Introduction to Renormalisation

João F. Melo

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

Abstract

These are introductory lecture notes aimed at beginning graduate students covering fundamental concepts and ideas behind the renormalisation group. Our main goal is to motivate it and then explore its consequences, in the context of quantum field theory. We shall reproduce the standard results without needing any mention of cancellation of infinities. To keep the story simple and compact, we'll ignore the role of symmetry in this game. We'll also restrict to scalar fields, and not mention neither fermions nor gauge fields. Moreover, we'll only briefly mention the connection with statistical field theory.

Knowledge of quantum mechanics and special relativity is essential. Prior knowledge of quantum field theory is not required, however, this is not designed to be an introduction to QFT, so it may not be the most pedagogical way to introduce the topic. Previous familiarity with path integrals is also not assumed.

CONTENTS

ACKNOWLEDGEMENTS

Useful References

Throughout these notes I haven't been very careful about citing the original references. To ameliorate this fact here is a list of the references I found useful in preparing them:

- Michael E. Peskin, Daniel V. Schroeder: *An Introduction to Quantum Field Theory*
  This is one of the standard references in QFT. Covers all the way from the basics to renormalisation and gauge theories. It has a very clear physical insight and covers many of the technical details explicitly. With respect to RG it follows quite a different path from these lectures, more historically inclined.

- David Tong: *Statistical Field Theory*
  In many aspects this was the main inspiration for these lecture notes and was my first exposure to the wonders of renormalisation. It is all about renormalisation in the context of phase transitions and condensed matter with only a hint of the quantum world. Cannot recommend it too highly to learn the mysteries of RG.

- David Skinner: *Quantum Field Theory II*
  These lectures cover some more advanced topics of quantum field theory in the path integral perspective. They are more on the mathematical side than what we'll cover, but offer a very clear view of what is going on. The first chapter on the path integral perspective of quantum nechanics and later on the section on the local potential approximation are very heavily based on David Skinner's notes.

- Hugh Osborn: *Advanced Quantum Field Theory*
  Another very useful reference for path integrals and quantum field yheory in general, following a more classical perspective on the subject. The appendix on the derivation of Feynman diagrams is based heavily on these notes.

- Timo Weigand: *Quantum Field Theory I+II*
  These are generic lecture notes covering two courses worth of QFT from the very basics to more advanced topics. They are quite clear and with a lot of physical insight. Highly recommended as an introduction to QFT. Although I didn't take anything in particular from them they were a constant companion to clear up some misconceptions.

- A. Zee: *Quantum Field Theory in a Nutshell*
  This is a charming book, almost purely conceptual and with very little mathematical rigour. It won't teach you how to compute 2-loop diagrams but it is the place to go when the physical interpretation isn't clear.

- Yuji Iragashi, Katsumi Itoh, Hidenori Sonoda: *Realization of Symmetry in the ERG Approach to Quantum Field Theory*
  Although this is a more generic review on the role of symmetry in Exact RG its first chapter has the clearest explanation of the Exact RG programme I ever came accross. It was the basis for the last section of these lecture notes.

- Timothy J. Hollowood: *6 Lectures on QFT, RG and SUSY*
  This is probably one of the best explanations of renormalisation in the QFT perspective on the market. Highly recommend the first few lectures to get a hold of what this is all about.

1. Introduction

In school (and in undergraduate physics degrees) we're taught many facts about the world, and it's a curious thing how stuff that seem quite relevant at that level soon become obsolete and other very simple ideas only grow stronger in importance as we continue on learning. An example of the latter will be the main focus of these lecture notes, the concept of scale. But there is one key difference between the concept of scale we meet as undergraduates and the concept of scale we'll delve into during this course. The former is a useful trick, the latter is probably one the deepest discoveries ever made in theoretical physics.

In classical physics, we're taught that physics is fundamentally scale invariant. We can always rescale our variables so that they are phrased in terms of dimensionless parameters, and, as long as those are fixed, everything should remain the same. Nonetheless, there are several hints to the contrary. First of all, it is very often easier to deal with an averaged, effective description rather than follow all details. Like when you use a Lagrange[1] multiplier to constrain a system, rather than calculating all forces. We're also taught about relativity and quantum mechanics and how they really only come about at scales different from our ordinary lives. So is physics fundamentally scale invariant or not?

Quite remarkably, we have a very definite answer to this question. Physics is, at its core, *not* scale invariant. The world looks different at different scales. Even more surprisingly, phrasing things in terms of dimensionless parameters does not help. Fundamental constants, even dimensionless ones, are not constant! They depend on the scale at which we perform our experiment. When people say "The fine structure constant has the value $\alpha \approx 1/137$" they are not correct in general, they are only correct at our ordinary energy scale.

If you think about it, this is quite bizarre. It seems to defy all our intuition built on dimensional analysis, and when even dimensional analysis doesn't work, what *can* we trust?? We have dimensionless couplings, but, somehow, we can trade them by functions of some esoteric scale that seems to appear out of nowhere. And even condensed matter physicists aren't safe! Near certain phase transitions they found some quantities that seemed to scale with *irrational* exponents! And they weren't known numbers like e or π, they were more like new fundamental *mathematical* constants. Even worse, how things scale is usually dictated by dimensional analysis, but all naive attempts failed massively. And, to add to the confusion, these exponents were the same for wildly different physical systems like magnets and gases. They called this mysterious phenomenon *universality*.

In due time we shall explain all these phenomena using the framework of the renormalisation group. But here is a hint of what's to come, it all boils down to a simple statement: field theories don't necessarily make sense to arbitrarily small length scales. If we try to make calculations in the continuum we simply get garbage. This "garbage" is the famous infinities that surround quantum field theory. However, if we introduce some fundamental scale in the game to define our theory, some minimal length beyond which our description breaks down, suddenly everything makes sense again. And using this framework we have been able to recover all those weird experimental facts to an astounding precision. And I do mean astounding. Renormalisation has the best prediction in the history of science. The *g*-factor of the the electron is the quantity that has the best agreement between data and theoretical prediction. Period. No caveats[2].

There is one issue with this. Renormalisation isn't simple. It isn't one of those ideas

---

[1] Joseph-Louis Lagrange: born Giuseppe Lodovico Lagrangia in 1736 in Turin, Piedmont-Sardinia; died in 1813 in Paris, France; Academic Advisors: Leonhard Euler, Giovanni Battista Beccaria

[2] To be fair renormalisation also has arguably the worst prediction in the history of science, which is the cosmological constant problem. However, whether that famous 10^{120} discrepancy really comes from renormalisation or by unjustified assumptions is a matter of heated debate

that you see it and think "how did I not see this before?". It is incredibly subtle. It took the better part of the 20$^{\text{th}}$ century to fully unfold, and, to some extent, we're still in the dark about many of its aspects. For this reason, there are lots of misconceptions and confusions surrounding renormalisation, which mostly stem from historical confusions. But all evidence points that it is the language of nature and we ought to understand it properly. It isn't a "hocus pocus" procedure like Feymman[3] once said, it isn't about sweeping infinities under a very large rug. It is about scale, and about phrasing predictions in terms of measurable quantities.

The point of these lectures is to convey precisely that. Stripping away all unnecessary technical details, we'll go to the simplest possible example, just a single real scalar field in flat spacetime. No discussion of symmetries, or of fermions[4]. Even at this level there is a lot to unpack, and lots of subtleties to confuse us. And the hope is that, when you go and learn about the more complex aspects of renormalisation, if ever you're confused about the meaning of it all, you can come back here, to the simple case, and then use that to understand what is actually happening in the more complicated one.

It won't be an easy journey, there are lots of dangers on the way, lot's of things one can get lost in. But, if you persevere, you'll come out with a new tool in the box, one that is incredibly powerful. So I hope you join me in this journey. See you on the other side.

---

[3]Richard Phillips Feynman: born in 1918 in Queens, USA; died in 1988 in Los Angeles, USA; Doctoral Advisor: John Archibald Wheeler; Nobel Prize in Physics 1965 together with Sin-Itiro Tomonaga and Julian Schwinger "for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles"

[4]Enrico Fermi: born in 1901 in Rome, Italy; died in 1954 in Chicago, USA; Academic Advisors: Luigi Puccianti, Max Born, Paul Ehrenfest; Nobel Prize in Physics 1938 "for his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons"

2. Path integrals in quantum mechanics

We start our journey of explaining the origin of the renormalisation group by delving into the path integral formulation of quantum mechanics. This is a very simple setting where the physics is clear and we have another very robust framework to compare with. We'll begin with deriving the path integral starting from the usual Hilbert[5] space formulation, and then we'll discuss how to recover basic quantum facts like the non-commutativity of position and momentum.

2.1. The quantum mechanical amplitude as a functional integral

We'll consider the simplest case. A single particle, with mass m, position $\boldsymbol{x}(t) \in \mathbb{R}^3$ for $t \in \mathbb{R}$, and subject to a potential $V(\boldsymbol{x})$ which does not depend on the momenta. The Hamiltonian[6] is

$$H(\boldsymbol{x}, \boldsymbol{p}) = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{x}) \tag{2.1}$$

where $\boldsymbol{p} = m\dot{\boldsymbol{x}}$.

In this case, there are no operator ordering ambiguities when going to the quantum theory. When more particles are present, or for momenta-dependent potentials, we have the usual ambiguities. This is a very hard problem to solve in general, but it has no bearing in the following discussion, hence, I'll ignore it.

We now do the usual thing, promote the canonical variables $\boldsymbol{x}$ and $\boldsymbol{p}$ to operators $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{p}}$, and impose the canonical commutation relations:

$$[\hat{x}_i, \hat{p}_j] = i\delta_{ij} \tag{2.2}$$

The Schrödinger[7] equation is then,

$$i\frac{\partial |\psi, t\rangle}{\partial t} = \hat{H} |\psi, t\rangle \tag{2.3}$$

which is solved by (since the Hamiltonian does not depend explicitly on time)

$$|\psi, t\rangle = e^{-i\hat{H}t} |\psi, 0\rangle \tag{2.4}$$

or in the position basis, i.e. written in terms of the wave-function

$$\psi(\boldsymbol{x}, t) = \langle \boldsymbol{x}| e^{-i\hat{H}t} |\psi, 0\rangle \tag{2.5}$$

Inserting the identity operator in the form $\int \mathrm{d}^3x\, |\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ we get

$$\psi(\boldsymbol{x}, t) = \int \mathrm{d}^3y\, \langle \boldsymbol{x}| e^{-i\hat{H}t} |\boldsymbol{y}\rangle \langle \boldsymbol{y}|\psi, 0\rangle = \int \mathrm{d}^3y\, \langle \boldsymbol{x}| e^{-i\hat{H}t} |\boldsymbol{y}\rangle\, \psi(\boldsymbol{y}, 0) \tag{2.6}$$

Therefore, the dynamics is entirely specified by the transition amplitude associated to a particle being prepared at $\boldsymbol{x}_\mathrm{i}$ at $t = 0$ and then found at $\boldsymbol{x}_\mathrm{f}$ at $t = T$:

---

[5]David Hilbert: born in 1862 in Königsberg or Wehlau, Kingdom of Prussia; died in 1943 in Göttingen, Germany; Doctoral Advisor: Ferdinand von Lindemann

[6]William Rowan Hamilton: born in 1805 in Dublin, Ireland; died in 1865 in Dublin, Ireland; Academic Advisor: John Brinkley

[7]Erwin Rudolf Josef Alexander Schrödinger: born in 1887 in Vienna, Austria-Hungary; died in 1961 in Vienna, Austria; Doctoral Advisor: Friedrich Hasenöhrl; Nobel Prize in Physics 1933 together with Paul Adrien Maurice Dirac "for the discovery of new productive forms of atomic theory"

$$\mathcal{A}_T(\boldsymbol{x}_{\mathrm{f}}, \boldsymbol{x}_{\mathrm{i}}) = \langle \boldsymbol{x}_{\mathrm{f}}| \, \mathrm{e}^{-\mathrm{i}\hat{H}T} \, |\boldsymbol{x}_{\mathrm{i}}\rangle \tag{2.7}$$

The trick to calculate this is to divide the interval $[0, T]$ in N slices of size $\Delta t = \frac{T}{N}$ and at the end take the limit $N \to \infty$. We then have

$$\mathcal{A}_T(\boldsymbol{x}_{\mathrm{f}}, \boldsymbol{x}_{\mathrm{i}}) = \langle \boldsymbol{x}_{\mathrm{f}}| \prod_{j=1}^{N} \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \, |\boldsymbol{x}_{\mathrm{i}}\rangle =$$

$$= \int \left(\prod_{j=1}^{N-1} \mathrm{d}^3 x_j \right) \langle \boldsymbol{x}_{\mathrm{f}}| \, \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \, |\boldsymbol{x}_{N-1}\rangle \, \langle \boldsymbol{x}_{N-1}| \, \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \, |\boldsymbol{x}_{N-2}\rangle \ldots \langle \boldsymbol{x}_1| \, \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \, |\boldsymbol{x}_{\mathrm{i}}\rangle \tag{2.8}$$

Where, in going to the last line, we once more inserted $\int \mathrm{d}^3 x \, |\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ between each Hamiltonian. Now we only need to evaluate expressions of the form

$$\langle \boldsymbol{x}_{i+1}| \, \mathrm{e}^{-\mathrm{i}\hat{H}\Delta t} \, |\boldsymbol{x}_i\rangle \tag{2.9}$$

for small Δt. To do that, we use two tricks. The first one is that Δt is small, hence, at leading order in Δt, the Baker[8]-Campbell[9]-Hausdorff[10] (BCH) formula gives

$$\exp\left(-\mathrm{i}\hat{H}\Delta t\right) \approx \exp\left(-\mathrm{i}\Delta t \frac{\hat{\boldsymbol{p}}^2}{2m}\right) \exp(-\mathrm{i}\Delta t V(\hat{\boldsymbol{x}})) \tag{2.10}$$

The second trick is inserting the identity in the form $\int \frac{\mathrm{d}^3 p}{(2\pi)^3} \, |\boldsymbol{p}\rangle\langle\boldsymbol{p}|$, which gives

$$\langle \boldsymbol{x}_{i+1}| \exp\left(-\mathrm{i}\hat{H}\Delta t\right) |\boldsymbol{x}_i\rangle = \int \frac{\mathrm{d}^3 p_i}{(2\pi)^3} \, \langle \boldsymbol{x}_{i+1}|\boldsymbol{p}_i\rangle \, \langle \boldsymbol{p}_i| \exp\left(-\mathrm{i}\hat{H}\Delta t\right) |\boldsymbol{x}_i\rangle \approx$$

$$\approx \int \frac{\mathrm{d}^3 p_i}{(2\pi)^3} \, \langle \boldsymbol{x}_{i+1}|\boldsymbol{p}_i\rangle \, \langle \boldsymbol{p}_i|\boldsymbol{x}_i\rangle \exp\left(-\mathrm{i}\Delta t \left(\frac{\boldsymbol{p}_i^2}{2m} + V(\boldsymbol{x}_i)\right)\right) \tag{2.11}$$

where we went from operators to eigenvalues by acting with $\exp(-\mathrm{i}\Delta t V(\hat{\boldsymbol{x}}))$ on the right and $\exp\left(-\mathrm{i}\Delta t \frac{\hat{\boldsymbol{p}}^2}{2m}\right)$ on the left.

Using $\langle \boldsymbol{p}|\boldsymbol{x}\rangle = \exp(-\mathrm{i}\,(\boldsymbol{p}\cdot\boldsymbol{x}))$ we get

$$\langle \boldsymbol{x}_{i+1}| \exp\left(-\mathrm{i}\hat{H}\Delta t\right) |\boldsymbol{x}_i\rangle \approx \int \frac{\mathrm{d}^3 p_i}{(2\pi)^3} \exp\left(\mathrm{i}\,(\boldsymbol{p}_i \cdot (\boldsymbol{x}_{i+1} - \boldsymbol{x}_i))\right) \exp\left(-\mathrm{i}\Delta t \left(\frac{\boldsymbol{p}_i^2}{2m} + V(\boldsymbol{x}_i)\right)\right) \tag{2.12}$$

This integral is a Gaussian[11] integral, hence, in principle, we know how to perform it. However, the coefficient of the quadratic term does not have a negative real part, it's pure

---

[8] Henry Frederick Baker: born in 1866 in Cambridge, England; died in 1956 in Cambridge England; Doctoral Advisor: Arthut Cayley

[9] John Edward Campbell: born in 1862 in Lisburn, Ireland; died in 1924 in Oxford, England

[10] Felix Hausdorff: born on 1868 in Breslau, Kindgom of Prussia; died in 1942 in Bonn, Germany; Doctoral Advisors: Heinrich Bruns, Adolph Mayer

[11] Johann Carl Friedrich Gauss: born in 1777 in Brunswick, Principality of Brunswick-Wolfenbüttel; died in 1855 in Göttingen, German Confederation; Doctoral Advisor: Johann Friedrich Pfaff

imaginary. Therefore this integral is not convergent. To solve this issue we use two tricks: Wick[12] rotation and analytic continuation.

The vague idea behind this is to think of the integral as a function of *complex* Δt instead of real (and positive). If we know the value of this function in a given open region, there is a unique procedure to assign values to this function in other points, so long as we don't cross any poles/branch cuts. This is the idea of *analytic continuation*.

In our case, we know the value of the integral so long as $\mathrm{Re}\{\mathrm{i}\Delta t\} > 0$, since, in this case, the integral is damped for $\boldsymbol{p}_i^2 \to \infty$. We then define $\Delta\tau = \mathrm{i}\Delta t$ and rotate to real (and positive) $\Delta\tau$. The variable τ is called *Euclidean*[13] *time*, and this substitution is called *Wick rotation*. We can then perform this integral because it is actually convergent. Because the final answer will make sense in an open region in the complex plane (the region $\mathrm{Re}\{\mathrm{i}\Delta t\} > 0$ is open) the mathematics of analytic continuation will ensure that, if we rotate back to real physical time Δt, the answer we get is unique.

I know this may feel a bit hand-wavy and not very rigorous, but there is actually serious mathematics behind this and, in the end, it works, i.e. it agrees with experiment. For the more complex case of quantum field theory, the status of Wick rotation is a bit shakier and it is still not entirely clear to what extent we are actually loosing any important physics by doing this. We shall mostly ignore all these issues and fully work with Euclidean time objects. Our integral then becomes[14]

$$\langle\boldsymbol{x}_{i+1}|\exp\left(-\Delta\tau\hat{H}\right)|\boldsymbol{x}_i\rangle \approx \int \frac{\mathrm{d}^3 p_i}{(2\pi)^3}\exp\left(-\Delta\tau\left(\frac{\boldsymbol{p}_i^2}{2m}+V(\boldsymbol{x}_i)\right)+\mathrm{i}\left(\boldsymbol{p}_i\cdot(\boldsymbol{x}_{i+1}-\boldsymbol{x}_i)\right)\right) =$$
$$= \left(\frac{m}{2\pi\Delta\tau}\right)^{\frac{3}{2}}\exp\left(-\Delta\tau\left(\frac{1}{2}m\frac{|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i|^2}{\Delta\tau^2}+V(\boldsymbol{x}_i)\right)\right) \quad (2.13)$$

The full amplitude is then (in Euclidean time $T_{\mathrm{E}} = \mathrm{i}T$)

$$K_{T_{\mathrm{E}}}(\boldsymbol{x}_{\mathrm{f}},\boldsymbol{x}_{\mathrm{i}}) = \langle\boldsymbol{x}_{\mathrm{f}}|\exp\left(-T_{\mathrm{E}}\hat{H}\right)|\boldsymbol{x}_{\mathrm{i}}\rangle =$$
$$= \lim_{N\to\infty}\int \prod_{i=1}^{N-1}\left[\mathrm{d}^3 x_i\left(\frac{m}{2\pi\Delta\tau}\right)^{\frac{3}{2}}\right]\exp\left(-\sum_{i=0}^{N-1}\Delta\tau\left(\frac{1}{2}m\frac{|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i|^2}{\Delta\tau^2}+V(\boldsymbol{x}_i)\right)\right)$$
$$(2.14)$$

where we renamed the amplitude in Euclidean time as $K_{T_{\mathrm{E}}}$ to emphasise the change to Euclidean time[15], and defined $\boldsymbol{x}_{\mathrm{i}} = \boldsymbol{x}_0$, $\boldsymbol{x}_{\mathrm{f}} = \boldsymbol{x}_N$.

This form of the amplitude is really suggestive. It would be very convenient if we could make the following identifications

---

[12] Gian Carlo Wick: born in 1909 in Turin, Kingdom of Italy; died in 1992 in Turin, Italy; Doctoral Advisor: Gleb Wataghin

[13] Euclid: born c. 325 BCE in an unknown place; died c. 270 BCE in (possibly) Alexandria, Hellenistic Egypt

[14] If you need some reminding of how to do Gaussian integrals refer to Appendix B

[15] Also, in Euclidean time we are essentially doing statistical physics and in that context this object would be the heat kernel.

$$\lim_{N \to \infty} \sum_{i=0}^{N-1} \Delta\tau \left(\frac{1}{2} m \frac{|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i|^2}{\Delta\tau^2} + V(\boldsymbol{x}_i) \right) \overset{?}{=} \int_0^{T_{\mathrm{E}}} \mathrm{d}\tau \left(\frac{1}{2} m \dot{\boldsymbol{x}}(\tau) + V(\boldsymbol{x}(\tau)) \right) = S_{\mathrm{E}}[\boldsymbol{x}]$$

(2.15)

$$\lim_{N \to \infty} \prod_{i=1}^{N-1} \left[\mathrm{d}^3 x_i \left(\frac{m}{2\pi\Delta\tau} \right)^{\frac{3}{2}} \right] \overset{?}{=} \mathcal{D}x$$

(2.16)

which we would identify as a Euclidean time version of the classical action and a functional integration measure. We'll attest the validity of these limits in the next section. Assuming this to be a legitimate procedure we write

$$K_{T_{\mathrm{E}}}(\boldsymbol{x}_{\mathrm{f}}, \boldsymbol{x}_{\mathrm{i}}) = \int_{\boldsymbol{x}_{\mathrm{i}},0}^{\boldsymbol{x}_{\mathrm{f}},T_{\mathrm{E}}} \mathcal{D}x \ \mathrm{e}^{-S_{\mathrm{E}}[\boldsymbol{x}]}$$

(2.17)

This expression can be interpreted as stating that to calculate the amplitude for some process we should sum over all possible paths that the particle can take, consistent with the measured initial and final positions, weighed with a factor given by the value of the classical action evaluated on the given path, with exponential suppression. This integral is dominated by the regions where the action is least, which means we recover classical mechanics by taking the saddle point approximation of this integral.[16]

2.2. Non-commutativity and the continuum limit

In the last chapter, we solved the Schrödinger equation by calculating the transition amplitude to go from one position to another. By writing this amplitude in terms of an integral over all possible paths, we fully determined the dynamics of our system. Given some initial state we can use the path integral to compute it's evolution, in principle at least[17]. But this is not all there is to it in quantum mechanics, we're missing operators.

If we're just asking about expectation values (or more generally correlation functions) at a particular time, then we're done. We can calculate the state hence we have the required probability distribution to calculate the expectation values, it's the whole point of calculating the state. However, we can imagine more complicated arrangements where the path integral comes in handy. Imagine that we know both the initial and the final states in some process (because we've measured them, or a little bird told us in advance), and we want to calculate the expectation value of an observable which is measured at some intermediate time t. This may seem like a ridiculously complicated arrangement, but when we move to quantum field theory, these kinds of statements are actually the only ones we can make, so bear with me.[18]

We'll consider the operator to be a function of $\hat{\boldsymbol{x}}$, and we'll denote it by $\hat{O}(\hat{\boldsymbol{x}})$, then this very complicated expectation value will be given by[19]

---

[16] Roughly, this approximation means assuming that the integral will just equal (the exponential of) the action evaluated at its extremum. Therefore the particle takes a single trajectory, the one that extremizes the action.

[17] Emphasis on *in principle*, path integrals are usually really hard to compute, as in, most problems in current theoretical physics can be traced back to a path integral that no one knows how to compute

[18] In that context we usually choose the initial and final states to be the vacuum, and the operator to be just a simple product of the fundamental fields

[19] We'll want to express this in terms of path integrals, where we don't have operators, only usual functions. In that context, the position is always a function of time, $\boldsymbol{x}(t)$, hence the notation on the LHS. On the RHS we're dealing with operators which can or cannot depend on time depending on the picture, hence the notation on the RHS, c.f. Appendix A

$$\langle O(\boldsymbol{x}(t))\rangle = \langle \boldsymbol{x}_{\mathrm{f}}|\,\mathrm{e}^{-\mathrm{i}\hat{H}(T-t)}\hat{O}(\hat{\boldsymbol{x}})\mathrm{e}^{-\mathrm{i}\hat{H}t}\,|\boldsymbol{x}_{\mathrm{i}}\rangle =$$

$$= \int \mathrm{d}^3x \,\langle \boldsymbol{x}_{\mathrm{f}}|\,\mathrm{e}^{-\mathrm{i}\hat{H}(T-t)}\hat{O}(\hat{\boldsymbol{x}})\,|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\,\mathrm{e}^{-\mathrm{i}\hat{H}t}\,|\boldsymbol{x}_{\mathrm{i}}\rangle =$$

$$= \int \mathrm{d}^3x \, O(\boldsymbol{x}) \,\langle \boldsymbol{x}_{\mathrm{f}}|\,\mathrm{e}^{-\mathrm{i}\hat{H}(T-t)}\,|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\,\mathrm{e}^{-\mathrm{i}\hat{H}t}\,|\boldsymbol{x}_{\mathrm{i}}\rangle =$$

$$= \int \mathrm{d}^3x \, O(\boldsymbol{x})\mathcal{A}_{T-t}(\boldsymbol{x}_{\mathrm{f}},\boldsymbol{x})\mathcal{A}_t(\boldsymbol{x},\boldsymbol{x}_{\mathrm{i}}) \tag{2.18}$$

The complicated exponential factors in the first line are there because we need to evolve the initial state until time t before we can act with the operator in order to perform the measurement, and then, to compare it with the known final state, we must evolve it the rest of the time interval from t to T. In the remaining steps we merely wrote it in terms of the transition amplitudes, which we know how to express in terms of path integrals. We then rotate to Euclidean time and write it as

$$\langle O(\boldsymbol{x}(\tau))\rangle = \int \mathrm{d}^3x \, O(\boldsymbol{x})K_{T_{\mathrm{E}}-\tau}(\boldsymbol{x}_{\mathrm{f}},\boldsymbol{x})K_\tau(\boldsymbol{x},\boldsymbol{x}_{\mathrm{i}}) =$$

$$= \int \mathrm{d}^3x \, O(\boldsymbol{x}) \int_{\boldsymbol{x},\tau}^{\boldsymbol{x}_{\mathrm{f}},T_{\mathrm{E}}} \mathcal{D}y \,\mathrm{e}^{-S_{\mathrm{E}}[\boldsymbol{y}]} \int_{\boldsymbol{x}_{\mathrm{i}},0}^{\boldsymbol{x},\tau} \mathcal{D}z \,\mathrm{e}^{-S_{\mathrm{E}}[\boldsymbol{z}]} =$$

$$= \int_{\boldsymbol{x}_{\mathrm{i}},0}^{\boldsymbol{x}_{\mathrm{f}},T_{\mathrm{E}}} \mathcal{D}y \, O(\boldsymbol{x}(\tau))\mathrm{e}^{-S_{\mathrm{E}}[\boldsymbol{y}]} \tag{2.19}$$

where in the last line we used the fact that integrating over all paths from from $\boldsymbol{x}_{\mathrm{i}}$ to $\boldsymbol{x}$, then integrating over all paths from $\boldsymbol{x}$ to $\boldsymbol{x}_{\mathrm{f}}$, and finally integrating over all possible intermediate points $\boldsymbol{x}$ is exactly the same thing as integrating over all paths from $\boldsymbol{x}_{\mathrm{i}}$ to $\boldsymbol{x}_{\mathrm{f}}$, provided the operator (which is now just a function) is evaluated at the correct time. This is called the *concatenation property*.

This is a bit surprising, the somewhat complicated arrangement we concocted, is actually a very natural thing to calculate with path integral methods. We're just calculating the expectation value of a function of position, which, in this context, is a function of time.

Now, what happens if we try to calculate the expectation value of a composite operator, one which is a product of several operators evaluated at different times? Said another way, what happens if we make multiple successive measurements at different times? The derivation proceeds exactly as in (2.18) inserting the identity in terms of the position basis in between every operator, but we must be careful. We make measurements in order, so we must insert operators such that operators evaluated at earlier times are to the right. Otherwise we get factors similar to

$$\langle \boldsymbol{x}|\,\mathrm{e}^{\hat{H}T_{\mathrm{E}}}\,|\boldsymbol{x}'\rangle \tag{2.20}$$

for $T_{\mathrm{E}} > 0$, which do not lead to convergent integrals. Defining $\mathcal{T}(\prod_i O_i(\tau_i))$ to be such *time-ordering* we get

$$\left\langle \mathcal{T}\left(\prod_i O_i(\boldsymbol{x}(\tau_i))\right)\right\rangle = \langle \boldsymbol{x}_{\mathrm{f}}|\,\mathrm{e}^{-\hat{H}T_{\mathrm{E}}}\mathcal{T}\left(\prod_i \mathrm{e}^{\hat{H}\tau_i}\hat{O}_i(\hat{\boldsymbol{x}})\mathrm{e}^{-\hat{H}\tau_i}\right)|\boldsymbol{x}_{\mathrm{i}}\rangle =$$

$$= \int_{\boldsymbol{x}_{\mathrm{i}}}^{\boldsymbol{x}_{\mathrm{f}}} \mathcal{D}x \,\mathrm{e}^{-S_{\mathrm{E}}[\boldsymbol{x}]}\prod_i O_i(\boldsymbol{x}(\tau_i)) \tag{2.21}$$

This is a very important point, the path integral naturally picks a preferred ordering of operators inside expectation values, the *time-ordering*. Since, inside the path integral, the operators became mere functions, everything commutes, there's no notion of ordering. However, a random product of operators in the path integral perspective becomes a time-ordered product of operators in the canonical quantisation perspective. If operators are evaluated at the same time they won't have a preferred ordering, however, since we're considering only functions of position, they commute even when viewed as operators acting on a Hilbert space, so there's no issue there. But what if we consider functions of momentum? Well, there are a few issues when considering functions of momentum that complicate matters

1. Hermiticity[20] is a bit bizarre once we go to Euclidean time. There are analogues, but it's a bit fiddly if you do the calculations in certain ways. We'll gently sweep this one under the rug, by doing the calculations in a way this isn't an issue.

2. For arbitrary potentials momentum is not conserved. This can show up in some calculations using path integrals where, if you do your calculations in some ways, it appears that there are additional contributions to the momentum proportional to derivatives of the potential. This is very subtle and has to do with the fact that, at least classically, the potential generates an impulse and contributes to the future momentum. I've just mentioned this here so that you don't go mad trying to understand where you went wrong if you do the calculation yourself. However, it's not very relevant for quantum field theory and we'll set $V = 0$

This illustrates what is simultaneously the greatest strength and the greatest weakness of the path integral. If you try to hack the path integral, you'll quickly get non-sense, you need to be very careful about what you mean from the get go in order to proceed. With operators, you can very easily be naughty and ignore some subtleties, and still get more or less the correct answer. What usually happens is that simple calculations and proofs are way easier with operators than path integrals, however, for genuinely subtle situations (like gauge theories), the path integral makes dead obvious from the start where you can go wrong, and doesn't let you go through until you truly understand what's going on. Whereas with operators it's very hard to see what assumption went wrong most of the times. It's a trade off that depends on the situation, for ordinary quantum mechanics, operators are the way to go, for quantum field theory, usually path integrals are easier, but in any given problem you should analyse it carefully.

So, because we have a lot of subtleties going on, I'll restrict to showing those that carry through to quantum field theory and simplify the rest. This means I'll set $V = 0$ and consider only one spatial dimension. Increasing the number of dimensions is easy, turning on the potential is not so easy. We'll proceed nevertheless by calculating the expectation value of the momentum.

First just a note on how to write momentum in Euclidean time

$$p = m\frac{\mathrm{d}x}{\mathrm{d}t} = \mathrm{i}m\frac{\mathrm{d}x}{\mathrm{d}\tau} = \mathrm{i}m\dot{x} \tag{2.22}$$

and also,

$$\langle x|\,\hat{p}\,|\psi\rangle = -\mathrm{i}\frac{\partial}{\partial x}\,\langle x|\psi\rangle \tag{2.23}$$

---

[20]Charles Hermite: born in 1822 in Dieuze, France; died in 1901 in Paris, France; Doctoral Advisor: Eugène Charles Catalan

The expectation value of the momentum is

$$\langle p(\tau)\rangle = \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\hat{p}\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle =$$

$$= \int \mathrm{d}x\, \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\,|x\rangle\langle x|\, \hat{p}\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle =$$

$$= \int \mathrm{d}x\, \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\,|x\rangle \left(-\mathrm{i}\frac{\partial}{\partial x}\,\langle x|\, \mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle\right) =$$

$$= \int \mathrm{d}x\, K_{T_{\mathrm{E}}-\tau}(x_{\mathrm{f}}, x)\left(-\mathrm{i}\frac{\partial K_{\tau}(x, x_{\mathrm{i}})}{\partial x}\right) \tag{2.24}$$

To calculate this derivative let's go to the discrete version of the heat kernel given by (2.14). Note that only the last term depends on the end point, hence we just need to calculate

$$\frac{\partial}{\partial x_{\tau}}K_{\Delta\tau}(x_{\tau}, x_{\tau-\Delta\tau}) \tag{2.25}$$

for $\Delta\tau \ll 1$. Plugging (2.14) in here we get

$$\frac{\partial}{\partial x_{\tau}}K_{\Delta\tau}(x_{\tau}, x_{\tau-\Delta\tau}) = \frac{\partial}{\partial x_{\tau}}\left(\sqrt{\frac{m}{2\pi\Delta\tau}}\exp\left(-\frac{m}{2}\frac{(x_{\tau}-x_{\tau-\Delta\tau})^2}{\Delta\tau}\right)\right) =$$

$$= -m\frac{x_{\tau}-x_{\tau-\Delta\tau}}{\Delta\tau}K_{\Delta\tau}(x_{\tau}, x_{\tau-\Delta\tau}) \tag{2.26}$$

Therefore,

$$\frac{\partial K_{\tau}(x, x_{\mathrm{i}})}{\partial x} = -m\left.\frac{\mathrm{d}x}{\mathrm{d}\tau}\right|_{\tau}K_{\tau}(x, x_{\mathrm{i}}) = \mathrm{i}p(\tau)K_{\tau}(x, x_{\mathrm{i}}) \tag{2.27}$$

Plugging back into (2.24) we finally get

$$\langle p(\tau)\rangle = \int \mathrm{d}x\, K_{T_{\mathrm{E}}-\tau}(x_{\mathrm{f}}, x)p(\tau)K_{\tau}(x, x_{\mathrm{i}}) =$$

$$= \int \mathrm{d}x\, p(\tau)\int_{x,\tau}^{x_{\mathrm{f}},T_{\mathrm{E}}}\mathcal{D}y\,\mathrm{e}^{-S_{\mathrm{E}}[y]}\int_{x_{\mathrm{i}},0}^{x,\tau}\mathcal{D}z\,\mathrm{e}^{-S_{\mathrm{E}}[z]} =$$

$$= \int_{x_{\mathrm{i}},0}^{x_{\mathrm{f}},T_{\mathrm{E}}}\mathcal{D}y\, p(\tau)\mathrm{e}^{-S_{\mathrm{E}}[y]} \tag{2.28}$$

It was a bit harder than before, but we ended up with a nice result, which is reassuring. It appears there is no difference whether it's position or momentum, you just insert it inside the path integral to calculate the expectation vale. However, there is an issue if we insert a function of both position and momentum. Let's say we want to insert the product of the two. If they're evaluated at different times there's no issue, everything goes exactly as expected, time ordering prevails. But, what if we insert them at the same time? Then we have a problem, in the operator perspective the result depends on the order in which we insert them, while in the path integral perspective they're ordinary functions and hence order cannot matter! How can we solve this?

We need to be clever, remember that the path integral picks an ordering, time-ordering. We'll exploit that by considering two different times τ_- and τ_+ such that $0 < \tau_- < \tau < \tau_+ < T_{\mathrm{E}}$. In this way we can force an ordering of the two operators by inserting them at different times. After that we make τ_- approach τ from below, and τ_+ approach τ from above; and calculate the difference between the two approaches. First let's examine what the canonical point of view has to say about this,

$$\langle x(\tau)p(\tau_-)\rangle = \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\hat{x}\mathrm{e}^{-\hat{H}(\tau-\tau_-)}\hat{p}\mathrm{e}^{-\hat{H}\tau_-}\,|x_{\mathrm{i}}\rangle \rightarrow \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\hat{x}\hat{p}\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle \quad (2.29)$$

$$\langle p(\tau_+)x(\tau)\rangle = \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau_+)}\hat{p}\mathrm{e}^{-\hat{H}(\tau_+-\tau)}\hat{x}\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle \rightarrow \langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}\hat{p}\hat{x}\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle \quad (2.30)$$

where the limits taken are $\tau_- \nearrow \tau$ and $\tau_+ \searrow \tau$. Subtracting the two we get

$$\langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}(T_{\mathrm{E}}-\tau)}[\hat{x},\hat{p}]\mathrm{e}^{-\hat{H}\tau}\,|x_{\mathrm{i}}\rangle = \mathrm{i}\,\langle x_{\mathrm{f}}|\, \mathrm{e}^{-\hat{H}T_{\mathrm{E}}}\,|x_{\mathrm{i}}\rangle = \mathrm{i}K_{T_{\mathrm{E}}}(x_{\mathrm{f}}, x_{\mathrm{i}}) \quad (2.31)$$

using the fact that

$$[\hat{x},\hat{p}] = \mathrm{i} \quad (2.32)$$

This is one of the most fundamental results in the quantum theory, it was our first step in setting up the problem. The path integral better be able to recover this. To do just that we'll move to the discrete version of the path integral given by (2.14). We'll focus on the integration around τ, taking $\tau_- = \tau - \Delta\tau$ and $\tau_+ = \tau + \Delta\tau$,

$$\int \mathrm{d}x_\tau\, K_{\Delta\tau}(x_{\tau+\Delta\tau}, x_\tau)\left(x_\tau m\mathrm{i}\frac{x_\tau - x_{\tau-\Delta\tau}}{\Delta\tau} - x_\tau m\mathrm{i}\frac{x_{\tau+\Delta\tau} - x_\tau}{\Delta\tau}\right)K_{\Delta\tau}(x_\tau, x_{\tau-\Delta\tau}) =$$

$$= -\mathrm{i}\int \mathrm{d}x_\tau\, x_\tau \frac{\partial}{\partial x_\tau}\Big(K_{\Delta\tau}(x_{\tau+\Delta\tau}, x_\tau)K_{\Delta\tau}(x_\tau, x_{\tau-\Delta\tau})\Big) =$$

$$= \mathrm{i}\int \mathrm{d}x_\tau\, K_{\Delta\tau}(x_{\tau+\Delta\tau}, x_\tau)K_{\Delta\tau}(x_\tau, x_{\tau-\Delta\tau}) = \mathrm{i}K_{2\Delta\tau}(x_{\tau+\Delta\tau}, x_{\tau-\Delta\tau}) \quad (2.33)$$

where in going to the second line we used (2.26), then we integrated by parts, and finally we used the concatenation property. The remaining integrals are easy to do, we're left with

$$\langle x(\tau)p(\tau_-) - p(\tau_+)x(\tau)\rangle \xrightarrow{\tau_{+/-}\to\tau} \mathrm{i}K_{T_{\mathrm{E}}}(x_{\mathrm{f}}, x_{\mathrm{i}}) \quad (2.34)$$

which agrees with the operator approach.

Wait, what? How was this possible? For differentiable functions that limit is exactly zero, no matter what we do. However, when computing the path integral using the discrete version we definitely don't obtain zero as an answer. The only possible solution is that the function $x(\tau)$ is a continuous but non-differentiable function, it actually must not be differentiable anywhere for (2.34) to make sense for all τ! This is against our common intuition, but then again, this is quantum mechanics, what isn't against our intuition in this field? But this means the limit in (2.15) doesn't make sense, there is no continuum action.

If (2.15) doesn't make sense, does (2.16) make sense? It turns out it also doesn't make sense. The reason for this is rather formal and beyond the main scope of these lectures but suffice to say that there exists no standard Lebesgue[21] measure for an infinite dimensional space. Does anything make sense in the continuum? Or do we need to always work with the discrete version of the path integral? It turns out that, although the measure and the action don't exist separately, the product between the measure and the kinetic part of the action has a well defined continuum limit. It constitutes a Wiener[22] measure, which is not a Lebesgue measure in the usual sense. This works so long as we use paths that are nowhere-differentiable. This is the reason why we were always capable of taking the continuum limit

---

[21]Henri Léon Lebesgue: born in 1875 in Beauvais, France; died in 1941 in Paris, France; Doctoral Advisor: Émile Borel

[22]Norbert Wiener: born in 1894 in Columbia, USA; died in 1964 in Stockholm, Sweden; Doctoral Advisor: Karl Schmidt

after we're done calculating and get sensible answers. And this is also connected to why standard quantum mechanics works without the need to discretise time.

However, this only works for non-relativistic quantum mechanics. When we move to quantum field theory things are not so simple, not all theories have a well defined continuum limit, and even for those that do, no one has been able to construct the analogue of the Wiener measure. We'll spend the rest of these lectures exploring how to deal with this and what does this tell us about our world.

3. The Wilsonian renormalisation group

In the last chapter, we've seen that in order to make sense of path integrals, we must make time a discrete variable and only at the end take the continuum limit. In the case of quantum field theory, our dynamical variables will depend on space as well as time so it makes sense to discretise space too. However, there is no Wiener measure to save us now, we must take this discretisation into account and fully understand its role in our description of Nature. There is no guarantee that some continuum limit exists for quantum field theory. This means there is a new scale in the game, the size of the spacing, call it a.

This raises many questions. First of all, what is this scale and what does it mean to say that we're discretising spacetime? Secondly, why this discretisation in particular, why do we choose a and not $a/2$ or $2a$, or why don't we use a smoother procedure? Thirdly, how to make contact with experiment, can we actually measure the effects of this scale or is it some theoretical trick? And finally, when can we remove this discretisation, i.e. when does our theory make sense in the continuum?

The rest of these notes will try to answer these questions. This first chapter will deal with the condensed matter or statistical point of view of effective field theories, which is also very relevant for the high energy/particle physics realm

3.1. Coarse graining

To be concrete, let's consider the most general theory with a single real scalar field $\phi(\boldsymbol{x})$ in d dimensions[23]. To do that, we write every possible term in the action compatible with the symmetries we impose. In particular we'll impose two symmetries, Euclidean group $E(d) \equiv ISO(d)$ composed of rotations and translations[24], and the discrete $\mathbb{Z}_2$ symmetry $\phi \to -\phi$. And I do mean every possible term: ϕ^{892} is there, and so is $\phi^7 \nabla^4 \phi^{2075}$ all with their respective couplings. However, $\partial_1 \phi$ isn't there because it isn't rotationally invariant, ϕ^3 is also forbidden because it is odd in ϕ, and $1/\phi$ is also not allowed because it blows up at $\phi = 0$. The first few terms in the action are,

$$S[\phi] = \int \mathrm{d}^d x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi)^2 + \frac{1}{2} m^2 \phi^2 + \frac{\lambda_4}{4!} \phi^4 + \dots \right) \tag{3.1}$$

where we fixed the coefficient in from of the kinetic term to be $1/2$ thereby fixing the overall scale of our field.

We could proceed to work directly in real space and use a lattice, however, it will be simpler to work in momentum space, which is entirely equivalent. Therefore, we'll regularise our theory by imposing that the fields have finite support in momentum space, i.e.

$$\tilde{\phi}(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda_0 \\ \tilde{\phi}(\boldsymbol{p}) & |\boldsymbol{p}| < \Lambda_0 \end{cases} \tag{3.2}$$

We write the partition function as

$$Z_{\Lambda_0} = \int \mathcal{D}\phi \, \mathrm{e}^{-S_{\Lambda_0}[\phi]} \tag{3.3}$$

leaving explicit the dependence of both the partition function and the action in Λ_0.

Note that we do not know how to define our theory without the regularisation, the cutoff is part of the *definition* of our theory. This is a very important point, the theory

---

[23]We assume that we have Wick rotated to Euclidean signature therefore $\boldsymbol{x} = (x^1, x^2, x^3, \dots, x^d)$, where $x^d = \tau$ is the Euclidean time.

[24]The Lorentzian signature counterpart would be the Poincaré group

is defined *at this scale*, we cannot answer questions about what happens above that scale, because our theory knows nothing of it. This is the clue to interpret the cutoff, it is the maximum scale to which we give credit to our theory, you can think of it as the resolution of our measurement apparatus. Above that scale we have absolutely no clue about what happens, this theory may continue to be valid, or it may be complete bollocks. But we can't access it with experiment so, who cares? I don't.

Now imagine that we have tested the theory with some experiments up to the scale Λ_0, so we know it works all the way up to that energy scale. But, we are only interested in describing phenomena at some scale way below that initial cutoff, we may not have the money to build an equally big accelerator. If this is the case we should be able to use a different theory, an *effective* theory with a different cutoff $\Lambda < \Lambda_0$ as long as we are only interested in energies $E \ll \Lambda$. Let's build that theory.

We'll split our field into two components, the useless/high energy/UV modes, ϕ^+, that have support between Λ and Λ_0 i.e.

$$\tilde{\phi}^+(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda_0 \\ \tilde{\phi}(\boldsymbol{p}) & \Lambda < |\boldsymbol{p}| < \Lambda_0 \\ 0 & |\boldsymbol{p}| < \Lambda \end{cases} \tag{3.4}$$

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

$$\tilde{\phi}^-(\boldsymbol{p}) = \begin{cases} 0 & |\boldsymbol{p}| > \Lambda \\ \tilde{\phi}(\boldsymbol{p}) & |\boldsymbol{p}| < \Lambda \end{cases} \tag{3.5}$$

The partition function becomes

$$Z_{\Lambda_0} = \int \mathcal{D}\phi^- \mathcal{D}\phi^+ \ e^{-S_{\Lambda_0}[\phi^+ + \phi^-]} \stackrel{!}{=} \int \mathcal{D}\phi^- \ e^{-S_{\Lambda}[\phi^-]} \tag{3.6}$$

where in the last line we expressed our wish of writing the partition function solely as an integration over the low energy modes. To do this we define the *effective action* as,

$$e^{-S_{\Lambda}[\phi^-]} = \int \mathcal{D}\phi^+ e^{-S_{\Lambda_0}[\phi^+ + \phi^-]} \tag{3.7}$$

Note the similarity with the way we defined the Wilsonian[25] effective action in (B.17), this is the Wilsonian effective action where the low-energy fields are considered as external sources in the path integral.

What is the form of this effective action? We started with the most general possible action, therefore we can only end up with an action that has exactly the same form but different coefficients

$$S_{\Lambda}[\phi^-] = \int \mathrm{d}^d x \left(\frac{1}{2} Z'_{\phi} (\boldsymbol{\nabla}\phi^-)^2 + \frac{1}{2} m'^2 (\phi^-)^2 + \frac{\lambda'_4}{4!} (\phi^-)^4 + \dots \right) \tag{3.8}$$

Let's stop for a moment and consider the implications. If we can only access lower energy scales, we'll "see"[26] different effective couplings in our theory. The masses and charges/couplings of the fundamental fields that make up our universe aren't fixed, they depend on the scales which we can access! If I didn't blow your mind just now, I don't know why do you study physics.

---

[25]Kenneth Geddes "Ken" Wilson: born in 1936 in Waltham, USA; died in 2013 in Saco, USA; Doctoral Advisor: Murray Gell-Mann; Nobel Prize in Physics 1982 "for his theory for critical phenomena in connection with phase transitions"

[26]We'll give a more precise meaning to this in the next chapter

As we're physicists we'd like to be able to compare the two theories, however, at this moment, that is impossible for two reasons. Firstly, the theory is defined *with* a cutoff, therefore if we just compared the two actions straight away we'd be comparing apples to oranges. Said another way, there is an implicit dependence on the cutoff inside the spacetime integration which would make our comparison meaningless. To get rid of this we rescale our spacetime variables as

$$\boldsymbol{x} \to \boldsymbol{x}' = \frac{\Lambda}{\Lambda_0}\boldsymbol{x}, \quad \boldsymbol{p} \to \boldsymbol{p}' = \frac{\Lambda_0}{\Lambda}\boldsymbol{p} \tag{3.9}$$

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

Secondly, rescaling the fields by an overall constant factor has no physical significance, therefore we must fix the overall scale of the field in some way between the two theories. The best way to do this depends on the particular theory you're considering, in our case, we'll fix the coefficient of the kinetic term to be exactly $1/2$ by defining

$$\phi'(\boldsymbol{x}') = \sqrt{Z_\phi}\phi^-(\boldsymbol{x}) \tag{3.10}$$

where $Z_\phi = \left(\frac{\Lambda_0}{\Lambda}\right)^{d-2} Z'_\phi$.

The final action is

$$S_\Lambda[\phi'] = \int \mathrm{d}^d x' \left(\frac{1}{2}(\boldsymbol{\nabla}'\phi')^2 + \frac{1}{2}m^2(\Lambda)\phi'^2 + \frac{\lambda_4(\Lambda)}{4!}\phi'^4 + \dots \right) \tag{3.11}$$

This procedure is the famous *renormalisation group* (RG). Which is a very bad name for it, since nothing is normalised more than once, and it is not a group since we cannot invert this procedure. But the name stuck and it would be criminal of me to call it any other thing.

To recap, the steps of the renormalisation group are the following:

1. Coarse graining: this means integrating out the degrees of freedom which are not accessible to yield an effective theory at some lower resolution

2. Fix the scale of the spacetime: this means rescaling the spacetime variables to make the dependence on the new cutoff explicit

3. Fix the scale of the field: this means removing the ambiguity in the overall scale factor of the field variables by some independent manner
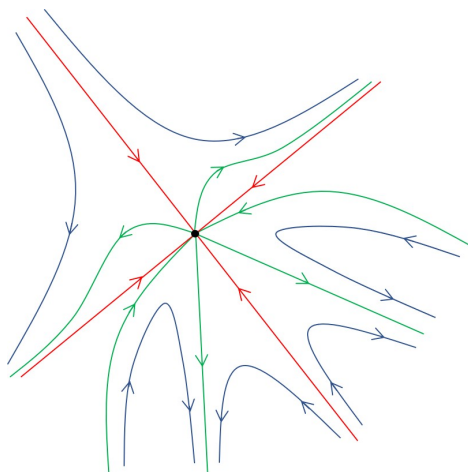

Figure 1: *RG flows in Theory Space*

To picture what's going on, imagine a space such that each point corresponds to a given value of all couplings, i.e. each point corresponds to a different theory, this is the *theory space*. Now, imagine we start at some point then we lower the scale of our theory by performing the three steps of RG, the end result is that we flow to some other point in theory space. Therefore this picture really has lots of flow lines that represent what happens under RG.

What are the possibilities for these flows? In general, this is an infinite dimensional space, and the flows can be rather complicated, understanding the shape of RG flows is something that occupies large swathes of modern theoretical physics. But there will be some special points for which the RG flow does nothing. These are *fixed points* and correspond to conformal field theories (CFT). So let us plant ourselves in such a point. Now imagine we deform our theory by changing a coupling by hand, moving away from the fixed point. What is the RG flow then? There are three possibilities:

1. We flow back towards the fixed point. In this case it is called an *irrelevant deformation* since those deformations don't matter for the low-energy physics. It will turn out that most of the deformations you can think of are of this type, which vastly simplifies our analysis. The region in theory space around the fixed point that corresponds to irrelevant deformations is called the *critical surface*, it is the set of points that under RG flow converge to our fixed point

2. We flow away from the fixed point. In this case it is called a *relevant deformation* since this is the kind of deformation that matters in the IR. There are typically only a few of these. The flow that emanates from a fixed point by turning on a relevant deformation is called the *renormalised trajectory*

3. Nothing happens. In this case they are called *marginal deformations* and signal the fact that the fixed point is not truly a point but rather a line or some more complicated object. Exactly marginal deformations are very rare, usually they are only approximately marginal, we'll see examples of this later on, in those cases they are either *marginally irrelevant* or *marginally relevant* deformations
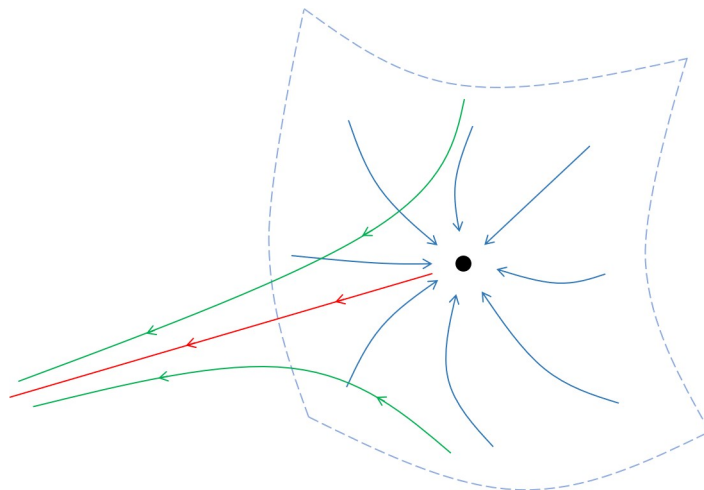

Figure 2: *Flow near a fixed point: in blue, irrelevant directions along the critical surface; in red, renormalised trajectory; in greed, generic direction*

If you start in some arbitrary point in theory space more complicated things can happen like flowing to a limit cycle, these kinds of exotic RG flows can be ruled out in most cases, but our grasp away from fixed points is rather limited.

The fact that this flow is highly converging on a renormalised trajectory is the reason for the phenomenon of universality we mentioned in the introduction, where distinct physical systems like gases and magnets behave in similar ways, they start from different points in theory space but end up flowing to the same trajectory. It is also what makes high energy physics really challenging because there is very little hope to extract the details of the UV if you can only do experiments deep in the IR.

This is all very pretty, but it's hard to get a feeling for that this procedure is actually doing. What always helps is a simple example. So we'll do the simplest possible thing. It's too simple to see a lot of the features we discussed, but we must start somewhere.

Consider just the first two terms in our action, i.e. a massive Klein[27]-Gordon[28] field:

$$
S_{\Lambda_0}[\phi] = \int \mathrm{d}^d x \frac{1}{2}\big((\boldsymbol{\nabla}\phi)^2 + m_0^2 \phi^2\big) =
$$
$$
= \int_{|\boldsymbol{p}|<\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2}\tilde{\phi}(-\boldsymbol{p})\big(\boldsymbol{p}^2 + m_0^2\big)\tilde{\phi}(\boldsymbol{p}) \tag{3.12}
$$

In this case, since ϕ^+ and ϕ^- have supports in momentum space that do not intersect, the action factors out completely,

$$
S_{\Lambda_0}[\phi^+ + \phi^-] = \int_{|\boldsymbol{p}|<\Lambda} \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2}\tilde{\phi}^-(-\boldsymbol{p})\big(\boldsymbol{p}^2 + m_0^2\big)\tilde{\phi}^-(\boldsymbol{p})+
$$
$$
+ \int_{\Lambda<|\boldsymbol{p}|<\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2}\tilde{\phi}^+(-\boldsymbol{p})\big(\boldsymbol{p}^2 + m_0^2\big)\tilde{\phi}^+(\boldsymbol{p}) \tag{3.13}
$$

therefore,

$$
\mathrm{e}^{-S_\Lambda[\phi^-]} = \int \mathcal{D}\phi^+ \mathrm{e}^{-S_{\Lambda_0}[\phi^+ + \phi^-]} =
$$
$$
= \mathcal{N} \exp\left(-\int_{|\boldsymbol{p}|<\Lambda} \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2}\tilde{\phi}^-(-\boldsymbol{p})\big(\boldsymbol{p}^2 + m_0^2\big)\tilde{\phi}^-(\boldsymbol{p})\right) \tag{3.14}
$$

where $\mathcal{N}$ is some unimportant constant we'll ignore in what follows. This means the first step is trivial. Now we rescale the momenta to yield,

$$
S_\Lambda[\phi^-] = \int_{|\boldsymbol{p}|<\Lambda_0} \frac{\mathrm{d}^d p'}{(2\pi)^d} \left(\frac{\Lambda}{\Lambda_0}\right)^d \frac{1}{2}\tilde{\phi}^-(-\boldsymbol{p})\left(\left(\frac{\Lambda}{\Lambda_0}\right)^2 \boldsymbol{p}'^2 + m_0^2\right)\tilde{\phi}^-(\boldsymbol{p}) \tag{3.15}
$$

hence,

$$
Z_\phi = \left(\frac{\Lambda}{\Lambda_0}\right)^{d+2}, \text{ and } \tilde{\phi}'(\boldsymbol{p}') = \left(\frac{\Lambda}{\Lambda_0}\right)^{\frac{d+2}{2}} \tilde{\phi}^-(\boldsymbol{p}) \tag{3.16}
$$

the final action is (dropping primes)

$$
S_\Lambda[\phi] = \int_{|\boldsymbol{p}|<\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2}\tilde{\phi}(-\boldsymbol{p})\left(\boldsymbol{p}^2 + \left(\frac{\Lambda}{\Lambda_0}\right)^2 m_0^2\right)\tilde{\phi}(\boldsymbol{p}) \tag{3.17}
$$

---

[27]Oskar Benjamin Klein: born in 1894 in Mörby, Sweden; died in 1977 in Stockholm, Sweden; Doctoral Advisor: Svante Arrhenius

[28]Walter Gordon: born in 1893 in Apolda, Germany; died in 1939 in Stockholm, Sweden; Doctoral Advisor: Max Planck

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

$$m^2(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^2 m_0^2 \tag{3.18}$$

If $m^2 = 0$ the RG procedure has no effect, therefore this is a fixed point, the *Gaussian fixed point*. In these notes we'll never venture too far away from this point in theory space.

A common way to capture equations like (3.18) is to differentiate wrt to the logarithm of the cutoff, this is the *beta-function*

$$\beta_{m^2}(\Lambda) = \Lambda \frac{\mathrm{d}m^2(\Lambda)}{\mathrm{d}\Lambda} = -2m^2(\Lambda) \tag{3.19}$$

The coefficient on the RHS is negative, which means m^2 *increases* in the IR, therefore it is a relevant deformation.

So far so boring right? Let's spice things up a little bit. Consider a generic small deformation around this fixed point $\sim \phi^n$, more concretely we add the following term to the action in momentum space (in this calculation we'll only consider powers of ϕ, but it's not very hard to extend this analysis to the case where we also add derivatives of ϕ, the conclusions are pretty much unchanged)

$$\int \prod_{i=1}^{n} \left(\frac{\mathrm{d}^d p_i}{(2\pi)^d} \tilde{\phi}(\boldsymbol{p}_i)\right) \lambda_{0,n} \delta^{(d)}\left(\sum_{i=1}^{n} \boldsymbol{p}_i\right) \tag{3.20}$$

We'll still assume the first step to be trivial (we'll correct this assumption in the next section), what is the effect of the second and third steps? The measure brings out a factor of $\left(\frac{\Lambda}{\Lambda_0}\right)^{nd}$ and the delta-function gives a factor of $\left(\frac{\Lambda_0}{\Lambda}\right)^d$ during the second step. The third step contributes $\left(\frac{\Lambda_0}{\Lambda}\right)^{n\frac{d+2}{2}}$. The end result is

$$\lambda_n(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^{\left(1-\frac{n}{2}\right)d+n} \lambda_{0,n} \Rightarrow \beta_{\lambda_n}(\Lambda) = -\left(\left(1-\frac{n}{2}\right)d+n\right)\lambda_n \tag{3.21}$$

It's clear that the behaviour of the beta function is dependent not only on the number of fields n but also on the number of dimensions d. The following table summarises 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

For $d = 1, 2$ the coefficient is always negative, this means all operators[29] of this type are relevant and if we turn any of them on we quickly flow away from the Gaussian fixed point and our analysis breaks down. For $d = 3, 4$ we have some relevant operators, one marginal, and then an infinite number of irrelevant operators. Because the irrelevant operators die off quite quickly, we can generally not turn them on at all since they won't matter for low energy physics, this means we really only need to deal with a handful of interactions, which is a massive simplification. For $d = 3$ however, the quartic term is relevant so we quickly

---

[29]Although we're in the path integral, and these are just functions, it's commonplace to give the name of operators to these deformations of the action, reminiscent of the canonical perspective

flow away from the fixed point, for $d = 4$ we have some hope of staying near the fixed point and use some sort of perturbation theory, we'll analyse this in more detail in the next section. In all higher dimensions the mass term is the only relevant operator hence the low energy theory is always free.

Even from this very simple analysis we start to get some physics out of this, for example, the real reason you spent so much time studying ϕ^4 theory in your QFT course is because it is the most general theory you can write down, in 4 dimensions, at low energies. All other interactions are literally irrelevant. However, the structure behind all this is still a bit obscure. What if I told you all this was only dimensional analysis? To see this, let's rephrase the three steps of RG in a slightly different way. We shall have the need to do this many times throughout this course in order to fully get our heads around this very subtle field, and, as is always the case in physics, every approach will have it's advantages and disadvantages, what it enlightens and what it obscures, but the only way to really understand a topic is to tackle it in as many ways as possible. So let's do it.

You may have noticed not all steps of RG are born equal. The first step sounds ridiculously harder than the other two. In comparison they seem like mere technicalities, whilst all the physics is in the first step, the coarse graining. And, although there is some truth in this, we have just seen how they can actually already lead to some interesting physics. Let's bring them to the centre stage. To do this, we must first realise how the third step has contributions both from integrating out fields, the Z'_ϕ, and from scaling the spacetime variables, the factor of $\left(\frac{\Lambda}{\Lambda_0}\right)^{d+2}$. In the following I'll split these contributions apart, calling the contribution from the coarse graining the *non-trivial part*, and the contribution from scaling the spacetime the *trivial part*[30].

The next trick is to express every coupling in terms of a dimensionless coupling using the cutoff to absorb the mass dimension of the coupling. We'll denote mass dimension by square brackets, i.e. the mass dimension of λ_n is denoted by $[\lambda_n]$ (this is sometimes also called the engineering dimension). Since $[x] = -1$, looking at the gradient term of the action we conclude that

$$[\phi] = \frac{d-2}{2} \tag{3.22}$$

and therefore, the mass dimension of the coupling to ϕ^n is

$$[\lambda_n] = \left(1 - \frac{n}{2}\right)d + n \tag{3.23}$$

and we write

$$\lambda_n = \Lambda^{[\lambda_n]} g_n \tag{3.24}$$

where g_n is a dimensionless coupling and Λ is the current cutoff of the theory.

Now we perform RG, for convenience I'll go back to real space and write,

$$\int \mathrm{d}^d x \, \Lambda_0^{[\lambda_n]} g_{0,n} \phi^n(\boldsymbol{x}), \qquad \lambda_{0,n} = \Lambda_0^{[\lambda_n]} g_{0,n} \tag{3.25}$$

where we made explicit the connection with our previous notation in terms of dimensionful couplings. After this we do step 1 and non-trivial part of step 3, to get

$$\int \mathrm{d}^d x \, \Lambda^{[\lambda_n]} g_n(\Lambda)(\phi^-)^n(\boldsymbol{x}), \qquad \lambda'_n = \Lambda^{[\lambda_n]} g_n(\Lambda) \tag{3.26}$$

---

[30]I should point out, these names are not standard, it's merely for my convenience

notice we had to keep the current cutoff out front for consistency. Now we can do step 2 and trivial part of step 3 to get

$$\int \mathrm{d}^d x' \, \Lambda_0^{[\lambda_n]} g_n(\Lambda) \phi'^n, \qquad \lambda_n(\Lambda) = \Lambda_0^{[\lambda_n]} g_n(\Lambda) \tag{3.27}$$

Wait, this last step did nothing to the dimensionless coupling, it merely changed the cutoff factor. This means we only need to track the change in g_n to get the full beta function. Tracking the change in the dimensionless coupling effectively does the trivial bits automatically. More explicitly, the beta function is

$$\beta_{\lambda_n}(\Lambda) = \Lambda \frac{\mathrm{d}\lambda_n}{\mathrm{d}\Lambda} = \Lambda_0^{[\lambda_n]} \Lambda \frac{\mathrm{d}g_n}{\mathrm{d}\Lambda} = \Lambda_0^{[\lambda_n]} \beta_n(\Lambda) \tag{3.28}$$

In this way of parametrising the flow we only need to do steps 1 and non-trivial 3 to get the full beta functions. This may seem like a big simplification but it can be a bit tricky to figure out exactly where all the factors go with this method, doing the full thing may be less prone to mistakes.

Let's look at our previous calculations once more. By neglecting steps 1 and non-trivial 3, we get $\lambda_{0,n} = \lambda_n'$, which, in terms of the dimensionless couplings means

$$\Lambda^{[\lambda_n]} g_n = \Lambda_0^{[\lambda_n]} g_{0,n} \Leftrightarrow g_n(\Lambda) = \left(\frac{\Lambda_0}{\Lambda}\right)^{[\lambda_n]} g_{0,n} \tag{3.29}$$

making it obvious that

$$\beta_n = -[\lambda_n] g_n \tag{3.30}$$

which is precisely the result we had previously.

This means the behaviour of RG near the Gaussian fixed point is in part determined by the mass dimension of the coupling. In particular, it also sheds some light on the meaning of steps 2 and 3. They're not mere technicalities, they are capturing the flow due to the intrinsic mass dimension of the coupling. Even if there was no functional integration (i.e. if we were dealing with a classical, rather than quantum, field theory) RG wouldn't stay silent, this contribution would still be there, albeit with a slightly different interpretation: how do the couplings change as we scale spacetime, paying due attention to the scale of the field in order to be able to make comparisons. For all these reasons, this contribution is called the *classical* or *scaling* contribution.

However, this analysis seems to only make sense near the Gaussian fixed point. In fact, this is not true, it is actually generic behaviour near any fixed point. To see why let's imagine we're at a fixed point, not necessarily the Gaussian fixed point, this may be a highly non-trivial interacting fixed point. How do we know we're at a fixed point? Well, if we perform RG and the couplings don't change, this means the beta functions are zero, $\beta_n^* = 0$. Away from this point in theory space, they will generically be non-zero, and very complicated functions that mix all the couplings in a highly non-linear fashion. However, since they vanish at that special point, we can linearise around it, and then diagonalise the resulting matrix transformation. We end up with

$$\beta_{g_a} = -\Delta_{g_a} g_a \tag{3.31}$$

where we used a different index a for the couplings to emphasise that the operators that they couple with are generically not simple powers of the fields, they can be rather complicated. Let $\mathcal{O}_a$ denote such an operator. This means the interaction terms in the action will look like,

22

$$\int \mathrm{d}^d x \, g_a \mathcal{O}_a \tag{3.32}$$

In analogy to the Gaussian fixed point, we call Δ_{g_a} the *scaling dimension* of the coupling, or equivalently (and more commonly) $\Delta_{\mathcal{O}_a} = d - \Delta_{g_a}$ the scaling dimension of the operator. This scaling dimension does not need to coincide with the naive mass dimension of the operator (which is also, somewhat rudely, called the engineering dimension), and the difference between the two is the *anomalous dimension*. We can also use these scaling dimensions to classify the operators into relevant, irrelevant, or marginal:

- $\beta_{g_a} > 0 \Leftrightarrow \Delta_{g_a} < 0 \Leftrightarrow \Delta_{\mathcal{O}_a} > d$: irrelevant deformation

- $\beta_{g_a} < 0 \Leftrightarrow \Delta_{g_a} > 0 \Leftrightarrow \Delta_{\mathcal{O}_a} < d$: relevant deformation

- $\beta_{g_a} = 0 \Leftrightarrow \Delta_{g_a} = 0 \Leftrightarrow \Delta_{\mathcal{O}_a} = d$: marginal deformation

So we see that near any fixed point, RG behaves exactly like we were just doing a scaling transformation. The only catch is that the actual, physical scaling dimensions need not coincide with the naive mass dimensions. Indeed, experimentally, these scaling dimensions can be measured through *critical exponents*. Quite famously, these exponents seemed quite bizarre, sometimes even irrational! RG managed to correctly explain them, we were merely measuring the *scaling* dimensions rather than the *mass* dimension. Dimensional analysis is saved by the present of a new dimensionful quantity to play with: the cutoff.

But wait a moment, what about the kinetic operator? We have insisted to have a canically normalised kinetic term, so it naively seems like it doesn't run, by definition. This is kind of true, however, there is some physics left to be extracted. The reason is that the field strength renormalisation Z_ϕ depends on the cutoff itself. So, in principle, we can define some sort of beta function for Z_ϕ. Even though the coupling by itself has little physical meaning, due to the arbitrariness in the overall scale of the field, the beta function will have a meaning, as the anomalous dimension of the field.

First, what happens to the field if we just make some naive scale transformation $\boldsymbol{x} \to \boldsymbol{x}' = \frac{\Lambda}{\Lambda_0}\boldsymbol{x}$? Well it changes according to the mass dimension, i.e.

$$\phi \to \phi' = \left(\frac{\Lambda}{\Lambda_0}\right)^{-[\phi]} \phi = \left(\frac{\Lambda}{\Lambda_0}\right)^{\frac{2-d}{2}} \phi \tag{3.33}$$

However, a real change of scale in our theory is given not by that simple transformation, but by an RG flow, and in that case, the transformation is

$$\phi \to \phi' = \sqrt{Z_\phi}\phi = \left(\frac{\Lambda}{\Lambda_0}\right)^{\frac{2-d}{2}} \sqrt{Z_\phi'}\phi \tag{3.34}$$

where both Z_ϕ and Z_ϕ' are functions of Λ. Expanding Λ about Λ_0 (and noting that $Z_\phi'(\Lambda_0) = 1$ by definition) gives

$$\phi' = \left(1 + \left(\frac{d-2}{2} - \frac{1}{2}\Lambda\frac{\mathrm{d}\log Z_\phi'(\Lambda)}{\mathrm{d}\Lambda}\bigg|_{\Lambda=\Lambda_0}\right)\left(1 - \frac{\Lambda}{\Lambda_0}\right) + \dots\right)\phi \tag{3.35}$$

We see that we get the naive mass dimension plus an extra contribution which involves the derivative of the logarithm of the field strength renormalisation. This prompts us to the define the field anomalous dimension as

$$\gamma_\phi = -\frac{1}{2}\Lambda\frac{\mathrm{d}\log Z'_\phi(\Lambda)}{\mathrm{d}\Lambda} \tag{3.36}$$

which gives the interpretation for the flow of the field strength renormalisation. It gives the anomalous dimension for how the field itself scales as we change scale.

And once more, as you are probably sick of hearing at this point, we can and we have measured the field anomalous dimension, and RG agrees with Nature.

3.2. Perturbative Renormalisation

In the last section, we have discussed the structure of RG, and calculated explicitly the flow near the Gaussian fixed point. However, we have neglected steps 1 and non-trivial 3, which means we never had to perform any functional integration, it came down to simple scaling and dimensional analysis. In fact, for these calculations, we didn't even need to invoke any partition function.

Now, it's finally time to grow up, take the path integral seriously, and tackle the quantum corrections to the beta functions. But we have a problem, if we start in an arbitrary point in theory space, we are entirely lost. The reason being we have no idea how to actually calculate a generic path integral. As path integrals are concerned, we can only do two things analytically: Gaussian integrals, and perturbation theory. Therefore, to make progress, we will assume we start with a weakly coupled theory, i.e. we assume that $g_{0,n} \ll 1$, $n > 2$. In this way we can use the techniques developed in Appendix B to evaluate the path integral.

Since the behaviour of the flows will depend on the number of dimensions, we shall organise our calculation starting from higher dimensions and progressively lowering the number of dimensions.

3.2.1 $d > 4$

This case is rather trivial. If we stay at weak coupling we can trust the classical scaling analysis to determine the character of our deformation. Using the results from the last section, we conclude that all interactions are irrelevant. Hence, if we move away from the Gaussian fixed point, we quickly focus on the free theory, all interactions die out in the IR. The theory is always free in the IR and scaling always works.

3.2.2 $d = 4$

In four dimensions life is more interesting. All interactions are irrelevant except for the quartic one. This means we can drop most interactions and just start with the following action:

$$S_{\Lambda_0}[\phi] = \int \mathrm{d}^4x \left(\frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \frac{1}{2}m_0^2\phi^2 + \frac{\lambda_0}{4!}\phi^4 \right) \tag{3.37}$$

where, despite using continuum notation, we implicitly cut off modes with $|\boldsymbol{p}| > \Lambda_0$.

Now we split the fields into high and low energy modes, as we did before, the only difference is that this time the action doesn't factorise neatly. This means we must calculate the full path integral given by

$$\mathrm{e}^{-S_\Lambda[\phi^-]} = \int \mathcal{D}\phi^+\mathrm{e}^{-S_{\Lambda_0}[\phi^++\phi^-]} \tag{3.38}$$

Before we attempt calculating this object let's massage this expression a little bit. Notice that, just as before, when expanding $S_{\Lambda_0}[\phi^+ + \phi^-]$ in powers of the fields, since the supports

in momentum space of the high and low energy fields don't intersect, there is no mixing at the quadratic level. This means we can make the split,

$$S_{\Lambda_0}[\phi^+ + \phi^-] = S_{\text{free}}[\phi^+] + S_{\text{free}}[\phi^-] + S_{\text{int}}[\phi^+, \phi^-] \tag{3.39}$$

where

$$S_{\text{free}}[\phi] = \int \mathrm{d}^4x \left(\frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + m_0^2\phi^2 \right) \tag{3.40}$$

$$S_{\text{int}}[\phi^+, \phi^-] = \int \mathrm{d}^4x \, \frac{\lambda_0}{4!}(\phi^+ + \phi^-)^4 \tag{3.41}$$

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

$$\mathrm{e}^{-\Delta S_\Lambda[\phi^-]} = \int \mathcal{D}\phi^+ \mathrm{e}^{-S_{\text{free}}[\phi^+] - S_{\text{int}}[\phi^+, \phi^-]} \tag{3.42}$$

Recalling Appendix B, we see that this is exactly like the Wilsonian effective action where the low energy fields serve as sources for the high energy fields, which are the only ones that can propagate. Therefore, we can calculate this object perturbatively by summing over all connected Feynman diagrams with a set number of external ϕ^- fields. We shall represent the external ϕ^- fields with black lines, and we won't attach propagators to these lines. All internal lines correspond to ϕ^+ fields being integrated over, they have propagators associated with them, and we shall represent them with blue lines to emphasise the difference.[31]

- **No external fields:**
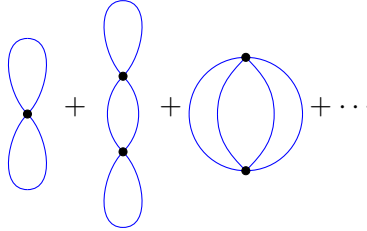
These diagrams are known as *vacuum bubbles*. They contribute with a scale dependent, but field *in*dependent, additive constant to the action. In a theory of gravity, these diagrams would have to be accounted for, since gravity couples to everything. And, in this context, this would be known as the *cosmological constant*. The name arises because, in our universe, this constant is very small, and hence, it is only relevant at cosmological scales. Correctly accounting for the smallness of the cosmological constant is one of the big unsolved problems in physics. For our purposes, however, this constant will not contribute to correlation functions, so we can neglect it.

- **Two external fields:**

The leading order contribution with two external fields is:

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

---

[31]Some authors use double line notation, or difference between solid and dotted lines. We prefer different colours since the other notations have other uses, respectively to denote matrix fields, and to distinguish between scalars and spinors

$$= -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - \Lambda^2 + m_0 \log\left(\frac{\Lambda^2 + m_0^2}{\Lambda_0^2 + m_0^2}\right)\right) \tag{3.43}$$

This means[32]

$$m'^2 = m_0^2 + \frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - \Lambda^2 + m_0 \log\left(\frac{\Lambda^2 + m_0^2}{\Lambda_0^2 + m_0^2}\right)\right) \tag{3.44}$$

Knowing that $g_2 = \Lambda^{-2}m'^2$, we could use (3.44) directly to calculate the beta function. However, it's a lot easier to go directly from the momentum integral, using the fact that

$$\Lambda\frac{\mathrm{d}}{\mathrm{d}\Lambda}\int_\Lambda^{\Lambda_0}\mathrm{d}k\, f(k) = -\Lambda f(\Lambda) \tag{3.45}$$

Either way, the end result is,

$$\beta_2 = \Lambda\frac{\mathrm{d}g_2}{\mathrm{d}\Lambda} = \Lambda\frac{\mathrm{d}}{\mathrm{d}\Lambda}(\Lambda^{-2}m'^2) = -2g_2 - \frac{\lambda_0}{16\pi^2}\frac{1}{1 + \Lambda^{-2}m_0^2} = -2g_2 - \frac{1}{16\pi^2}\frac{g_4}{1 + g_2} \tag{3.46}$$

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.

- **Four external fields:**

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

$$= -\lambda_0 + \frac{\lambda_0^2}{2}\int_\Lambda^{\Lambda_0}\frac{\mathrm{d}^4k}{(2\pi)^4}\left(\frac{1}{k^2 + m_0^2}\frac{1}{(k + p_1 + p_2)^2 + m_0^2}+\right.$$
$$+\frac{1}{k^2 + m_0^2}\frac{1}{(k + p_1 + p_4)^2 + m_0^2}+$$
$$\left.+\frac{1}{k^2 + m_0^2}\frac{1}{(k + p_1 + p_3)^2 + m_0^2}\right) \tag{3.47}$$

Hmm... this time the integral depends on external momenta, what does that mean? Since these integrals are all finite we can do a Taylor series in external momenta. We'll end up with terms that look like $\sim (\tilde{\phi}^-)^2 p^2 (\tilde{\phi}^-)^2$ which in real space correspond to $\sim (\phi^-)^2 \partial^2 (\phi^-)^2$. This signals the fact that when we start integrating out degrees of freedom, we will generate all possible terms consistent with the symmetries we impose, our simplification of considering only a handful of terms seems to break down. However, all these interactions are classically irrelevant, which means scaling will come to the rescue

---

[32]There is an annoying minus sign coming from the fact the Wilsonian action is defined as *minus* the logarithm of the partition function, if you look carefully at Appendix B you will spot the bastard

and these terms will be completely swamped by the classical contribution, quickly dying off. Therefore, we won't need to consider them.

Looking at the limit when $\boldsymbol{p}_a = \boldsymbol{0}$, $a = 1, \ldots, 4$, we get

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

$$= -\lambda_0 + \lambda_0^2 \frac{3}{32\pi^2} \left(\frac{m_0^2}{m_0^2 + \Lambda_0^2} - \frac{m_0^2}{m_0^2 + \Lambda^2} + \log\left(\frac{m_0^2 + \Lambda_0^2}{m_0^2 + \Lambda^2} \right) \right) \tag{3.48}$$

therefore

$$\lambda' = \lambda_0 \left(1 - \lambda_0 \frac{3}{32\pi^2} \left(\frac{m_0^2}{m_0^2 + \Lambda_0^2} - \frac{m_0^2}{m_0^2 + \Lambda^2} + \log\left(\frac{m_0^2 + \Lambda_0^2}{m_0^2 + \Lambda^2} \right) \right) \right) \tag{3.49}$$

You may be concerned that we are considering a $O(\lambda_0^2)$ correction here, while we only considered until $O(\lambda_0)$ in the mass term. However, the original term here is already $O(\lambda_0)$, the correction still only has an single *extra* power of λ_0. This feature is completely generic, we should always count loops instead of vertices when calculating a Wilsonian effective action. This can be rigorously shown using elementary graph theory but it's not very relevant for our purposes so we won't delve into it.

The beta function is then

$$\beta_4 = \Lambda \frac{\mathrm{d}g_4}{\mathrm{d}\Lambda} = \Lambda \frac{\mathrm{d}\lambda'}{\mathrm{d}\Lambda} = \frac{3}{16\pi^2} \frac{\Lambda^4 \lambda_0^2}{(\Lambda^2 + m_0^2)^2} = \frac{3}{16\pi^2} \frac{g_4^2}{(1 + g_2)^2} \tag{3.50}$$

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

All further interactions are irrelevant and hence are irrelevant for the rest of this section[33]. What can we conclude from this analysis?

Looking at the beta function for the quartic coupling, we see that it is positive, this means this coupling is actually *marginally irrelevant* rather than purely marginal. Therefore, $d = 4$ behaves similarly to $d > 4$ in the sense that all interactions are irrelevant, and in the IR we always flow to a free theory. This phenomenon is known as *triviality*. The only difference between $d = 4$ and $d > 4$ is that the quartic interaction dies off much slower in $d = 4$, it only decays logarithmically, instead of with a power law. We shall have much more to say about these kinds of decays and triviality later on when we discuss the continuum limit, but keep these features in mind.

---
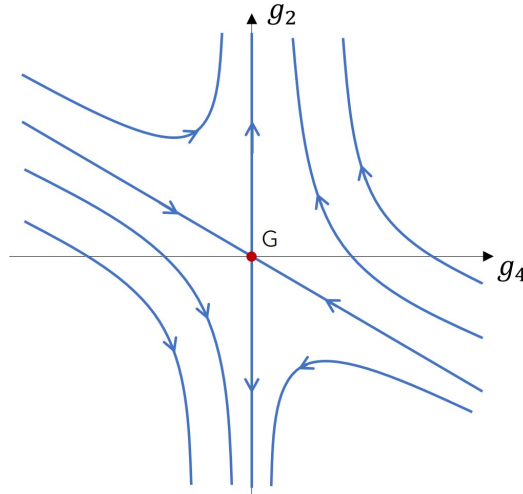
[33] Pun intended

Notice that, if you just turn on this coupling, you won't simply flow back to the Gaussian fixed point, there is some mixing between this coupling and the quadratic coupling. We will in fact flow to a free *massive* theory rather than the Gaussian fixed point. In order to hit the Gaussian fixed point we must turn on a particular combination of g_2 and g_4.

3.2.3 $d < 4$

We would like to complete our analysis and study the RG flow for $d = 1, 2, 3$. However, perturbation theory is of little help. All these theories feature relevant interactions, which grow as we flow to the IR, rendering our weak coupling approximation useless.

It turns out that the $d = 1, 2$ cases are exactly solvable. In $d = 1$ we can describe the flows in terms of ordinary quantum mechanics and tunnelling prevents us from reaching a fixed point. In $d = 2$ we can use sophisticated conformal field theory techniques to figure out there is actually an infinite number of fixed points. However, these details are beyond the scope of this course. To examine $d = 3$ we use a trick called the *epsilon expansion*. It consists of considering $d = 4 - \epsilon$ and doing perturbation theory for $\epsilon \ll 1$.

And by now you surely think I've gone mad. What the hell to I mean by taking the dimension to be $4 - \epsilon$? And haven't I noticed 1 is not actually $\ll 1$?

The answer to the first question is that we don't mean considering an action or a path integral in non-integer dimensions. We're just going to *formally* extend the integrals in perturbation theory, and the beta functions, to non-integer dimensions, and then expand in powers of ϵ. The answer to the second question is shhhhh. Or more explicitly, blah, blah, resummation, blah, blah.

It's straightforward to generalise the beta functions for arbitrary d, we get

$$\beta_2 = -2g_2 - \frac{\text{Vol}(S^{d-1})}{32\pi^4} \frac{g_4}{1 + g_2} \tag{3.51}$$

$$\beta_4 = (d - 4)g_4 + \frac{3\text{Vol}(S^{d-1})}{32\pi^4} \frac{g_4^2}{(1 + g_2)^2} \tag{3.52}$$

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

$$\beta_2 = -2g_2 - \frac{1}{16\pi^2} \frac{g_4}{1 + g_2} \tag{3.53}$$

$$\beta_4 = -\epsilon g_4 + \frac{3}{16\pi^2} \frac{g_4^2}{(1 + g_2)^2} \tag{3.54}$$

The remarkable thing about these beta functions is that there is an additional fixed point beyond the Gaussian fixed point, this is called the Wilson-Fisher[34] fixed point. To find it, we set the beta functions to zero and solve the resulting system of equations:

$$0 = -2g_2^* - \frac{1}{16\pi^2} \frac{g_4^*}{1 + g_2^*} \tag{3.55}$$

$$0 = -\epsilon g_4^* + \frac{3}{16\pi^2} \frac{g_4^{*2}}{(1 + g_2^*)^2} \tag{3.56}$$

---

[34]Michael Ellis Fisher: born in 1931 in Fyzabad, Trinidad and Tobago; Doctoral Advisor: Cyril Domb

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

$$g_2^* = -\frac{1}{6}\epsilon \tag{3.57}$$

$$g_4^* = \frac{16\pi^2}{3}\epsilon \tag{3.58}$$

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

$$g_2 = g_2^* + \delta g_2 \tag{3.59}$$

$$g_4 = g_4^* + \delta g_4 \tag{3.60}$$

and expand the beta functions up to linear order in δg_2, δg_4, and ϵ. The end result is

$$\Lambda \frac{\mathrm{d}}{\mathrm{d}\Lambda} \begin{pmatrix} \delta g_2 \\ \delta g_2 \end{pmatrix} = \begin{pmatrix} -2 + \frac{\epsilon}{3} & -\frac{1}{16\pi^2}(1 + \frac{\epsilon}{6}) \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} \delta g_2 \\ \delta g_4 \end{pmatrix} \tag{3.61}$$

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:
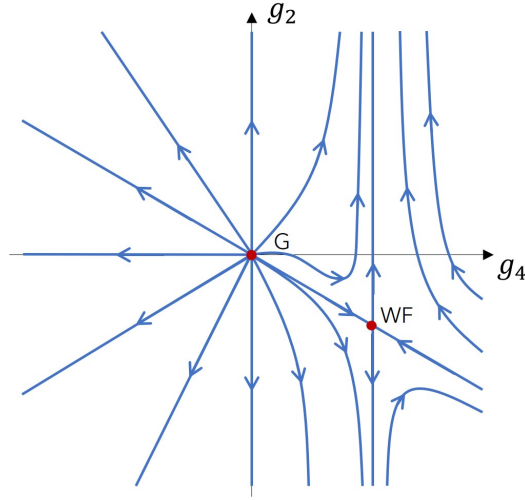

Figure 4: *RG flows for $d = 4 - \epsilon$*

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. This is still an asymptotically free direction. 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.

These flows are very important for two reasons. First of all this is a flow of a physical theory, the Ising model, or equivalently, gases in the liquid-gas phase transition. Both these systems exhibit second order phase transitions which characteristically are scale invariant, i.e. they correspond to fixed points. And, because the Gaussian fixed point is actually unstable, in the real world, we will most definitely hit the WF fixed point. This has many experimental consequences, and correctly explaining them was one the great successes of RG.

The second reason is that it illustrates many different features of RG flows. I've been emphasising all these names of triviality, asymptotic freedom and asymptotic safety because they are very important for high energy physics. The reason for this is that, if our world has no intrinsic scale, if spacetime is, in fact, continuous, then if we go up in the UV we must hit some fixed point, otherwise, at some point, the theory breaks down. However, the theories that describe the experiments at low energies are interacting, which means we should avoid theories that are entirely trivial, such as ϕ^4 in $d = 4$. We want either asymptotically free, or asymptotically safe theories. Most remarkably, quantum chromodynamics (QCD), the theory of strong interactions, is asymptotically free, which means it can be valid at arbitrarily high energies. It also means it is more weakly coupled the higher up we go in the energy scale, and it is very strongly coupled in the IR, which causes very severe theoretical challenges such as confinement and hadronisation. It is still an open problem whether the full standard model of particle physics has a fixed point in the UV.

Further, now you are capable of understanding one of the problems with quantum gravity. If we take the action for general relativity and apply the same procedure, we find that all interactions are irrelevant, which makes it very hard to analyse at very high energies and naively the whole theory breaks down. Many approaches to quantum gravity make it have a physical limit above which physics makes no sense to solve this issue. This basically stops the RG flow before it gets too bad. Another possibility is that GR actually comes from a highly non-trivial interacting fixed point in the UV, i.e. it is an asymptotically safe theory, this is the asymptotic safety program for quantum gravity. They try to characterise this fixed point using very sophisticated RG techniques and figure out whether or not we can construct such a theory. So far it is inconclusive which of these leads in the right path, only time will tell.

3.3. The Local Potential Approximation

So far we've been able to do two things: Gaussian integrals and perturbation theory. And while those are really powerful, there remains the question: can we do more? The short answer is: no. The long answer is: kinda. What we can do is use other approximations that are not weak coupling to get some results. However, the trick is to use those approxima-

tions to reduce the actual calculations to either Gaussian integrals, or Feynman diagrams. This whole program of trying to examine the RG flows when weak coupling is not a good approximation is called *exact renormalisation group*, and we'll go back to it when we discuss the continuum limit. But for now, let's just look at a simple example, to get a feel for how these things work. We won't extract much more physics than we already did just using perturbation theory, but we'll have a slightly different take on the derivation, which is always useful.

There are two approximations in what follows, the first one is merely for convenience, it can be dropped or changed without affecting the ideas too much, but it will make things easier for us, for the time being. It is the *local potential approximation*, and it amounts to considering that the effective action at the scale Λ is given by

$$S_\Lambda[\phi] = \int \mathrm{d}^d x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi)^2 + V(\phi) \right) \tag{3.62}$$

where $V(\phi)$ is a local function of ϕ, i.e. it doesn't depend on derivatives of ϕ. This is an approximation, as we've seen, in general, additional derivative interactions will be generated as we flow to the IR. For $d > 2$ they are irrelevant, so we can, hopefully, neglect them.

The second approximation is the crucial one, we give up on the idea of integrating out all modes between Λ and Λ_0. Instead, we focus on extracting just the beta functions. That is, we do an infinitesimal lowering of the cutoff $\Lambda = \Lambda_0 - \delta\Lambda$.

Doing the usual split $\phi = \phi^+ + \phi^-$ we get

$$S_{\Lambda_0}[\phi^+ + \phi^-] = S_{\Lambda_0}[\phi^-] +$$
$$+ \int \mathrm{d}^d x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi^+)^2 + \phi^+ V'(\phi^-) + \frac{1}{2} (\phi^+)^2 V''(\phi^-) + \frac{1}{6} (\phi^+)^3 V^{(3)}(\phi^-) + \dots \right) \tag{3.63}$$

To calculate this it seems we would have to resort to Feynman diagrams again, which wouldn't be too much of a help. But now our second approximation kicks in. Since $\Lambda = \Lambda_0 - \delta\Lambda$, then each high energy loop will come with an integral in momentum similar to

$$\int_{\Lambda_0 - \delta\Lambda}^{\Lambda_0} \frac{\mathrm{d}^d k}{(2\pi)^d} \cdots \approx \delta\Lambda \frac{\Lambda_0^{d-1}}{(2\pi^D)} \int_{S^{d-1}} \mathrm{d}\Omega \dots \tag{3.64}$$

where the last integral is over a unit $(d-1)$-sphere.

This means each loop is of order $\delta\Lambda$, then, in our approximation, we only need to consider linear order in $\delta\Lambda$, i.e. we only need to consider 1-loop diagrams[35], which is already a massive simplification. But instead of jumping in and calculating those diagrams, we can do even more progress using some elementary graph theory.

Consider a particular diagram with a number of vertices, v_i, with i powers of ϕ^+ and arbitrary powers of ϕ^-. Then, if that diagram has e edges and l loops, by Euler's identity, it obeys

$$e - \sum_i v_i = l - 1 \tag{3.65}$$

---

[35] You may be concerned by the fact that this argument does not seem to rule out tree diagrams. However, with a little bit more thought you can see that they also don't matter. We'll see shortly that, in the local potential approximation, the external momenta should be sent to zero, however, then there is no way any internal momenta can be high energy in a tree diagram, which means they also don't contribute

By our reasoning above we know that $l = 1$. Further, since all edges must be high energy, and there can be no external ϕ^+ lines, all edges must end in the the vertices, and then we have

$$2e = \sum_i i v_i \tag{3.66}$$

Putting everything together gives

$$\sum_i \frac{i-2}{2} v_i = 0 \tag{3.67}$$

This means that the only diagrams that can contribute are those whose vertices have exactly 2 ϕ^+ lines, therefore we can truncate (3.63) and just consider

$$S_{\Lambda_0}[\phi^+ + \phi^-] = S_\Lambda[\phi^-] + \int \mathrm{d}^d x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi^+)^2 + \frac{1}{2}(\phi^+)^2 V''(\phi^-) \right) \tag{3.68}$$

This is a Gaussian integral! However, $V''(\phi^-)$, which would play the role of mass, is spacetime dependent which complicates everything. It is time for our local potential approximation to come to the rescue. Because the whole point is not that we start with a local potential, that we can always do, there's no approximation there, the real approximation is to impose we end with a local potential. This truly is an approximation, that is, it isn't exactly true, it's just that those extra derivative interactions will be irrelevant for $d > 2$, so we're not too far off.

But how do we implement that? Thinking in terms of diagrams again, we want to set all external momenta to zero to get the momentum independent piece. This is equivalent to treating $V''(\phi^-)$ as if it wasn't spacetime dependent. Then we have a usual Gaussian integral, which we know how to do. For convenience let's rewrite (B.3).

$$Z[\boldsymbol{M}] = \int \mathrm{d}^n \phi\, e^{-\frac{1}{2}(\boldsymbol{\phi}\cdot\boldsymbol{M}\boldsymbol{\phi})} = \sqrt{\frac{(2\pi)^n}{\det \boldsymbol{M}}} \tag{3.69}$$

using the identity $\log(\det \boldsymbol{M}) = \mathrm{tr}(\log \boldsymbol{M})$, we can write

$$Z[\boldsymbol{M}] \sim \exp\left(-\frac{1}{2} \mathrm{tr}(\log \boldsymbol{M}) \right) \tag{3.70}$$

To apply this to our case, we go to momentum space (treating $V''(\phi^-)$ as a constant), then our analogue of $\boldsymbol{M}$ is

$$M(\boldsymbol{p}, \boldsymbol{p}') = \delta^{(d)}(\boldsymbol{p} + \boldsymbol{p}')(\boldsymbol{p}^2 + V''(\phi^-)) \tag{3.71}$$

The log is defined in terms of the eigenvalues of M, and the trace just means putting $\boldsymbol{p} = -\boldsymbol{p}'$ and integrating over $\boldsymbol{p}$. Therefore

$$\mathrm{tr}(\log(M)) = \int_{\Lambda_0 - \delta\Lambda}^{\Lambda_0} \frac{\mathrm{d}^d p}{(2\pi)^d} \delta^{(d)}(\boldsymbol{0}) \log\left(\boldsymbol{p}^2 + V''(\phi^-)\right) \tag{3.72}$$

Since $\delta\Lambda \ll \Lambda_0$ the integral is easy to do

$$\mathrm{tr}(\log(M)) = \delta\Lambda\, \Lambda_0^{d-1} \frac{\mathrm{Vol}(S^{d-1})}{(2\pi)^d} \delta^{(d)}(\boldsymbol{0}) \log\left(\Lambda_0^2 + V''(\phi^-)\right) \tag{3.73}$$

The $\delta^{(d)}(\boldsymbol{0})$ is also easy to interpret, it is a volume factor, or equivalently an integration over spacetime. In the end we have

$$\text{tr}(\log(M)) = \delta\Lambda\,\Lambda_0^{d-1}\frac{\text{Vol}(S^{d-1})}{(2\pi)^d}\int \mathrm{d}^d x\,\log\big(\Lambda_0^2 + V''(\phi^-)\big) \tag{3.74}$$

This means we have

$$S_\Lambda[\phi^-] = \int \mathrm{d}^d x\left(\frac{1}{2}(\boldsymbol{\nabla}\phi^-)^2 + V(\phi^-) + k\,\delta\Lambda\,\Lambda_0^{d-1}\log\big(\Lambda_0^2 + V''(\phi^-)\big)\right) \tag{3.75}$$

where

$$k = \frac{\text{Vol}(S^{d-1})}{2(2\pi)^d} = \frac{1}{(4\pi)^{\frac{d}{2}}\Gamma\big(\frac{d}{2}\big)} \tag{3.76}$$

To get the beta functions we need to do steps 2 and 3, which give us

$$S_\Lambda[\phi'] = \int \mathrm{d}^d x'\left(\frac{1}{2}(\boldsymbol{\nabla}'\phi')^2 + \left(\frac{\Lambda_0}{\Lambda}\right)^{d-n_\phi\frac{d-2}{2}} V(\phi') + \right.$$

$$\left. + k\,\delta\Lambda\,\Lambda_0^{d-1}\left(\frac{\Lambda_0}{\Lambda}\right)^{d-n'_\phi\frac{d-2}{2}}\log\big(\Lambda_0^2 + V''(\phi^-)\big)\right) =$$

$$= \int \mathrm{d}^d x'\left(\frac{1}{2}(\boldsymbol{\nabla}'\phi')^2 + V(\phi') + \delta\Lambda\left(\left(d - n_\phi\frac{d-2}{2}\right)\Lambda_0^{-1}V(\phi') + \right.\right.$$

$$\left.\left. + k\Lambda_0^{d-1}\log\big(\Lambda_0^2 + V''(\phi')\big)\right) + O(\delta\Lambda^2)\right) \tag{3.77}$$

where n_ϕ counts the number of fields inside V, how do we do that? Just think of it as an operator

$$n_\phi V(\phi) = \phi V'(\phi) \tag{3.78}$$

Keeping in mind the minus sign in $\Lambda = \Lambda_0 - \delta\Lambda$, we have the following expression for the beta function of the potential

$$\Lambda\frac{\mathrm{d}V(\phi)}{\mathrm{d}\Lambda} = -d\cdot V(\phi) + \frac{d-2}{2}\phi V'(\phi) - \Lambda^d\log\big(\Lambda^2 + V''(\phi)\big) \tag{3.79}$$

We could also expand V in a Taylor series and track the evolution of the individual couplings. We write

$$V(\phi) = \sum_{n=1}^\infty \Lambda^{d-n(d-2)}\frac{g_{2n}}{(2n)!}\phi^{2n} \tag{3.80}$$

By expressing V in terms of dimensionless couplings as usual we should not use (3.79) directly since it encapsulates scaling already. We should instead use the expression just with step 1. Plugging things in you eventually reach the conclusion that

$$\Lambda\frac{\mathrm{d}g_{2n}}{\mathrm{d}\Lambda} = (n(d-2) - d)g_{2n} - k\Lambda^{n(d-2)}\frac{\partial^{2n}}{\partial\phi^{2n}}\log\big(\Lambda^2 + V''(\phi)\big)\bigg|_{\phi=0} \tag{3.81}$$

Just for concreteness the first few terms in this expansion are (to confirm we haven't done anything too silly, it's also a straightforward exercise to find out which 1-loop diagrams would correspond to each term)

$$\Lambda \frac{\mathrm{d}g_2}{\mathrm{d}\Lambda} = -2g_2 - k\frac{g_4}{1 + g_2}$$

$$\Lambda \frac{\mathrm{d}g_4}{\mathrm{d}\Lambda} = (d-4)g_4 - k\frac{g_6}{1 + g_2} + 3k\frac{g_4^2}{(1 + g_2)^2} \tag{3.82}$$

$$\Lambda \frac{\mathrm{d}g_6}{\mathrm{d}\Lambda} = (2d-6)g_6 - k\frac{g_8}{1 + g_2} + 15k\frac{g_4 g_6}{(1 + g_2)^2} - 30k\frac{g_4^3}{(1 + g_2)^3}$$

Before we move on, I must address the elephant in the room. At this point, you may be a bit uneasy. First of all, this whole procedure seems rather ill justified[36], and, to some extent, it is. If we're too far away from the Gaussian fixed point we really can't trust the mass dimensions to tell whether the couplings are relevant or irrelevant. Therefore, strictly speaking, if we have too strong coupling there may be more subtle effects that make the derivative interactions matter. Also, we made all this argument in terms of diagrams, so there may be non-perturbative effects kicking in that we will never capture with diagrams. All this to say this isn't very mathematically rigorous, and we can't really claim these beta functions are the whole story, or even that they are non-perturbative in nature. The strongest statement we can do about these beta functions is that if we start with a theory without derivative interactions the beta functions for the local interactions are given by (3.81) to all orders in perturbation theory. More geometrically, if you imagine, in theory space, the space of all theories that have no derivative interactions as a plane, we know the projection of the beta functions parallel to this plane at any point in that plane, modulo non-perturbative effects.

Secondly, this doesn't seem like a huge gain to perturbation theory, like, why didn't we just use 1-loop Feynman diagrams to compute (3.82) once we new that just to compute beta functions we could restrict to 1-loop? Well, we did get every term in one go, which is nice. But that's not the real reason, there is hidden power in (3.79). This equation is really useful, both because it deals directly with the potential and not with the individual couplings, and because it is valid to all orders in perturbation theory, even away from weak coupling (with the caveats from the above paragraph). It is still very hard to solve, however, it is more amenable to plug into a computer. For example, this formulation of the flow has been used to search for fixed points directly in $d = 3$, without resorting to the epsilon expansion, and indeed one finds an interacting fixed point whose properties agree precisely with the prediction of the epsilon expansion for the Wilson-Fisher fixed point which gives some faith to both approaches.

The real conclusion is that, despite all its flaws, it is a take on the RG flow slightly different from standard perturbation theory. And the truth is, computing RG flows is bloody hard, there's no way around it. So, we grab everything we can get our hands on, to get different glimpses into the wilderness of theory space. In the next chapter we will collect a few more perspectives under our belt, more experimentally driven. And, historically, these perspectives actually came first, but they were clouded in misconceptions and confusion, and only with what we learned in this chapter about Wilson's view of the Renormalisation Group can we truly understand what is going on.

---

[36]Or, as we say in Portugal, *javardo*

4. Counterterms and the continuum limit

In the last chapters we have delved quite deeply into the role of scale in quantum field theory, how, to even define a theory, we need to introduce some scale into the game. We saw how our theories depend on this scale through the study of renormalisation group flows and even calculated explicitly these flows, in some simple cases, elucidating the structure of scalar field theory, and exemplifying important techniques. But we still have a lot of questions left to answer, and, historically, these were actually the first questions to be asked and answered. The structure we exposed previously took some more decades to fully unfold.

The first question is experimental. Yes, we have showed that the couplings in the Lagrangian change as we flow down to the IR. But is this just some theoretical slight of hand, or can we actually measure RG? And related to that, do we really need this scale or is it just because we're too dumb to write the theory without it? Finally, we shall close our lectures with a long promised proof, how does our theory depend on different choices of regularisation schemes? Are some schemes better than others, or is it all the same in the end?

4.1. The flow of correlation functions

We wish to connect with the real world of experiments. To do that, we need to connect RG flows to observable quantities. But what can we observe? Well, in QFT experiments, there are two things we can measure: cross sections and decay rates. These can be reduced to the calculation of scattering amplitudes (S-matrix elements), which, in turn, using the LSZ (Lehmann[37]-Symanzik[38]-Zimmermann[39]) reduction formula and Wick rotation, can be reduced to the calculation of Euclidean time correlation functions[40].

All this to say we'll consider correlation function as our observables, knowing there are standard methods to convert these into things experimentalists like. So let's calculate the flow of correlation functions! We'll begin with the simplest, the partition function. But before we jump into it, we need to discuss another thing first. What does the partition function depend on? Well, firstly, it cannot depend on the fields directly since we have integrated over them. What the partition function will depend on, is the many choices we had to make in defining our theory, namely, the various couplings (where mass is counted as a coupling), the spacetime manifold, and the field configuration space. In our case, we'll restrict the spacetime manifold to be Euclidean space with a flat metric. Further, the way we'll see the dependence on the field space is purely in terms of the cutoff (remember that the point of the cutoff is precisely to change which fields we're integrating over). Therefore we write

$$Z \equiv Z(\Lambda_0; g_{0,a}) \tag{4.1}$$

What happens when we change the scale? By construction, nothing can happen. We're only changing the order in which we integrate our fields, starting with the high energy ones. This means

$$Z(\Lambda; g_n(\Lambda)) = Z(\Lambda_0; g_{0,a}) \tag{4.2}$$

[37] Harry Lehmann: born in 1924 in Güstrow, Germany; died in 1998 in Hamburg, Germany; Doctoral Advisor: Friedrich Hund

[38] Kurt Symanzik: born in 1923 in Lyck, Germany; died in 1983 in Hamburg, Germany; Doctoral Advisor: Werner Heisenberg

[39] Wolfhart Zimmermann: born in 1928 in Freiburg im Breisgau, Germany; died in 2016 in Munich, Germany; Doctoral Advisors: Emanuel Sperner, Wilhelm Süss

[40] Any QFT textbook will have this in detail

or, infinitesimally,

$$\Lambda \frac{\mathrm{d}Z(\Lambda, g_a(\Lambda))}{\mathrm{d}\Lambda} = \left(\Lambda \frac{\partial}{\partial \Lambda} + \beta_a \frac{\partial}{\partial g_a}\right) Z(\Lambda; g_a(\Lambda)) = 0 \tag{4.3}$$

where we substituted $\beta_a = \Lambda \frac{\partial g_a}{\partial \Lambda}$.

This is the first example of a Callan[41]-Symanzik (CS) equation. It tells us that the couplings in the effective action change so as to precisely balance the change in the field configuration space (cutoff), and, in the end, leave the partition function unchanged.

Now, let's tackle correlation functions. Using the original theory with canonically normalised kinetic terms, and cutoff Λ_0, the n-point correlator is given by

$$\langle \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n)\rangle = \frac{1}{Z} \int \mathcal{D}\phi \; \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n) \mathrm{e}^{-S_{\Lambda_0}[\phi]} =$$
$$= \Gamma_{\Lambda_0}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_{0,a}) \tag{4.4}$$

If the fields are inserted at positions such that $|\boldsymbol{x}_i - \boldsymbol{x}_j| \gg \frac{1}{\Lambda}$, then we should be able to use the low energy effective theory to compute this correlator. However, we must be careful with field strength renormalisation, i.e.

$$\langle \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n)\rangle = \frac{1}{Z} \int \mathcal{D}\phi' \; \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n) \mathrm{e}^{-S_\Lambda[\phi']} \tag{4.5}$$

where, as usual, $\phi' = \sqrt{Z_\phi}\phi^-$ and we're using dimensionless couplings to avoid spatial rescalings. Then

$$\langle \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n)\rangle = \frac{1}{Z} Z_\phi^{-\frac{n}{2}}(\Lambda) \int \mathcal{D}\phi' \; \phi'(\boldsymbol{x}_1) \dots \phi'(\boldsymbol{x}_n) \mathrm{e}^{-S_{\Lambda_0}[\phi']} =$$
$$= Z_\phi^{-\frac{n}{2}}(\Lambda) \Gamma_\Lambda^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_a(\Lambda)) \tag{4.6}$$

Equating (4.4) and (4.6) gives

$$\Gamma_{\Lambda_0}^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_{0,a}) = Z_\phi^{-\frac{n}{2}}(\Lambda) \Gamma_\Lambda^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_a(\Lambda)) \tag{4.7}$$

or, infinitesimally,

$$\Lambda \frac{\mathrm{d}}{\mathrm{d}\Lambda} \left(Z_\phi^{-\frac{n}{2}}(\Lambda) \Gamma_\Lambda^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_a(\Lambda)) \right) = 0 \Leftrightarrow$$
$$\Leftrightarrow \left(\Lambda \frac{\partial}{\partial \Lambda} + \beta_a \frac{\partial}{\partial g_a} + n\gamma_\phi\right) \Gamma_\Lambda^{(n)}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; g_a(\Lambda)) = 0 \tag{4.8}$$

where we've used the field anomalous dimension, $\gamma_\phi = -\frac{1}{2}\Lambda \frac{\partial \log Z_\phi(\Lambda)}{\partial \Lambda}$.

This is another CS equation that tells us the couplings and field renormalisation precisely balance the changing scale to leave the observable correlation functions invariant. This makes perfect physical sense, since the theories are equivalent, any observable quantity must agree. It's more a matter of which version of the theory is more convenient to use.

But hold on a minute! This seems to say RG is a theoretical trick. A procedure to go from one theory to another to choose the most useful one, or to gauge which theories should we bother writing, in which the actual measurable quantities are unchanged. This

---

[41]Curtis Gove Callan Jr.: born in 1942 in North Adams, USA; Doctoral Advisor: Sam Treiman

is completely true, however, somewhat paradoxically, the reason why RG isn't measurable will tell us how to measure RG.

Let's go back a moment and understand how RG isn't measurable. After all, we have calculated that the mass of the field should depend on the scale for example. Can't experimentalists measure masses? Yes, they can perfectly measure masses, they just look for peaks in their histograms. Or, in our language, they measure 2-point functions as a function of the momentum, and then call "mass" the value of the momentum for which the correlator diverges. However, that point doesn't have to be the thing we called m. We have been saying mass all along, but really it was just the coupling for the quadratic term in the Lagrangian. We haven't really checked if that m was the actual, physical mass of the field.

Let's go back to theory space. In the last chapter, the point of view was that we have defined the theory at some scale Λ_0 by whatever means necessary. By defining a theory we specified the field configuration space (which will depend on Λ_0), the spacetime manifold, and the values of the couplings. Then, we used RG to get an equivalent effective description at some other sale Λ, which gave us some new values for couplings. This procedure tracks a trajectory in theory space. Now, obviously, we're free to parametrise this flow however we want. Instead of focusing on the UV theory given by divine intervention, we can focus on the IR observables, and use that to parametrise the flow. That is, instead of defining the couplings to compute amplitudes in terms of those couplings, we define the couplings in terms of some amplitudes. Then we can use those values to calculate more amplitudes, which could be, for example, the same amplitude but at a different energy scale.
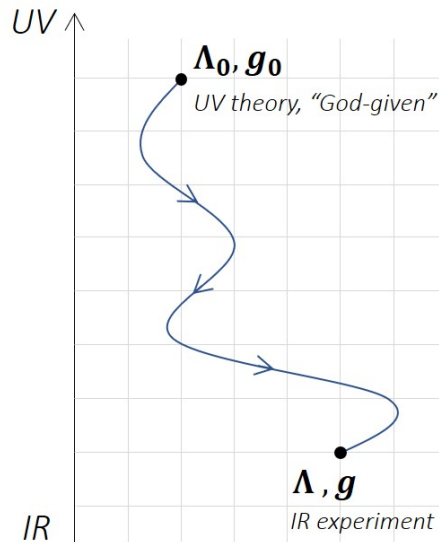
There is a catch however, we cannot invert RG. Once we've integrated out fields there is no way to integrate them back in. Those modes are gone. This seems like a fatal flaw for high energy physics. Sure you can use experiments at high energies to then calculate low energy amplitudes but that seems very much like our original picture, just with a possible change of basis. And strictly speaking, this is all we can do if we assume nothing more. This is very much the reason why we keep building bigger and bigger accelerators, we don't know what's behind the corner at higher energies.

However, we can be clever about it. Let's assume this theory extends beyond the energies we have so far probed. This may not be true, but let's assume it is. What we can then do is run RG from that hypothetical high energy down to our low energy experiments. With this information, we can then tune the high energy coupling such that the low energy

amplitude agrees with experiment. This seems like a silly idea. What have we gained from this? First of all, we'll see shortly that this is actually a very practical way to calculate RG flows, and to parametrise the flow in terms of observables. Further, this allows us to leave our comfort zone and ask the question "if this theory were valid up to higher energies, what would it look like?", then we can make predictions about high energy experiments, albeit predictions that can very well not be true, but it gives a guide as to what we're looking for. Finally, this point of view will lend itself quite naturally to the continuum limit. This high energy cutoff is almost begging to be removed at this stage. We'll purposefully delay that removal to the next section to emphasise the relevance and validity of these techniques even when we're not taking the continuum limit.

Now it's time to do some calculations. Although these techniques will be valid in more general circumstances, since we're taking the high energy physics point of view, which is more concerned with 4 dimensions, let's consider 4-dim ϕ^4 theory. Also, this is the theory where perturbation theory is more reliable, which simplifies our discussion.

We could just use the results from Chapter 3 and try to find the correct m_0 and λ_0 such that it reproduces some given amplitude by giving it some Λ_0 dependence. Although there's nothing wrong with that approach there is a simpler way to phrase the calculations. We split the action into two pieces, usually called the *bare* action, and the *counterterm* action.

$$S_{\Lambda_0}[\phi] = S_{\mathrm{B}}[\phi] + S_{\mathrm{CT}}[\phi] \tag{4.9}$$

where

$$S_{\mathrm{B}}[\phi] = \int \mathrm{d}^4x \left(\frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \frac{1}{2}m_0^2\phi^2 + \frac{\lambda_0}{4!}\phi^4 \right) \tag{4.10}$$

$$S_{\mathrm{CT}}[\phi] = \int \mathrm{d}^4x \left(\frac{1}{2}\,\delta Z_\phi\,(\boldsymbol{\nabla}\phi)^2 + \frac{1}{2}\,\delta m^2\,\phi^2 + \frac{\delta\lambda}{4!}\phi^4 \right) \tag{4.11}$$

and we assume that all dependence on the cutoff is in the counterterm action. We shall also assume that each counterterm contribution in a diagram counts as a loop. That is, tree-level contributions from the counterterms come at the same order as 1-loop contributions from the bare action. This may seem odd but there's no physics in this, it is purely a computational choice. After all, we're going to tune the counterterms anyway so they'll come at the order that we wish. You'll see in practice how this help.

There are many different ways we could fix the counterterms and the bare couplings depending on the context we're doing the calculations. These choices are called *renormalisation schemes*. The particular scheme we'll choose in this chapter is called the *on-shell* scheme since it's the one closer to experiment. But it's not the only valid choice, for other contexts there can be more convenient schemes.

We shall now go through both the 2- and 4-point functions in this theory. Both of them will have different annoying subtleties that we'll have to address, but, hopefully, in this simpler theory, we won't have muddled the physics too much. Just note a few details which are common to both. Firstly, we'll be calculating full correlation functions in the high energy theory, we're not integrating out modes, so our propagators will have momentum that goes all the way from 0 to Λ_0. Secondly, we'll consider only *connected* correlation functions, the reason for this is that the disconnected pieces are fairly trivial. Physically, they correspond to when particles don't "see" each other, it's once again part of the standard bag of techniques to go from connected correlators to physical amplitudes. With those points in mind let's do the calculations.

4.1.1 2-Point Function

The irritating technicality with this correlator is the fact that the propagator has the mass in the denominator instead of the numerator. Therefore we need to do some trickery. In general, the diagrams that contribute to the 2-point function, even beyond leading order are (excluding counterterms for the time being)

$$(4.12)$$

We see that some of the diagrams are basically copies of each other. For example, the third term looks like two copies of the second. This suggests we should be able to only consider "distinct" diagrams. We say that a diagram is one-particle irreducible (1PI) if it has no bridges, where a bridge is a line that once removed makes the diagram disconnected. For example, the second is 1PI but the third isn't, because, if we remove the middle propagator, we get two disconnected diagrams (two copies of the second term). We can then write a series for the 1PI diagrams with two external legs:

$$(4.13)$$

We'll call this expansion $\Pi(\boldsymbol{p}^2)$ ($\boldsymbol{p}$ is the external momenta, and we have amputated the two external propagators), this is often called the *self-energy* of the field. Using this notation the propagator is then:

$$
\begin{aligned}
&= \frac{1}{\boldsymbol{p}^2 + m_0^2} + \frac{1}{\boldsymbol{p}^2 + m_0^2}\Pi(\boldsymbol{p}^2)\frac{1}{\boldsymbol{p}^2 + m_0^2} + \frac{1}{\boldsymbol{p}^2 + m_0^2}\Pi(\boldsymbol{p}^2)\frac{1}{\boldsymbol{p}^2 + m_0^2}\Pi(\boldsymbol{p}^2)\frac{1}{\boldsymbol{p}^2 + m^2} + \dots \\
&= \frac{1}{\boldsymbol{p}^2 + m_0^2 - \Pi(\boldsymbol{p}^2)}
\end{aligned}
\tag{4.14}
$$

where in going to the last line we summed the geometric series.

Now all we're left to do is calculate $\Pi(\boldsymbol{p}^2)$. To 1-loop order it is given by (including counterterms)

$$
\Pi(\boldsymbol{p}^2) = \underbrace{}_{} + \frac{-\boldsymbol{p}^2\,\delta Z_\phi}{\otimes} + \frac{-\,\delta m^2}{\otimes} =
$$

$$
= -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m_0^2\log\left(1 + \frac{\Lambda_0^2}{m_0^2}\right)\right) - \boldsymbol{p}^2\,\delta Z_\phi - \delta m^2
\tag{4.15}
$$

where we've used the results from (3.43), setting Λ to 0.

It is finally time to go to experiment to fix δZ_ϕ and δm^2. Experimentally, there is a simple pole when the external momentum is $\boldsymbol{p}^2 = -m_{\text{phys}}^2$, further, we require that pole to

have unit residue to fix the field strength renormalisation. But wait a moment! We're in Euclidean signature, $\boldsymbol{p}^2$ is positive, so it doesn't seem very reliable to set it equal to *minus* the square of a real number. But there's an easy solution to that, you just Wick rotate back to Lorentzian[42] signature, which, if you use the $(-, +, +, +)$ sign convention[43], means you have to do absolutely nothing! The expression looks exactly the same, except all vectors are now vectors in Minkowski[44] space, which just means they'll loose the boldface. And it makes perfect sense that in order to compare to experiment we'd need to go back to the physical Lorentzian world.

In the end, in terms of the self-energy, the two necessary requirements are

$$m_0^2 - \Pi(-m_{\text{phys}}^2) = m_{\text{phys}}^2 \tag{4.16}$$

$$\left.\frac{\partial \Pi(p^2)}{\partial p^2}\right|_{p^2 = -m_{\text{phys}}^2} = 0 \tag{4.17}$$

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

$$\delta Z_\phi = 0 \tag{4.18}$$

$$\delta m^2 = m_{\text{phys}}^2 - m_0^2 - \frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m_0^2 \log\left(1 + \frac{\Lambda_0^2}{m_0^2}\right)\right) \tag{4.19}$$

we can further choose $m_0^2 = m_{\text{phys}}^2 = m^2$ to get the requirement $\Pi(-m^2) = 0$ and

$$\delta m^2 = -\frac{\lambda_0}{32\pi^2}\left(\Lambda_0^2 - m^2 \log\left(1 + \frac{\Lambda_0^2}{m^2}\right)\right) \tag{4.20}$$

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.

Note, however, than we can still extract the mass beta function from this calculation. The reason being that the quadratic coupling is now $m^2 + \delta m^2$, and we have fully specified its dependence on the cutoff such that it yields the physical mass m^2. This physical mass is just a result of some measurement, therefore it can't depend on the cutoff Λ_0, all cutoff dependence is explicit. More concretely, we have, to leading order in λ_0,

$$\beta_2 = \Lambda_0 \frac{\mathrm{d}g_2}{\mathrm{d}\Lambda_0} = \Lambda_0 \frac{\mathrm{d}}{\mathrm{d}\Lambda_0}[\Lambda_0^{-2}(m^2 + \delta m^2)] = -2g_2 - \frac{g_4}{16\pi^2}\frac{1}{1 + g_2} \tag{4.21}$$

where we take $\frac{\mathrm{d}m^2}{\mathrm{d}\Lambda_0} = 0$ and $g_4 = \lambda_0$.

As a final comment, notice that this calculation hinged on setting $p^2 = -m^2$, hence if m^2 is higher than the cutoff energy then this step makes no sense, since we don't trust the amplitude up to those energies. In this case, physically, we simply haven't probed high

---

[42]Hendrik Antoon Lorentz: born in 1853 in Arnhem, Netherlands; died in 1928 in Haarlem, Netherlands; Doctoral Advisor: Pieter Rijke; Nobel Prize in Physics 1902 together with Pieter Zeeman "in recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena"
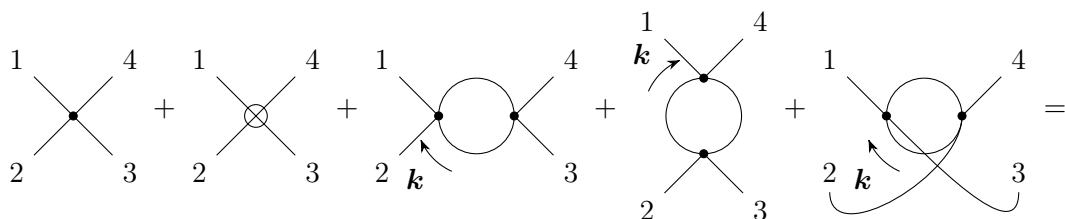
[43]Also known as the correct sign convention.

[44]Hermann Minkowski: born in 1864 in Aleksotas, Russian Empire; died in 1909 in Göttingen, German Empire; Doctoral Advisor: Ferdinand von Lindemann

enough energies to be able to measure the mass. And, looking at the propagator, notice that the k^2 term is negligible in comparison with the m^2 term, if m^2 is above the cutoff, which means that the field essentially can't propagate. This is the reasoning behind the common practice of integrating out fields whose masses are well above the energies we can probe. Even though this may seem like a somewhat brutal approximation, it has some merits. For example, by integrating out the W and Z bosons we can go from the standard model to Fermi's theory of weak interactions, which is a very useful approximation and was very important historically.

4.1.2   4-Point Function

The diagrams that contribute to this correlator are (up to 1-loop order)

$$= -\lambda_0 - \delta\lambda + \frac{\lambda_0^2}{2} \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \left(\frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_2)^2 + m^2} + \right.$$
$$+ \frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_4)^2 + m^2} +$$
$$\left. + \frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_3)^2 + m^2} \right) \tag{4.22}$$

where we have used the results from the previous section to use the physical mass m.

Now we have some dependence on the external momenta, which makes this correlator much more interesting than the previous one. We will at last understand how to measure RG. But first we need to evaluate this integral, keeping the dependence on the external momentum. To do that, we'll need the following trick due to Feynman (which you can check by explicit computation of the RHS):

$$\frac{1}{AB} = \int_0^1 \mathrm{d}x \, \frac{1}{(xA + (1-x)B)^2} \tag{4.23}$$

which allows us to combine two propagators, e.g. (where we define $\boldsymbol{p}_{12} = \boldsymbol{p}_1 + \boldsymbol{p}_2$)

$$\frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_{12})^2 + m^2} = \int_0^1 \mathrm{d}x \, \frac{1}{\left[x\left((\boldsymbol{k} + \boldsymbol{p}_{12})^2 + m^2 \right) + (1-x)(\boldsymbol{k}^2 + m^2) \right]^2} =$$
$$= \int_0^1 \mathrm{d}x \, \frac{1}{\left[\boldsymbol{k}^2 + m^2 + x\left(2(\boldsymbol{k} \cdot \boldsymbol{p}_{12}) + \boldsymbol{p}_{12}^2 \right) \right]^2} =$$
$$= \int_0^1 \mathrm{d}x \, \frac{1}{\left[(\boldsymbol{k} + x\boldsymbol{p}_{12})^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2 \right]^2} \tag{4.24}$$

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

$$\int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{\boldsymbol{k}^2 + m^2} \frac{1}{(\boldsymbol{k} + \boldsymbol{p}_1 + \boldsymbol{p}_2)^2 + m^2} =$$
$$= \int_0^1 \mathrm{d}x \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{1}{\left[\boldsymbol{l}^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} \tag{4.25}$$

In rigour, 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. And anyway, we can only really trust this theory if we probe it at energies much lower than the cutoff. Hence, we will approximate and integrate $\boldsymbol{l}$ over the region $|\boldsymbol{l}| < \Lambda_0$. This is mainly for theoretical convenience, in practice you could always to the full integral numerically.

Therefore we write

$$\int_0^1 \mathrm{d}x \int^{\Lambda_0} \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{1}{\left[\boldsymbol{l}^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} =$$
$$= \int_0^1 \mathrm{d}x \frac{\mathrm{Vol}(S^3)}{(2\pi)^4} \int_0^{\Lambda_0} \mathrm{d}l \frac{l^3}{\left[l^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2\right]^2} =$$
$$= \int_0^1 \mathrm{d}x \frac{1}{16\pi^2} \left(\log\left(\frac{\Lambda_0^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2}{m^2 + x(1-x)\boldsymbol{p}_{12}^2} \right) + \frac{m^2 + x(1-x)\boldsymbol{p}_{12}^2}{\Lambda_0^2 + m^2 + x(1-x)\boldsymbol{p}_{12}^2} - 1 \right) \tag{4.26}$$

Doing similar manipulations with the other three integrals, and keeping only the terms that don't vanish in the limit $\Lambda_0 \to \infty$, since those are the only ones we have reliably computed in our approximation, we get for the full 4 point amplitude:

$$\mathcal{A}(s,t,u) = -\lambda_0 - \delta\lambda + \frac{\lambda_0^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)s} \right) + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)t} \right) + \right.$$
$$\left. + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)u} \right) - 3 \right) \tag{4.27}$$

where we've Wick rotated to Lorentzian signature, and used the Mandelstam[45] variables

$$s = -p_{12}^2, \qquad t = -p_{14}^2, \qquad u = -p_{13}^2 \tag{4.28}$$

which, due to Lorentz invariance, completely specify four particle scattering kinematics.

Now, imagine we performed some experiment with kinematics s_0, t_0, u_0, we can then define the physical coupling at that energy scale to be

$$\lambda_{\mathrm{phys}} = -\mathcal{A}(s_0, t_0, u_0) \tag{4.29}$$

which means

$$\delta\lambda = \lambda_{\mathrm{phys}} - \lambda_0 + \frac{\lambda_0^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)s_0} \right) + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)t_0} \right) + \right.$$
$$\left. + \log\left(\frac{\Lambda_0^2}{m^2 - x(1-x)u_0} \right) - 3 \right) \tag{4.30}$$

---

[45] Stanley Mandelstam: born in 1928 in Johannesburg, South Africa; died in 2016 in Berkeley, USA; Doctoral Advisor: Paul Taunton Matthews

Once again we can choose $\lambda_{\text{phys}} = \lambda_0 = \lambda$ for simplicity[46]. For theoretical purposes it is often convenient to set $s_0 = t_0 = u_0 = 0$, even though this point isn't actually experimentally accessible. In this case we are essentially parametrising the RG flow in terms of the asymptotic value of the coupling infinitely far into the IR.

As we did for the quadratic coupling we can calculate the beta-function.

$$\beta_4 = \Lambda_0 \frac{\mathrm{d}}{\mathrm{d}\Lambda_0}(\lambda + \delta\lambda) = \frac{3g_4^2}{16\pi^2} \tag{4.31}$$

where we substituted $g_4 = \lambda$.

In this case we don't get the exact same result as we did before, this is due to our approximate way of calculating the integral, that is, we assumed that Λ_0 was much bigger than any other energy scale, which, given our caveats in the previous section, also means $g_2 \ll 1$, and taking the $g_2 \ll 1$ limit on the beta function from Chapter 3 gives us (4.31). In this limit we still recover the essential physics, namely that this beta function is positive, and therefore this is a marginally irrelevant operator.

Note that this beta function does not depend on the particular point s_0, t_0, u_0 used to define the coupling. This makes perfect physical sense, we're asking about how the high energy theory coupling depends on the cutoff in order to yield the same IR physics, it shouldn't depend on where exactly are we fixing the IR coupling.

So far this is looking very similar to the 2-point function story, but now comes the cool part. The full amplitude, after plugging in the chosen counterterm is:

$$\mathcal{A}(s,t,u) = -\lambda + \frac{\lambda^2}{32\pi^2} \int_0^1 \mathrm{d}x \left(\log\left(\frac{m^2 - x(1-x)s_0}{m^2 - x(1-x)s} \right) + \log\left(\frac{m^2 - x(1-x)t_0}{m^2 - x(1-x)t} \right) + \right.$$
$$\left. + \log\left(\frac{m^2 - x(1-x)u_0}{m^2 - x(1-x)u} \right) \right) \tag{4.32}$$

There are several things we can note about this amplitude. Firstly, we can confidently give this to an experimentalist. Everything is perfectly well defined in terms of experimentally measurable (i.e. physical) quantities. Secondly, the coupling λ will depend on the precise energies (s_0, t_0, u_0) we used to define it, however, the amplitude itself can't change, it's a physically measurable quantity, we're just parametrising it differently. This means we can write a kind of Callan-Symanzik equation, but now with respect to the low energy scales. For simplicity, let's set $s_0 = t_0 = u_0 = \mu^2$ (this is not necessary but it makes the expressions neater), then we can write

$$\mu \frac{\mathrm{d}\mathcal{A}}{\mathrm{d}\mu} = \left(\mu \frac{\partial}{\partial\mu} + \mu \frac{\partial\lambda}{\partial\mu} \frac{\partial}{\partial\lambda} \right) \mathcal{A} = 0 \tag{4.33}$$

where we can define a sort of experimental/low energy beta function as

$$\beta_\lambda^{\text{exp}} = \mu \frac{\partial\lambda}{\partial\mu} \tag{4.34}$$

Using (4.33) we can calculate this (to leading order in λ),

$$\beta_\lambda^{\text{exp}} = \frac{3\lambda^2}{16\pi^2} \int_0^1 \mathrm{d}x \frac{-x(1-x)\mu^2}{m^2 - x(1-x)\mu^2} \approx \frac{3\lambda^2}{16\pi^2} \tag{4.35}$$

---

[46]In fact, even though they're conceptually different quantities, in general we can always make this choice of putting the physical coupling as the parameter in the Lagrangian. This choice is sometimes known as *renormalised perturbation theory*, whereas keeping λ_0 unfixed is known as *bare perturbation theory*. Both names are completely uninformative

where in the last step we approximated $\mu^2 \gg m^2$ which is true if we're probing energies much higher than the mass of the particle.

We have now computed two beta functions which look rather similar both in shape and in interpretation but they are telling us different things. First, their similarities arise because they both obey a CS equation, and, due to dimensional analysis, there are only so many ways you can play with a dimensionful parameter. They also have very similar interpretation, they are both capturing the variation of the coupling such that the physics stays the same. These similarities are a big source of confusion in the literature, but they are in fact different concepts. So let us delve into it.

The low energy beta function is, in some sense, the more physical one. This one captures the dependence of the amplitude with the energy, which means it can be measured. And in fact, for theories which correspond to physical models such as QCD or QED, these have been measured, verifying the theoretical predictions to an astonishing accuracy! For example, the electron charge (or equivalently the fine structure constant) is the relevant coupling for the interaction between photons and electrons in QED and is predicted to be a marginally irrelevant parameter just like $\lambda\phi^4$. Sure enough this dependence has been measured and the electron charge is smaller at higher energies!

Further, note that you can't use this to play tricks with perturbation theory. Since μ^2 is basically a free parameter, it seems like, in principle, we could set it so that the coupling is very small, helping perturbation theory. However, if then you're probing energies too different from that μ^2, the logarithms that appear inside the integral will grow, ruining everything. So you really need to use physical measurements for this. This doesn't mean we have somehow lost predictability, we can still predict the dependence of the amplitude with the energy scale, and further we can use these amplitudes to calculate higher point amplitudes.

The original high energy beta function is a bit trickier. At first glace it appears to be completely unphysical. It is merely a theoretical construct to compensate for the UV cutoff needed to define the theory. And sure enough, in the final result the cutoff dependence completely disappeared. However, this result was only valid for low energies, far away from the cutoff. If we started to probe energies closer to the cutoff we would most likely start to see it's effects, usually as non-localities in the theory. And at that point we would have to re-examine our beliefs, because it can either be that the theory is the ultimate theory and we're dealing with a discrete spacetime[47], or perhaps there is new physics that comes into play, like a new very massive field. So even though it is not directly measurable it still has a lot of important physics.

4.2. The continuum limit

In the last section we have constructed an algorithm to phrase perturbation theory in terms of physically measurable quantities, which involved the use of counterterms to tune the high energy couplings such that it will reproduce the necessary low energy physics. This is all fine and dandy but along the line something miraculous happened. The end result was independent of the UV cutoff. This means that, at the end of the calculation, we can set it to be whatever we want it to be without changing the actual measurable predictions. What if we set it to infinity? There seems to be no problem with this. OK, and what if instead of waiting all this time to phrase things in terms of physically measurable quantities I just try to work with a theory without a UV cutoff? For instance, let's look at the 2-to-2 particle scattering, the integrand goes like

---

[47]This is not suggesting some esoteric quantum gravity theory, for example, it can be a theory to describe some condensed matter system, which usually has an underlying lattice.

$$\int \mathrm{d}^4 k \, \frac{1}{k^4} \sim \int \mathrm{d}k \, \frac{1}{k} \tag{4.36}$$

which diverges if we have no cutoff!

This is one example of the famous infinities that seemingly plague quantum field theory. But with all the tools we have developed, it won't be very hard to deal with them. In this case it is obvious what went wrong, even for ordinary QM we can't just take the continuum limit carelessly, we must be very careful in our calculations and only at the end take the appropriate limit. The cool thing is that if we're careful, computing things in a well defined theory, phrase everything in terms of physically measurable quantities, and wait until the very end to take the $\Lambda_0 \to \infty$ limit we get a sensible answer. However, it is not at all obvious this behaviour is generic. Sure we can introduce counterterms to parametrise the RG flow, but who is to say the final answer doesn't depend on the cutoff? If the theory is fundamentally discrete, for example, there's no conceptual uneasiness in having physical answers depending on the precise lattice spacing.

It turns out it isn't a generic behaviour. Some theories are "nice", and, after introducing the relevant counterterms, we can remove the cutoff and still get finite answers, and some theories are "less nice" and we we can't completely remove the cutoff dependence. At this point it is helpful to use a slightly less careful nomenclature when discussing these issues. Instead of focusing on the cutoff and talking about whether or not we can remove the cutoff dependence, it is easier to think in terms of divergences. That is, we compute diagrams (including counterterms) with a finite cutoff, and we talk about whether the final amplitude diverges or not as we take the infinite cutoff limit. Then the counterterms play a double role, they are both trying to absorb that infinity and to phrase the answer in terms of physical quantities. Note that this is just a change of words, if you feel uneasy about this mindset of removing infinities, you can always go back to the previous picture of always having a finite cutoff and doing the limit carefully just at the end.

At this point, all QFT textbooks go through a rather involved discussion about how to tell whether or not a diagram will diverge and how bad is the divergence. It involves lots of subtleties with graph theory and the fact that some diagrams appear to converge but then they have a subgraph that is actually divergent. It is all incredibly messy, and the worst thing is, even after all that hard work, most textbooks don't end up using this technology at all! The reason being the actual proof of the theorem is so incredibly fiddly that it is completely not worth the effort.

In my opinion, that discussion of subgraphs and divergences doesn't really help in understanding the final theorem, so we'll skip it and go straight to the end result. If you want more details you can find them in any self-respecting QFT textbook. The end result is encapsulated in the following theorem due to Bogoliubov[48], Parasiuk[49], Hepp[50], 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*

---

[48] Nikolay Nikolayevich Bogolyubov: born in 1909 in Nizhny Novgorod, Russian Empire; died in 1992 in Moscow, Russian Federation; Doctoral Advisor: Nikolay Krylov

[49] Ostap Stepanovich Parasyuk; born in 1921 in Belka, Republic of Poland; died in 2007 in Kyiv, Ukraine

[50] Klaus Hepp: born in 1936 in Kiel, Germany; Doctoral Advisor: Res Jost

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*

This is a crucial result in perturbative QFT, but it is also one of the most controversial (and arguably misinterpreted). There are several claims made about the nature of this theorem, some are a bit misguided and others are just flat out wrong, we'll enumerate a few below to help navigate the literature (especially when older).

- *Non-renormalisable theories have infinities which cannot be removed*: This is one the of the "flat out wrong" type, we can perfectly take care of the infinities in non-renormalisable theories, we just need an infinite number of couplings to do it.

- *Non-renormalisable theories are not predictive*: The idea behind this is the interpretation that to fix each counterterm we need a relevant experiment, then, since we need an infinite number of counterterms we would need an infinite number of experiments to be able to predict anything. This is technically true but only if you want to take the continuum limit. You can perfectly truncate to just a finite number of terms, keeping the cutoff dependence, then scaling will tell us the higher order terms will matter less and less for lower energies. At that level we can perfectly make predictions using non-renormalisable theories, you just have an additional parameter, the cutoff. Arguably they are even more predictive than renormalisable theories, since renormalisable theories are almost completely insensitive to UV physics, whereas non-renormalisable theories have much sharper bounds on where new physics comes into play, which is around the scale of the cutoff. And this has been widely exploited in the past for the discovery of the W, Z, and the Higgs bosons.

- *Non-renormalisable theories are inconsistent*: This was a belief that was held for a long time but is now rather out dated. Usually people complain that either locality or unitarity is broken in these theories. And to some extent this is true. However, it is not because of some fundamental inconsistency, it is merely the statement that these theories appear to not have a continuum limit. But this is perfectly fine so long as we have some fundamental reason behind it (like a lattice) or if we just take them as effective field theories at low energy. It is a bit hubris to think our puny theory is valid up to arbitrarily large scales, it is perfectly fine to deal with theories that don't have a continuum limit

- *The BPHZ theorem determines which theories have a sensible continuum limit*: This one is not entirely wrong, it's just a bit too quick of a conclusion. First of all, it is fundamentally perturbative in nature. Secondly, it deals with the consistency of a particular algorithm. All it states is for which theories will the counterterm algorithm work and remove the cutoff dependence, and for which theories it doesn't work. It doesn't mean the theory itself makes or doesn't make sense in the continuum.

Essentially it is only asking how sensitive your theory is to the UV physics and classifying whether theories are less sensitive (renormalisable) or more sensitive (non-renormalisable). We'll see below how a theory can be renormalisable, but at the same time, even at the perturbative level appear not to have a sensible continuum limit.

A final note before moving on. Not everyone agrees with the classification stated above. Some people call what we've been describing as renormalisability the *power counting* notion of renormalisability, preferring the term renormalisable to mean "theories with a continuum limit". This is partially to address the misconceptions stated above, and is a perfectly fine distinction, one just has to be careful when navigating different conventions.

Although the BPHZ theorem is a good start when we want to ask the question of the continuum limit, as we stated above, it isn't the whole story. The real world is far more complex than that. So then, which theories do have a continuum limit? To be entirely honest, the answer is not known in its full generality. Mostly what we have is very strong evidence pointing one or the other way but very rarely do we have an actual loophole-free proof.[51] However, we can still make some progress and understand which class of theories would have a continuum limit, or said another way, how to spot if a theory has a continuum limit. To do that, let's bring back our old picture of the RG flows in theory space.
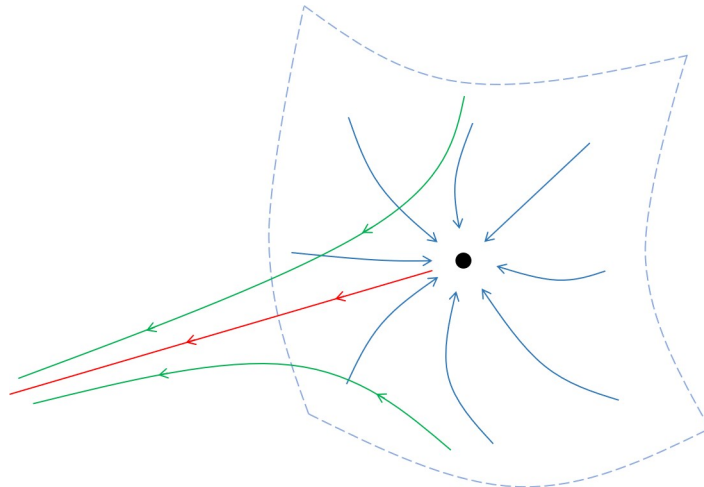

Figure 5: *RG flows near a fixed point*

How to probe the continuum limit using this? Technically, the RG flow only goes one way, from high energies to low energies, but let's imagine we can run it backwards, fixing a point in the IR and then increasing the cutoff as we go along. This is indeed the difficulty, we don't really know how to do this is practice, but we can always use our imagination. What we want to know is which theories give sensible answers as we send the energy scale to infinity.

The first example is the trivial one. If we sit at a fixed point, then RG flow does absolutely nothing to the theory, so obviously we can also go backwards and we stay in the same theory. Therefore, conformal field theories have a sensible continuum limit. Like I said, this is a very trivial example, they don't depend on the energy scale in the first place so nothing interesting happens.

Now imagine we sit at the critical surface, these are theories that have a fixed point in the IR. Then, as we flow to higher energies they get further away from the fixed point.

---

[51]In fact, one of the famous Clay Mathematics Institute's Millenium Prize Problems (worth 10^6 \$) is to prove Yang-Mills theories, i.e. non-Abelian gauge theories, have a continuum limit.

There are two additional possibilities: either they come from another UV fixed point and are really a renormalised trajectory with respect to that CFT; or they go all the way to infinity. In the latter case, we have a theory with infinite couplings in the high energy limit, which is obviously bonkers. These theories don't have a continuum limit.

Finally, we could have a theory that sits in the renormalised trajectory of some UV CFT. These theories have a well defined limit as we go to higher and higher energies, the asymptotic fixed point, hence, they make sense at arbitrarily high energy scales and they have a continuum limit. Usually, these theories are further classified by whether or not they are interacting, a classification we have alluded to in Chapter 3. If there are no interactions everywhere along the flow it is said to be trivial, if there are interactions along the flow but the asymptotic CFT is the Gaussian fixed point it is called asymptotically free, and if the CFT itself is also interacting we call it asymptotically safe.

Now we can truly see how the BPHZ classification can fool us. For instance, ϕ^4 theory in 4 dimensions has a marginally irrelevant operator, therefore, strictly speaking it appears to not have a continuum limit[52]. However the BPHZ theorem classified this theory as "nice", and indeed we were able to remove the cutoff dependence entirely. What is going on?

The reason for this is that the quartic coupling isn't irrelevant but merely marginally irrelevant. This means its divergence is very soft, in fact it is logarithmic. To see this let's recall the beta function (in the massless limit)

$$\beta_4 = \frac{3g_4^2}{16\pi^2} \tag{4.37}$$

which can be solved to yield

$$g_4(\Lambda) = -\frac{16\pi^2}{16\pi^2 C + 3\log\Lambda} \tag{4.38}$$

where C is some integration constant. Naively, we would have to parametrise this in terms of some value for g_4 at a particular scale. However, there is a natural scale which we can use to parametrise the flow, the scale at which the coupling diverges. This is given by

$$16\pi^2 C + 3\log\Lambda_{\phi^4} = 0 \tag{4.39}$$

substituting back into (4.38) gives

$$g_4(\Lambda) = \frac{16\pi^2}{3\log\left(\frac{\Lambda_{\phi^4}}{\Lambda}\right)} \tag{4.40}$$

This phenomenon when the coupling appears to diverge at a finite energy scale is called a Landau[53] pole. In our case, since this is a 1-loop computation we shouldn't really interpret this as true divergence but more as a signal that perturbation theory breaks down since the theory becomes strongly coupled. Also, the fact that we managed to trade a dimensionless coupling with an energy scale is known as *dimensional transmutation* and used to be a big mystery.

This gives us the clue we need to interpret the BPHZ result. The divergence is so incredibly soft that, if we start in the weak coupling regime, we need to probe energies exponentially higher than that to see any flow whatsoever. Therefore, when perturbation

---

[52]Of course this is a statement about a perturbative calculation up to 1-loop, so we haven't really proved a continuum limit doesn't exist, only that it is very unlikely.

[53]Lev Davidovich Landau: born in 1908 in Baku, Russian Empire; died in 1968 in Moscow, Soviet Union; Academic Advisor: Niels Bohr; Nobel Prize in Physics 1962 "for his pioneering theories for condensed matter, especially liquid helium"

theory is valid, which is also when the counterterm algorithm makes sense, this coupling barely runs at all and it will appear that we can confidently remove this cutoff. It is purely an artefact of the algorithm.

In the real world we have examples for all these kinds of flows. Fermi's theory for the weak interactions was non-renormalisable, which signalled the presence of the W and Z bosons. QCD, the theory of strong interactions, appears to be asymptotically free so it can indeed be valid at arbitrarily high energy scales. QED on the other hand behaves closer in spirit with ϕ^4 theory, it appears to not have a continuum limit, however, the energy scale at which it would break is $\Lambda_{\mathrm{QED}} \sim 10^{286}$ GeV, which is so ludicrously high it is completely unimportant phenomenologically. The standard model of particle physics which is our best theory to describe the fundamental interactions is renormalisable, however, it is still an open problem whether or not it has a continuum limit. Nowadays, it is generally believed it is but an effective theory, and people are working hard to search for signals of irrelevant operators which could signal new physics, all searches so far have not proved successful. And finally, general relativity is non-renormalisable so it naively appears to not have a continuum limit, there are many approaches to try to circumvent this, by having a discrete theory (e.g. loop quantum gravity), by having new physics at the Planck[54] scale (e.g. string theory), or by saying it actually comes from a highly non-trivial UV CFT (e.g. asymptotic safety). So far, there is no compelling experimental evidence to confirm or disprove any of these alternatives

4.3. Scheme (in)dependence

Throughout these lectures, we made a bunch of choices, namely, the precise regulator (hard momentum cutoff), and the renormalisation scheme (on-shell) with the promise that the overall physics did not depend on those choices. It is finally the time to address such concerns. Some of them are quite easy to deal with, so we'll start with that.

First, the renormalisation scheme. We chose a physical scheme such that the couplings would relate to actual experimentally measurable quantities. However, we could have chosen other schemes, that is, we could have fixed the counterterms in a different way. They are, after all, arbitrary. If we want to do old school subtracting infinities renormalisation, we would still need the infinite cutoff limit to make sense, but the finite part of the counterterm is always free to choose. What do different choices mean? Well, it's very easy to see they are essentially a reparametrisation of the RG trajectory with possibly a change of basis in theory space. That will clearly not change the overall shape of the flow, e.g. the existence of fixed points, or if we're flowing towards or away from a fixed point. So we can sleep quietly with regards to this one. The on-shell scheme is a good way to keep the physics more explicit, but, in practical calculations, other choices may be more convenient.

Secondly, imagine we are *not* taking the continuum limit, we are keeping the regulator there to define our theory. Then, physical answers must obviously depend on the precise nature of the regulator. For renormalisable theories and for low enough energies, that dependence will go away. However, for non-renormalisable theories, or if we're probing energies near the regulator scale, that dependence will be there and it is absolutely physical. We should take care to use a regulator that corresponds to the actual physical system we're studying.

Finally, imagine we are indeed dealing with a theory with a well defined continuum

---

[54]Karl Ernst Ludwig Marx Planck: born in 1858 in Kiel, Duchy of Holstein; died in 1947 in Göttingen, Germany; Doctoral Advisor: Alexander von Brill, Gustav Kirchhoff, Hermann von Helmholtz; Nobel Prize in Physics 1918 "in recognition of the services he rendered to the advancement of Physics by his discovery of energy quanta"

limit. Will the physics change if we use a different style of regulator? Or said another way, will the answer to a physical question depend on precisely how we take the limit? We would like the answer to those two questions to be a sound *no*, but we should check. And indeed we will prove that, with the added bonus of showcasing some possibly useful techniques along the way.

First of all, we need to be able to encode a more generic regularisation procedure. So, instead of explicitly turning off high momentum modes, we can simply dampen their contribution to the path integral. If we introduce sufficient damping, there will be no UV divergences to worry about. To that effect, we change the kinetic term to

$$S_{\Lambda_0}[\phi] = \int \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2} \tilde{\phi}(-\boldsymbol{p}) \frac{\boldsymbol{p}^2 + m^2}{K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right)} \tilde{\phi}(\boldsymbol{p}) + S_{\mathrm{int},\Lambda_0}[\phi] \tag{4.41}$$

where K is the function that will do the necessary damping. Ideally, we would demand the following properties:

1. $K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right)$ is a smooth non-increasing positive function of $\frac{\boldsymbol{p}^2}{\Lambda_0^2}$

2. $K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right) = 1$ for $\boldsymbol{p}^2 < \Lambda_0^2$

3. $K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right) \to 0$ sufficiently fast as $\frac{\boldsymbol{p}^2}{\Lambda_0^2} \to \infty$
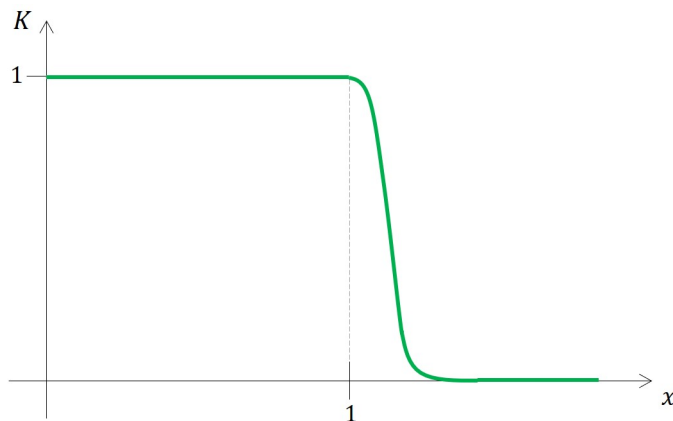

Figure 6: *Example of an appropriate $K(x)$*

We say ideally because some of them can be tweaked without loosing the main picture, for example, we can give up smoothness (using a hard momentum cutoff does so), or we could instead demand the function to be very close to 1 rather than exactly 1 for $\boldsymbol{p}^2 < \Lambda_0^2$. The third one however is completely non-negotiable, that's where the physics lies.[55]

Now using the identity

---

[55] This style of regulator captures every possible regulator that actually presents a definition for the path integral. However, some commonly used regulators don't really do that. Instead, they are fundamentally perturbative and only take care of making the Feynman integrals finite. They work well for renormalisable theories, if we're interested in the continuum limit (or for low enough energies), because, in that case, we won't care about the UV beta function, and then any renormalisation scheme will make the IR flow physical. An example of this is dimensional regularisation, which involves analytically continuing integrals to non-integer dimensions and then removing the poles as $d \to 4$. These regulators are often the more convenient choice since they can more easily preserve the symmetries of the theory, e.g. gauge invariance. However, they are far less rigorous (and far less physical) and I won't deal with them in these notes.

$$\int \mathcal{D}\phi_1 \, \mathcal{D}\phi_2 \, \exp\left[-\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\left[\frac{1}{A(\boldsymbol{p})}\tilde{\phi}_1(\boldsymbol{p})\tilde{\phi}_2(-\boldsymbol{p}) + \frac{1}{B(\boldsymbol{p})}\tilde{\phi}_2(\boldsymbol{p})\tilde{\phi}_2(-\boldsymbol{p})\right] + S_{\mathrm{int}}[\phi_1 + \phi_2]\right] =$$

$$= \int \mathcal{D}\phi \, \exp\left[-\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\frac{1}{A(\boldsymbol{p}) + B(\boldsymbol{p})}\tilde{\phi}(\boldsymbol{p})\tilde{\phi}(-\boldsymbol{p}) - S_{\mathrm{int}}[\phi]\right]\cdot$$

$$\cdot \int \mathcal{D}\phi' \, \exp\left[\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\left(\frac{1}{A(\boldsymbol{p})} + \frac{1}{B(\boldsymbol{p})}\right)\tilde{\phi}'(\boldsymbol{p})\tilde{\phi}'(-\boldsymbol{p})\right] \tag{4.42}$$

where

$$\tilde{\phi} = \tilde{\phi}_1 + \tilde{\phi}_2 \tag{4.43}$$

$$\tilde{\phi}' = -\frac{B}{A+B}\tilde{\phi}_1 + \frac{A}{A+B}\tilde{\phi}_2 \tag{4.44}$$

We can write (by appropriately choosing A and B, and neglecting the ϕ' integral since it only contributes with a field independent constant)

$$\int \mathcal{D}\phi \, \mathrm{e}^{-S_{\Lambda_0}[\phi]} = \int \mathcal{D}\phi^- \mathcal{D}\phi^+ \exp\left[-\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\boldsymbol{p}^2 + m^2}{K\left(\frac{\boldsymbol{p}}{\Lambda}\right)}\tilde{\phi}^-(\boldsymbol{p})\tilde{\phi}^-(-\boldsymbol{p}) - \right.$$

$$\left. -\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\boldsymbol{p}^2 + m^2}{K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right) - K\left(\frac{\boldsymbol{p}}{\Lambda}\right)}\tilde{\phi}^+(\boldsymbol{p})\tilde{\phi}^+(-\boldsymbol{p}) - S_{\mathrm{int},\Lambda_0}[\phi^+ + \phi^-]\right] \tag{4.45}$$

Then, using the shorthand notation $K(\boldsymbol{p}) \equiv K\left(\frac{\boldsymbol{p}}{\Lambda}\right)$ and $K_0(\boldsymbol{p}) \equiv K\left(\frac{\boldsymbol{p}}{\Lambda_0}\right)$, we define the interaction part of the effective action as,

$$\mathrm{e}^{-S_{\mathrm{int},\Lambda}[\phi^-]} = \int \mathcal{D}\phi^+ \exp\left[-\frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\boldsymbol{p}^2 + m^2}{K_0 - K}\tilde{\phi}^+(\boldsymbol{p})\tilde{\phi}^+(-\boldsymbol{p}) - S_{\mathrm{int},\Lambda_0}[\phi^+ + \phi^-]\right] \tag{4.46}$$

and therefore the full action,

$$S_\Lambda[\phi^-] = \frac{1}{2}\int \frac{\mathrm{d}^dp}{(2\pi)^d}\frac{\boldsymbol{p}^2 + m^2}{K}\tilde{\phi}^-(\boldsymbol{p})\tilde{\phi}^-(-\boldsymbol{p}) + S_{\mathrm{int},\Lambda}[\phi^-] \tag{4.47}$$

All this together then establishes how to perform the RG procedure with a smooth regulator. It perfectly agrees with our intuition about how RG should work. We managed to have a simple splitting between high and low energy modes. The low energy mode looks like the original one but with a different cutoff scale, and the high energy mode only has support in the interval $[\Lambda, \Lambda_0]$. Note however a few differences, in this time we are not changing the field configuration space, the transformations were mere shifts, all the physics is in the quadratic damping. Therefore, there is no implicit dependence on the scale in the spacetime integration or in the field definition. This doesn't mean we can't perform the second and third steps of RG, we can recover both scaling and the field anomalous dimension. However, they don't play any role in what follows so we will ignore them. Finally, we chose to regulate the full quadratic term rather than just the kinetic term, this simplifies our story when we go to perturbative Feynman diagrams calculations, however, it is just a choice, we could have chosen otherwise without affecting the main points.

Given the above considerations about what depends and doesn't depend on the cutoff, it's a straightforward exercise to derive the differential form of the RG equation (4.46). This is known as the Polchinski[56] equation:

$$\Lambda\frac{\partial}{\partial\Lambda}S_{\text{int},\Lambda}[\phi] = \int\frac{\mathrm{d}^dp}{(2\pi)^d}\frac{1}{2}\frac{\Delta\!\left(\frac{\boldsymbol{p}}{\Lambda}\right)}{\boldsymbol{p}^2+m^2}\left\{\frac{\delta S_{\text{int},\Lambda}}{\delta\tilde\phi(-\boldsymbol{p})}\frac{\delta S_{\text{int},\Lambda}}{\delta\tilde\phi(\boldsymbol{p})}-\frac{\delta^2 S_{\text{int},\Lambda}}{\delta\tilde\phi(-\boldsymbol{p})\,\delta\tilde\phi(\boldsymbol{p})}\right\} \tag{4.48}$$

where $\Delta\!\left(\frac{\boldsymbol{p}}{\Lambda}\right)=\Lambda\frac{\partial}{\partial\Lambda}K\!\left(\frac{\boldsymbol{p}}{\Lambda}\right)$.

This equation is entirely equivalent to (4.46), it's just the usual transition between integral and differential forms. This means we can calculate RG flows entirely based on the Polchinski equation, rather than trying to compute Feynman integrals. To a large extent, what we did in Section 3.3 was precisely that, just dressed in a slightly different language.

Now, to proceed, we need to be able to compare (connected) correlation functions computed with both actions. To that effect we define the Wilsonian effective action with a source term, i.e. the generating functional of connected correlation functions:

$$\mathrm{e}^{-W_\Lambda[J]} = \int\mathcal{D}\phi\,\exp\left[-S_\Lambda[\phi]+\int\frac{\mathrm{d}^dp}{(2\pi)^d}\tilde J(-\boldsymbol{p})\tilde\phi(\boldsymbol{p})\right] = \int\mathcal{D}\phi\,\mathrm{e}^{-S_\Lambda[\phi;J]} \tag{4.49}$$

Using (4.46), we can relate the Wilsonian effective actions calculated at different energies (we shall omit the argument of the various functions appearing for brevity and because we're only to write the quadratic piece explicitly, in which case one of the fields/sources is evaluated at $\boldsymbol{p}$ and the other at $-\boldsymbol{p}$).

$$\mathrm{e}^{-W_{\Lambda_0}[J]} = \int\mathcal{D}\phi\,\mathrm{e}^{-S_{\Lambda_0}[\phi;J]} =$$

$$= \int\mathcal{D}\phi\,\exp\left[-\int\frac{\mathrm{d}^dp}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K_0}\tilde\phi\tilde\phi-\tilde J\tilde\phi\right)-S_{\text{int},\Lambda_0}[\phi]\right] =$$

Doing the rotation as in (4.45), where the source term behaves as an interaction,

$$= \int\mathcal{D}\phi^-\mathcal{D}\phi^+\,\exp\left[-\int\frac{\mathrm{d}^dp}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K}\tilde\phi^-\tilde\phi^-+\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K_0-K}\tilde\phi^+\tilde\phi^+-\right.\right.$$
$$\left.\left.-\tilde J\!\left(\tilde\phi^-+\tilde\phi^+\right)\right)-S_{\text{int},\Lambda_0}[\phi^-+\phi^+]\right] =$$

Completing the square

$$= \int\mathcal{D}\phi^-\mathcal{D}\phi^+\,\exp\left[-\int\frac{\mathrm{d}^dp}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K}\tilde\phi^-\tilde\phi^-+\right.\right.$$
$$\left.+\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K_0-K}\left(\tilde\phi^+-\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde J\right)\left(\tilde\phi^+-\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde J\right)-\tilde J\tilde\phi^--\frac{1}{2}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde J\tilde J\right)-$$
$$\left.-S_{\text{int},\Lambda_0}[\phi^-+\phi^+]\right] =$$

[56]Joseph Gerard Polchinski Jr.: born in 1954 in White Plains, USA; died in 2018 in Santa Barbara, USA; Doctoral Advisor: Stanley Mandelstam

Defining $\tilde{\phi}' = \tilde{\phi}^+ - \frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}$

$$= \int \mathcal{D}\phi^- \mathcal{D}\phi' \exp\left[-\int \frac{\mathrm{d}^d p}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K}\tilde{\phi}^-\tilde{\phi}^- + \frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K_0-K}\tilde{\phi}'\tilde{\phi}' - \tilde{J}\tilde{\phi}^- - \right.\right.$$
$$\left.\left. - \frac{1}{2}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\tilde{J}\right) - S_{\mathrm{int},\Lambda_0}\left[\tilde{\phi}^- + \tilde{\phi}' + \frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\right]\right] =$$

Using (4.46) to perform the ϕ' integration

$$= \int \mathcal{D}\phi^- \exp\left[-\int \frac{\mathrm{d}^d p}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K}\tilde{\phi}^-\tilde{\phi}^- - \tilde{J}\tilde{\phi}^- - \frac{1}{2}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\tilde{J}\right) - \right.$$
$$\left. - S_{\mathrm{int},\Lambda}\left[\tilde{\phi}^- + \frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\right]\right]$$

Defining $\tilde{\phi}'' = \tilde{\phi}^- + \frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}$

$$= \int \mathcal{D}\phi'' \exp\left[-\int \frac{\mathrm{d}^d p}{(2\pi)^d}\left(\frac{1}{2}\frac{\boldsymbol{p}^2+m^2}{K}\tilde{\phi}''\tilde{\phi}'' - \frac{K_0}{K}\tilde{J}\tilde{\phi}'' + \frac{1}{2}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\frac{K_0}{K}\tilde{J}\tilde{J}\right) - S_{\mathrm{int},\Lambda}[\phi'']\right] =$$
$$= \int \mathcal{D}\phi'' \exp\left[-S_\Lambda[\phi''] + \int \frac{\mathrm{d}^d p}{(2\pi)^d}\frac{K_0}{K}\left(\tilde{J}\tilde{\phi}'' - \frac{1}{2}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\tilde{J}\right)\right] =$$
$$= \exp\left[-W_\Lambda\left[\frac{K_0}{K}\tilde{J}\right] - \frac{1}{2}\int \frac{\mathrm{d}^d p}{(2\pi)^d}\frac{K_0}{K}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}\tilde{J}\right] \tag{4.50}$$

After all this, we get our desired relation between the Wilsonian effective actions calculated in both theories:

$$W_{\Lambda_0}[J] = W_\Lambda\left[\frac{K_0}{K}\tilde{J}\right] + \frac{1}{2}\int \frac{\mathrm{d}^d p}{(2\pi)^d}\frac{K_0}{K}\frac{K_0-K}{\boldsymbol{p}^2+m^2}\tilde{J}(\boldsymbol{p})\tilde{J}(-\boldsymbol{p}) \tag{4.51}$$

From this, we can use (B.22) to compare connected correlation functions in both theories:

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2)\right\rangle_{\Lambda_0} = \frac{K_0^2}{K^2}\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2)\right\rangle_\Lambda - \frac{K_0-K}{\boldsymbol{p}_1^2+m^2}\frac{K_0}{K}\delta^{(d)}(\boldsymbol{p}_1+\boldsymbol{p}_2) \tag{4.52}$$

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n)\right\rangle_{\Lambda_0} = \left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n)\right\rangle_\Lambda \prod_{i=1}^{n}\frac{K_0(\boldsymbol{p}_i)}{K(\boldsymbol{p}_i)} \tag{4.53}$$

where $n > 2$, and in the first line K and K_0 can be evaluated at either $\boldsymbol{p}_1$ or $\boldsymbol{p}_2$, the distinction is irrelevant.

Note that, if we use a smooth regulator that doesn't vanish at a finite value of the momentum, these relations are perfectly invertible. We are not loosing any information. It seems like we could use correlation functions at the scale Λ to calculate correlation functions at the scale Λ_0, this would be true if were actually able to calculate the correlators exactly, however, since we only have access to perturbation theory it is not that straightforward to invert them.

We now have everything we need under our belts. We take the limit $\Lambda_0 \to \infty$, which means $K_0 \to 1$. Then imagine we deform K to $K + \delta K$, where $\delta K \ll 1$, and that we have an action S_Λ with that deformed regulator. Then (4.53) gives, to leading order in δK,

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle_\infty = \frac{1}{K^2} \left\langle \left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_1)\left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle_\Lambda +$$
$$+ \left(1 - \frac{1}{K} + \frac{\delta K}{K^2}\right)\frac{1}{\boldsymbol{p}_1^2 + m^2}\delta^{(d)}(\boldsymbol{p}_1 + \boldsymbol{p}_2) \tag{4.54}$$

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n) \right\rangle_\infty = \left\langle \left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_1)\cdots\left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_n) \right\rangle_\Lambda \prod_{i=1}^n \frac{1}{K(\boldsymbol{p}_i)} \tag{4.55}$$

From this action we can construct a new one by

$$S'_\Lambda = S_\Lambda + \int \frac{\mathrm{d}^d p}{(2\pi)^d}\left(\frac{\delta K}{K}\tilde{\phi}(\boldsymbol{p})\frac{\delta S_\Lambda}{\delta\tilde{\phi}(-\boldsymbol{p})} + \frac{1}{2}\frac{\delta K}{\boldsymbol{p}^2 + m^2}\left(\frac{\delta S_\Lambda}{\delta\tilde{\phi}(\boldsymbol{p})}\frac{\delta S_\Lambda}{\delta\tilde{\phi}(-\boldsymbol{p})} + \frac{\delta^2 S_\Lambda}{\delta\tilde{\phi}(-\boldsymbol{p})\,\delta\tilde{\phi}(\boldsymbol{p})}\right)\right)$$
$$\tag{4.56}$$

This new action will have the following correlation functions, to leading order in δK[57]:

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle'_\Lambda = \left\langle \left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_1)\left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle_\Lambda +$$
$$+ \frac{\delta K}{\boldsymbol{p}_1^2 + m^2}\delta^{(d)}(\boldsymbol{p}_1 + \boldsymbol{p}_2) \tag{4.57}$$

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n) \right\rangle'_\Lambda = \left\langle \left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_1)\cdots\left(1 - \frac{\delta K}{K}\right)\tilde{\phi}(\boldsymbol{p}_n) \right\rangle_\Lambda \tag{4.58}$$

Therefore

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle_\infty = \frac{1}{K^2}\left\langle \tilde{\phi}(\boldsymbol{p}_1)\tilde{\phi}(\boldsymbol{p}_2) \right\rangle'_\Lambda + \frac{1 - \frac{1}{K}}{\boldsymbol{p}_1^2 + m^2}\delta^{(d)}(\boldsymbol{p}_1 + \boldsymbol{p}_2) \tag{4.59}$$

$$\left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n) \right\rangle_\infty = \left\langle \tilde{\phi}(\boldsymbol{p}_1)\cdots\tilde{\phi}(\boldsymbol{p}_n) \right\rangle'_\Lambda \prod_{i=1}^n \frac{1}{K(\boldsymbol{p}_i)} \tag{4.60}$$

This result means S'_Λ satisfies the RG equation (4.46) (or equivalently the Polchinski equation) with regulator K and at the same time gives exactly the same continuum limit as S_Λ. Therefore regulating with K or $K + \delta K$ gives exactly the same continuum limit. Alas, the physics does not depend on our choice of regulator, as we wanted in the first place.

---

[57]To see this you should write the correlation function in terms of a path integral, expand the exponential to leading order in δK, then use $\frac{\delta S_\Lambda}{\delta\tilde{\phi}}\mathrm{e}^{-S_\Lambda} = -\frac{\delta}{\delta\tilde{\phi}}\mathrm{e}^{-S_\Lambda}$ and similarly for the second derivative, and finally integrate by parts wrt $\tilde{\phi}$. The 2-point function should be straightforward, the second term in (4.56) only gives a disconnected contribution to higher-point correlators which does not contribute to the connected correlation function

5. Historical Perspective and Concluding Remarks

During these lectures, we have done a path that is almost the exact opposite to the way renormalisation was unveiled historically. Physicists were quick to develop the basic framework of QFT after the discovery of quantum mechanics. Much of the underlying structure was known by the end of the 1930s, however, they were almost at the point of giving up on the whole endeavour! The reason being they were getting infinity for any calculation beyond tree-level. Dirac[58] even said in 1937 "Because of its extreme complexity, most physicists will be glad to see the end of QED".

Later, after the war, a new generation of physicists, such as Feynman, Schwinger[59], Tomonaga[60], and Dyson[61] took matters into their own hands and realised there was some structure behind the infinities. They eventually managed to develop a systematic procedure to sweep away the infinities under a very large rug, earning Feynman, Schwinger, and Tomonaga the 1965 Nobel[62] prize. In our language, they developed the idea of counterterms, and phrasing perturbation theory in terms of physical observables.

However, everyone was rather displeased with this procedure, Feynman, the one who first came up with it, famously called it a "dippy process". Further studies by Gell-Mann[63] and Low[64] in 1954 revealed some hints of a group action, and how one could define an effective electron charge at a certain energy scale. And later the Callan-Symanzik equations were written down, long before any interpretation of integrating out degrees of freedom was available.

Simultaneously, in the condensed matter community, there was a lot of confusion surrounding second order phase transitions. Namely the phenomena of universality, and anomalous dimensions were a huge mystery. Later on, Leo Kadanoff[65] introduced the idea of blocking, which is essentially RG in real space.

The final picture, which we presented in Chapter 3, only came along with Kenneth Wilson in the 1970s, it was him that unveiled the full picture of RG flows, theory space, and integrating out high energy modes. It was also him, together with Michael Fisher that first used the epsilon expansion[66]. The Polchinski equation and the exact RG formalism we presented in Sections 3.3 and 4.3 came along much later, in 1980s and 1990s.

This was a very complex history (of which I have only grazed the very surface). It took almost fifty years since the infinities first appeared to when they were completely

---

[58] Paul Adrien Maurice Dirac: born in 1902 in Bristol, England; died in 1984 in Tallahasee, USA; Doctoral Advisor: Ralph Fowler; Nobel Prize in Physics 1933 together with Erwin Schrödinger "for the discovery of new productive forms of atomic theory"

[59] Julian Seymour Schwinger: born in 1918 in New York City, USA; died in 1994 in Los Angeles, USA; Doctoral Advisor: Isidor Isaac Rabi; Nobel Prize in Physics 1965 together with Sin-Itiro Tomonaga and Richard P. Feynman "for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles"

[60] Sin-Itiro Tomonaga: born in 1906 in Tokyo, Japan; died in 1979 in Tokyo, Japan; Academic Advisor: Yoshio Nishina; Nobel Prize in Physics 1965 together with Julian Schwinger and Richard P. Feynman "for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles"

[61] Freeman John Dyson: born in 1923 in Crowthorne, England; Academic Advisor: Hans Bethe

[62] Alfred Bernhard Nobel: born in 1833 in Stockholm, Sweden; died in 1896 in Sanremo, Italy

[63] Murray Gell-Mann: born in 1929 in Manhattan, USA; died in 2019 in Santa Fe, USA; Doctoral Advisor: Victor Weisskopf; Nobel Prize in Physics 1969 "for his contributions and discoveries concerning the classification of elementary particles and their interactions"

[64] Francis Eugene Low: born in 1924 in New York City, USA; died in 2007 in Haverford, USA; Doctoral Advisor: Hans Bethe

[65] Leo Philip Kadanoff: born in 1937 in New York City, USA; died in 2015 in Chicago, USA; Doctoral Advisor: Paul Martin

[66] The paper in question has, arguably, the best title ever: "Critical Exponents in 3.99 Dimensions"

understood, and even longer before they started appearing in standard textbooks. And for a good reason! This is an incredibly subtle subject. Even now, when many more techniques are available, and we have a much better grasp on the wilderness of theory space, there is a lot of confusion surrounding renormalisation.

I hope these lectures have been useful to dispel some of that confusion. Using the most basic example, the one for which all these methods were originally designed to deal with, we can get a clear picture of what is going on. So let us summarise the broad structure of what has been covered:

- To define a quantum field theory we need to introduce some kind of regulator that has an intrinsic scale associated with it (e.g. the cutoff), otherwise we get non-sensical answers.

- This gives us a way to define an action whereby we change that scale to some lower scale by integrating out degrees of freedom, ending up with an effective description at lower energies. This is the Renormalisation Group.

- What RG boils down to is changing the couplings of our theory as we change the energy scale such that the observable low energy dynamics are unchanged. This gives us a flow in theory space that let's us ask (and possibly answer) questions about the structure of theories themselves and their relations

- Using counterterms, we can phrase all this in terms of physically measurable quantities. However, we're forced to introduce a second scale, the scale at which we do experiments. These two scales, the cutoff and the experimental, behave in very similar manners but are conceptually distinct. Most notably, the experimental scale is always observable, whereas the cutoff may not be.

- Finally, using all these tools, we can probe how sensible theories are to the UV data, and further, for which theories can we sensibly take the continuum limit. It is here the only place where infinities may appear.

But this is not the whole story! We have left out one of the most crucial parts of field theories, symmetries. The relationship between scale and symmetry is deep and fruitful. There is a beautiful story about how symmetries dictate which phase we're in, or which universality class. And this is relevant far away from condensed matter, particle physicists must also worry about this to be able to describe, for example, the quark-gluon plasma. There is also an intricate connection between renormalisation and anomalies, that is, symmetries present in the classical theory but not in the quantum one.

And finally, the deepest connections are with the two kinds of symmetries that aren't really symmetries: gauge redundancies and dualities. Gauge theories are the pinnacle of our understanding of physics, underlying both the standard model of particle physics and general relativity, but they are bizarre in the sense they are defined in terms of things we can't observe, even in principle. Gauge symmetries, should really be thought of as redundancies in the way we write the theory, not as physical symmetries. There is a lot to learn from the interplay between gauge invariance and RG flow, and a lot of it is still very much in the realm of ignorance.

Dualities are quite a different beast. These are symmetries not within a given theory but that relate seemingly different theories. And I do mean different, for example, the AdS/CFT correspondence is a duality between a strongly coupled quantum field theory and classical gravity! There are a lot of examples out there and it is a very active field of research to understand whether there's some deep mystery of nature or just a mathematical accident.

But in every example, scale and renormalisation play a crucial role in our understanding of Nature.

A. Notation and 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[67] (eV). They are related by

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

We'll denote Euclidean signature vectors with boldface, e.g. $\boldsymbol{x} = (x^1, x^2, x^3, \dots)$, and the spatial indices by Latin letters starting from the middle of the alphabet $i, j, k, \cdots = 1, 2, 3, \dots$. 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 $\mu, \nu, \cdots = 0, 1, 2, 3, \dots$. Any other kind of index such as internal indices will be denoted by Latin letters starting from the beginning of the alphabet. Einstein's[68] summation convention for repeated indices will be used unless explicitly stated.

For Fourier[69] transforms we use the following conventions:

$$\tilde{f}(k) = \int \mathrm{d}x \, f(x) \mathrm{e}^{-\mathrm{i}kx} \tag{A.2}$$

$$f(x) = \int \frac{\mathrm{d}k}{2\pi} \tilde{f}(k) \mathrm{e}^{\mathrm{i}kx} \tag{A.3}$$

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

$$\langle \boldsymbol{p} | \boldsymbol{p}' \rangle = (2\pi)^d \delta^{(d)}(\boldsymbol{p} - \boldsymbol{p}') \tag{A.4}$$

$$\int \frac{\mathrm{d}^d p}{(2\pi)^d} |\boldsymbol{p}\rangle\langle\boldsymbol{p}| = \mathbb{1} \tag{A.5}$$

$$\langle \boldsymbol{x} | \boldsymbol{p} \rangle = \mathrm{e}^{-\mathrm{i}(\boldsymbol{p} \cdot \boldsymbol{x})} \tag{A.6}$$

Further, throughout the text (but mostly in Chapter 2) we'll merrily swap between Heisenberg[70] 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.

$$\hat{O}(t) = \mathrm{e}^{\mathrm{i}\hat{H}t}\hat{O}(0)\mathrm{e}^{-\mathrm{i}\hat{H}t}, \quad \hat{O} \equiv \hat{O}(0) \tag{A.7}$$

$$|\psi, t\rangle = \mathrm{e}^{-\mathrm{i}\hat{H}t} |\psi, 0\rangle, \quad |\psi\rangle \equiv |\psi, 0\rangle \tag{A.8}$$

---

[67] Alessandro Giuseppe Antonio Anastasio Volta: born in 1754 in Como, Duchy of Milan; died in 1827 in Como, Lombardy-Venetia

[68] Albert Einstein: born in 1879 in Ulm, German Empire; died in 1955 in Princeton, USA; Doctoral Advisor: Alfred Kleiner; Nobel Prize in Physics 1921 "for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect"

[69] Jean-Baptiste Joseph Fourier: born in 1768 in Auxerre, Kingdom of France; died in 1830 in Paris, Kingdom of France; Academic Advisor: Joseph-Louis Lagrange

[70] Werner Karl Heisenberg: born in 1901 in Würzburg, German Empire; died in 1976 in Munich, West Germany; Doctoral Advisor: Arnold Sommerfeld; Nobel Prize in Physics 1932 "for the creation of quantum mechanics, the application of which has, inter alia, led to the discovery of the allotropic forms of hydrogen"

B. Path integrals in quantum field theory

This appendix is intended for those who are not very familiar with the path integral in the context of quantum field theory, especially how to write the path integral in terms of Feynman diagrams. We assume that we have Wick rotated[71] to Euclidean signature therefore $\boldsymbol{x} = (x^1, x^2, x^3, \ldots, x^d)$, where $x^d = \tau$ is the Euclidean time.

B.1. The partition function

In quantum field theory, the fundamental degrees of freedom are functions of both space and time, i.e. fields $\phi_a(\boldsymbol{x})$, where a is some internal index. We can then define actions for these fields and write the transition amplitude in terms of a path integral, as we did for ordinary quantum mechanics. However, we'll be most interested in amplitudes where $t_\mathrm{i} \to -\infty$ and $t_\mathrm{f} \to \infty$, therefore, inside the action, we integrate over the entirety of spacetime. This new transition amplitude is called *partition function* due to the connection with statistical field theory, and we'll write it as

$$Z = \int \mathcal{D}\phi_a \, \mathrm{e}^{-S[\phi_a]} \tag{B.1}$$

As we learned in Chapter 2, this is not a very well defined object if we allow the fields to vary continuously over the spacetime variables. We also saw we could solve this by discretising our spacetime, effectively making the number of degrees of freedom finite, and turning the functional integration into a finite number of ordinary integrations. Because of this, henceforth, I shall be lazy and use finite dimensional notation to perform the necessary integrations, at the end there will be a section highlighting the main results using continuum notation.[72]

This means fields will be denoted by $\boldsymbol{\phi} \in \mathbb{R}^n$, such that all complicated index structure and spacetime dependence is summarised into one index. I'll restrict to real fields since complex fields can be decomposed in terms of real fields[73], and moreover, as in the rest of the notes, I won't deal with fermions.

B.1.1 Free Theory

Let's start with the simplest case, a free theory, i.e. an action at most quadratic in the fields. Consider a positive definite symmetric matrix $\boldsymbol{M}$, and the following partition function

$$Z[\boldsymbol{M}] = \int \mathrm{d}^n\phi \, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\phi} \cdot \boldsymbol{M}\boldsymbol{\phi})} \tag{B.2}$$

To evaluate this integral note that, because $\boldsymbol{M}$ is symmetric, there is an orthogonal matrix $\boldsymbol{O}$ such that $\boldsymbol{O}^\mathrm{T}\boldsymbol{M}\boldsymbol{O} = \boldsymbol{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Then we make the change of variables $\boldsymbol{\phi}' = \boldsymbol{O}\boldsymbol{\phi}$, which leaves the measure invariant since $|\det \boldsymbol{O}| = 1$, to get

$$Z[\boldsymbol{M}] = \int \mathrm{d}^n\phi' \, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\phi}' \cdot \boldsymbol{D}\boldsymbol{\phi}')} = \prod_{i=1}^n \int \mathrm{d}\phi_i' \, \mathrm{e}^{-\frac{1}{2}\phi_i'^2 \lambda_i} = \prod_{i=1}^n \sqrt{\frac{2\pi}{\lambda_i}} = \sqrt{\frac{(2\pi)^n}{\det \boldsymbol{M}}} \tag{B.3}$$

---

[71]Refer to Chapter 2 for motivation and details

[72]This also has the side effect of making my manipulations more well defined in the eyes of a pure mathematician, but that's not the main motivation. Laziness is definitely a bigger factor.

[73]This is a bit of a lie since there are way more convenient ways to deal with complex fields than to reduce them to real fields, but it won't be very important for the main text

It is completely straightforward to generalise this result by adding a linear term, also called a *source term*,

$$Z[\boldsymbol{M}, \boldsymbol{J}] = \int \mathrm{d}^n \phi \, \mathrm{e}^{-\frac{1}{2}(\phi \cdot \boldsymbol{M}\phi) + (\phi \cdot \boldsymbol{J})} \tag{B.4}$$

We can then complete the square by performing the transformation $\phi' = \phi - \boldsymbol{M}^{-1}\boldsymbol{J}$, which is a simple shift, so no Jacobian[74] penalty,

$$Z[\boldsymbol{M}, \boldsymbol{J}] = \int \mathrm{d}^n \phi' \, \mathrm{e}^{-\frac{1}{2}(\phi' \cdot \boldsymbol{M}\phi') + \frac{1}{2}(\boldsymbol{J} \cdot \boldsymbol{M}^{-1}\boldsymbol{J})} = \mathrm{e}^{\frac{1}{2}(\boldsymbol{J} \cdot \boldsymbol{M}^{-1}\boldsymbol{J})} \sqrt{\frac{(2\pi)^n}{\det \boldsymbol{M}}} \tag{B.5}$$

This gives us a neat trick to calculate expectation values of functions of ϕ,

$$\langle f(\phi) \rangle = \frac{1}{Z[\boldsymbol{M}]} \int \mathrm{d}^n \phi \, f(\phi) \mathrm{e}^{-\frac{1}{2}(\phi \cdot \boldsymbol{M}\phi)} \tag{B.6}$$

Using the trick,

$$f(\phi) = f\left(\frac{\partial}{\partial \boldsymbol{J}}\right) \mathrm{e}^{(\phi \cdot \boldsymbol{J})}\bigg|_{\boldsymbol{J}=0} \tag{B.7}$$

we write

$$\langle f(\phi) \rangle = \frac{1}{Z[\boldsymbol{M}]} \int \mathrm{d}^n \phi \, f\left(\frac{\partial}{\partial \boldsymbol{J}}\right) \mathrm{e}^{-\frac{1}{2}(\phi \cdot \boldsymbol{M}\phi) + (\phi \cdot \boldsymbol{J})}\bigg|_{\boldsymbol{J}=0} = \frac{1}{Z[\boldsymbol{M}]} f\left(\frac{\partial}{\partial \boldsymbol{J}}\right) Z[\boldsymbol{M}, \boldsymbol{J}]\bigg|_{\boldsymbol{J}=0} =$$
$$= f\left(\frac{\partial}{\partial \boldsymbol{J}}\right) \mathrm{e}^{\frac{1}{2}(\boldsymbol{J} \cdot \boldsymbol{M}^{-1}\boldsymbol{J})}\bigg|_{\boldsymbol{J}=0} \tag{B.8}$$

This means we only need to know the partition function with a source term to get every possible expectation value. For that reason the partition function with a source term is often called the *generating functional*. To get a feel for this expression let's consider some examples. You should try to apply the formula (B.8) carrying all the derivatives yourself to make sure you understand where this all comes from,

- $\langle \phi_i \rangle = 0$

- $\langle \phi_i \phi_j \rangle = M_{ij}^{-1}$

- $\langle \phi_i \phi_j \phi_k \phi_l \rangle = M_{ij}^{-1} M_{kl}^{-1} + M_{ik}^{-1} M_{jl}^{-1} + M_{il}^{-1} M_{jk}^{-1}$

The latter two have a diagrammatic representation. We represent a factor of M_{ij}^{-1} by a line from i to j:

$$i \text{ ———— } j$$

Then to calculate a generic correlation function you add an external point for each field inserted and connect them, summing all the possible alternatives. For example, for the 4-point function above we can draw

[74]Carl Gustav Jacob Jacobi: born in 1804 in Potsdam, Kingdom of Prussia; died in 1851 in Berlin, Kingdom of Prussia; Doctoral Advisor: Enno Dirksen

60

It's not very hard to convince yourself that this procedure will work for all correlation functions. The diagrams we just introduced are the famous *Feynman diagrams*, and this result for the free theory is called *Wick's theorem*.

B.1.2 Interacting Theory

But these are boring Feynman diagrams, what we really want to see is what happens when we turn on interactions. Turning on interactions is equivalent to adding a potential term to the action,

$$Z[\boldsymbol{M}, \boldsymbol{J}, V] = \int \mathrm{d}^n \phi \, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\phi} \cdot \boldsymbol{M} \boldsymbol{\phi}) + (\boldsymbol{\phi} \cdot \boldsymbol{J}) - V(\boldsymbol{\phi})} \tag{B.9}$$

where we have included a source term both for generality, and because we know it may come in handy in the future. Also, we assume $V(\boldsymbol{0}) = 0$, $\left.\frac{\partial V}{\partial \phi_i}\right|_{\phi_i = 0} = 0$, and $\left.\frac{\partial^2 V}{\partial \phi_i \partial \phi_j}\right|_{\phi_i = 0} = 0$, which is tantamount to ignoring a field independent constant which doesn't affect the dynamics, absorbing the linear term into the source, and absorbing the quadratic term into $\boldsymbol{M}$. These assumptions could be dropped, but, in most setups, they are equivalent to expanding around the correct vacuum, which is a physical requirement.

For an arbitrary potential, we do not know how to calculate this integral.[75] What we can do is approximate, namely we approximate that V is, in some sense, a "small" perturbation. It is hard to be more specific than this without being given an actual expression for V, but, in general, there will be coefficients inside the function and those coefficients are the small parameters. The details will depend on the case but, assuming such a statement can be made, we can expand in powers of V. And, although we can do that straight away, we shall be clever about it.

First, use the trick in (B.7) to write

$$Z[\boldsymbol{M}, \boldsymbol{J}, V] = \mathrm{e}^{-V\left(\frac{\partial}{\partial \boldsymbol{J}}\right)} \int \mathrm{d}^n \phi \, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\phi} \cdot \boldsymbol{M} \boldsymbol{\phi}) + (\boldsymbol{\phi} \cdot \boldsymbol{J})} = Z[\boldsymbol{M}] \mathrm{e}^{-V\left(\frac{\partial}{\partial \boldsymbol{J}}\right)} \mathrm{e}^{\frac{1}{2}\left(\boldsymbol{J} \cdot \boldsymbol{M}^{-1} \boldsymbol{J}\right)} \tag{B.10}$$

The $Z[\boldsymbol{M}]$ factor is not important, it will disappear in actual correlation function calculations. The physics is in the two exponentials, we could expand them right away but that still is too complicated. We can be even more clever about it by using this identity.

Claim.

$$G\left(\frac{\partial}{\partial \boldsymbol{J}}\right) F(\boldsymbol{J}) = \left.F\left(\frac{\partial}{\partial \boldsymbol{\phi}}\right) G(\boldsymbol{\phi}) \mathrm{e}^{(\boldsymbol{\phi} \cdot \boldsymbol{J})}\right|_{\boldsymbol{\phi} = \boldsymbol{0}} \tag{B.11}$$

Proof. As always in physics, when in doubt, you Fourier transform and see what happens.[76] Let $\tilde{G}(\boldsymbol{k})$ and $\tilde{F}(\boldsymbol{k})$ denote the Fourier transforms of G and F respectively, then the LHS of (B.11) can be rewritten as

$$\int \frac{\mathrm{d}^n k_1}{(2\pi)^n} \tilde{G}(\boldsymbol{k}_1) \mathrm{e}^{\mathrm{i}\left(\frac{\partial}{\partial \boldsymbol{J}} \cdot \boldsymbol{k}_1\right)} \int \frac{\mathrm{d}^n k_2}{(2\pi)^n} \tilde{F}(\boldsymbol{k}_2) \mathrm{e}^{\mathrm{i}(\boldsymbol{J} \cdot \boldsymbol{k}_2)} =$$
$$= \int \frac{\mathrm{d}^n k_1}{(2\pi)^n} \frac{\mathrm{d}^n k_2}{(2\pi)^n} \tilde{G}(\boldsymbol{k}_1) \tilde{F}(\boldsymbol{k}_2) \mathrm{e}^{\mathrm{i}(\boldsymbol{J} \cdot \boldsymbol{k}_2)} \mathrm{e}^{-(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)} \tag{B.12}$$

---

[75] We could plug it into a computer but, as Eugene Wigner once said "It is nice to know that the computer understands the problem. But I would like to understand it too."

[76] If that fails resort to waving hands, works best in seminars.

Similarly, the RHS becomes

$$
\int \frac{\mathrm{d}^n k_2}{(2\pi)^n} \tilde{F}(\boldsymbol{k}_2) \mathrm{e}^{\left(\frac{\partial}{\partial \phi} \cdot \boldsymbol{k}_2\right)} \int \frac{\mathrm{d}^n k_1}{(2\pi)^n} \tilde{G}(\boldsymbol{k}_1) \mathrm{e}^{\mathrm{i}(\phi \cdot \boldsymbol{k}_1)} \mathrm{e}^{(\phi \cdot \boldsymbol{J})} \Big|_{\phi=\boldsymbol{0}} =
$$

$$
= \int \frac{\mathrm{d}^n k_1}{(2\pi)^n} \frac{\mathrm{d}^n k_2}{(2\pi)^n} \tilde{G}(\boldsymbol{k}_1) \tilde{F}(\boldsymbol{k}_2) \mathrm{e}^{(\phi \cdot (\mathrm{i}\boldsymbol{k}_1 + \boldsymbol{J}))} \mathrm{e}^{-(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)} \mathrm{e}^{\mathrm{i}(\boldsymbol{J} \cdot \boldsymbol{k}_2)} \Big|_{\phi=\boldsymbol{0}} =
$$

$$
= \int \frac{\mathrm{d}^n k_1}{(2\pi)^n} \frac{\mathrm{d}^n k_2}{(2\pi)^n} \tilde{G}(\boldsymbol{k}_1) \tilde{F}(\boldsymbol{k}_2) \mathrm{e}^{\mathrm{i}(\boldsymbol{J} \cdot \boldsymbol{k}_2)} \mathrm{e}^{-(\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)} \tag{B.13}
$$

Which is equal to (B.12), finishing out proof. $\qquad\square$

Now we apply this new result to (B.10) to get

$$
Z[\boldsymbol{M}, \boldsymbol{J}, V] = Z[\boldsymbol{M}] \, \mathrm{e}^{\frac{1}{2}\left(\frac{\partial}{\partial \phi} \cdot \boldsymbol{M}^{-1} \frac{\partial}{\partial \phi}\right)} \mathrm{e}^{-V(\phi) + (\boldsymbol{J} \cdot \phi)} \Big|_{\phi=\boldsymbol{0}} \tag{B.14}
$$

Finally we expand both exponentials to get a perturbative expansion. Let's first analyse the case $\boldsymbol{J} = \boldsymbol{0}$. You can think of this as just calculating the actual partition function. We'll use the notation,

$$
V_{i_1 \ldots i_k} = \frac{\partial}{\partial \phi_{i_1}} \cdots \frac{\partial}{\partial \phi_{i_k}} V(\phi) \Big|_{\phi=\boldsymbol{0}} \tag{B.15}
$$

The first few terms in the expansion are (Check!):

$$
\frac{Z[\boldsymbol{M}, V]}{Z[\boldsymbol{M}]} = 1 - \frac{1}{8} M_{ij}^{-1} M_{kl}^{-1} V_{ijkl} - \frac{1}{48} M_{ij}^{-1} M_{kl}^{-1} M_{mn}^{-1} V_{ijklmn} + \frac{1}{8} M_{ij}^{-1} M_{kl}^{-1} M_{mn}^{-1} V_{ijk} V_{lmn} +
$$

$$
+ \frac{1}{12} M_{ij}^{-1} M_{kl}^{-1} M_{mn}^{-1} V_{ikm} V_{jln} + \frac{1}{384} M_{ij}^{-1} M_{kl}^{-1} M_{mn}^{-1} M_{pq}^{-1} V_{ijklmnpq} + \ldots \tag{B.16}
$$

You might be worried about expanding in the propagator exponential as well as the potential since we made no assumption about propagator, only about the potential. For an arbitrary potential you'd be right, our procedure isn't very legitimate, however, most potentials we deal with have a finite (and small) number of terms. Therefore, eventually, the derivatives will simply vanish, terminating the series and rendering the expansion consistent. The reason why we only consider such simple potentials is explained in Chapter 3.

This is progress but we are not happy yet, we want to rephrase this in terms of Feynman diagrams, like we did for the free theory. We'll represent a factor of M_{ij}^{-1} by a line joining two points, the *propagator*

$$
i \underline{\hspace{3cm}} j
$$

Since the end result is a scalar, all indices must be contracted. From the first exponential the $\boldsymbol{M}$ indices are contracted with derivatives that will act on V. This means we have a second kind of term which joins up several lines through derivatives of the potential, we call it a *vertex*. It comes with a factor of $-V_{i_1 \ldots i_k}$ and we draw it as (the dots denote the other lines)
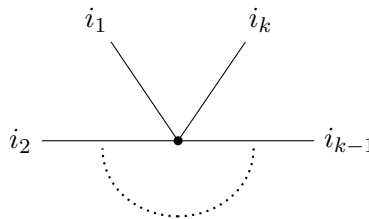
Hence, the terms in (B.16) become the following diagrams,

$$1 + \text{∞} + \text{⊛} + \text{⊖} + \text{⊖} + \text{⊛} + \ldots$$
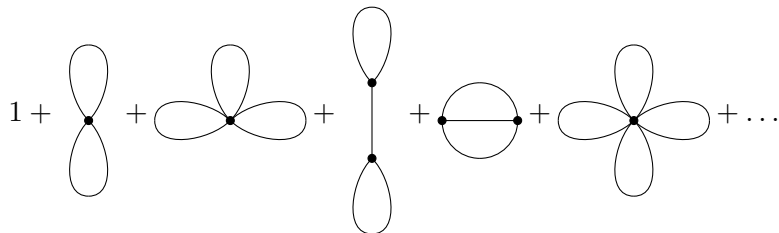
But we're still missing a piece, how to get those factors out front? For that we need a bit of group theory. Let's analyse a generic term in the double expansion with n powers of the potential and p propagators. Schematically, it looks like $\sim \frac{1}{2^p p!}(M^{-1})^p \partial^{2p} \frac{1}{n!} V^n$. Since $V(\mathbf{0}) = 0$, we'll need to act with the derivatives on all factors of V which will cancel the $1/n!$ term. Further, due to the Leibniz[77] rule, the derivatives will simply generate all possible terms with that number of Vs and that number of derivatives. Thinking in terms of diagrams, the action of the derivatives on the potential will generate all possible diagrams with n vertices and $2p$ lines, where we must label the lines and vertices uniquely. The propagators will then pair up these lines.

Now comes the complication, all these diagrams will be distinct if we take into account their labelling. However, some of them will be completely identical up to relabelling, we say they are *topologically* identical, and, since in the end the indices are contracted, the labelling doesn't 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 and are all identical. This means that, after contraction, we can freely swap two indices of the same propagator, which corresponds to flipping the propagator's ends in the diagram, or exchanging two propagators. The terms/diagrams that give the same contribution are the ones who are mapped to one another by this kind of action.

It's easy to see that 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! We're almost there. 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. Now let's prove this theorem[78]

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 $|O_x||S_x| = |G|$*

Proof. Let y be an arbitrary element of O_x, and let $S_{xy} = \{g \in G : gx = y\}$. Pick $h \in S_{xy}$ and define a map $\phi : S_x \to S_{xy}$ by $\phi : g \to hg$. The map $\psi : S_{xy} \to S_x$ defined by

---

[77]Gottfried Wilhelm (von) Leibniz: born in 1646 in Hanover, Holy Roman Empire; died in 1716 in Hanover, Holy Roman Empire; Doctoral Advisor: Bartholomäus Leonhard von Schwendendörffer; Academic Advisors: Erhard Weigel, Jakob Thomasius, Christiaan Huygens

[78]I present the proof mainly for completion and because it's really short. If you don't follow it don't worry, nothing else depends on this, or on similar reasoning, alternatively there's a very nice explanation on `https://gowers.wordpress.com/2011/11/09/group-actions-ii-the-orbit-stabilizer-theorem/` (which is where I found this statement of the proof)

$\psi : u \to h^{-1}u$ inverts ϕ, so $|S_x| = |S_{xy}|$ for every $y \in O_x$. But the sets S_{xy} with $y \in O_x$ form a partition of G. It follows that $|G| = |S_x||O_x|$. $\qquad\square$

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. Check you can reproduce the symmetry factors of the diagrams shown above.[79]

There's only one thing remaining, sources. But this is really easy, sources don't depend of ϕ hence exactly one derivative needs to act on it. Diagrammatically, we represent this by
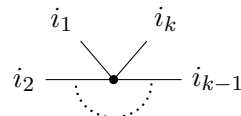
$$J_i \;\bullet\!\!\!\rule[2pt]{2cm}{0.4pt}$$

To calculate actual correlation functions you just need to use the trick in (B.7). This means you calculate the diagrams in the presence of sources, and then differentiate with respect to the sources. For example, for the 4-point function, you'll need to differentiate exactly 4 times wrt the sources, hence, you need to calculate all diagrams with exactly 4 sources. Usually, since we just want to correlation functions of simple powers of the fields, we skip the sources bit and add external points to our diagrams. For the 4-point function we'd add 4 external ϕ_i points.

To summarise, here's how to associate a QFT calculation with a diagram:

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 M_{ij}^{-1}: $\quad i \rule[2pt]{1.5cm}{0.4pt} j$

   - You join several internal lines in a *vertex*. A vertex with k lines carries a factor of $-V_{i_1\ldots i_k}(\mathbf{0})$:

   $$i_2 \rule[2pt]{1cm}{0.4pt}\!\!\!\!\overset{\displaystyle i_1 \qquad i_k}{\underset{\cdots}{\bullet}}\!\!\!\!\rule[2pt]{1cm}{0.4pt} i_{k-1}$$

   - Contract all indices

   - Divide by the symmetry factor, which is the size of the stabiliser, i.e. the number of actions you can perform on a labelled diagram that leave it completely unchanged (including the labelling), where you can either flip a propagator's ends or exchange it with another propagator.

B.2. The Wilsonian effective action

To calculate the partition function we needed to calculate all diagrams, even those who were disconnected. For example, even though we didn't draw it, the following diagram would also contribute:

---

[79]If you're having difficulty reproducing them, here's a hint: the factors shown (in order) are equal to: $\frac{1}{2^3}$, $\frac{1}{2^3 3!}$, $\frac{1}{2^3}$, $\frac{1}{2\cdot 3!}$, $\frac{1}{2^4 4!}$
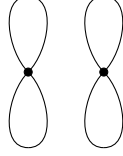
This is clearly not very efficient since it is just two copies of the same diagram. It would be really nice if we had a way of just considering connected diagrams. It turns out we do, it's the *Wilsonian effective action*:

$$W[\boldsymbol{J}] = -\log Z[\boldsymbol{J}] \tag{B.17}$$

where from now on we define $Z[\boldsymbol{J}] \equiv Z[\boldsymbol{M}, \boldsymbol{J}, V]$ and $Z_0 \equiv Z[\boldsymbol{M}]$ for convenience.

This object is related to the free energy of statistical mechanics, and will be very useful in the main text. But for now, let's see how to phrase it in terms of connected Feynman diagrams. 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

$$S_\Gamma = \prod_a n_a! S_a^{n_a} \tag{B.18}$$

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

$$F(\Gamma) = \prod_a F(\Gamma_a)^{n_a} \tag{B.19}$$

Then we can write the partition function as

$$\frac{Z[\boldsymbol{J}]}{Z_0} = \sum_{\Gamma \in \text{disconnected}} \frac{F(\Gamma)}{S_\Gamma} = \sum_{\{n_a\}} \prod_a \frac{F(\Gamma_a)^{n_a}}{n_a! S_a^{n_a}} = \prod_a \sum_{n_a=0}^{\infty} \frac{1}{n_a!} \left(\frac{F(\Gamma_a)}{S_a} \right)^{n_a} =$$

$$= \prod_a \exp\left(\frac{F(\Gamma_a)}{S_a} \right) = \exp\left(\sum_a \frac{F(\Gamma_a)}{S_a} \right) = \exp\left(\sum_{\Gamma \in \text{connected}} \frac{F(\Gamma)}{S_\Gamma} \right) \tag{B.20}$$

Therefore

$$W[\boldsymbol{J}] = W_0 - \sum_{\Gamma \in \text{connected}} \frac{F(\Gamma)}{S_\Gamma} \tag{B.21}$$

where $W_0 = -\log Z_0$.

As promised the Wilsonian effective action can be built out of connected Feynman diagrams. Consequently, we can use this to calculate *connected* n-point correlation functions as

$$\langle \phi_{i_1} \dots \phi_{i_n} \rangle_{\text{connected}} = -\frac{\partial}{\partial J_{i_1}} \dots \frac{\partial}{\partial J_{i_n}} W[\boldsymbol{J}] \bigg|_{\boldsymbol{J}=\boldsymbol{0}} \tag{B.22}$$

65

for this reason the Wilsonian effective action is also called the generating functional for correlation functions.

There is one further generating functional we could build, which would generate 1PI correlation functions. We won't need it for the main text hence I won't cover it. Just for reference it is called the *quantum effective action* and is related to the Gibbs[80] free energy of statistical mechanics.

B.3. An example: ϕ^4 theory

Just to illustrate how this all works in practice, we'll consider the following very simple theory: ϕ^4 with a single real scalar field in d dimensions. The action for this theory is

$$S[\phi] = \int \mathrm{d}^d x \left(\frac{1}{2} (\boldsymbol{\nabla}\phi)^2 + \frac{1}{2} m^2 \phi^2 + \frac{\lambda}{4!} \phi^4 \right) \tag{B.23}$$

The partition function with a source term is

$$Z[J] = \int \mathcal{D}\phi \ \mathrm{e}^{-S[\phi] + \int \mathrm{d}^d x J(\boldsymbol{x})\phi(\boldsymbol{x})} \tag{B.24}$$

Correlation functions can be calculated via

$$\langle \phi(\boldsymbol{x}_1) \dots \phi(\boldsymbol{x}_n) \rangle = \frac{\delta}{\delta J(\boldsymbol{x}_1)} \cdots \frac{\delta}{\delta J(\boldsymbol{x}_n)} Z[J] \tag{B.25}$$

where $\frac{\delta}{\delta J}$ denotes the functional derivative defined by the essential rule

$$\frac{\delta J(\boldsymbol{x})}{\delta J(\boldsymbol{y})} = \delta^{(d)}(\boldsymbol{x} - \boldsymbol{y}) \tag{B.26}$$

The quadratic term can be integrated by parts to give

$$\int \mathrm{d}^d x \frac{1}{2} \left((\boldsymbol{\nabla}\phi)^2 + m^2 \phi^2 \right) = \int \mathrm{d}^d x \frac{1}{2} \phi(-\nabla^2 + m^2)\phi \tag{B.27}$$

In the continuum, instead of a matrix we have an operator, which you can think of as an infinite-dimensional matrix. It's inverse is the Green's function,

$$(-\nabla^2 + m^2)\Delta_{\mathrm{F}}(\boldsymbol{x}, \boldsymbol{y}) = \delta^{(d)}(\boldsymbol{x} - \boldsymbol{y}) \tag{B.28}$$

which we can solve easily by taking a Fourier transform,

$$(\boldsymbol{p}^2 + m^2)\tilde{\Delta}_{\mathrm{F}}(\boldsymbol{p}) = 1 \Leftrightarrow \tilde{\Delta}_{\mathrm{F}}(\boldsymbol{p}) = \frac{1}{\boldsymbol{p}^2 + m^2} \tag{B.29}$$

Also, the potential is just $V(\phi) = \frac{\lambda}{4!}\phi^4$, hence all it's derivatives will vanish except for the 4th one. This means our Feynman rules will be[81],

- Propagator:

$$\boldsymbol{x} \ \text{———} \ \boldsymbol{y} \ = \Delta_{\mathrm{F}}(\boldsymbol{x} - \boldsymbol{y}) \tag{B.30}$$

---

[80]Josiah Willard Gibbs: born in 1839 in New Haven, USA; died in 1903 in New Haven, USA; Doctoral Advisor: Hubert Anson Newton

[81]Usually scalars are represented with dotted lines, however, since we will only be dealing with scalars we can represent them by full lines with no chance of confusion with fermions

- Vertex:

$$\times \quad = -\lambda \qquad (B.31)$$

- Contracting indices means integrating over the position of internal vertices

Alternatively we can write this in momentum space, the action becomes,

$$S[\tilde{\phi}] = \int \frac{\mathrm{d}^d p}{(2\pi)^d} \frac{1}{2} \tilde{\phi}(-\boldsymbol{p})(\boldsymbol{p}^2 + m^2)\tilde{\phi}(\boldsymbol{p}) + \frac{\lambda}{4!} \int \prod_{i=1}^{4} \left(\frac{\mathrm{d}^d p_i}{(2\pi)^d} \tilde{\phi}(\boldsymbol{p}_i) \right) \delta^{(d)} \left(\sum_{i=1}^{4} \boldsymbol{p}_i \right) \qquad (B.32)$$

The structure of the delta functions means we should impose momentum conservation along a line and at each vertex. The Feynman rules then become:

- Propagator:

$$\xrightarrow{\boldsymbol{p}} \quad = \frac{1}{\boldsymbol{p}^2 + m^2} \qquad (B.33)$$

- Vertex:

$$\times \quad = -\lambda \qquad (B.34)$$

- After imposing conservation of momenta at each vertex there may be undetermined momenta in a closed loop. You should integrate over that momenta.

- External sources in momentum space are usually denoted not by points but by external lines with a momentum label, there should be no propagator associated to those lines.