following Jacobian:

$$J(x,y) = \delta^{(4)}(x-y)\mathrm{e}^{-\mathrm{i}\alpha(x)\gamma^5} \tag{7.50}$$

Because the path integral is over fermionic variables we get a Jacobian factor of $(\det J)^{-1}$ rather than $\det J$. Moreover for the axial rotation the Jacobian from ψ and $\bar{\psi}$ are the same[47], all in all:

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} \to (\det J)^{-2}\mathcal{D}\psi\mathcal{D}\bar{\psi} \tag{7.51}$$

Using $\log\det J = \mathrm{Tr}\log J$ we then write

[47] For the vector-like rotation they would have different signs in the exponential and cancel out, therefore the vector-like symmetry is not anomalous.

$$(\det J)^{-2} = \exp\left[2\mathrm{i}\int \mathrm{d}^4x\,\alpha(x)\,\mathrm{Tr}\,\delta^{(4)}(x-x)\gamma^5\right] \tag{7.52}$$

Looking at this expression there are two quick arguments for what the result should be:

- The first is that it should be zero because $\mathrm{Tr}\,\gamma^5 = 0$
- The second is that it should be infinite because we have a $\delta^{(4)}(0)$

It turns out that both of these are a bit too quick. Let us first focus on the infinity. We want to somehow regulate this infinity, to do that, it is instructive to write the delta function in Fourier space:

$$\delta^{(4)}(x-y) = \int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{\mathrm{i}k\cdot(x-y)} \tag{7.53}$$

which for $x = y$ is just equal to the volume of the spacetime, which is infinite, so we need to introduce a regulator as usual.

However, we do not want to impose some hard cutoff on k, this is because it would not be gauge invariant. Instead we want to impose some condition involving the gauge covariant derivative D_μ. We therefore choose the following regulator

$$\delta^{(4)}(x-y) = \lim_{\Lambda\to\infty}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{(\mathrm{i}\not{D}_x)^2/\Lambda^2}\mathrm{e}^{\mathrm{i}k\cdot(x-y)} \tag{7.54}$$

which also has the added bonus that it relates nicely to (7.46). The subscript x represents that the partial derivatives act on x rather than y and also that the gauge field inside it is a function of x: $A_\mu(x)$.

Now the integral has been regulated and is finite (at least for finite Λ), moreover the matrix structure in $\not{D}$ will also prevent the first argument above. All we have left to do is to compute.

We first use the fact that (Check yourself!)

$$f(\partial_x)\mathrm{e}^{\mathrm{i}kx} = \mathrm{e}^{\mathrm{i}kx}f(\partial + \mathrm{i}k) \tag{7.55}$$

to write

$$\delta^{(4)}(x-y) = \lim_{\Lambda\to\infty}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{\mathrm{i}k\cdot(x-y)}\mathrm{e}^{(\mathrm{i}\not{D}_x - \not{k})^2/\Lambda^2} \tag{7.56}$$

where any partial derivatives acting on the far right gives zero.

We now use our γ-matrix technology to evaluate $(\mathrm{i}\not{D}_x - \not{k})^2$. The main identities we shall need are[48] (Exercise!):

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

[48] The second item should be viewed as a definition for $S^{\mu\nu}$

- $[\gamma^\mu, \gamma^\nu] = 4\mathrm{i}S^{\mu\nu}$
- $\not{k}^2 = -k^2$
- $[D_\mu, D_\nu] = -\mathrm{i}gF_{\mu\nu}$

We can then write

$$\begin{aligned}(\mathrm{i}\not{D}_x - \not{k})^2 =& \not{k}^2 - \mathrm{i}\{\not{k}, \not{D}\} - \not{D}^2 = \\ =& -k^2 - \mathrm{i}\{\gamma^\mu, \gamma^\nu\}k_\mu D_\nu - \gamma^\mu\gamma^\nu D_\mu D_\nu \\ =& -k^2 + 2\mathrm{i}(k\cdot D) + D^2 + 2\mathrm{i}S^{\mu\nu}F_{\mu\nu}\end{aligned} \tag{7.57}$$

where in going to the last line we also used $2\gamma^\mu\gamma^\nu = \{\gamma^\nu, \gamma^\nu\} + [\gamma^\mu, \gamma^\nu]$.

Plugging this into the Dirac-δ and rescaling k by Λ we get:

$$\delta^{(4)}(x-y) = \lim_{\Lambda\to\infty} \Lambda^4 \int \frac{\mathrm{d}^4k}{(2\pi)^4} \mathrm{e}^{\mathrm{i}\Lambda k\cdot(x-y)} \mathrm{e}^{-k^2} \mathrm{e}^{2\mathrm{i}(k\cdot D)/\Lambda + D^2/\Lambda^2 + 2\mathrm{i}S^{\mu\nu}F_{\mu\nu}/\Lambda^2} \tag{7.58}$$

and therefore

$$\mathrm{Tr}\,\delta^{(4)}(x-x)\gamma^5 = \lim_{\Lambda\to\infty} \Lambda^4 \int \frac{\mathrm{d}^4k}{(2\pi)^4} \mathrm{e}^{-k^2}\,\mathrm{Tr}\Big\{\mathrm{e}^{2\mathrm{i}(k\cdot D)/\Lambda + D^2/\Lambda^2 + 2\mathrm{i}S^{\mu\nu}F_{\mu\nu}/\Lambda^2}\gamma^5\Big\} \tag{7.59}$$

Now we expand the exponential. We need exactly four inverse powers of Λ to cancel the overall Λ^4 and yield a finite limit. Further, in order for the trace to not vanish we need at least 4 other γ-matrices (and an even number of them). Altogether the only term in the exponential that contributes is:

$$\frac{1}{2}\frac{(gS^{\mu\nu}F_{\mu\nu})^2}{\Lambda^4} \tag{7.60}$$

and so

$$\mathrm{Tr}\,\delta^{(4)}(x-x)\gamma^5 = \int \frac{\mathrm{d}^4k}{(2\pi)^4} \mathrm{e}^{-k^2} F_{\mu\nu}F_{\rho\sigma}\,\mathrm{Tr}\big\{S^{\mu\nu}S^{\rho\sigma}\gamma^5\big\} \tag{7.61}$$

The spin trace can be evaluated to be

$$\mathrm{Tr}\big\{S^{\mu\nu}S^{\rho\sigma}\gamma^5\big\} = -\frac{1}{4}\,\mathrm{Tr}\big\{\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^5\big\} = \mathrm{i}\epsilon^{\mu\nu\rho\sigma} \tag{7.62}$$

To evaluate the integral over k we Wick rotate and then back:

$$\int \frac{\mathrm{d}^4k}{(2\pi)^4}\mathrm{e}^{-k^2} = \mathrm{i}\int \frac{\mathrm{d}^4k_E}{(2\pi)^4}\mathrm{e}^{-k_E^2} = \mathrm{i}\bigg(\int \frac{\mathrm{d}x}{2\pi}\mathrm{e}^{-x^2}\bigg)^4 = \frac{\mathrm{i}}{16\pi^2} \tag{7.63}$$

Plugging these two results in, we get

$$\operatorname{Tr}\delta^{(4)}(x-x)\gamma^5 = -\frac{g^2}{32\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \tag{7.64}$$

which is neither zero nor infinite!

Now we plug this back into (7.52)

$$(\det J)^{-2} = \exp\left[-\frac{\mathrm{i}g^2}{16\pi^2}\int \mathrm{d}^4x\,\alpha(x)\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\right] \tag{7.65}$$

Including this transformation of the measure and the transformation of the action given by (7.49) we get

$$Z[A] = \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\ \mathrm{e}^{\mathrm{i}S}\exp\left[-\mathrm{i}\int \mathrm{d}^4x\,\alpha(x)\left(\frac{g^2}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} + \partial_\mu\mathcal{J}^\mu_A(x)\right)\right] \tag{7.66}$$

Given that $Z[A]$ cannot change we must conclude that

$$\partial_\mu\mathcal{J}^\mu_A = -\frac{g^2}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \tag{7.67}$$

Note how we never had to assume any expansion in powers of g, this result is *exact*[49].

Beyond being a neat example of an anomaly this case is particularly interesting for a few reasons. Firstly, it is experimentally relevant. In the real world axial charge is not conserved, and the precise relationship we derived is crucial for calculating the amplitude for certain processes which do not conserve axial charge.

Additionally it has consequences for the consistency of chiral gauge theories. A gauge symmetry represents a redundancy in our description. It cannot be broken or we loose mathematical consistency. However, if we had tried to gauge the axial symmetry, the anomaly would mean that the gauge symmetry did not survive quantisation and everything we did was inconsistent. In order to have chiral gauge theories we need to very carefully choose the matter content to cancel the anomaly and ensure that the gauge symmetry is not anomalous. That cancellation is so precise that for the case of the Standard Model it uniquely fixes the charges for all elementary particles! It is anomaly cancellation which explains the quantisation of charge that we observe in Nature.

[49]The most remarkable thing is that actually the 1-loop calculation would yield the same result, all higher looks actually vanish. The reason for this has to do with the Atiyah-Singer index theorem which relates certain topological quantities to geometrical quantities, which will ensure that the integral of the LHS of the anomaly is actually an integer. This in turn will mean that it cannot receive a "small" correction like higher order loops could have given.

8 Abelian gauge theories

In the previous chapter we focused on global symmetries. Namely what are the consequences for quantum field theories when there is in fact a symmetry, and how a symmetry of the classical theory may not survive quantisation. However, there are other kinds of symmetries. In this chapter we will focus on gauge symmetries, which should be better thought as a redundancy in our description.

We will focus on theories such as QED whose gauge group is $U(1)$. These theories are quite simple, but still have enough structure to serve as a jumping off point to the more complicated non-Abelian gauge theories such as Yang-Mills.

8.1 The QED action

When deriving the chiral anomaly we wrote down the QED action and discussed its global symmetries. However, there is an additional symmetry of that action we did not discuss. Let us reproduce the action for clarity:

$$S = \int \mathrm{d}^4x \left(\bar{\psi} \mathrm{i} \not{D} \psi - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \right) \tag{8.1}$$

where

$$D_\mu = \partial_\mu - \mathrm{i}gA_\mu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{8.2}$$

If we just promote the non-anomalous, vector-like symmetry parameter to a function of spacetime:

$$\psi(x) \to \mathrm{e}^{-\mathrm{i}\alpha(x)} \psi(x) \tag{8.3}$$

$$\bar{\psi}(x) \to \mathrm{e}^{\mathrm{i}\alpha(x)} \bar{\psi}(x) \tag{8.4}$$

it is no longer a symmetry (in fact it is a trick to derive an expression for $\mathcal{J}_V$), however, we can shift A_μ by

$$A_\mu(x) \to A_\mu(x) - \frac{1}{g} \partial_\mu \alpha(x) \tag{8.5}$$

and now the *combined* transformation *is* a symmetry!

To make these expressions a bit cleaner it is common to rescale A_μ by

$$A_\mu \to \frac{1}{g} A_\mu \tag{8.6}$$

so that

$$S = \int \mathrm{d}^4x \left(\bar{\psi} \mathrm{i} \not{D} \psi - \frac{1}{4g^2} F^{\mu\nu} F_{\mu\nu} \right) \tag{8.7}$$

where

$$D_\mu = \partial_\mu - \mathrm{i}A_\mu, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{8.8}$$

and the symmetry is

$$\psi(x) \to \mathrm{e}^{-\mathrm{i}\alpha(x)} \psi(x) \tag{8.9}$$

$$\bar{\psi}(x) \to \mathrm{e}^{\mathrm{i}\alpha(x)}\bar{\psi}(x) \tag{8.10}$$

$$A_\mu(x) \to A_\mu(x) - \partial_\mu\alpha(x) \tag{8.11}$$

At first sight you might be unfazed by this symmetry. However, if you dig a bit deeper you will see some peculiarities. First of all, there is no Noether current associated to this symmetry, whatever method you use to compute it would yield zero. Further, if you are serious about making A_μ dynamical you should invert the kinetic term to compute its propagator:

$$\frac{1}{4g^2}F^{\mu\nu}F_{\mu\nu} = \frac{1}{2g^2}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu) \tag{8.12}$$

which is not invertible!

In fact, the position dependent symmetry above ensures the kinetic term is annihilated by $A_\mu(x) = \partial^\mu\alpha(x)$ for *any* $\alpha(x)$. How can we even proceed?

The trick is that these kinds of symmetries must be understood differently. Rather than a usual symmetry which relates the dynamics of two separate quantum states. We should understand this as a *redundancy* in our description. States related by the above symmetry are not related, they are *the same state*. There is no physical distinction between two different gauges, they are equivalent descriptions of the same state. This is why we call it a *gauge* symmetry or sometimes even gauge redundancy.

Let us call the space of all vector fields $A_\mu(x)$: $\mathcal{A}$. The above paragraph means we do not want to perform the path integral over $\mathcal{A}$:

$$\int_{\mathcal{A}} \mathcal{D}A\ \mathrm{e}^{\mathrm{i}S} \tag{8.13}$$

Instead, we want to perform the integration over vector fields modulo gauge transformations G, that is:

$$\int_{\mathcal{A}/G} \mathcal{D}A\ \mathrm{e}^{\mathrm{i}S} \tag{8.14}$$

However, it is most convenient to instead pick a gauge and integrate over vector fields with satisfy our gauge conditions:

$$\int_{\mathcal{A}|_{\text{gauge condition}}} \mathcal{D}A\ \mathrm{e}^{\mathrm{i}S} \tag{8.15}$$

The question that remains is how do we ensure that we do not pick up a Jacobian factor from (8.14) to (8.15). There are two methods: the first means going back to the Hamiltonian formulation, doing the full quantisation with constraints and deriving a path integral from scratch. This is the most rigorous method because it also ensures we are picking the correct starting point. This is not the method we will use, you can see the details in Weinberg's books.

Instead we will take a different route which focuses more on the path integral. It is much quicker but it has the downside that we have to pick a starting point. From this method alone we do not know which of (8.14) or (8.15) is the appropriately normalised one.

8.2 Fadeev-Popov determinant: a 0-dim toy model

To gain some intuition behind gauge fixing let us take a look at a similar method for finite-dimensional integrals (or zero dimensional quantum field theories). Consider the following integral:

$$\int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, \mathrm{e}^{-S(x,y)} \tag{8.16}$$

where the function $S : \mathbb{R}^2 \to \mathbb{R}^2$ is in fact just a function of the radial variable $r = \sqrt{x^2 + y^2}$, *i.e.* $S(x, y) = h(r)$. The rotational $SO(2)$ symmetry is the analogue of the gauge symmetry, and $S(x, y)$ is the analogue of the action.

For this case it is obvious to just change to polar coordinates and write:

$$\int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, \mathrm{e}^{-S(x,y)} = 2\pi \int_0^\infty \mathrm{d}r\, r\ \mathrm{e}^{-h(r)} \tag{8.17}$$

We want to do the analogue of this computation for the field theory. However, it is not at all clear what the right set of variables (in our case r) would be, let alone the correct integration measure in that space ($\mathrm{d}r\, r$ in our case). So let us write this computation in a more complicated way which we could generalise to QFT.

With the insight that we often want to fix a gauge, let us consider a curve C in the (x, y)-plane. This curve should intersect circles of constant radius exactly once.

INSERT FIGURE

This curved can be defined by a condition $f(x, y) = 0$, the statement that the curve intersects all circles of constant radius exactly once translates to the following conditions on the function f on $\mathbb{R}^2$:

- For any point $\boldsymbol{x} \in \mathbb{R}^2$ there exists a rotation $R \in SO(2)$ such that $f(R\boldsymbol{x}) = 0$. This means that for any point we can rotate it (at constant radius) until we hit the curve, *i.e.* $f = 0$.
- The curve only intersecting each orbit once means $f(R\boldsymbol{x}) = f(\boldsymbol{x})$ if and only if R is the identity.

The last condition in particular means that f is *not* rotationally invariant. This function is the analogue of the gauge condition, and the the curve C the analogue of the gauge slice.

Now we would like to write an integral on the lines of

$$\int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, \delta(f(x, y))\ \mathrm{e}^{-S(x,y)} \tag{8.18}$$

but this integral depends on the choice of f. For example, just rescaling $f \to cf$ for some number $c \in \mathbb{R}$ the delta function changes by $\delta(f) \to \frac{1}{|c|}\delta(f)$.

To take care of this dependence we define the following object, the *Fadeev-Popov determinant* (the name will make more sense shortly):

$$\Delta_{\mathrm{FP}}[f] = \left.\frac{\partial f(R_\theta \boldsymbol{x})}{\partial \theta}\right|_{\theta=0} \tag{8.19}$$

where $R_\theta \in SO(2)$ rotates the point $\boldsymbol{x} \in \mathbb{R}^2$ by an angle θ.

Let us then consider the integral:

$$\int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, |\Delta_{\mathrm{FP}}[f]| \delta(f(x,y))\ \mathrm{e}^{-S(x,y)} \tag{8.20}$$

The FP determinant scales inversely to the Dirac-δ so it is clearly invariant under rescalings of f. Similarly, the derivative in its definition is purely with respect to θ so actually we can also rescale f by a function of r only, $c(r)$ and the integral would still be invariant. In fact, we can even show in full generality that the previous integral is completely independent of the choice of the gauge-fixing function f.

Let us consider two gauge slices C_1 and C_2 with respective defining functions f_1 and f_2. Because they intersect each circle of constant radius only once, we can always rotate the curves into one another, as long as we allow ourselves the freedom to rotate by a different amount at different radii. For the functions, that means we can write:

$$f_1(\boldsymbol{x}) = c(r) f_2(R(r)\boldsymbol{x}) \tag{8.21}$$

where $c(r)$ is some function of the radius and $R(r) \in SO(2)$ is the radius-dependent rotation that connects the two curves.

Let $\boldsymbol{x}' = R(r)\boldsymbol{x}$, then, the invariance under rescalings by functions of the radius $c(r)$ means

$$|\Delta_{\mathrm{FP}}[f_1(\boldsymbol{x})]| \delta(f_1(\boldsymbol{x})) = \big|\Delta_{\mathrm{FP}}[f_2(\boldsymbol{x}')]\big| \delta(f_2(\boldsymbol{x}')) \tag{8.22}$$

The action is rotationally invariant, therefore $S(\boldsymbol{x}) = S(\boldsymbol{x}')$. We only have to show that the measure is also invariant, *i.e.* that

$$\mathrm{d}x\, \mathrm{d}y = \mathrm{d}x'\, \mathrm{d}y' \tag{8.23}$$

To do so we just compute the Jacobian explicitely. The transformation is:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta(r)) & \sin(\theta(r)) \\ -\sin(\theta(r)) & \cos(\alpha(r)) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{8.24}$$

and it is a straightforward computation to check the measure is invariant.

Therefore we have an integral which is the same for any choice of f, we now just have to show that there is one choice of f that makes it agree with the original integral. Let us take $f(\boldsymbol{x}) = y$. Then $f(R_\theta \boldsymbol{x}) = -x \sin\theta + y \cos\theta$ and thus

$$\Delta_{\mathrm{FP}}[f] = \frac{\partial}{\partial\theta}(-x\sin\theta + y\cos\theta)\bigg|_{\theta=0} = -x \tag{8.25}$$

and therefore

$$\int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, |\Delta_{\mathrm{FP}}[f]| \delta(f(x,y))\ \mathrm{e}^{-S(x,y)} = \int_{\mathbb{R}^2} \mathrm{d}x\, \mathrm{d}y\, |x| \delta(y)\ \mathrm{e}^{-S(x,y)} =$$

$$= \int_{\mathbb{R}} \mathrm{d}x\, |x|\ \mathrm{e}^{-h(|x|)} = 2\int_0^\infty \mathrm{d}r\, r\ \mathrm{e}^{-h(r)} \tag{8.26}$$

Which agrees with our original integral! The only difference is the normalisation, we do not have the factor of 2π but that is a good thing because for QFT it would be infinite anyway. There is an extra factor of 2 because our curve actually hits every circle of constant radius twice rather than once. This last behaviour is called a *Gribov ambiguity* and occurs rather generically in gauge theories, however, we shall not concern ourselves too much with this and shove under the normalisation rug (which admitedly is rather large at this point).

In this last step we can actually see the weakness with this method, we need something to compare it with to know it gives the right answer. The only way to answer that for field theories is to go through the painful constrained Hamiltonian system quantisation.

In summary we have shown that our original integral is given by

$$\frac{1}{\mathrm{Vol}(SO(2))}\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mathrm{e}^{-S(x,y)} = \int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,|\Delta_{\mathrm{FP}}[f]|\delta(f(x,y))\ \mathrm{e}^{-S(x,y)} \tag{8.27}$$

If the gauge invariance has n parameters we have to instead insert a factor of

$$|\Delta_{\mathrm{FP}}[f]| \prod_{a=1}^{n} \delta(f_a(\boldsymbol{x})) \tag{8.28}$$

where

$$\Delta_{\mathrm{FP}} = \det\left(\frac{\partial f_a(g_\theta \boldsymbol{x})}{\partial \theta^b}\right)\bigg|_{\theta^b=0} \tag{8.29}$$

where g_θ is the gauge transformation corresponding to θ^b, such that $g_0 = 1$. This motivates calling it a determinant.

8.3 Ghosts and the BRST symmetry

Now let us apply the ideas of the previous section to QED. We shall choose the Lorenz gauge, for which, the gauge fixing functional is:

$$f[A_\mu(x)] = \partial_\mu A^\mu(x) \tag{8.30}$$

and therefore the Fadeev-Popov determinant is:

$$\Delta_{\mathrm{FP}}[f] = \det\left(\frac{\delta f[A_\mu(x) - \partial_\mu \alpha(x)]}{\delta \alpha(y)}\right) = \det\Big(-\partial_\mu \partial^\mu \delta^{(D)}(x-y)\Big) \tag{8.31}$$

where det is a functional determinant.

For the case of QED this determinant is independent of the gauge field A_μ and so it will only contribute as a normalisation. We will nevertheless proceed with calculating it because it will actually be important to understand some of the physics. Further, for non-Abelian gauge theories (and for curved backgrounds) the determinant is not just a normalisation, so it is worth knowing how to deal with it.

To evaluate the determinant we just need to remind ourselves of (3.21). A functional determinant can be computed from a Gaussian integral with fermionic variables. We can therefore write:

$$\Delta_{\text{FP}}[f] = \int \mathcal{D}c\, \mathcal{D}\bar{c}\, \exp\left[\mathrm{i}\int \mathrm{d}^D x\, \mathrm{d}^D y\, \bar{c}(x) \frac{\delta f[A_\mu(x) - \partial_\mu \alpha(x)]}{\delta \alpha(y)} c(y)\right] \tag{8.32}$$

$$= \int \mathcal{D}c\, \mathcal{D}\bar{c}\, \exp\left[-\mathrm{i}\int \mathrm{d}^D x\, \bar{c}(x)\partial_\mu \partial^\mu c(x)\right] \tag{8.33}$$

Note how despite being fermionic the auxiliary fields c and $\bar{c}$ are scalars rather than spinors. This might naively seem to be a violation of the spin statistic theorem, however, neither of these fields will correspond to physical particles, they are merely auxiliary fields in the path integral so the spin statistics doesn't apply to them. For these reasons it is customary to call these fields *Fadeev-Popov ghosts*[50].

The Dirac-δ is also not very amenable to Feynman diagrams, so it is useful to also write it as a path integral by introducing some auxiliary fields. For this we just have to remember the Fourier representation of a Dirac-δ:

$$\delta(x) = \int \frac{\mathrm{d}k}{2\pi} \mathrm{e}^{-\mathrm{i}kx} \tag{8.34}$$

In analogy to this we write:

$$\delta[f[A_\mu(x)]] = \int \mathcal{D}h\, \exp\left[-\mathrm{i}\int \mathrm{d}^D x\, h(x) f[A_\mu(x)]\right] =$$

$$= \int \mathcal{D}h\, \exp\left[-\mathrm{i}\int \mathrm{d}^D x\, h(x)\partial_\mu A^\mu(x)\right] \tag{8.35}$$

where h is another auxiliary field (this time bosonic) sometimes called the *Nakanishi-Lautrup field*.

Putting it all together, our path integral now looks like:

$$\int \mathcal{D}A\, \mathcal{D}\psi\, \mathcal{D}\bar{\psi}\, \mathcal{D}c\, \mathcal{D}\bar{c}\, \mathcal{D}h\, \mathrm{e}^{\mathrm{i}S_{\text{total}}[A,\psi,\bar{\psi},c,\bar{c},h]} \tag{8.36}$$

where

$$S_{\text{tot}} = S_{\text{QED}} + S_{\text{ghost}} + S_{\text{gf}} \tag{8.37}$$

$$S_{\text{QED}} = \int \mathrm{d}^4 x \left(\bar{\psi}\mathrm{i}\not{D}\psi - \frac{1}{4g^2} F^{\mu\nu}F_{\mu\nu}\right) \tag{8.38}$$

$$S_{\text{ghost}} = -\int \mathrm{d}^4 x\, \bar{c}(x)\partial_\mu \partial^\mu c(x) \tag{8.39}$$

[50] Sometimes it is often to call c the ghost and $\bar{c}$ the anti-ghost because there is a ghost number conservation law if c is taken to have ghost number 1 and $\bar{c}$ is taken to have ghost number -1.

$$S_{\rm gf} = -\int \mathrm{d}^4x\, h(x)\partial_\mu A^\mu(x) \tag{8.40}$$

There are however, a few puzzles with this action. The first is of course, renormalisation. After all of this, we still need to introduce a regulator, and flowing in RG will generate every possible term consistent with the symmetries of the theory, however, the final action we have wrote down does not seem to have gauge invariance! After all, that was the whole point of the gauge fixing. But how can we ensure we do not generate gauge breaking terms such as $A_\mu A^\mu$? The second puzzle is the auxiliary fields. We want them to just be auxiliary, and we want them to not contribute to the final states, but how exactly do we impose that?

The solution to both of these puzzles is the same and it lies in the fact the final action we we wrote down has an extra symmetry, the *BRST symmetry* due to Becchi, Rouet, Stora and Tyurin.

$$\delta A_\mu(x) = -\epsilon\, \partial_\mu c(x) \tag{8.41}$$

$$\delta\psi(x) = -\mathrm{i}\epsilon\, c(x)\psi(x) \tag{8.42}$$

$$\delta\bar{\psi}(x) = \mathrm{i}\epsilon\, c(x)\bar{\psi}(x) \tag{8.43}$$

$$\delta c(x) = 0 \tag{8.44}$$

$$\delta\bar{c}(x) = \epsilon h(x) \tag{8.45}$$

$$\delta h(x) = 0 \tag{8.46}$$

where ϵ is an anti-commuting Grassmann number.

For the physical fields A_μ, ψ, and $\bar{\psi}$ this transformation is just a gauge transformation with parameter $\alpha(x) = \epsilon c(x)$ (which, as the product of two Grassmann variables is bosonic). However, the transformation of $\bar{c}$ exactly compensates the variation of the gauge fixing term[51]. Because of our gauge fixing our total action is *not* gauge invariant, however, it is *BRST* invariant. This surviving global symmetry is exactly was is needed to ensure we do not break gauge invariance by quantisation.

The most important consequence is also the easiest to spot. The transformation of all fields except for the matter fields ψ and $\bar{\psi}$ is *linear*. Remembering the discussion at the end of Section 7.2, we therefore conclude that the quantum effective action for A_μ, c, $\bar{c}$, and h is invariant under the *same* symmetry. This fact alone is enough to rule out gauge-breaking terms such as a mass term $\sim A_\mu A^\mu$ appearing at the quantum level. BRST symmetry ensures gauge invariance!

The matter fields however transform *non-linearly* under BRST. The above argument does not apply and we would expect the symmetry of the quantum effective action to be different than the above. However, because the ghost fields $\bar{c}$ and c do not couple with anything else the expectation values $\langle c\psi\rangle$ and $\langle c\bar{\psi}\rangle$ just reduce to the produce of the 1-point functions $\langle c\rangle\langle\psi\rangle$ and $\langle c\rangle\langle\bar{\psi}\rangle$. Therefore, despite the symmetry being non-linear,

[51] For non-Abelian gauge symmetries the transformation for c would need to be non-trivial.

because the ghosts are decoupled from the remaining fields one can still show that the 1PI effective action is invariant under the *same* BRST symmetry. Gauge invariance survives quantisation and renormalisation![52]

There are two final properties of the BRST transformation which are of paramount importance.

The first one, is that it is nil-potent, that is, $\delta^2\chi_a = 0$ where χ_a is any of the fields A_μ, ψ, $\bar{\psi}$, c, $\bar{c}$, or h. Which is quite straightforward to check explicitely by using the facts that $\delta c = 0$ and $c^2 = 0$, the last one being true because c is Grassmann-valued.

The second property concerns the ghost and gauge-fixing terms:

$$-\int \mathrm{d}^4x\, \delta(\bar{c}\partial_\mu A^\mu) = -\int \mathrm{d}^4x\, (\epsilon h \partial_\mu A^\mu - \bar{c}\epsilon\partial_\mu\partial^\mu c) = \epsilon(S_{\text{ghost}} + S_{\text{gf}}) \tag{8.47}$$

where in the last equality we used the fact ϵ is Grassmann-valued to write $\bar{c}\epsilon = -\epsilon\bar{c}$.

This means the terms that arose from the gauge fixing are in fact just a BRST transformation of another quantity! Terms like this are called *BRST-exact*. In analogy with exact differential forms. In a similar vein we call BRST-invariant operators, *i.e.* those that satisfy $\delta\mathcal{O} = 0$, *BRST-closed*. The nil-potency of the BRST transformation therefore implies that all BRST-exact operators are also BRST-closed.

Applying (7.13) to the BRST transformation, and remembering that $\epsilon^2 = 0$ implies there no non-linear terms, we find:

$$0 = \sum_{i=1}^{n} \left\langle (\delta\mathcal{O}_i) \prod_{j\neq i} \mathcal{O}_j \right\rangle \tag{8.48}$$

if all operators are BRST-invariant except for $\mathcal{O}_1$ we then find

$$0 = \left\langle (\delta\mathcal{O}_1) \prod_{j=2}^{n} \mathcal{O}_j \right\rangle \tag{8.49}$$

therefore the correlator of a BRST-exact operator $\delta\mathcal{O}_2$ with any number of BRST-invariant operators vanishes.

In particular, we can add any BRST-exact operator to the action without changing the outcome of any gauge/BRST-invariant correlators. In fact, this is one way to show the path integral is independent of the choice of gauge fixing operator. Looking at (8.47) we can see that choosing a different gauge fixing condition would just differ by a BRST-exact term.

Additionally, one can just add the following term[53]

[52] For non-Abelian gauge theories the ghosts are coupled to the gauge fields and BRST symmetry is hopelessly non-linear. However, using techniques far beyond the scope of our lectures (detailed in Weinberg's chapter 17), one can still show that gauge invariance survives renormalisation. The crux of the calculation is not only the original BRST symmetry, but also the vector-like *global* symmetry, which is both linear and non-anomalous! These, together with ghost number conservation massively constrain the original theory to the point where only gauge-invariant counterterms are possible.

[53] The factor of g^2 is slightly different than what is usually seen in the literature, this is due to the differing normalisations for A_μ. What we have written is appropriate for the normalisation where g^2 sits in the kinetic term rather than inside the covariant derivative.

$$\frac{\xi g^2}{2}\int \mathrm{d}^4x\,\delta(\bar{c}h) = \frac{\xi g^2}{2}\int \mathrm{d}^4x\,h^2 \tag{8.50}$$

The choice of using this term is called an R_ξ-gauge. In particular choosing $\xi = 1$ is the *Feynman gauge*.

The path integral for the Nakanishi-Lautrup field then becomes

$$\int \mathcal{D}h\ \exp\left[\mathrm{i}\int \mathrm{d}^4x\left(-h(x)\partial_\mu A^\mu(x) + \frac{\xi g^2}{2}h(x)^2\right)\right] = \mathcal{N}\exp\left[-\mathrm{i}\int \mathrm{d}^4x\,\frac{1}{2\xi g^2}(\partial_\mu A^\mu)^2\right] \tag{8.51}$$

where we recognised the fact this was a Gaussian integral and completed the square by defining

$$h = H + \frac{1}{\xi g^2}\partial_\mu A^\mu \tag{8.52}$$

Adding this to the kinetic term we get:

$$\begin{aligned}&\int \mathrm{d}^4x\left(-\frac{1}{4g^2}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\xi g^2}(\partial_\mu A^\mu)\right) = \\ &= \int \mathrm{d}^4x\,\frac{1}{2g^2}A_\mu\left(\eta^{\mu\nu}\partial_\rho\partial^\rho - \left(1-\frac{1}{\xi}\right)\partial^\mu\partial^\nu\right)A_\nu\end{aligned} \tag{8.53}$$

This is now invertible! And what is more we do not have to deal with diagrams involving h. If we choose $\xi = 1$, the kinetic term is particularly simple, it is just the wave operator $\partial^\mu\partial_\mu$. These R_ξ-gauges are particularly convenient for doing perturbative calculations due to their simple Feynman rules, but understanding where Fadeev-Popov ghosts come from and the subtleties of BRST-invariance were crucial to known *why* this gives the correct answer.